Semiconductor Device Characterization and Modelling for Effective Design of 5G Front-Ends

by

Ahmed Raslan

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	David Root Research Fellow, Keysight Technologies
Supervisor:	Slim Boumaiza Professor, Department of Electrical and Computer Engineering, University of Waterloo
Internal Member:	Safieddin Safavi-Naeini Professor, Department of Electrical and Computer Engineering, University of Waterloo
Internal Member:	Ajoy Opal Professor, Department of Electrical and Computer Engineering, University of Waterloo
Internal-External Member:	James Martin Associate Professor, Department of Physics and Astronomy, University of Waterloo

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.

Ahmed Raslan

Abstract

The realization of fifth-generation mobile communication will require transceiver front-ends that can handle data-rates of multi-gigabits per second. To provide such data-rates, bandwidth in the hundreds of megahertz will need to be available; hence, moving to millimeter wave frequencies is a must. Utilizing millimeter waves will provide a very high level of integration to the transceiver system, and concepts such as massive multiple-input-multiple-output could be exploited to enable higher data capacities. In order to design such a front-end, accurate simulation tools for both the system and circuit levels are needed. The first is needed because a high-level of integration prevents any opportunity for post-fabrication tweaking; accurate simulation will optimize a first-pass design. The second is essential to preserve the overall efficiency of the system; configurations incorporating a large number of power amplifiers (PAs) will require new design methodologies to enable linear and efficient operation. Unfortunately, existing transistor models cannot be used for these simulations as they cannot predict the device behaviour under real-life operating conditions (i.e., under modulated signal stimulus).

The objective of this thesis is to propose a transistor model that can be used in circuit simulators under any stimulus, including modulated signals. This means that the proposed model will need to be accurate both globally and locally. While the former property relates to the compact modelling approach, the latter relates to behavioural modelling. Thus, the proposed model will bridge the gap between the two modelling approaches.

This thesis starts by studying existing compact and behavioural modelling techniques for radio frequency power transistors. These techniques will be grouped based on the common properties imposed by the model structure (i.e., model formulation and extraction measurements) to help with problem identification. Based on this study, high-order network parameters (HONPs) are proposed. HONPs represent a new set of Volterra-based network parameters that are capable of completely describing weakly-nonlinear (WNL) behaviours of the transistor, and hence guaranteeing the local accuracy of the model. It will be proven, in both measurement and simulation environments, that these parameters exhibit the same properties of linear network parameters represented in power independency, and solution continuity and uniqueness. Therefore, HONPs represent a true extension of linear network parameters. Furthermore, HONPs will be extracted using continuous-wave (CW) nonlinear vector network analyzer (NVNA) measurements, and their ability to predict the device WNL behaviours will be demonstrated using wideband multi-tone stimuli in both simulation and measurement environments.

Next, a novel compact model that utilizes these parameters as a building block will be proposed. New model formulations and extraction techniques will be presented. Model im-

plementation will be performed in two parts. First, a small-signal model will be generated using a new layered-neural network (NN) technique that allows testing different topologies for the extrinsic and intrinsic shells, in an automated manner. Second, an intrinsic HONPs-based large-signal model will be constructed using a novel NN technique that allows optimizing a nonlinear function to fit first and high-order derivatives simultaneously.

Finally, the proposed model will be tested under different circuit simulators (direct-current, small-signal, harmonic balance and envelope simulators). Modulated signal validations will be shown in the simulation environment. Also, a two-tone measurement validation using a class-AB PA will be presented.

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to my supervisor, Professor Slim Boumaiza, for his never-ending help and support throughout my research. I would also like to thank him for introducing me to this field and giving me the chance to be involved in different research projects.

I have also had the honour to be part of an amazing research group. A special thanks to Marwen Ben Rejeb for all the engaging technical and non-technical discussions, and also for reviewing Chapter 3 of my thesis and providing me with insightful comments. I would also like to thank Yushi Hu for helping me applying the proposed techniques on his power amplifier by providing the schematics. Finally, I would like to thank Chad Gillease from Keysight for providing me with commands that enabled implementing my proposed model on the PNA-X.

Last but not least, I would like to show my gratitude to my family and friends. Their unconditional love and support over the years have made this journey possible.

Dedication

To my family and friends

Table of Contents

List of Tables

List of Figures

Nomenclature

List of Abbreviations

5G	5^{th} Generation Mobile Networks
A-Wave	Incident Power-Wave
ACPR	Adjacent Channel Power Ratio
ADS	Keysight's Advanced Design System Circuit Simulator
B-Wave	Reflected Power-Wave
BW	Bandwidth
CAD	Computer-Aided Design
CW	Continuous-Wave
DAC	Data Access Component
DC	Direct Current
DIDO	Dual-Input-Dual-Output
DPD	Digital Predistortion
DUT	Device Under Test
DynaFET	Dynamic FET
EVM	Error Vector Magnitude
GaN	Gallium Nitride

HB	Harmonic Balance
HEMT	High Electron Mobility Transistor
HONPs	High-Order Network Parameters
HOSPs	High-Order Network Parameters in the S Form
HOYPs	High-Order Network Parameters in the Y Form
IMD	Intermodulation Distortion
IoT	Internet of Things
LSM	Large-Signal Model
LSOP	Large-Signal Operating Point
M2M	Machine to Machine
MIMO	Multiple-Input-Multiple-Output
mm-waves	Millimetre Waves
MSE	Mean Squared Error
NMSE	Normalized Mean-Square-Error
NN	Neural Network
NVNA	Nonlinear Vector Network Analyzer
PAPR	Peak to Average Power Ratio
PA	Power Amplifier
RF	Radio Frequency
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
SSD	Symbolically Defined Device
VNA	Vector Network Analyzer

WCDMA Wideband Code Division Multiple Access

WNL Weakly-Nonlinear

Chapter 1

Introduction

1.1 Motivation

Fifth generation (5G) mobile communication has been attracting a lot of attention for the past few years. Although it has not yet been standardized, there is a global consensus that it should provide data-rates of multi-gigabits per second in order to support concepts like machine to machine (M2M) communication and the "Internet of Things" (IoT) [1]. This data-rate requirement cannot be achieved without moving to millimetre wave (mm-waves) frequencies for two reasons. First, operating at mm-waves enables bandwidth (BW) in the hundreds of megahertz [2]. Second, since the wave-length becomes very small, concepts such as massive multiple-input-multiple-output (MIMO) can enable higher data capacities [3, 4]. While there is ongoing research targeting implementation of the 5G standard that examines the physical layer up to the application layer, this thesis targets the transistor model - an essential element for the design of the front-end.

Computer-aided design (CAD) tools are critical to achieving an optimized first-pass design that reduces time to market, yet moving to 5G will be challenging to the transistor models used by these tools. This is because the high level of integration required for 5G front-ends will demand that accurate system-level simulation be incorporated into the design process as post-fabrication tweaking will no longer be an option. Unfortunately, existing transistor models cannot be used in such simulations because they are not trustworthy when simulating real-life operating conditions, i.e., when simulated using modulated signals. More importantly, achieving power efficiency for the whole system will be challenging because of the large number of power amplifiers (PAs) that will be included. This challenge is based upon the traditional methodology used to design PAs, which con-

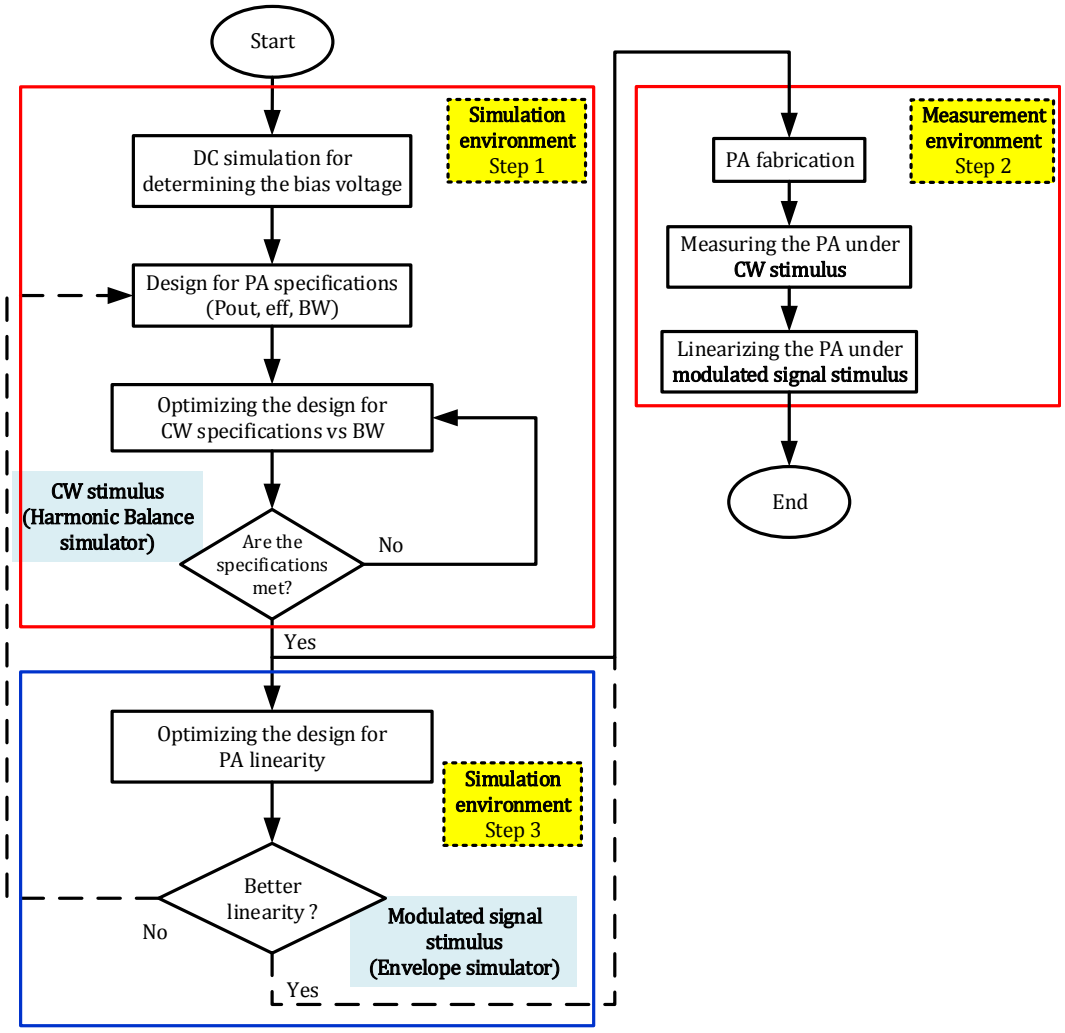


Figure 1.1: PAs design methodology flowchart. The red boxes (steps 1 and 2) illustrates the conventional design methodology that aims to optimize PAs efficiency. The blue box illustrates the extra step that needs to be added for simultaneous efficiency and linearity optimization.

sists of steps 1 and 2 in Figure 1.1. Historically, there has always been a tradeoff between efficiency and linearity, i.e., the more efficient the PA, the more nonlinear it becomes and vice versa [5]. Since the PA is the most power-hungry block in the transmitter, the conventional approach has been to optimize the PA to meet the efficiency specification. The linearity specification, on the other hand, is addressed by another block called the linearizer

which consumes extra power. While this additional power consumption is acceptable in current mobile devices and base stations, the move to massive MIMO schemes will make the linearizing blocks the dominant source of power consumption (each PA would require a separate linearizer), and degrade the overall efficiency of the system. Therefore, an extra step that allows concurrent design for both efficiency and linearity has to be added (step 3 in Figure 1.1). Again, existing transistor models cannot be used for such a design approach as they cannot simulate operation under modulated signals.

1.2 Problem Statement

Existing transistor models are incapable of predicting the behaviour under modulated signal stimuli. As a demonstration of this statement, a 20 W Doherty PA operating from 1.72 GHz to 2.27 GHz [6] [1] was used. The PA was designed using 10 W and 25 W gallium nitride (GaN) high electron mobility transistor (HEMT) packaged devices from Cree to meet the efficiency and output power requirements over the desired BW. A continuous-wave (CW) stimulus was used for designing and measuring the PA, using steps 1 and 2 in Figure 1.1, and good agreement between simulations and measurements was achieved. A two-step test was performed as follows. The first step was to examine the transistor model's ability to predict the spectrum by exciting the PA in both simulation and measurement environments with a 5 MHz wideband code division multiple access (WCDMA) signal generated around a 2 GHz center frequency and reporting the output. The results are shown in Figure 1.2. The second step was to examine the model's ability to cancel the nonlinearity exhibited by the fabricated PA which requires an accurate prediction of both the amplitude and phase of the spectrum. This was achieved by using the measured output signal from the first step to linearize the fabricated PA [2] and generate a predistorted signal. The first step was then repeated using this predistorted signal as input (Figure 1.3).

It is obvious that there is a good agreement in the fundamental band between the measured and simulated outputs for both cases, as can be seen in the zoomed-in portion of the spectrum in Figure 1.3 (b). On the other hand, the third and fifth intermodulation bands (from ± 2.5 MHz to ± 12.5 MHz) show clear disagreement (especially in the second case where the measured PA output is linearized, and the simulated output is not). These observations lead to the following conclusions. First, the transistor model did a great job at predicting the device's behaviour in the fundamental band - the equivalent of simulating it under CW stimulus - as the linear response is most dominant in this band. Second, since

[1] This PA was designed by one of the EmRG [7] members, so the design schematic is accessible.
[2] Volterra-based digital predistortion (DPD) was used to linearize the PA.

the passive circuitry around the PA has nothing to do with generating intermodulation distortion (IMD) products, it cannot be blamed for the disagreement in the intermodulation bands. In other words, the transistor model failed to predict the device performance under modulated signal stimuli. Therefore, it cannot be used for accurate system-level simulations or efficient/linear PA design.

This thesis bridges the gap between transistor compact models and black-box behavioural models by presenting a novel transistor model that is accurate both globally and locally. While the first property allows using the model to design circuits, the later allows predicting the behaviour of the transistor under modulated signal stimuli. The proposed model utilizes the high-order network parameters (HONPs), also proposed in this work, as the model's building block. These parameters enable describing the device -at any biasing point- by a weakly nonlinear (WNL) circuit and hence provide accurate calculations of the current/voltage (I/V) derivatives. Therefore, the model should be able to predict any WNL behaviour, including the behaviour under modulated signal stimuli. The proposed model shows excellent ability in predicting the transistor's large-signal behaviour under modulated signal stimuli in the simulation environment, and the behaviour under multi-tone stimuli in the measurement environment.

1.3 Thesis Outline

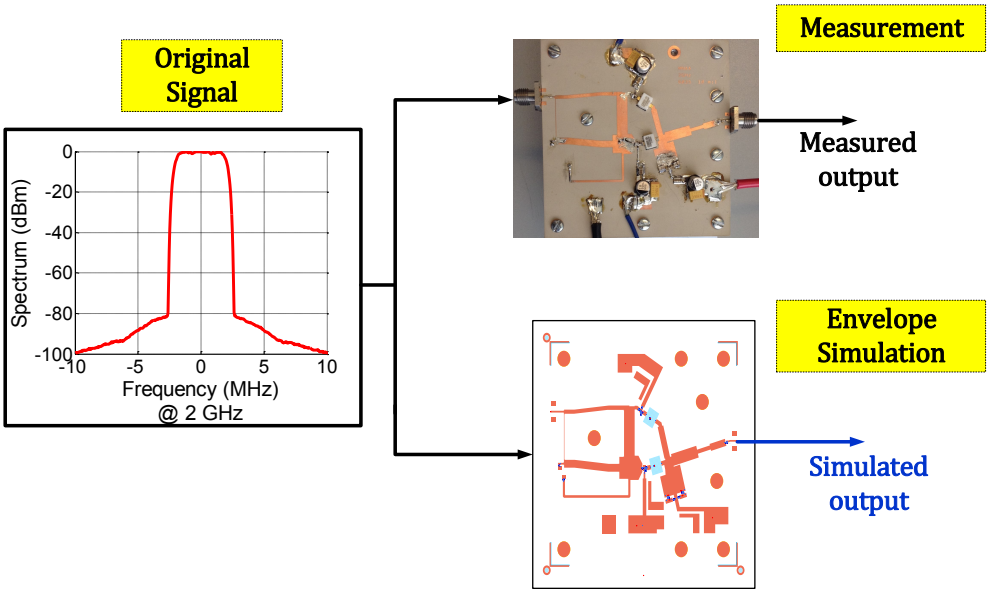
This thesis is organized as follows. Chapter 2 presents a general overview of nonlinear modelling approaches that exist in the literature. Linear measurement-based compact models are explained, and then state-of-the-art nonlinear measurement-based models are described. Furthermore, some of the most successful nonlinear behavioural modelling approaches will be investigated (in order to bring the local accuracy property to transistor models).

Chapter 3 proposes a new set of parameters (HONPs) that can completely describe WLN behaviours of nonlinear systems. HONPs formulations (in both Y and S forms), properties and extraction techniques are presented. Also, HONPs ability to predict WNL behaviours due to wideband multi-tone signal excitations is demonstrated in both simulation and measurement environments.

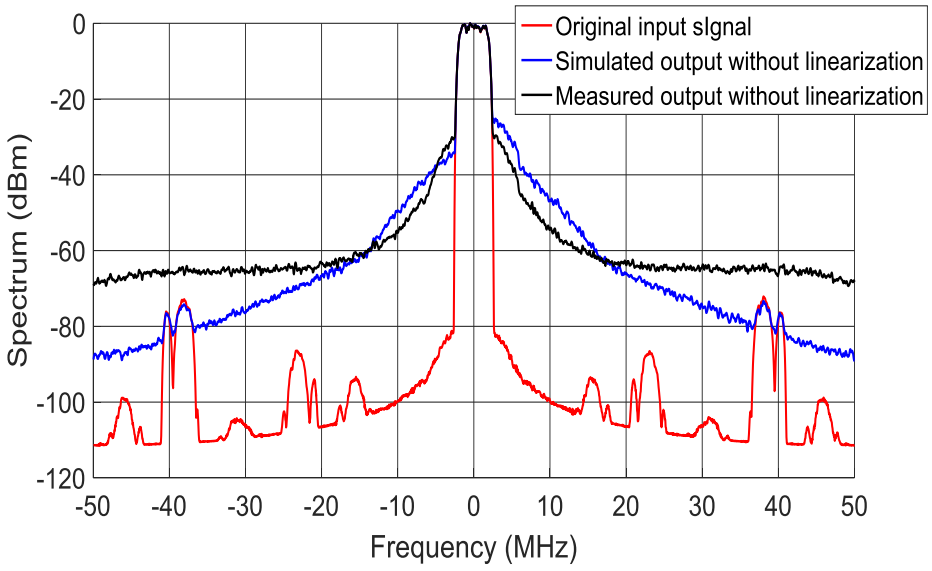
Chapter 4 presents a novel technique to build technology-independent transistor small-signal-models. A Layered-neural networks (NNs) technique that allows trying different topologies for extrinsic and intrinsic shells is explained. Also, model implementation and validation in Keysight ADS software are demonstrated.

Chapter 5 extends HONPs concept, proposed in Chapter 3, to handle strong nonlinear behaviours through a novel HONPs-based transistor compact model. A new NN technique that allows constructing nonlinear functions from their derivatives is presented. Furthermore, the proposed model implementation in Keysight ADS software and validations under different circuit simulators (DC, S-parameters, Harmonic Balance, and Envelope simulators) are discussed.

The contribution of this work to the field is discussed in Chapter 6. Besides, possible directions for future research work on this topic are discussed.

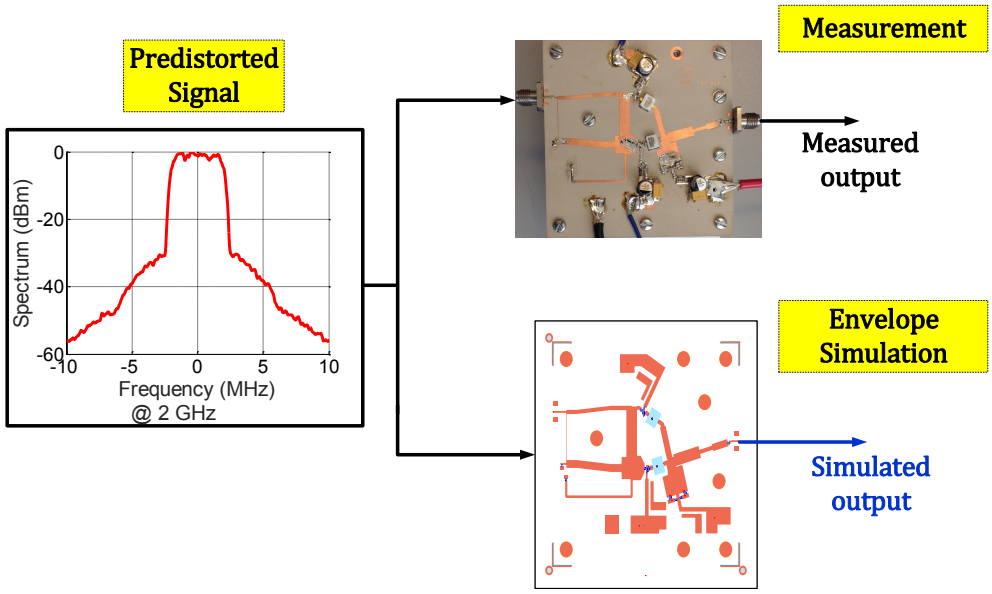


(a)

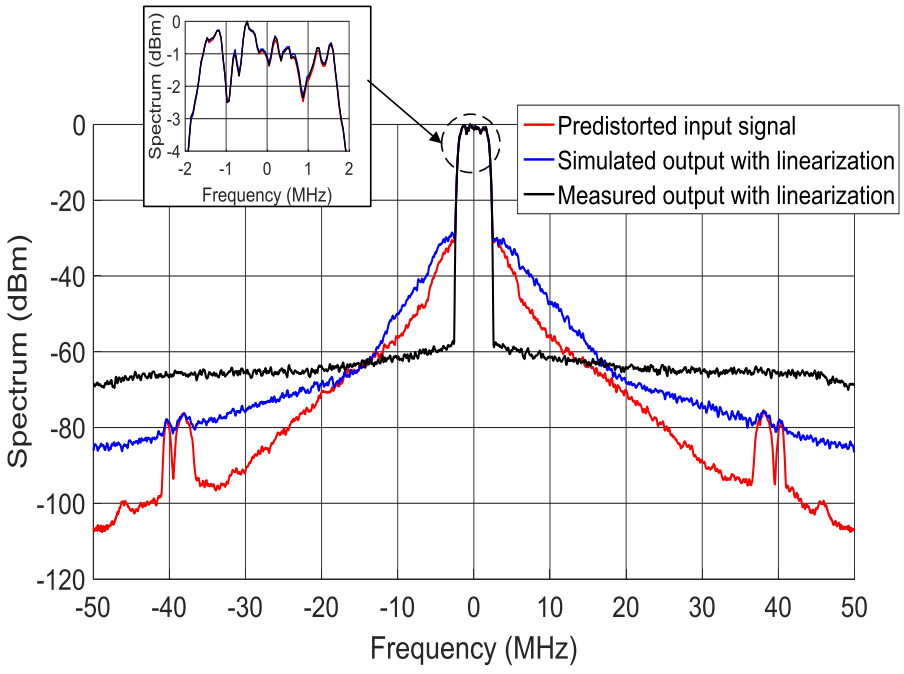


(b)

Figure 1.2: (a) Stimulating the PA in both measurement and simulation environments with a 5 MHz WCDMA signal around 2 GHz, and (b) the spectrum of the input signal and both outputs.



(a)



(b)

Figure 1.3: (a) Stimulating the PA in both measurement and simulation environments with a predistorted 5 MHz WCDMA signal around 2 GHz, and (b) the spectrum of the input signal and both outputs.

Chapter 2

Nonlinear Modelling Techniques

2.1 Introduction

Device modelling approaches can be seen as occupying different places on a continuum, as shown in Figure 2.1, the extremes being physics-based modelling (also known as white-box modelling) and behavioural modelling (also known as black-box modelling). Physics-based models assume the availability of all the detailed information (geometry, dimensions, material properties, doping profile, band structure, etc.) of the device being modelled [8]. It then discretizes the transistor geometry and finds the solution to a set of partial differential equations at each cell. These equations are formulated based on Maxwell's equation in the case of modelling passive structures (i.e., the extrinsic shell of the transistor), and quantum-mechanical equations to describe the intrinsic part of the transistor [9][1]. This approach of modelling is used on the device-level by semiconductor foundries. Although it is the most general one, it requires a substantial amount of computational power for simulation which makes it inefficient to use in circuit-level simulations.

Behavioural models, on the other hand, are constructed based on observations of the input-output signal mapping of the device. This mapping can be as simple as using one function to model a static single-input-single-output (SISO) system or using two nonlinear differential equations to model the transistor as a dynamic dual-input-dual-output (DIDO) system. This approach is ideal for intellectual property protection. It is also computationally efficient (it takes very little simulation time and has good convergence properties),

[1]Extrinsic and intrinsic shells are common terminologies used in transistor modelling and will be explained in later sections of this chapter.

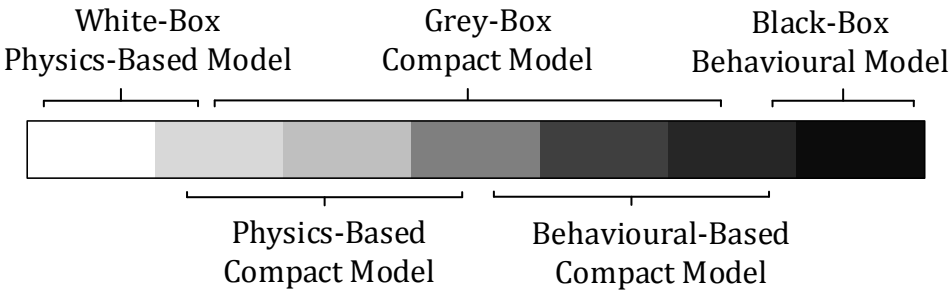


Figure 2.1: The modelling spectrum.

thus used in system-level simulations. However, behavioural models are locally accurate (because they are built with no knowledge of the internal construction of the system). Therefore, restricting the validity of these kinds of models to the conditions under which they were extracted; as such, they cannot be used in the PA design phase.

Compact models include everything in between the above two extremes. They are a compromise between the generality of physics-based models and the computational efficiency of behavioural models. Because of their ability to represent the device under a broader zone of input stimuli with an acceptable level of accuracy, compact models represent the best option for circuit-level simulations. It is obvious from Figure 2.1, that "compact model" is a very loose term. It can be particularly confusing when talking about behavioural-based compact models which are quite distinct from black-box behavioural models. The best way to define "compact model" is through the objective for which the model was built, i.e., to design circuits (Step 1 in Figure 1.1). Therefore, compact models are expected to work under different circuit simulators [direct current (DC), small-signal and large-signal simulators].

From the discussion in Section 1.2, one can notice the use of two types of models. First, a compact model was used to design the PA, and then a behavioural model was built to linearize it. Since this thesis aims to bridge the gap between the two approaches, the remainder of this chapter presents a review of the existing compact and behavioural modelling techniques for radio frequency (RF) power transistors. They will be grouped based on the common properties imposed by the model structure (formulation and extraction measurements) to help with problem identification, and hence pave the way to the proposed model[2].

[2]Since compact and behavioural modelling approaches have different objectives, no comparison will be provided in this chapter. Yet, each approach will be compared separately with its equivalence in the proposed work (Table 3.3 and Table 5.5).

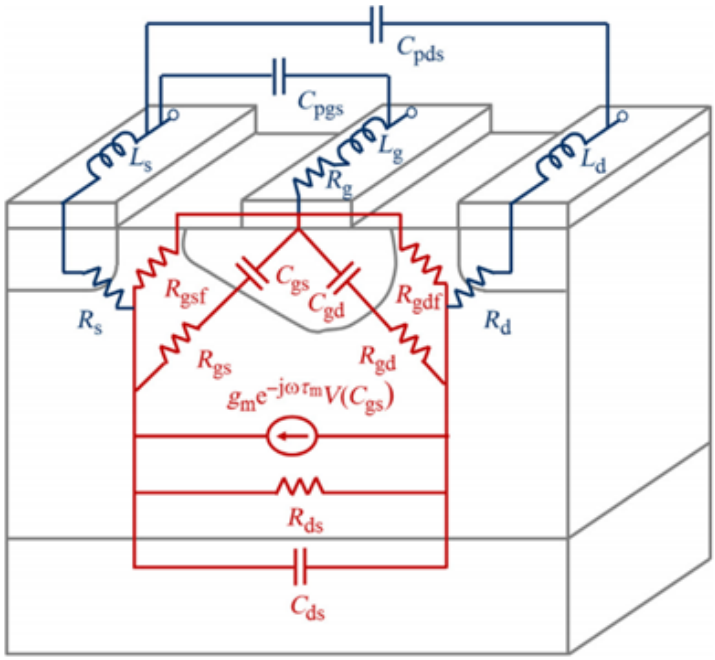
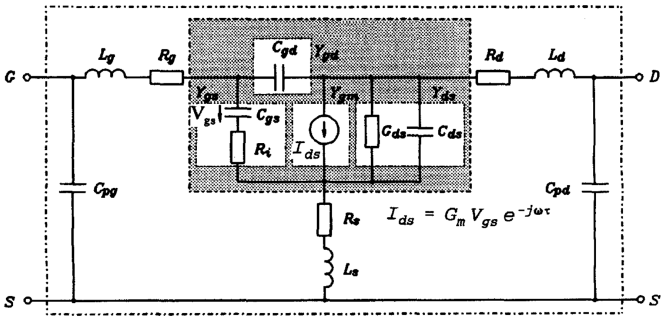


Figure 2.2: An example of a physics-inspired equivalent circuit of the transistor [10]. The blue and red elements represent the extrinsic and intrinsic shells, respectively.

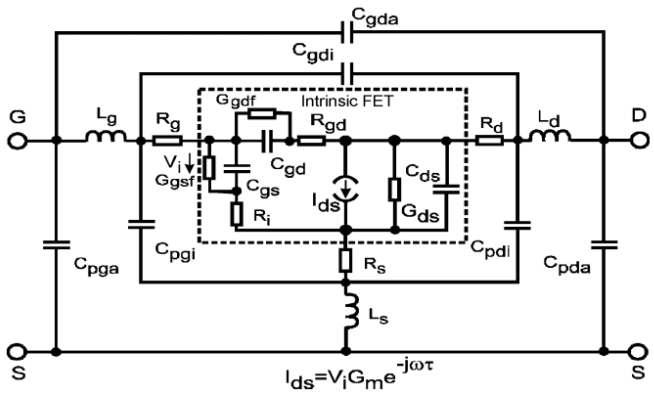
2.2 Compact Models

2.2.1 Linear Measurement-Based Compact Models

Transistor compact modelling approach approximates the transistor's large-signal behaviour using an equivalent circuit that is divided into two shells: an active intrinsic shell and a passive extrinsic shell. While the first represents the part at which transistor action takes place (the channel), the later represents everything outside the channel (often called transistor parasitics) [9]. The two-shell assumption is fundamental to this modelling approach as it gives the equivalent circuit more physical significance (Figure 2.2) and hence simplifies the extraction of its elements through a two-tier process. First, extrinsic shell elements are extracted using ColdFET technique [11], where the transistor channel is turned off. Second, intrinsic shell elements are extracted analytically using bias-dependent S-parameters measurements. The most important step in this process is to find an equivalent circuit that matches the transistor's behaviour (S-parameters versus both frequency and biasing grids). Therefore, different topologies (for each shell) exist in the literature for different de-



(a)



(b)

Figure 2.3: Examples of different extrinsic and intrinsic topologies used in transistor compact modelling: (a) 15-elements [12], and (b) 22-elements [13] models. G, D, and S stand for the gate, drain and source, respectively.

vice technologies, e.g. 15-elements and 22-elements topologies are shown in Figure 2.3 (a) and (b), respectively. It is worth mentioning that even within the same topology, optimization techniques are a must [13, 14] for the following reasons. First, analytical expressions for intrinsic shell elements cannot be found if the shell has more than eight elements[3]. Second, solution uniqueness cannot be achieved due to unavoidable measurement uncertainties [12]. Finally, given a prior-knowledge of the application, compact model accuracy can be increased by optimizing some elements using large-signal measurements [15].

In order to save time and effort in compact model extraction and implementation, a

[3]Since intrinsic shell extraction depends on S-parameters, only four complex equations are available. Therefore, only eight elements have to exist in order for the system to have a unique solution.

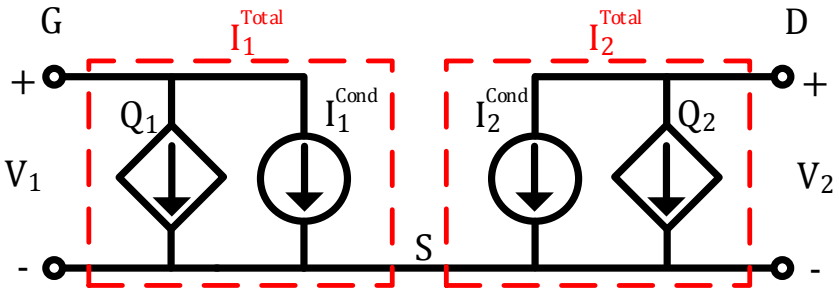


Figure 2.4: The equivalent circuit for the intrinsic transistor in the technology-independent approach. G, D, and S stands for the gate, drain and source, respectively.

technology-independent approach was introduced in [16]. The equivalent circuit for the intrinsic transistor is shown in Figure 2.4 and consists of conduction and displacement current sources per each port. These sources are found by integrating bias-dependent Y-parameters. The frequency-domain constitutive relationships are expressed in (2.1) and Y-parameters represent the first order Taylor expansion around DC biasing points (2.2). Although this equivalent circuit has less physical significance, ColdFET technique is still used to add more degrees of freedom in order to match the transistor's behaviour [17].

$$\begin{aligned} I_1^{Total}(V_1, V_2) &= I_1^{Cond}(V_1, V_2) + j\omega Q_1(V_1, V_2) \\ I_2^{Total}(V_1, V_2) &= I_2^{Cond}(V_1, V_2) + j\omega Q_2(V_1, V_2) \end{aligned} \tag{2.1}$$

$$\begin{bmatrix} I_1^{Total}(v_1, v_2) \\ I_2^{Total}(v_1, v_2) \end{bmatrix} = \begin{bmatrix} I_1^{Total}(V_1^{DC}, V_2^{DC}) \\ I_2^{Total}(V_1^{DC}, V_2^{DC}) \end{bmatrix} + [Y] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad : \tag{2.2}$$

$$[Y] = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} \frac{dI_1^{Cond}}{dV_1} + j\omega\frac{dQ_1}{dV_1} & \frac{dI_1^{Cond}}{dV_2} + j\omega\frac{dQ_1}{dV_2} \\ \frac{dI_2^{Cond}}{dV_1} + j\omega\frac{dQ_2}{dV_1} & \frac{dI_2^{Cond}}{dV_2} + j\omega\frac{dQ_2}{dV_2} \end{bmatrix}_{\substack{V_1=V_1^{DC} \\ V_2=V_2^{DC}}}$$

where,

I_p^{Total} is the frequency-domain total current at port p

I_p^{Cond} is the frequency-domain conduction current at port p

Q_p is a charge source that represents the frequency-domain

displacement current at port p

V_p is the large-signal voltage at port p

v_p is the small-signal voltage at port p

V_p^{DC} is the DC voltage at port p

$[Y]$ is the frequency domain Y-parameters at (V_1^{DC}, V_2^{DC})

The linear measurement-based compact modelling approach has a strong mathematical foundation. Long-term memory effects (thermal and trapping) can be added as in [18–21]. It has a systematic way of extraction that requires a simple measurement setup. However, compact models built using this approach have very bad prediction capability of nonlinear behaviours such as harmonic and intermodulation tone generation.

2.2.2 Nonlinear Measurement-Based Compact Models

The evolution of measurement setups [22, 23] led by nonlinear vector network analyzers (NVNAs), gave a whole different prospect to device modelling techniques. Although this approach uses the same constitutive relationships given by (2.1) to model the transistor, it is different from the linear measurement-based approach. Instead of linearizing the expressions around each biasing point, this approach builds the time-domain version of (2.1) using the load-line concept. This concept exploits the NVNA's capacity to measure the large-signal time-domain voltage and current waveforms. Figure 2.5 shows an example of three voltage and current waveforms projected at the I_2/V_2 plane producing three load-lines. Hundreds of these load-lines should be generated to fully cover the modelling space represented by the blue-shaded area in Figure 2.5, thus fulfilling the generality requirement described in section 2.1. These load-lines are gathered using a sophisticated active source/load-pull setup [24]. The idea is to intelligently sweep the power and phase of the incident power-waves (A-waves) as well as the DC biasing points, to generate current and voltage waveforms that result in load-lines covering the required modelling space. Load-lines are usually generated using two frequencies in order to simplify the fitting of the constructed time-domain data. While the first frequency is chosen to be low enough such that the behaviour of the conduction current is dominant, the second is chosen to be high enough to do the opposite, i.e., to make the displacement current dominant.

One of the most successful models that use this approach is Dynamic FET (DynaFET) model. Initially proposed in [25], the model was extracted by measuring 8542 load-lines at

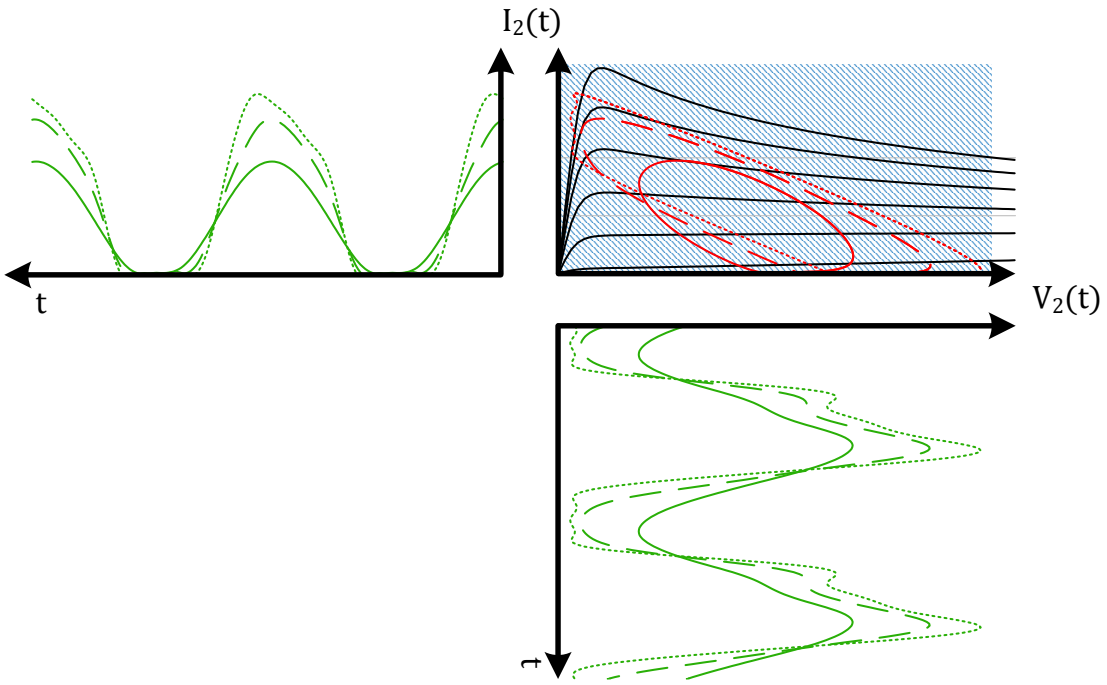


Figure 2.5: An illustration of projecting load-lines on the instantaneous I_2/V_2 plot. The green, red, and black lines represent voltage and current time domain waveforms, the projected load lines, and the DC I/V curves, respectively. The blue-shaded area represents the modelling space.

two frequencies: 2 GHz and 4 GHz, with power levels ranging from small-signal[4] up to an 8 dB compression point. Further improvements were outlined in [27] where the model was extracted by performing large-signal measurements at 100 MHz and S-parameters at 1.9 GHz. Self-heating and charge trapping were added to this model by introducing auxiliary variables that are functions of the instantaneous dissipated power and the average terminal voltages, respectively. In other words, by adding more degrees of freedom to (2.1). Time-domain waveforms were constructed by measuring the first six harmonics in addition to the DC component.

There are two main advantages of such an approach. First, the device is modelled under large-signal operating conditions, and hence, more nonlinearities are stimulated and

[4]The small-signal power level is defined as a 20 dB back-off from the 1 dB compression point, because at this point the compression level is 0.01 dB [26].

modelled. Second, driving the device really hard, in terms of RF power[5], allows it to be modelled in regions closer to the breakdown without damaging it (as it enters this region for a very short period). On the other hand, the models generated using this approach suffer from a major problem that prevents them from accurately predicting the transistor behaviour under modulated signals. This is because using the load-line concept to generate compact models does not excite the full dynamics of the high-order partial derivatives of the ports currents. These high-order derivatives are responsible for all nonlinear behaviours needed to predict the performance under modulated signals, as will be explained later in the proposed solution (Chapter 3 and Chapter 5).

Finally, it is worth mentioning that not all transistor compact models can be classified into only one of the categories above. The dynamic-bias model [15, 28], for example, is a linear measurement-based compact model that is locally tuned using the load-line concept (assuming a priori-knowledge of the application).

2.3 Behavioural Models

2.3.1 The Volterra Series

The Volterra series is considered one of the most rigorous theories that deal with modelling WNL behaviours [29]. It describes the total WNL response of the system as a summation of different order responses (2.3) that are related to the input through Volterra kernels (nonlinear impulse responses). Since dealing with the multidimensional convolution in (2.3) is a complicated process, Volterra kernels can be expressed in the frequency-domain as nonlinear transfer functions (2.5) by assuming the excitation has the general multi-tone form in (2.4) [30].

$$y(t) = \sum_{n=1}^{N} y_n(t) : \tag{2.3}$$

$$y_n(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) x(t-\tau_1) \dots x(t-\tau_n) d\tau_1 \dots d\tau_n$$

$$x(t) = \frac{1}{2} \sum_{\substack{q=-Q \\ q \neq 0}}^{Q} X_q e^{j\omega_q t} \tag{2.4}$$

[5]While typical GaN devices can handle up to 5 W/mm of DC power, this approach can derive them to reach an instantaneous power of 20 W/mm [27].

$$y_n(t) = \frac{1}{2^n} \sum_{q_1=-Q}^{Q} \dots \sum_{q_n=-Q}^{Q} X_{q_1} \dots X_{q_n} H_n(\omega_{q_1} \dots \omega_{q_n}) e^{j(\omega_{q_1}+\dots+\omega_{q_n})t} \tag{2.5}$$

where,

$y(t)$ is the total WNL response

$y_n(t)$ is the n^{th}-order response

$x(t)$ is the multi-tone excitation

n is the response-order

X_q is the complex amplitude of $x(t)$ at ω_q

Q is the total number of tones in $x(t)$

$h_n(\tau_1, \dots, \tau_n)$ is the n^{th}-order nonlinear impulse response

$H_n(\omega_{q_1} \dots \omega_{q_n})$ is the n^{th}-order nonlinear transfer function

The Volterra series is very accurate in predicting WNL behaviours such as intermodulation distortion (IMD) bands. Therefore, different derivations of it are used in PAs linearization [31, 32]. For example, a Volterra-based DPD was utilized to linearize the PA in Section 1.2. However, it is difficult to use the Volterra series in transistor compact modelling for the following reasons. First, being a black-box behavioural model limits the validity of the Volterra kernels to the conditions under which they were extracted. Second, the application of the Volterra series is limited to the WNL region of the device. In order to handle strong nonlinear behaviours, high-order kernels need to be included which affects the convergence of the series [33]. Finally, any attempt to extract Volterra kernels to cover the entire transistor modelling space (and hence achieve the generality requirement) will result in an explosion of the number of kernels.

2.3.2 Poly-Harmonic Distortion (PHD) -based Models

Poly-Harmonic distortion (PHD) concept [34, 35] was proposed to extend linear S-parameters to predict nonlinear behaviours. It starts with the concept of multi-harmonic spectral mapping- which states that if a nonlinear system is excited with a stimulus (A-wave) that has spectral components on a harmonic grid, the response (B-Wave) will have spectral components on the same grid- to derive the PHD describing function (2.6). $F_{p,k}$ is a nonlinear function that relates the k^{th} spectral component of B-wave at port p to all the spectral components of A-wave at all ports. According to (2.2), Y-parameters represent the first-order Taylor expansion of the device describing functions around the DC operating point.

Similarly, PHD-based models represent the first-order Taylor expansion of the describing function (2.6) around the system's large-signal steady-state due to both the stimulus and the response. This new concept of operating points was referred to as the large-signal operating point (LSOP). To guarantee the analyticity of $F_{p,k}$ (and hence its linearizability), the complex amplitudes of the spectral components were added to the functional dependency resulting in (2.7). Next, time-invariant mapping was assumed which allowed normalizing the phases of all spectrum components with respect to the phase of the fundamental tone of A-wave at port 1 as described by (2.8). Finally, the linearization of (2.8) is performed around the LSOP to reach the formulation of X-parameters[6] (2.9) [36, 37]. $X^{(F)}$ represent the large-signal response which is equivalent to the DC currents in (2.2). $X^{(S)}_{p,k;q,l}$ and $X^{(T)}_{p,k;q,l}$ are interpreted in a similar way to S-parameters, i.e., they represent the contribution of the l^{th}-harmonic of A-wave at port q to the k^{th} harmonic of B-wave at port p.

$$B_{p,k} = F_{p,k}(A_{1,1}, A_{1,2}, \dots, A_{2,1}, A_{2,2}, \dots) \tag{2.6}$$

$$B_{p,k} = F_{p,k}(A_{1,1}, A^*_{1,1}, A_{1,2}, A^*_{1,2}, \dots, A_{2,1}, A^*_{2,1}, A_{2,2}, A^*_{2,2}, \dots) \tag{2.7}$$

$$B_{p,k} = F_{p,k}\big(|A_{1,1}|, A_{1,2}P^{-1}, A^*_{1,2}P^2, \dots, A_{2,1}P^{-1}, A^*_{2,1}P^1, A_{2,2}P^{-2}, A^*_{2,2}P^2, \dots\big)P^k \tag{2.8}$$

$$B_{p,k} = X^{(F)}_{p,k}(LSOP)P^k + \sum_{q,l}\left[X^{(S)}_{p,k;q,l}(LSOP)a_{q,l}P^{k-1} + X^{(T)}_{p,k;q,l}(LSOP)a^*_{q,l}P^{k+1}\right] \tag{2.9}$$

where,

$a_{q,l}$ is the small-signal A-wave at port q and harmonic l

$A_{q,l}$ is the large-signal A-wave at port q and harmonic l

$B_{p,k}$ is the large-signal B-wave at port p and harmonic k

$P = e^{j\phi(A_{1,1})}$ is the the phase of $A_{1,1}$

It is worth mentioning that X-parameters were first introduced to model well-matched nonlinear devices. Therefore, for a 2-port device, only DC voltages at both ports ($V_{1,0}$ and $V_{2,0}$) and the fundamental tone of the stimulus ($A_{1,1}$) were considered the contributors to the LSOP, i.e. $LSOP = (V_{1,0}, V_{2,0}, |A_{1,1}|)$. In order to handle strong nonlinear behaviours that results from poor matching, load-dependent X-parameters were introduced by extending the LSOP to include the the fundamental tone of A-wave at the output port ($A_{2,1}$), i.e. $LSOP = (V_{1,0}, V_{2,0}, |A_{1,1}|, A_{2,1})$ [38]. Furthermore, Dynamic kernels were introduced

[6] X-parameters is a PHD-based model commercialized by Keysight.

to model the dynamics of the nonlinear device [38, 39]. Also, frequency scalability was added by introducing the admittance-space X-parameters (XY-parameters) [40].

A multi-harmonic source/load-pull setup [41] is needed in order to extract PHD-based models [42]. First, an LSOP needs to be defined, and then an extraction tone is injected to sweep all other variables sequentially. Other PHD-based models include Cardiff model [43] and Pade's model [44]. While the first uses the polar form of the PHD describing function to separate between phase and amplitude nonlinearities (which leads to a quicker model identification), the second use Pade's approximation instead of the polynomials as basis functions to extend the modelling space. PHD-based models have demonstrated an excellent ability to model nonlinear behaviours. However, all these models are limited by the PHD principle itself when it comes to solution generality. This is because high-order responses are treated as first-order Taylor expansions of a multidimensional function (2.6)-(2.9) which results in the following problems:

- It is quite difficult to determine which variables should be included in the LSOP when handling strong nonlinearities, especially if there is no a priori-knowledge of the application [37].
- Any attempt to generalize the solution, by including more variables in the LSOP (in order to be used as a compact model) will result in an explosion of the modelling space [42].
- The formulation of the PHD-based models in (2.6) as a multi-dimensional function makes the model prediction highly dependent on the implementation and the interpolation techniques used by the simulator [45].

2.4 Conclusion

The basic assumption, as well as the fundamental limitation, in linear measurement-based compact models, is that it is possible to recover a nonlinear function by measuring only the DC and first-order components of its Taylor series expansion. Another significant disadvantage of this approach is the non-systematic method used to come up with equivalent circuits because elements have to be empirically added to the model whenever the transistor technology is changed. Nonlinear measurement-based compact models, on the other hand, depend on stimulating as many nonlinearities and dynamics in the transistor as the characterization setup can manage. Although this seems to be the right approach to modelling as it captures the nonlinear effects without any assumptions, it does not excite

the full dynamics of the high-order partial derivatives of the ports currents. Thus, these models lack the capacity to predict the device behaviour under modulated signal stimuli.

Black-box behavioural models are very accurate, yet their application is limited to the conditions under which they are extracted. It is also difficult to cover the modelling space required by compact models without having an explosion in the number of kernels. Finally, even if one manages to extract these kernels, there is no mathematical framework that allows utilizing them in transistor compact modelling.

In conclusion, compact models are hungry for more data that includes the transistor's nonlinear and dynamical behaviours. The main problem is the lack of mathematical formulations that allows incorporating these data into one model that can be extracted systematically. It is worth mentioning that none of the compact models mentioned in this chapter was validated under modulated signal stimuli.

Chapter 3

High-Order Network Parameters

3.1 Introduction

In this chapter, we take a step back and start with extending the original 2-port linear network theory [46] to the WNL case. A Volterra-based system is used to derive a new set of parameters that, once identified, can be used to describe WNL behaviours of the device under test (DUT) under any stimuli. These superset parameters are called high-order network parameters (HONPs) and will be derived in two forms: high-order Y-parameters (HOYPs) and high-order S-parameters (HOSPs). It will be proven, in both simulation and measurement environments, that these high-order parameters are the correct extension of the linear parameters (i.e. Y- and S-parameters) as they do not depend on the load termination, power, or the stimulus used in the extraction process. In other words, they only depend on the DUT. It will also be shown that HONPs inherit the main properties of the linear parameters represented in power independency, continuity of the behaviour across the modelling space (i.e. frequency and biasing grids), and solution uniqueness. These properties are the main reason for the popularity of linear parameters. While the first two properties limit the explosion of the modelling space[1] and allow interpolating the parameters, and hence reducing measurement time, the last one is of particular importance as it allows identifying any linear system with absolute mathematical certainty. A misleading understanding of S-parameters is that they are extracted in the DUT linear region of operation. In this chapter, it will be shown that HONPs are extracted while the nonlinear DUT is operating in the same region as S-parameters. Therefore, this region will

[1]The modelling space for linear DUTs is represented by the frequency. For nonlinear DUTs , e.g. a transistor, it is defined by both frequency and biasing points.

be referred to as the WNL region. Finally, a multi-tone verification is presented to prove HONPs capabilities in predicting WNL behaviours.

3.2 Modelling Domain

The relationship between the degree of the nonlinearity and the order of the response is an important modelling concept that can greatly affect the results if not dealt with properly. Nonlinear models can be represented in either one of the following domains: voltage/current (V/I) or incident/reflected power-waves (A/B). While V/I domain is used to implement models in circuit simulators, A/B domain is used to collect measurement data at microwave frequencies. This is because, at microwave frequencies, signal generators can only set the input incident power-wave rather than the input voltage. In order to illustrate the nature of the problem, the following example is considered. Equation (3.1) describes a third-degree nonlinear conductance where a, b and c are arbitrary constants. This element is excited with CW voltage source, CW current source, and CW power source, respectively, under HB simulator[2] (Figure 3.1). The spectrum of the simulated voltage and current at the nonlinear conductance are shown in Figure 3.2. Since there is no feedback in this circuit and the nonlinearity is limited to the third-degree, the response is expected to have components up to the third-order mixing product only. While this is true when using a CW voltage source, stimulating the nonlinear element with a CW current source results in an infinite-order mixing products (Figure 3.2 (a), and (b), respectively). This behaviour can be justified by reversing (3.1) using the *formula of power series reversion* [47], which results in an infinite-degree nonlinear resistor (3.2). This means that a third-degree nonlinear conductance behaves as a third-degree nonlinearity only if excited with a voltage source. Since the models considered in this thesis work at microwave frequencies, power sources are used to excite the DUT with A-waves. Therefore, both voltage and current will have infinite-order mixing products [Figure 3.2 (c)]. It is worth mentioning that although both domains are linearly related through (3.3), the choice of the modelling domain will affect the extracted HONPs (as will be shown in the next sections).

$$i(t) = av(t) + bv^2(t) + cv^3(t) \tag{3.1}$$

$$v(t) = Ai(t) + Bi^2(t) + Ci^3(t) + \dots \tag{3.2}$$

$$\begin{aligned} v(t) &= (A(t) + B(t))\sqrt{Z_o} \\ i(t) &= (A(t) - B(t))/\sqrt{Z_o} \end{aligned} \tag{3.3}$$

[2]Keysight's Advanced Design System (ADS) circuit simulator was used.

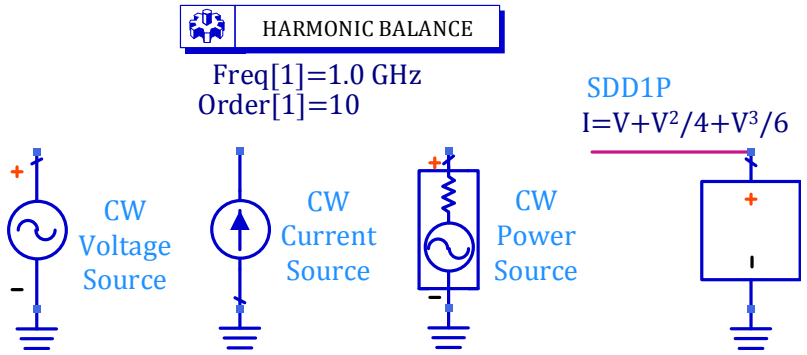


Figure 3.1: The schematic of a nonlinear conductance excited using CW voltage source, CW current source, and CW power source in HB simulator.

where,

$$aA = 1, \qquad a^3B = -b, \qquad a^5C = 2b^2 - ac, \quad \ldots$$

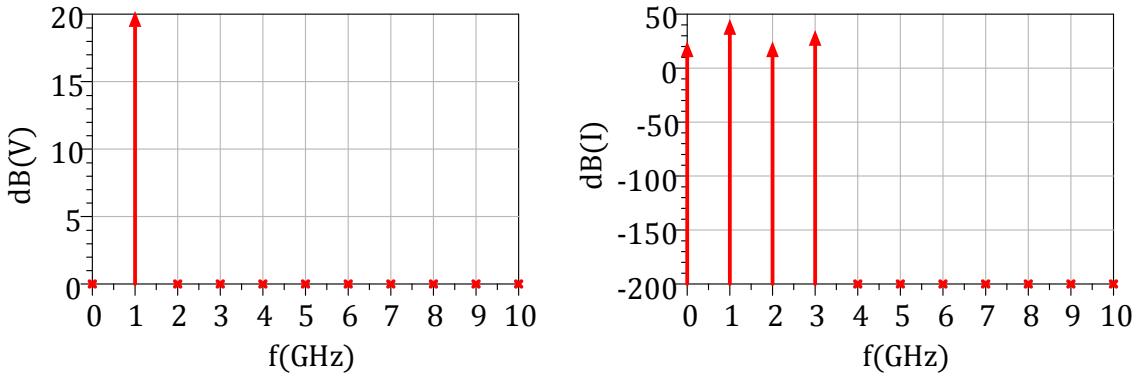
$V(t)$ is the time-domain voltage

$i(t)$ is the time-domain current

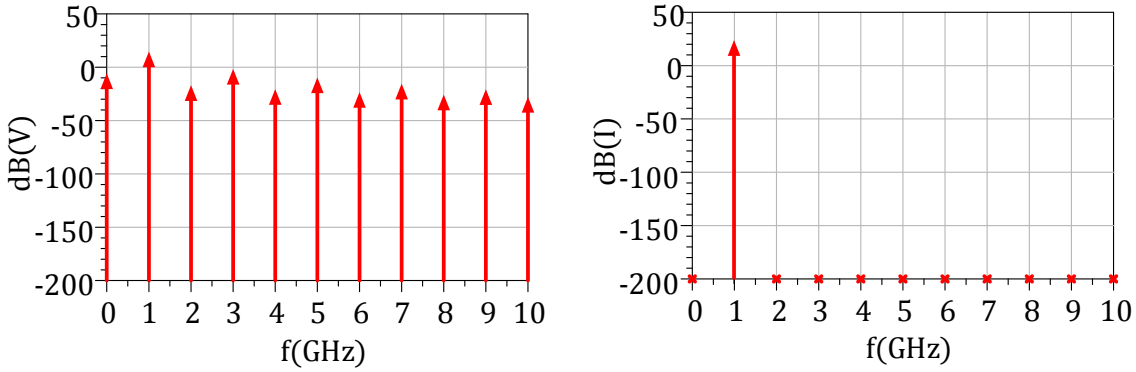
$A(t)$ is the time-domain incident power-wave

$B(t)$ is the time-domain reflected power-wave

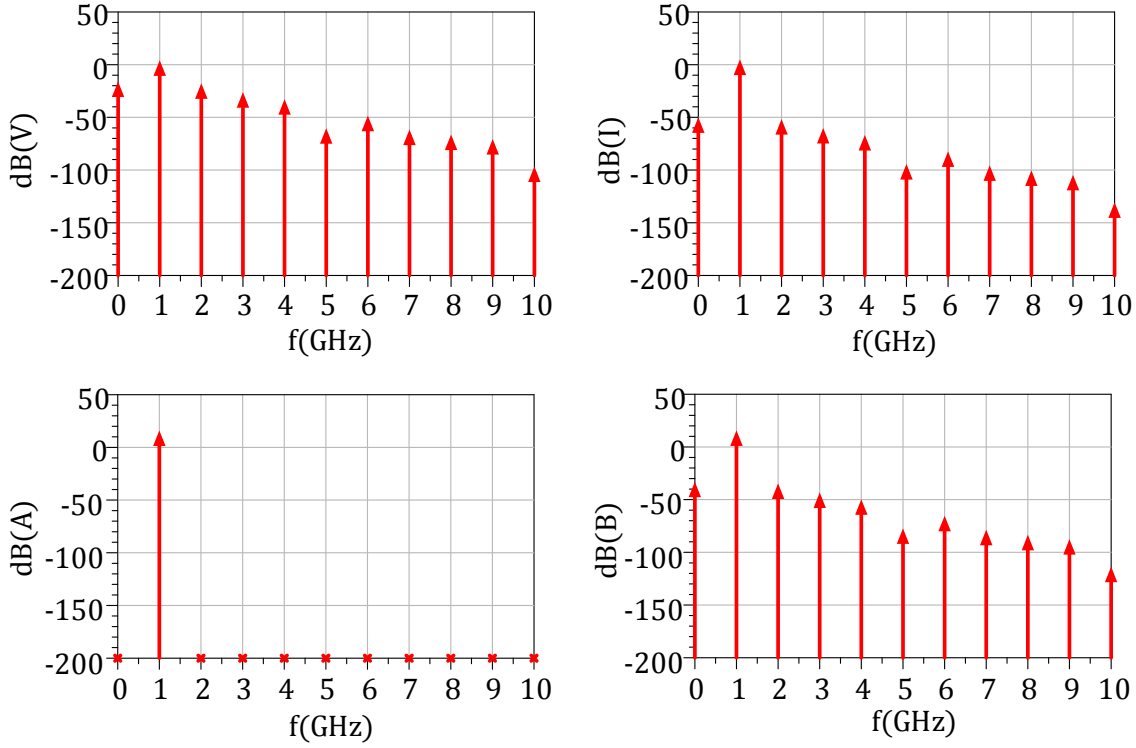
Z_o is the characteristic impedance of the system.



(a)



(b)



(c)

Figure 3.2: A comparison between voltage, current, and power-wave spectrums when exciting the nonlinear element shown in Figure 3.1 with: (a) CW voltage source, (b) CW current source, and (c) CW power source.

3.3 High-Order Network Parameters (HONPs) Formulations

In this section, the original 2-port linear network theory [46] is extended to the WNL case. Although many equivalent representations exist in the linear case (e.g. Y, Z, S, H, etc), only Y-parameters and S-parameters will be the points of focus here. This is because the first is essential for compact model implementation in available circuit simulators[3], and the second is irreplaceable when it comes to measurements performed at microwave frequencies.

