Cyprus University of Technology

杭州電子科技大学
HANGZHOU DIANZI UNIVERSITY

Master's thesis

Behavioral Modeling Techniques for RF Devices Based on Physical TCAD Model

JUNWEI YE

Limassol, February 2024

MSc in Electronics Science and Technology

CYPRUS UNIVERSITY OF TECHNOLOGY

Faculty of Engineering and Technology

Department of Electrical Engineering, Computer Engineering, and Informatics

Master’s thesis

Behavioral Modeling Techniques for RF Devices Based on Physical TCAD Model

JUNWEI YE

Supervisor

Neophytos Lophitis

Limassol, February 2024

Approval Form

Master's thesis

Behavioral Modeling Techniques for RF Devices Based on Physical TCAD Model

Presented by

JUNWEI YE

Supervisor: Neophytos Lophitis

Member of the committee: Committee member 1

Member of the committee: Committee member 2

Cyprus University of Technology
Limassol, February 2024

Copyrights

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors for their invaluable guidance and continuous support throughout the course of this research.

I am deeply thankful to Professor Jiangtao Su at Hangzhou Dianzi University, my Chinese advisor, for his patient supervision, insightful suggestions, and academic rigor. His dedication to scientific research and generous mentorship have provided me with solid guidance and constant encouragement at every stage of my graduate studies.

I would also like to extend my heartfelt thanks to Professor Neophytos Lophitis at the Cyprus University of Technology, my international supervisor. During the joint training period, his expertise, international perspective, and constructive feedback greatly broadened my academic horizon and strengthened my research capacity. I truly appreciate his support in research discussions, academic writing, and intercultural collaboration.

I am grateful to both Hangzhou Dianzi University and the Cyprus University of Technology for offering me the opportunity to participate in this joint education program. This experience has enabled me to grow in a cross-cultural academic environment and to benefit from the strengths of both institutions in engineering education and scientific research.

My sincere appreciation also goes to my fellow labmates and colleagues, whose collaboration and companionship made my research journey enjoyable and productive.

Last but not least, I am profoundly thankful to my family for their unwavering love, patience, and encouragement. Their support has been the strongest pillar behind all my academic pursuits.

This thesis is a reflection of the collective efforts, insights, and kindness of all those who have supported me along the way. Thank you all.

ABSTRACT

Behavioral modeling of radio frequency (RF) devices has become an essential technique for accurately predicting device performance in complex operational scenarios. This paper presents a comprehensive methodology that integrates physical Technology Computer-Aided Design (TCAD) simulations, X-parameter extraction, and neural network-based prediction models. Initially, TCAD is employed to simulate the internal physical behavior of RF devices, providing a detailed understanding of carrier transport, thermal effects, and electric field distributions. Using these simulations, behavioral models are constructed, enabling the extraction of X-parameters, which effectively characterize the nonlinear dynamics and frequency-domain interactions of RF devices. To enhance predictive accuracy and computational efficiency, a neural network model is developed to learn the relationship between input signals and device responses based on the extracted X-parameters. The proposed approach demonstrates high fidelity in predicting device behavior under various operational conditions, offering a robust framework for RF device design and optimization. This study highlights the synergy between physical simulation, parameter extraction, and machine learning for advancing RF device modeling techniques.

Keywords: Behavioral modeling,TCAD simulation,neural networks,nonlinear modeling.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TCAD	Technology Computer-Aided Design
PHD model	Poly-harmonic distortion
RF	Radio Frequency

1 Introduction

With the rapid development of modern electronics, compound semiconductor materials such as gallium nitride (GaN), gallium arsenide (GaAs), and indium phosphide (InP) are increasingly being applied in fields such as power electronics, radio-frequency (RF) communications, and optoelectronic devices. Compared to traditional silicon-based semiconductors, compound semiconductor materials exhibit superior performance because of their unique electrical and thermal properties, such as better high-frequency characteristics, higher breakdown voltage, and lower conduction loss. However, designing and optimizing these devices pose significant challenges, particularly in complex operating conditions such as high frequencies and high power. Effective evaluation and prediction of the performance of these devices has become a critical technical issue in semiconductors [1].

In circuit design, the use of computer-aided design (CAD) software has become an essential step for simulation and analysis. These tools provide valuable insights prior to the actual fabrication of circuits, significantly reducing design costs and shortening development cycles. However, achieving this requires an accurate circuit model, as only a reliable model can offer meaningful guidance for design [2].

Traditionally, microwave modeling methods have been primarily categorized into two approaches: physics-based models and empirical models [3, 4]. Physics-based models involve a detailed understanding of the materials, structural parameters, and process parameters used in transistor devices to derive their voltage-current characteristics. However, due to the difficulty in accurately characterizing physical phenomena, the practical application of physics-based models in circuit design remains limited. In contrast, empirical models use fitted mathematical functions [5] to approximate device characteristics, sacrificing the physical interpretability of parameters in exchange for improved computational efficiency. Yet, this trade-off diminishes their utility in guiding device design.