3.3.1 High-Order Y-Parameters (HOYPs) Formulation

A 2-port linear network can be fully described by a set of four parameters. Figure 3.3 shows two topologies for Y-Parameters. Although these representations are equivalent in the linear case, they are not in the nonlinear case. For example, if each element is described by a polynomial [as in (3.1)], the ports currents of the first topology would have terms containing $v_1 v_2$ which are absent in the second topology. Also, port 1 current of the second topology would have terms containing v_2^2 that do not exist in the first topology[4]. In order to maintain the generality of the theory, all high-order terms should be included and hence the topology in Figure 3.4 is proposed. The relationship between ports voltages and currents is described by (3.4)

$$\begin{bmatrix} i_1(t) \\ i_2(t) \end{bmatrix} = \begin{bmatrix} y_{11}^{(1)}(t) & y_{12}^{(1)}(t) \\ y_{21}^{(1)}(t) & y_{22}^{(1)}(t) \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} + \begin{bmatrix} y_{11}^{(2)}(t) & y_{12}^{(2)}(t) & y_{13}^{(2)}(t) \\ y_{21}^{(2)}(t) & y_{22}^{(2)}(t) & y_{23}^{(2)}(t) \end{bmatrix} \begin{bmatrix} v_1^2(t) \\ v_2^2(t) \\ v_1(t)v_2(t) \end{bmatrix} \tag{3.4}$$

[3]Most of the existing commercial circuit simulators use Modified Nodal Analysis method to formulate systems of differential equations, which will then be solved using numerical techniques. Therefore, implementing compact models in the nodal form, i.e. ports currents are functions of ports voltages, helps in the simulation convergence [47].

[4]This example shows the missing terms considering only second-order nonlinearities.

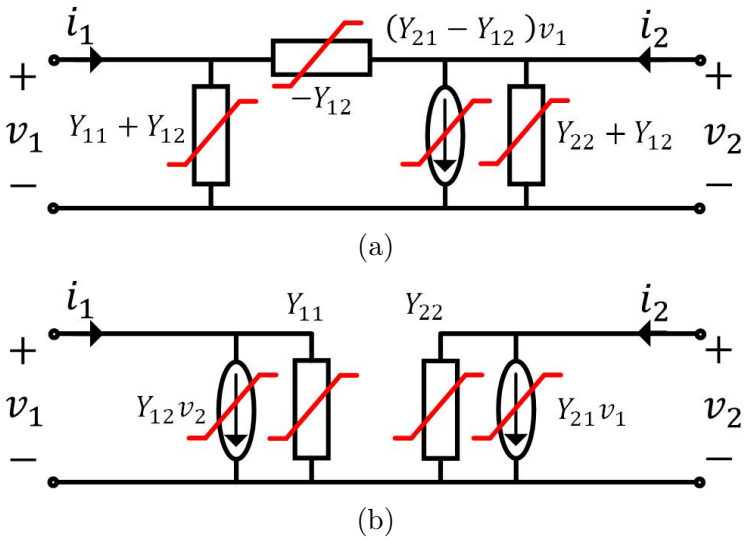


Figure 3.3: Y-parameters equivalent circuits: (a) output-circuit-dependent-generator, and (b) two-dependent-generator form [48].

$$+ \begin{bmatrix} y_{11}^{(3)}(t) & y_{12}^{(3)}(t) & y_{13}^{(3)}(t) & y_{14}^{(3)}(t) \\ y_{21}^{(3)}(t) & y_{22}^{(3)}(t) & y_{23}^{(3)}(t) & y_{24}^{(3)}(t) \end{bmatrix} \begin{bmatrix} v_1^3(t) \\ v_2^3(t) \\ v_1^2(t)v_2(t) \\ v_1(t)v_2^2(t) \end{bmatrix}$$

where $v_p(t)$ and $i_p(t)$ represent the time-domain voltages and currents at port p. $y_{kl}^{(n)}(t)$ represents the time-domain n^{th}-order network parameter and the subindexes k and l are used similarly as in the original 2-port linear network theory. Although the system described by (3.4) represents a DIDO Volterra system with dynamic kernels, a different mathematical treatment will be applied here [49] to allow both solution generality and systematic extraction procedure that are necessary for compact model formulation and extraction (a full analysis for the proposed system using Volterra Series is shown in Appendix A).

First, without loss of generality, the dynamics of these parameters are assumed to be approximated by the first and high-order time-derivatives and time-integrals of the ports voltages, which is a generalized case of the quasi-static assumption[5]. Therefore, the

[5]The quasi-static approach assumes that both conduction-currents and charge sources in Figure 2.4 are static functions of the controlling voltages [10, 50], i.e., $m = 0, 1$ in (3.5). The purpose of this assumption is to simplify the transformation from time-domain to frequency-domain without using convolution.

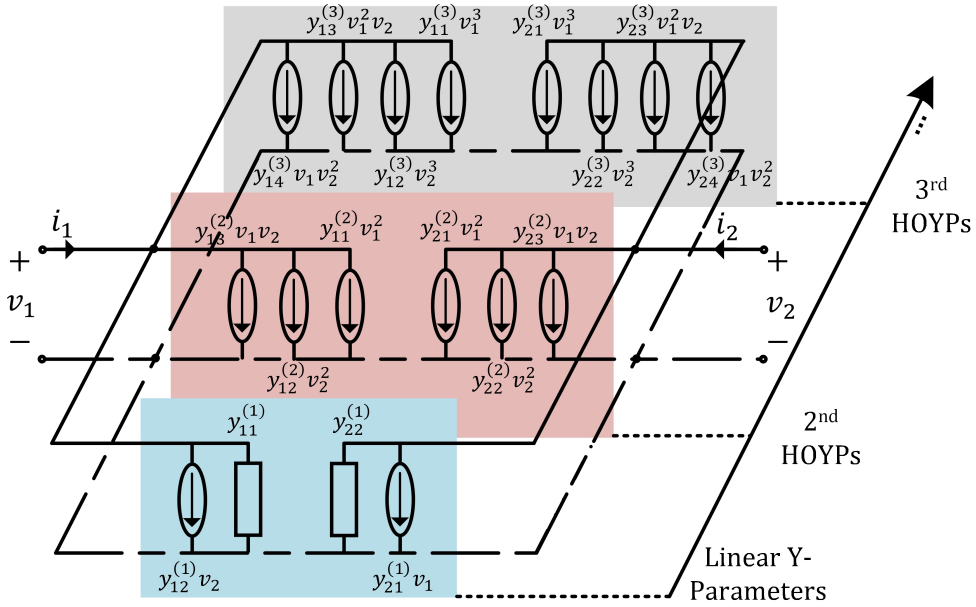


Figure 3.4: First (blue), second (red), and third (grey) HOYPs equivalent circuit. Time dependency is dropped for ease of notation.

time-domain high-order network parameters $y_{kl}^{(n)}(t)$ in (3.4) become the integro-differential operators $Y_{kl}^{(n)}(P)$ described by (3.5), where P is a differential operator as shown in (3.6)[6]. The positive and negative values of m represent the m^{th}-order time-derivative and time-integral, respectively.

$$Y_{kl}^{(n)}(P) = \sum_{m=-\infty}^{\infty} Y_{kl,m}^{(n)} P^m \tag{3.5}$$

$$Pv(t) = \frac{d}{dt}v(t) \qquad \text{and} \qquad P^{-1}v(t) = \int_{-\infty}^{t} v(z)dz \tag{3.6}$$

The next step is to solve the system of nonlinear equations (3.4) using the Method of Nonlinear Currents [30]. Both voltages and currents are divided into components with different orders, and hence the superposition principle can be applied. Considering up to the N^{th}-order mixing products and assuming multi-tone excitations at both ports, ports

[6]Operator notation is used here to simplify equations readability.

voltages and currents can be expressed as in (3.7).

$$\begin{aligned} v_p(t) &= \sum_{n=1}^{N} v_{p,n}(t) : \qquad v_{p,n}(t) = \frac{1}{2} \sum_{\substack{q=-Q_n \\ q\neq 0}}^{Q_n} V_{p,q} e^{j2\pi f_q t} \\ i_p(t) &= \sum_{n=1}^{N} i_{p,n}(t) : \qquad i_{p,n}(t) = \frac{1}{2} \sum_{\substack{q=-Q_n \\ q\neq 0}}^{Q_n} I_{p,q} e^{j2\pi f_q t} \end{aligned} \tag{3.7}$$

where,

$p = 1, 2$ is the port number

$v_p(t)$ is the total time-domain voltage at port p

$i_p(t)$ is the total time-domain current at port p

$v_{p,n}(t)$ is the n^{th}-order time-domain voltage at port p

$i_{p,n}(t)$ is the n^{th}-order time-domain current at port p

$V_{p,q}(t)$ is the complex amplitude of $v_p(t)$ at tone q

$I_{p,q}(t)$ is the complex amplitude of $i_p(t)$ at tone q

Q_1 is the number of fundamental tones in the multi-tone excitation

Q_n is the number of all n^{th}-order mixing product (harmonic and IMD) tones

In order to illustrate the concept, both the degree of nonlinearities and the order of the response will be limited to three ($N = 3$). Therefore substituting by (3.5) and (3.7) in (3.4) will result in (3.8), where the three current components represent first, second, and third-order responses of the WNL system (3.9)-(3.11).

$$\begin{bmatrix} i_1(t) \\ i_2(t) \end{bmatrix} = \begin{bmatrix} i_{1,1}(t) \\ i_{2,1}(t) \end{bmatrix} + \begin{bmatrix} i_{1,2}(t) \\ i_{2,2}(t) \end{bmatrix} + \begin{bmatrix} i_{1,3}(t) \\ i_{2,3}(t) \end{bmatrix} \tag{3.8}$$

$$\begin{bmatrix} i_{1,1}(t) \\ i_{2,1}(t) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(P) & Y_{12}^{(1)}(P) \\ Y_{21}^{(1)}(P) & Y_{22}^{(1)}(P) \end{bmatrix} \begin{bmatrix} v_{1,1}(t) \\ v_{2,1}(t) \end{bmatrix} \tag{3.9}$$

$$\begin{bmatrix} i_{1,2}(t) \\ i_{2,2}(t) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(P) & Y_{12}^{(1)}(P) \\ Y_{21}^{(1)}(P) & Y_{22}^{(1)}(P) \end{bmatrix} \begin{bmatrix} v_{1,2}(t) \\ v_{2,2}(t) \end{bmatrix} + \begin{bmatrix} Y_{11}^{(2)}(P) & Y_{12}^{(2)}(P) & Y_{13}^{(2)}(P) \\ Y_{21}^{(2)}(P) & Y_{22}^{(2)}(P) & Y_{23}^{(2)}(P) \end{bmatrix} \begin{bmatrix} v_{1,1}^2(t) \\ v_{2,1}^2(t) \\ v_{1,1}(t)v_{2,1}(t) \end{bmatrix} \tag{3.10}$$

$$\begin{bmatrix} i_{1,3}(t) \\ i_{2,3}(t) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(P) & Y_{12}^{(1)}(P) \\ Y_{21}^{(1)}(P) & Y_{22}^{(1)}(P) \end{bmatrix} \begin{bmatrix} v_{1,3}(t) \\ v_{2,3}(t) \end{bmatrix} + \begin{bmatrix} Y_{11}^{(2)}(P) & Y_{12}^{(2)}(P) & Y_{13}^{(2)}(P) \\ Y_{21}^{(2)}(P) & Y_{22}^{(2)}(P) & Y_{23}^{(2)}(P) \end{bmatrix} \begin{bmatrix} 2v_{1,1}(t)v_{1,2}(t) \\ 2v_{2,1}(t)v_{2,2}(t) \\ v_{1,1}(t)v_{2,2}(t) + v_{2,1}(t)v_{1,2}(t) \end{bmatrix} + \begin{bmatrix} Y_{11}^{(3)}(P) & Y_{12}^{(3)}(P) & Y_{13}^{(3)}(P) & Y_{14}^{(3)}(P) \\ Y_{21}^{(3)}(P) & Y_{22}^{(3)}(P) & Y_{23}^{(3)}(P) & Y_{24}^{(3)}(P) \end{bmatrix} \begin{bmatrix} v_{1,1}^3(t) \\ v_{2,1}^3(t) \\ v_{1,1}^2(t)v_{2,1}(t) \\ v_{1,1}(t)v_{2,1}^2(t) \end{bmatrix} \tag{3.11}$$

Since the application of the integro-differential operators (3.5) to the exponential terms in (3.7) results in $(j2\pi f)^m$, rewriting (3.9)-(3.11) at fundamental, second and third-order mixing product tones results in (3.12)-(3.14), respectively. $[Y^{(1)}(f)]$, $[Y^{(2)}(f)]$, and $[Y^{(3)}(f)]$ are the first[7], second and third HOYPs in frequency-domain (3.15) - (3.17).

$$\frac{1}{2}\sum_{q=-Q}^{Q} \begin{bmatrix} I_{1,1_q} \\ I_{2,1_q} \end{bmatrix} e^{j2\pi f_q t} = \frac{1}{2}\sum_{q=-Q}^{Q} \left[Y^{(1)}(f_q)\right] \begin{bmatrix} V_{1,1_q} \\ V_{2,1_q} \end{bmatrix} e^{j2\pi f_q t} \tag{3.12}$$

$$\frac{1}{2}\sum_{q=-Q}^{Q} \begin{bmatrix} I_{1,2_q} \\ I_{2,2_q} \end{bmatrix} e^{j2\pi f_q t} = \frac{1}{2}\sum_{q=-Q}^{Q} \left[Y^{(1)}(f_q)\right] \begin{bmatrix} V_{1,2_q} \\ V_{2,2_q} \end{bmatrix} e^{j2\pi f_q t} + \frac{1}{2^2}\sum_{q_1=-Q}^{Q}\sum_{q_2=-Q}^{Q} \left[Y^{(2)}(f_{q1}+f_{q2})\right] \begin{bmatrix} V_{1,1_{q1}}V_{1,1_{q2}} \\ V_{2,1_{q1}}V_{2,1_{q2}} \\ V_{1,1_{q1}}V_{2,1_{q2}} \end{bmatrix} e^{j(f_{q1}+f_{q2})t} \tag{3.13}$$

[7]First HOYPs represent the traditional linear Y-parameters.

$$\frac{1}{2}\sum_{q=-Q}^{Q}\begin{bmatrix}I_{1,3_q}\\ I_{2,3_q}\end{bmatrix}e^{j2\pi f_q t}=\frac{1}{2}\sum_{q=-Q}^{Q}\left[Y^{(1)}(f_q)\right]\begin{bmatrix}V_{1,2_q}\\ V_{2,2_q}\end{bmatrix}e^{j2\pi f_q t}$$
$$+\frac{1}{2^2}\sum_{q_1=-Q}^{Q}\sum_{q_2=-Q}^{Q}\left[Y^{(2)}(f_{q1}+f_{q2})\right]\begin{bmatrix}V_{1,1_{q1}}V_{1,2_{q2}}+V_{1,1_{q2}}V_{1,2_{q1}}\\ V_{2,1_{q1}}V_{2,2_{q2}}+V_{2,1_{q2}}V_{2,2_{q1}}\\ V_{1,1_{q1}}V_{2,2_{q2}}+V_{2,1_{q2}}V_{1,2_{q1}}\end{bmatrix}e^{j(f_{q1}+f_{q2})t} \tag{3.14}$$

$$+\frac{1}{2^3}\sum_{q_1=-Q}^{Q}\sum_{q_2=-Q}^{Q}\sum_{q_3=-Q}^{Q}\left[Y^{(3)}(f_{q1}+f_{q2}+f_{q3})\right]\begin{bmatrix}V_{1,1_{q1}}V_{1,1_{q2}}V_{1,1_{q3}}\\ V_{2,1_{q1}}V_{2,1_{q2}}V_{2,1_{q3}}\\ V_{1,1_{q1}}V_{1,1_{q2}}V_{2,1_{q3}}\\ V_{1,1_{q1}}V_{2,1_{q2}}V_{2,1_{q3}}\end{bmatrix}e^{j(f_{q1}+f_{q2}+f_{q3})t}$$

$$\left[Y^{(1)}(f)\right]=\begin{bmatrix}Y_{11}^{(1)}(f) & Y_{12}^{(1)}(f)\\ Y_{21}^{(1)}(f) & Y_{22}^{(1)}(f)\end{bmatrix} \tag{3.15}$$

$$\left[Y^{(2)}(f)\right]=\begin{bmatrix}Y_{11}^{(2)}(f) & Y_{12}^{(2)}(f) & Y_{13}^{(2)}(f)\\ Y_{21}^{(2)}(f) & Y_{22}^{(2)}(f) & Y_{23}^{(2)}(f)\end{bmatrix} \tag{3.16}$$

$$\left[Y^{(3)}(f)\right]=\begin{bmatrix}Y_{11}^{(3)}(f) & Y_{12}^{(3)}(f) & Y_{13}^{(3)}(f) & Y_{14}^{(3)}(f)\\ Y_{21}^{(3)}(f) & Y_{22}^{(3)}(f) & Y_{23}^{(3)}(f) & Y_{24}^{(3)}(f)\end{bmatrix} \tag{3.17}$$

Important Remarks

- Since HOYPs equivalent circuit (Figure 3.4) represents a Volterra system, one can notice the following similarities between the derivation above and the original Volterra series formulation mentioned in Section 2.3.1. First, both HONPs and Volterra formulations (2.3) describes the total WNL response of the system as a summation of different order responses. Second, the use of the multi-tone formulation helped in transforming the time-domain Volterra kernels into frequency-domain kernels. Similarly, multi-tone formulations along with the generalized quasi-static assumption helped in transforming the time-domain HOYPs into frequency-domain HOYPs. Again, the relationship between HOYPs and DIDO Volterra kernels can be found in Appendix A.
- Equation (3.15) represents the original Y-parameters for linear networks. The high-order parameters described by (3.16) and (3.17) are the ones responsible for all

second- and third-order WNL behaviours of the system. This means that the responses at the baseband and second-harmonic band are described by the same unique set of second-order network parameters. Also, the responses at the third-IMD in the fundamental band and the third-harmonic band are described by the same unique set of third-order network parameters.

- The first terms in (3.13) and (3.14) represent the effect of the linear network on the high-order responses, and they are important for two reasons. First, as mentioned in Section 3.2, the use of power sources in the simulation environment to stimulate the WNL system (3.4) will generate infinite-order tones at the input voltages. Second, in the measurement environment, even though HONPs will be written in A/B domain (as will be seen in the next subsection), infinite-order tones will still appear at the input A-wave due to spur tones generated by any real-life source and the mismatch at the DUT's ports.

- First-order Y-parameters describe the relationship between ports first-order currents and voltages only, i.e.

$$\begin{bmatrix} I_{1,n} \\ I_{2,n} \end{bmatrix} \neq \left[Y^{(1)}\right] \begin{bmatrix} V_{1,n} \\ V_{2,n} \end{bmatrix} \qquad : n > 1 \tag{3.18}$$

  Instead, one should use (3.12)-(3.14).

3.3.2 High-Order S-Parameters (HOSPs) Formulation

HOSPs can be found by describing the original WNL system (Figure 3.4) in terms of ports A and B-waves instead of ports voltages and currents. Following the same procedure in (3.4) - (3.11), first, second and third HOSPs ($[S^{(1)}(f)]$, $[S^{(2)}(f)]$, and $[S^{(3)}(f)]$) will be described by (3.19) - (3.24) in the frequency-domain. It should be noted that although HOSPs and HOYPs are mathematically equivalent, they are not when extracted in the measurement environment as will be shown in the coming sections.

$$\frac{1}{2}\sum_{q=-Q}^{Q} \begin{bmatrix} B_{1,1_q} \\ B_{2,1_q} \end{bmatrix} e^{j2\pi f_q t} = \frac{1}{2}\sum_{q=-Q}^{Q} \left[S^{(1)}(f_q)\right] \begin{bmatrix} A_{1,1_q} \\ A_{2,1_q} \end{bmatrix} e^{j2\pi f_q t} \tag{3.19}$$

$$\frac{1}{2}\sum_{q=-Q}^{Q} \begin{bmatrix} B_{1,2_q} \\ B_{2,2_q} \end{bmatrix} e^{j2\pi f_q t} = \frac{1}{2}\sum_{q=-Q}^{Q} \left[S^{(1)}(f_q)\right] \begin{bmatrix} A_{1,2_q} \\ A_{2,2_q} \end{bmatrix} e^{j2\pi f_q t}$$

$$+\frac{1}{2^2}\sum_{q_1=-Q}^{Q}\sum_{q_2=-Q}^{Q}\left[S^{(2)}(f_{q1}+f_{q2})\right]\begin{bmatrix}A_{1,1_{q1}}A_{1,1_{q2}}\\A_{2,1_{q1}}A_{2,1_{q2}}\\A_{1,1_{q1}}A_{2,1_{q2}}\end{bmatrix}e^{j(f_{q1}+f_{q2})t} \tag{3.20}$$

$$\frac{1}{2}\sum_{q=-Q}^{Q}\begin{bmatrix}B_{1,3_q}\\B_{2,3_q}\end{bmatrix}e^{j2\pi f_q t}=\frac{1}{2}\sum_{q=-Q}^{Q}\left[S^{(1)}(f_q)\right]\begin{bmatrix}A_{1,2_q}\\A_{2,2_q}\end{bmatrix}e^{j2\pi f_q t}$$
$$+\frac{1}{2^2}\sum_{q_1=-Q}^{Q}\sum_{q_2=-Q}^{Q}\left[S^{(2)}(f_{q1}+f_{q2})\right]\begin{bmatrix}A_{1,1_{q1}}A_{1,2_{q2}}+A_{1,1_{q2}}A_{1,2_{q1}}\\A_{2,1_{q1}}A_{2,2_{q2}}+A_{2,1_{q2}}A_{2,2_{q1}}\\A_{1,1_{q1}}A_{2,2_{q2}}+A_{2,1_{q2}}A_{1,2_{q1}}\end{bmatrix}e^{j(f_{q1}+f_{q2})t} \tag{3.21}$$

$$+\frac{1}{2^3}\sum_{q_1=-Q}^{Q}\sum_{q_2=-Q}^{Q}\sum_{q_3=-Q}^{Q}\left[S^{(3)}(f_{q1}+f_{q2}+f_{q3})\right]\begin{bmatrix}A_{1,1_{q1}}A_{1,1_{q2}}A_{1,1_{q3}}\\A_{2,1_{q1}}A_{2,1_{q2}}A_{2,1_{q3}}\\A_{1,1_{q1}}A_{1,1_{q2}}A_{2,1_{q3}}\\A_{1,1_{q1}}A_{2,1_{q2}}A_{2,1_{q3}}\end{bmatrix}e^{j(f_{q1}+f_{q2}+f_{q3})t}$$

$$\left[S^{(1)}(f)\right]=\begin{bmatrix}S^{(1)}_{11}(f) & S^{(1)}_{12}(f)\\S^{(1)}_{21}(f) & S^{(1)}_{22}(f)\end{bmatrix} \tag{3.22}$$

$$\left[S^{(2)}(f)\right]=\begin{bmatrix}S^{(2)}_{11}(f) & S^{(2)}_{12}(f) & S^{(2)}_{13}(f)\\S^{(2)}_{21}(f) & S^{(2)}_{22}(f) & S^{(2)}_{23}(f)\end{bmatrix} \tag{3.23}$$

$$\left[S^{(3)}(f)\right]=\begin{bmatrix}S^{(3)}_{11}(f) & S^{(3)}_{12}(f) & S^{(3)}_{13}(f) & S^{(3)}_{14}(f)\\S^{(3)}_{21}(f) & S^{(3)}_{22}(f) & S^{(3)}_{23}(f) & S^{(3)}_{24}(f)\end{bmatrix} \tag{3.24}$$

3.4 HONPs Extraction Techniques and Properties

In this section, different extraction techniques are investigated. The objective is to show that HONPs are a unique set of parameters that do not depend on the stimulus power nor type, nor port termination. HONPs only depend on the system and, once identified, can fully describe the WNL behaviour of it. In the simulation environment, HONPs will be extracted using CW HB simulations. These parameters will then be applied to predict the

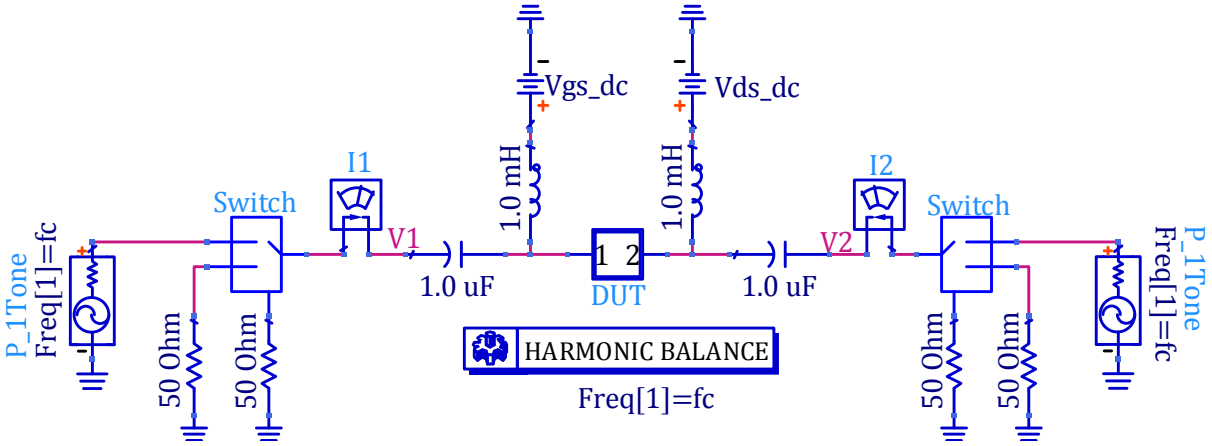


Figure 3.5: HONPs extraction setup in Keysight ADS using a CW HB simulation.

system response under multi-tone excitation in Envelope simulator (in the WNL region). A generic GaN HFET transistor model will be used as the DUT to mimic the real-life measurement scenario.

3.4.1 Extraction and Validation Setups in Simulation Environment

Keysight ADS software is used for HONPs extraction and validation. Figure 3.5 shows the extraction setup using a CW HB simulator, where power sources are used to excite port 1 and 2 simultaneously. A Verilog-A switch is implemented to allow only one source to be connected at a time while the other port is terminated by a 50 Ω load. Since HONPs extraction procedure requires many customized sweeps, as will be shown in the next subsections, MATLAB is utilized to control ADS simulations and collect the simulated data.

Figure 3.6 shows a multi-tone validation setup using Envelope simulator, where baseband I and Q files are uploaded to Data Access Components (DACs) then up-converted to the carrier frequency f_c using an IQ modulator. ADS is a circuit simulator, and hence it solves for ports voltages and currents. A and B-waves are then calculated using (3.3).

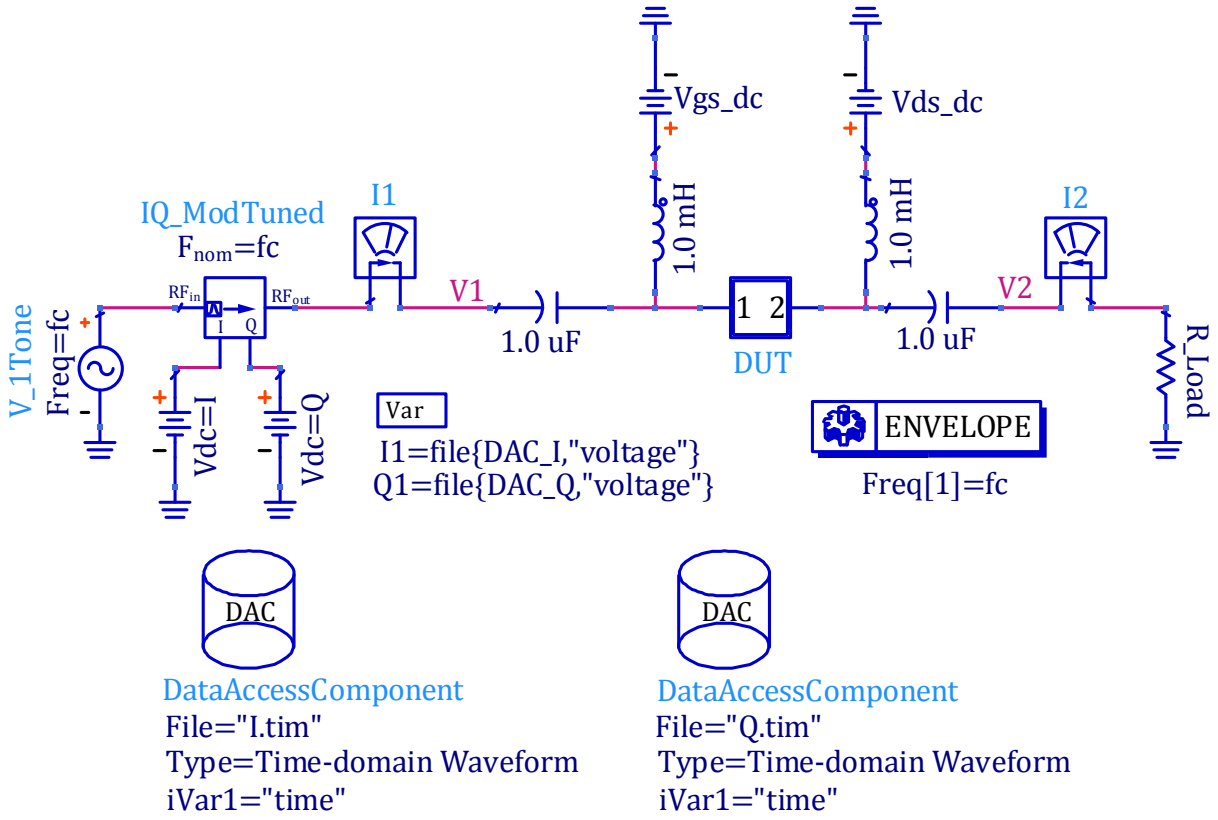


Figure 3.6: HONPs validation setup using Envelop simulation.

3.4.2 HONPs Extraction Techniques

Second and third HONPs are extracted sequentially using (3.13)-(3.14) or (3.20)-(3.21) in the Y or S form, respectively. While first-order parameters are extracted using two independent measurements, second-/third- order parameters need at least three/four independent measurements to be extracted. This is achieved by simultaneously stimulating both ports with a CW excitation and controlling the power or phase-difference between A-waves at ports 1 and 2 (A_1 and A_2) at the fundamental tone, as in Figure 3.7 (a) and (b), respectively. It is worth mentioning that the first two measurements in Figure 3.7 (a) are simply the ones used to extract traditional S-parameters (one port is excited at a time, while the other port is terminated by a 50 Ω load). This does not mean that S-parameters have to be extracted under 50 Ω terminations. It is just a practical method to measure them without the need for power calibration [in measurements 3 and 4 Figure 3.7 (a)] or NVNA measurements [in measurements 1 to 4 in Figure 3.7 (b)]. Controlling

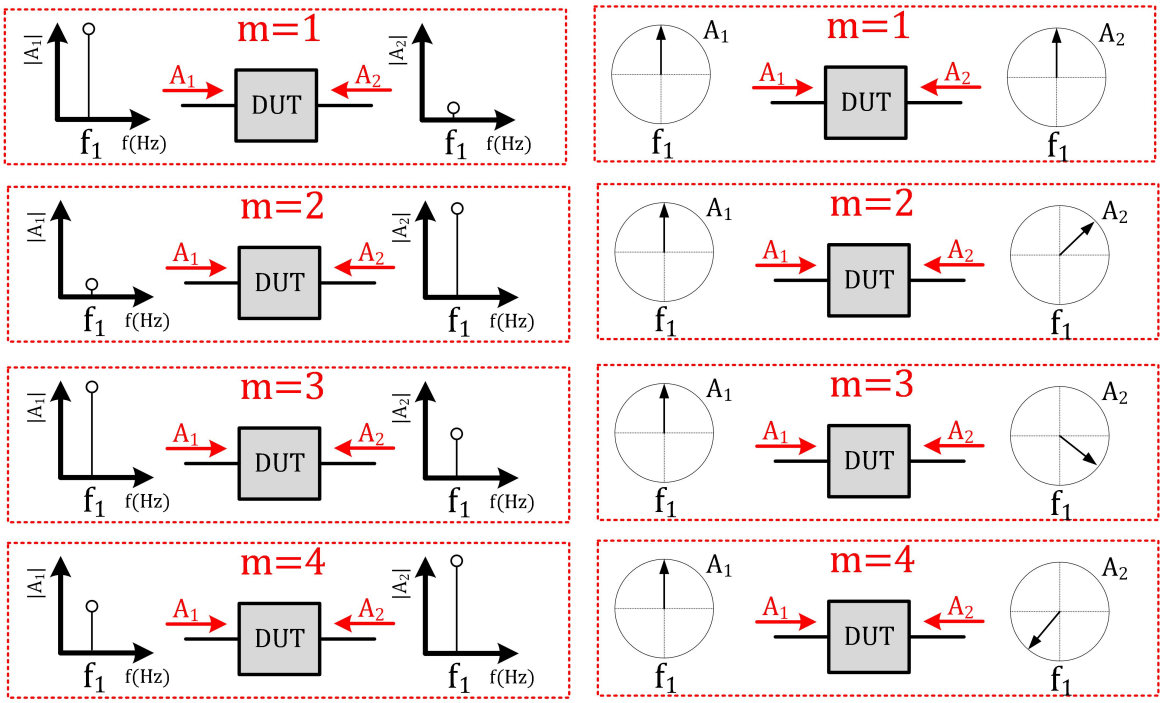


Figure 3.7: HONPs extraction techniques: (a) Sweeping the power of the fundamental tone between both sources, and (b) sweeping the phase-difference between the fundamental tones at both sources. A_1 and A_2 are plotted at the fundamental frequency as amplitudes only in (a) and in the polar form in (b). m is a measurement sweep index.

the phase-difference between A_1 and A_2 is preferred for the following reasons.

- More measurements are needed to make the system of equations in (3.13)-(3.14)/(3.20)-(3.21) more well-conditioned. As in the case of traditional network parameters, some elements are weaker than the others (e.g. S_{12} is much weaker than S_{21} in the case of a transistor) and an ill-conditioned system will result in false identification of these weak elements. While this might not be too problematic in the linear case, it is critical in the WNL case because we are interested in modelling a behaviour that is at least 40 dBc less than the dominant linear behaviour. It was found empirically that ten measurements represent a good compromise between HONPs extraction stability and measurement time. Achieving this number of measurements by controlling the power between the two ports is not practical due to dynamic range limitations of the

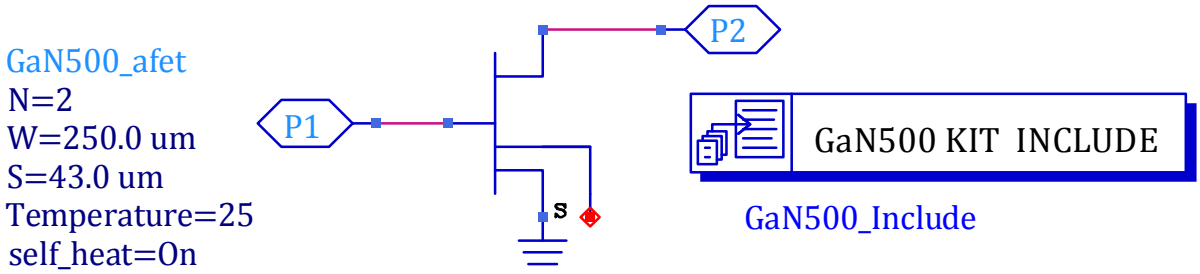


Figure 3.8: A generic GaN HFET model used as a DUT in the simulation environment.

measurement setup, the WNL extraction region (will be discussed in the following subsection), and the minimum step a real power source can achieve.

- HONPs are claimed to be power-independent (similar to traditional network parameters). Therefore, sweeping the power of the sources will be used to extract HONPs at different power levels (phase-difference between A_1 and A_2 will be used to extract them at each power level), and hence prove this property.

3.4.3 Simulated HONPs Properties: Power-Independency

A generic GaN HFET model (Figure 3.8) biased at (V_{gs},V_{ds}) of (-3.2, 12) V was used in the simulation environment as a DUT (Figure 3.5). Ten random phase-difference values were generated and first, second and third HONPs were extracted in both Y and S forms at 800 MHz fundamental frequency, as described in the previous subsection.

Linear network parameters extraction region is defined by the amount of compression in the fundamental tone. To measure S-parameters, for example, the signal has to be backed-off such that the maximum compression in the fundamental tone is 0.01 dB [26]. This means that setting the stimuli power to any value below this compression level will result in the same set of S-parameters, and this is why S-parameters are power independent. The same concept applies to HONPs; there is a specific extraction region where these parameters are power-independent. Figure 3.9 shows simulated A and B-waves at 800 MHz fundamental frequency versus power sweep (at one of the phase sweeps used for HONPs extraction). Both ports are simultaneously excited with a CW power source (A_1 and A_2-waves), and fundamental, second-harmonic, and third-harmonic tones are measured. In order to show how these tones compress against the stimuli power, the slopes are calculated with respect to A_1 fundamental tone power (Figure 3.10). Ideal HONPs extraction region is defined

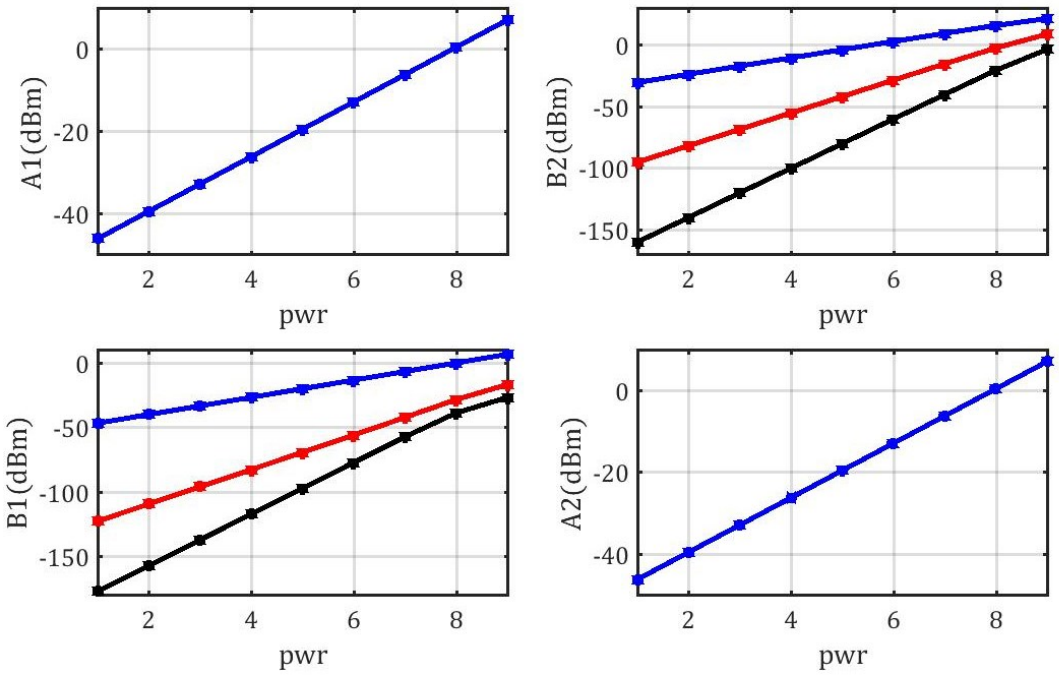
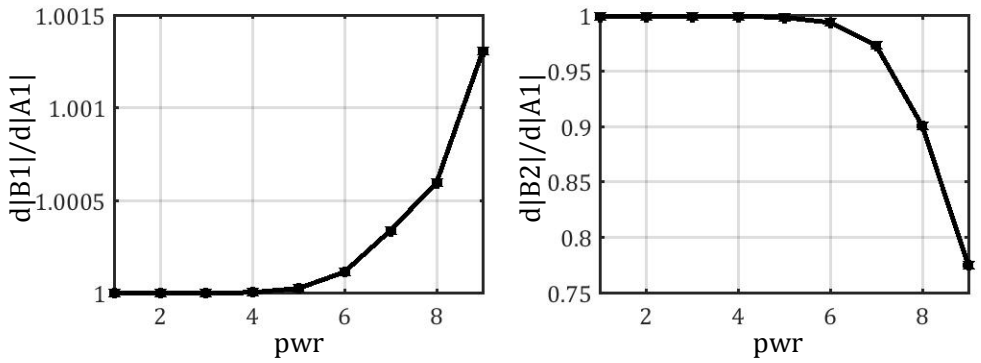


Figure 3.9: An example of simulated A and B-waves at 800 MHz fundamental frequency. Fundamental, second-harmonic, and third-harmonic tones are represented by blue, red, and black lines, respectively. The x-axis (pwr) represents the number of power sweep points at the same fundamental frequency and the same phase-difference between A_1 and A_2 fundamental tones.

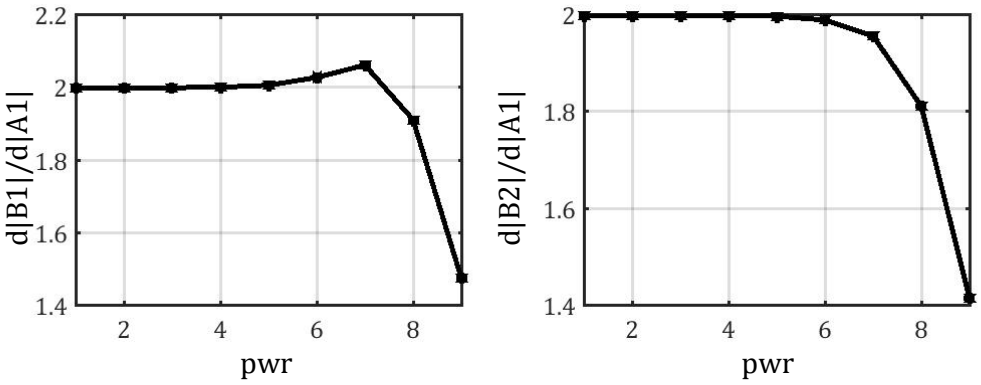
where fundamental, second-harmonic, and third-harmonic tones have slopes of one, two, and three with respect to the fundamental tone input power.

It should be noted that high-order tones compress faster than fundamental tones. It was found, in both simulation and measurement environments, that HONPs are best extracted with a maximum compression of 0.2 dB in the high-order tones. Finally, Figure 3.11 and 3.12 show the real and imaginary parts of the extracted first, second, and third HOYPs [8] and HOSPs versus the fundamental tone power of the input stimuli (A_1 and A_2). While these figures show a constant behaviour of HONPs against 30 dB range of input A_1 and A_2 power (the power-independence property), in real life, this range is very narrow because of the limited dynamic range of the measurement setup (will be shown in next sections).

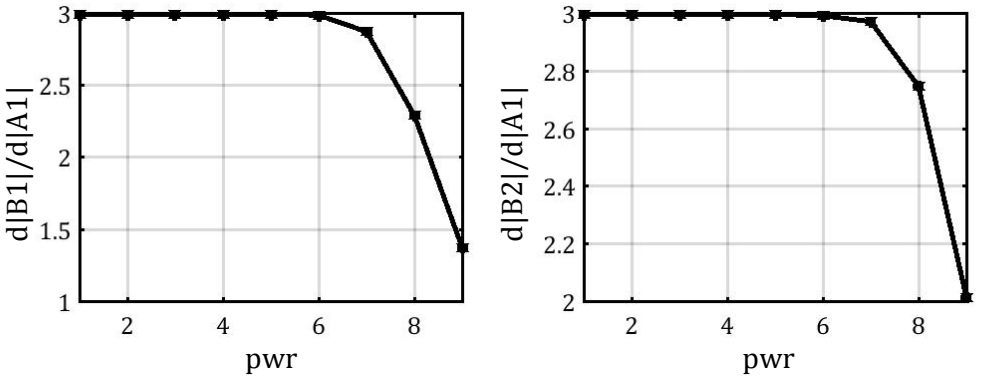
[8]While first HOYPs are measured in Ampere/Volt (A/V), second and third HOYPs are measured in A/V^2 and A/V^3, respectively. The units of HOSPs can be found similarly.



(a)

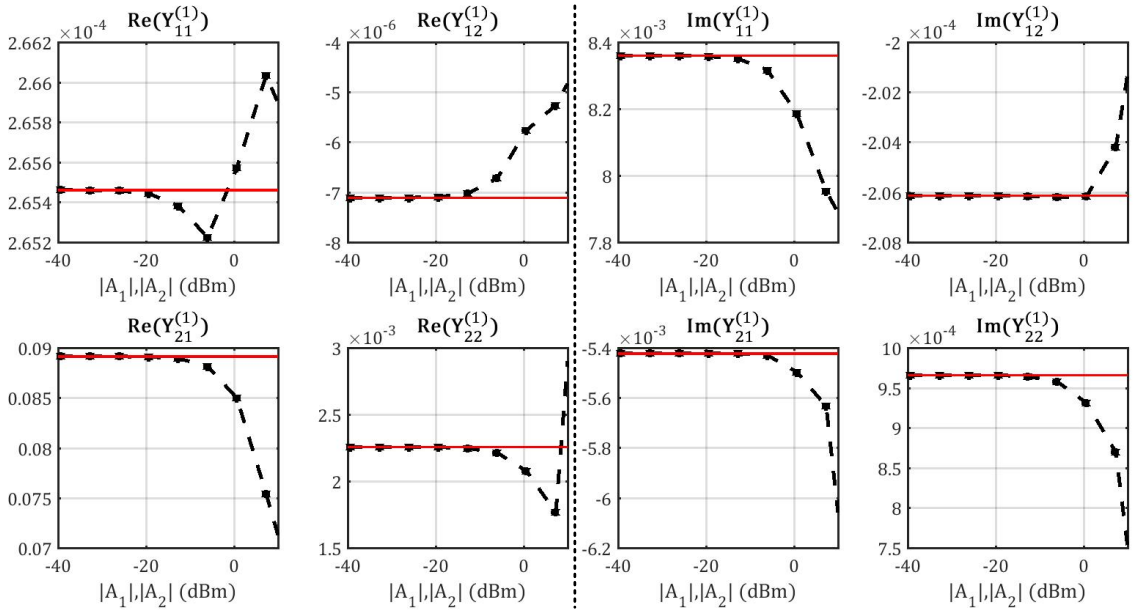


(b)

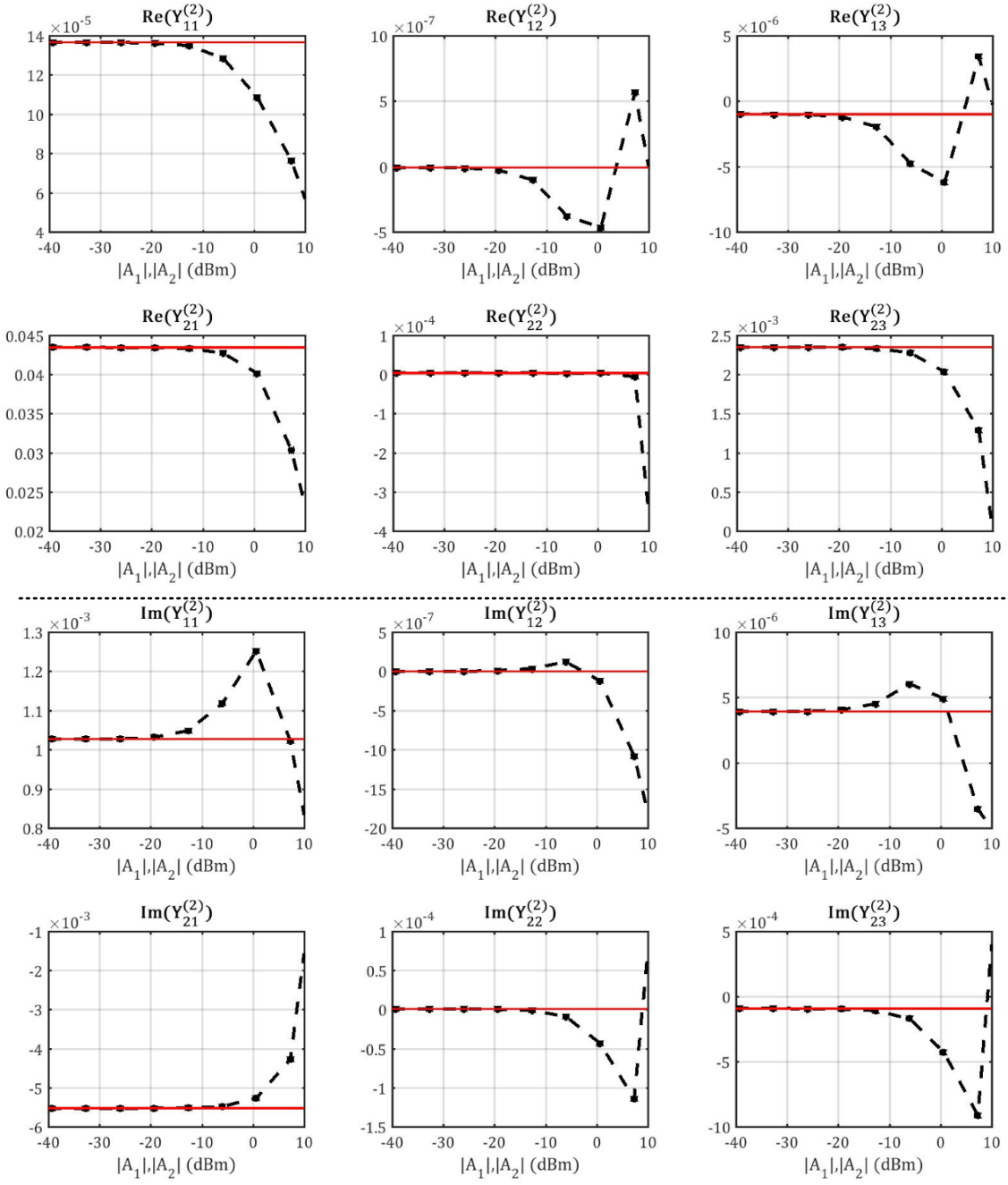


(c)

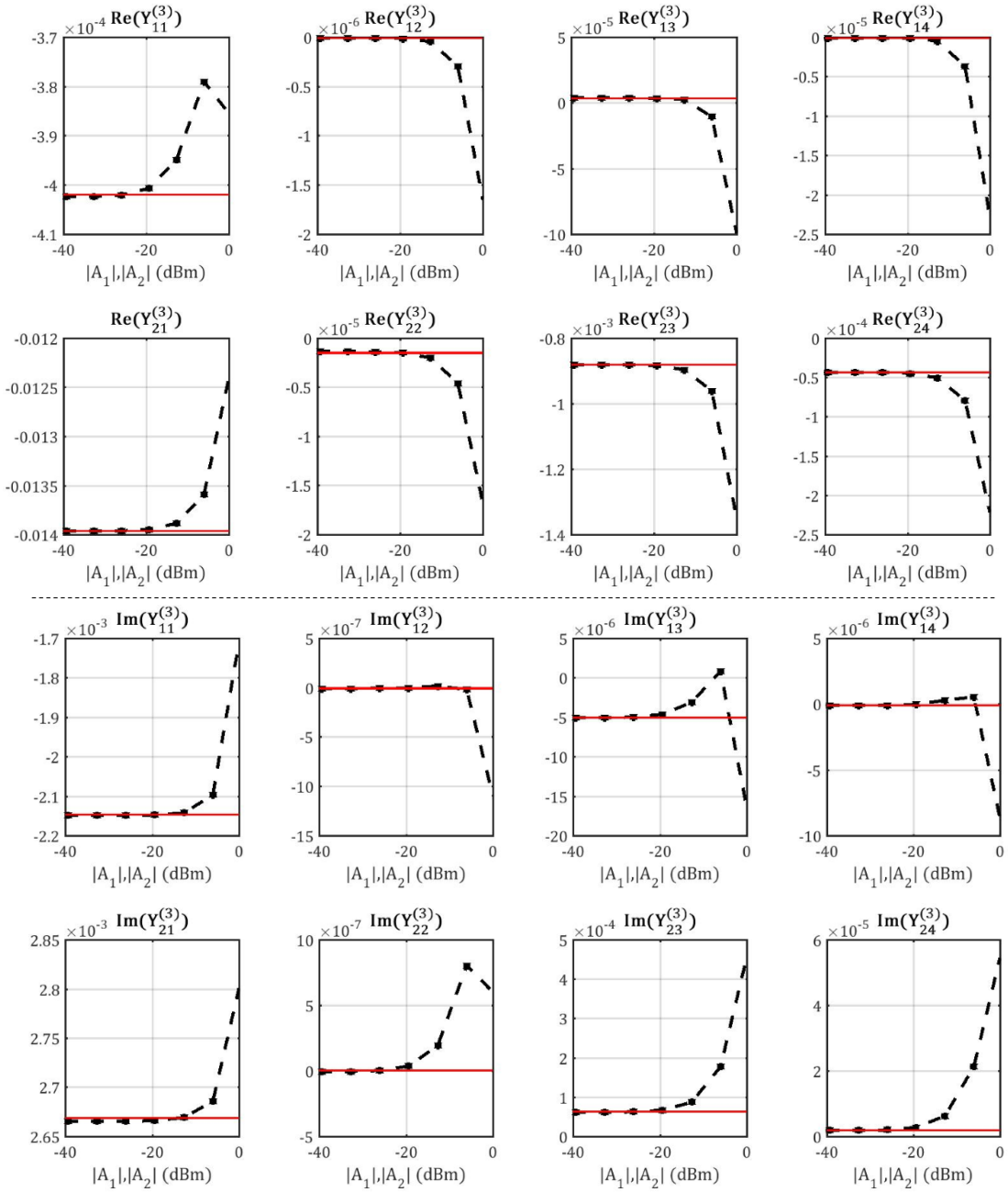
Figure 3.10: The slops of (a) fundamental tone, (b) second-harmonic tone, and (c) third-harmonic tone of both B_1 and B_2 with respect to A_1 fundamental tone power. The x-axis (*pwr*) represents the number of power sweep points at the same fundamental frequency and the same phase-difference between A_1 and A_2 fundamental tones.


Re(Y11(1))
×10-4
2.662
2.66
2.658
2.656
2.654
2.652
Re(Y12(1))
×10-6
-4
-5
-6
-7
-8
Im(Y11(1))
×10-3
8.4
8.2
8
7.8
Im(Y12(1))
×10-4
-2
-2.02
-2.04
-2.06
-2.08
Re(Y21(1))
0.09
0.085
0.08
0.075
0.07
Re(Y22(1))
×10-3
3
2.5
2
1.5
Im(Y21(1))
×10-3
-5.4
-5.6
-5.8
-6
-6.2
Im(Y22(1))
×10-4
10
9.5
9
8.5
8
7.5
-40
-20
0
|A1|,|A2| (dBm)

(a)

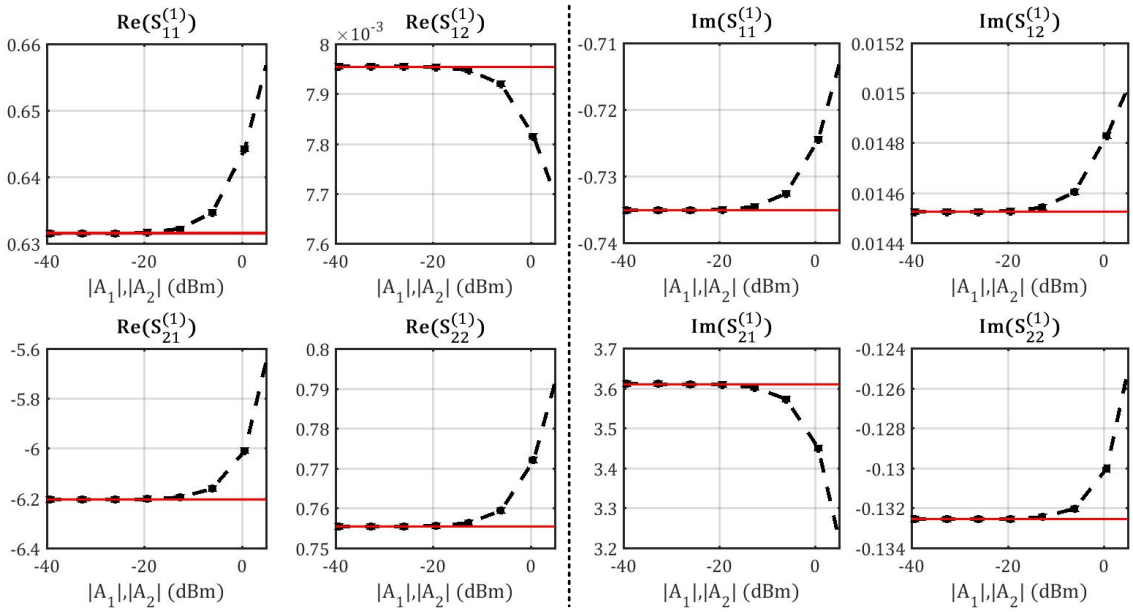

Re(Y(2)11)
Re(Y(2)12)
Re(Y(2)13)
Re(Y(2)21)
Re(Y(2)22)
Re(Y(2)23)
Im(Y(2)11)
Im(Y(2)12)
Im(Y(2)13)
Im(Y(2)21)
Im(Y(2)22)
Im(Y(2)23)
|A1|,|A2| (dBm)

(b)

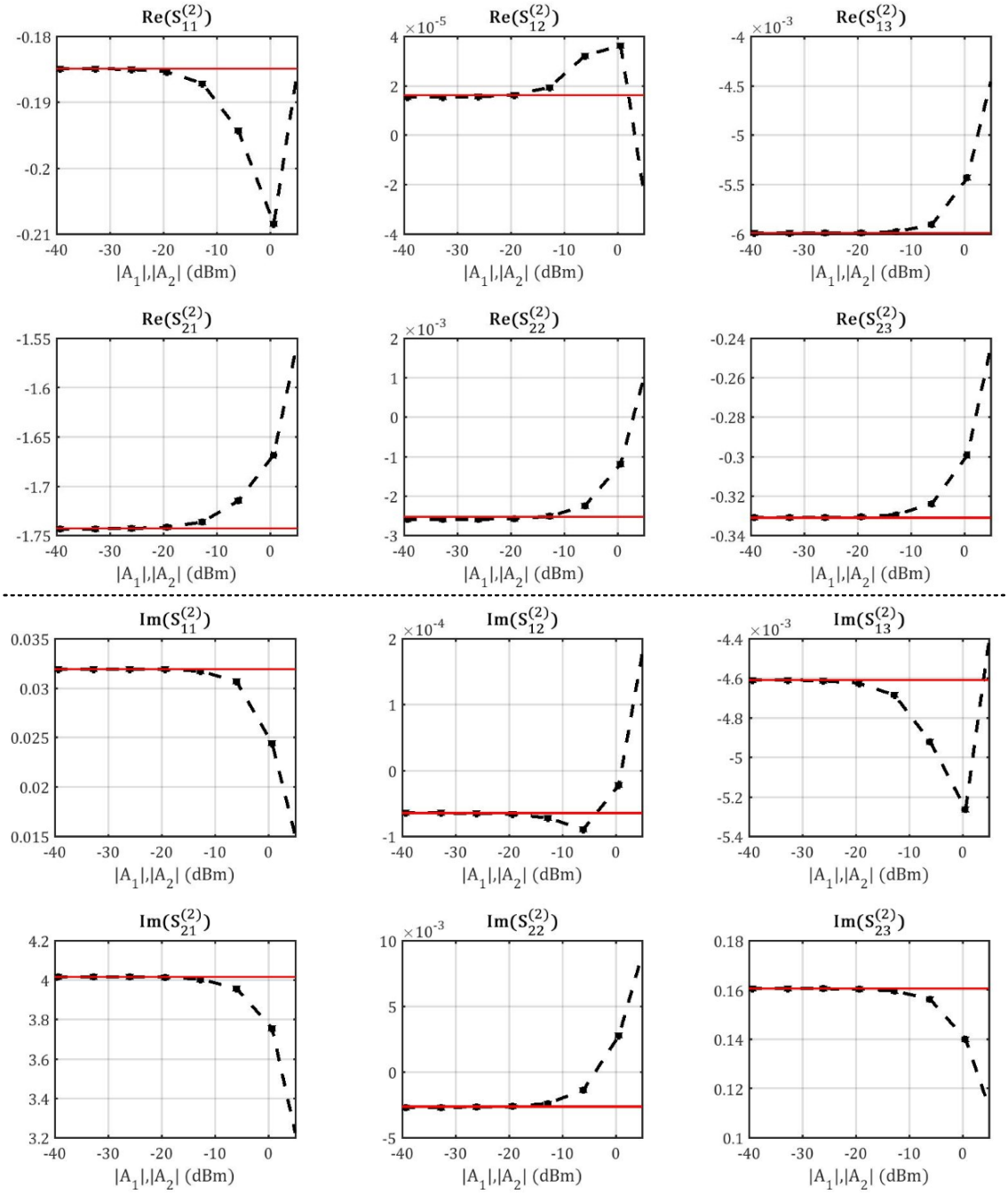


(c)

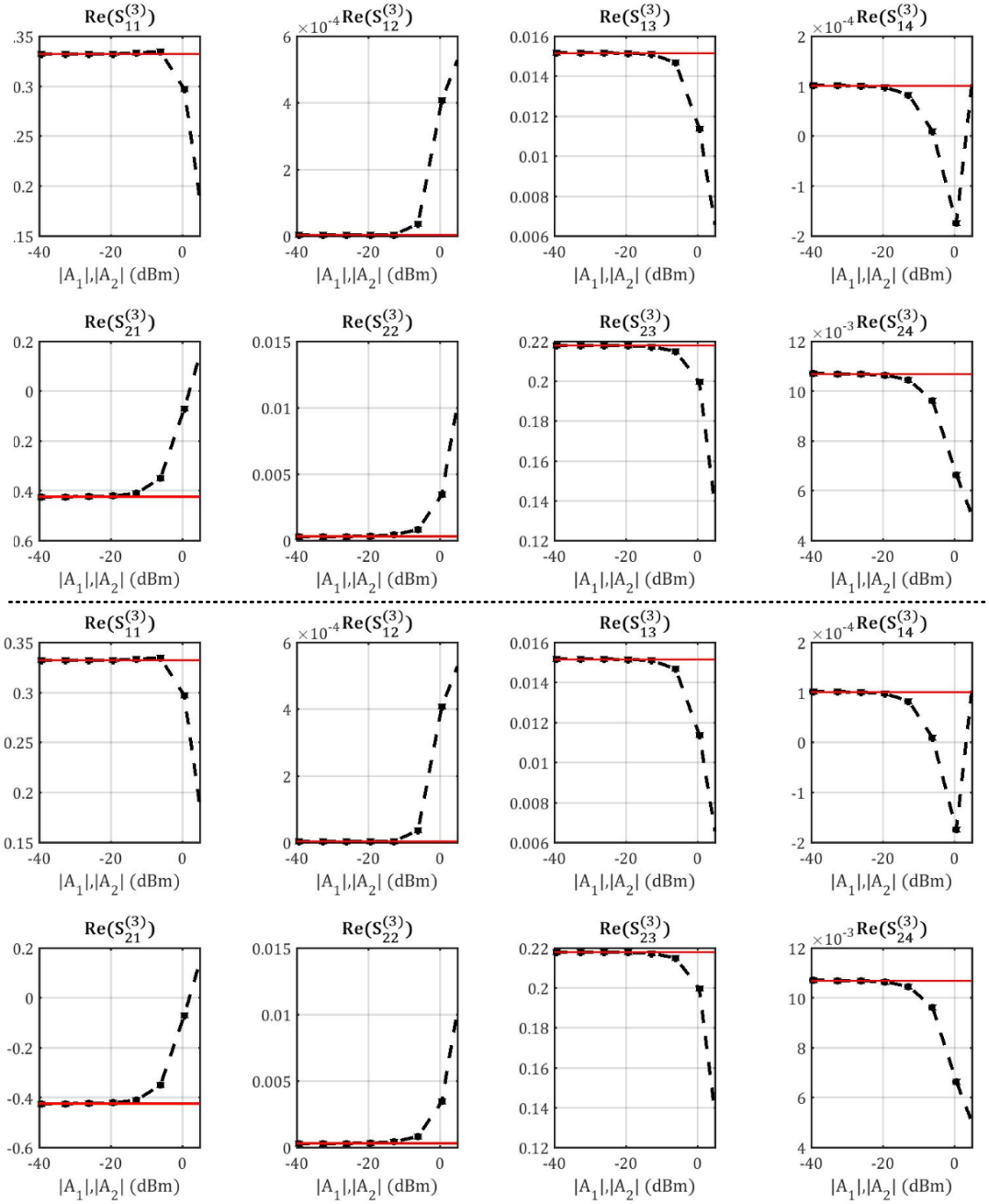
Figure 3.11: Extracted real and imaginary parts of (a) first, (b) second, and (c) third HOYPs at (V_{gs},V_{ds})=(-3.2, 12) V and 800 MHz fundamental frequency versus fundamental tone stimuli power. The red line represents the element value that will be used.