Both physics-based and empirical models require precise measurements of S-parameters. However, in real-world electromagnetic environments, input and output signals often deviate from simple linear relationships. As the performance demands of communication systems increase, the limitations of traditional S-parameter models become particularly evident when devices operate under saturated conditions with high power, high efficiency, and nonlinear outputs. In such large-signal conditions, RF devices exhibit various nonlinear phenomena, including gain compression, harmonic distortion, intermodulation distortion, self-heating effects, and memory effects. To account for these effects, physics-based and empirical models require additional parameters, which increases model complexity and makes parameter extraction more challenging.

Behavioral modeling has emerged as a promising research direction in recent years [6]. As a black-box modeling approach, behavioral models do not require knowledge of the internal structure or equivalent circuit of a device. Instead, they rely solely on measured input and output signals to establish an equivalent model of the device's characteristics. By selecting an appropriate model structure and identifying model parameters based on port information, behavioral modeling creates a functional representation of the device. Since it avoids the need to understand the physical relationships among internal components, behavioral modeling helps protect intellectual property and prevent reverse engineering. Compared to physics-based and empirical models, behavioral models offer significant advantages, including reduced time costs, higher modeling efficiency, and sufficient accuracy. As a result, they have found widespread application in modeling RF front-end circuits and devices. This fast and efficient modeling approach addresses the growing need for nonlinear system modeling and facilitates further research and applications in RF engineering.

Traditional physical TCAD (Technology Computer-Aided Design) models can accurately simulate the internal physical behaviors of devices, such as carrier transport, temperature distribution, and electric field distribution. However, due to their high computational complexity and resource requirements, physical models often become inefficient for large-scale device integration and system-level simulations. Behavioral models, as a ”black-box” modeling approach, characterize the external behavior of devices, simplifying the modeling process, and improving simulation efficiency. They are widely used in integrated circuits, RF devices, and other fields. By combining physical TCAD models with behavioral models, it is possible to achieve efficient device and system-level simulations while maintaining model accuracy, thus providing more effective tools for device design and [7].

1.1 Aims and Objectives

Enhancing Modeling Efficiency and Accuracy through the Combination of Physical TCAD Models and Behavioral Models Physical TCAD models provide detailed and precise predictions of device performance by accurately describing internal physical processes, such as carrier diffusion and electric field distribution. However, due to their computational complexity, especially in multi-dimensional modeling, these models require substantial computational resources. Behavioral models, as a ”black-box” modeling approach, derive the external behavior of devices from test data and extract characteristic parameters, avoiding the need for complex physical simulations. By combining physical TCAD models and behavioral models, it is possible to retain the high accuracy of physical models while leveraging behavioral models to simplify simulation calculations, thereby improving overall modeling efficiency and accuracy.

Enhancing Flexibility and Scalability in Device Design Behavioral models provide a high-level abstraction that can represent the electrical behavior of devices across multiple dimensions, making them suitable for various operating environments. This simplified and efficient modeling approach enables designers to perform device design and system-level simulations more rapidly and easily extend these models to multi-device systems and complex circuits. This flexibility is particularly valuable for large-scale integrated circuits, RF systems, and other applications.

Reducing Development Time and Costs By combining physical TCAD models with behavioral models, device performance evaluation can be conducted rapidly during the design phase, avoiding the need for extensive physical prototype fabrication and testing inherent in traditional experimental methods. Behavioral models characterize external behavior based on test data, minimizing the focus on internal details. This allows for fast simulations at a lower computational cost, significantly shortening the development cycle and reducing overall costs during the development process.

Supporting Research and Applications of Novel Compound Semiconductor Devices and Systems With the emergence of novel compound semiconductor materials (e.g., gallium nitride [GaN], gallium arsenide [GaAs]) and innovative device structures, new design methods and simulation tools are critical to advancing technology. Combining physical TCAD models with behavioral models enables more accurate evaluation of these new materials and devices, providing stronger support for their development. In particular, in specialized applications such as high-frequency and high-power scenarios, behavioral models can quickly assess the electrical performance of devices without involving complex physical simulations, offering effective technical support for the design and application of novel devices.

Improving Simulation Convergence and Computational Efficiency A key advantage of behavioral models is their superior simulation convergence, especially in multi-dimensional modeling. Compared to physical TCAD models, behavioral models simplify the model structure, allowing for multiple simulations in a shorter time frame, helping designers quickly find optimal solutions. Therefore, the hybrid modeling approach combining behavioral models and physical TCAD models can greatly enhance simulation efficiency and convergence while maintaining accuracy.

1.2 Research Questions

The primary research questions of this thesis are summarized as follows:

1. **Effectiveness of Behavioral Models**
   - How can behavioral models accurately represent the external characteristics of RF devices while maintaining prediction accuracy comparable to physical TCAD models?
   - What are the applicability and limitations of behavioral models under multidimensional operating conditions, such as high frequency and high power?
2. **Integration of Physical and Behavioral Models**
   - What strategies can effectively integrate physical TCAD models with behavioral models to achieve both high accuracy and high efficiency in modeling?
   - How can the trade-offs between physical accuracy and computational efficiency be identified in different modeling scenarios?
3. **Optimization of Modeling Efficiency and Accuracy**
   - How can hybrid modeling approaches maximize computational efficiency while maintaining precision?
   - Which aspects of physical TCAD models can be substituted by behavioral models to reduce computational resource consumption?
4. **System-Level Applications and Scalability**
   - How scalable are behavioral models for multi-device systems or complex circuits?
   - How does the hybrid modeling framework perform in supporting the design and optimization of large-scale integrated circuits and RF systems in system-level simulations?
5. **Modeling Methods for Novel Materials and Structures**
   - How can the combined modeling framework remain adaptable to novel compound semiconductor materials (e.g., GaN, InP) and innovative device structures?
   - How can these new materials and devices' electrical performance be evaluated and optimized, particularly under extreme operating conditions such as high temperature and high voltage?

6. **Practical Value of the Modeling Framework**
    - What specific impacts does the hybrid modeling approach have in reducing the development cycle and cost in practical design workflows?
    - What are the specific demands and challenges for the modeling framework in different application scenarios, such as high-frequency communications and power electronics?

1.3 Contribution

Contributions

The primary contributions of this thesis are summarized as follows:

1. **Proposed a Hybrid Modeling Framework** An innovative hybrid modeling framework was proposed, integrating physical TCAD models and behavioral models to fully utilize the strengths of both approaches, achieving high accuracy and computational efficiency in RF device modeling.
2. **Improved Modeling Efficiency and Scalability** The hybrid modeling approach was demonstrated to significantly reduce the computational resources required for multi-dimensional RF device simulations while maintaining the accuracy of physical models. The framework also enhances scalability for multi-device systems and complex circuits.
3. **Optimized Device Performance Evaluation and System-Level Simulations** By combining the efficiency of behavioral models and the accuracy of physical models, a novel method was developed for rapid RF device performance evaluation, supporting system-level simulations in large-scale integrated circuits and RF systems.
4. **Supported Research and Applications of Novel Materials and Devices** A tailored hybrid modeling approach was developed for novel compound semiconductor materials (e.g., GaN, GaAs) and innovative device structures, providing technical support for their property evaluation and optimization.
5. **Reduced Development Cycle and Cost** The hybrid modeling framework significantly shortened the development cycle of RF devices through rapid performance evaluation and simplified simulation workflows, thereby reducing time and cost in the development process.
6. **Validated Practical Application Value of the Framework** The practical applicability of the hybrid modeling framework was validated through real-world case studies, demonstrating its potential in critical areas such as high-frequency communications and power electronics, providing valuable support for technological advancements in the industry.

1.4 Structure of the Thesis

This thesis is organized into five chapters, as outlined below:

This thesis consists of five main chapters that collectively address the research problem, methodology, and findings. The first chapter lays the groundwork by introducing the research background, problem, objectives, and contributions. It also provides an overview of the thesis structure.

The second chapter offers a comprehensive review of the relevant literature, highlighting the theoretical foundations of physical TCAD models and behavioral models. It identifies current limitations and challenges in the field, establishing a clear direction for the research.

The third chapter details the methodology employed in this study, focusing on the integration of physical TCAD and behavioral models to develop a hybrid modeling framework. This chapter also discusses the tools and techniques used for data collection and validation.

The fourth chapter presents the experimental results derived from the proposed hybrid modeling approach. It includes a comparative analysis of the hybrid model's performance against traditional methods and discusses its practical applications and implications.

Finally, the fifth chapter concludes the thesis by summarizing the key findings, emphasizing the contributions to both theory and practice, and offering recommendations for future research and potential extensions of the work.

1.5 Summary

This study advances innovation and development in semiconductor device modeling technology by combining physical TCAD models with behavioral models. This approach not only deepens the understanding of the internal behaviors of compound semiconductor devices but also provides more advanced tools for efficient device design. By leveraging this innovative modeling methodology, the accuracy and efficiency of semiconductor device design can be significantly improved, driving modeling technologies toward higher levels of efficiency and precision.

Behavioral models enhance the ability to evaluate and optimize device performance by precisely characterizing the external behaviors of devices. They allow for rapid performance evaluation and optimization without involving complex physical calculations, which is especially significant for compound semiconductor devices operating under high-frequency and high-power conditions. By integrating physical TCAD models, external characteristics remain highly consistent with the underlying physical processes, ensuring more reliable performance assessments.

With the continuous development of novel compound semiconductor materials, effectively evaluating and optimizing their performance has become a critical challenge for academia and industry alike. The combined modeling techniques of physical TCAD and behavioral models provide more accurate tools for the research and application of these new materials. This is particularly valuable in fields such as high-frequency communications, power electronics, and optoelectronics, where they can accelerate the development and deployment of novel devices.