Re($S_{11}^{(1)}$)
Re($S_{12}^{(1)}$)
Im($S_{11}^{(1)}$)
Im($S_{12}^{(1)}$)
Re($S_{21}^{(1)}$)
Re($S_{22}^{(1)}$)
Im($S_{21}^{(1)}$)
Im($S_{22}^{(1)}$)
×10^-3
|A1|,|A2| (dBm)

(a)


Re($S_{11}^{(2)}$)
Re($S_{12}^{(2)}$)
Re($S_{13}^{(2)}$)
Re($S_{21}^{(2)}$)
Re($S_{22}^{(2)}$)
Re($S_{23}^{(2)}$)
Im($S_{11}^{(2)}$)
Im($S_{12}^{(2)}$)
Im($S_{13}^{(2)}$)
Im($S_{21}^{(2)}$)
Im($S_{22}^{(2)}$)
Im($S_{23}^{(2)}$)
$\times 10^{-5}$
$\times 10^{-3}$
$\times 10^{-4}$
$\times 10^{-3}$
$\times 10^{-3}$
$|A_1|,|A_2|$ (dBm)

(b)



(c)

Figure 3.12: Extracted real and imaginary parts of (a) first, (b) second, and (c) third HOSPs at (V_{gs},V_{ds})=(-3.2, 12) V and 800 MHz fundamental frequency versus fundamental tone stimuli power. The red line represents the element value that will be used.

3.4.4 Simulated HONPs Properties: Solution Continuity and Uniqueness

The same simulation setup described in the previous subsection was used here, where the fundamental frequency of both A_1 and A_2 was swept from 10 MHz to 6 GHz with a frequency step of 40 MHz. Figures 3.13 and 3.14 show extracted first, second, and third HOYPs and HOSPs, respectively. As mentioned in Subsection 3.4.2, more measurements are needed in order for the system of equations used in HONPs extraction to be more well-defined. This is demonstrated by extracting both HOYPs and HOSPs using four and ten phase-difference values between A_1 and A_2 fundamental frequency (red and black curves in Figure 3.13 and 3.14, respectively). It is obvious that the dominating elements (the ones with larger values) are not affected by increasing the number of measurements. However, the weaker elements need more measurements to be correctly identified. This can be demonstrated mathematically by monitoring the conditioning number of the system. Table 3.1 shows a comparison between the conditioning number of the system of linear equations used in extracting second and third HOSPs and HOYPs using four and ten measurements at 800 MHz fundamental frequency. It can be observed that increasing the number of measurements results in a better-conditioned system, and hence more trusted results. Yet, a more important observation is that extracting HONPs in the S form [using (3.20) - (3.21)] is always more reliable than extracting them in the Y form [using (3.13) - (3.14)] at microwave frequencies. This is because sources at microwave frequencies set incident power-waves, not voltages, and hence the voltage spectrum will contain tones at all the harmonic frequencies. These unwanted tones will make the second term in (3.13) and second and third terms in (3.14), more effective than those in (3.20) and (3.21). That is why the system conditioning number is significantly lower in the case of HOSPs than HOYPs. It is also clear that HOSPs are almost not affected by increasing the number of measurements (Figure 3.14). Have CW voltage sources been used instead of CW power sources in the extraction schematic (Figure 3.5), HOYPs would have lower conditioning number than HOSPs. However, this is an impractical case, and that is why one must pay attention to the concept of modelling domain discussed in Section 3.2. Finally, the formulations in (3.13) - (3.14) are linear with respect to the coefficients. This proves that there is a unique solution for that system of equations. Also, the fact that the conditioning number in Table 3.1 is very close to one indicates that the solution is stable.

To examine HONPs continuous behaviour across the biasing grid, V_{gs} and V_{ds} were swept from -4 to 0 V, and from 0 to 20 V, respectively. While a uniform step of 0.25 V was used for V_{gs} sweep, a non-uniform step was used to cover V_{ds} sweep with only ten points[9].

[9]Steps of 0.5 and 4 V were used to sweep V_{ds} from 0 to 2 V, and from 4 to 20 V, respectively.

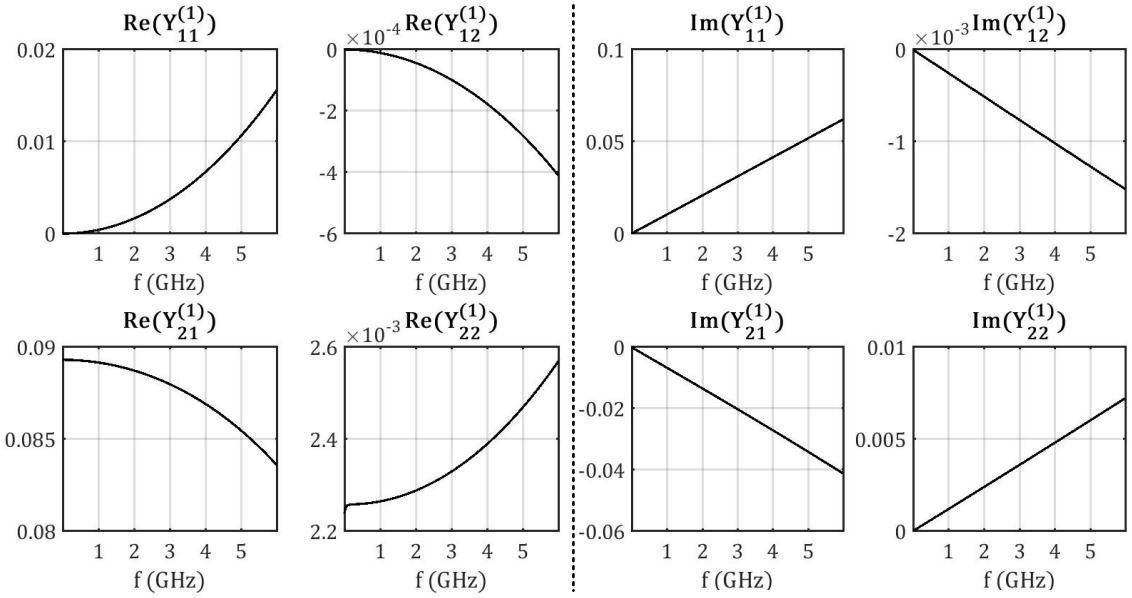
	4 measurements	10 measurements
Second HOSPs	8.7	1.79
Third HOSPs	431	2.36
Second HOYPs	461	275
Third HOYPs	1.6e5	5.6e3

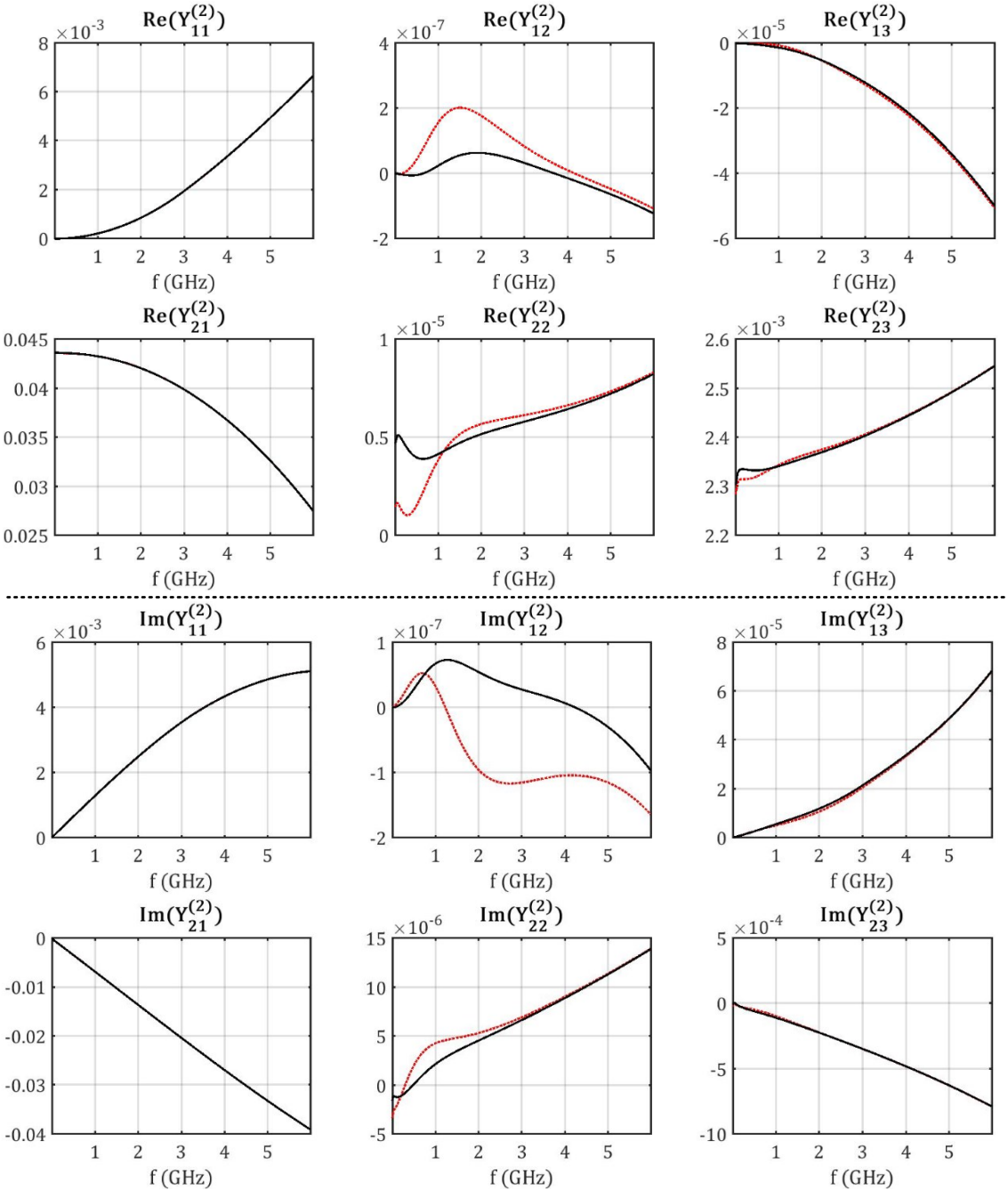
Table 3.1: A comparison between the conditioning number of the system of linear equations used to extract both second and third HOSPs and HOYPs using four and ten measurements in the simulation environment.

Figure 3.15 (a), (b), and (c) show the amplitude of the extracted first, second, and third HOSPs versus frequency, V_{gs}, and V_{ds}[10]. Specific planes are plotted across this 3D modelling space where the elements' amplitudes are represented using colour maps. The continuity of the behaviour can be observed from the smooth transitions between colours (elements' amplitudes) within a particular plane and also at the intersection between different planes. While this behaviour is expected from first HOSPs (traditional S-parameters), it is evident that high-order parameters exhibit the same continuous behaviour across the modelling space. Finally, Figure 3.15 (b), and (c) are of great importance as they sum up all second and third-order nonlinear[11] behaviours for this DUT (within the shown modelling space).

[10]Linear interpolations were used to refine both frequency and biasing grids, and only part of the modelling space is shown for figures clarity.

[11]The term nonlinear was used instead of WNL because nonlinear behaviours can be found using HONPs at different biasing points, as will be shown in Chapter 5.


$\mathbf{Re(Y_{11}^{(1)})}$
0.02
0.01
0
1 2 3 4 5
f (GHz)
$\times 10^{-4}$ $\mathbf{Re(Y_{12}^{(1)})}$
0
-2
-4
-6
1 2 3 4 5
f (GHz)
$\mathbf{Im(Y_{11}^{(1)})}$
0.1
0.05
0
1 2 3 4 5
f (GHz)
$\times 10^{-3}$ $\mathbf{Im(Y_{12}^{(1)})}$
0
-1
-2
1 2 3 4 5
f (GHz)
$\mathbf{Re(Y_{21}^{(1)})}$
0.09
0.085
0.08
1 2 3 4 5
f (GHz)
$\times 10^{-3}$ $\mathbf{Re(Y_{22}^{(1)})}$
2.6
2.4
2.2
1 2 3 4 5
f (GHz)
$\mathbf{Im(Y_{21}^{(1)})}$
0
-0.02
-0.04
-0.06
1 2 3 4 5
f (GHz)
$\mathbf{Im(Y_{22}^{(1)})}$
0.01
0.005
0
1 2 3 4 5
f (GHz)


Re(Y(2)11)
×10-3
f (GHz)
Re(Y(2)12)
×10-7
f (GHz)
Re(Y(2)13)
×10-5
f (GHz)
Re(Y(2)21)
f (GHz)
Re(Y(2)22)
×10-5
f (GHz)
Re(Y(2)23)
×10-3
f (GHz)
Im(Y(2)11)
×10-3
f (GHz)
Im(Y(2)12)
×10-7
f (GHz)
Im(Y(2)13)
×10-5
f (GHz)
Im(Y(2)21)
f (GHz)
Im(Y(2)22)
×10-6
f (GHz)
Im(Y(2)23)
×10-4
f (GHz)

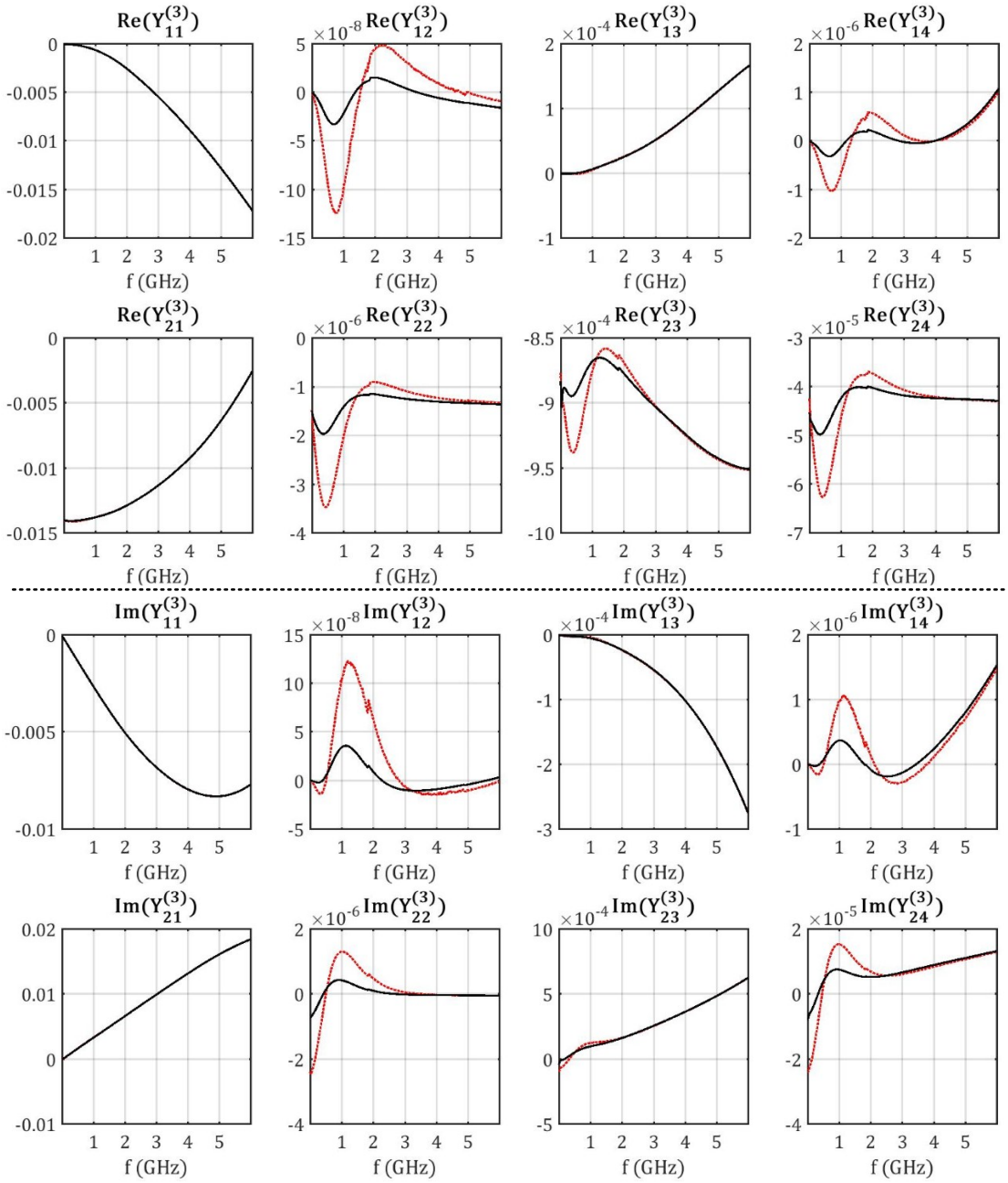
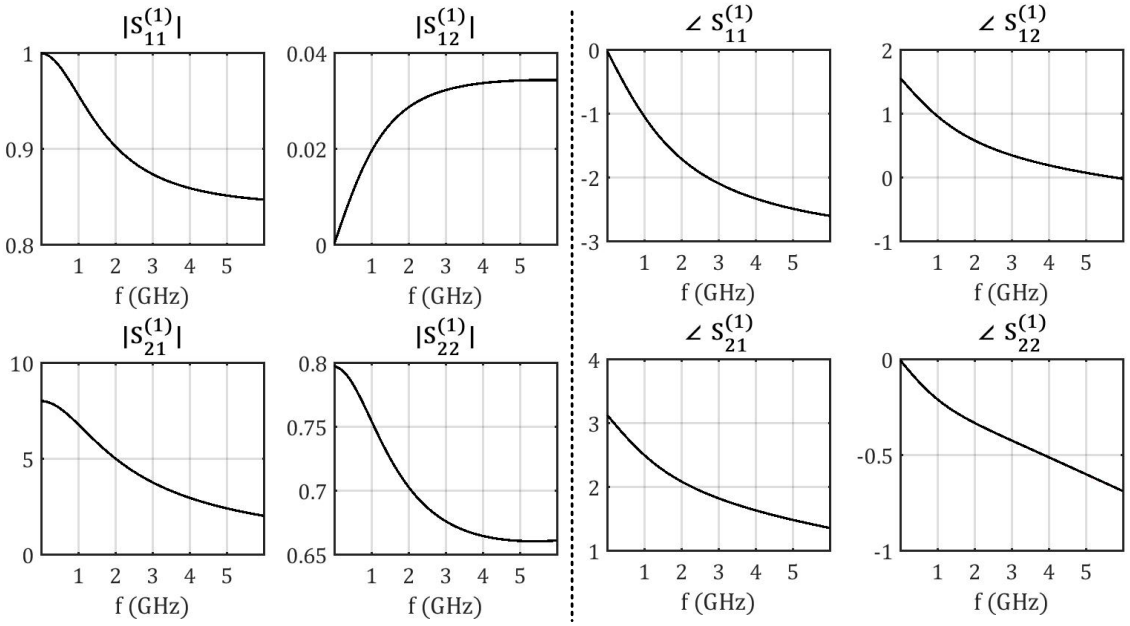
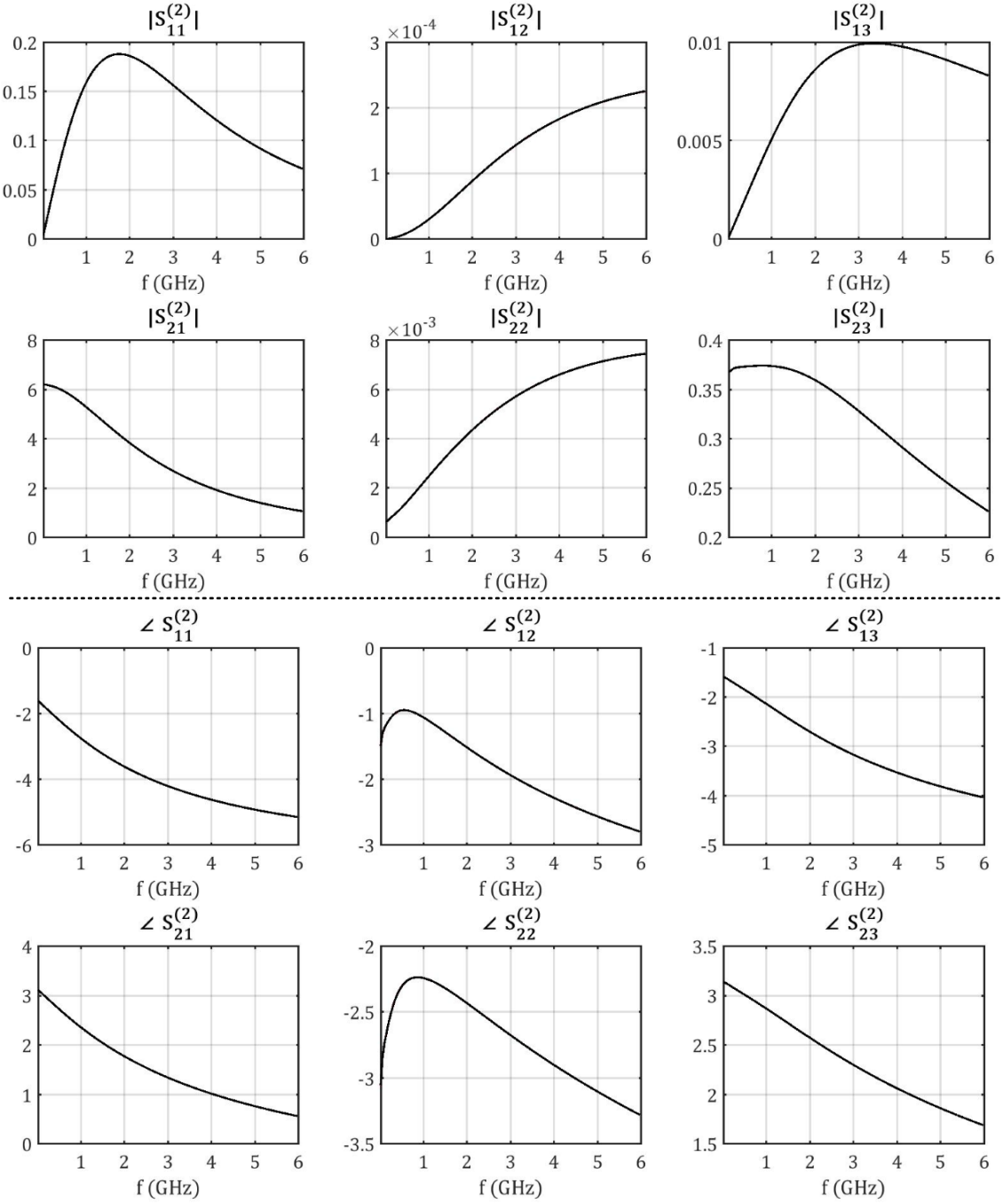


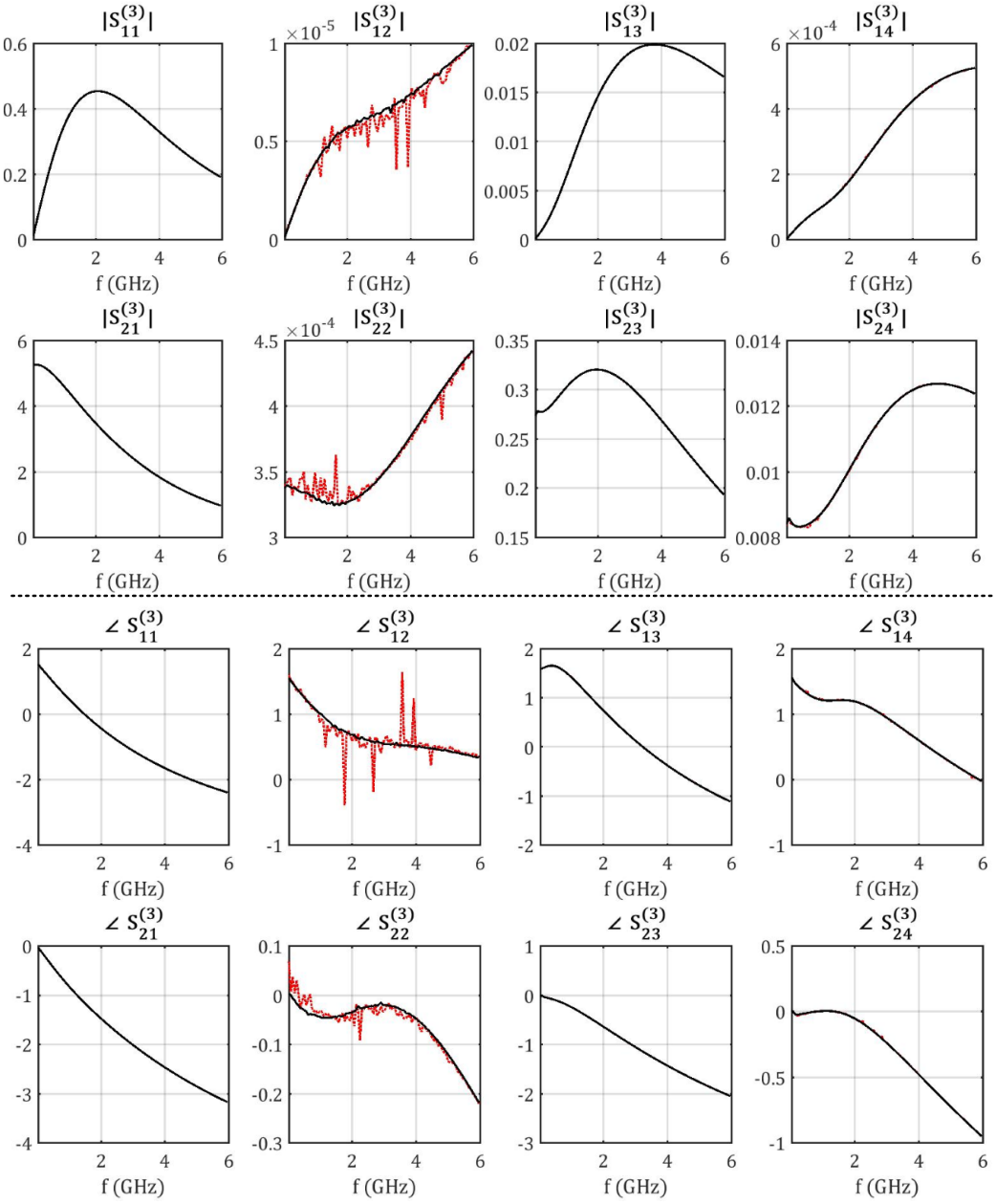
Figure 3.13: Extracted real and imaginary parts of (a) first, (b) second, and (c) third HOYPs at (V_{gs},V_{ds})=(-3.2, 12) V versus frequency. The red and black curves show the extracted elements using four and ten phase-difference values between A_1 and A_2 fundamental tones, respectively.


$|S_{11}^{(1)}|$
$|S_{12}^{(1)}|$
$\angle S_{11}^{(1)}$
$\angle S_{12}^{(1)}$
$|S_{21}^{(1)}|$
$|S_{22}^{(1)}|$
$\angle S_{21}^{(1)}$
$\angle S_{22}^{(1)}$
f (GHz)

(a)

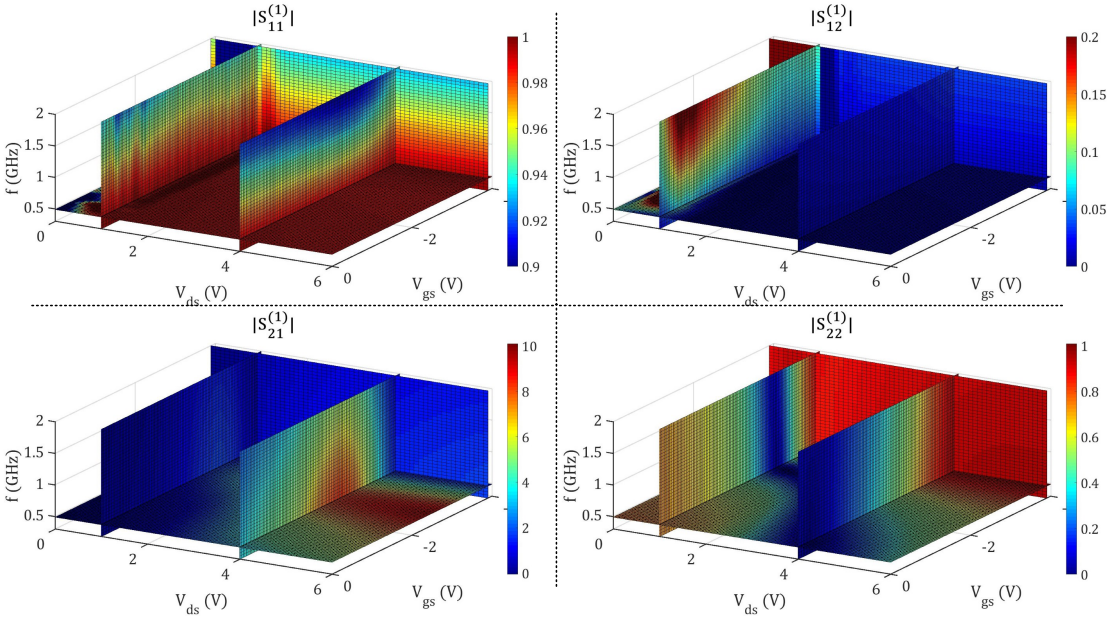

$|S_{11}^{(2)}|$
$|S_{12}^{(2)}|$
$|S_{13}^{(2)}|$
$|S_{21}^{(2)}|$
$|S_{22}^{(2)}|$
$|S_{23}^{(2)}|$
$\angle S_{11}^{(2)}$
$\angle S_{12}^{(2)}$
$\angle S_{13}^{(2)}$
$\angle S_{21}^{(2)}$
$\angle S_{22}^{(2)}$
$\angle S_{23}^{(2)}$
f (GHz)
$\times 10^{-4}$
$\times 10^{-3}$

(b)

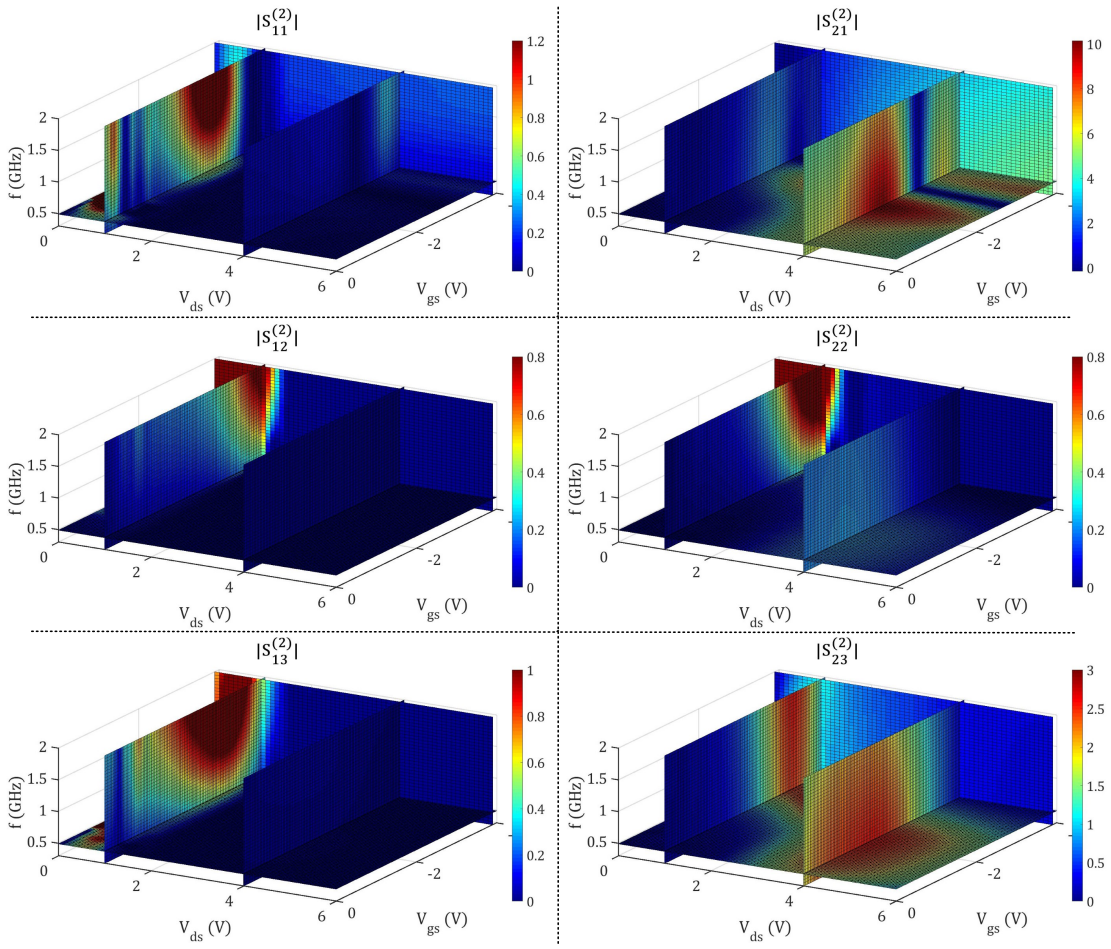


(c)

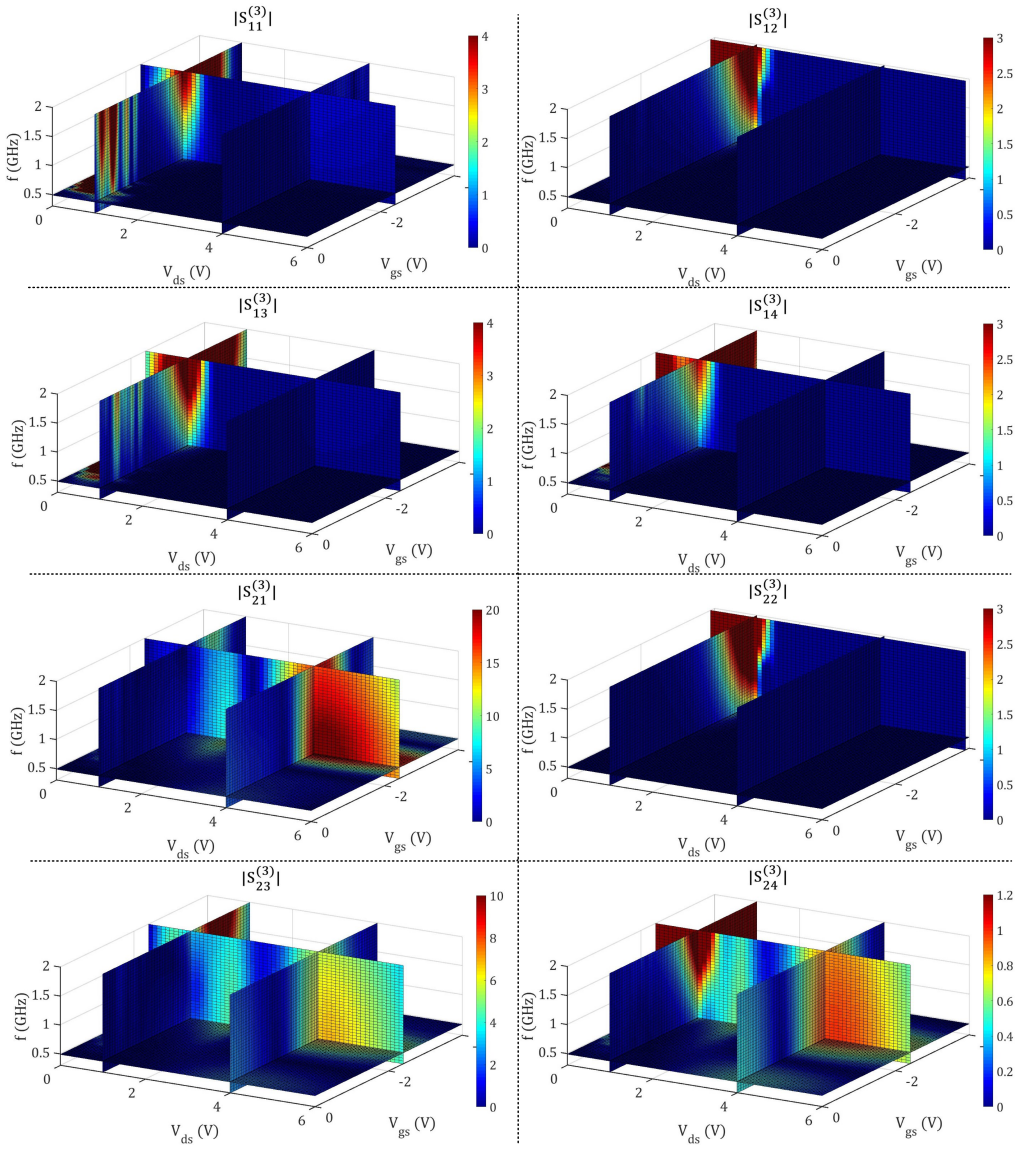
Figure 3.14: Extracted real and imaginary parts of (a) first, (b) second, and (c) third HOSPs at (V_{gs},V_{ds})=(-3.2, 12) V versus frequency. The red and black curves show the extracted elements using four and ten phase-difference values between A_1 and A_2 fundamental tones, respectively.



(a)



(b)



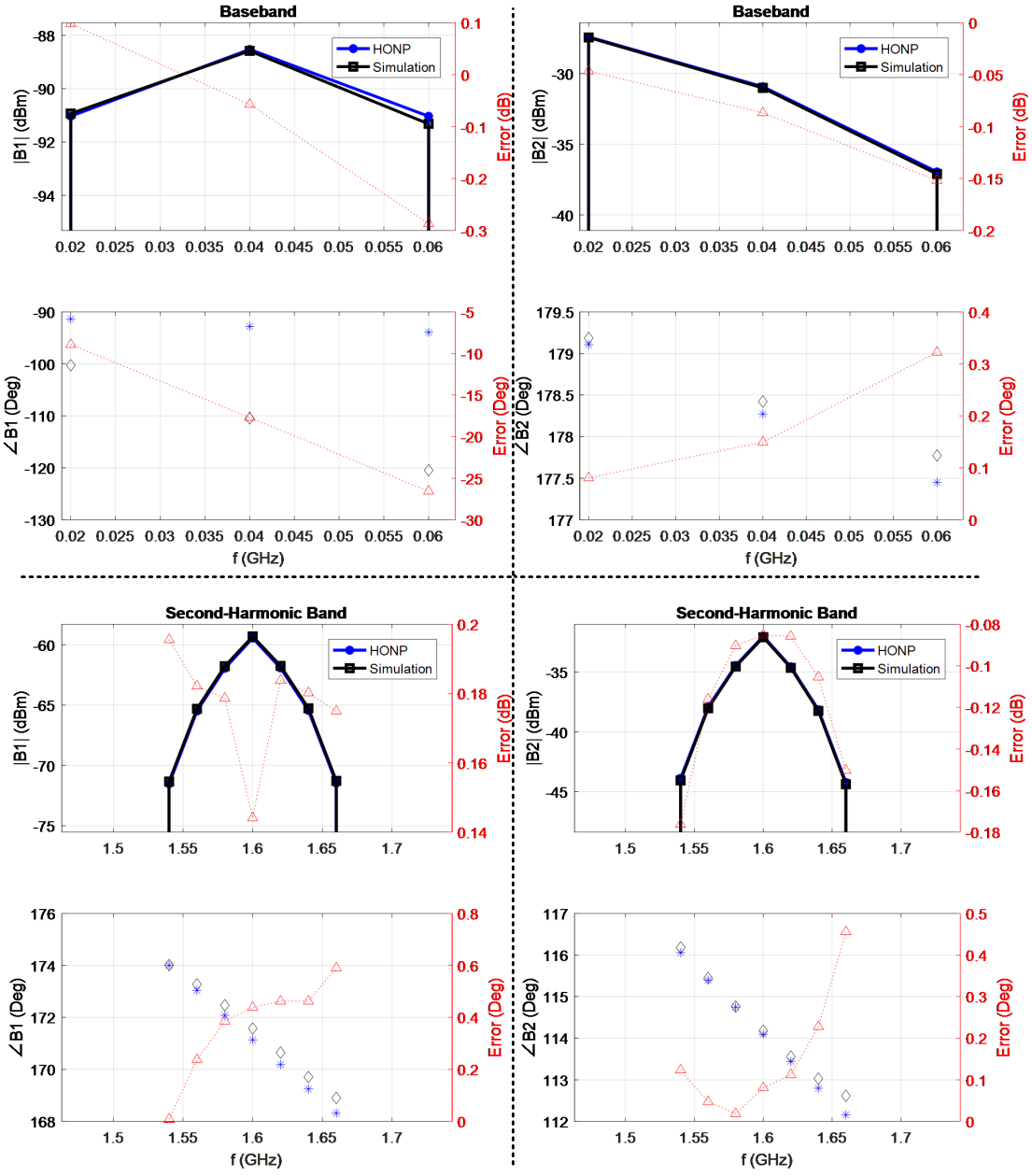
(c)

Figure 3.15: Amplitude (represent by the colour map) of the extracted: (a) linear S-parameters, (b) second HOSPs, and (c) third HOSPs versus frequency, V_{gs}, and V_{ds} in the simulation environment. These particular planes are shown to prove HOSPs continuity across the modelling space.

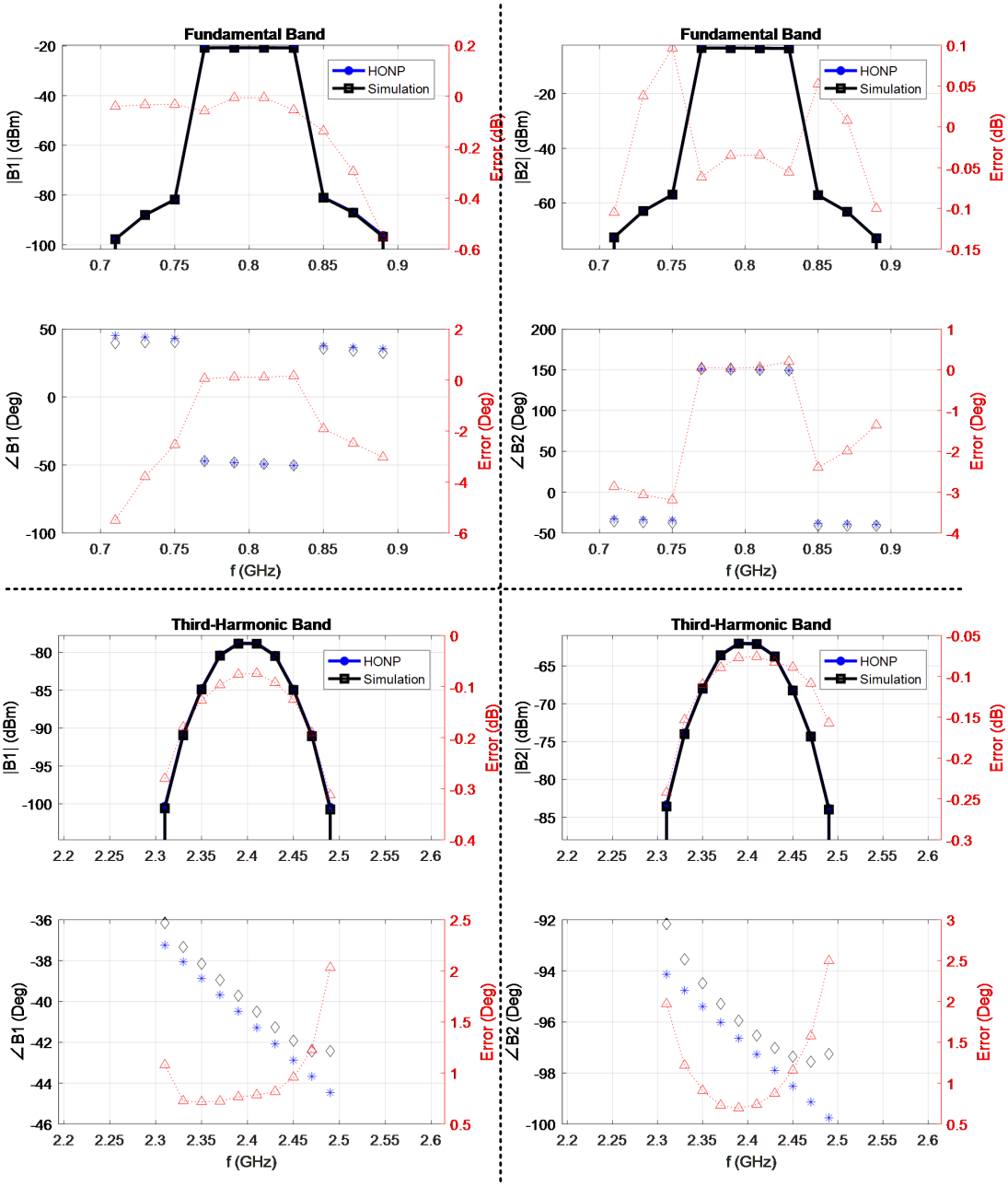
3.4.5 Prediction of WNL Response Due To Multi-tone Excitation Using HONPs

In order to verify that HONPs [12] can predict WNL behaviours, the same DUT (Figure 3.8) biased at (V_{gs},V_{ds}) of (-3.2, 12) V was stimulated with a 20 MHz spaced 4-tone signal around 800 MHz under Envelope simulator (Figure 3.6). Simulated A_1 and A_2 and extracted HOSPs (at the same biasing point) were then used in (3.20) and (3.21) to predict second-order behaviours (baseband and second-harmonic band), and third-order behaviours (third-IMD in the fundamental band and third-harmonic band). Figure 3.16 (a) and (b) shows simulated and predicted second- and third-order behaviours for both B_1 and B_2, due to a 4-tone excitation at 2.6 dBm average input power and 6.02 dB PAPR. HONPs demonstrated an excellent ability in WNL behaviours predictions where a maximum amplitude error of 0.5 dB and a few degrees phase error can be observed. The average input power was then increased to 8.6 dBm and simulated and predicted second- and third-order B_1 and B_2 behaviours were plotted in Figure 3.17. Although the average input power was 6 dB more than the first case, the error in amplitude prediction increased to 1 dB. These results confirm that HONPs are power independent. Figure 3.18 shows a case where the average input power was further increased to 10.5 dBm. It is obvious that HONPs at this biasing point could not predict the WNL behaviours of the DUT. This is because the WNL assumption is not valid anymore and a new set of HONPs at a different biasing point was excited. It is also worth noting that the first-order behaviour (represented in fundamental tones) was also badly predicted, which again means that a different set of linear S-parameters was excited. The use of multiple HONPs at different biasing points to predict strong nonlinear behaviours will be discussed in Chapter 5.

[12] Modulated signal validations and linearization results are provided in Appendix B.

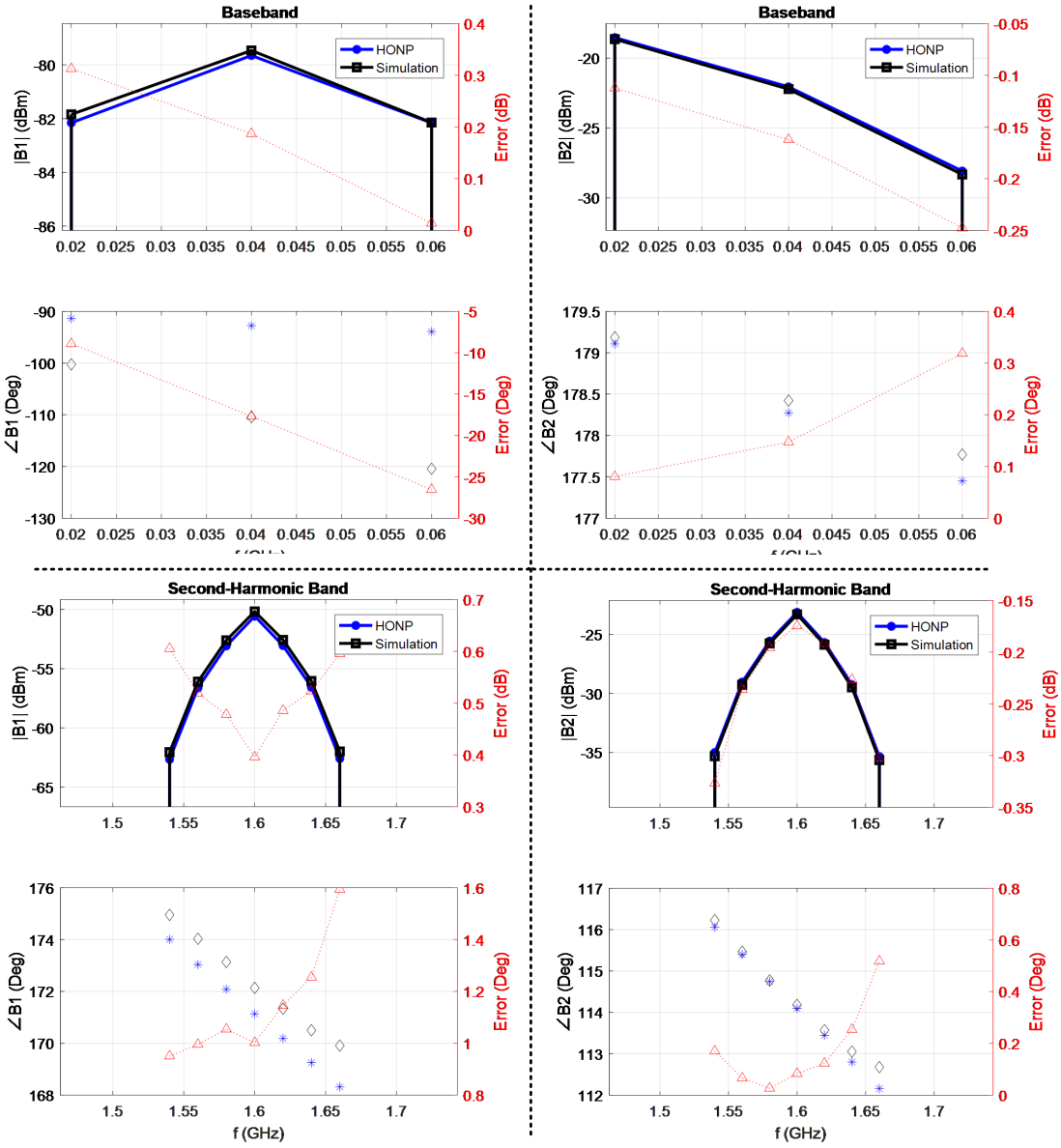

Baseband
HONP
Simulation
|B1| (dBm)
Error (dB)
Baseband
HONP
Simulation
|B2| (dBm)
Error (dB)
∠B1 (Deg)
Error (Deg)
f (GHz)
∠B2 (Deg)
Error (Deg)
f (GHz)
Second-Harmonic Band
HONP
Simulation
|B1| (dBm)
Error (dB)
Second-Harmonic Band
HONP
Simulation
|B2| (dBm)
Error (dB)
∠B1 (Deg)
Error (Deg)
f (GHz)
∠B2 (Deg)
Error (Deg)
f (GHz)

(a)

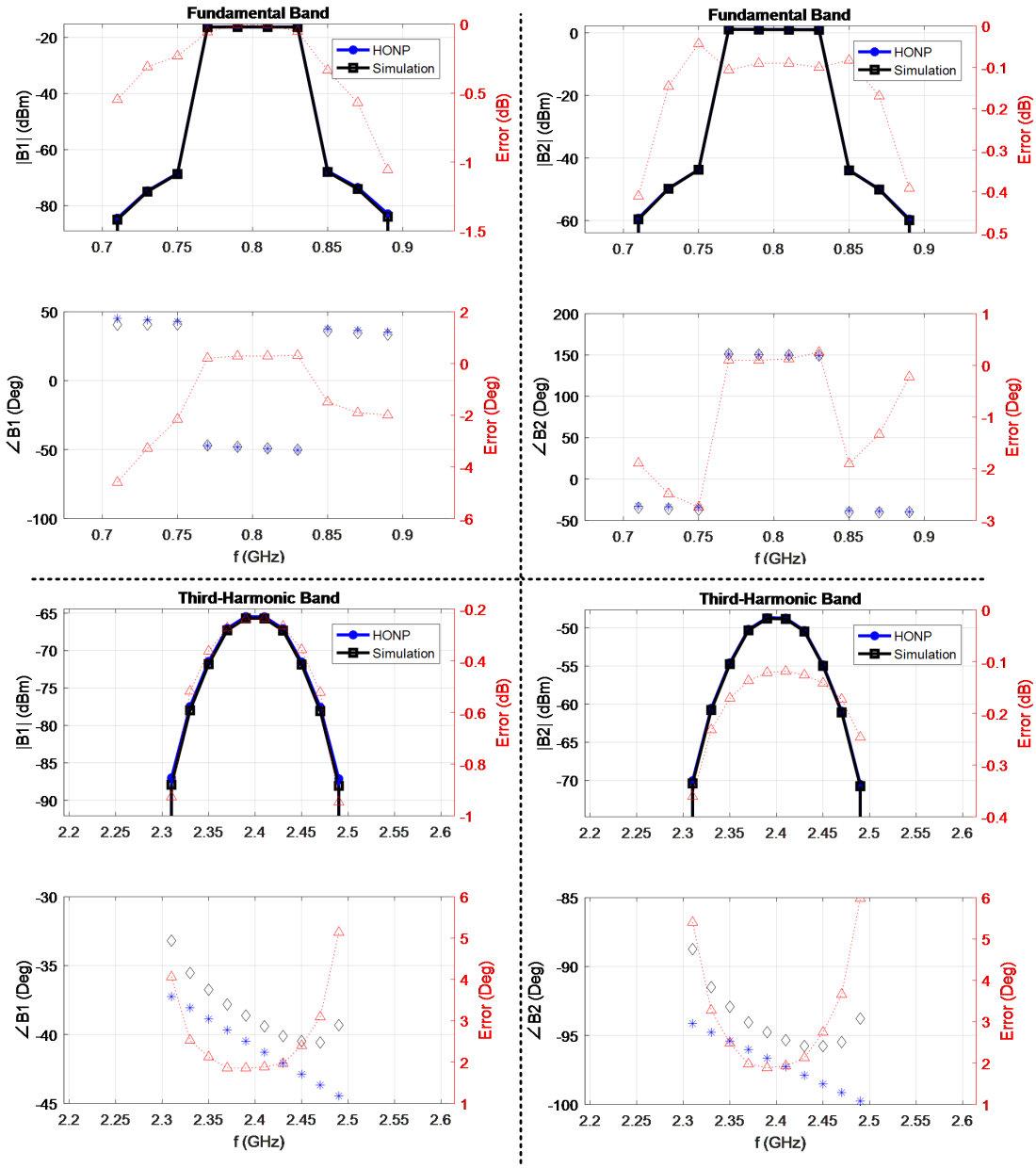


(b)

Figure 3.16: Simulated and HONPs predicted: (a) second-order behaviours, and (b) third-order behaviours for both B_1 and B_2. The stimulus is a 4-tone signal at 2.6 dBm average input power and 6.02 dB PAPR.