In industries such as communications, energy, and power, the increasing demand for high-performance devices necessitates rapid responses to market needs and optimized device performance. By employing the combined modeling approach, the design and validation processes for high-performance devices can be greatly accelerated, enabling quicker market entry and fostering further technological innovation in the industry.

The integration of behavioral models also brings advantages to system-level design and integration, making it feasible to apply them to multi-device integrated circuits and complex systems. By combining physical TCAD models with behavioral models, comprehensive evaluations at the system level are possible, which helps improve the design efficiency and performance of integrated circuits and multi-device systems, further advancing the field of integrated circuits.

The combined modeling approach aims to improve the accuracy and efficiency of device design, promoting the application of compound semiconductor devices in high-frequency and high-power scenarios. This research provides more efficient tools for developing novel devices, theoretical support for semiconductor device design and optimization, and enhanced capabilities for system-level integration. Ultimately, it facilitates technological advancements across the semiconductor industry chain, offering robust support for designing high-performance electronic systems and promoting the widespread adoption of novel materials and devices.

2 Literature Review

2.1 Introduction

Behavioral modeling is a typical black-box modeling technique. In this method, the model is constructed without relying on detailed knowledge of the internal structure of the device [8]. Instead, it is based on analyzing the input and output data characteristics of the device. By collecting input-output data and using it to establish a black-box model, a mathematical model can be generated to characterize the behavior of the device under test. This method is particularly suitable for devices with unclear or unknown internal structures. Using this approach, it is possible to rapidly explore and analyze the external characteristics of the device without a deep understanding of its internal details, providing valuable references for device design and optimization.

The primary advantages of this method are its simplicity and ease of implementation. Even for devices lacking detailed physical information, an accurate model can still be constructed by applying input signals and measuring the corresponding output responses. This makes it an effective and widely used modeling technique for many unknown devices. Figure 2.1 illustrates the topology of the behavioral model. Although the internal structure of the model remains uncertain, it can still generate corresponding output data based on the applied input signals, thereby effectively characterizing the device's behavior.

In the behavioral modeling of power transistors, actual measurement data and simulation results are typically used to derive complete model parameters through appropriate mathematical formulas and methods. The accuracy of a power transistor's behavioral model depends on the precision of the measurement data. Therefore, the behavioral modeling of power transistors can be divided into two critical stages:

1. **Measurement Scheme Design:** First, a reasonable measurement scheme must be designed. This includes determining the port excitation conditions, load settings, control of external environmental factors, and ensuring the detailed accuracy of the test data. Ensuring the precision of the measurement data is crucial for providing high-quality foundational data for subsequent modeling.

2. **Mathematical Model Construction:** Second, a mathematical model matching the nonlinear behavior of the device needs to be established based on the measured data. Power transistors exhibit

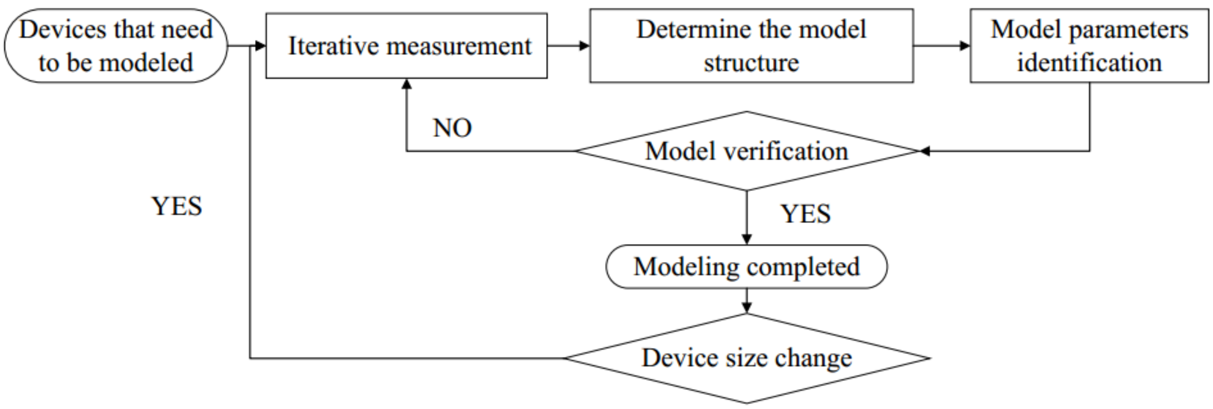

Figure 2.1: Flow chart of behavior model establishment

significant nonlinear characteristics. Researchers must analyze and process the measurement data to extract nonlinear parameters and incorporate them into the mathematical model to more accurately describe and predict the behavior of power transistors.

The close integration of these two stages is key to establishing an accurate behavioral model for power transistors. Only by ensuring the accuracy of the measurement data and combining it with an appropriate mathematical model can the operational characteristics of power transistors be deeply understood, thereby providing strong support for their optimized design.