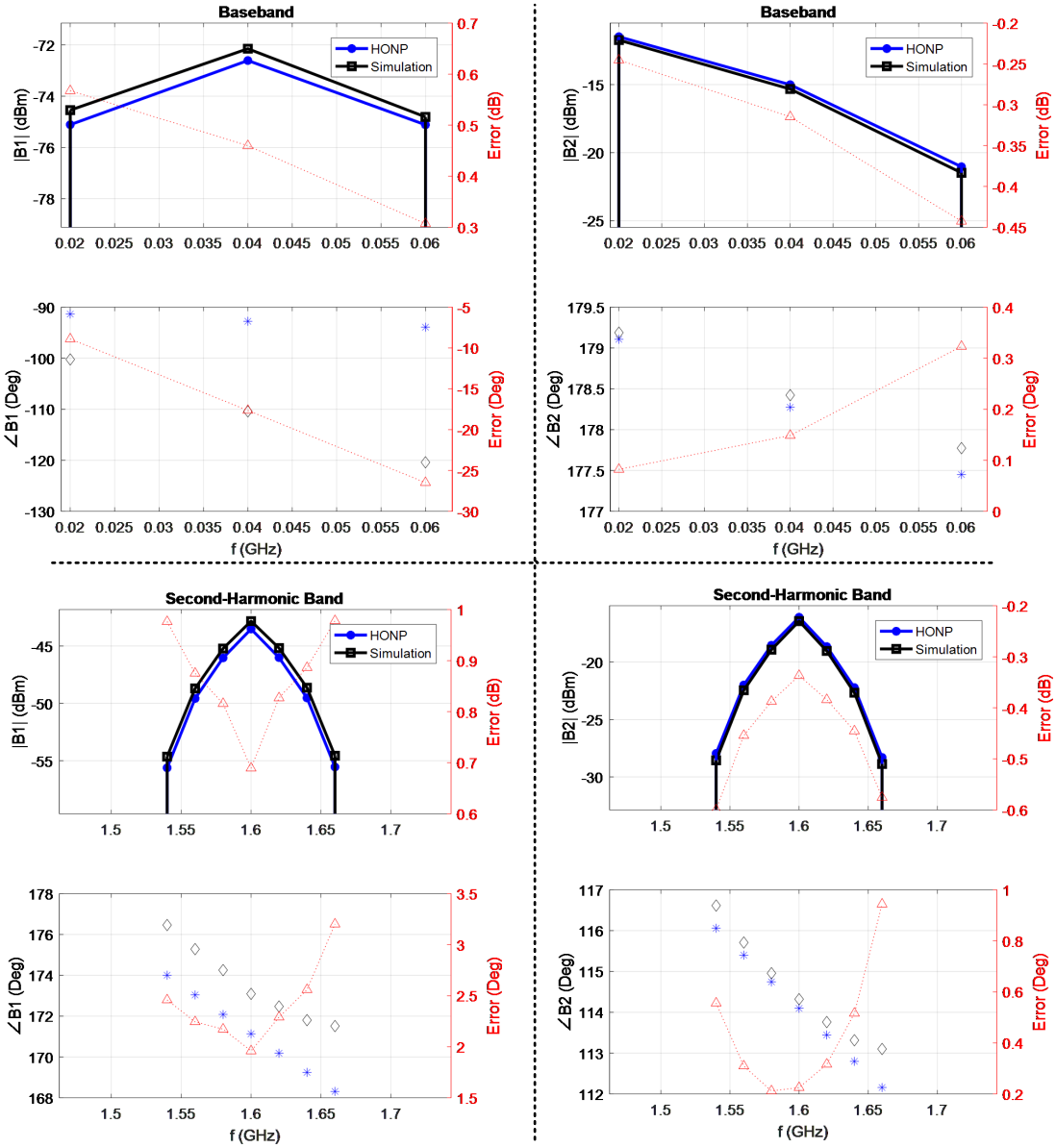

Baseband
HONP
Simulation
|B1| (dBm)
Error (dB)
Baseband
HONP
Simulation
|B2| (dBm)
Error (dB)
∠B1 (Deg)
Error (Deg)
∠B2 (Deg)
Error (Deg)
Second-Harmonic Band
HONP
Simulation
|B1| (dBm)
Error (dB)
Second-Harmonic Band
HONP
Simulation
|B2| (dBm)
Error (dB)
∠B1 (Deg)
Error (Deg)
f (GHz)
∠B2 (Deg)
Error (Deg)
f (GHz)

(a)

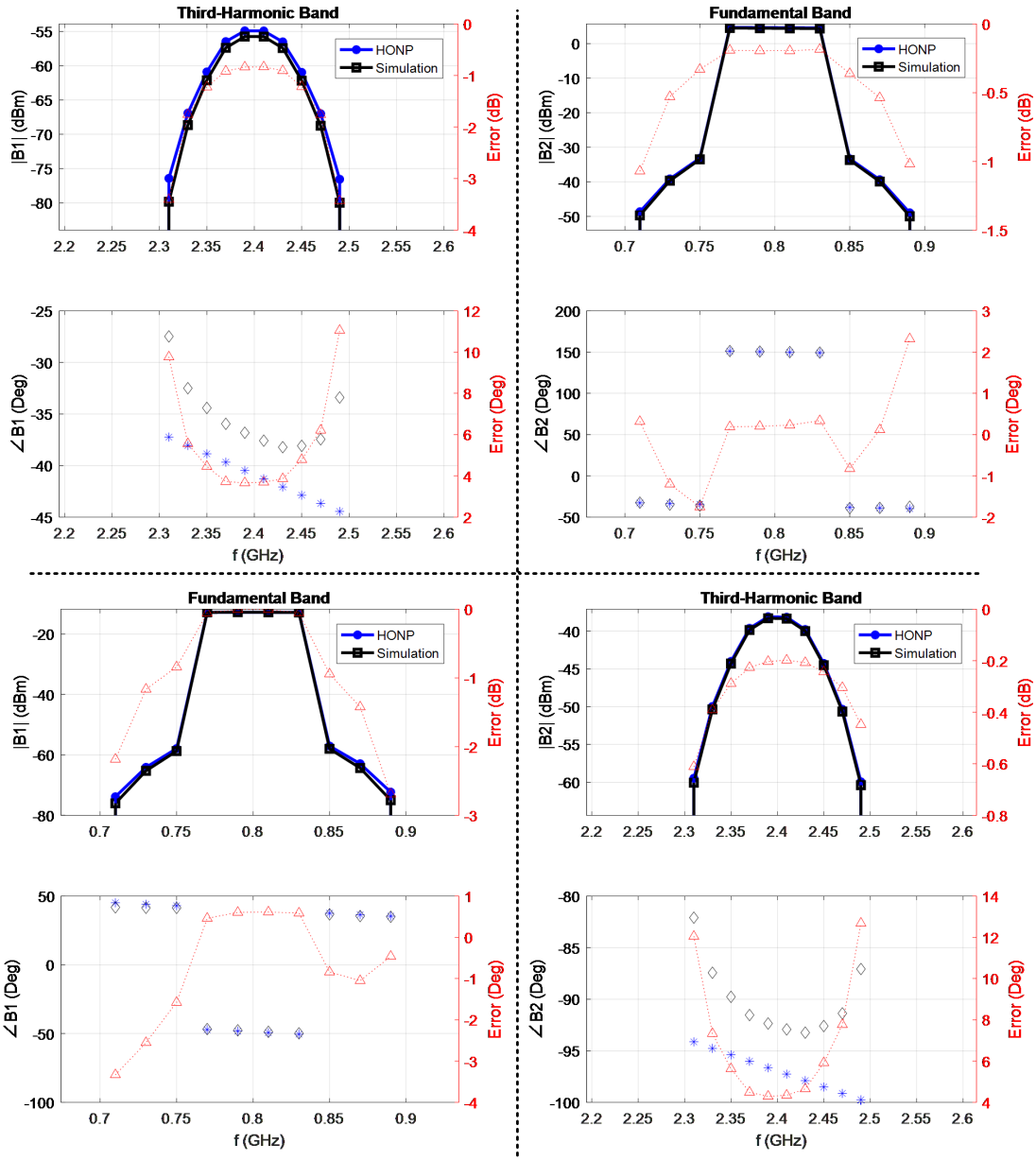


(b)

Figure 3.17: Simulated and HONPs predicted: (a) second-order behaviours, and (b) third-order behaviours for both B_1 and B_2. The stimulus is a 4-tone signal at 8.6 dBm average input power and 6.02 dB PAPR.


Baseband
HONP
Simulation
|B1| (dBm)
Error (dB)
Baseband
HONP
Simulation
|B2| (dBm)
Error (dB)
∠B1 (Deg)
Error (Deg)
f (GHz)
∠B2 (Deg)
Error (Deg)
f (GHz)
Second-Harmonic Band
HONP
Simulation
|B1| (dBm)
Error (dB)
Second-Harmonic Band
HONP
Simulation
|B2| (dBm)
Error (dB)
∠B1 (Deg)
Error (Deg)
f (GHz)
∠B2 (Deg)
Error (Deg)
f (GHz)

(a)



(b)

Figure 3.18: Simulated and HONPs predicted: (a) second-order behaviours, and (b) third-order behaviours for both B_1 and B_2. The stimulus is a 4-tone signal at 10.5 dBm average input power and 6.02 dB PAPR.

3.5 HONPs Extraction and Validation in Measurement Environment

In this section, HONPs are extracted and validated in the measurement environment. Cree CGH40006P 6 Watt GaN HEMT biased at (V_{gs}, V_{ds}) of (-2.4, 3) V was used as a DUT. Since the measurement setup has dynamic range limitations, a biasing point that stimulates strong nonlinear behaviours of the DUT was chosen. First, extraction and validation setups will be explained. Then, it will be demonstrated that measured HONPs have the same properties discussed in the previous Section (the simulation environment). Finally, HONPs ability to predict WNL behaviours will be validated using multi-tone excitations.

3.5.1 Extraction and Validation Setups in Measurement Environment

Figure 3.19 shows a photo of the measurement setup, where different configurations are used as follows:

- **CW extraction setup**

  HONPs were extracted using Keysight N5242 PNA-X through NVNA software. The PNA-X was configured as shown in Figure 3.20 (a), where internal sources 1 and 2 feed ports 1 and 3, respectively. Both sources were used simultaneously to stimulate the DUT with a CW excitation, and the phase-difference between these two sources was controlled. The internal attenuators were configured to keep signal levels at the PNA-X receivers below -20 dBm, and hence guarantee their linearity.

- **Two-tone validation setup**

  PNA-X internal sources outputs were combined (through an internal signal combiner) and routed to feed port 1, Figure 3.20 (b). Each source controls one tone of the 2-tone stimulus, and hence tone power, center frequency and tone spacing can be easily controlled.

- **Multi-tone validation setup**

  Keysight Arbitrary Waveform Generator (AWG) M8190A was used as a multi-tone source along with the PNA-X internal receivers. A broadband Balun was used to combine the AWG output and its inverse in order to generate signals with a better

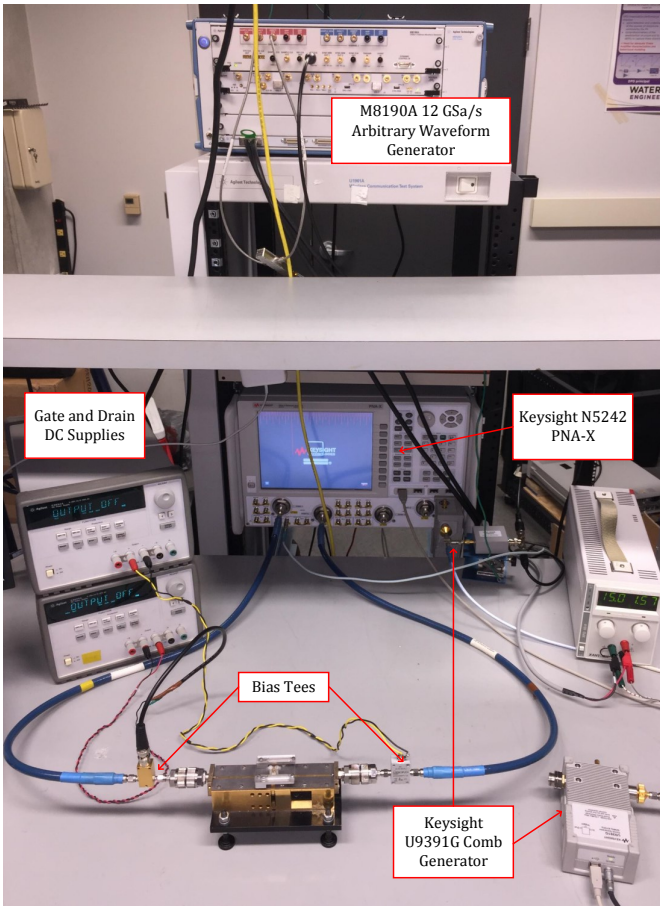
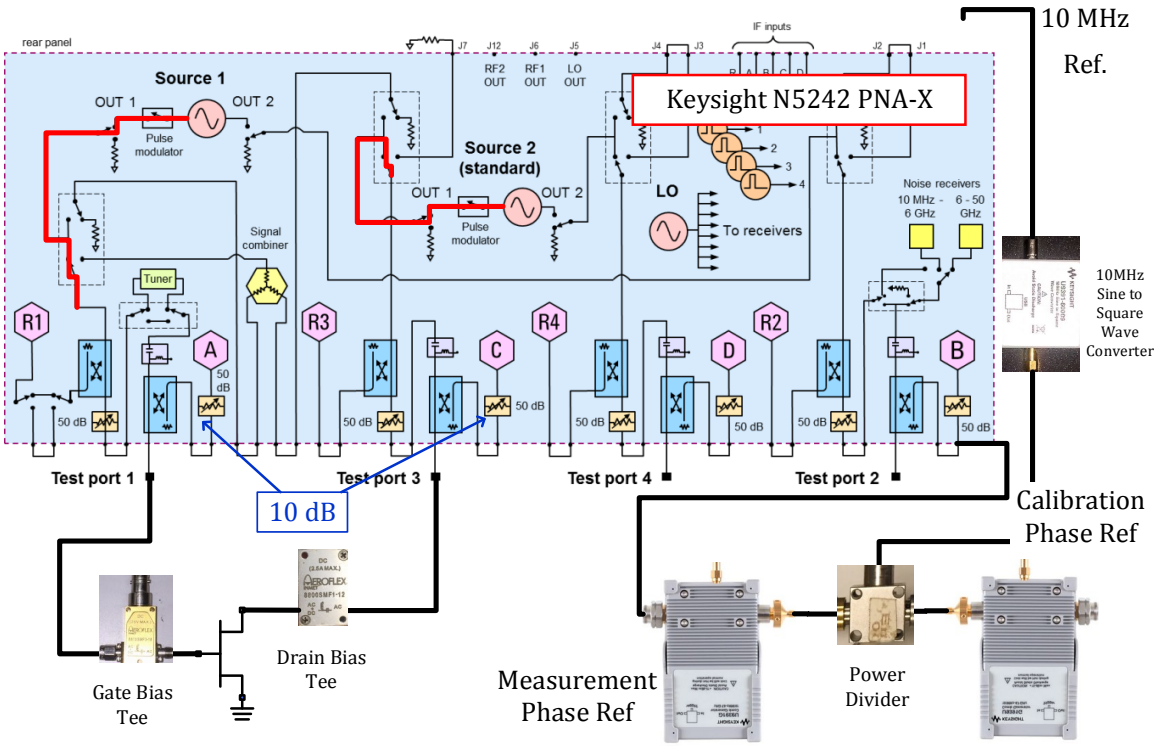


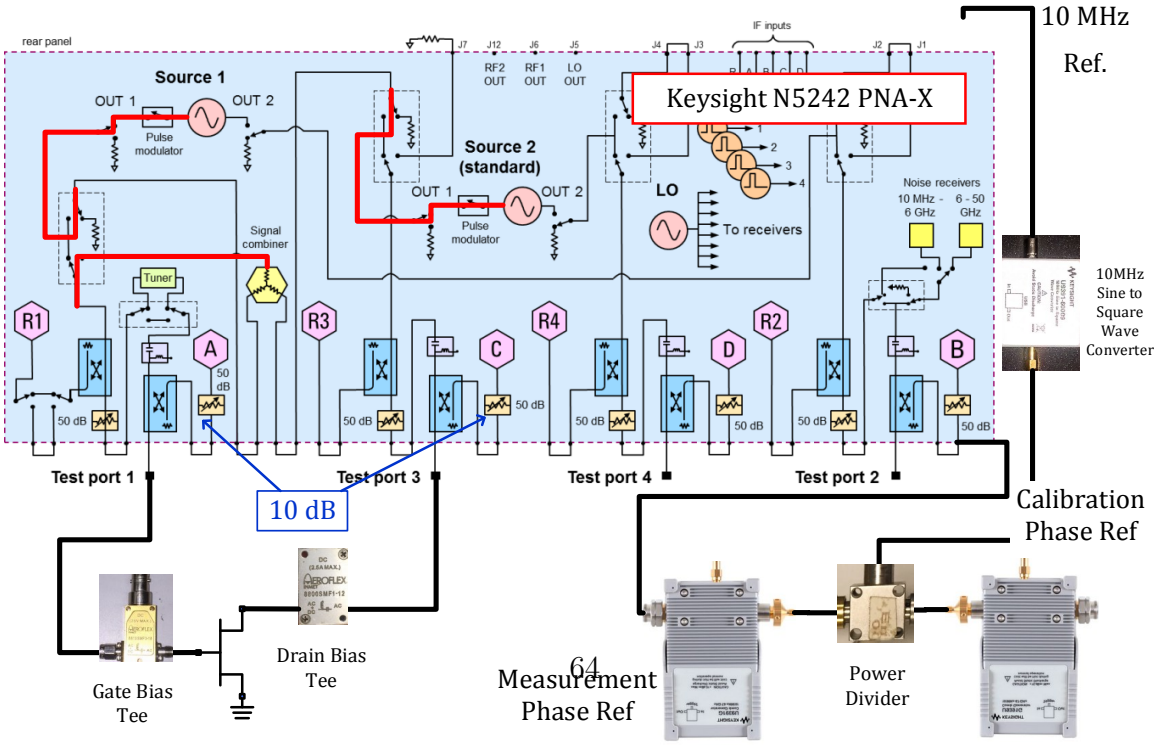
Figure 3.19: A photo of the measurement setup used in HONPs extraction and validation.

signal-to-noise ratio (SNR). This multi-tone signal was fed to the PNA-X as shown in Figure 3.20 (c).

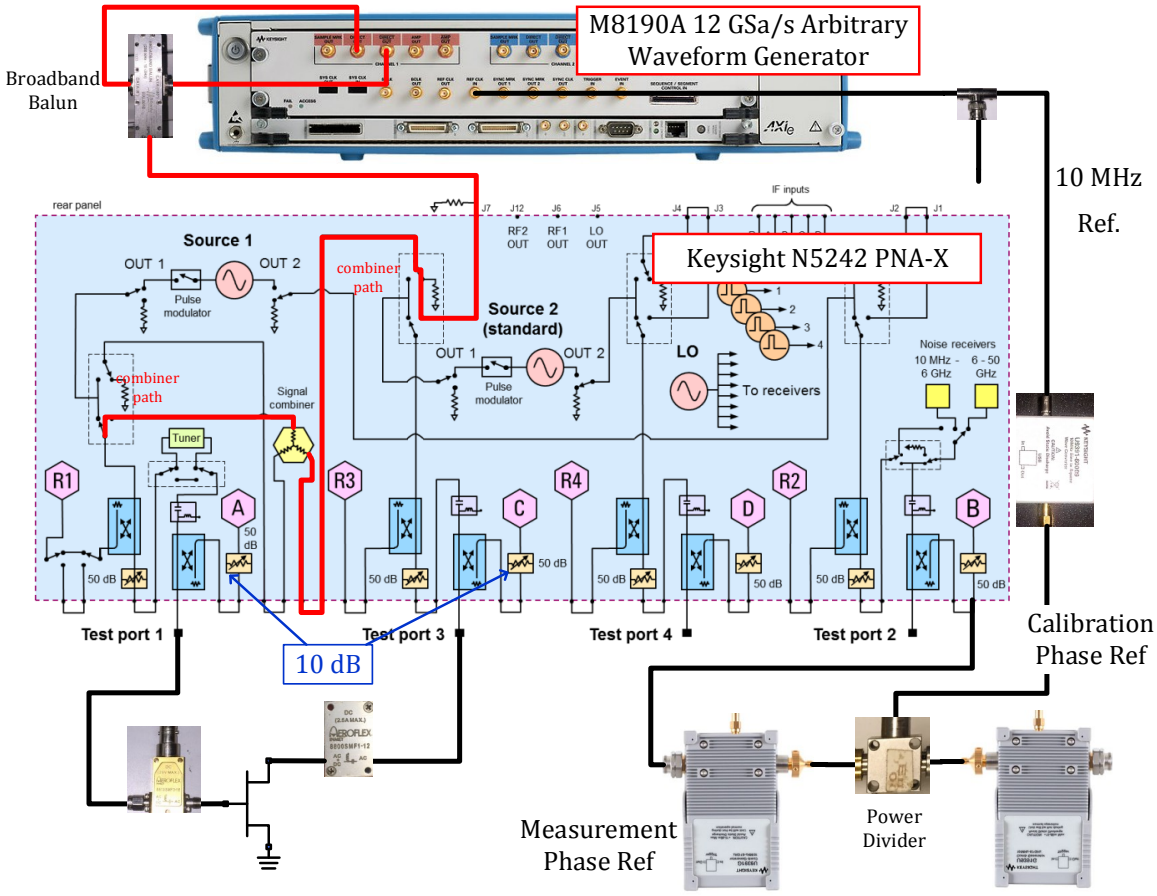
All these configurations use NVNA measurements which require absolute calibrations (vector, power, and phase calibrations). The receivers' IFBW was set to 10 Hz which represents a good compromise between measurement time and dynamic range. Keysight U9391G comb-generator generates a 10 MHz frequency grid with stable phase relationships to sequentially measure the phase of the tones. A 10 MHz sine to square wave converter was used to feed the comb generator for more phase stability at high frequency. Finally, two bias tees were used for DC biasing.


10 MHz
Ref.
rear panel
IF inputs
Source 1
OUT 1
OUT 2
Pulse modulator
RF2 OUT
RF1 OUT
LO OUT
Keysight N5242 PNA-X
Source 2 (standard)
LO
To receivers
Noise receivers
10 MHz - 6 GHz
6 - 50 GHz
Signal combiner
Tuner
10MHz Sine to Square Wave Converter
R1
R3
R4
R2
A
B
C
D
50 dB
Test port 1
Test port 3
Test port 4
Test port 2
10 dB
Calibration Phase Ref
Gate Bias Tee
Drain Bias Tee
Measurement Phase Ref
Power Divider

(a)


10 MHz
Ref.
rear panel
IF inputs
Source 1
OUT 1
OUT 2
Pulse modulator
RF2 OUT
RF1 OUT
LO OUT
Keysight N5242 PNA-X
Source 2 (standard)
LO
To receivers
Noise receivers
10 MHz - 6 GHz
6 - 50 GHz
Signal combiner
Tuner
10MHz Sine to Square Wave Converter
R1
R3
R4
R2
A
B
C
D
50 dB
Test port 1
Test port 3
Test port 4
Test port 2
10 dB
Calibration Phase Ref
Gate Bias Tee
Drain Bias Tee
Measurement Phase Ref
Power Divider

(b)



(c)

Figure 3.20: Different configurations of the HONPs measurement setup: (a) CW Extraction setup, (b) 2-tone validation setup, and (c) multi-tone validation setup. The red lines show PNA-X internal routing for each measurement configuration.

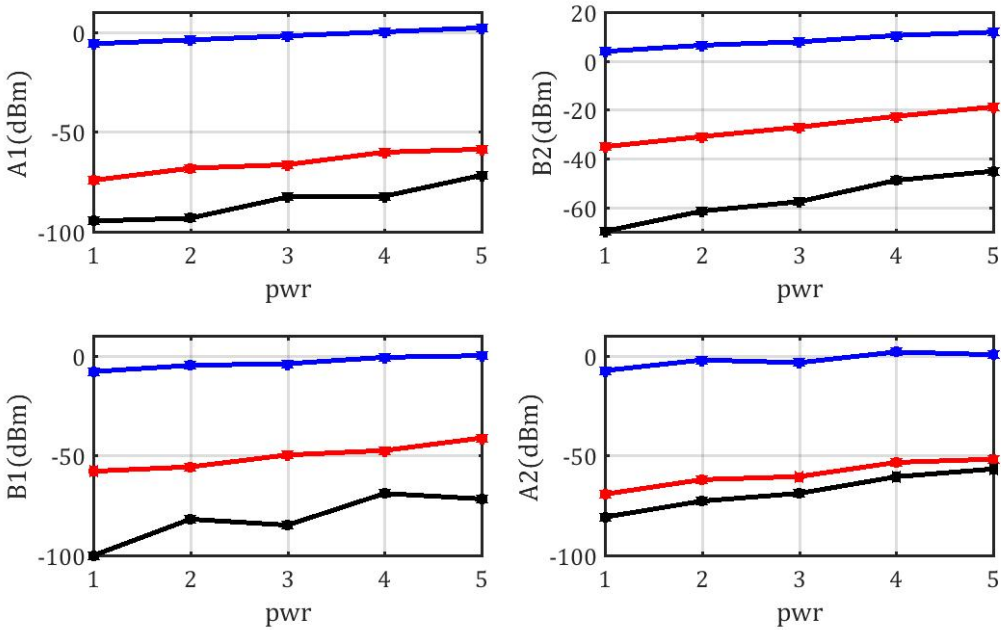


Figure 3.21: An example of measured A and B-waves at 600 MHz fundamental frequency. Fundamental, second-harmonic, and third-harmonic tones are represented by blue, red, and black lines, respectively. The x-axis (pwr) represents the number of power sweep points at the same fundamental frequency and the same phase-difference between A_1 and A_2 fundamental tones.

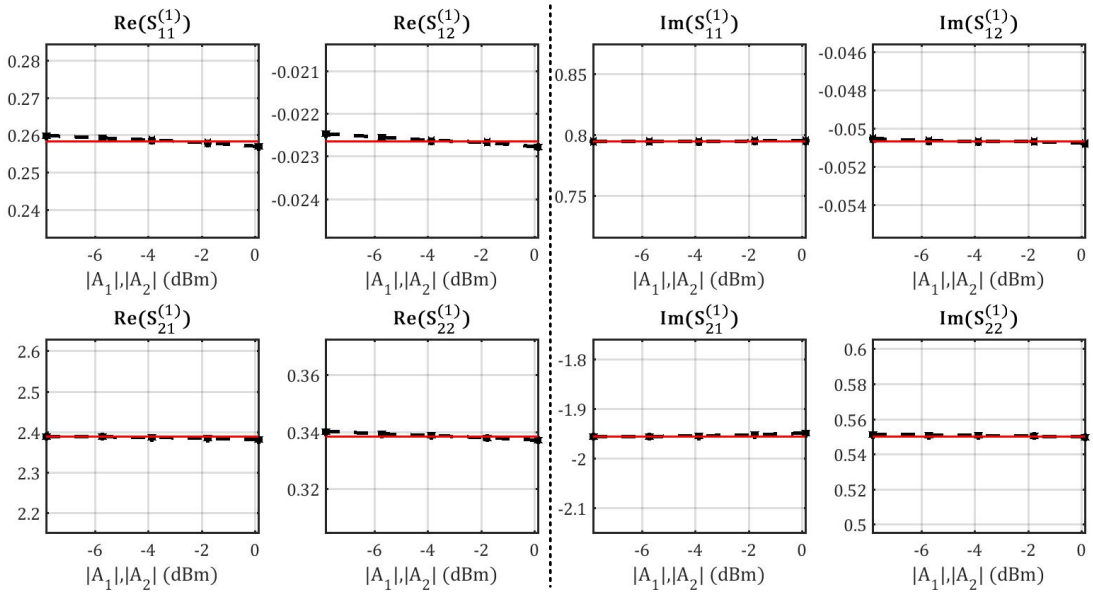
3.5.2 Measured HONPs Properties: Power-Independency

Second and third HOSPs[13] were extracted at five different A_1 and A_2 power levels (using 10 phase-difference values), where both PNA-X internal sources were set to 600 MHz fundamental frequency, and their powers were swept from -8 to 0 dBm [using the setup in Figure 3.20 (a)]. Figure 3.21 shows the measured fundamental, second-harmonic and third-harmonic tones for A and B-waves at both ports. One should notice the following significant differences between these measured waves and the simulated waves (Figure 3.9). First, A_1 and A_2 have second and third-harmonic components because of the spur tones generated by any real-life signal generator. Second, a measured n^{th}-order tone does not have a slope of n versus the input power (as in Figure 3.10). Therefore, identifying

[13]As mentioned in the previous section, HOSPs extraction is more trusted than HOYPs extraction at microwave frequencies. Thus, only HOSPs will be considered here.

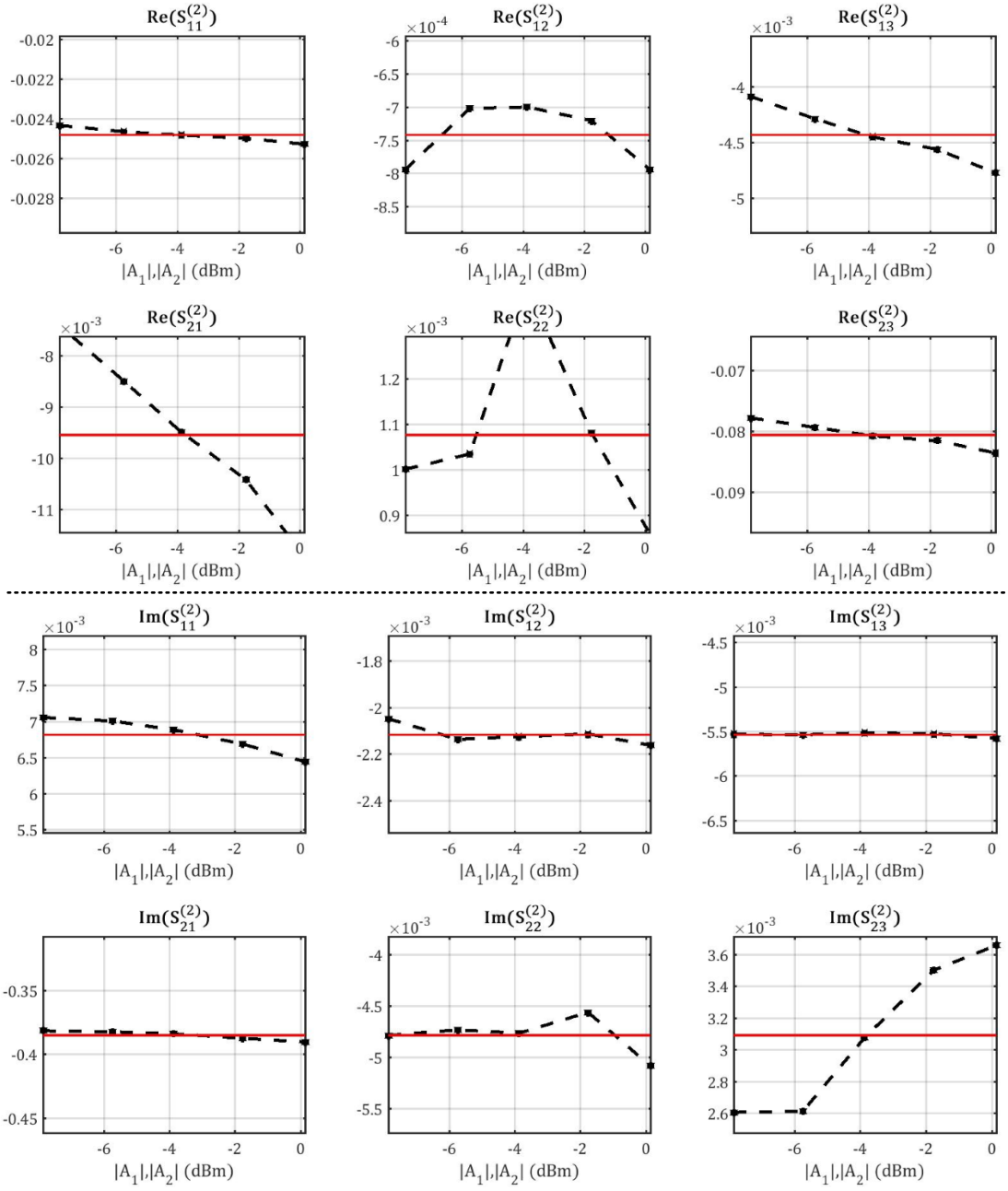
the extraction region in the measurement environment is more complicated than in the simulation environment[14].

Extracted first, second, and third HOSPs are plotted versus the input power in Figure 3.22 (a), (b), and (c), respectively. It is obvious that low-order parameters are more stable versus input power than high-order parameters. This is because measurement noise affects high-order tones more than low-order ones, as well as extraction error propagation from low- to high-order parameters. Also, within the same set of parameters, weak elements can be easily dominated by stronger elements, and hence wrongly identified [e.g. S^3_{14} and S^3_{22} in Figure 3.22 (c)]. This may represent a problem if many weak elements exist within the same set of parameters and the only solution is to make the system more well-conditioned by adding more measurements (more phase sweeps).

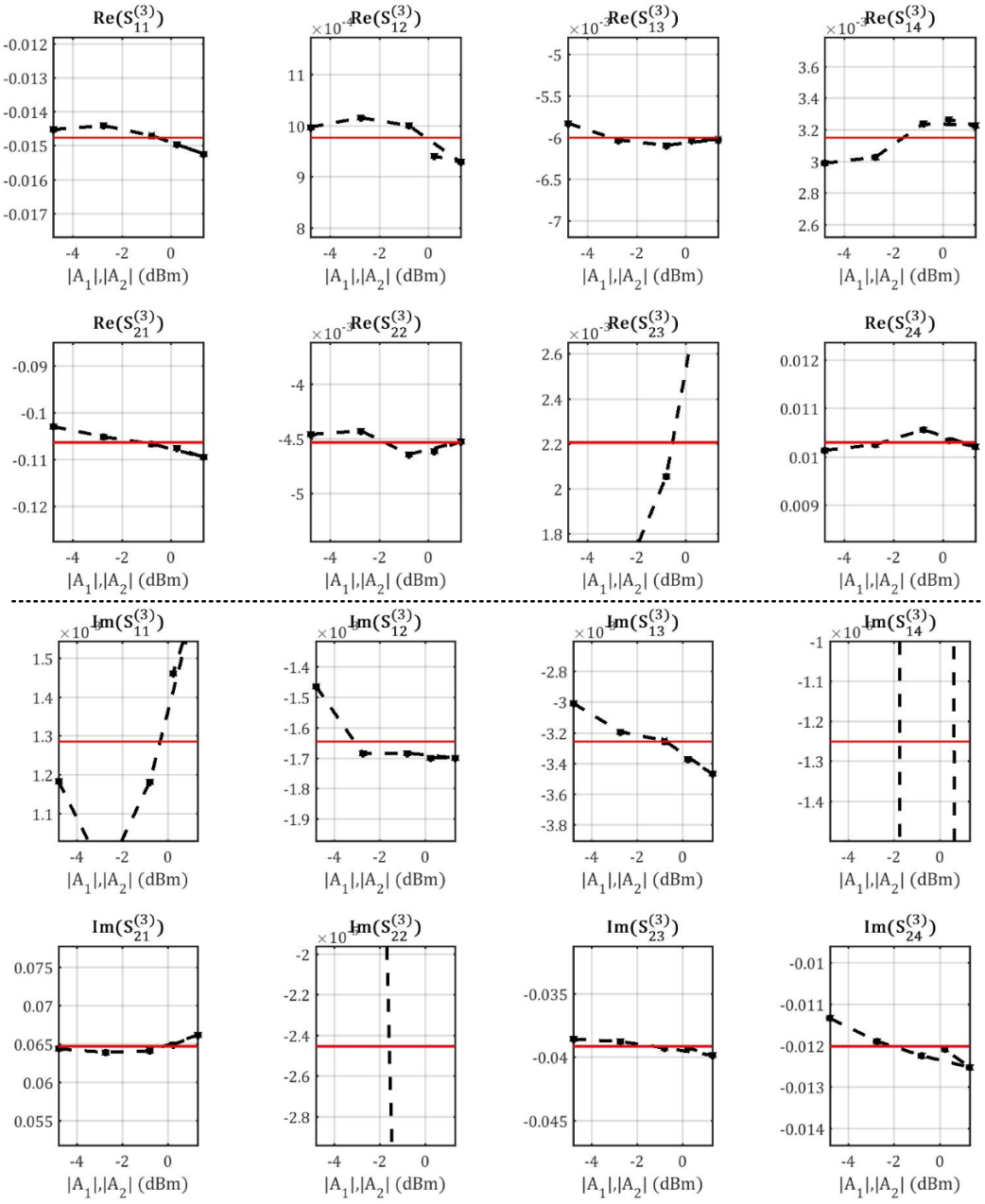


(a)

[14]A MATLAB code was optimized to identify the right WNL extraction region with minimum power sweeps.


Re(S11(2))
Re(S12(2))
Re(S13(2))
Re(S21(2))
Re(S22(2))
Re(S23(2))
Im(S11(2))
Im(S12(2))
Im(S13(2))
Im(S21(2))
Im(S22(2))
Im(S23(2))
|A1|,|A2| (dBm)

(b)



(c)

Figure 3.22: The extracted real and imaginary parts of (a) first, (b) second, and (c) third HOSPs at (V_{gs},V_{ds})=(-2.4, 3) V and 600 MHz fundamental frequency versus fundamental tone stimuli power. The red line represents the element value that will be used.

3.5.3 Measured HONPs Properties: Solution Continuity and Uniqueness

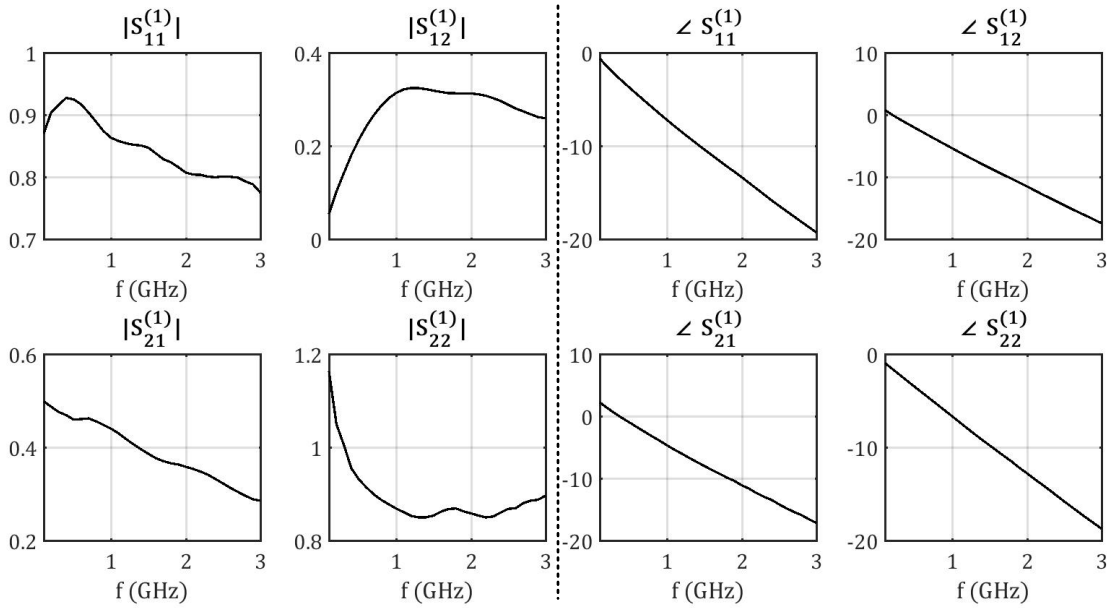
The frequency of the fundamental tones of A_1 and A_2 was swept from 10 MHz to 3 GHz with a step of 20 MHz [Figure 3.20 (a)]. At each fundamental frequency, the correct extraction power level was identified then ten phase-difference values were used to extract second and third HOSPs. The black curves in Figure 3.23 (a), (b), and (c) show the extracted first, second and third HOSPs versus frequency, respectively. HOSPs were extracted again using only four phase-difference values (red curves in Figure 3.23). As mentioned before, using more phase values is very important for correct identification of the weak elements. Table 3.2 shows a comparison between the system conditioning number in the case of four and ten phase-difference values at 600 MHz fundamental frequency. It is obvious that using ten values results in a more trusted HOSPs extraction. Again, the fact that the system conditioning number is close to one is an indication of the solution stability.

	4 measurements	10 measurements
Second HOSPs	12	1.19
Third HOSPs	512	2.86

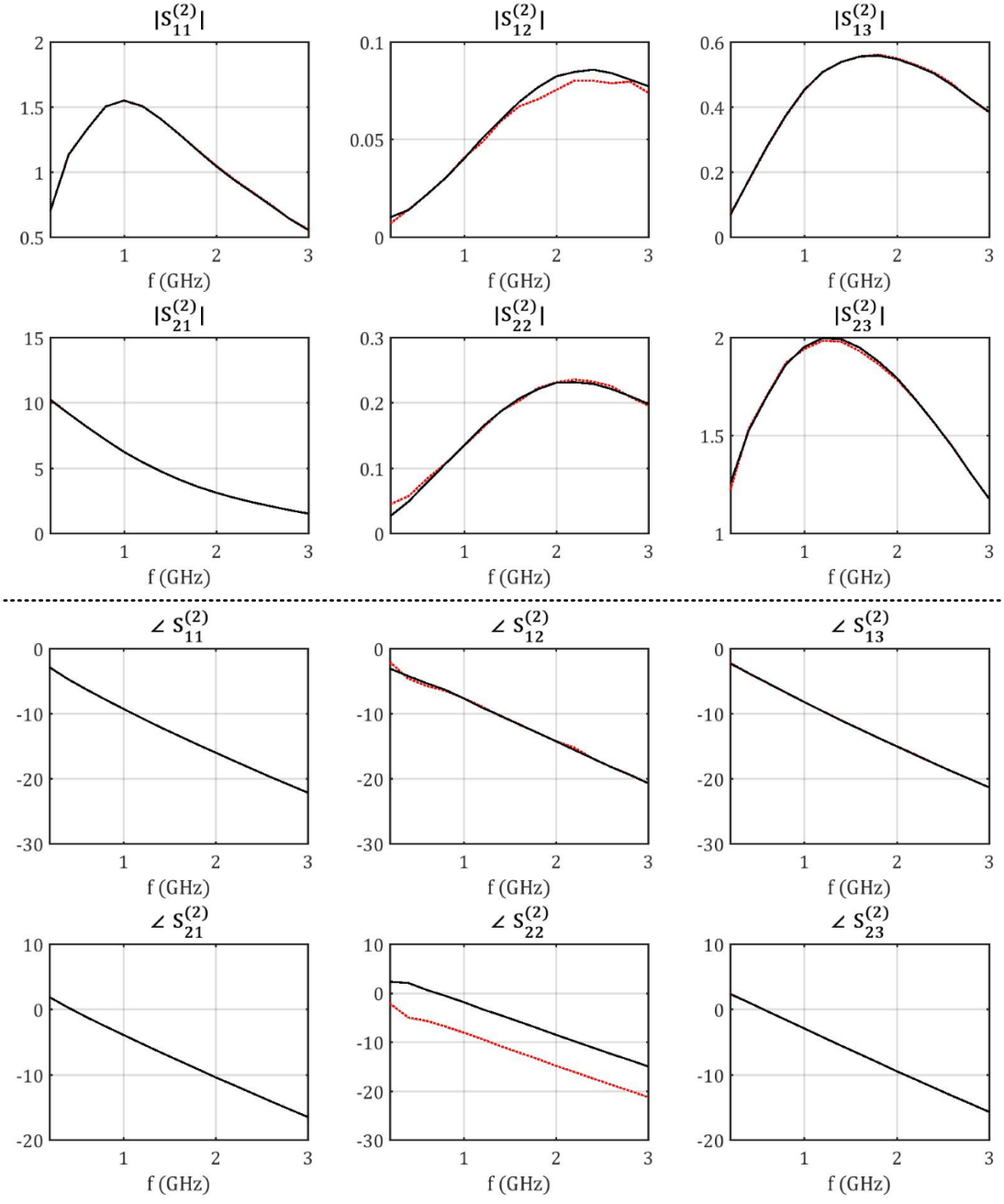
Table 3.2: A comparison between the conditioning number of the system of linear equations used to extract both second and third HOSPs using four and ten measurements in the measurement environment.

The behaviour of measured HOSPs was then examined across the biasing grid by sweeping V_{gs} and V_{ds} from -4 to 0 V, and from 0 to 50 V, respectively [15]. Figure 3.24 (a), (b), and (c) shows the extracted amplitudes of first, second, and third HOSPs versus part of frequency and biasing grids, where different planes were plotted in each dimension to show the continuity of the behaviour that was expected from the simulation environment. Again, Figure 3.24 (b) and (c) can completely describe second- and third-order nonlinear behaviours of the DUT (within the shown modelling space).

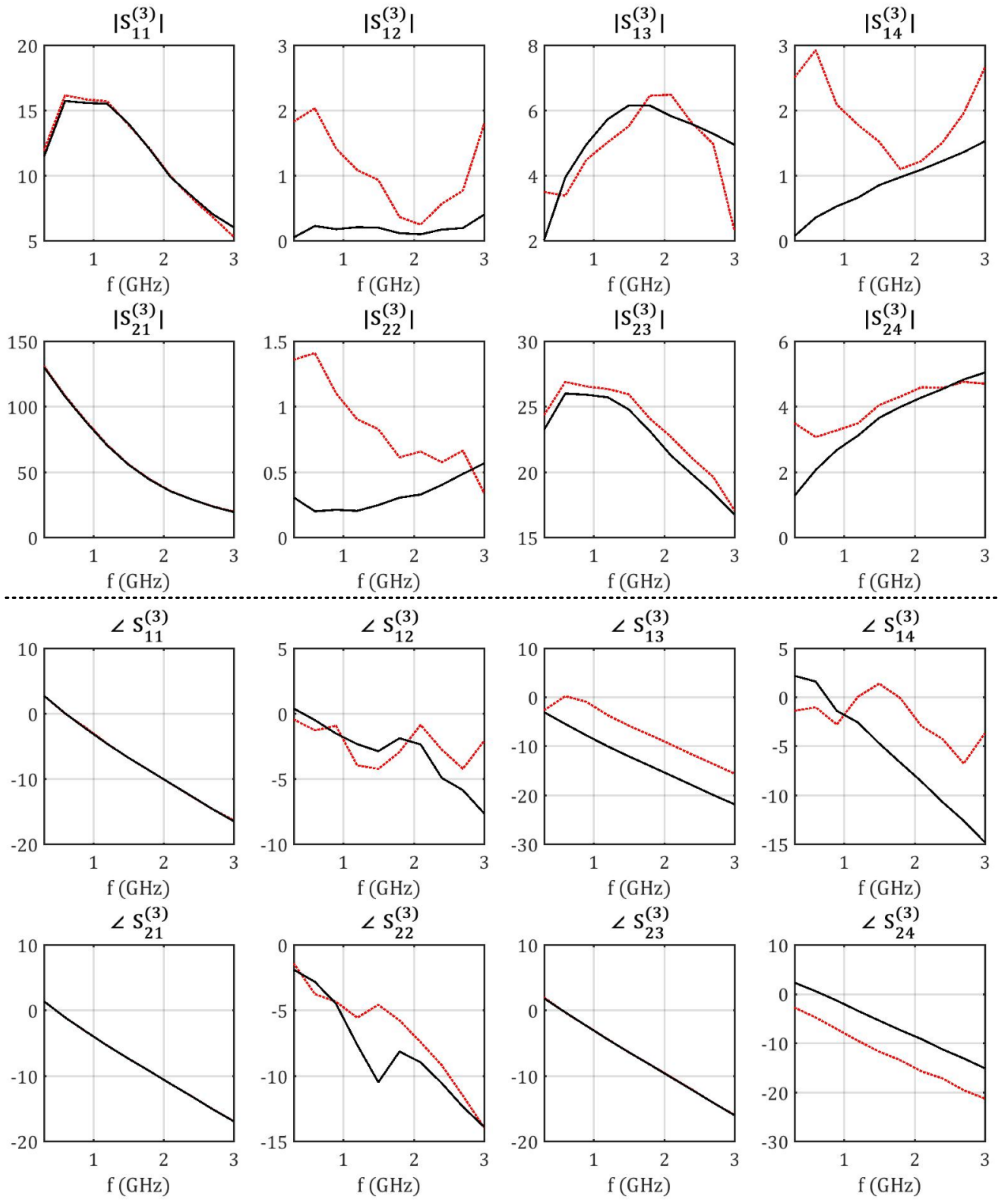
[15] To reduce measurement time, the frequency step for these sweeps was increased to 50 MHz.


$|S_{11}^{(1)}|$
$|S_{12}^{(1)}|$
$\angle S_{11}^{(1)}$
$\angle S_{12}^{(1)}$
$|S_{21}^{(1)}|$
$|S_{22}^{(1)}|$
$\angle S_{21}^{(1)}$
$\angle S_{22}^{(1)}$
f (GHz)

(a)

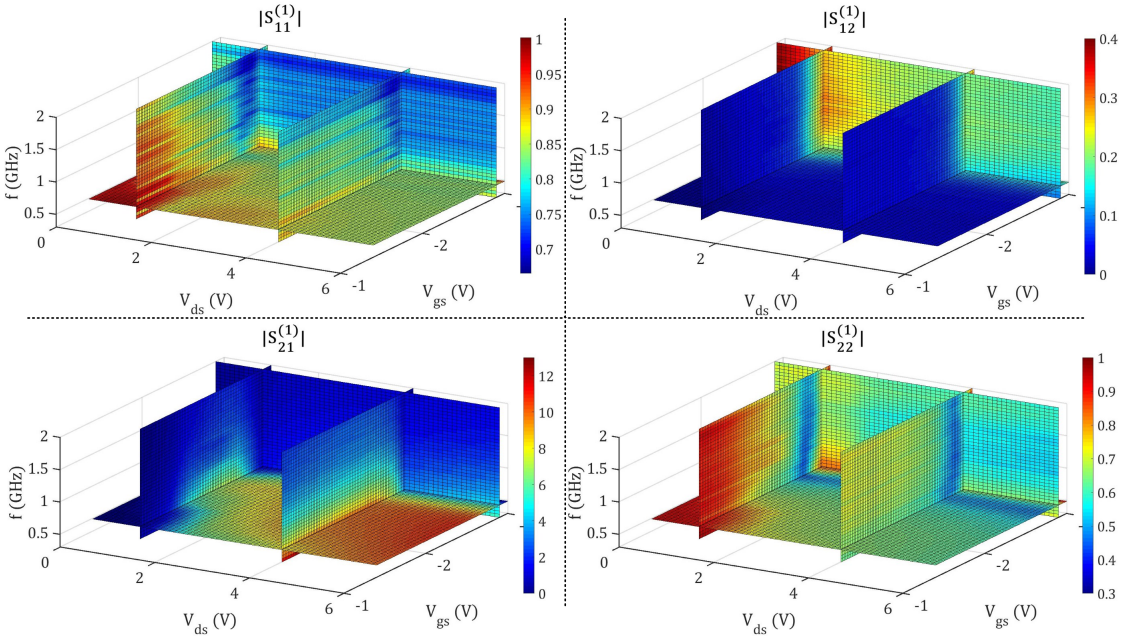

$|S_{11}^{(2)}|$
$|S_{12}^{(2)}|$
$|S_{13}^{(2)}|$
$|S_{21}^{(2)}|$
$|S_{22}^{(2)}|$
$|S_{23}^{(2)}|$
$\angle S_{11}^{(2)}$
$\angle S_{12}^{(2)}$
$\angle S_{13}^{(2)}$
$\angle S_{21}^{(2)}$
$\angle S_{22}^{(2)}$
$\angle S_{23}^{(2)}$
f (GHz)

(b)

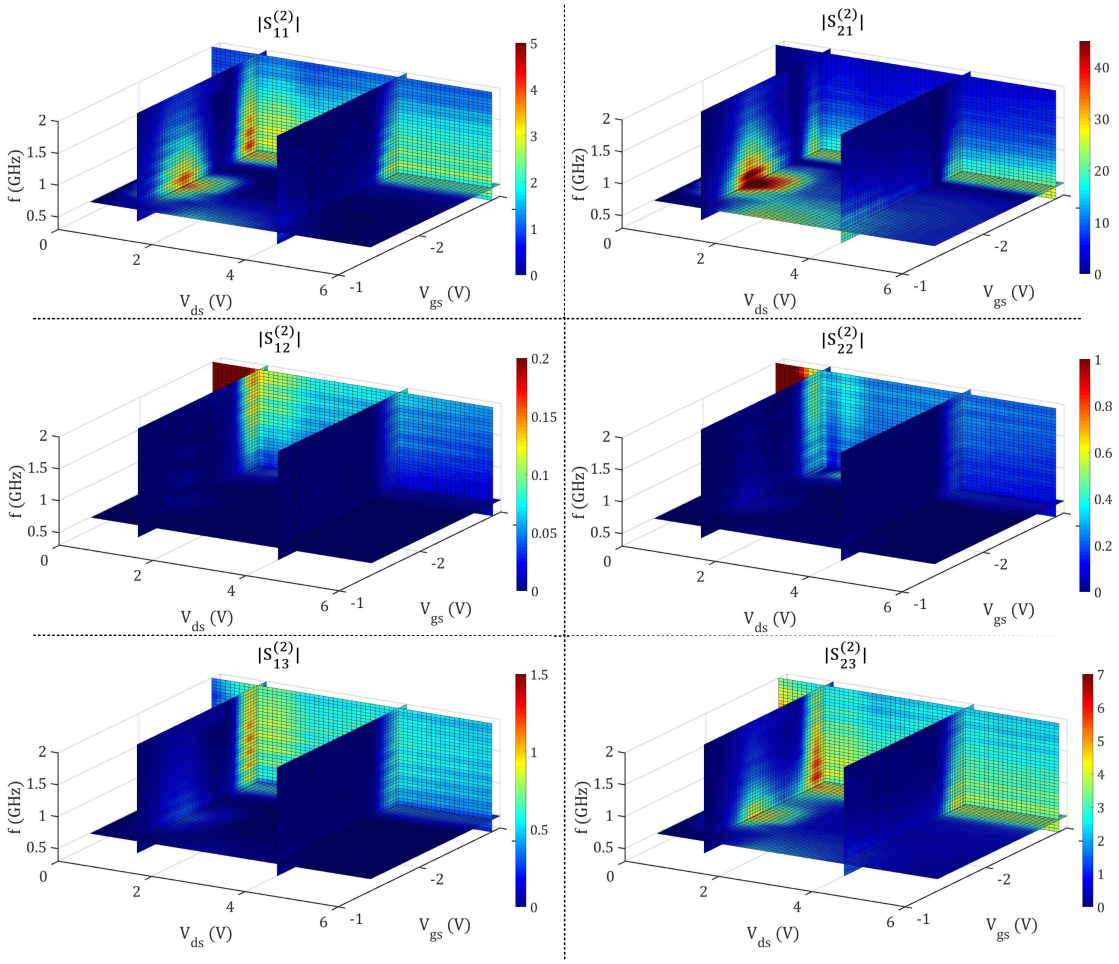


(c)

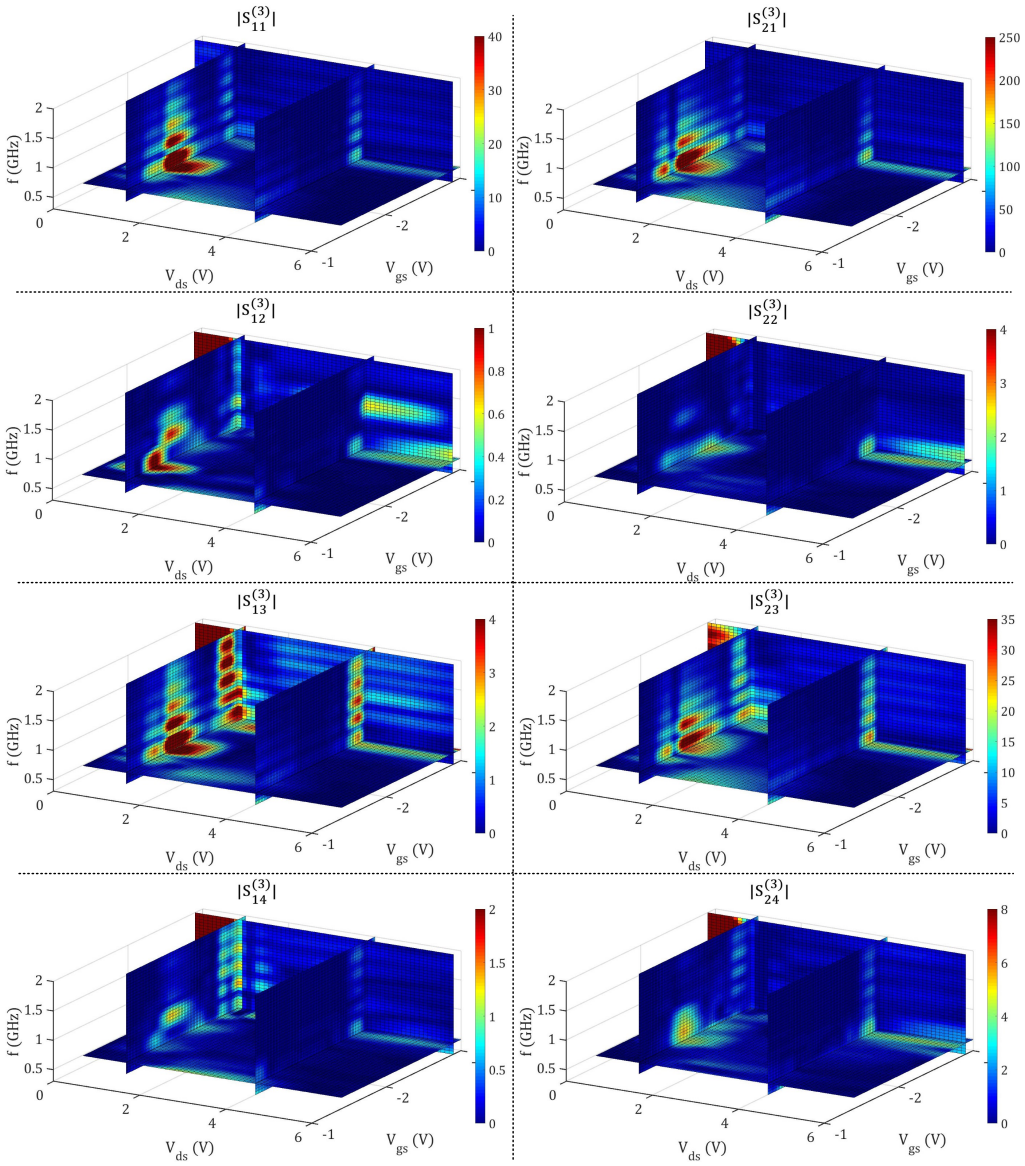
Figure 3.23: Extracted real and imaginary parts of (a) first, (b) second, and (c) third HOSPs at (V_{gs},V_{ds})=(-2.4, 3) V versus frequency. The red and black curves show the extracted elements using four and ten phase-difference values between A_1 and A_2 fundamental tones, respectively.



(a)



(b)



(c)

Figure 3.24: Amplitude of the extracted: (a) linear S-parameters, (b) second HOSPs, and (c) third HOSPs versus frequency, V_{gs}, and V_{ds} in the measurement environment.

3.5.4 Prediction of WNL Response Due To Multi-tone Excitation Using HONPs

HONPs capability to predict WNL behaviours was tested using different multi-tone excitations in the measurement environment. The same DUT [Cree CGH40006P 6 Watt GaN HEMT biased at (V_{gs}, V_{ds}) of (-2.4, 3) V] used for HONPs extraction was used here for validation.

First, a 20 MHz spaced 2-tone signal with its center frequency swept from 340 to 700 MHz (with a step of 20 MHz) was used to excite the DUT [using the measurement setup in Figure 3.20 (b)]. Measured and predicted second and third-harmonic bands of both B_1 and B_2 are plotted in Figure 3.25. Although a good agreement between measurements and HONPs predictions can be seen, one should notice the following:

- While black and blue curves represent measured and predicted behaviours, the green curve represents the effect of the linear network on the spur tones generated by the source. It is critical to plot these green curves in the measurement environment because in some cases, source nonlinearities might dominate WNL behaviours (e.g. when using a driver). In other words, the first terms in (3.20) and (3.21) will dominate the other terms. In this scenario, blue and black curves will still be in agreement, but it would be due to traditional S-parameters, not the high-order parameters.

- Baseband and third-IMD behaviours are not reported in Figure 3.25. PNA-X internal couplers roll-off is too large below 100 MHz, and it reaches -70 dB at 10 MHz. Therefore, measuring a second-harmonic tone within the WNL extraction region is impossible. Thus, the second HONPs at 20 MHz were not correctly identified, and that is why baseband and third-IMD behaviours could not be predicted. This problem can be solved by using a baseband extension for the PNA-X.

- The error in predicting tones below -90 dBm is too large because of the limited dynamic range of all measurement equipment. This is obvious when predicting B_1 third-harmonic band behaviours for frequencies higher than 1.6 GHz. This error can be reduced by decreasing the receivers IFBW, yet the measurement time will significantly increase. This tradeoff solely depends on the application.

Second, a wideband 8-tone stimulus (with 20 MHz frequency spacing) centred around 570 MHz was used to stimulate the DUT [Figure 3.20 (c)]. The signal PAPR was 7.3 dB, and two average input powers were used: 4.9 and 7.6 dBm. Figures 3.27 and 3.28 show measured and predicted second and third- order behaviours of both B_1 and B_2 for both average input powers, respectively. The following remarks should be noted.

- Figure 3.26 shows the measured spectrums for A and B-waves. While A_1 spectrum contains unwanted spur tones generated by the AWG, A_2 spectrum contains tones at all IMD and harmonic bands due to port 2 mismatch.
- The baseband prediction error of B-waves becomes acceptable at frequencies higher than 80 MHz. This is because PNA-X internal couplers roll-off allows correct identification of second HONPs.
- Although the error in the baseband propagates to the third-IMD band, a good agreement between measured and predicted amplitudes can be noticed due to the correct prediction of some of the baseband tones.

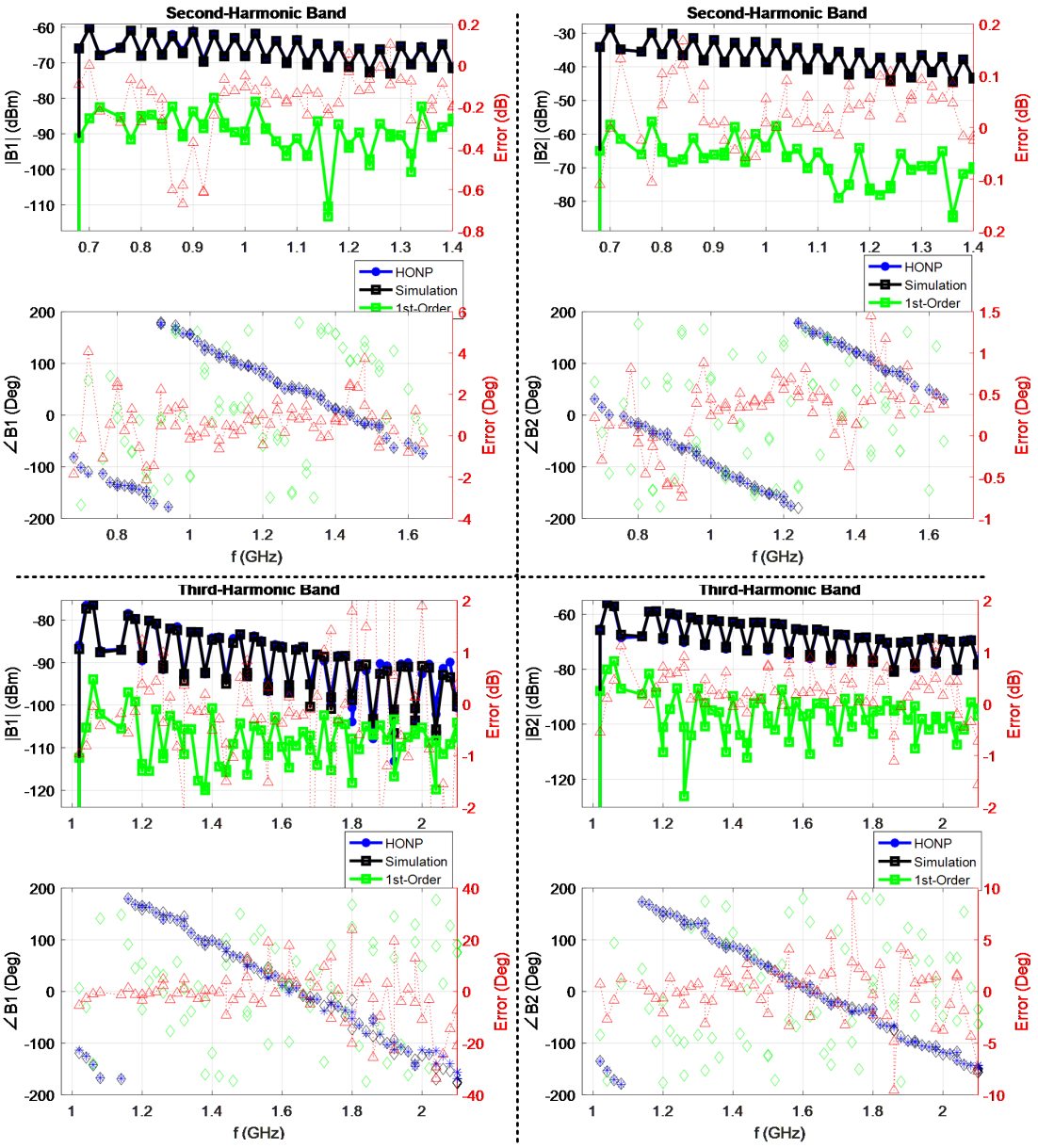
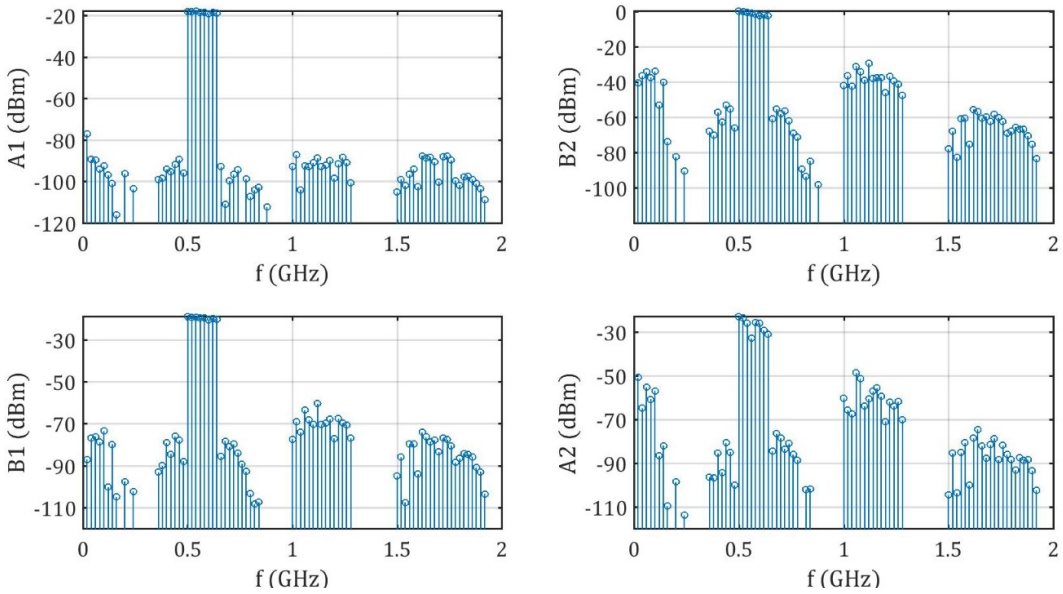
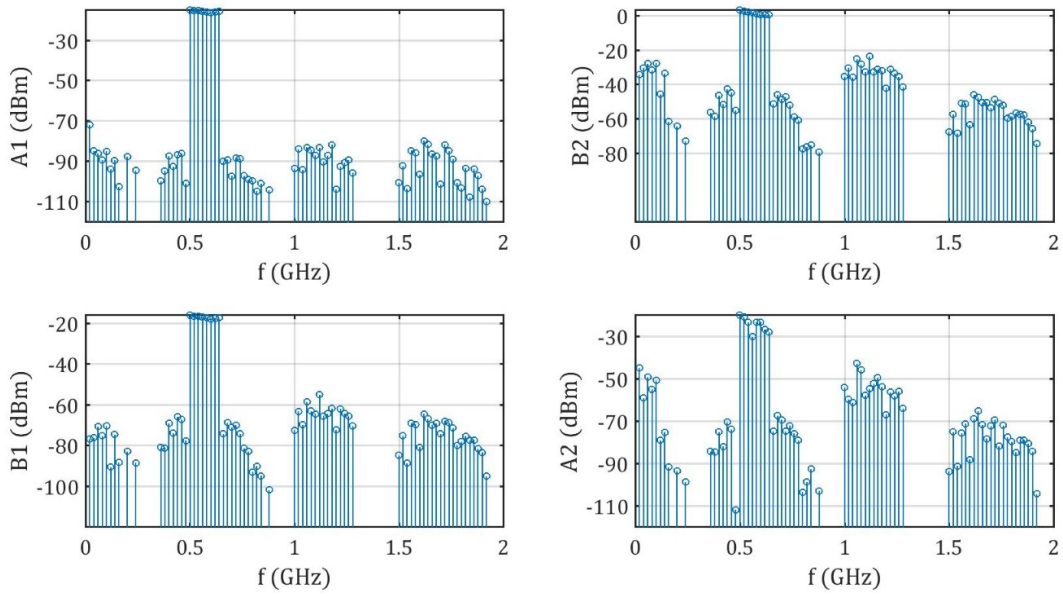


Figure 3.25: Measured and HONPs predicted: (a) second-harmonic band and (b) third-harmonic band behaviours for both B_1 and B_2. The stimulus is a 2-tone signal with 20 MHz tone spacing, and the center frequency swept from 340 to 700 MHz.

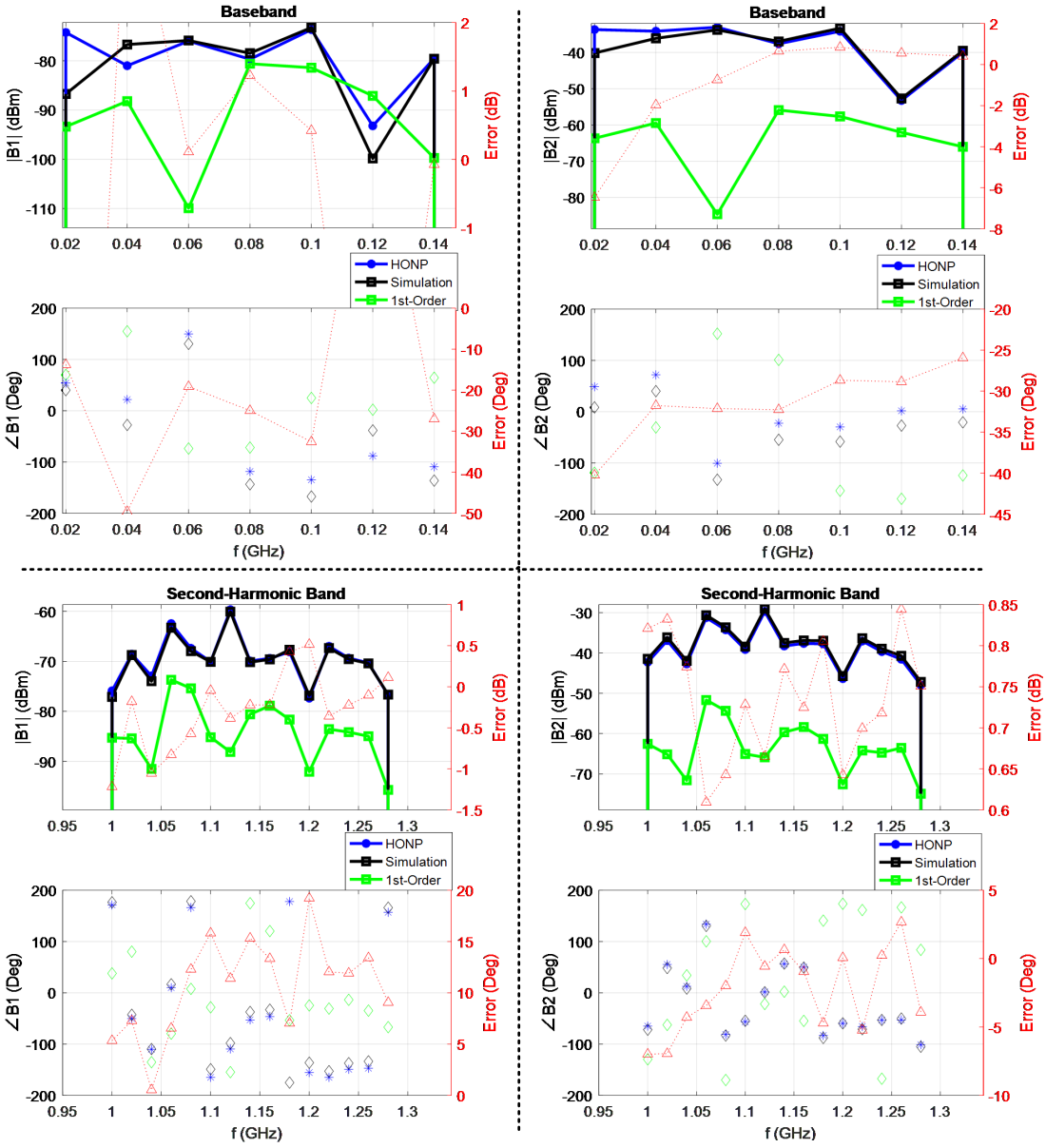


(a)

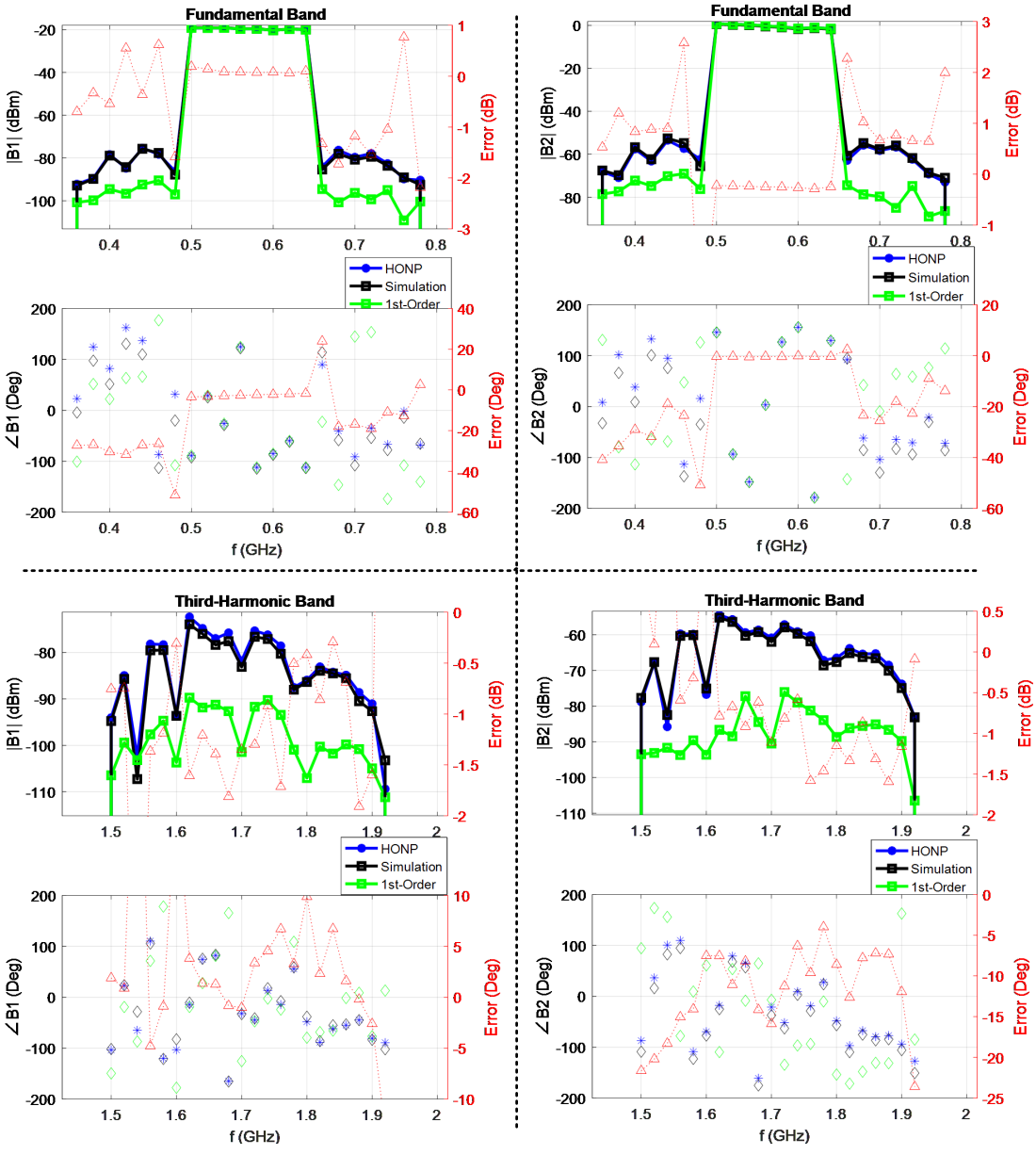


(b)

Figure 3.26: Measured A and B-waves spectrums for 160 MHz 8-tone stimulus centred around 570 MHz with average input powers: (a) 4.9 dBm and (b) 7.7 dBm.

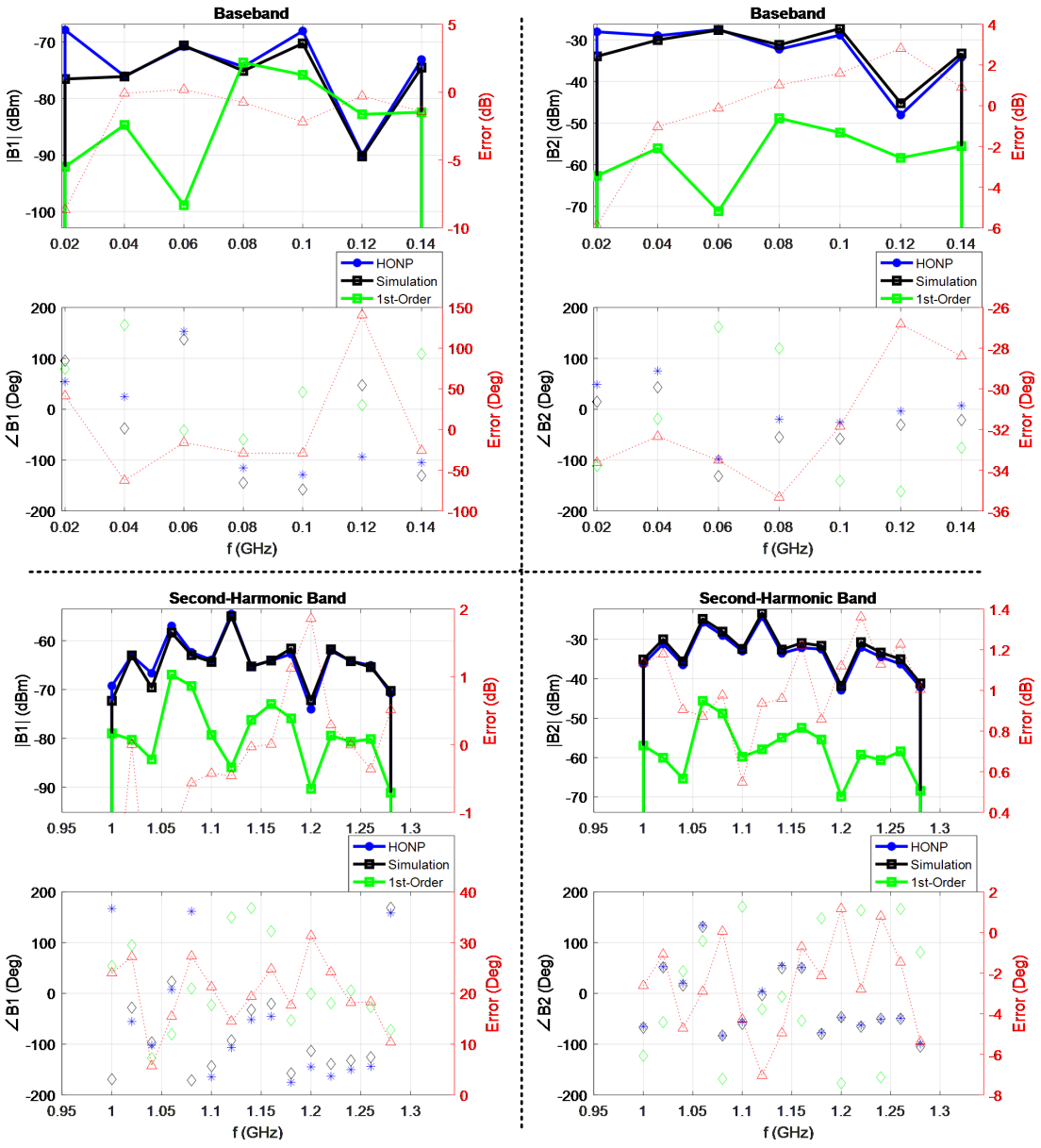

Baseband
|B1| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B1 (Deg)
Error (Deg)
f (GHz)
Baseband
|B2| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B2 (Deg)
Error (Deg)
f (GHz)
Second-Harmonic Band
|B1| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B1 (Deg)
Error (Deg)
f (GHz)
Second-Harmonic Band
|B2| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B2 (Deg)
Error (Deg)
f (GHz)

(a)

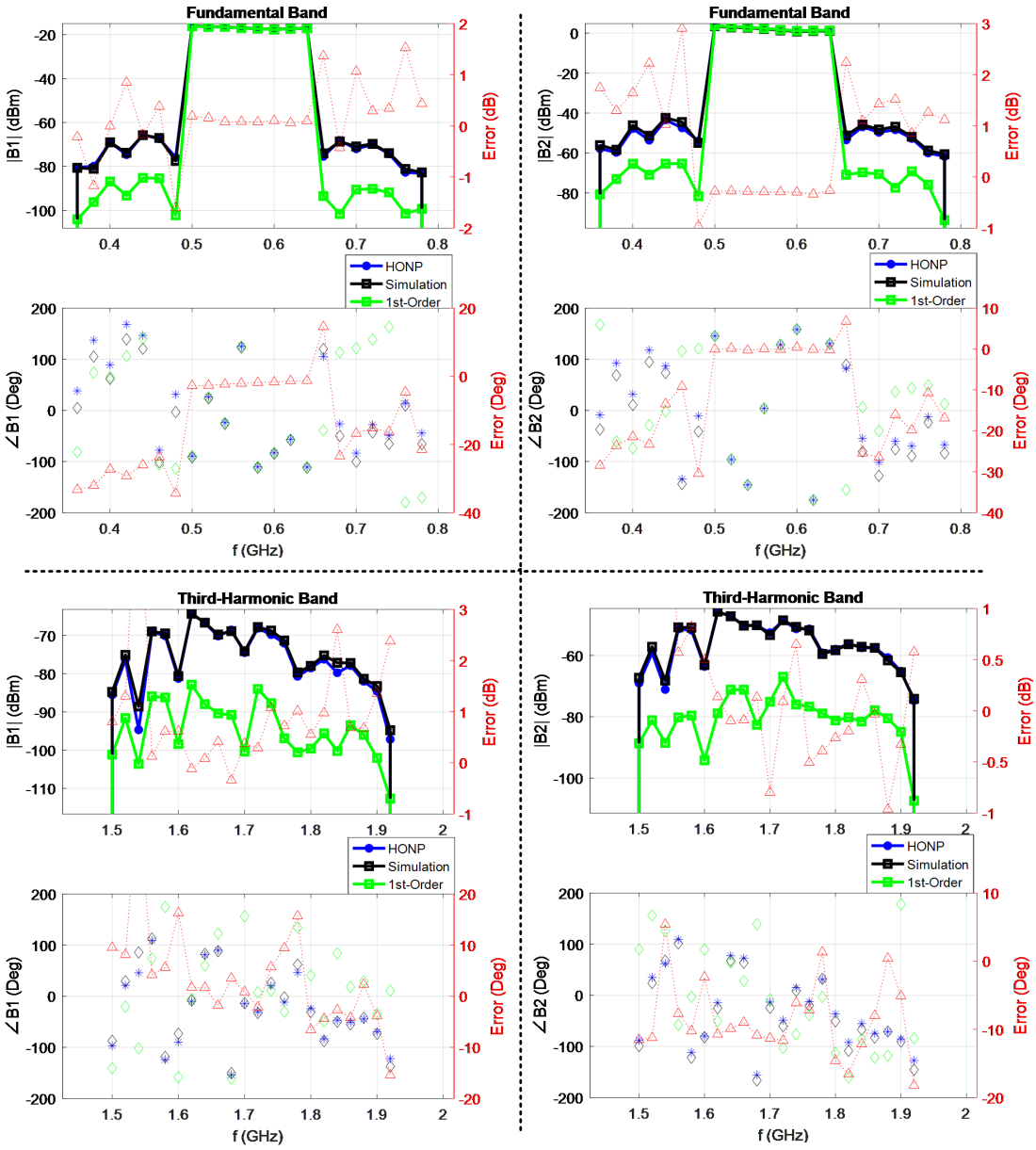


(b)

Figure 3.27: Measured and HONPs predicted: (a) second-order behaviours, and (b) third-order behaviours for both B_1 and B_2. The stimulus is an 8-tone signal at 4.9 dBm average input power and 7.6 dB PAPR.


Baseband
|B1| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B1 (Deg)
Error (Deg)
f (GHz)
Baseband
|B2| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B2 (Deg)
Error (Deg)
f (GHz)
Second-Harmonic Band
|B1| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B1 (Deg)
Error (Deg)
f (GHz)
Second-Harmonic Band
|B2| (dBm)
Error (dB)
HONP
Simulation
1st-Order
∠B2 (Deg)
Error (Deg)
f (GHz)

(a)



(b)

Figure 3.28: Measured and HONPs predicted: (a) second-order behaviours, and (b) third-order behaviours for both B_1 and B_2. The stimulus is an 8-tone signal at 7.6 dBm average input power and 7.3 dB PAPR.

3.6 Comparison

It is worth highlighting that HONPs represent Volterra kernels under the generalized quasi-static assumption. This reduces the dimensionality of the kernels to one. For example, a third-order Volterra kernel for a DIDO system is a six-dimensional function as it describes all combinations between first-order tones at both ports that will result in a third-order tone[16]. However, HONPs formulations presented in this chapter maintains the domain of the function from exploding, and hence allow utilizing these parameters to build compact models (as will be explained in Chapter 5).

Table 3.3 provides a comparison between the proposed HONPs and X-parameters. This comparison is based on the work presented in this chapter and the discussion in Section 2.3.2.

3.7 Conclusion

A novel set of parameters (HONPs) that can completely describe WNL behaviours of any nonlinear system were proposed. These parameters are believed to be the true extension of traditional 2-port linear network parameters and were formulated in two forms: HOYPs and HOSPs. HONPs were validated in both simulation and measurement environments to have the following properties:

- HONPs are power independent in the WNL extraction region. While the lower limit of this region is limited by the measurement setup dynamic range, the upper limit is defined by the compression in both fundamental and high-order tones.
- HONPs are continuous across the modelling space, and hence can easily be interpolated. In this work, the modelling space is defined by frequency and biasing grids as the objective is to build a transistor compact model.
- HONPs formulations are linear with respect to the coefficients which proves the uniqueness of them. Also, the conditioning number of the system of linear equations used in HONPs extraction is close to one. Thus, it proves these parameters are stable and can describe the WNL behaviour with great mathematical certainty.

[16]For more details, please refer to Appendix A.

Finally, the proposed HONPs showed excellent ability in predicting WNL behaviours due to wideband multi-tone signal excitations. HONPs also showed great ability in predicting WNL behaviours under modulated signals and proved that they can be used for lioearizating the DUT (Appendix B).

Table 3.3: A comparison between the proposed HONPs and X-parameters.

	X-parameters	HONPs
Theory	• PHD-based model	• Volterra-based model
Measurement Setup	• NVNA measurements • Source/load-pull setup • High power capabilities	• NVNA measurements
Strengths	• Strong nonlinear behaviours are modelled through load-dependent X-parameters • It inherently provides large-signal validations	• Strong nonlinear behaviours are modelled through bias-dependent HONPs (Chapter 5) • Power independent • Continuous behaviour across the modelling space
Drawbacks	• The modelling space can easily explode • Sensitive to the implementation and interpolation techniques	• Sensitive to the measurement noise • Affected by the limited dynamic range

Chapter 4

Compact Model Implementation Part 1: Linear Measurement-Based Compact Model Implementation using Neural Networks

4.1 Introduction

The extraction process of linear measurement-based compact models mentioned in Section 2.2.1 requires the following steps. First, an equivalent circuit has to be assumed for both extrinsic and intrinsic shells. Second, initial values for extrinsic shell elements are found (using ColdFET) and de-embedded. Intrinsic shell values are then extracted for all biasing points (using HotFET), and then the whole process has to be repeated until the resulting S-parameters match the measured ones (within an acceptable error). Now, what if the desired performance cannot be met? The answer is to try a different topology that has more elements, i.e. more degrees of freedom.

This chapter proposes a novel technology-independent linear-measurement based compact model. A powerful implementation technique using neural networks (NNs) that enables testing different topologies for intrinsic and extrinsic shells -in an automated manner- is proposed. The transistor's equivalent circuit is represented by a custom NN that is built as cascaded layers, and both extrinsic and intrinsic elements are extracted in one step (at

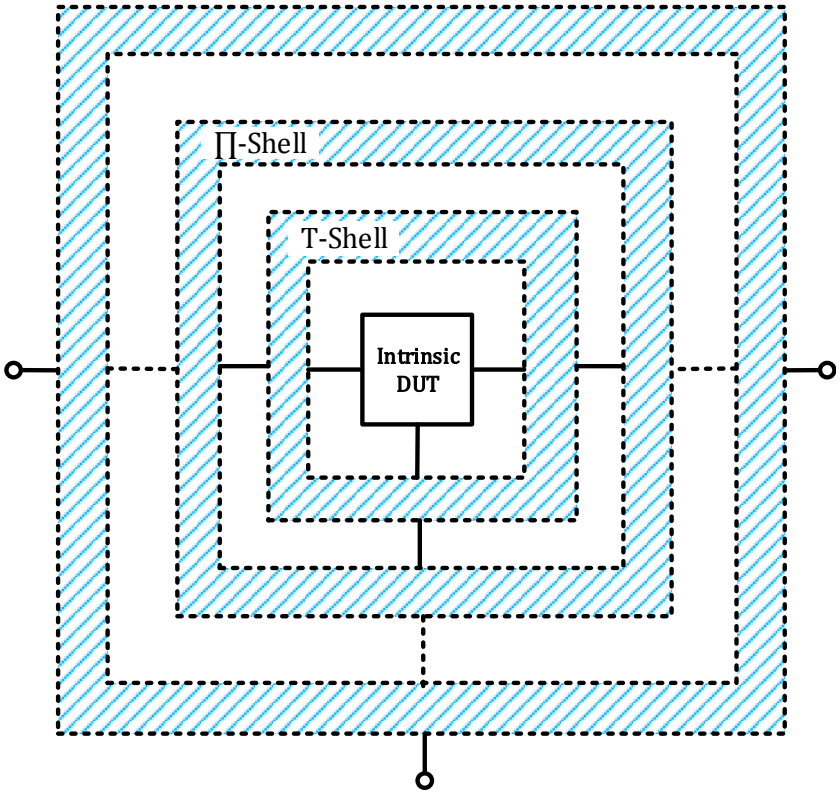


Figure 4.1: A schematic of the proposed linear-measurement based compact model. The blue-shaded area represents the extrinsic shell generation as successive T and Π-shells.

all biasing and frequency points). Thus, there is no need for going back and forth between ColdFET and HotFET. The proposed technique uses a general intrinsic topology in order to avoid derivation of complex analytical expressions, as well as a general extrinsic layered-topology.

4.2 Intrinsic and Extrinsic Shells Implementation using Neural Networks

The proposed compact model is built as layers (Figure 4.1). This is achieved by generating custom NNs in MATLAB to map the relationships between ports voltages and currents

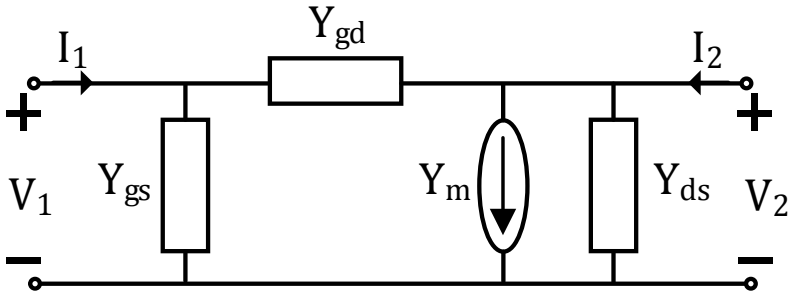


Figure 4.2: The Y-parameters equivalent circuit used as the intrinsic DUT of the proposed compact model.

using linear Y and Z-parameters[1]. First, an intrinsic shell is created followed by an extrinsic shell that is composed of successive T and Π-shells.

4.2.1 Intrinsic Shell NN

Technology-Independent compact models use any of the linear Y-parameters equivalent circuits shown in Figure 3.3 as the intrinsic shell. In this work, the output-circuit-dependent-generator form will be used (Figure 4.2), where the four complex admittances (Y_{gs}, Y_{gd}, Y_m, and Y_{ds}) are related to Y-parameters[2] (in frequency-domain) through (4.1).

$$\begin{bmatrix} Y_{11}(f) & Y_{12}(f) \\ Y_{21}(f) & Y_{22}(f) \end{bmatrix} = \begin{bmatrix} Y_{gs}(f) + Y_{gd}(f) & -Y_{gd} \\ Y_m(f) - Y_{gd} & Y_{ds}(f) + Y_{gd}(f) \end{bmatrix} \tag{4.1}$$

In physics-based compact models, the intrinsic shell topology is modified (by adding more physical elements, e.g. resistances, inductances, capacitances, diodes, etc) until it matches the DUT intrinsic behaviour as shown in Figure 2.3 (a) and (b). The technology-independent approach, however, matches the DUT intrinsic behaviour using the same topology for the intrinsic shell. This is achieved through two proposed implementation techniques that include the use of behavioural (non-physical) elements.

The first technique utilizes the generalized quasi-static assumption mentioned in Section 3.3.1 to describe the four admittances in Figure 4.2 using (4.2). This approach is important because it simplifies the transformation between frequency-domain and time-domain by

[1]It is worth noting that this procedure is the opposite of extrinsic shell de-embedding [11, 51].

[2]In this chapter, only linear network parameters are considered because the objective is to build a linear measurement-based compact model.

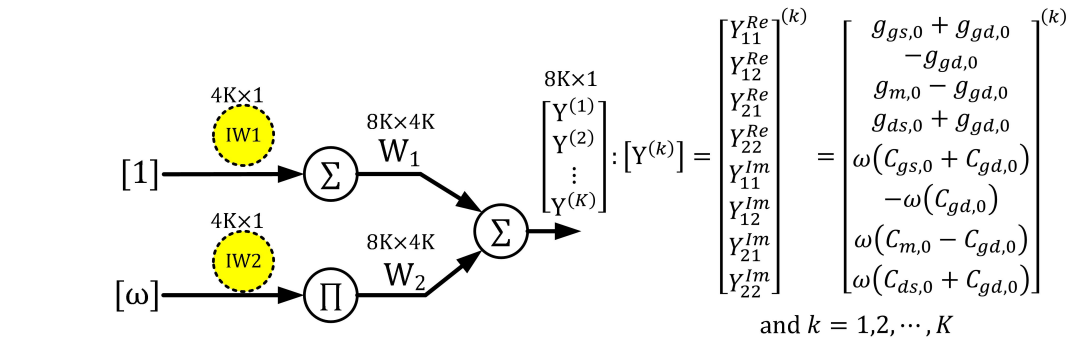
replacing $(j\omega)^m$ with d^m/dt^m. Therefore, a compact model that is based on frequency-domain Y-parameters can still be implemented in time-domain. It is worth mentioning that terms with $m = 0$ in (4.2) are of special importance as they represent a bridge between physics-based and technology-independent compact models. Thus, optimizing an extrinsic shell to force the intrinsic DUT to have only these terms might enhance the prediction capability of the model. Admittance-space X-parameters (XY-parameters) [40], for example, were proved to be frequency scalable even though X-parameters are not. Although both X and XY-parameters have the same information, the implementation of XY-parameters (using only $j\omega$ as frequency dependency) allows them to predict the behaviour at frequencies different than the ones used in extraction. Finally, quasi-static intrinsic shells can be implemented in time-domain (using Verilog-A) or in frequency-domain [using Symbolically Defined Device (SDD) components in ADS].

$$Y_x(f) = \sum_{m=0}^{\infty} [(j\omega)^{2m} g_{x,m} + (j\omega)^{2m+1} C_{x,m}] \quad : x \in \{gs, ds, m, gd\} \tag{4.2}$$

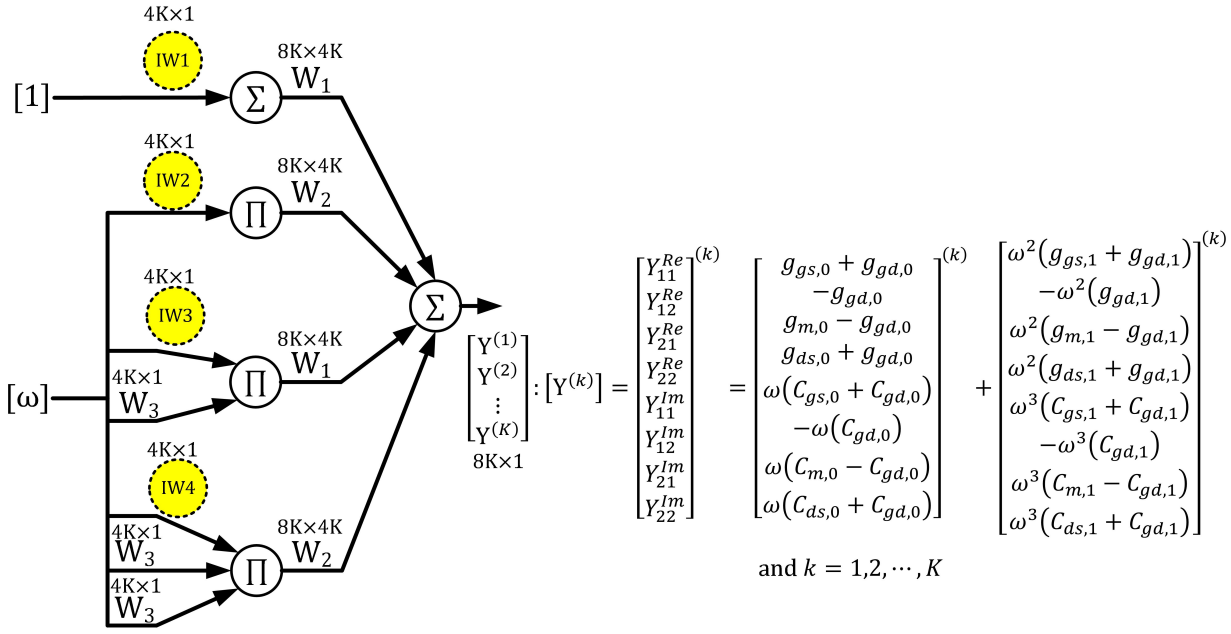
Figure 4.3 (a) and (b) shows two NNs that represent the quasi-static intrinsic shell in case of $m = 0$ and $m = 0, 1$ (4.2), respectively. These NNs have the angular frequency (ω) as input (in addition to a vector of ones) and the real and imaginary parts of Y-parameters as targets. Only input weights (IW's) (4.3) are trained while other weights (W's) are vectors and matrices of fixed values [Appendix C (C.1)]. Since intrinsic shell elements are bias-dependent, a superset is used in (4.3) as a biasing index, where K is the total number of DC biasing points.

$$\begin{aligned}
IW1 &= \begin{bmatrix} g^{(1)}_{gs,0} & g^{(1)}_{gd,0} & g^{(1)}_{m,0} & g^{(1)}_{ds,0} & \cdots & g^{(K)}_{gs,0} & g^{(K)}_{gd,0} & g^{(K)}_{m,0} & g^{(K)}_{ds,0} \end{bmatrix}^T \\
IW2 &= \begin{bmatrix} C^{(1)}_{gs,0} & C^{(1)}_{gd,0} & C^{(1)}_{m,0} & C^{(1)}_{ds,0} & \cdots & C^{(K)}_{gs,0} & C^{(K)}_{gd,0} & C^{(K)}_{m,0} & C^{(K)}_{ds,0} \end{bmatrix}^T \\
IW3 &= \begin{bmatrix} g^{(1)}_{gs,1} & g^{(1)}_{gd,1} & g^{(1)}_{m,1} & g^{(1)}_{ds,1} & \cdots & g^{(K)}_{gs,1} & g^{(K)}_{gd,1} & g^{(K)}_{m,1} & g^{(K)}_{ds,1} \end{bmatrix}^T \\
IW4 &= \begin{bmatrix} C^{(1)}_{gs,1} & C^{(1)}_{gd,1} & C^{(1)}_{m,1} & C^{(1)}_{ds,1} & \cdots & C^{(K)}_{gs,1} & C^{(K)}_{gd,1} & C^{(K)}_{m,1} & C^{(K)}_{ds,1} \end{bmatrix}^T
\end{aligned} \tag{4.3}$$

The second technique exploits NNs ability to replicate any behaviour. Thus, eight NNs can be used to reach eight expressions that represent the real and imaginary parts of the four admittances in Figure 4.2. These expressions are then implemented in ADS using SDD components. It is worth noting that this technique can only be implemented in frequency-domain. Also, having complex frequency dependencies in the model slows the



(a)



(b)

Figure 4.3: The proposed NNs that represent the quasi-static intrinsic shell using (4.2) for: (a) $m = 0$, and (b) $m = 0, 1$.

simulation speed and might affect its convergence. Therefore, this technique is preferred to be used with nonlinear circuits that have complex frequency behaviours. For example, it can be used to build nonlinear models for PAs where input and output matching networks generate frequency behaviours that cannot be matched using the first technique. Another use for this technique is to generate large-signal-models. The detailed implementation of

this technique will be discussed in the next chapter (Section 5.2.2).

4.2.2 Extrinsic Π and T-Shells NNs

Figure 4.4 (a) shows a Π-shell, which consists of three conductances (g_1 g_2, and g_3) and three capacitances (C_1 C_2, and C_3), added to the DUT. The relationship between the resulting total Y-parameters, DUT Y-parameters and Π-shell components is described by (4.4).

$$\begin{aligned}\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}^{Total} &= \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}^{DUT} \\ &+ \begin{bmatrix} g_1 + g_2 & -g_2 \\ -g_2 & g_3 + g_2 \end{bmatrix} + j\omega \begin{bmatrix} C_1 + C_2 & -C_2 \\ -C_2 & C_3 + C_2 \end{bmatrix}\end{aligned} \tag{4.4}$$

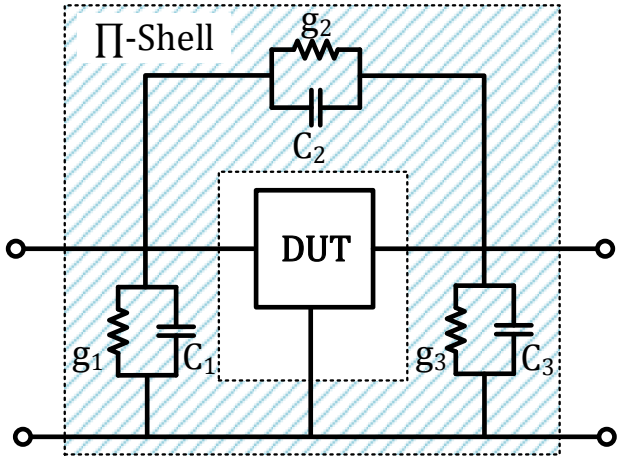
Figure 4.4 (b) shows the proposed NN representation for (4.4), where the inputs are ω, a vector of ones, and the DUT's Y-parameters, and the outputs are the real and imaginary parts of total Y-parameters. Since extrinsic shell elements are bias-independent, only six elements represented as input weights $IW1$ and $IW2$ (4.5) are trained while other weights (W's) are vectors and matrices of fixed values [Appendix C (C.2)].

$$\begin{aligned} IW1 &= \begin{bmatrix} g_1 & g_2 & g_3 \end{bmatrix}^T \\ IW2 &= \begin{bmatrix} C_1 & C_2 & C_3 \end{bmatrix}^T \end{aligned} \tag{4.5}$$

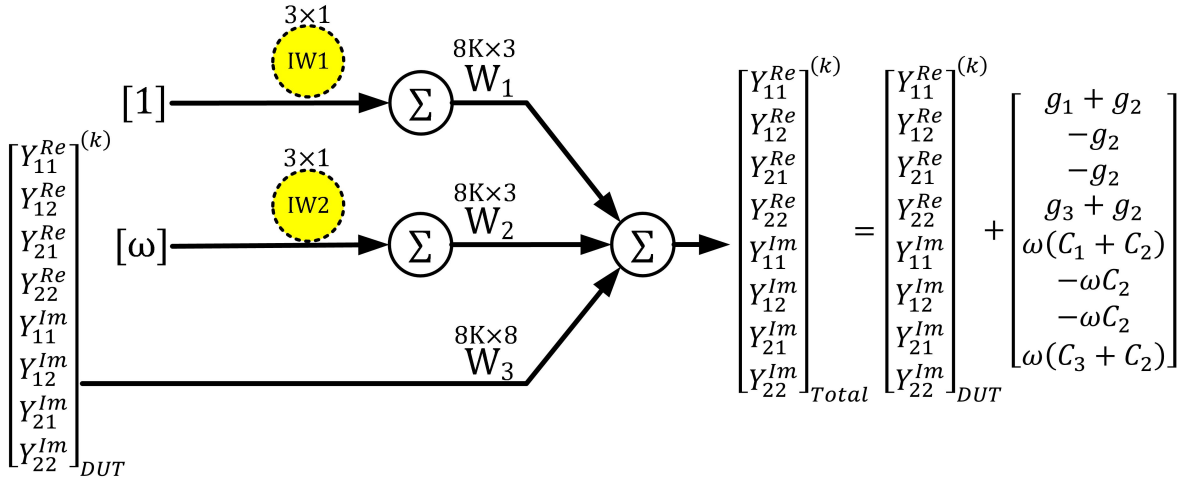
Similarly, Figure 4.5 (a) shows a T-shell, that consists of three resistances (R_1 R_2, and R_3) and three inductances (L_1 L_2, and L_3), added to the DUT. The relationship between the resulting total Z-parameters, DUT Z-parameters, and T-shell components is described by (4.6), where the proposed NN representation is shown in Figure 4.5 (b). Again, only six elements represented as input weights $IW1$ and $IW2$ (4.7) are trained while other weights (W's) are vectors and matrices of fixed values [Appendix C (C.3)].

$$\begin{aligned}\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}^{Total} &= \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}^{DUT} \\ &+ \begin{bmatrix} R_1 + R_2 & R_2 \\ R_2 & R_3 + R_2 \end{bmatrix} + j\omega \begin{bmatrix} L_1 + L_2 & L_2 \\ L_2 & L_3 + L_2 \end{bmatrix}\end{aligned} \tag{4.6}$$

$$\begin{aligned} IW1 &= \begin{bmatrix} R_1 & R_2 & R_3 \end{bmatrix}^T \\ IW2 &= \begin{bmatrix} L_1 & L_2 & L_3 \end{bmatrix}^T \end{aligned} \tag{4.7}$$

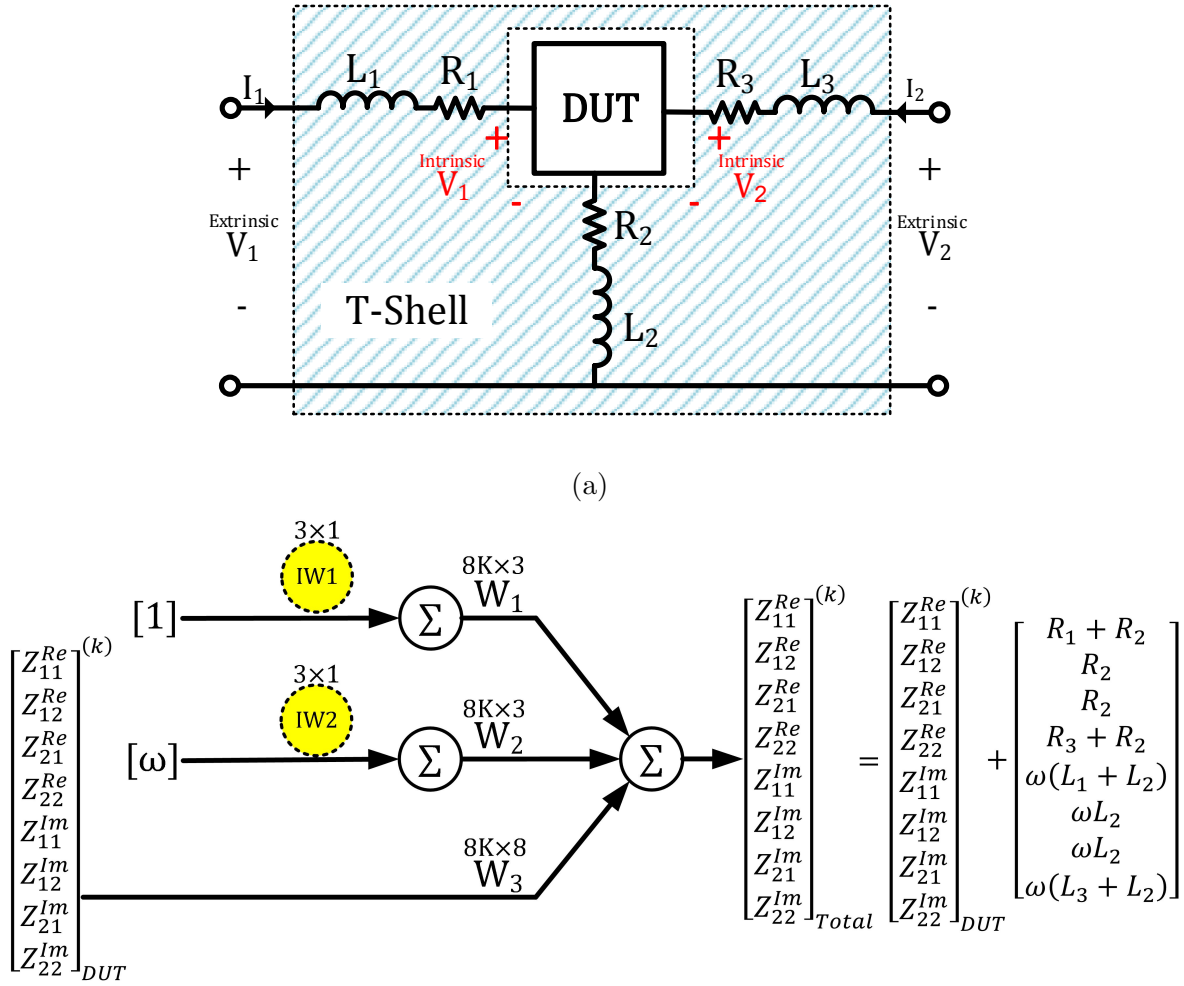


(a)



(b)

Figure 4.4: The addition of a Π-extrinsic shell to the DUT: (a) schematic, and (b) proposed NN representation.



(a)

(b)

Figure 4.5: The addition of a T-extrinsic shell to the DUT: (a) schematic, and (b) proposed NN representation.

4.2.3 Generating the Full Neural Network Algorithm

Figure 4.6 shows the algorithm used for building the compact model using the proposed layered NNs technique. First, an intrinsic layer is generated with $m = 0$, followed by one T and one Π-extrinsic shells (T-Π-shell). The generated NN is then initialized and trained using Levenberg-Marquardt back-propagation algorithm with a target mean squared error (MSE) of 1e-6, where all Y-parameters (versus both biasing and frequency grids) are used as targets. If the required MSE could not be reached, then another T-Π-shell is added (up to three shells). If the generated network is still not sufficient to meet the required MSE, the intrinsic-layer is updated to include $m = 0, 1$ and then the same steps are repeated. The following remarks should be noted.

- Y-parameters of the intrinsic shell need to be converted to Z-parameters before adding the first extrinsic T-shell (4.6). This is achieved by using the NN shown in Figure 4.7, where the input is a complex 2×2 matrix $[A]$ and the output is the inverted version of this matrix $[A]^{-1}$. All weights (W's) are fixed for this NN [Appendix C (C.4)]. This matrix inversion NN has to be inserted between each extrinsic T and Π-shells as well.

- The back-propagation algorithm is used for NNs training to minimize a cost function that includes either Y, Z, or S-parameters[3]. Equation (4.8), for example, shows a cost function described by the MSE between real and imaginary parts of measured Y-parameters and those generated by the NN (associated with a superset NN). K is the total number of DC biasing points and $x \in \{11, 12, 21, 22\}$.

$$MSE = \frac{1}{K} \sum_{k=1}^{K} \sum_{x} [Real(Y_x - Y_x^{NN})^2 + Imag(Y_x - Y_x^{NN})^2] \tag{4.8}$$

  Z-parameters can be used in the cost function instead of Y-parameters by choosing the last extrinsic layer to be a T-shell or a Π-shell with a matrix inversion NN (Figure 4.7). Finally, to use S-parameters in the cost function, another NN that converts Y-parameters to $\Delta\times$S-parameters is used (Figure 4.8), where Δ is described by (4.9), and W's have fixed values [Appendix C (C.5)].

$$\Delta = (1 + Z_o Y_{11})(1 + Z_o Y_{22}) - Z_o^2 Y_{12} Y_{21} \tag{4.9}$$

  where Z_o is the system characteristic impedance.

[3]The last layer of the extrinsic shell determines the type of network parameters used as targets and consequently used in the cost function.

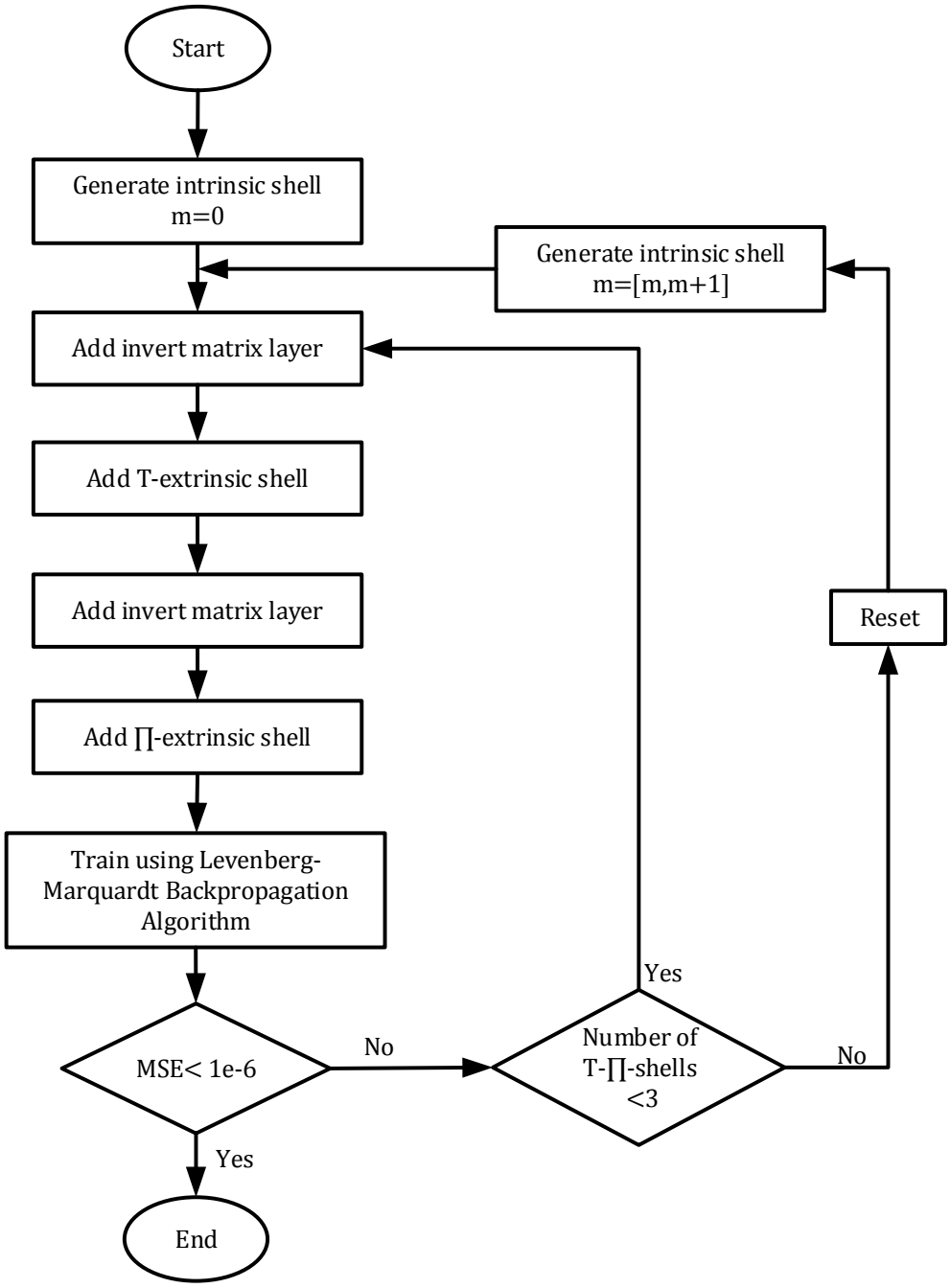


Figure 4.6: A block diagram explaining the algorithm used for generating the proposed model using NNs.

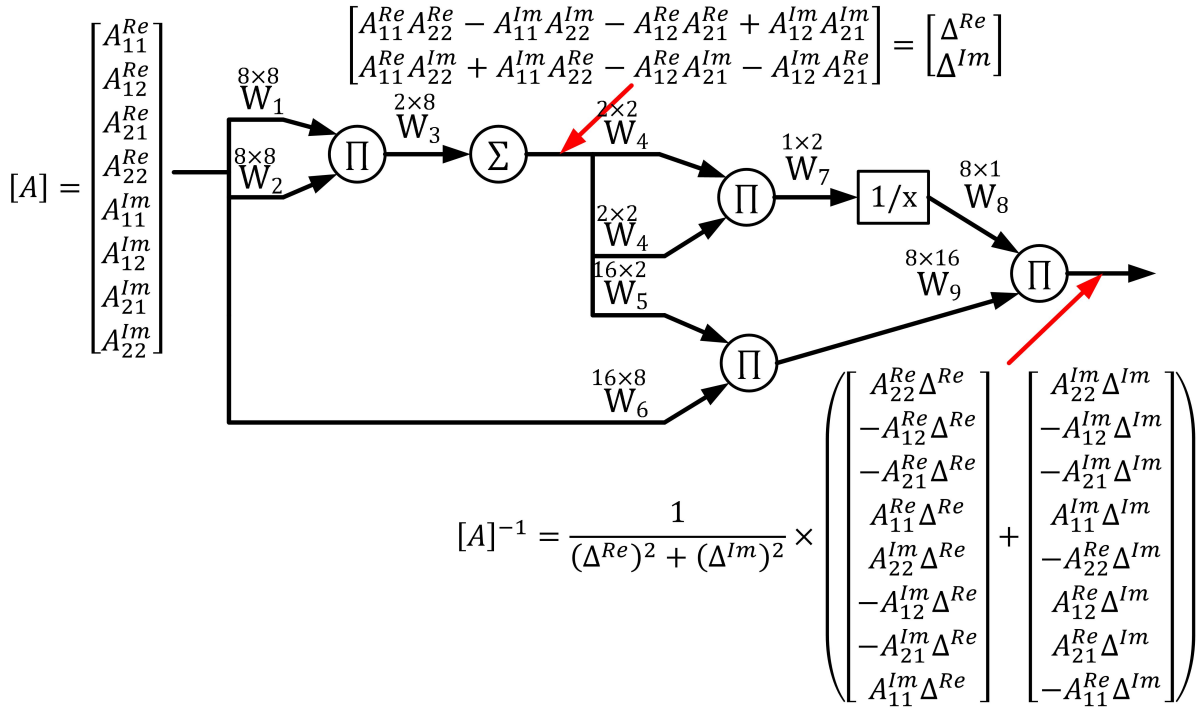


Figure 4.7: The proposed NN for inverting complex 2×2 matrices.

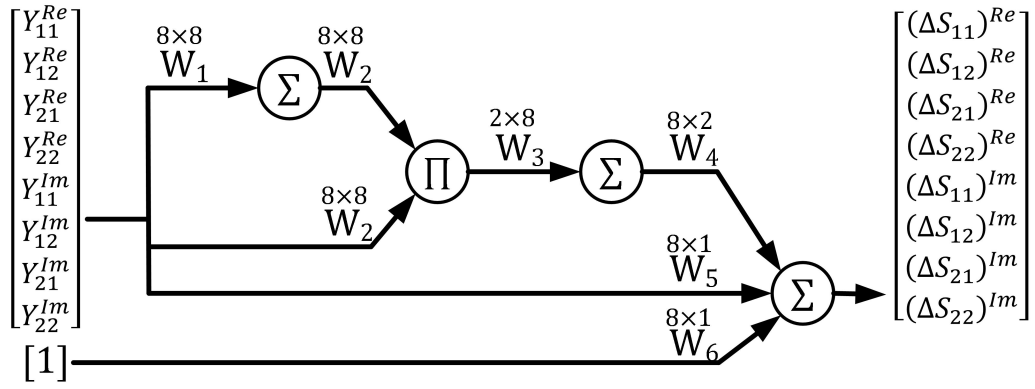


Figure 4.8: The proposed NN to convert Y-parameters to $\Delta \times$S-Paraemters, where Δ is described by (4.9).

- Table 4.1 shows the number of layers added by each NN shell in addition to the number of elements that will be trained. It should be noted that the matrix inversion NN (Figure 4.7) is the bottleneck of the proposed compact model implementation as it adds eight extra layers each time it is being used. Also, it uses a $1/x$ activation function, hence slowing down both training speed and convergence.

	# of layers	# of elements that will be trained
Quasi-static intrinsic shell $m = 0$	3	$8 \times K$
Quasi-static intrinsic shell $m = 0, 1$	5	$16 \times K$
Extrinsic Π-shell	3	6
Extrinsic T-shell	3	6
2×2 complex matrix inversion	8	0
Y-parameters to $\Delta\times$S-parameters conversion	4	0

Table 4.1: The number of layers and elements that will be trained added by each NN shell, where K is the total number of DC biasing points.

- In order to improve the NN training, all capacitances and inductances, as well as input angular frequency (ω) are normalized with respect to the maximum ω value.

- Since intrinsic shell elements are functions of DC biasing points, attention should be paid when using resistances in extrinsic T-shells as the intrinsic DC values become different than the extrinsic DC values (Figure 4.5). As a result, intrinsic shell elements become functions of a different DC biasing grid described by (4.10). It should be noted that the new biasing grid is nonuniform (even if the original biasing grid used by the measurement setup was uniform), which makes it more challenging to train compact models. Therefore, all extrinsic T-shells will have only inductances.

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}_{DC}^{Intrinsic} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}_{DC}^{Extrinsic} - \begin{bmatrix} R_1 + R_2 & R_2 \\ R_2 & R_3 + R_2 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}_{DC} \tag{4.10}$$

	L1 (nH)	L2 (nH)	L3 (nH)	C1 (pF)	C2 (pF)	C3 (pF)
1st T-shell	0.87	0.046	0.88			
1st Π-shell				0.41	-0.005	0.4

Table 4.2: The extracted extrinsic T-Π-shell elements values of the generated model.

4.3 Validation

For validation purpose, simulated S-parameters were generated for Cree CGH40010P 10 Watt GaN HEMT using ADS. V_{gs} and V_{ds} were swept from -4 to 0 V, and from 0 to 50 V, respectively. While a uniform step of 0.5 V was used for V_{gs} sweep, a nonuniform step was used to cover V_{ds} range using 12 points. The frequency was swept from 0.1 to 4 GHz with a step of 0.2 GHz.

The algorithm described in the previous section was applied. Figure 4.9 shows the final NN (compact model), where an intrinsic layer with $m = 0, 1$ and one T-Π-extrinsic shell were sufficient to achieve the required MSE. Table 4.2 and Figure 4.10 show the extracted extrinsic and intrinsic elements, respectively. Although the extracted inductances and capacitances could be forced to have positive values and ColdFET could have been used to initialize the extrinsic shell, this was not considered as the objective was to build a technology independent compact model (as stated in Section 4.1). The generated model was implemented in ADS using SDD components, as shown in Figure 4.11 (a). S-parameters were then simulated using biasing and frequency grids finer than the grids[4] used in training (to test the interpolation capability of the proposed technique), and compared to original S-parameters. The proposed technique showed a great ability to match the required S-parameters with a total percentage error of 2.8 % (the generalized error expression (4.11) was used [52]). A sample of original and simulated S-parameters at four biasing points is shown in Figure 4.11 (b)-(e).

$$\%Error = 100 \times \sqrt{\frac{\sum_{k=1}^{K} \sum_{x} |S_x - S_x^{NN}|^2}{\sum_{k=1}^{K} \sum_{x} |S_x^2|}} \quad : x \in \{11, 12, 21, 22\} \tag{4.11}$$

Finally, it is worth mentioning that the quasi-static intrinsic shell technique should be limited to $m \leq 2$. As shown in Figure 4.10, g_1's and C_1's values are in the range of 1e-23 and 1e-33 because they are multiplied by ω^2 and ω^3, respectively. These ω values are too big and might cause large-signal simulators to diverge. An alternative approach is to use

[4]The new V_{gs}, V_{ds}, and frequency steps were 0.2 V, 1 V and 0.1 GHz, respectively.

the second technique for intrinsic shells (Section 4.2.1), where a solution is guaranteed to be found.