2.2 S parameter

When a power transistor operates within its linear region, S-parameters can provide highly accurate descriptions of circuit characteristics [9]. S-parameters, or scattering parameters, are key parameters in microwave transmission and include S11, S21, S12, and S22. S11 represents the input reflection coefficient, which indicates the return loss at the input; S22 is the output reflection coefficient, reflecting the return loss at the output. S21 denotes the forward transmission coefficient, often used to describe gain, while S12 represents the reverse transmission coefficient, reflecting isolation. These parameters help accurately analyze the performance of power transistors and their associated circuits.In this case, the S-parameters of the a and b waves are expressed as follows: see Equation 2.1.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \tag{2.1}$$

2.3 Hot-S Parameters

As the input signal increases, power transistors exhibit nonlinear characteristics and memory effects under real operating conditions, making S-parameters insufficient for accurately describing their behavior [10]. To address this limitation, the concept of hot-S parameters was proposed, building upon traditional S-parameters. Unlike traditional S-parameters, which are measured under single-source excitation conditions, hot-S parameters incorporate dual-source excitation. A large signal is used to activate the device's operating state, followed by the measurement of the response to a small signal [11, 12]. This method enables hot-S parameters to better describe the behavior of nonlinear devices while maintaining some similarities to traditional S parameters.Hot-S parameter are expressed as follows:see Equation 2.2.

$$\begin{bmatrix} b_1(f_s) \\ b_2(f_s) \end{bmatrix} = \begin{bmatrix} hotS_{11} & hotS_{12} \\ hotS_{21} & hotS_{22} \end{bmatrix} \begin{bmatrix} a_1(f_s) \\ a_2(f_s) \end{bmatrix}. \tag{2.2}$$

However, this simplified expansion method only considers the effect of the large-signal input f_c and does not fully account for the intermodulation products between the large signal f_c and the small signal f_s. To refine the model, an additional small signal with the same frequency as the large driving signal is introduced at the other end of the device. Since the power of the input small signal is much lower than that of the driving large signal, we can treat this small signal as a perturbation signal without altering the device's operating state. This assumption holds under the condition that the higher-order harmonic responses at each port are neglected, and only the fundamental frequency responses are retained. Under these conditions, the Hot-S parameters can be expressed as:see Equation 2.3.

$$\begin{bmatrix} b_1(f_s) \\ b_2(f_s) \end{bmatrix} = \begin{bmatrix} hotS_{11} & hotS_{12} \\ hotS_{21} & hotS_{22} \end{bmatrix} \times \begin{bmatrix} a_1(f_c) \\ a_2(f_c) \end{bmatrix} + \begin{bmatrix} T_{12} \\ T_{22} \end{bmatrix} e^{j2\varphi(a_1(f_c))} \text{conj}(a_2(f_c)). \tag{2.3}$$

In the above equation, the introduction of T_{12} and T_{22} accounts for the nonlinear response of the perturbation signal and its conjugate component near the large-signal operating point. This approach partially linearizes the nonlinear behavior of the system. However, it is important to note that since the equation only considers the fundamental responses, it cannot fully capture the influence of the nonlinear intermodulation products.

2.4 PHD model

As an advancement of the extended thermal S-parameters, Verspecht et al. introduced a novel large-signal scattering parameter technique, referred to as the poly-harmonic distortion (PHD) model [13, 14]. The PHD model is rigorously derived through mathematical analysis based on the description function.

The description function characterizes a time-invariant system, where any delay in the incident wave corresponds to an identical delay in the scattered wave. This behavior manifests as a linear phase shift within the frequency domain. In this framework, the phase of the fundamental excitation signal serves as the reference, and the phase operator P is defined as:

$$P = e^{j\varphi_{A_{11}}}. \tag{2.4}$$

Here, the separation of phase and amplitude becomes feasible, ultimately leading to Formula as fellow:

$$B_{pm} = F_{pm}\left(|A_{11}|, A_{12}P^{-2}, A_{13}P^{-3}, \ldots, \times A_{21}P^{-1}, A_{22}P^{-2}, \ldots\right) P^{p+m}. \tag{2.5}$$

After performing this transformation, the primary independent variable A_{11} is simplified from a complex number to a positive real value, significantly reducing the complexity of subsequent mathematical operations (as shown in Figure 2.2).