4.4 Conclusion

A novel technique to build a technology-independent linear-measurement based compact model was proposed, where layered-NNs were used to try different topologies for extrinsic and intrinsic shells. This technique eliminates the need to derive complex analytical expressions for each topology. It also extracts both extrinsic and intrinsic shell elements simultaneously without going back and forth between ColdFET and HotFET. Finally, the ability of the proposed model to reproduce S-parameters was validated in the simulation environment.

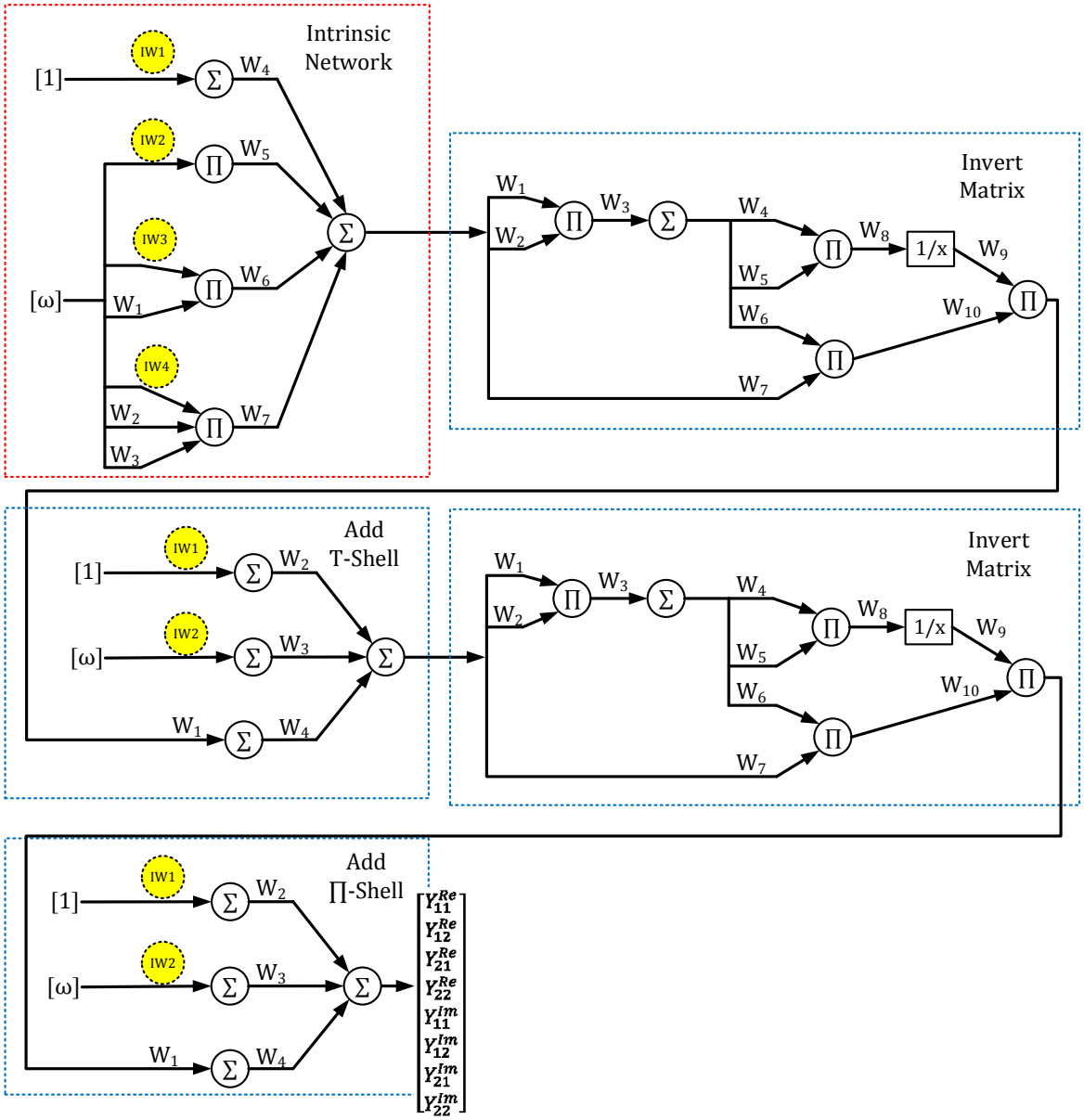
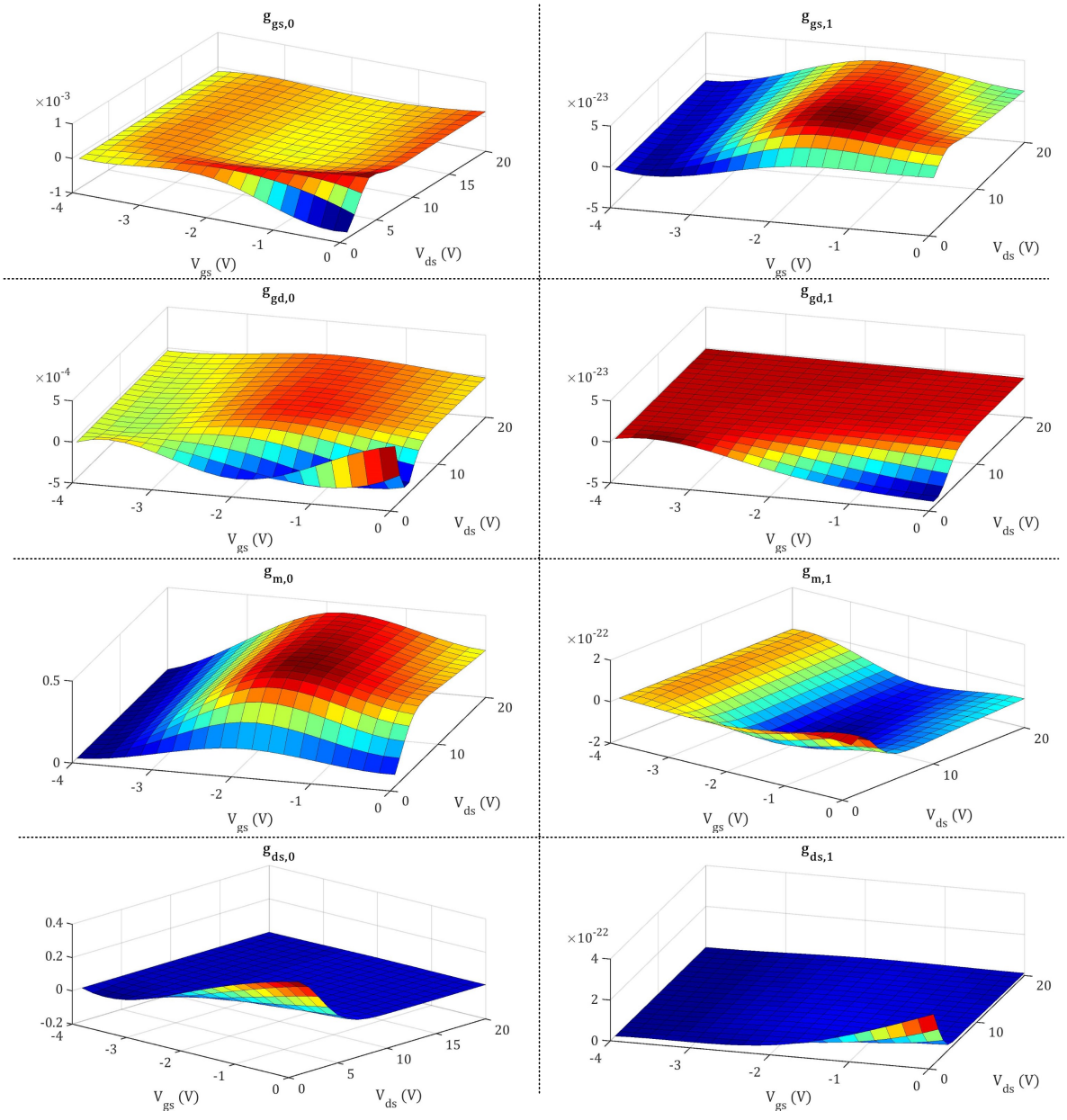
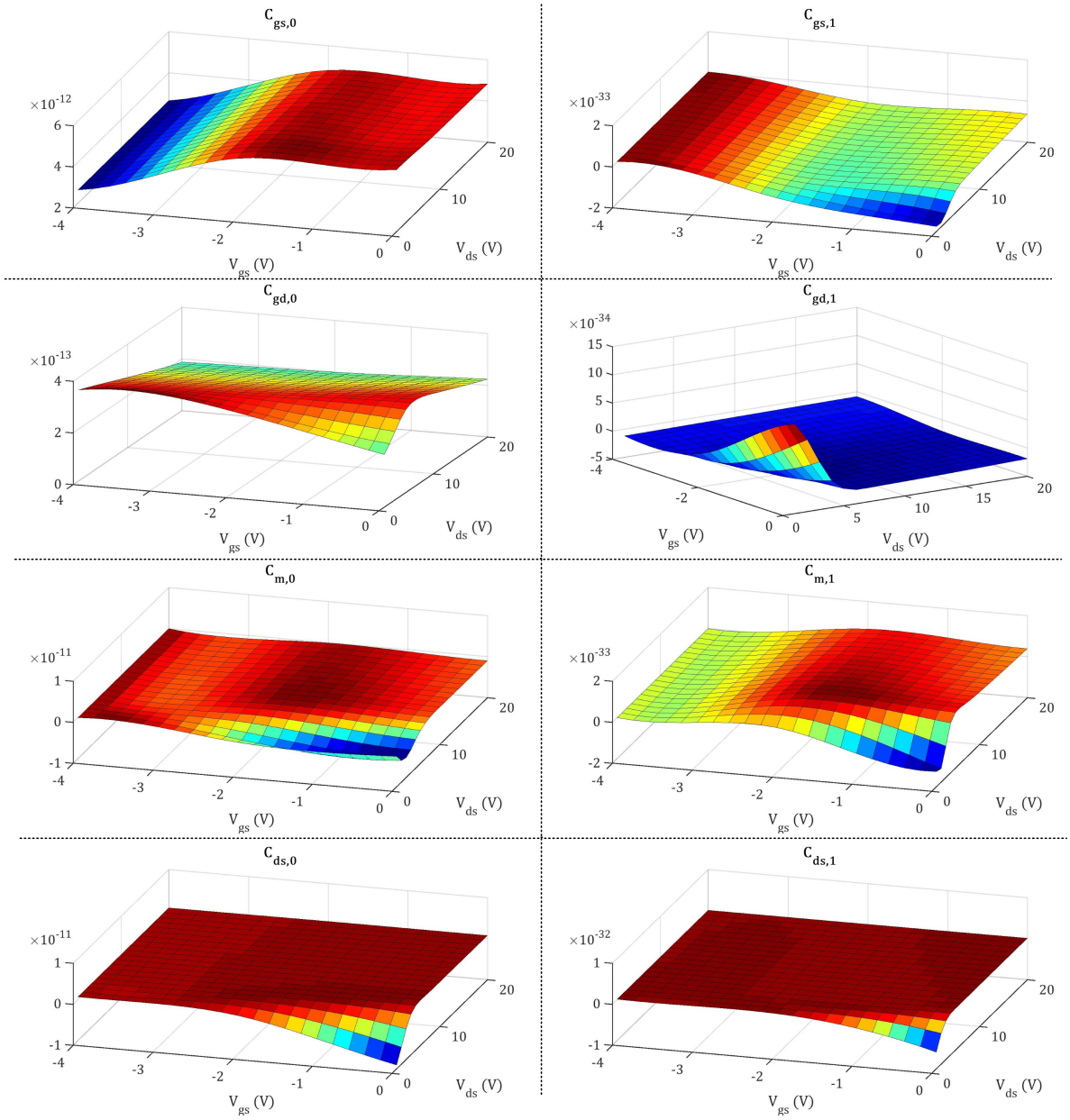


Figure 4.9: The generated NN using the proposed layered-NNs technique.


$g_{gs,0}$
$\times 10^{-3}$
1
0
-1
-4
-3
-2
-1
0
V_{gs} (V)
0
5
10
15
20
V_{ds} (V)
$g_{gs,1}$
$\times 10^{-23}$
5
0
-5
-4
-3
-2
-1
0
V_{gs} (V)
0
10
20
V_{ds} (V)
$g_{gd,0}$
$\times 10^{-4}$
5
0
-5
-4
-3
-2
-1
0
V_{gs} (V)
0
10
20
V_{ds} (V)
$g_{gd,1}$
$\times 10^{-23}$
5
0
-5
-4
-3
-2
-1
0
V_{gs} (V)
0
10
20
V_{ds} (V)
$g_{m,0}$
0.5
0
-4
-3
-2
-1
0
V_{gs} (V)
0
10
20
V_{ds} (V)
$g_{m,1}$
$\times 10^{-22}$
2
0
-2
-4
-3
-2
-1
0
V_{gs} (V)
0
10
20
V_{ds} (V)
$g_{ds,0}$
0.4
0.2
0
-0.2
-4
-3
-2
-1
0
V_{gs} (V)
0
5
10
15
20
V_{ds} (V)
$g_{ds,1}$
$\times 10^{-22}$
4
2
0
-4
-3
-2
-1
0
V_{gs} (V)
0
10
20
V_{ds} (V)

(a)



(b)

Figure 4.10: The extracted intrinsic elements versus the biasing grid: (a) conductances, and (b) capacitances.

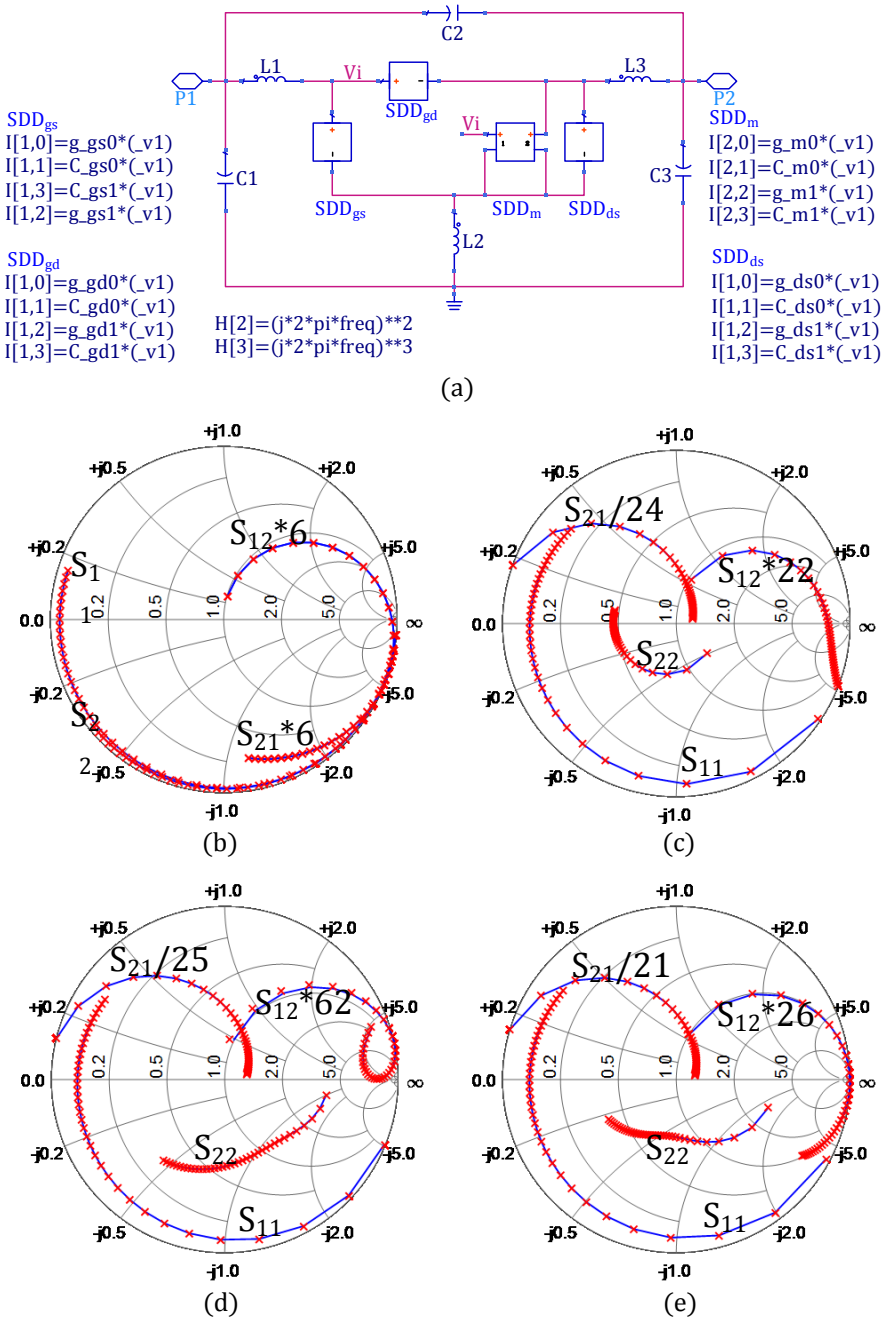


Figure 4.11: (a) The implementation of the proposed compact model in ADS. Original [red (x)] and simulated [blue (-)] S-parameters at four biasing points (b) $(V_{gs}, V_{ds}) = (-4, 1)$ V, (c) $(V_{gs}, V_{ds}) = (-2.5, 5)$ V, (d) $(V_{gs}, V_{ds}) = (-1.5, 50)$ V, and (e) $(V_{gs}, V_{ds}) = (0, 20)$ V.

Chapter 5

Compact Model Implementation Part 2: HONPs-Based Compact Model

5.1 Introduction

The advancements in transistor compact modelling have been driven by the ever-increasing challenges and intensified requirements circuit designers face. In the particular case of high-power transistors, used in power amplifiers (PA) for wireless communication, compact models have mainly aimed for accurately predicting both output power and efficiency. Generally, they can be divided into two main categories. The first category includes compact models that are extracted using DC and S-parameters measurements (linear measurement-based compact models mentioned in Chapter 4). Recently, however, they have been outperformed by the second category, which takes advantage of the nonlinear vector network analyzer (NVNA) measurements to enhance the model's performance under large-signal conditions (nonlinear measurement-based compact models mentioned in Section 2.2.2). State-of-the-art NVNA-based compact models, e.g. the DynaFET model [25, 27] and the Dynamic-bias model [15, 28], have demonstrated a great ability to predict device performance under continuous wave (CW) stimulus at the fundamental and harmonic frequencies, which is adequate to design for output power and efficiency. However, they fall short in enabling the optimization of the PA linearity at its design stage under modulated signal stimuli.

A well-known fact about behavioural modelling is that "N^{th}-order I/V derivatives are

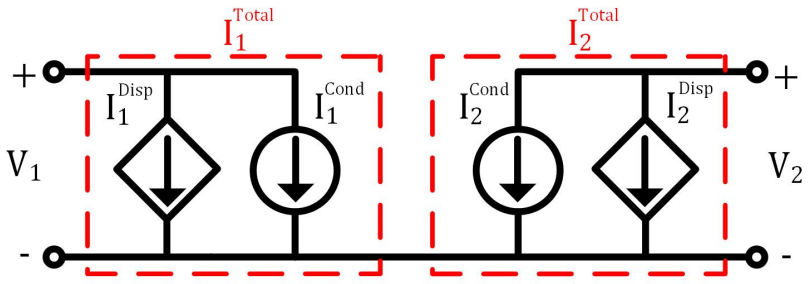


Figure 5.1: The 2-Port LSM used in compact model implementation.

needed to model N^{th}-order distortion" [53]. Failing to meet this condition is the main reason why existing compact models do not achieve the required accuracy under modulated signal stimuli. In this chapter, a novel large-signal-model (LSM) that bridges the gap between compact and behavioural models is proposed. The proposed LSM utilizes high-order network parameters (HONPs), introduced in Chapter 3, as the compact model building block. Therefore, enables describing the device, at any biasing point, by a weakly nonlinear (WNL) circuit, and hence enabling accurate calculations of I/V derivatives. The upcoming sections discuss the model formulation and implementation using neural networks (NNs). Also, the model validation in both simulation and measurement environments will be demonstrated.

5.2 HONPs-Based Compact Model Formulation and Implementation

Compact model formulations have to be driven with both model implementation and model extraction in mind. Since most of the existing commercial circuit simulators use Modified Nodal Analysis method, implementing compact models in the nodal form, i.e. ports currents are functions of ports voltages, helps in the simulation convergence [47]. Thus, the 2-port LSM shown in Figure 5.1 is used in the literature to describe transistor compact models as two nonlinear current sources (I_1^{Total} and I_2^{Total}). Each current source is then divided into conduction (I^{Cond}) and displacement (I^{Disp}) current sources to simplify model extraction (5.1), where I^{Disp} is expressed as a time derivative of a charge source (Q). In technology-independent linear measurement-based compact models, for example, conduction and displacement currents are constructed from the real and imaginary parts of Y-parameters, respectively. In nonlinear measurement-based compact models, they are

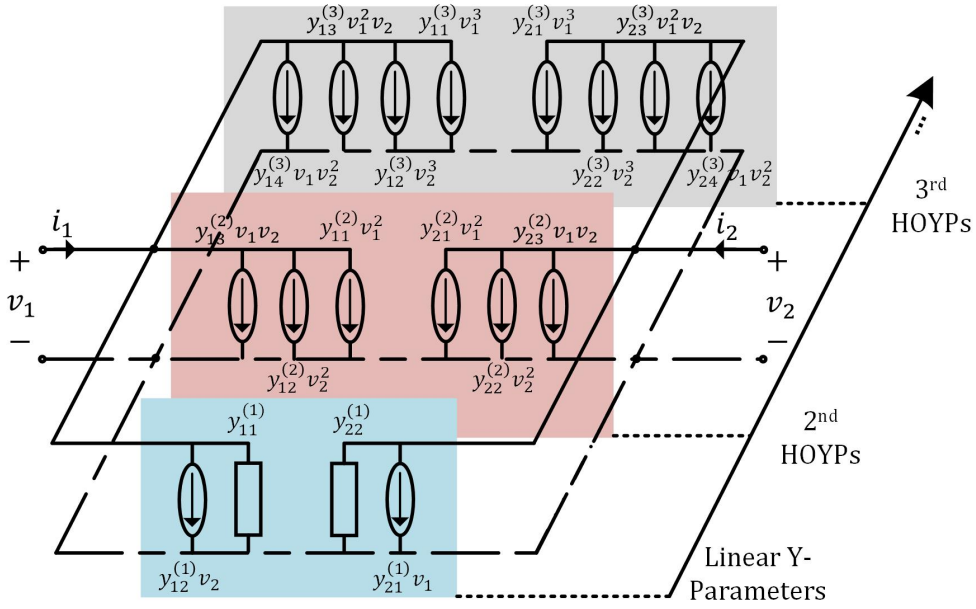


Figure 5.2: First (blue), second (red), and third (grey) HOYPs equivalent circuit.

identified separately by exciting the transistor with low-frequency and high-frequency stimuli (as mentioned in Section 2.2.2).

$$\begin{aligned} I_p^{Total}[V_1(t), V_2(t)] &= I_p^{Cond}[V_1(t), V_2(t)] + I_p^{Disp}[V_1(t), V_2(t)] \\ &= I_p^{Cond}[V_1(t), V_2(t)] + \frac{d}{dt} Q_p[V_1(t), V_2(t)] \end{aligned} \tag{5.1}$$

where,

I_p^{Total} is the large-signal time-domain current at port p
I_p^{Cond} is the large-signal time-domain connection current at port p
I_p^{Dips} is the large-signal time-domain displacement current at port p
Q_p is the large-signal time-domain charge source at port p
V_p is the large-signal time-domain voltage at port p
$p = 1, 2$ is the port number.

The proposed compact model uses HONPs (Figure 5.2) as the model's building block, and hence conduction and displacement current sources are built in frequency-domain from

the real and imaginary parts of HONPs. Each one of these sources can be described as a nonlinear function of the ports voltages and frequency, $I(V_1, V_2, f)$. Since frequency and voltages cannot be implemented as indexes of the same nonlinear function in circuit simulators, the original function $I(V_1, V_2, f)$ is divided into two nonlinear functions: $X(V_1, V_2)$ and $Z(f)$, as described by (5.2)[1]. While the first function describes HONPs as static current derivatives with respect to voltage, the later takes care of their dynamics. The details of these two functions and their implementations are discussed in the following subsections.

$$I_p(V_1, V_2, f) = X_p(V_1, V_2) Z_p(f) \tag{5.2}$$
where $p = 1, 2$ is the port number.

5.2.1 High-Order Current Derivatives NN: $X(V_1, V_2)$

Linear Y-parameters represent the first-order derivatives of ports currents with respect to ports voltages (2.2) [16]. Similarly, by comparing HONPs in the Y-form (HOYPs) (3.4) to the two-dimensional Taylor Expansion (TE) of $I_p(V_1, V_2, f)$ around $(V_{1,dc}, V_{2,dc})$ (5.3), HOYPs represent currents high-order derivatives with respect to voltages (5.4)-(5.6). $[Y^{(1)}]$, $[Y^{(2)}]$, and $[Y^{(3)}]$ are first, second, and third-HOYPs and the dependency of X on V_1 and V_2 is removed for notation simplicity.

$$TE = \sum_{\substack{0 \leqslant l,m \leqslant n \\ l+m \leqslant n}} \frac{1}{l!m!} \frac{\partial^{l+m} I_p(V_1, V_2, f)}{\partial V_1^l \partial V_2^m}\bigg|_{\substack{V_1 = V_{1,dc} \\ V_2 = V_{2,dc}}} \times v_1^l v_2^m \tag{5.3}$$

$$\left[Y^{(1)}\right] = \begin{bmatrix} Y_{11}^{(1)} & Y_{12}^{(1)} \\ Y_{21}^{(1)} & Y_{22}^{(1)} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial X_1}{\partial V_1} Z_1(f) & \dfrac{\partial X_1}{\partial V_2} Z_1(f) \\ \dfrac{\partial X_2}{\partial V_1} Z_2(f) & \dfrac{\partial X_2}{\partial V_2} Z_2(f) \end{bmatrix} \tag{5.4}$$

$$\left[Y^{(2)}\right] = \begin{bmatrix} Y_{11}^{(2)} & Y_{12}^{(2)} & Y_{13}^{(2)} \\ Y_{21}^{(2)} & Y_{22}^{(2)} & Y_{23}^{(2)} \end{bmatrix} \tag{5.5}$$

$$= \begin{bmatrix} \dfrac{1}{2}\dfrac{\partial^2 X_1}{\partial V_1^2} Z_1(f) & \dfrac{1}{2}\dfrac{\partial^2 X_1}{\partial V_2^2} Z_1(f) & \dfrac{\partial^2 X_1}{\partial V_1 V_2} Z_1(f) \\ \dfrac{1}{2}\dfrac{\partial^2 X_2}{\partial V_1^2} Z_2(f) & \dfrac{1}{2}\dfrac{\partial^2 X_2}{\partial V_2^2} Z_2(f) & \dfrac{\partial^2 X_2}{\partial V_1 V_2} Z_2(f) \end{bmatrix}$$

[1]For the proposed compact model final formulation, please refer to (5.12).

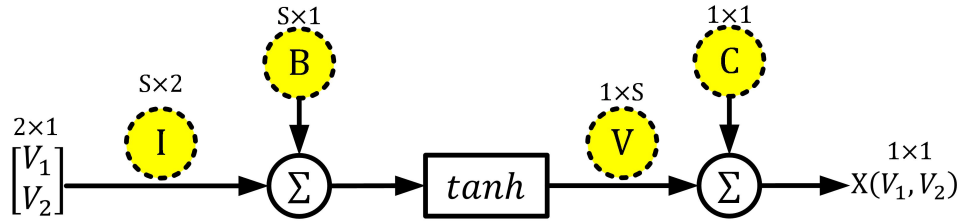


Figure 5.3: A one-hidden-layer feedforward NN with S hidden neurons that represents $X(V_1, V_2)$. Matrix notation is used to describe I, V, B and C, where their sizes are shown above each circle.

$$\begin{aligned}\left[Y^{(3)}\right] &= \begin{bmatrix} Y_{11}^{(3)} & Y_{12}^{(3)} & Y_{13}^{(3)} & Y_{14}^{(3)} \\ Y_{21}^{(3)} & Y_{22}^{(3)} & Y_{23}^{(3)} & Y_{24}^{(3)} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3!}\frac{\partial^3 X_1}{\partial V_1^3}Z_1(f) & \frac{1}{3!}\frac{\partial^3 X_1}{\partial V_2^3}Z_1(f) & \frac{1}{2}\frac{\partial^3 X_1}{\partial V_1^2 V_2}Z_1(f) & \frac{1}{2}\frac{\partial^3 X_1}{\partial V_1 V_2^2}Z_1(f) \\ \frac{1}{3!}\frac{\partial^3 X_2}{\partial V_1^3}Z_2(f) & \frac{1}{3!}\frac{\partial^3 X_2}{\partial V_2^3}Z_2(f) & \frac{1}{2}\frac{\partial^3 X_2}{\partial V_1^2 V_2}Z_2(f) & \frac{1}{2}\frac{\partial^3 X_2}{\partial V_1 V_2^2}Z_2(f) \end{bmatrix}\end{aligned} \tag{5.6}$$

Unavoidable measurements uncertainties make $X(V_1, V_2)$ non-conservative over the modelling space. Therefore, $X(V_1, V_2)$ cannot be calculated by directly integrating the derivatives (5.4)-(5.6) because the integration is path-dependent. Hence, forcing the integration through a particular path will degrade the model's accuracy thus wasting the benefits gained by using HONPs. This problem is referred to as the integrability condition [9], and few solutions were presented in the literature for the linear parameters case. In [17], integration path (IP) was included in the modelling space resulting in a 3D function $X(V_1, V_2, \text{IP})$, which was then implemented as a 3D table-based model. A more elegant solution was introduced in [54], where an adjoint neural network (NN) was simultaneously trained along with the original NN to provide the charge functions given only the partial derivative data. This technique was then used in Keysight NeuroFET Model [55].

In this work, the integrability problem is more complicated because of the introduction of high-order derivatives represented by HOYPs. A new NN technique -that allows optimizing a nonlinear function to simultaneously fit first and high-order derivatives- is proposed. Figure 5.3 shows a one-hidden-layer feedforward NN with S neurons and a hyperbolic tangent ($tanh$) activation function. This NN has two inputs (V_1, V_2) and one output $[X(V_1, V_2)]$ and can be described by (5.7), where $I_{s,1}$, $I_{s,2}$, V_s and B_s are the s^{th} neuron's parameters and C is a constant. These parameters are shown in the matrix form (5.8)

in Figure 5.3 and are trained, using any of the back-propagation algorithms, to generate $X(V_1, V_2)$ that matches specific training target with an acceptable error.

$$X(V_1, V_2) = \sum_{s=1}^{S} V_s tanh(I_{s,1}V_1 + I_{s,2}V_2 + B_s) + C \tag{5.7}$$

$$\begin{aligned} I &= \begin{bmatrix} I_{1,1} & I_{2,1} & \dots & I_{S,1} \\ I_{1,2} & I_{2,2} & \dots & I_{S,2} \end{bmatrix}^T \\ V &= \begin{bmatrix} V_1 & V_2 & \dots & V_S \end{bmatrix} \\ B &= \begin{bmatrix} B_1 & B_2 & \dots & B_S \end{bmatrix}^T \end{aligned} \tag{5.8}$$

While $X(V_1, V_2)$ is the unknown in this problem, its derivatives can be identified from HOYPs. Therefore, (5.7) is derived up to the n^{th}-order and the results are shown in (5.9), where each n^{th}-order derivative gives $n+1$ expressions. Custom NNs were then built using MATLAB to map these expressions, Figure 5.4. The following remarks should be noted:

- The original parameters I, V, and B (showed in yellow circles) are the only parameters that will be trained while other weights (W's) are vectors and matrices with fixed values [Appendix D (D.1)].
- n^{th}-order derivative networks use n^{th}-order derivative hyperbolic tangent $[tanh^{(n)}]$ as activation functions. It is worth mentioning that these functions are non-monotonic, Figure 5.5. Therefore, an active-range was assigned to each function to limit their training to a range where they exhibit a monotonic behaviour. This technique speeds up the training and prevents it from being trapped in local minima.
- Simultaneous training for the first, second, and up to the n^{th}-order derivatives versus K biasing points is illustrated in Figure 5.4, where the red rectangles show inputs and targets at corresponding biasing points.
- Since different orders of HOYPs have different orders of magnitudes, training weights were adjusted to give all HOYPs the same weight in the training cost function.

$$\begin{aligned} \frac{\partial^n X(V_1, V_2)}{\partial V_1^n} &= \sum_{s=1}^{S} V_s I_{s,1}^n tanh^{(n)}(I_{s,1}V_1 + I_{s,2}V_2 + B_s) \\ \frac{\partial^n X(V_1, V_2)}{\partial V_1^{n-1} V_2} &= \sum_{s=1}^{S} V_s I_{s,1}^{n-1} I_{s,2} tanh^{(n)}(I_{s,1}V_1 + I_{s,2}V_2 + B_s) \end{aligned}$$

$$\vdots \tag{5.9}$$

$$\frac{\partial^n X(V_1,V_2)}{\partial V_1 V_2^{n-1}} = \sum_{s=1}^{S} V_s I_{s,1} I_{s,2}^{n-1} tanh^{(n)}(I_{s,1}V_1 + I_{s,2}V_2 + B_s)$$

$$\frac{\partial^n X(V_1,V_2)}{\partial V_2^{n}} = \sum_{s=1}^{S} V_s I_{s,2}^{n} tanh^{(n)}(I_{s,1}V_1 + I_{s,2}V_2 + B_s)$$

$$\text{where } n = 1, 2, \ldots, N$$

5.2.2 Dynamics NN: $Z(f)$

HONPs dynamics refer to any frequency behaviour they exhibit (Figure 3.13) and are modelled in two steps. First, $Z(f)$ is implemented using another one-hidden-layer feedforward NN (Figure 5.6) with R neurons and a *tanh* activation function (5.10), where J_r, U_r and D_r are the r^{th} neuron's parameters and E is a constant [matrix notation is shown in (5.11)]. The input to this NN is frequency (f), and the output is multiplied by the derivatives of $X(V_1, V_2)$ as in (5.4)-(5.6).

$$Z(f) = \sum_{r=1}^{R} U_r tanh(J_r f + D_r) + E \tag{5.10}$$

$$\begin{aligned} J &= \begin{bmatrix} J_1 & J_2 & \ldots & J_R \end{bmatrix}^T \\ U &= \begin{bmatrix} U_1 & U_2 & \ldots & U_R \end{bmatrix} \\ D &= \begin{bmatrix} D_1 & D_2 & \ldots & D_R \end{bmatrix}^T \end{aligned} \tag{5.11}$$

Figure 5.7 shows the implementation of the multiplication process between the derivatives NN (Figure 5.4) and the dynamics NN (Figure 5.6), where W's are fixed wights [Appendix D (D.2)]. The inputs and targets are illustrated using the matrices in the black dotted rectangles for K biasing points and L frequency points (the red rectangles represent inputs and targets at corresponding biasing-frequency points).

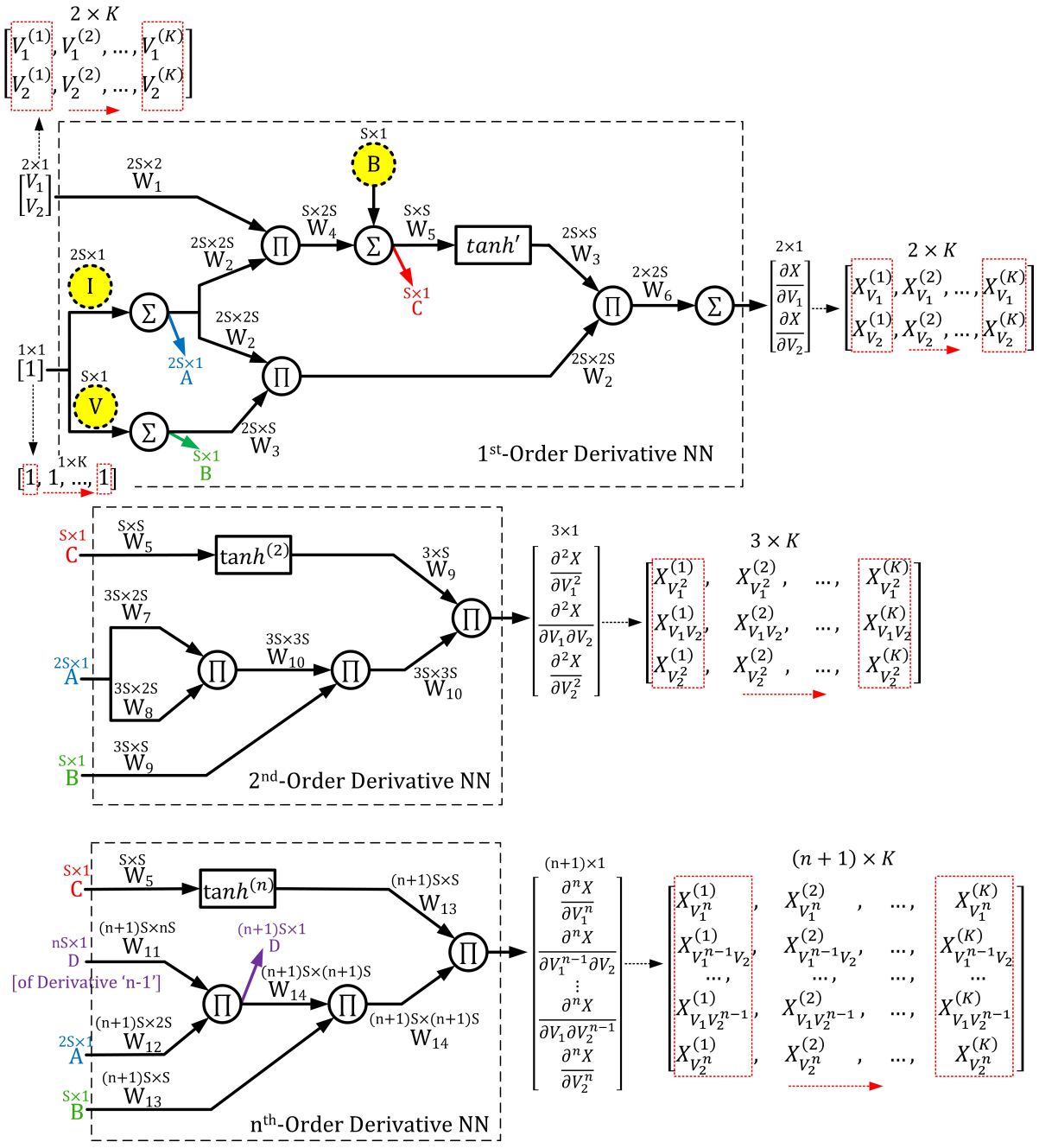


Figure 5.4: The proposed NN for the first, second, and up to the n^{th}-order derivatives (matrix notation is used) of $X(V_1, V_2)$. The red rectangles show inputs and targets at corresponding biasing points.

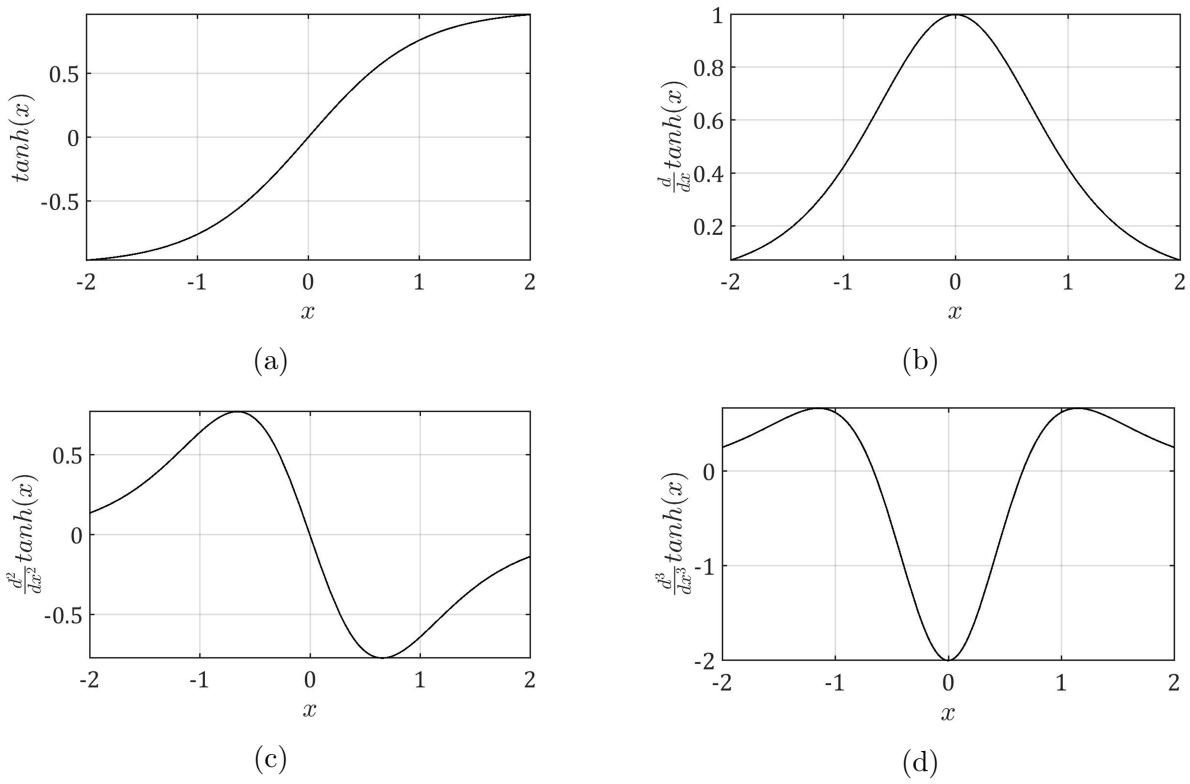


Figure 5.5: Activation functions used in (a) original, (b) first-order derivative, (c) second-order derivative, and (d) third-order derivative NNs.

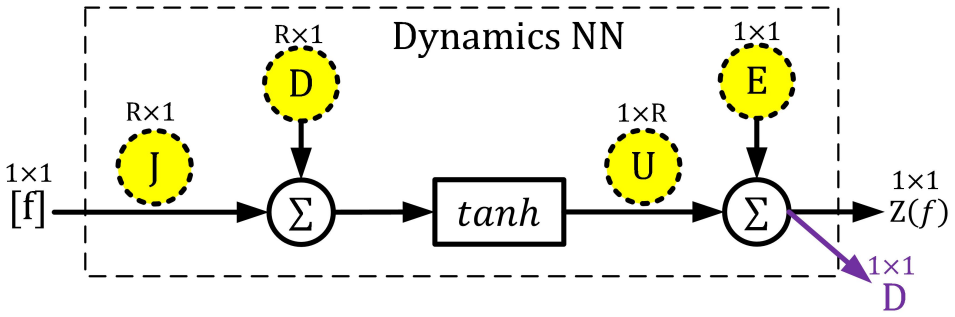


Figure 5.6: A one-hidden-layer feedforward NN with R hidden neurons that represents $Z(f)$. Matrix notation is used to describe J, U, D and E, where their sizes are shown above each circle.

Second, as shown in Chapter 3, HONPs do not have the same frequency dependency, and hence one more degree of freedom needs to be added. For example, first-order derivatives of $I_1(V_1, V_2, f)$ with respect to V_1 and V_2 result in $Y_{11}^{(1)}$ and $Y_{12}^{(1)}$, respectively. Since the behaviour of these two parameters (at a particular biasing point) versus frequency is not the same, (5.4) does not have enough degrees of freedom to model them. Therefore, another parameter M is introduced as in (5.12), and the final implementation is shown in Figure 5.8. W's are fixed wights [Appendix D (D.3)], and the inputs are same as in Figure 5.7. It is worth mentioning that increasing M by one adds a complete set of derivatives and dynamics NNs, i.e. a complete set of the parameters shown in yellow circles.

$$I_p(V_1, V_2, f) = \sum_{m=1}^{M} X_{p,m}(V_1, V_2) Z_{p,m}(f) \tag{5.12}$$
$$\text{where } p = 1, 2 \text{ is the port number.}$$

5.2.3 DC Solution

One of the most important aspects of modelling is solution consistency, i.e. the same result is expected when simulating the model under different simulators [56]. Since the model is built from bias-dependent HONPs, correct DC solution is crucial for correct model performance. The problem arises from the fact that I^{Cond} is constructed from RF measurements, and hence it is different from I^{Cond} at DC [I^{Disp} has no effect at DC (5.1)]. In [16, 57], the DC solution was implemented behaviourally by defining a frequency below which RF dispersion effects are neglected as described by (5.13), where I_{DC}^{Cond} and I_{RF}^{Cond} are DC and RF conduction currents, respectively, and $H(f)$ has low-pass characteristics. A circuit implementation of (5.13) can be realized using a choke coil to separate between I_{DC}^{Cond} and I_{RF}^{Cond} [17]. Another implementation of the DC solution was carried out in time-domain by describing I^{Cond} as DC current and adding RF dispersion effects as a small disturbance around it [58, 59].

$$\begin{aligned} I^{Total}(V_1, V_2, f) &= I_2^{Cond}(V_1, V_2, f) + I_2^{Disp}(V_1, V_2, f) \\ &= H(f) I_{DC}^{Cond}(V_1, V_2) + [1 - H(f)] I_{RF}^{Cond}(V_1, V_2) + j\pi f Q(V_1, V_2) \end{aligned} \tag{5.13}$$

In this work, DC solution is realized as shown in Figure 5.9, where an NN that represents (5.12) is added to the NN in Figure 5.8. This implementation forces the generated nonlinear function to have correct derivatives at RF (i.e. HOYPs) and also correct values at DC (i.e. DC solution). It should be noted that the final NN with the DC constraint has the same

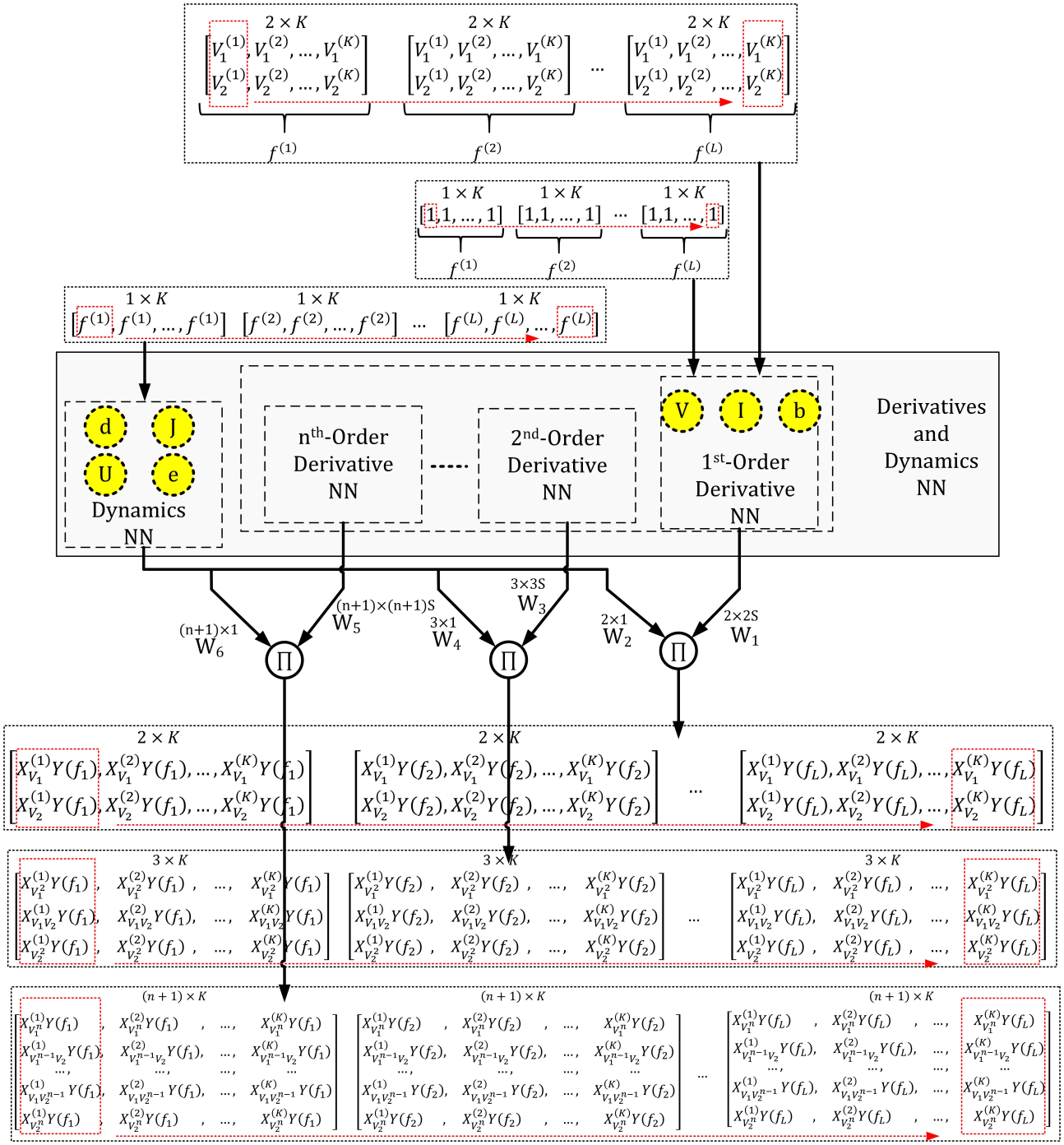


Figure 5.7: Implementation of the multiplication process between derivatives NN (Figure 5.4) and dynamics NN (Figure 5.6).

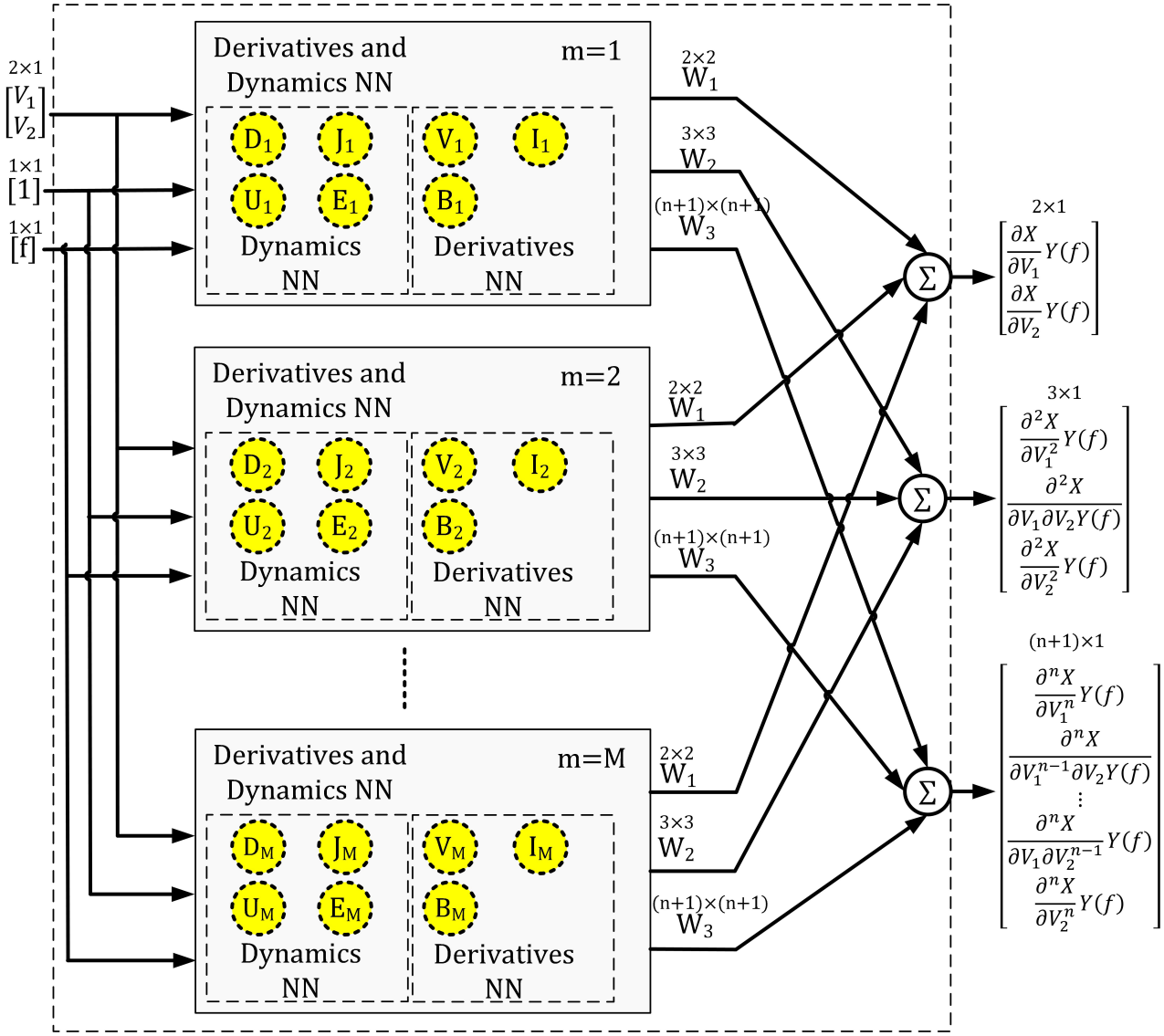


Figure 5.8: Final implementation of HONPs-based compact model (5.12).

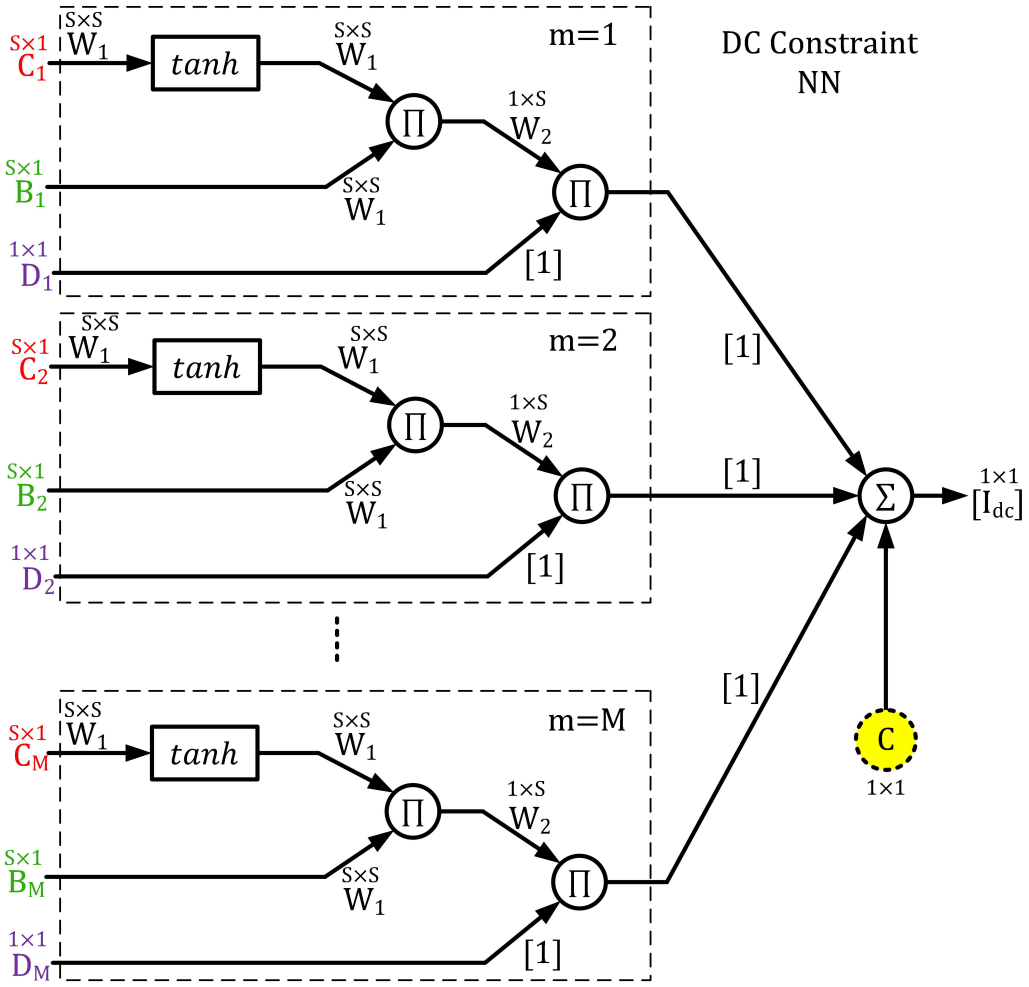


Figure 5.9: The proposed NN realization of DC the solution.

inputs in Figure 5.7, but with zeros added to the frequency vector to represent DC, and the corresponding target is I_{DC}^{Cond}.

5.2.4 Remarks

- The generated compact model has four current sources (5.1), each source is implemented as a separate NN (Figure 5.8) with different S, R, and M values. The generated functions are then implemented in Keysight ADS software using Symbol-

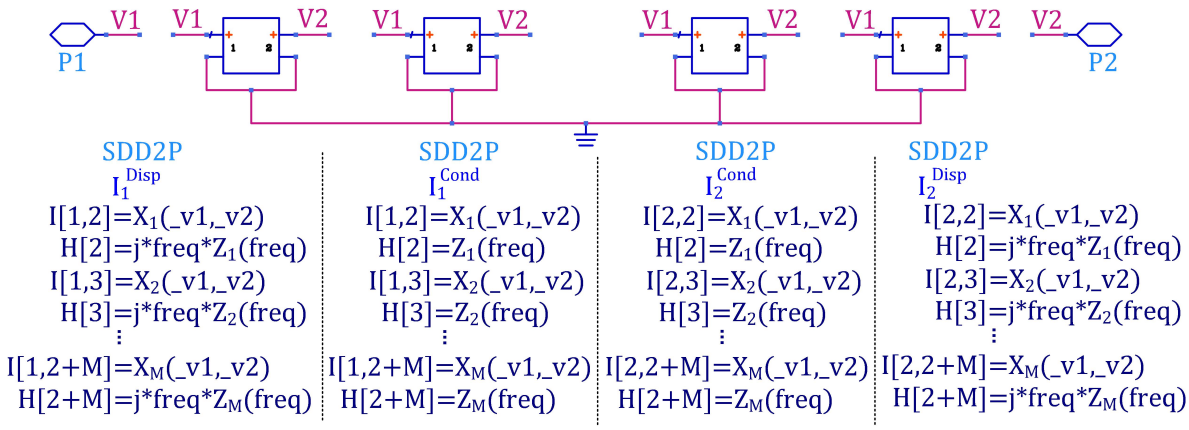


Figure 5.10: The compact model implementation in ADS.

ically Defined Device (SDD) components, as shown in Figure 5.10[2]. DC constraint NNs (Figure 5.9) are added to both I_1^{Cond} and I_2^{Cond} to force correct DC solutions. I_1^{Disp} and I_2^{Disp}, however, are trained as Q sources, i.e. their values are divided by ω. Therefore, no DC constraint NN needs to be added as their contribution at DC is guaranteed to be zero (5.1). It is worth mentioning that I^{Cond} and I^{Disp} are generated from the real and imaginary parts of HOYPs, respectively.

- Although the proposed dynamics NN (Figure 5.6) can model any frequency behaviour, using a large number of neurons (R) may cause convergence problems in large-signal simulators (Harmonic Balance and Envelope simulators). Thus, utilizing the layered-technique (proposed in Chapter 4) to optimize an extrinsic shell that puts less stress on the intrinsic shell implementation is recommended. For example, optimizing an extrinsic shell to eliminate resonance behaviours from intrinsic HOYPs results in smaller R needed for the dynamics NN.

- The proposed compact model implementation can be realized in time-domain by utilizing the generalized quasi-static assumption mention in Section 3.3.1. This implementation will maintain the causality of the model. However, it is not guaranteed to find an extrinsic shell that forces all HOYPs dynamics [$Z(f)$'s] to have $(\jmath\omega)^n$ frequency dependency.

[2]The generated expressions for both X and Z function are too complex to be implemented manually in ADS. Therefore, this schematic is generated by MATLAB as a Netlist file and then loaded into ADS circuit simulator.

	# of layers	Elements that will be trained
1^{st}-order derivative	9	B , I , V
each n^{th}-order derivative where $n > 1$	5	None
Dynamics NN	2	J , D , U , E
DC constraint NN	4	C

Table 5.1: The number of layers added and elements that will be trained by each NN in the proposed technique (for $M = 1$). The size of each element is indicated in the corresponding figure for each NN.

- The DC solution can be realized without the DC constraint NN by adding zeros to the input frequency vector in Figure 5.7 and adding DC Y-parameters to the corresponding targets. C can then be adjusted to shift the resulting output to meet the actual solution. However, having the DC solution as a separate target results in better training for the model.

- Levenberg-Marquardt back-propagation algorithm is used for NNs training. The objective is to minimize an MSE cost function that includes all derivatives (and DC solutions in the case of I_1^{Cond} and I_2^{Cond}) as targets. Since HOYPs were proven to be continuous over V_1, V_2, and frequency (Chapter 3), they are interpolated over a finer biasing-frequency grid. Therefore, more data is available to train the NN which improves the accuracy of the generated model. It is worth mentioning that low-order models require interpolating HOYPs over a finer biasing grid than high-order models.

- The generated model can be interpolated by circuit simulators -within its modelling space- and is expected to correctly predict up to the n^{th}-order behaviours. However, it does not have any extrapolation capability.

- Table 5.1 shows the number of layers added and elements that will be trained by each NN in the proposed technique (for $M = 1$). High-order models take more time to be trained. This is because the model's complexity, represented by the number of layers and targets, increases with its order and also because of the nature of the activation functions used in the implementation (Figure 5.5).

- Table 5.2 summarizes all degrees of freedom in the proposed NN technique.

	Role
S	• The number of neurons in the derivatives NN. • Affected by the number of derivatives and the number of biasing points.
R	• The number of neurons in the dynamics NN. • Affected by the frequency behaviour of a particular parameter.
M	• The number of derivatives and dynamics NNs used. • Affected by the difference in frequency behaviour between different parameters.

Table 5.2: A summary of all degrees of freedom in the proposed NN technique.

5.3 Validation

5.3.1 Theoretical Validations

As a proof concept, a Cree CGH40010P 10 Watt GaN HEMT was used in Keysight ADS simulation environment as a DUT. V_{gs} was swept from -4 to 0 V with a step of 0.5 V, and V_{ds} was swept from 0 to 56 V with 23 points. First, second, and third-HOYPs were extracted using CW stimulus, as illustrated in Chapter 3, from 10 MHz to 3 GHz (15 points were used to sweep the specified frequency range). The following three compact models were constructed using the proposed NN technique[3]:

- 1^{st}-Order Model: built from first-HOYPs (linear Y-parameters).
- 2^{nd}-Order Model: built from first and second-HOYPs.
- 3^{rd}-Order Model: built from first, second, and third-HOYPs.

For the three models, S, R, and M values -for the four current sources in Figure 5.10- were optimized to give a good compromise between training time and accuracy (Table 5.3). It is important to note that "N^{th}-order models are capable of predicting up to N^{th}-HONPs". For illustration purpose, the resulting expression for I_2^{Cond} in each model was derived up to the third-order derivatives, resulting in the real parts of first, second and third-HOYPs. These derivatives were then compared to the original HOYPs used for training, and the resulting MSE is reported in Table 5.4. It is obvious that 3^{rd}-Order Model was the only

[3]Since this is the same DUT used in Section 4.3, the extrinsic circuit in Table 4.2 was de-embedded. Therefore, the proposed NN technique was used to model the intrinsic part of the DUT.

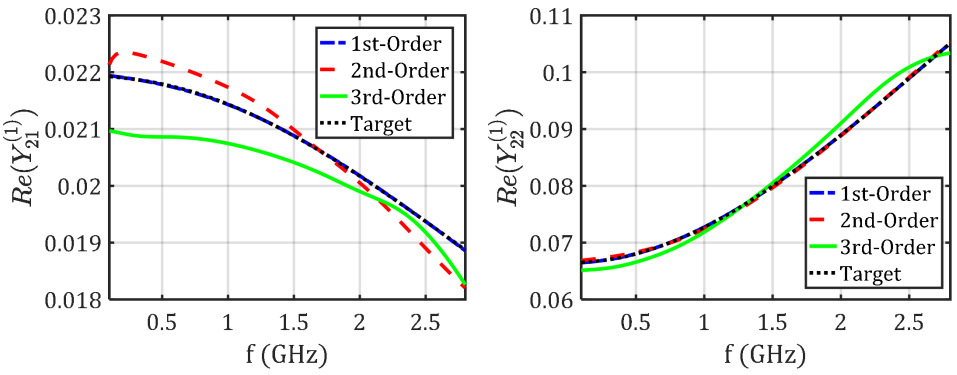
	1^{st}-Order Model	2^{nd}-Order Model	3^{rd}-Order Model
I_1^{Cond}	(15 , 2 , 2)	(15 , 2 , 3)	(20 , 2 , 6)
I_1^{Disp}	(15 , 2 , 2)	(15 , 2 , 3)	(20 , 2 , 6)
I_2^{Cond}	(15 , 2 , 3)	(15 , 2 , 4)	(20 , 2 , 6)
I_2^{Disp}	(15 , 2 , 2)	(15 , 2 , 3)	(20 , 2 , 6)

Table 5.3: The values of (S , R , M) for each current source in the three models.

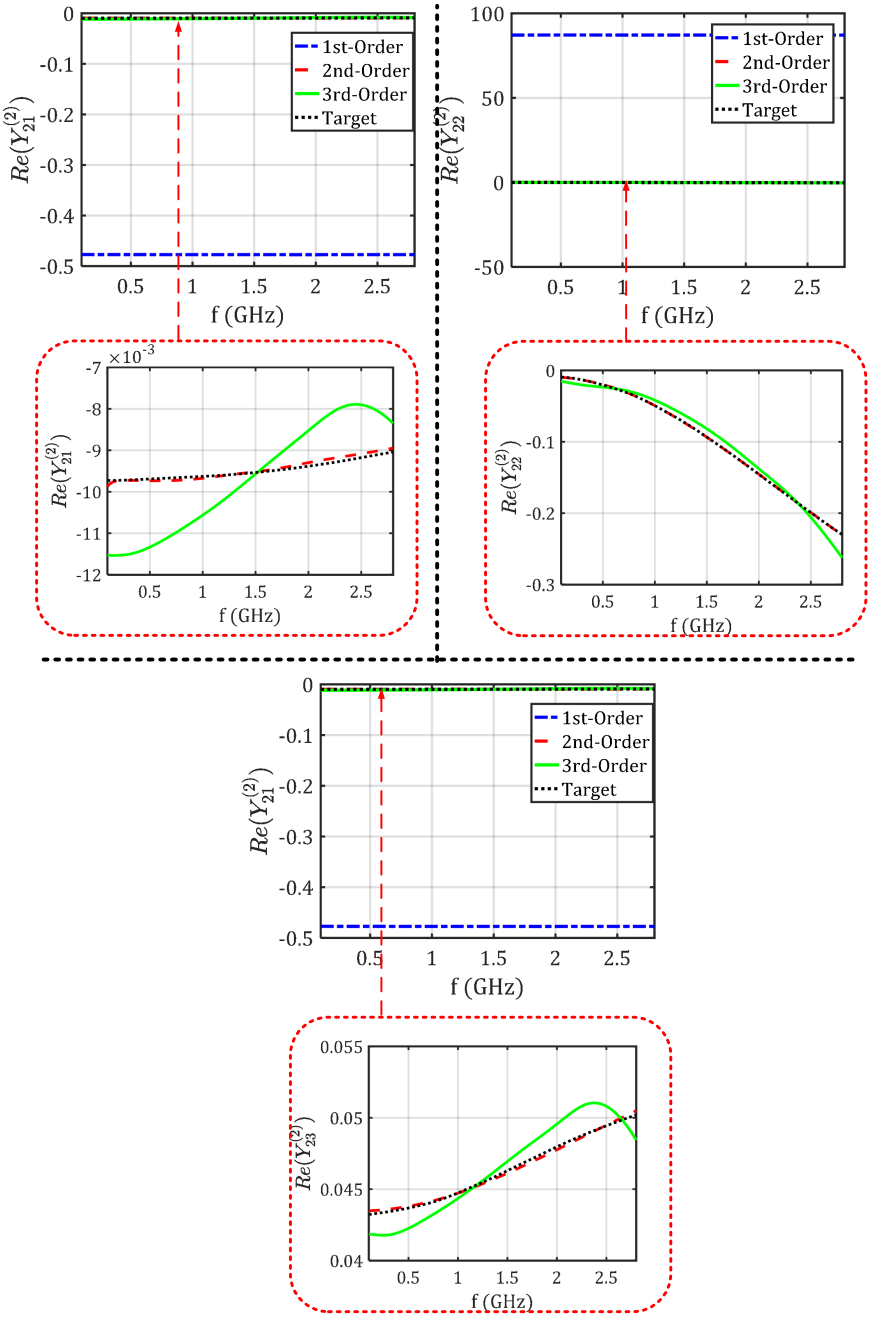
model that could predict all HOYPs (with an acceptable error). Figure 5.11 (a), (b), and (c) show these derivatives versus the original target used in training at (V_{gs},V_{ds}) of $(-2.5, 5)$ V. It is worth mentioning that false prediction of I/V derivatives (i.e. HOYPs) results in false prediction of the corresponding high-order behaviours, as will be illustrated in the next validations.

MSE	1^{st}-Order Model	2^{nd}-Order Model	3^{rd}-Order Model
$Re[Y_{2,1}^{(1)}]$, $Re[Y_{2,2}^{(1)}]$	2.7e-8	2.6e-7	4.1e-6
$Re[Y_{2,1}^{(2)}]$, $Re[Y_{2,2}^{(2)}]$, $Re[Y_{2,3}^{(2)}]$	910	1.6e-7	3.6e-6
$Re[Y_{2,1}^{(3)}]$, $Re[Y_{2,2}^{(3)}]$, $Re[Y_{2,3}^{(3)}]$, $Re[Y_{2,4}^{(3)}]$	1.1e5	4.3e6	8.1e-6

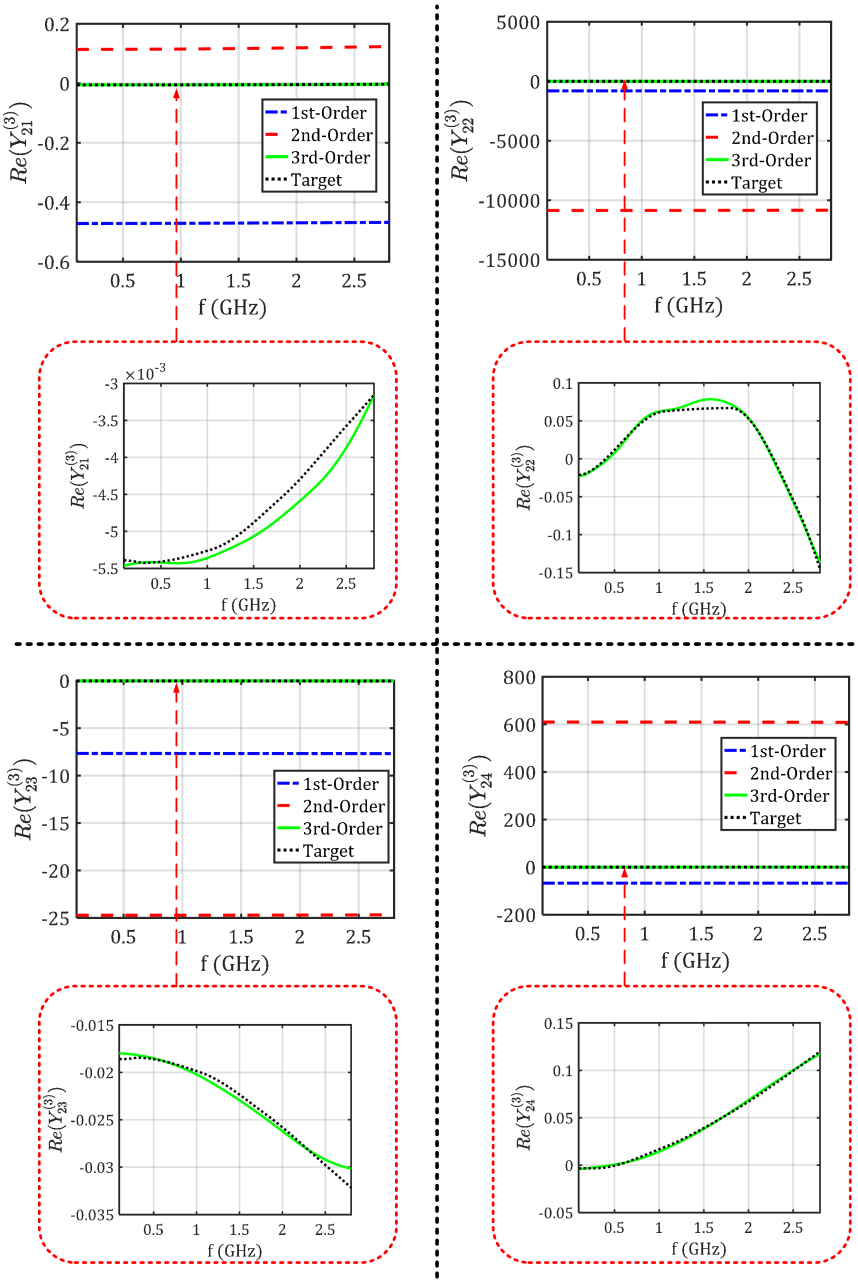
Table 5.4: An example of HOYPs prediction error by different compact model orders.



(a)


1st-Order
2nd-Order
3rd-Order
Target
$Re(Y_{21}^{(2)})$
$Re(Y_{22}^{(2)})$
$Re(Y_{23}^{(2)})$
f (GHz)
$\times 10^{-3}$

(b)



(c)

Figure 5.11: Predictions of (a) first-HOYPs, (b) second-HOYPs, and (c) third-HOPYs [versus frequency at (V_{gs}, V_{ds}) of $(-2.5, 5)$ V] using different compact model orders. The dotted red rectangles represent zoomed-in versions of the corresponding plots.

5.3.2 Simulation Validations

The 3^{rd}-Order Model extracted in the previous subsection was used under different circuit simulators in ADS software as follows. Figure 5.12 (a) shows simulated DC IV curves for the generated model and the original DUT. The excellent match between both curves indicates that the DC constraint NN (Figure 5.9) was able to force I_2^{Cond} to have a correct DC solution. Small-signal simulations were then performed on the generated model. Figure 5.12 (b)-(e) shows simulated S-parameters for 3^{rd}-Order Model and the DUT at four DC biasing points. The ability of the model to correctly predict S-parameters is due to the fact it was able to correctly predict first-order I/V derivatives.

In order to show the ability of the proposed model to predict strong nonlinear behaviours, a class-AB PA was designed at 0.9 GHz fundamental frequency using Cree model. The transistor was biased at (V_{gs},V_{ds}) of $(-2.75, 28)$ V and the simulated output power and power added efficacy (PAE) were 39 dBm and 67.4 %, respectively. Both 1^{st}-Order Model and 3^{rd}-Order Model were simulated under exactly the same PA circuit conditions. Figure 5.13 (a)-(d) shows the simulated output power under CW stimulus at the fundamental, second, third, and fifth-harmonic tones. The following should be noted:

- 3^{rd}-Order Model showed better performance in the fundamental band because both 1^{st} and 3^{rd}-Order Models were trained using the same HOYPs data. The vertical shift in the performance of the 1^{st}-Order Model is due to a training issue, which can be solved by interpolating first-HOYPs over a finer biasing grid before training.
- 1^{st}-Order Model failed to predict the behaviour at the second and third-harmonic tones because it could not predict second and third-HOYPs (Table 5.4).
- As expected, both models failed to predict the behaviour at the fifth-harmonic tone. However, 3^{rd}-Order Model performance is better which enhances the simulation convergence.

Finally, a 20 MHz 4C-WCDMA signal was used to stimulate the PA. The signal's PAPR was 7.14 dB, and the simulated PA peak power was 38.6 dBm. The simulated output spectrum at the fundamental band is shown in Fig 5.13 (e). While the 1^{st}-Order Model failed to predict the spectrum, the 3^{rd}-Order Model succeeded at predicting it almost to the end of the third-IMD (IMD3) band. The (normalized mean-square-error) NMSE of the output IQ signal (compared to Cree model output) is -15.3 and -30.6 dB for 1^{st}- and 3^{rd}-Order Models, respectively.

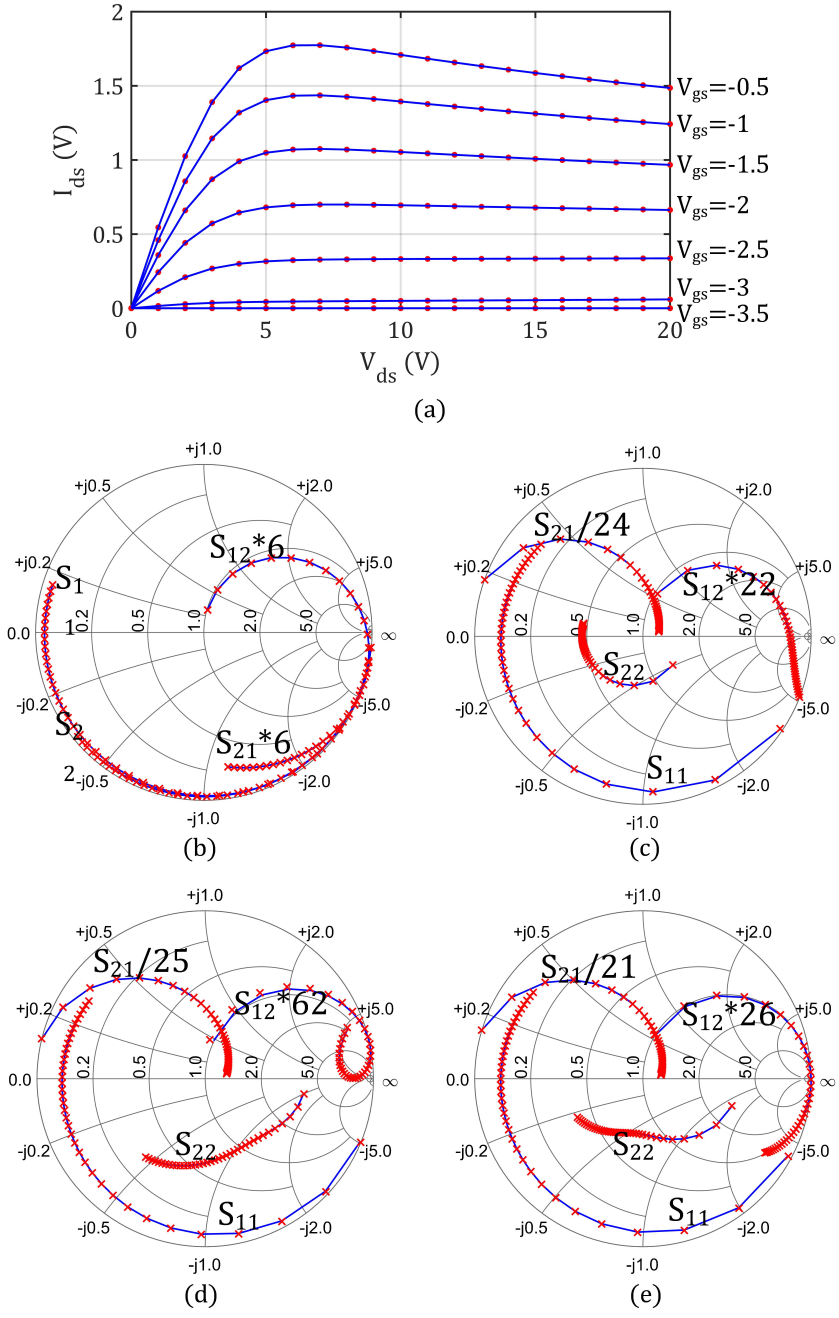


Figure 5.12: Original [red (x)] and simulated [blue (-)]:(a) DC IV curves, and S-parameters at (b) $(V_{gs}, V_{ds}) = (-4, 1)$ V, (c) $(V_{gs}, V_{ds}) = (-2.5, 5)$ V, (d) $(V_{gs}, V_{ds}) = (-1.5, 50)$ V, and (e) $(V_{gs}, V_{ds}) = (0, 20)$ V.

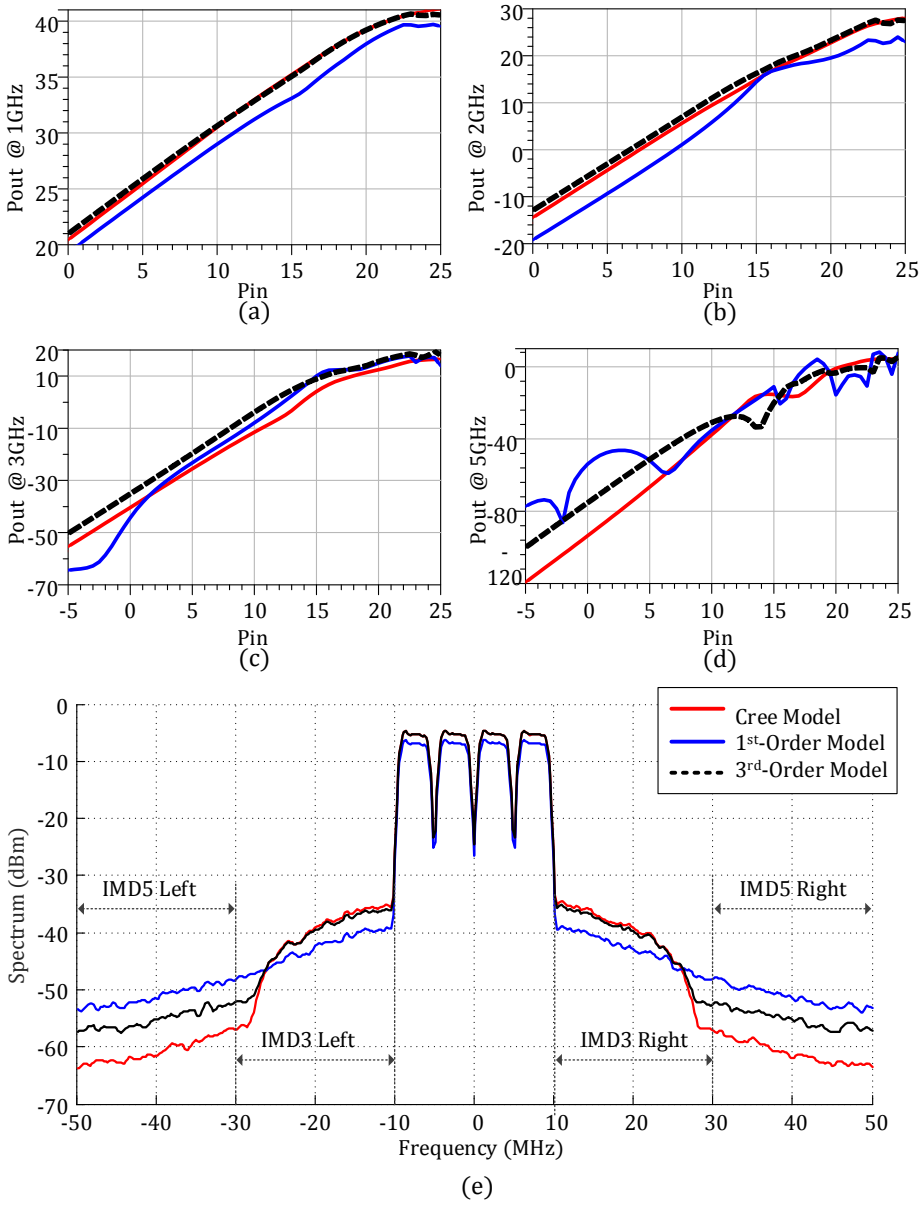


Figure 5.13: The simulated output power for Cree model and both 1^{st} and 3^{rd}-Order Models under CW stimulus at (a) fundamental, (b) second, (c) third, and (d) fifth-harmonic frequencies. (e) The simulated output spectrum of the fundamental band under 4C-WCDMA signal.

5.3.3 Measurement Validations

First, second, and third-HOYPs were measured for Cree CGH40006P 6 Watt GaN HEMT[4] using the setup in Figure 3.19. V_{gs}, V_{ds}, and frequency were swept from -4 to 0 V, from 0 to 50 V, and from 10 MHz to 3 GHz, respectively. Both 1^{st} and 3^{rd}-Order Models were generated and used in Keysight ADS software to simulate a fabricated class-AB PA, biased at V_{gs} of 25 V and I_{ds} of 100 mA [60][5]. A 20 MHz 2-tone signal centred around 850 MHz was used to stimulate the PA. The following was noticed:

- 3^{rd}-Order Model demonstrated better capability in predicting nonlinear behaviours such as gain compression and IMD3 components within the fundamental band, Figure 5.14 (a) and (b), respectively.
- As mentioned in Section 3.5.4, it is practically impossible to measure second-HONPs below 100 MHz due to the PNA-X internal couplers roll-off. The false identification of second-HONPs at 20 MHz reflects directly in the model ability to predict IMD3. This problem can be solved by adding a baseband extension to the PNA-X.
- The limited dynamic range of the measurement setup makes it difficult to identify HONPs in the saturation region of the transistor (because it is less nonlinear than the triode region). This means that the contribution of HONPs in this region to strong nonlinear behaviours can be neglected. However, the proposed NN needs data in this region to be trained. This problem affects the accuracy of the generated model and will be tackled in the future work[6].

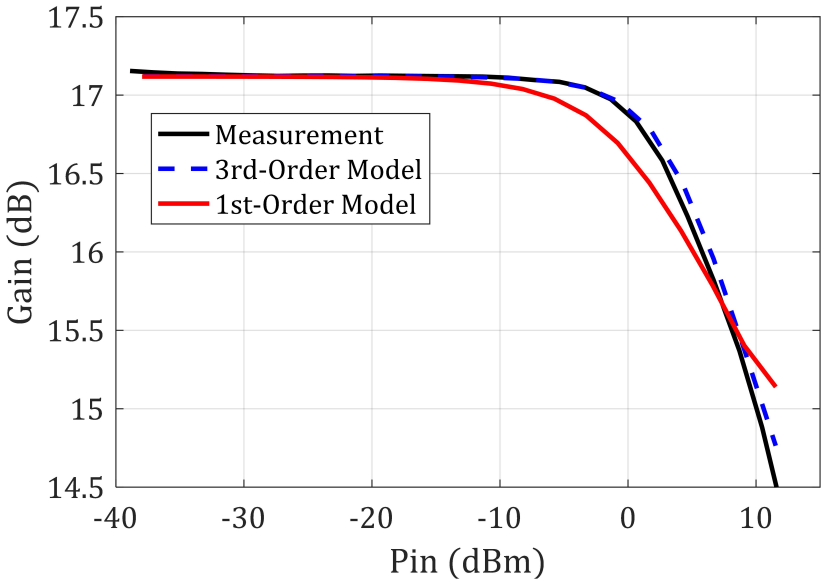
5.4 Comparison

Table 5.5 shows a comparison between the proposed HONPs-based compact model and state-of-the art large-signal models in the literature.

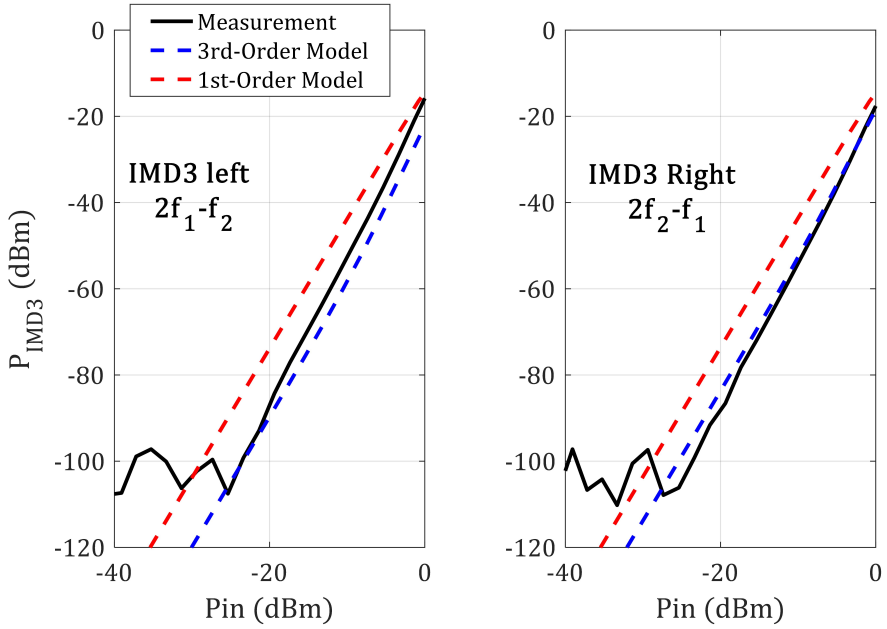
[4]These are the same measurement shown in Section 3.5 and will be used here to generate the proposed compact model.

[5]This PA was designed by one of the EmRG [7] members, so the design schematic is accessible.

[6]A simple solution is to decrease the receivers IFBW, yet the measurement time will significantly increase (especially in the case of generating compact models were long sweeps are required to cover the entire modelling space).



(a)



(b)

Figure 5.14: Measured PA (a) gain at 0.85 GHz, and (b) left and right IMD3 components, both versus input power.

Table 5.5: A comparison between the proposed HONPs-based compact model and state-of-the art LSMs

	Linear measurement-based models	DynaFET model	The proposed model
Theory	• First-order Taylor Expansion	• Load-line concept	• Volterra-based model
Required Data	• Bias-dependent S-parameters	• Voltages and currents time-domain waveforms	• Bias-dependent HONPs
Measurement Setup	• VNA measurements	• NVNA measurements • Source/load-pull setup • High-power capabilities	• NVNA measurements
Strengths	• It can be tuned using the load-line concept • Thermal and trapping effects can be included • Systematic extraction	• It inherently provides large-signal validations • It allows modelling breakdown region • Thermal and trapping effects are included	• Systematic extraction • It measures most of the DUT nonlinearities and dynamics • It works under modulated signals
Drawbacks	• Poor predictive capabilities for nonlinear behaviours	• Nonsystematic extraction • It does not stimulate all of the DUT dynamics	• Sensitive to the measurement noise • Affected by the limited dynamic range

5.5 Conclusion

An HONPs-based compact model that can predict strong nonlinear behaviours under modulated signals was proposed. It handles the transistor behaviour at any biasing point as a WNL circuit described by HONPs. The model extraction is carried out using low-power NVNA-based CW measurement data. The proposed model demonstrated an excellent ability to predict the transistor's large-signal behaviour under modulated signal stimuli in the simulation environment. It also showed a good capacity to predict the measured AM/AM characteristics and IMD3 of the corresponding class AB power amplifier under multi-tone stimuli.

Chapter 6

Conclusion

The realization of 5G front-ends will require accurate simulations for both system and circuit levels. While first is needed because the high-level of integration prevents any opportunity for post-fabrication tweaking, the second is essential to preserve the overall efficiency of the system. Unfortunately, existing transistor models cannot be used for these simulations as they cannot predict the device behaviour under real-life operating conditions (i.e., under modulated signal stimulus). The main aim of this thesis was to propose and develop a transistor compact model that can predict the behaviour under modulated signal stimuli.

As a first step, HONPs -a new set of Volterra-based parameters- were proposed. These parameters were proven, in both simulation and measurement environments, to be power independent, continuous across the modelling space and unique. HONPs formulation allows using them to cover the required modelling space without having an explosion in the number of parameters. HONPs were extracted systematically using low power CW NVNA measurements and were capable of predicting the device WNL behaviour under wideband multi-tone stimuli.

Next, a novel compact model was developed in two steps. First, a new small-signal modelling technique was proposed using layered-NN. This technique allows optimizing both intrinsic and extrinsic topologies in order to fit the device small-signal performance. Second, a new large-signal model that utilizes HONPs as a building block was proposed. The implementation of this model was performed by introducing a new NN technique that allows constructing nonlinear functions from their derivatives. The proposed model is both globally and locally accurate and was validated under DC, small-signal and Harmonic Balance simulators. Its ability to predict strong nonlinear behaviour was demonstrated using

a CW harmonic Balance simulator. A modulated signal validation was also performed in the simulation environment using a 20 MHz 4C-WCDMA signal under envelope simulator. Also, a two-tone validation was performed in the measurement environment using a fabricated class-AB PA.

6.1 Summary of Contributions

The objective of this thesis was to develop a novel transistor compact model that can be used in circuit simulators under modulated signal stimuli. This goal was achieved through the following key contributions:

1. A novel set of parameters that allows describing the nonlinear system WNL behaviours were presented. The formulation of these parameters allows the solution generality that is necessary for transistor compact modelling. Also, they are systematically extracted using low power CW NVNA measurements.

2. A small-signal modelling technique was proposed that allows optimizing intrinsic and extrinsic topologies without the need for driving complex analytical expressions or going back and forth between ColdFET and HotFET.

3. A novel HONPs-based large-signal model for the intrinsic transistor was proposed that allows building nonlinear functions from their derivatives (i.e., from HONPs). The generated model can accurately calculate current/voltage derivatives, and hence predicting the behaviour under modulated signal stimuli.

6.2 Future Work

In this thesis, we have presented a new set of parameters that simplifies characterizing and analyzing nonlinear systems, and as an application, a novel transistor compact model was proposed. This work can be extended into several lines of research.

The focus of this work was on HONPs presented in Chapter 3. A baseband extension can be added to the measurement setup to allow accurate extraction of the second-HONPs, and hence accurate prediction of the third-IMD products in the fundamental band. It would be interesting to investigate the ability of these parameters to linearize nonlinear systems under ultra-wideband stimuli using mixer-based receiver characterization. This

is of great importance to the predistortion field, especially after moving to 5G and being limited by the receiver's bandwidth [1]. Also, HONPs ability to predict the stability of nonlinear systems under large-signal operating conditions can be studied.

Another interesting topic for future research would be studying the effect of the extrinsic shell on the frequency behaviour of HONPs. In Chapter 4, a layered-technique was proposed for optimizing the model's equivalent circuit shells (intrinsic and extrinsic) to fit linear Y-parameters. Including high-order parameters in this technique would be necessary for improving the HONPs-based compact model proposed in Chapter 5. Also, the proposed NN technique that builds nonlinear functions from the derivatives should be optimized to handle noisy and missing measurements. This, along with the baseband extension of the measurement setup, shall allow testing the model under modulated signal stimuli in the measurement environment. Lastly, it would be worthwhile to investigate adding thermal and trapping effects to this model, which would include studying HONPs under different stimuli, such as pulsed and dynamic-biased stimuli.

[1] Simulation validations is provided in Appendix B.

Appendix A

DIDO Volterra and HONPs

A.1 System Description

This appendix illustrates the relationship between HOYPs (proposed in Chapter 3) and Volterra kernels. Only first and second-order kernels will be derived as they are sufficient to convey the concept without dealing with complicated mathematical expressions. Figure A.1 shows HOYPs equivalent circuit, which is described by (A.1)[1]. Applying the generalized quasi-static assumption for the elements dynamics result in (A.2), where operator notation is used (3.6). HOYPs are assumed to be known [(3.15)-(3.17)] as the objective here is to solve for Volterra kernels.

It is worth mentioning that there is a difference between extracting Volterra kernels and applying them to find the WNL response of the system. For example, while Volterra kernels have to be extracted using non-commensurate signals, they can be applied when exciting the system with commensurate signals. Also, when extracting DIDO Volterra kernels, the signal at both ports will be assumed to have components at the same frequency grid. However, once these kernels are extracted, they can be applied to find the WNL solution even if both ports are excited with different signals[2]. This can be easily achieved using the Method of Nonlinear Currents, as the superposition principle can be applied.

[1] In order to simplify the equations readability, the definitions of all the symbols used in this appendix equations are listed in Section A.4.

[2] The same applies to HONPs extraction and application as these parameters are based on the Volterra Series.

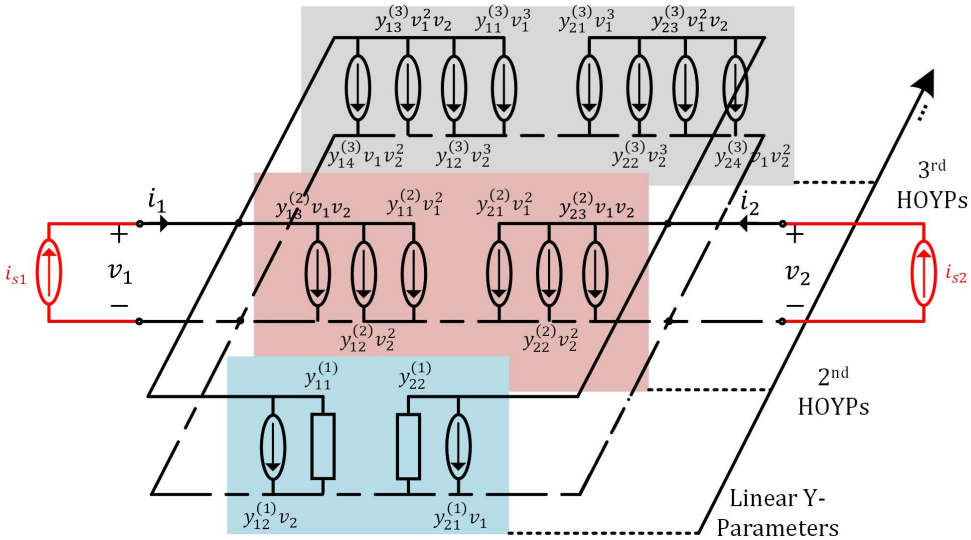


Figure A.1: First (blue), second (red), and third (grey) HOYPs equivalent circuit. Time dependency is dropped for ease of notation. i_{s1} and i_{s2} represent the input sources used to extract Volterra kernels.

$$\begin{bmatrix} i_1(t) \\ i_2(t) \end{bmatrix} = \begin{bmatrix} y_{11}^{(1)}(t) & y_{12}^{(1)}(t) \\ y_{21}^{(1)}(t) & y_{22}^{(1)}(t) \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} + \begin{bmatrix} y_{11}^{(2)}(t) & y_{12}^{(2)}(t) & y_{13}^{(2)}(t) \\ y_{21}^{(2)}(t) & y_{22}^{(2)}(t) & y_{23}^{(2)}(t) \end{bmatrix} \begin{bmatrix} v_1^2(t) \\ v_2^2(t) \\ v_1(t)v_2(t) \end{bmatrix}$$
$$+ \begin{bmatrix} y_{11}^{(3)}(t) & y_{12}^{(3)}(t) & y_{13}^{(3)}(t) & y_{14}^{(3)}(t) \\ y_{21}^{(3)}(t) & y_{22}^{(3)}(t) & y_{23}^{(3)}(t) & y_{24}^{(3)}(t) \end{bmatrix} \begin{bmatrix} v_1^3(t) \\ v_2^3(t) \\ v_1^2(t)v_2(t) \\ v_1(t)v_2^2(t) \end{bmatrix} \tag{A.1}$$

$$\begin{bmatrix} i_1(t) \\ i_2(t) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(P) & Y_{12}^{(1)}(P) \\ Y_{21}^{(1)}(P) & Y_{22}^{(1)}(P) \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} + \begin{bmatrix} Y_{11}^{(2)}(P) & Y_{12}^{(2)}(P) & Y_{13}^{(2)}(P) \\ Y_{21}^{(2)}(P) & Y_{22}^{(2)}(P) & Y_{23}^{(2)}(P) \end{bmatrix} \begin{bmatrix} v_1^2(t) \\ v_2^2(t) \\ v_1(t)v_2(t) \end{bmatrix}$$

$$+\begin{bmatrix} Y_{11}^{(3)}(P) & Y_{12}^{(3)}(P) & Y_{13}^{(3)}(P) & Y_{14}^{(3)}(P) \\ Y_{21}^{(3)}(P) & Y_{22}^{(3)}(P) & Y_{23}^{(3)}(P) & Y_{24}^{(3)}(P) \end{bmatrix} \begin{bmatrix} v_1^3(t) \\ v_2^3(t) \\ v_1^2(t)v_2(t) \\ v_1(t)v_2^2(t) \end{bmatrix} \tag{A.2}$$

A.2 DIDO Volterra Formulations

Time-domain SISO Volterra formulation and the transformation between time-domain kernels (nonlinear impulse responses) and frequency-domain kernels (nonlinear transfer functions) were shown in (2.3)-(2.5). Similarly, one could start with the time-domain DIDO Volterra formulation, with time-domain kernels in (A.3) [61], and use the multi-tone formulation in (A.4) to reach the formulation with frequency-domain kernels in (A.5). The semicolon in these expressions separates between port 1 and port 2 parameters. In order to find Volterra kernels for the WNL system described by (A.2), the Harmonic Input Method will be used [30, 62].

$$v_p(t) = \sum_{\substack{n_1=0 \\ n_2=0 \\ (n_1,n_2)\neq(0,0)}}^{\substack{n_1=N_1 \\ n_2=N_2}} v_{p,n_1n_2}(t) : \tag{A.3}$$

$$\begin{aligned} v_{p,n_1n_2}(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} & h_{p,n_1n_2}(\tau_{11}, \dots, \tau_{1n_1}; \tau_{21}, \dots, \tau_{2n_2}) \\ & .i_1(t-\tau_{11}) \dots i_1(t-\tau_{1n_1}).i_2(t-\tau_{21}) \dots i_2(t-\tau_{2n_2}) \\ & .d\tau_{11} \dots d\tau_{1n_1}.d\tau_{21} \dots d\tau_{2n_2} \end{aligned}$$

$$i_p(t) = \frac{1}{2} \sum_{\substack{q=-Q \\ q\neq 0}}^{Q} I_{p,q} e^{j\omega_q t} \tag{A.4}$$

$$\begin{aligned} v_{p,n_1n_2}(t) = \frac{1}{2^{n_1+n_2}} \sum_{q_1=-Q}^{Q} \dots \sum_{q_{n_1}=-Q}^{Q} \sum_{r_1=-Q}^{Q} \dots \sum_{r_{n_2}=-Q}^{Q} & I_{1,q_1} \dots I_{1,q_{n_1}}.I_{2,r_1} \dots I_{2,r_{n_2}} \\ .H_{p,n_1n_2}(\omega_{q_1} \dots \omega_{q_{n_1}}; \omega_{r_1} \dots \omega_{r_{n_2}}) & e^{j(\omega_{q_1}+\dots+\omega_{q_{n_1}}+\omega_{r_1}+\dots+\omega_{r_{n_2}})t} \end{aligned} \tag{A.5}$$

It is worth mentioning that assuming both ports are excited with signals that have components at the same frequency grid (Section A.1) simplifies the process of extracting Volterra kernels. This is because Volterra kernels will be described by $H_{p,n_1n_2}(\omega_{q_1}\dots\omega_{q_n})$ instead of $H_{p,n_1n_2}(\omega_{q_1}\dots\omega_{q_{n_1}};\omega_{r_1}\dots\omega_{r_{n_2}})$, where both of them are defined in Section A.4. In the following subsections, both first and second-order Volterra kernels will be derived using the mentioned assumption. However, the general solution to Volterra kernels (when both ports are excited with signals that have components at different frequency grid) will be mentioned in Section A.2.3.

A.2.1 First-Order DIDO Volterra Kernels $[H_1(\omega_1)]$

First-order frequency-domain kernels (first-order transfer functions) are found as follows. First, the inputs (i_{s1} and i_{s2}) in Figure A.1 are assumed to be 1-tone unite amplitude complex signals (A.6). These inputs are substituted in (A.5) to find the overall voltage response at port p described by (A.3), where $N_1 = 1$ and $N_2 = 1$. Second, the resulting voltages (A.7) are inserted in the original system of integro-differential equations (A.2). Since the application of the integro-differential operators in (A.2) to the exponential terms in (A.5) results in $(j\omega)^m$, (A.2) can be rewritten as (A.8). Finally, first-order kernels can be found by equating the first-order terms (terms with $e^{j\omega_1 t}$) in (A.8), because sinusoidal functions are orthogonal with respect to time. The resulting kernels are expressed in the matrix form by (A.9).

$$\begin{bmatrix} i_{s1}(t) \\ i_{s2}(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{j\omega_1 t} \tag{A.6}$$

$$\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} = \begin{bmatrix} H_{1,10}(\omega_1) & H_{1,01}(\omega_1) \\ H_{2,10}(\omega_1) & H_{2,01}(\omega_1) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{j\omega_1 t} \tag{A.7}$$

$$\begin{aligned} \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{j\omega_1 t} = & \begin{bmatrix} Y_{11}^{(1)}(\omega_1) & Y_{12}^{(1)}(\omega_1) \\ Y_{21}^{(1)}(\omega_1) & Y_{22}^{(1)}(\omega_1) \end{bmatrix} \begin{bmatrix} H_{1,10}(\omega_1) & H_{1,01}(\omega_1) \\ H_{2,10}(\omega_1) & H_{2,01}(\omega_1) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{j\omega_1 t} \\ & + \begin{bmatrix} Y_{11}^{(2)}(2\omega_1) & Y_{12}^{(2)}(2\omega_1) & Y_{13}^{(2)}(2\omega_1) \\ Y_{21}^{(2)}(2\omega_1) & Y_{22}^{(2)}(2\omega_1) & Y_{23}^{(2)}(2\omega_1) \end{bmatrix} \end{aligned} \tag{A.8}$$

$$\cdot \begin{bmatrix} H_{1,10}(\omega_1)^2 & H_{1,01}(\omega_1)^2 & 2H_{1,01}(\omega_1)H_{1,10}(\omega_1) \\ H_{2,10}(\omega_1)^2 & H_{2,01}(\omega_1)^2 & 2H_{2,01}(\omega_1)H_{2,10}(\omega_1) \\ H_{1,10}(\omega_1)H_{2,10}(\omega_1) & H_{1,01}(\omega_1)H_{2,01}(\omega_1) & H_{1,01}(\omega_1)H_{2,10}(\omega_1) + H_{1,10}(\omega_1)H_{2,01}(\omega_1) \end{bmatrix}$$

$$\cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} e^{j2\omega_1 t} + \dots$$

$$\big[H_1(\omega_1)\big] = \begin{bmatrix} H_{1,10}(\omega_1) & H_{1,01}(\omega_1) \\ H_{2,10}(\omega_1) & H_{2,01}(\omega_1) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(\omega_1) & Y_{12}^{(1)}(\omega_1) \\ Y_{21}^{(1)}(\omega_1) & Y_{22}^{(1)}(\omega_1) \end{bmatrix}^{-1} \tag{A.9}$$

A.2.2 Second-Order DIDO Volterra Kernels $[H_2(\omega_1, \omega_2)]$

The extraction procedure of second-order frequency-domain kernels (second-order transfer functions) is similar to the procedure described in the previous subsection. The inputs (i_{s1} and i_{s2}) in Figure A.1 are assumed to be 2-tone unite amplitude complex signals (A.10). Repeating the same procedure in (A.7)-(A.9), and then equating the second-order terms (terms with $e^{j(\omega_1+\omega_2)t}$) yields (A.11).

$$\begin{bmatrix} i_{s1}(t) \\ i_{s2}(t) \end{bmatrix} = \sum_{q=1}^{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{j\omega_q t} \tag{A.10}$$

$$\begin{aligned} \big[H_2(\omega_1, \omega_2)\big] &= \begin{bmatrix} H_{1,20}(\omega_1, \omega_2) & H_{1,02}(\omega_1, \omega_2) & H_{1,11}(\omega_1, \omega_2) \\ H_{2,20}(\omega_1, \omega_2) & H_{2,02}(\omega_1, \omega_2) & H_{2,11}(\omega_1, \omega_2) \end{bmatrix} \\ &= -\big[H_1(\omega_1 + \omega_2)\big]\big[Y^{(2)}(\omega_1 + \omega_2)\big]\big[A\big([H_1(\omega_1)], [H_1(\omega_2)]\big)\big] \end{aligned} \tag{A.11}$$

where,

$$\big[H_1(\omega_1 + \omega_2)\big] = \begin{bmatrix} H_{1,10}(\omega_1 + \omega_2) & H_{1,01}(\omega_1 + \omega_2) \\ H_{2,10}(\omega_1 + \omega_2) & H_{2,01}(\omega_1 + \omega_2) \end{bmatrix},$$

$$\big[Y^{(2)}(\omega_1 + \omega_2)\big] = \begin{bmatrix} Y_{11}^{(2)}(\omega_1 + \omega_2) & Y_{12}^{(2)}(\omega_1 + \omega_2) & Y_{13}^{(2)}(\omega_1 + \omega_2) \\ Y_{21}^{(2)}(\omega_1 + \omega_2) & Y_{22}^{(2)}(\omega_1 + \omega_2) & Y_{23}^{(2)}(\omega_1 + \omega_2) \end{bmatrix},$$

$$\big[A\big([H_1(\omega_1)],[H_1(\omega_2)]\big)\big] =$$
$$\begin{bmatrix} H_{1,10}(\omega_1)H_{1,10}(\omega_2) & H_{1,01}(\omega_1)H_{1,01}(\omega_2) & H_{1,01}(\omega_1)H_{1,10}(\omega_2)+H_{1,01}(\omega_2)H_{1,10}(\omega_1) \\ H_{2,10}(\omega_1)H_{2,10}(\omega_2) & H_{2,01}(\omega_1)H_{2,01}(\omega_2) & H_{2,01}(\omega_1)H_{2,10}(\omega_2)+H_{2,01}(\omega_2)H_{2,10}(\omega_1) \\ H_{1,10}(\omega_1)H_{2,10}(\omega_2) & H_{1,01}(\omega_1)H_{2,01}(\omega_2) & H_{1,01}(\omega_1)H_{2,10}(\omega_2)+H_{1,10}(\omega_1)H_{2,01}(\omega_2) \end{bmatrix}$$

A.2.3 DIDO Volterra Kernels: General Solution

The purpose of the black-box behavioural modelling approach is to quickly identify the systems kernels and use them to predict the response under the same conditions. Therefore, one might be interested in the general solution of Volterra kernels with no assumptions on the type of excitation. The derivation will not be shown here. However, the full solution to first and second-order Volterra Kernels are expressed by (A.12) and (A.13). It should be noted that these expressions can be reduced to the simpler expressions in the previous section if both ports are assumed to be excited with signals that have components at the same frequency grid.

$$\big[H_1(\omega_q;\omega_r)\big] = \begin{bmatrix} H_{1,10}(\omega_q;0) & H_{1,01}(0;\omega_r) \\ H_{2,10}(\omega_q;0) & H_{2,01}(0;\omega_r) \end{bmatrix} \tag{A.12}$$

where,

$$\begin{bmatrix} H_{1,10}(\omega_q;0) \\ H_{2,10}(\omega_q;0) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(\omega_q) & Y_{12}^{(1)}(\omega_q) \\ Y_{21}^{(1)}(\omega_q) & Y_{22}^{(1)}(\omega_q) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} H_{1,01}(0;\omega_r) \\ H_{2,01}(0;\omega_r) \end{bmatrix} = \begin{bmatrix} Y_{11}^{(1)}(\omega_r) & Y_{12}^{(1)}(\omega_r) \\ Y_{21}^{(1)}(\omega_r) & Y_{22}^{(1)}(\omega_r) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\big[H_2(\omega_{q_1},\omega_{q_2};\omega_{r_1},\omega_{r_2})\big] =$$
$$\begin{bmatrix} H_{1,20}(\omega_{q_1},\omega_{q_2};0,0) & H_{1,02}(0,0;\omega_{r_1},\omega_{r_2}) & H_{1,11}(\omega_{q_1},0;\omega_{r_1},0) \\ H_{2,20}(\omega_{q_1},\omega_{q_2};0,0) & H_{2,02}(0,0;\omega_{r_1},\omega_{r_2}) & H_{2,11}(\omega_{q_1},0;\omega_{r_1},0) \end{bmatrix} \tag{A.13}$$

where,

$$\begin{bmatrix} H_{1,20}(\omega_{q_1},\omega_{q_2};0,0) \\ H_{2,20}(\omega_{q_1},\omega_{q_2};0,0) \end{bmatrix} = -\big[H_1(\omega_{q_1}+\omega_{q_2};0)\begin{bmatrix} Y_{11}^{(2)}(\omega_{q_1}+\omega_{q_2}) \\ Y_{21}^{(2)}(\omega_{q_1}+\omega_{q_2}) \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$
$$.\big[A\big([H_1(\omega_{q_1};0)],[H_1(\omega_{q_2};0)]\big)\big]$$

$$\begin{bmatrix} H_{1,02}(0,0;\omega_{r_1},\omega_{r_2}) \\ H_{2,02}(0,0;\omega_{r_1},\omega_{r_2}) \end{bmatrix} = -\big[H_1(0;\omega_{r_1}+\omega_{r_2})\begin{bmatrix} Y_{12}^{(2)}(\omega_{r_1}+\omega_{r_2}) \\ Y_{22}^{(2)}(\omega_{r_1}+\omega_{r_2}) \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$
$$.\big[A\big([H_1(0;\omega_{r_1})],[H_1(0;\omega_{r_2})]\big)\big]$$

$$\begin{bmatrix} H_{1,11}(\omega_{q_1},0;\omega_{r_1},0) \\ H_{2,11}(\omega_{q_1},0;\omega_{r_1},0) \end{bmatrix} = -\big[H_1(\omega_{q_1}+\omega_{r_1};\omega_{q_1}+\omega_{r_1})\begin{bmatrix} Y_{13}^{(2)}(\omega_{q_1}+\omega_{r_1}) \\ Y_{23}^{(2)}(\omega_{q_1}+\omega_{r_1}) \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$
$$.\big[A\big([H_1(\omega_{q_1};\omega_{r_1})],[H_1(\omega_{q_1};\omega_{r_1})]\big)\big]$$

A.3 Conclusion

The relationship between the proposed HONPs and DIDO Volterra kernels was illustrated. It is evident that identifying HONPs is sufficient to identify Volterra kernels. However, HONPs formulation allows solution generality without suffering an explosion in the dimensionality of parameters.

A.4 Definitions

- $i_p(t)$ is the total time-domain WNL current at port p.
- $v_p(t)$ is the total time-domain WNL voltage at port p.
- $[y^{(n)}(t)]$ is the time-domain n^{th}-HOYPs in the matrix form.
- $[Y^{(n)}(f)]$ is the frequency-domain n^{th}-HOYPs in the matrix form.
- $P = \dfrac{d}{dt}$ is a differential operator.
- n_p is the response-order of port p.

- $v_{p,n_1n_2}(t)$ is the time-domain voltage component of order $n_1 + n_2$ at port p.
- $I_{p,q}$ is the complex amplitude of $i(t)$ at ω_q and port p.
- Q is the total number of tones in $i(t)$.
- $h_{p,n_1n_2}(\tau_{11}, \ldots, \tau_{1n_1}; \tau_{21}, \ldots, \tau_{2n_2})$ is the nonlinear impulse response of order $n_1 + n_2$.
- $(\tau_{11}, \ldots, \tau_{1n_1}; \tau_{21}, \ldots, \tau_{2n_2})$ represents all combinations between the time delays at port 1 and port 2, i.e., all combinations between $\tau_{11}, \ldots, \tau_{1n_1}$ and $\tau_{21}, \ldots, \tau_{2n_2}$.
- $H_{p,n_1n_2}(\omega_{q_1} \ldots \omega_{q_{n_1}}; \omega_{r_1} \ldots \omega_{r_{n_2}})$ is the nonlinear transfer function of order $n_1 + n_2$.
- $(\omega_{q_1} \ldots \omega_{q_{n_1}}; \omega_{r_1} \ldots \omega_{r_{n_2}})$ represents all combinations between different tones at port 1 and port 2, i.e., all combinations between $\omega_{q_1} \ldots \omega_{q_{n_1}}$ and $\omega_{r_1} \ldots \omega_{r_{n_2}}$.
- $H_{p,n_1n_2}(\omega_{q_1} \ldots \omega_{q_n})$ is the nonlinear transfer function of order n_1+n_2 when both ports are excited with signals that have same frequency grids.
- $[H_1(\omega)]$ is the frequency-domain first-order DIDO Volterra kernels in the matrix form, when both ports are excited with signals that have same frequency grids.
- $[H_2(\omega_1, \omega_2)]$ is the frequency-domain second-order DIDO Volterra kernels in the matrix form, when both ports are excited with signals that have same frequency grids.
- $[H_1(\omega_q; \omega_r)]$ is the general frequency-domain first-order DIDO Volterra kernels in the matrix form.
- $[H_2(\omega_{q_1}, \omega_{q_2}; \omega_{r_1}, \omega_{r_2})]$ is the general frequency-domain second-order DIDO Volterra kernels in the matrix form.

Appendix B

HONPs Modulated Signal Validations in Simulation Environment

B.1 Introduction

This appendix illustrates the ability of the proposed HONPs to predict nonlinear behaviours under modulated signal stimuli in the simulation environment. It also provides a simulation proof that these parameters can be used to linearize nonlinear DUTs. This is of great importance because HONPs are extracted using mixer-based receivers. Thus, using them for linearization solves the receiver bandwidth limitation problem that the predistortion field faces, especially after moving to 5G.

This appendix is organized as follows. First, HONPs implementation as a behavioural model in Keysight ADS circuit simulator is illustrated. Second, modulated signal validations are provided. Then, a new extraction technique is discussed to improve HONPs capability to linearize strongly nonlinear systems. Finally, the conclusion is presented.

B.2 Model Implementation

In Chapter 5, neural networks were utilized to generate a large-signal nonlinear model using HONPs at multiple bias points. In this section, HONPs at a single bias point are

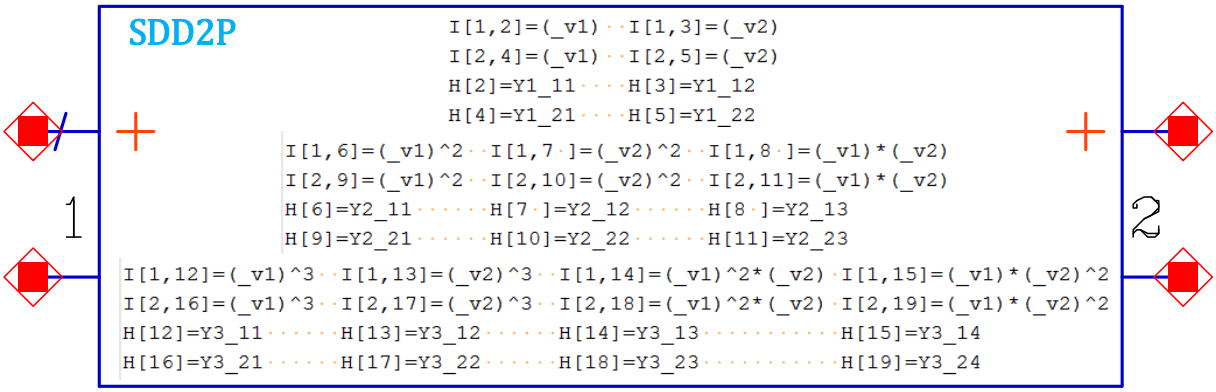


Figure B.1: HONPs implementation as a behavioural model in Keysight ADS circuit simulator.

implemented as a standalone behavioural model. This was achieved by implementing equation (3.4) in Keysight ADS circuit simulator using a two-port Symbolically Defined Device (SDD2P) component, as shown in Figure B.1. Extracted first, second, and third-HONPs were then implemented as table-based parameters versus frequency using Data Access Components (DACs).

B.3 Modulated Signal Validations

In order to verify that HONPs can predict WNL behaviours under modulated signal stimuli, the same DUT used in Chapter 3 (Figure 3.8) is used here. It was biased at (V_{gs},V_{ds}) of (-3.2, 12) V and first, second, and third-HONPs were extracted in the frequency range of 10 MHz to 6 GHz (40 frequency points were used). The HONPs model in Figure B.1 was generated and excited using two modulated signal stimuli under Envelope simulator (Figure 3.6):

- **Signal 1**: a 20 MHz 4C-WCDMA signal with 7.14 dB PAPR around 1.4 GHz [Figure B.2 (a)]
- **Signal 2**: a 50 MHz 10C-WCDMA signal with 10 dB PAPR around 800 MHz [Figure B.2 (b)]

Table B.1 compares between simulated first, second and third-order behaviours of the DUT and the HONPs model due to **Signal 1** and **Signal 2**, at four average input power

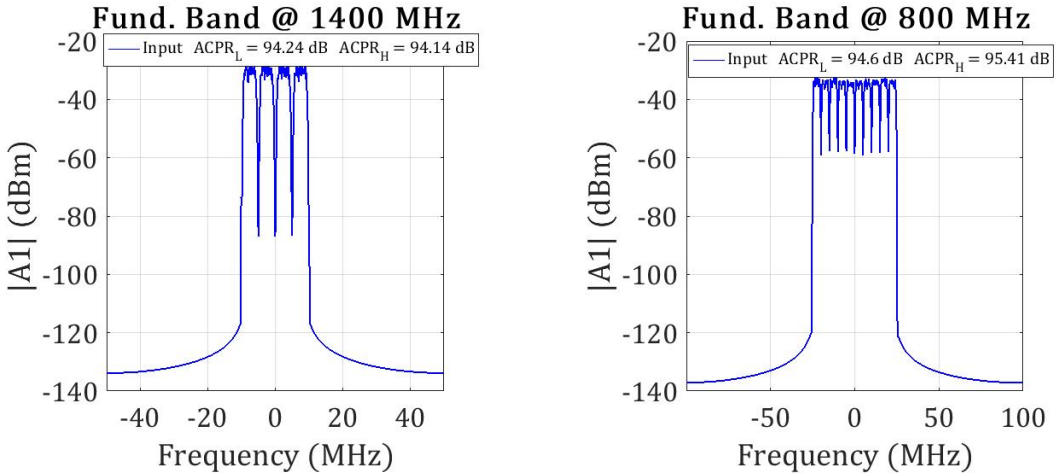


Figure B.2: (a) **Signal 1**: a 20 MHz 4C-WCDMA signal with 7.14 dB PAPR around 1.4 GHz, and (b) **Signal 2**: a 50 MHz 10C-WCDMA signal with 10 dB PAPR around 800 MHz.

levels. An example of the simulated B_1 and B_2 spectrums due to **Signal 1** at -8.9 dBm input average power is shown in Figure B.4. The following remarks should be noted:

- Since the generated HONPs model is a third-order model, only first, second, and third-order responses are expected to be well-predicted. For example, fourth-order responses in both the baseband and the second-harmonic band are poorly predicted. Also, fifth-order responses in the fundamental and third-harmonic bands are poorly predicted (these different-order responses are illustrated in Figure B.4). Therefore, the NMSE values reported in Table B.1 represent only second and third-order responses in the specified bands.