Assuming that for a device under test (DUT), the fundamental input signal A_{11} represents a large-signal excitation, while other harmonic components act as small-signal excitations, the operating point of the DUT is determined solely by A_{11}. Under this condition, the harmonic superposition principle (as illustrated in Figure 2) can be applied to linearize these small harmonic signals. Here, the mutual influence of small signals and their conjugate components is negligible, and their output effects appear in a positively correlated form. Consequently, the final PHD model can be described as follows:

$$B_{pm} = \sum_{qn} S_{pq,mn}\left(|A_{11}|\right) P^{p+m-n} A_{qn} + \sum_{qn} T_{pq,mn}\left(|A_{11}|\right) P^{p+m+n} \text{conj}(A_{qn}). \tag{2.6}$$

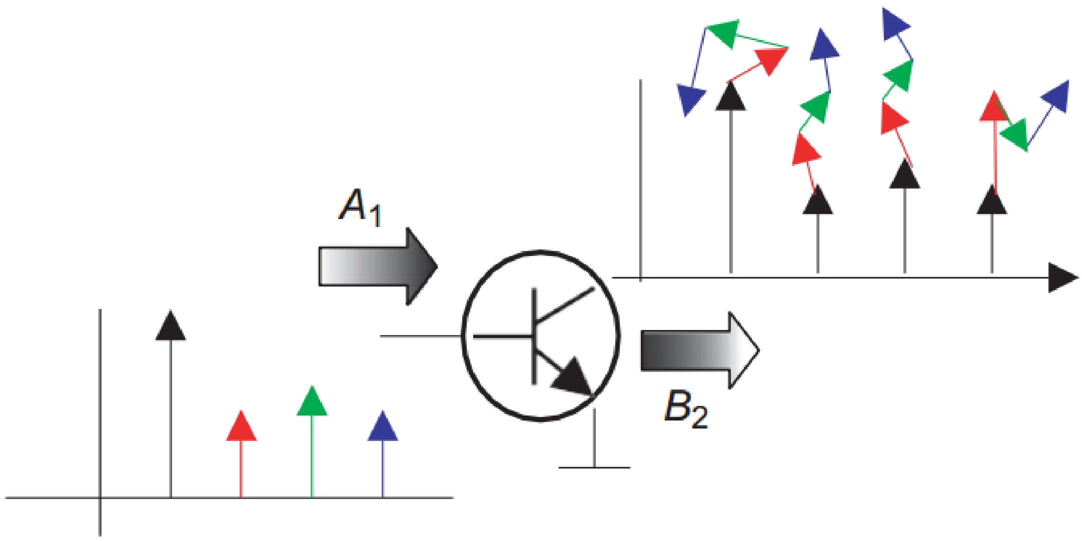

Figure 2.2: The harmonic superposition principle

2.5 X parameter

In order to enhance and commercialize the model, Jan and Agilent jointly filed a patent for the X-parameter behavioral model. The description formula for this model is given as:

$$B_{ef} = X_{ef}^{(F)}(|A_{11}|)P^f + \sum_{gh\neq 11} X_{ef,gh}^{(S)}(|A_{11}|)A_{gh}P^{f-h} + \sum_{gh\neq 11} X_{ef,gh}^{(T)}(|A_{11}|)A_{gh}^{*}P^{f+h}. \tag{2.7}$$

In Equation 2.7, the first term solely depends on the large-signal operating point. At this stage, only the large-signal excitation is applied to the DUT, while the amplitude and phase data at all relevant frequency points are captured. By comparing these measurements to the amplitude and phase of the excitation signal, the amplitude ratio and phase shift across different frequencies can be determined to extract the model parameters.

The X_S and X_T terms describe the effects of the perturbation small signal and its conjugate component on the port's scattering waves, respectively. While the reflection coefficient is not explicitly modeled, the nonlinear effects caused by load impedance mismatches are incorporated in the form of small signal disturbances. As illustrated in Figure 2.3), it can be observed that as A_{11} increases, system nonlinearities become more pronounced. This effect leads to greater compression and distortion, influencing system behavior, including rotation and scaling factors.

According to the preceding derivation, the X-parameter's large-signal operating point is determined entirely by the device's nonlinear state. This includes both the input signal and the DC bias conditions of the large signal A_{11} (expressed as $|A_{11}|, DC$). Since the model derivation assumes a single large-signal excitation, the X-parameter model is best suited for devices with proper impedance matching. However, in cases where significant load impedance mismatches occur, causing the DUT to operate under strong nonlinear conditions, the X-parameter model may exhibit substantial prediction errors.

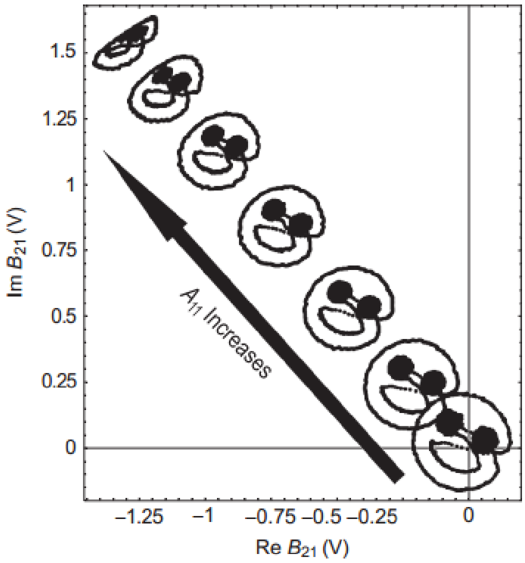

Figure 2.3: The increase of the incident wave distorts the smiling face

2.6 Cardiff Model

The development of the X-parameter model led to the creation of the Cardiff model, which introduces a new modeling strategy. This approach combines a truth look-up table with a polynomial-based behavioral model [15, 16]. By integrating measurement-based and model-based methods, the Cardiff model facilitates the conversion of measured voltage and current waveforms into model parameters, allowing for smooth incorporation into CAD software.

The Cardiff model extends the classical X-parameter model, and its formulation is expressed as:

$$b_k = \sum_{m=0}^{n-1} C_{k,m} \left(\frac{Q}{P}\right)^m a_1 + \sum_{m=0}^{n-1} U_{k,m} \left(\frac{P}{Q}\right)^m a_2. \tag{2.8}$$

Compared to load-dependent X-parameters, the Cardiff model accounts for the amplitude $|A_{21}|$ of the incident wave at the output and separates the amplitude and phase of the large-signal incident wave into two distinct components [17, 18]. Here, the ratio Q/P is introduced as an independent variable to represent the phase difference. Model parameters C and U are determined by integrating and rotating phase operators of various orders, enabling the model to cover a broader range on the Smith chart. This expansion significantly improves the prediction range by employing interpolation and extrapolation techniques.

However, one limitation of the Cardiff model is its inability to capture high-order harmonics of the reflected wave. As a result, it is primarily suited for simulations involving fundamental wave outputs, such as AM-AM and AM-PM distortions or fundamental wave load-pull. In strongly nonlinear conditions, this model may exhibit increased prediction errors. To address this, the single-input Cardiff model has been extended to handle multi-tone and two-tone excitations, facilitating the prediction of mixed-order intermodulation products. The Cardiff model is also applicable to multi-input single-output (MISO) power amplifiers. In such cases, where multiple excitation signals with varying amplitudes and phases interact, the model can be expressed as:

$$B_{p,h} = (\angle A_{1,1})^h \cdot \sum_m \sum_n \sum_x K_{p,h,m,n,x} \cdot |A_{1,1}|^x \cdot |A_{2,1}|^m \cdot \left(\frac{\angle A_{2,1}}{\angle A_{1,1}} \right)^n . \tag{2.9}$$

In this formulation, subscripts x and m represent the amplitude variation ranges of the two signals, while n denotes the range of phase variations. To ensure that each harmonic aligns with the fundamental signal phase, the Cardiff model normalizes all phases to the fundamental phase of $A_{1,1}$. The coefficients are separated into amplitude-dependent components, eliminating phase dependence and simplifying the mathematical representation. This normalization enhances the time invariance of the model and reduces computational complexity.

Experimental validation has demonstrated the Cardiff model's excellent interpolation capabilities and accuracy while significantly reducing the size of the required dataset. However, the re-normalization of traveling waves adds complexity to parameter extraction. To further enhance the model's versatility, DC bias has been incorporated into the Cardiff model [19], improving its adaptability and reducing testing overhead.

2.7 Summary

This chapter presented advanced behavioral modeling techniques for RF devices, starting with the black-box **behavioral modeling** approach, which characterizes device behavior based on input-output relationships. The limitations of traditional **S-parameters** in nonlinear conditions were addressed with **Hot-S parameters**, which incorporate large-signal excitation and perturbation signals. The **PHD model** further extended this concept by separating amplitude and phase components to describe harmonic effects. The **X-parameter model** was introduced as a commercialized approach for large-signal nonlinear modeling, while the **Cardiff model** expanded X-parameters with polynomial behavior modeling and truth look-up tables, improving prediction accuracy for multi-tone and mixed-order systems. These techniques collectively enhance the accuracy and efficiency of RF device characterization and design.

3 Research Methodology

The previous chapter reviewed commonly used behavioral modeling methods and their development in the context of RF device modeling, with a particular focus on advanced techniques such as S-parameters, Hot-S parameters, PHD models, X-parameter models, and Cardiff models. These models have significantly improved the accuracy and efficiency of characterizing nonlinear device behavior. To further explore the modeling mechanisms and practical applications of power amplifiers, this chapter introduces the fundamental principles of power amplifiers and X-parameters, along with their relevance in large-signal nonlinear modeling. It systematically presents the key steps involved in model extraction, including excitation configuration, data acquisition, and parameter fitting, and further discusses the implementation and validation of the models in simulation environments. This provides essential technical support for subsequent device performance analysis and optimization design.

3.1 Introduction to the Fundamental Theory of Power Amplifiers

With the continuous advancement of semiconductor technology and the widespread application of electronic devices, power amplifiers (PAs) in wireless communication systems are evolving toward more systematic and industrialized technological development. Currently, PA research covers multiple aspects, including the application of various types of transistors and system-level studies tailored for different application scenarios [20]. This field encompasses devices, modeling, circuits, and testing. This section introduces the classification of microwave transistor devices and the fundamental performance metrics of power amplifiers.