- Left and right adjacent channel power ratio ($ACPR_L$ and $ACPR_R$, respectively) were used as an indication of whether the DUT is operating in the WNL region or the strongly nonlinear region. The HONPs model demonstrated a good ability to predict WNL behaviours for ACPR values below 40 dBc.

- B_1-wave baseband spectrum values are too small (below -110 dBm) to be correctly predicted by the model. Therefore, it was not reported in Table B.1.

In order to validate HONPs ability to linearize nonlinear DUTs, digital predistortion

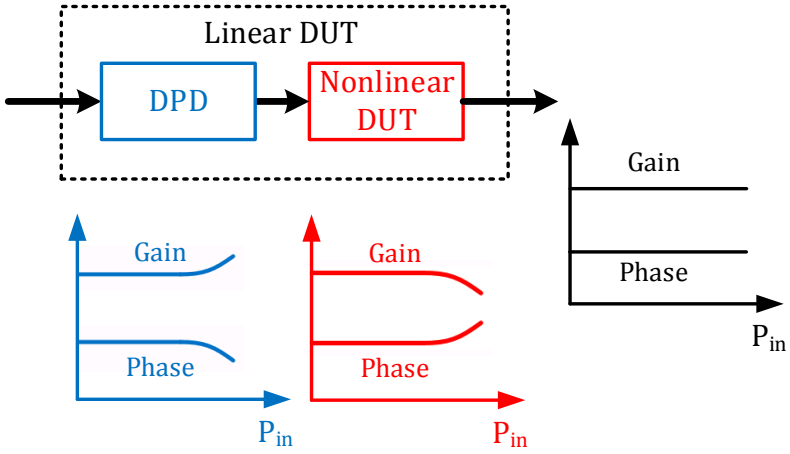


Figure B.3: An illustration of DPD concept.

(DPD) was used. DPD incorporates a nonlinear function before the nonlinear DUT to pre-distort the input signal. The resulting cascaded system of both the DPD and the nonlinear DUT behaves linearly in terms of the output gain and phase versus the input power (P_{in}), as shown in Figure B.3 [63]. DPD was applied to the DUT for the different stimuli and input average power levels discussed earlier, using the simulated B_2-wave, and the results are reported in the lower part of Table B.1. Three metrics are used to describe the performance of the DPD: NMSE and error vector magnitude (EVM) between the output and the original input, and the ACPR reduction. It is evident that a good linearization (i.e. NMSE and EVM are below -40 dB and 1%, respectively, and ACPR reduction by 10 dB) was achieved for ACPR values below 40 dBc. Deriving the DUT with higher input average power resulted in a poor prediction of the output and hence poor linearization results (simulation # 4 of Table B.1). Figure B.5 shows B_2-wave spectrums before and after applying DPD with the ACPR value in each case. Figure B.9 shows both gain distortion and phase distortion (AM/PM) before and after applying DPD for the different stimuli and input levels reported in Table B.1. It is worth mentioning that HONPs showed a great ability for linearizing the DUT as long as it is working in the WNL region. Also, the model shows the same performance regardless the type of the stimulus, which is predicted from the discussion presented in Chapter 3.

Stimulus		**Signal 1**				**Signal 2**			
Simulation #		1	2	3	4	1	2	3	4
P_{in} (dBm)		-14.9	-8.9	5.4	-2.9	-17.7	-11.7	-8.2	-5.7
P_{out} (dBm)		0.7	6.7	10.1	12.4	-4.3	5.3	8.7	11.1
$ACPR_L$ (dBc)		61	49	42	38	59	47	41	37
$ACPR_R$ (dBc)		61	49	42	38	58	46	40	36
NMSE B_1	F (dB)	-48	-48	-44	-35	-37	-37	-36	-33
	$2H$ (dB)	-33	-25	-28	-16	-31	-27	-24	-14
	$3H$ (dB)	-35	-18	-7	0.6	-28	-13	-3	3
	$IMD3$ (dB)	-55	-33	-18	-7.7	-13	-12	-3	3
NMSE B_2	F (dB)	-50	-49	-47	-31	-41	-40	-36	-28
	$2H$ (dB)	-34	-22	-18	-9	-30	-20	-13	-7
	BB (dB)	-29	-24	-40	-10	-19	-18	-14	-9
	$3H$ (dB)	-34	-21	-15	-5	-29	-18	-10	-3
	$IMD3$ (dB)	-52	-31	-17	-7	-21	-15	-8	-2
w/o	$NMSE$ (dB)	-44.6	-41.4	-35	-31	-40.9	-36.6	-31.5	-27.9
	EVM (%)	0.58	0.84	1.7	2.8	0.8	1.4	2.6	4
with DPD	$ACPR_L$ (dBc)	89	71	56	44	85	66	51.6	41
	$ACPR_R$ (dBc)	87	70	55	44	81	63	50	40
	$NMSE$ (dB)	-49	-44	-42	-36	-42.6	-42.4	-39.3	-32
	EVM (%)	0.35	0.63	0.72	1.6	0.74	0.76	1.07	2.5

Table B.1: The resulting NMSE in the first, second, and third-order behaviours between B_1 and B_2-waves of the DUT and the HONPs model. F, $2H$, BB, $3H$, and $IMD3$ refer to the fundamental band, second-harmonic band, baseband, third-harmonic band and IMD3 band, respectively. The lower part of the table shows both NMSE and EVE without (w/o) and with DPD.

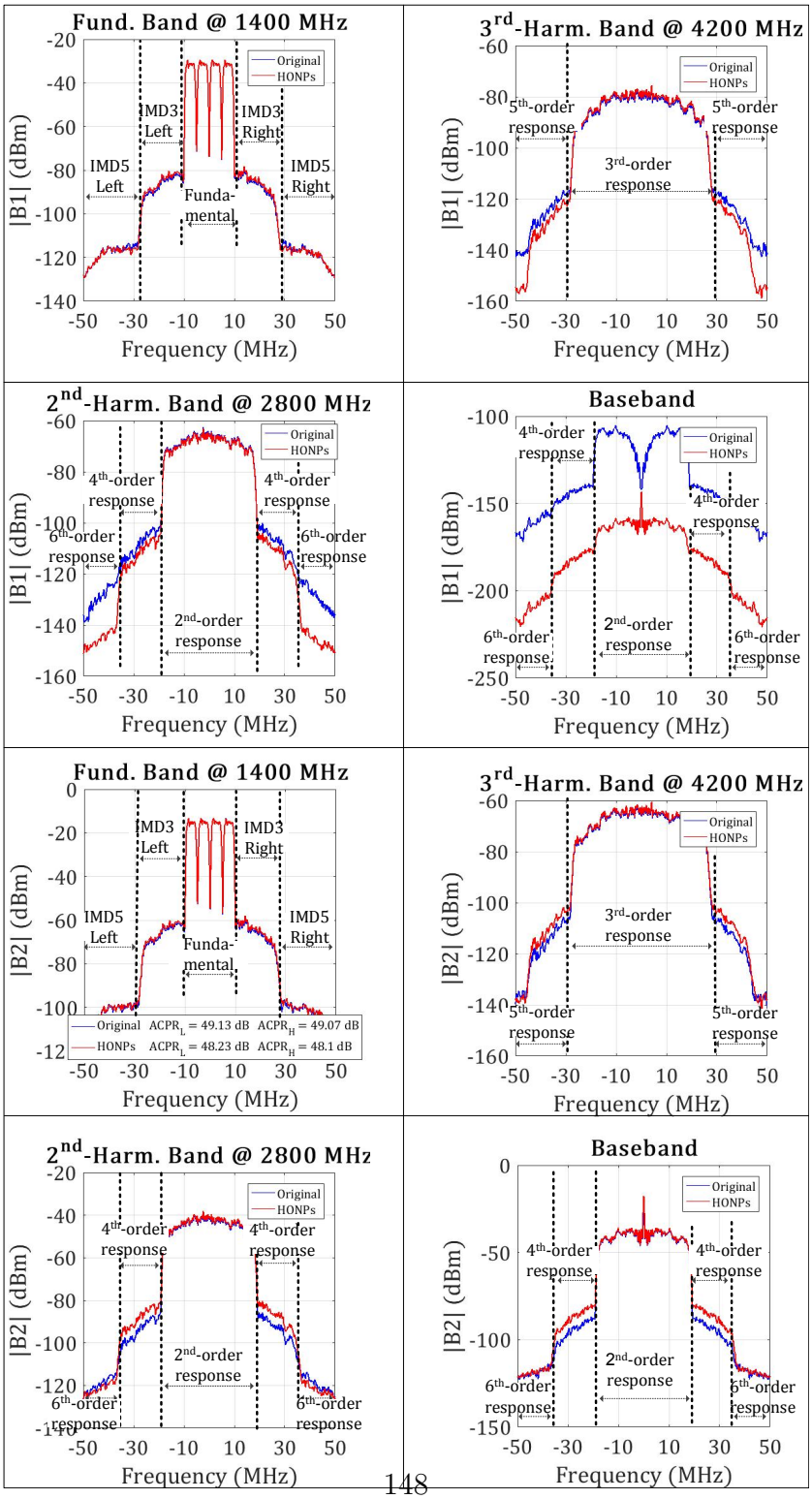


Figure B.4: Simulated and HONPs predicted B_1 and B_2-waves spectrum at the fundamental band, second-harmonic band, baseband, third-harmonic band due to **Signal 1** stimulus.

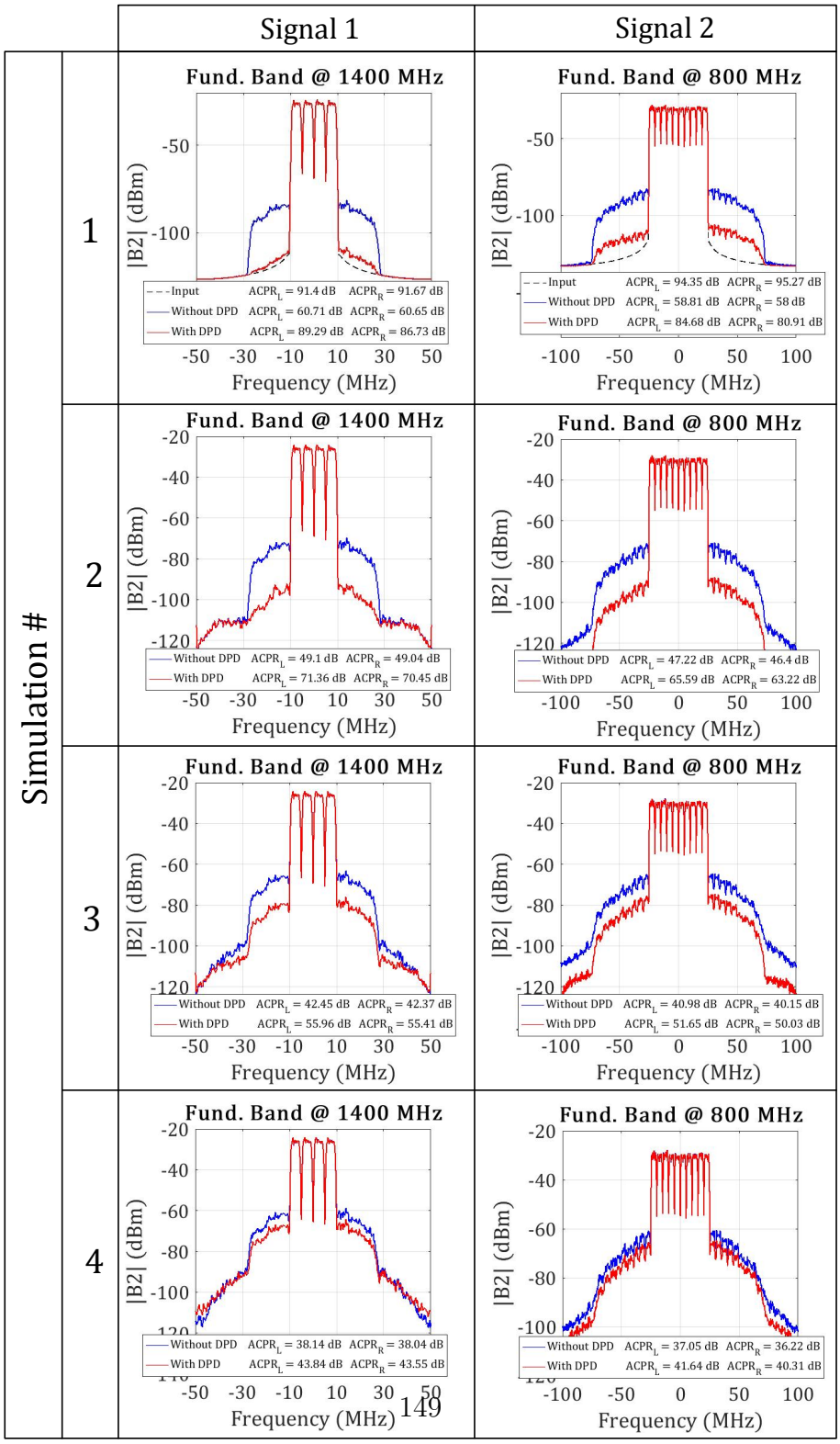
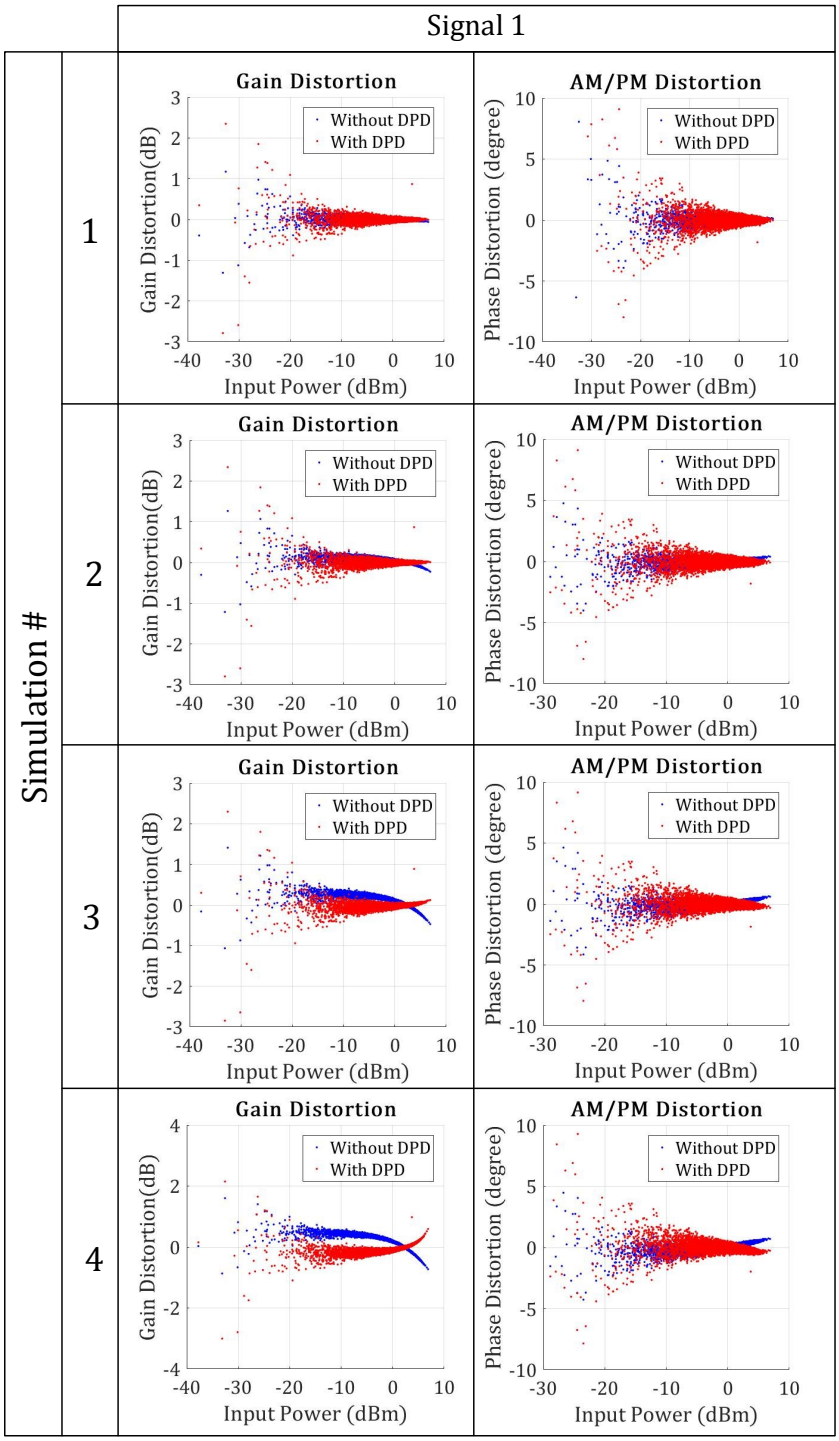
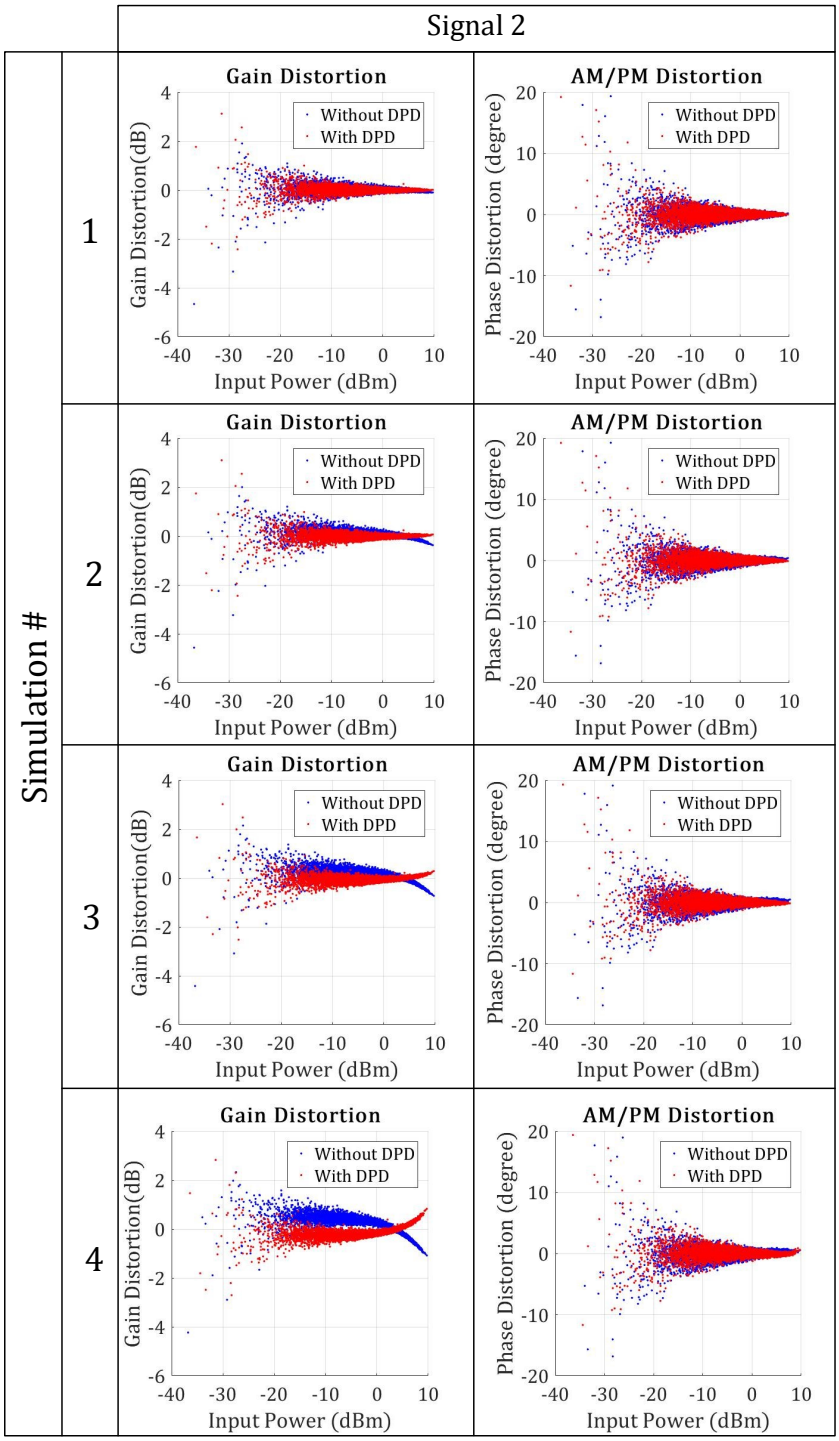


Figure B.5: B_2-wave spectrums before and after applying DPD (for the stimuli and input average power levels reported in Table B.1).


Signal 1
Simulation #
1
2
3
4
Gain Distortion
AM/PM Distortion
Without DPD
With DPD
Gain Distortion(dB)
Phase Distortion (degree)
Input Power (dBm)

(a)



(b)

Figure B.6: Gain distortion and AM/PM versus the input power before and after applying DPD due to: (a) **Signal 1**, and (b) **Signal 2** stimuli (for the input average power levels reported in Table B.1).

B.4 Linearizing Strongly Nonlinear Systems using HONPs

The previous section showed that the HONPs model works well for both predicting and linearizing WNL behaviours (defined by ACPR values below 40 dBc) due to modulated signal stimuli. This section extends HONPs ability to linearize responses with ACPR values up to 30 dBc. This was achieved by borrowing the large-signal operating point (LSOP) concept from X-parameters and applying it to HONPs as follows. DPD assumes a priori-knowledge of the type of the stimulus. Therefore, a large-signal tone with an average power of 2.6 dBm was applied at 1.4 GHz (which are the average input power and the center frequency of **Signal 1** stimulus, respectively)[1]. This large-signal tone in addition to the DC biasing point were assumed to be the LSOP, and HONPs were extracted using the same technique explained in Chapter 3. Table B.2 shows both prediction and linearization results, and the following should be noted:

- Stimulating the DUT with a large signal tone at a fixed frequency while using small-signal tones at both ports to sweep the required frequency range generates IMD tone. Only the small-signal fundamental tone and its harmonics are used for HONPs extraction. Also, it is important to maintain a non-commensurate relationship between the large-signal tone and the small-signal tone frequencies. In other words, the large-signal IMDs and harmonics should not fall on top of the small-signal harmonics.

- Extracting the HONPs model around an LSOP showed an improved ability to predict and linearize nonlinear behaviours with ACPR up to 30 dBc. Figure B.8 shows B_2-wave spectrums before and after applying DPD for the different stimuli and input power levels reported in Table B.2.

- The poor prediction of second-order behaviours in both the baseband and the second-harmonic band indicates that a fourth-order model needs to be generated.

- It was proved in Chapter 3 that HONPs are power independent, and hence they have the same value as long as they are extracted in the WNL region (Figure 3.11). Extracting HONPs around an LSOP is more difficult than in the original case as it is more complicated to identify the extraction range of these parameters. Figure B.7 shows the extracted parameters versus the fundamental tone input power. It is evident that the injection of a large-signal tone reduces the region where these parameters are power independent.

[1] The same was applied for **Signal 2** stimulus.

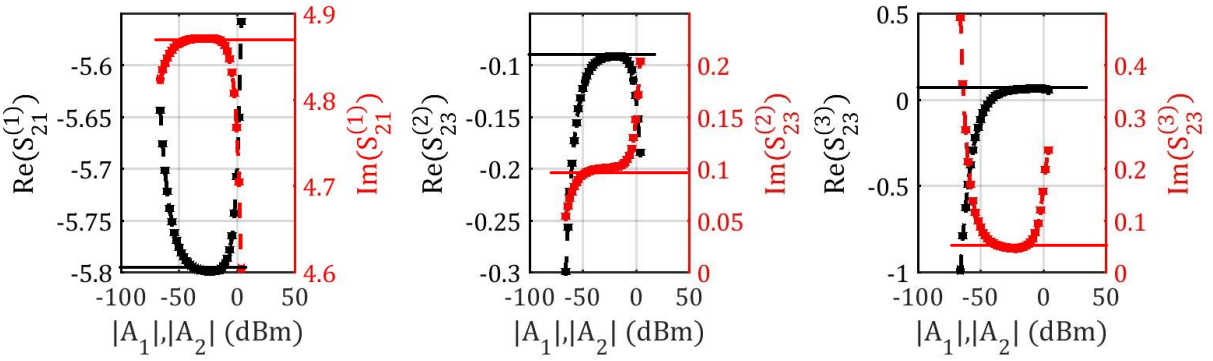


Figure B.7: The real and imaginary parts of the extracted HOSPs around an LSOP versus fundamental tone input power at 1.1 GHz. The solid lines represent the correct value of the element.

- Figure B.9 shows both gain distortion and AM/PM before and after applying DPD for the different stimuli and input levels reported in Table B.2.

B.5 Conclusion

HONPs at a single bias point were used as a standalone behavioural model. Their ability to predict and linearize WNL behaviours due to modulated signal stimuli were validated in the simulation environment. The model proved to be stimulus-independent, which confirms the results in Chapter 3. Finally, HONPs were extracted around an LSOP in order to tune the model towards DPD applications. Good linearization results were reported for ACPR values up to 30 dBc.

Stimulus		**Signal 1**				**Signal 2**			
Simulation #		1	2	3	4	1	2	3	4
P_{in} (dBm)		-11.4	-4	-0.1	2.6	-17.7	-6.8	-3.7	-1.4
P_{out} (dBm)		5.9	13.1	16.7	18.9	1.17	11.9	14.7	16.7
$ACPR_L$ (dBc)		59.5	44.5	35.8	29.7	65	42	36	31
$ACPR_R$ (dBc)		59.4	44.4	35.7	29.6	64	42	35	30
NMSE B_1	F (dB)	-38	-36.6	-35	-34	-36	-35	-34	-34
	$2H$ (dB)	11	-14.4	-6	-7.1	11.3	7.1	-5	-6
	$3H$ (dB)	1.15	-1.9	-10.7	-14.3	1.9	-4	-11	-11
	$IMD3$ (dB)	-35.2	-24.5	-22.3	-20.5	0.8	-3	-10	-10
NMSE B_2	F (dB)	-25	-24	-23	-21	-25	-23	-23	-21
	$2H$ (dB)	2.4	4.1	8	8.2	3	5	6	5
	BB (dB)	3.5	6	7	-1.2	3	6	6	1.5
	$3H$ (dB)	-19	-21	-25	-13.8	-19	-20	-20	-12
	$IMD3$ (dB)	-25	-23	-22	-15.8	-16	-24	-26	-16
w/o	$NMSE$ (dB)	-43.8	-37.3	-28.8	-22.6	-41	-32.8	-26.5	-21.8
	EVM (%)	0.64	1.4	3.64	7.4	0.89	2.3	4.7	8.1
with DPD	$ACPR_L$ (dBc)	73	62.9	52.7	43.5	80	60	50	43
	$ACPR_R$ (dBc)	73	62.4	52.3	43.1	80	59	50	42
	$NMSE$ (dB)	-48.6	-43.7	-42	-36	-42.1	-41.5	-39	-34
	EVM (%)	0.37	0.65	0.8	1.6	0.78	0.8	1.1	2

Table B.2: The resulting NMSE in the first, second, and third-order behaviours between B_1 and B_2-waves of the DUT and the HONPs model. F, $2H$, BB, $3H$, and $IMD3$ refer to the fundamental band, second-harmonic band, baseband, third-harmonic band and IMD3 band, respectively. The lower part of the table shows both NMSE and EVE without (w/o) and with DPD.

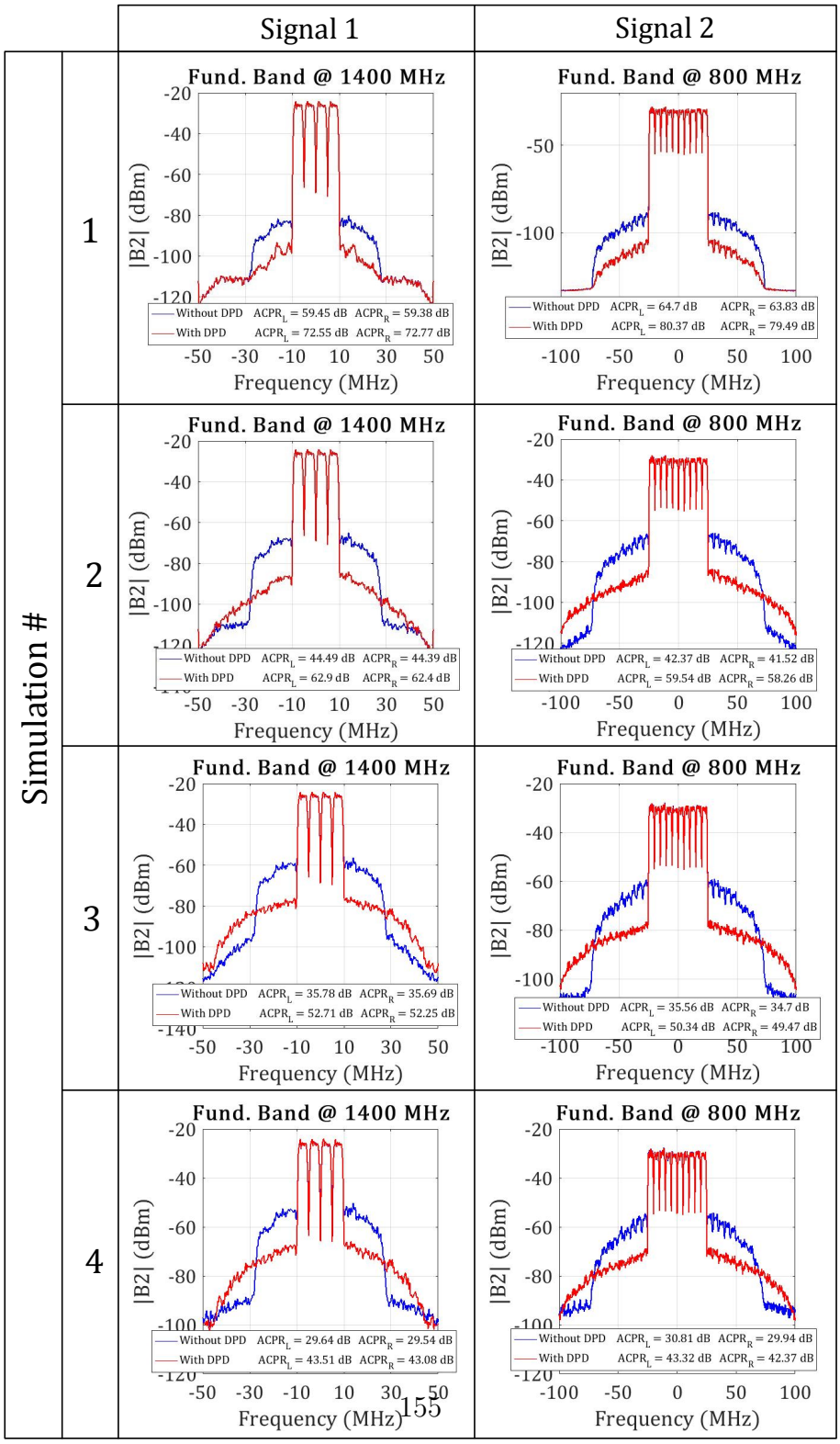
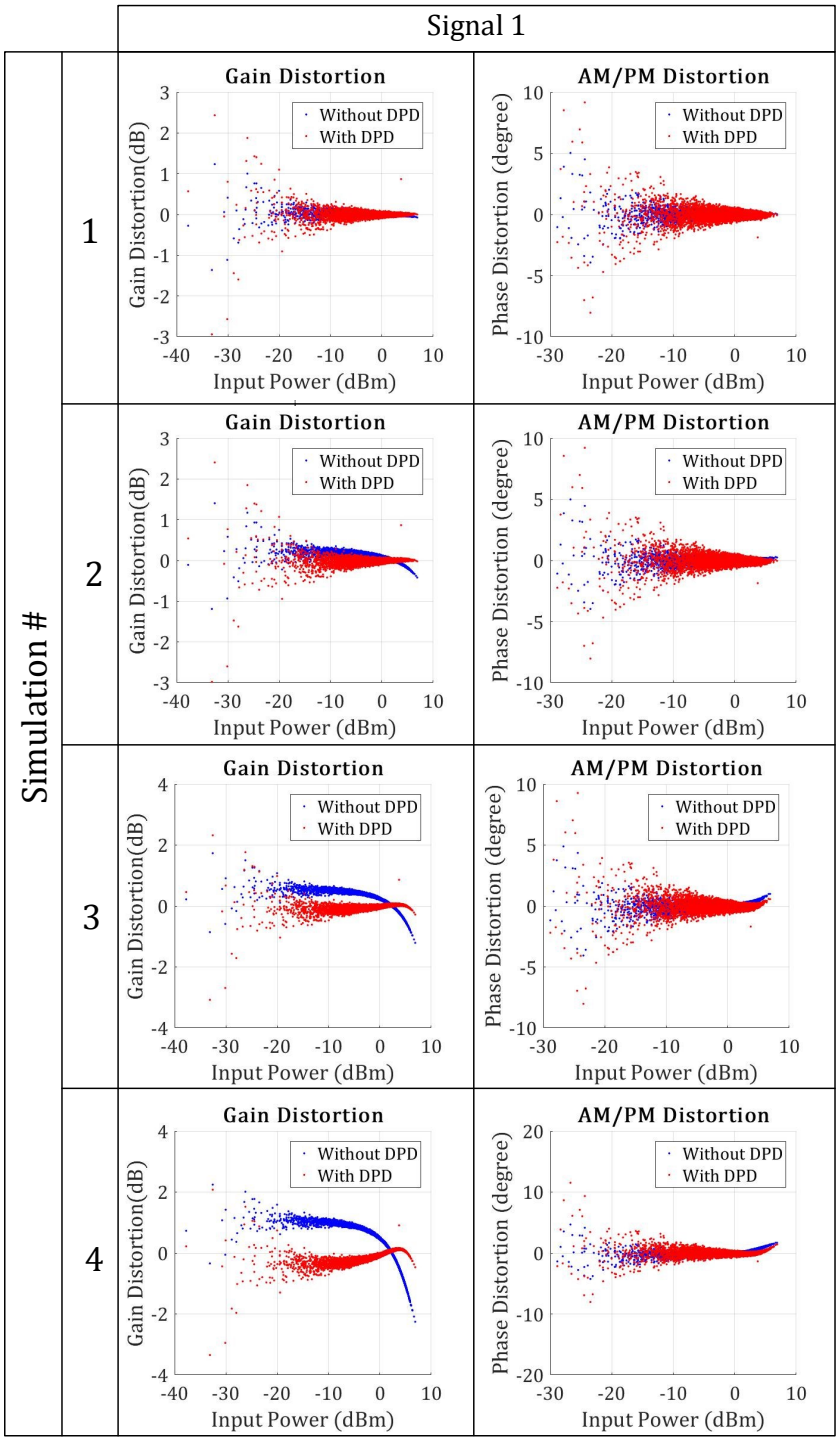
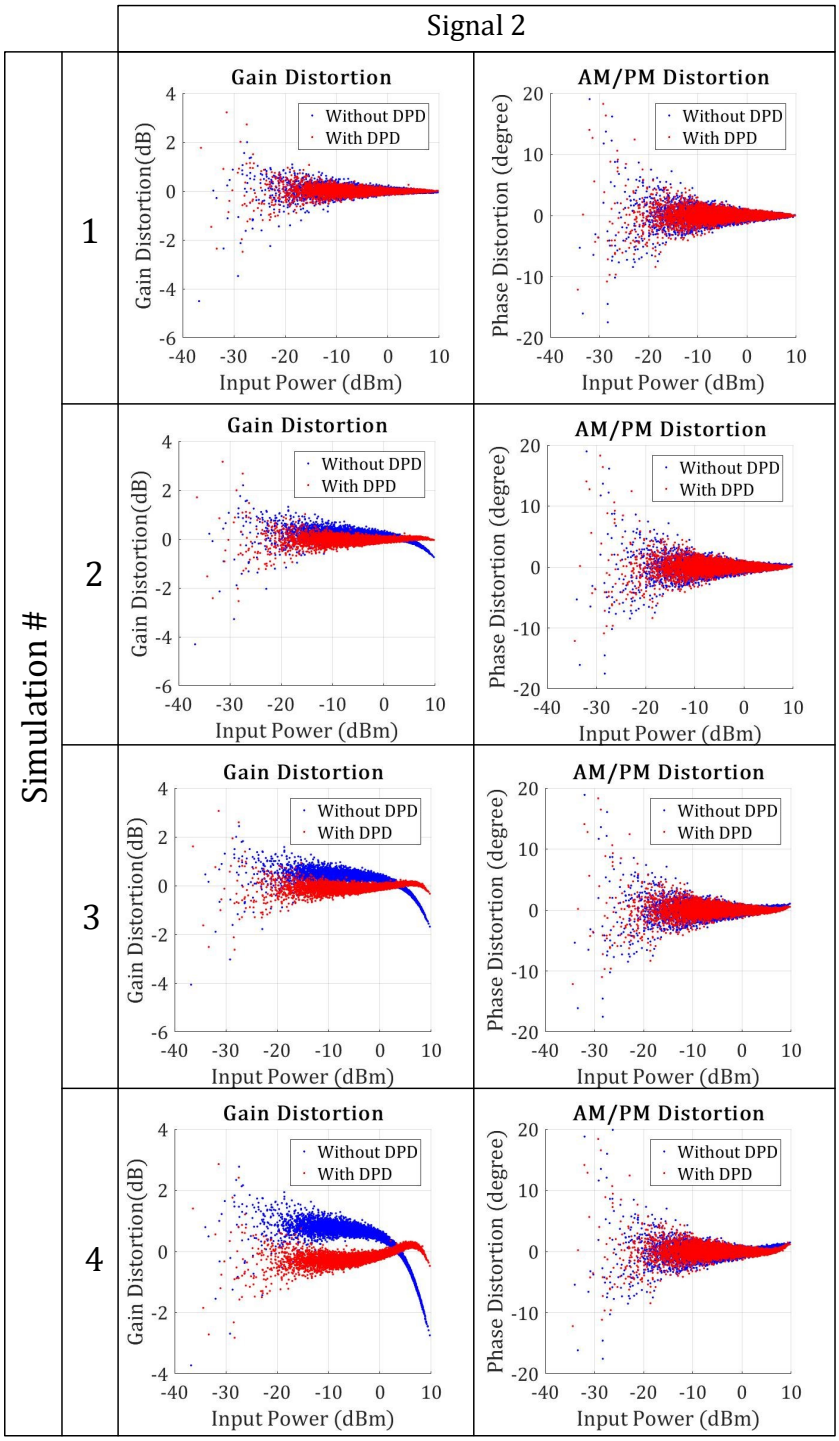


Figure B.8: B_2-wave spectrums before and after applying DPD (for the stimuli and input average power levels reported in Table B.2).


Signal 1
Simulation #
1
2
3
4
Gain Distortion
AM/PM Distortion
Without DPD
With DPD
Gain Distortion(dB)
Phase Distortion (degree)
Input Power (dBm)

(a)



(b)

Figure B.9: Gain distortion and AM/PM versus the input power before and after applying DPD due to: (a) **Signal 1**, and (b) **Signal 2** stimuli (for the input average power levels reported in Table B.2).

Appendix C

Chapter 4 Neural Networks Fixed Weights

- Figure 4.3

$$[W_1^{(k)}] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad [W_2^{(k)}] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad [W_3^{(k)}] = \left[\text{ones}(4\times 1)\right] \tag{C.1}$$

- Figure 4.4

$$[W_1^{(k)}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad [W_2^{(k)}] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad [W_3^{(k)}] = [\text{diagonal}(8\times 8)] \tag{C.2}$$

- Figure 4.5

$$[W_1^{(k)}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad , \quad [W_2^{(k)}] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad , \quad [W_3^{(k)}] = [\text{diagonal}(8\times 8)] \tag{C.3}$$

- Figure 4.7

$$[W_1] = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad , \quad [W_2] = [\text{diagonal}(8\times 8)] \ ,$$

$$[W_3] = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix} \quad , \quad [W_4] = [\text{diagonal}(2\times 2)] \ ,$$

$$[W_5] = \begin{bmatrix} \text{ones}(8\times 1) & \text{zeros}(8\times 1) \\ \text{zeros}(8\times 1) & \text{ones}(8\times 1) \end{bmatrix} \quad , \quad [W_7] = \begin{bmatrix} 1 & 1 \end{bmatrix} \ , \tag{C.4}$$

$$[W_8] = \begin{bmatrix} \text{ones}(8\times 1) \end{bmatrix} \ , \quad [W_9] = \begin{bmatrix} \text{diagonal}(8\times 8) & \text{diagonal}(8\times 8) \end{bmatrix} \ ,$$

$$[W_6] = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Figure 4.8

$$[W_1] = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad [W_2] = [\text{diagonal}(8\times 8)] \,,$$

$$[W_3] = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix}, \quad [W_4] = -Z_o^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{C.5}$$

$$[W_5] = Z_o \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}, \quad [W_6] = -Z_o^2 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where, Z_o is the system characteristic impedance.

Appendix D

Chapter 5 Neural Networks Fixed Weights

- Figure 5.4

$$[W_1] = \begin{bmatrix} \text{ones(S}\times\text{1)} & \text{zeros(S}\times\text{1)} \\ \text{zeros(S}\times\text{1)} & \text{ones(S}\times\text{1)} \end{bmatrix} , \quad [W_2] = \begin{bmatrix} \text{diagonal(2S}\times\text{2S)} \end{bmatrix} ,$$

$$[W_3] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} \\ \text{diagonal(S}\times\text{S)} \end{bmatrix} , \quad [W_4] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \end{bmatrix} ,$$

$$[W_5] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} \end{bmatrix} , \quad [W_6] = \begin{bmatrix} \text{ones(1}\times\text{S)} & \text{zeros(1}\times\text{S)} \\ \text{zeros(1}\times\text{S)} & \text{ones(1}\times\text{S)} \end{bmatrix} ,$$

$$[W_7] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} & \text{zeros(S}\times\text{S)} \\ \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \\ \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \end{bmatrix} , \quad [W_8] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} & \text{zeros(S}\times\text{S)} \\ \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \\ \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \end{bmatrix} ,$$

$$[W_9] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} \\ \text{diagonal(S}\times\text{S)} \\ \text{diagonal(S}\times\text{S)} \end{bmatrix} \text{ and its size} = (n+1)S \times S, \quad [W_{10}] = \begin{bmatrix} \text{diagonal(3S}\times\text{3S)} \end{bmatrix} ,$$

$$[W_{11}] = \begin{bmatrix} [W_{11}^{(1)}] & [W_{11}^{(2)}] \\ [W_{11}^{(3)}] & [W_{11}^{(4)}] \end{bmatrix} , \quad [W_{12}] = \begin{bmatrix} [W_{12}^{(1)}] \\ [W_{12}^{(2)}] \end{bmatrix} : \tag{D.1}$$

$$[W_{11}^{(1)}] = \begin{cases} \text{diagonal}\Big(\Big(\frac{n+1}{2}\Big)S \times \Big(\frac{n+1}{2}\Big)S\Big) & : n \text{ is odd} \\ \text{diagonal}\Big(\Big(\frac{n}{2}+1\Big)S \times \Big(\frac{n}{2}+1\Big)S\Big) & : n \text{ is even} \end{cases},$$

$$[W_{11}^{(2)}] = \begin{cases} \text{zeros}\Big(\Big(\frac{n+1}{2}\Big)S \times \Big(\frac{n-1}{2}\Big)S\Big) & : n \text{ is odd} \\ \text{zeros}\Big(\Big(\frac{n}{2}+1\Big)S \times \Big(\frac{n}{2}-1\Big)S\Big) & : n \text{ is even} \end{cases},$$

$$[W_{11}^{(3)}] = \begin{cases} \text{zeros}\Big(\Big(\frac{n+1}{2}\Big)S \times \Big(\frac{n-1}{2}\Big)S\Big) & : n \text{ is odd} \\ \text{zeros}\Big(\Big(\frac{n}{2}\Big)S \times \Big(\frac{n}{2}\Big)S\Big) & : n \text{ is even} \end{cases},$$

$$[W_{11}^{(4)}] = \begin{cases} \text{diagonal}\Big(\Big(\frac{n+1}{2}\Big)S \times \Big(\frac{n+1}{2}\Big)S\Big) & : n \text{ is odd} \\ \text{diagonal}\Big(\Big(\frac{n}{2}\Big)S \times \Big(\frac{n}{2}\Big)S\Big) & : n \text{ is even} \end{cases},$$

$$[W_{12}^{(1)}] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} & \text{zeros(S}\times\text{S)} \\ \text{diagonal(S}\times\text{S)} & \text{zeros(S}\times\text{S)} \\ \cdots & \\ \text{diagonal(S}\times\text{S)} & \text{zeros(S}\times\text{S)} \end{bmatrix} \quad \text{and its size} = \begin{cases} \Big(\frac{n+1}{2}\Big)S \times 2 & : n \text{ is odd} \\ \Big(\frac{n}{2}\Big)S \times 2 & : n \text{ is even} \end{cases},$$

$$[W_{12}^{(2)}] = \begin{bmatrix} \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \\ \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \\ \cdots & \\ \text{zeros(S}\times\text{S)} & \text{diagonal(S}\times\text{S)} \end{bmatrix} \quad \text{and its size} = \begin{cases} \Big(\frac{n+1}{2}\Big)S \times 2 & : n \text{ is odd} \\ \Big(\frac{n}{2}+1\Big)S \times 2 & : n \text{ is even} \end{cases}$$

$$[W_{13}] = \begin{bmatrix} \text{diagonal(S}\times\text{S)} \\ \text{diagonal(S}\times\text{S)} \\ \cdots \\ \text{diagonal(S}\times\text{S)} \end{bmatrix}, \quad [W_{14}] = \big[\text{diagonal([(n+1)S}\times\text{(n+1)S]}\big]$$

- Figure 5.7

$$[W_1] = \begin{bmatrix} \text{ones}(1\times\text{S}) & \text{zeros}(1\times\text{S}) \\ \text{zeros}(1\times\text{S}) & \text{ones}(1\times\text{S}) \end{bmatrix} \ , \quad [W_2] = \big[\text{ones}(2\times1)\big] \ , \quad [W_6] = \big[\text{ones}((\text{n}+1)\times1)\big]$$

$$[W_3] = \begin{bmatrix} \text{ones}(1\times\text{S}) & \text{zeros}(1\times\text{S}) & \text{zeros}(1\times\text{S}) \\ \text{zeros}(1\times\text{S}) & \text{ones}(1\times\text{S}) & \text{zeros}(1\times\text{S}) \\ \text{zeros}(1\times\text{S}) & \text{zeros}(1\times\text{S}) & \text{ones}(1\times\text{S}) \end{bmatrix} \ , \quad [W_4] = \big[\text{ones}(3\times1)\big] \ , \tag{D.2}$$

$$[W_5] = \begin{bmatrix} \text{ones}(1\times\text{S}) & \text{zeros}(1\times\text{S}) & \dots & \text{zeros}(1\times\text{S}) \\ \text{zeros}(1\times\text{S}) & \text{ones}(1\times\text{S}) & \dots & \text{zeros}(1\times\text{S}) \\ \dots & \dots & \dots & \dots \\ \text{zeros}(1\times\text{S}) & \text{zeros}(1\times\text{S}) & \dots & \text{ones}(1\times\text{S}) \end{bmatrix} \quad \text{and its size} = (n+1)\times(n+1)S$$

- Figure 5.8

$$[W_1] = \big[\text{diagonal}(2\times2)\big] \ , \quad [W_2] = \big[\text{diagonal}(3\times3)\big] \ ,$$

$$[W_3] = \big[\text{diagonal}((\text{n}+1)\times(\text{n}+1))\big] \tag{D.3}$$

- Figure 5.9

$$[W_1] = \big[\text{diagonal}(\text{S}\times\text{S})\big] \ , \quad [W_2] = \big[\text{ones}(1\times\text{S})\big] \tag{D.4}$$

References

[1] S. Dahmen-Lhuissier, "ETSI - 5G mobile system requirements discussed at ETSI Future Mobile Summit."

[2] T. S. Rappaport, R. W. H. Jr, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*, 1st ed. Upper Saddle River, NJ: Prentice Hall, Sep. 2014.

[3] J. Thompson, X. Ge, H.-C. Wu, R. Irmer, H. Jiang, G. Fettweis, and S. Alamouti, "5g wireless communication systems: prospects and challenges [Guest Editorial]," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 62–64, Feb. 2014.

[4] Iwpc, "Evolutionary & Disruptive Visions: Towards Ultra High Capacity Networks," Tech. Rep., Apr. 2014.

[5] S. C. Cripps, *RF Power Amplifiers for Wireless Communications, Second Edition*, 2nd ed. Artec House, May 2006.

[6] M. Naseri, "Extended bandwidth doherty power amplifier for carrier aggregated signals," Master's thesis, 2014. [Online]. Available: http://hdl.handle.net/10012/8665

[7] "Emerging Radio Systems Group | University of Waterloo," https://ece.uwaterloo.ca/~emrg/, accessed: 2016-01-18.

[8] J. Pedro and T. Reis Cunha, "Predictable Behavior: Behavioral Modeling from Measured Data," *IEEE Microwave Magazine*, vol. 15, no. 6, pp. 75–90, Sep. 2014.

[9] P. Aaen, *Modeling and Characterization of RF and Microwave Power FETs*, 1st ed. Cambridge: Cambridge University Press, Jun. 2011.

[10] G. Crupi and D. Schreurs, *Microwave De-embedding: From Theory to Applications*, 1st ed. Amsterdam ; Boston: Academic Press, Dec. 2013.

[11] G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, "A new method for determining the FET small-signal equivalent circuit," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, no. 7, pp. 1151–1159, Jul. 1988.

[12] G. Kompa, "Modeling of dispersive microwave FET devices using a quasi-static approach," *International Journal of Microwave and Millimeter-Wave Computer-Aided Engineering*, vol. 5, no. 3, pp. 173–194, May 1995.

[13] A. Jarndal and G. Kompa, "A new small-signal modeling approach applied to GaN devices," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 11, pp. 3440–3448, Nov. 2005.

[14] A. Jarndal, "Genetic algorithm based extraction method for distributed small-signal model of GaN HEMTs," in *2010 IEEE International Conference on Semiconductor Electronics (ICSE)*, Jun. 2010, pp. 41–44.

[15] G. Avolio, A. Raffo, I. Angelov, V. Vadala, G. Crupi, A. Caddemi, G. Vannini, and D.-P. Schreurs, "Millimeter-Wave FET Nonlinear Modelling Based on the Dynamic-Bias Measurement Technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2526–2537, Nov. 2014.

[16] D. Root, S. Fan, and J. Meyer, "Technology Independent Large Signal Non Quasi-Static FET Models by Direct Construction from Automatically Characterized Device Data," in *Microwave Conference, 1991. 21st European*, vol. 2, Sep. 1991, pp. 927–932.

[17] Y. Long, Y.-X. Guo, and Z. Zhong, "A 3-D Table-Based Method for Non-Quasi-Static Microwave FET Devices Modeling," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 10, pp. 3088–3095, Oct. 2012.

[18] J. King and T. Brazil, "Nonlinear Electrothermal GaN HEMT Model Applied to High-Efficiency Power Amplifier Design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 444–454, Jan. 2013.

[19] J. Joh, J. del Alamo, U. Chowdhury, T.-M. Chou, H.-Q. Tserng, and J. Jimenez, "Measurement of Channel Temperature in GaN High-Electron Mobility Transistors," *IEEE Transactions on Electron Devices*, vol. 56, no. 12, pp. 2895–2901, Dec. 2009.

[20] J. King and T. Brazil, "Equivalent circuit GaN HEMT model accounting for gate-lag and drain-lag transient effects," in *2012 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR)*, Jan. 2012, pp. 93–96.

[21] F. van Raay, R. Quay, M. Seelmann-Eggebert, D. Schwantuschke, D. Peschel, M. Schlechtweg, and O. Ambacher, "New Low-Frequency Dispersion Model for AlGaN/GaN HEMTs Using Integral Transform and State Description," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 154–167, Jan. 2013.

[22] W. Van Moer and L. Gomme, "NVNA versus LSNA: enemies or friends?" *IEEE Microwave Magazine*, vol. 11, no. 1, pp. 97–103, Feb. 2010.

[23] G. Pailloncy, G. Avolio, M. Myslinski, Y. Rolain, M. Bossche, and D. Schreurs, "Large-Signal Network Analysis Including the Baseband," *IEEE Microwave Magazine*, vol. 12, no. 2, pp. 77–86, Apr. 2011.

[24] F. M. Ghannouchi and M. S. Hashmi, *Load-Pull Techniques with Applications to Power Amplifier Design*, 2013th ed. New York: Springer, Jun. 2012.

[25] J. Xu, J. Horn, M. Iwamoto, and D. Root, "Large-signal FET model with multiple time scale dynamics from nonlinear vector network analyzer data," in *Microwave Symposium Digest (MTT), 2010 IEEE MTT-S International*, May 2010, pp. 417–420.

[26] J. P. Dunsmore, *Handbook of Microwave Component Measurements: with Advanced VNA Techniques*. Chichester, West Sussex, United Kingdom: Wiley, Oct. 2012.

[27] J. Xu, S. Halder, F. Kharabi, J. McMacken, J. Gering, and D. Root, "Global dynamic FET model for GaN transistors: DynaFET model validation and comparison to locally tuned models," in *ARFTG Microwave Measurement Conference (ARFTG), 2014 83rd*, Jun. 2014, pp. 1–6.

[28] G. Avolio, A. Raffo, I. Angelov, G. Crupi, G. Vannini, and D. Schreurs, "A novel technique for the extraction of nonlinear model for microwave transistors under dynamic-bias operation," in *Microwave Symposium Digest (IMS), 2013 IEEE MTT-S International*, Jun. 2013, pp. 1–3.

[29] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Krieger Pub Co, Aug. 2006.

[30] S. A. Maas, *Nonlinear Microwave and RF Circuits, 2nd Edition*, 2nd ed. Boston, MA: Artech House Publishers, Jan. 2003.

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

[32] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic Deviation Reduction-Based Volterra Behavioral Modeling of RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[33] P. Roblin, *Nonlinear RF Circuits and Nonlinear Vector Network Analyzers: Interactive Measurement and Design Techniques*. Cambridge ; New York: Cambridge University Press, Jul. 2011.

[34] D. E. Root, J. Verspecht, D. Sharrit, J. Wood, and A. Cognata, "Broad-band polyharmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 11, pp. 3656–3664, Nov. 2005.

[35] Verspecht and D. E. Root, "Polyharmonic distortion modeling," *IEEE Microwave Magazine*, vol. 7, no. 3, pp. 44–57, Jun. 2006.

[36] D. E. Root, J. Verspecht, J. Horn, and M. Marcu, *X-Parameters: Characterization, Modeling, and Design of Nonlinear RF and Microwave Components*, 1st ed. Cambridge, United Kingdom: Cambridge University Press, Nov. 2013.

[37] D. Bespalko, "Modular Nonlinear Characterization System and Large-Signal Behavioral Modelling of Unmatched Transistors for Streamlined Power Amplifier Design," Jan. 2016. [Online]. Available: https://uwspace.uwaterloo.ca/handle/10012/10179

[38] J. Verspecht, D. Gunyan, J. Horn, J. Xu, A. Cognata, and D. E. Root, "Multitone, Multi-port, and Dynamic Memory Enhancements to PHD Nonlinear Behavioral Models from Large-signal Measurements and Simulations," in *2007 IEEE/MTT-S International Microwave Symposium*, Jun. 2007, pp. 969–972.

[39] J. Verspecht, D. Root, and T. Nielsen, "Dynamic X-parameters: Behavioral modeling in the presence of long term memory effects," in *2012 The 7th German Microwave Conference*, Mar. 2012, pp. 1–4.

[40] D. E. Root, R. M. Biernacki, M. Marcu, M. Koh, and P. J. Tasker, "Frequency-scalable nonlinear behavioral transistor model from single frequency X-parameters based on time-reversal transformation properties (INVITED)," in *2015 86th ARFTG Microwave Measurement Conference*, Dec. 2015, pp. 1–7.

[41] F. M. Ghannouchi and M. S. Hashmi, *Load-Pull Techniques with Applications to Power Amplifier Design*. New York: Springer, Jun. 2012.

[42] A. Amini, "Analysis of power transistor behavioural modeling techniques suitable for narrow-band power amplifier design," Master's thesis, 2012. [Online]. Available: http://hdl.handle.net/10012/7038

[43] S. P. Woodington, R. S. Saini, D. Willams, J. Lees, J. Benedikt, and P. J. Tasker, "Behavioral model analysis of active harmonic load-pull measurements," in *2010 IEEE MTT-S International Microwave Symposium*, May 2010, pp. 1–1.

[44] J. Cai, J. B. King, B. M. Merrick, and T. J. Brazil, "Pad-Approximation-Based Behavioral Modeling," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4418–4427, Dec. 2013.

[45] B. Pichler, N. Leder, G. Magerl, and H. Arthaber, "On the dependability of load dependent X-parameter models for varying loads," in *2017 Integrated Nonlinear Microwave and Millimetre-wave Circuits Workshop (INMMiC)*, Apr. 2017, pp. 1–3.

[46] D. M. Pozar, *Microwave Engineering*. Wiley, Nov. 2011.

[47] P. Rodrigues, *Computer-Aided Analysis of Nonlinear Microwave Circuits*, 1st ed. Boston: Artech House Publishers, Jul. 1999.

[48] P. M. Chirlian, *Basic network theory*. McGraw-Hill, 1969.

[49] D. D. Weiner and J. F. Spina, *The Sinusoidal Analysis and Modeling of Weakly Nonlinear Circuits: With Application to Nonlinear Interference Effects*. Van Nostrand Reinhold, 1980.

[50] J. C. Pedro and N. B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Artech House, Dec. 2002.

[51] D. M. Rudolph, C. Fager, and D. E. Root, *Nonlinear Transistor Model Parameter Extraction Techniques*, Dec. 2011.

[52] R. Brady, C. Oxley, and T. Brazil, "An Improved Small-Signal Parameter-Extraction Algorithm for GaN HEMT Devices," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 7, pp. 1535–1544, Jul. 2008.

[53] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*. Boston: Artech House, Feb. 2003.

[54] J. Xu, M. C. E. Yagoub, R. Ding, and Q. J. Zhang, "Exact adjoint sensitivity analysis for neural-based microwave modeling and design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 1, pp. 226–237, Jan. 2003.

[55] J. Xu, D. Gunyan, M. Iwamoto, A. Cognata, and D. E. Root, "Measurement-Based Non-Quasi-Static Large-Signal FET Model Using Artificial Neural Networks," in *2006 IEEE MTT-S International Microwave Symposium Digest*, Jun. 2006, pp. 469–472.

[56] D. Root and B. Hughes, "Principles of Nonlinear Active Device Modeling for Circuit Simulation," in *ARFTG Conference Digest-Winter, 32nd*, vol. 14, Dec. 1988, pp. 1–24.

[57] I. Schmale and G. Kompa, "An improved physics-based nonquasi-static FET-model," in *Microwave Conference, 1997. 27th European*, vol. 1, Sep. 1997, pp. 328–330.

[58] A. Jarndal, B. Bunz, and G. Kompa, "Accurate Large-Signal Modeling of AlGaN-GaN HEMT Including Trapping and Self-Heating Induced Dispersion," in *IEEE International Symposium on Power Semiconductor Devices and IC's, 2006. ISPSD 2006*, Jun. 2006, pp. 1–4.

[59] A. Jarndal and G. Kompa, "Large-Signal Model for AlGaN/GaN HEMTs Accurately Predicts Trapping- and Self-Heating-Induced Dispersion and Intermodulation Distortion," *IEEE Transactions on Electron Devices*, vol. 54, no. 11, pp. 2830–2836, Nov. 2007.

[60] Y. Hu and S. Boumaiza, "Power-Scalable Wideband Linearization of Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 5, pp. 1456–1464, May 2016.

[61] M. Ben Rejeb, "Behavioral Modeling of Unmatched Nonlinear Devices Driven with Modulated Signal Stimuli using Volterra Series," Sep. 2014. [Online]. Available: https://uwspace.uwaterloo.ca/handle/10012/8855

[62] F. Giannini and G. Leuzzi, *Nonlinear Microwave Circuit Design*. John Wiley & Sons, Oct. 2004.

[63] F. M. Ghannouchi, O. Hammi, and M. Helaoui, *Behavioral Modeling and Predistortion of Wideband Wireless Transmitters*. Chichester, West Sussex, United Kingdom: Wiley, Jul. 2015.