3.1.1 Classification of Transistors

Microwave transistors serve as the core components in high-power amplifier (PA) systems, and their technical characteristics have a direct impact on the overall PA performance. Therefore, a solid understanding of microwave transistor fundamentals is essential for anyone involved in device modeling.

Microwave transistors are generally categorized based on their structure and operating principles into two main types: junction transistors and field-effect transistors (FETs). Junction transistors include bipolar junction transistors (BJTs), typically fabricated with a single semiconductor material such as silicon, and heterojunction bipolar transistors (HBTs), which use compound semiconductors [21]. BJTs, also known as triodes, are made up of three regions (NPN or PNP), namely the emitter, base, and collector. They are commonly used in the lower frequency range (0.1 GHz to 4 GHz) due to their low cost and adequate performance. HBTs, on the other hand, are built with heterojunctions formed from different semiconductor materials, such as InP or AlGaAs/GaAs. Compared to BJTs, HBTs offer higher emission efficiency and faster carrier transport, allowing them to operate at frequencies exceeding 100 GHz in some cases [21].

Field-effect transistors are unipolar devices in which current is carried by a single type of charge carrier—electrons in n-channel FETs and holes in p-channel FETs. While BJTs are current-controlled devices, FETs are controlled by voltage. Structurally, a FET consists of three terminals: gate, source, and drain.

The current between the source and drain is modulated by the gate voltage, acting similarly to a voltage-dependent resistor. Various types of FETs include MESFETs [22, 23] (metal-semiconductor FETs), MOSFETs [24] (metal-oxide-semiconductor FETs), HEMTs [25] (high electron mobility transistors), and PHEMTs [26] (pseudomorphic HEMTs).

Among these, HEMTs have seen rapid development in China and are considered a leading technology in the FET domain. In particular, GaN-based HEMTs are regarded as highly promising for RF power applications due to their excellent high-frequency, high-power, and high-efficiency performance. These devices also offer substantial design flexibility, making them well-suited for modern wireless communications, radar systems, and RF power amplifiers. However, challenges remain: characteristics such as self-heating, floating-body effects, and trap-related phenomena—stemming from the wide bandgap properties of GaN—can significantly impact device behavior and pose difficulties for accurate modeling .

3.1.2 Fundamental Performance Metrics

In radio frequency (RF) systems, the power amplifier serves as the final stage in the transmitter chain, playing an essential role in boosting the power of modulated signals. Its performance has a direct impact on the effectiveness and reliability of systems such as wireless communications, radar, and satellite links. To meet the demands of coverage, efficiency, and signal fidelity, understanding the core performance parameters of RF power amplifiers is critical.

During the design process, several key metrics require particular attention—most notably power gain, the 1 dB compression point, and power-added efficiency (PAE). These indicators influence not only the amplifier's ability to deliver sufficient output power over distance but also the overall efficiency of the system, its ability to preserve signal quality, and its potential to minimize interference with adjacent signals. A precise evaluation and optimization of these parameters enable designers to develop RF circuits that support robust, energy-efficient, and high-performance signal transmission [45].

Transmission gain and power gain are defined as follows:

Assume the input power is P_{in}, the available power from the source is P_{av}, and the output power is P_{out}. Then, the transmission gain G_t and the power gain G_p are defined respectively as:

$$G_t = \frac{P_{\text{out}}}{P_{\text{av}}} \tag{3.1}$$

$$G_p = \frac{P_{\text{out}}}{P_{\text{in}}} \tag{3.2}$$

The following describes the 1 dB compression point and the third-order intercept point (IP3):

Power amplifiers exhibit different performance characteristics under varying input signal levels. When operating under small-signal conditions, the gain remains relatively constant and linear. However, as the input signal strength increases, nonlinearities between the input and output begin to emerge, causing the gain to gradually decrease and eventually leading to output power saturation.

A critical parameter used to describe this nonlinear behavior is the **1 dB compression point** ($P_{1\text{dB}}$), which is defined as the output power level at which the gain drops by 1 dB from its ideal linear value. It is commonly used to evaluate the linearity of power amplifiers and serves as an essential indicator of system signal fidelity. The 1 dB compression point can be mathematically expressed as:

$$P_{1\mathrm{dB}} = P_{\text{out, uncompressed}} - 1\ \mathrm{dB} \tag{3.3}$$

where $P_{1\mathrm{dB}}$ is the output power at the compression point, and $P_{\text{out, uncompressed}}$ is the ideal output power without gain compression.

Another important indicator of linearity is the **Third-Order Intercept Point (IP3)**. It is derived from the intersection point between the extrapolated fundamental signal output line and the extrapolated third-order intermodulation product line. The IP3 is widely used to assess the degree of intermodulation distortion in power amplifiers. It represents the hypothetical output power level where the third-order products would reach the same amplitude as the fundamental output signal, assuming both continue increasing linearly.

A larger distance between the IP3 and the 1 dB compression point (and the saturation point) indicates better amplifier linearity, as it reflects a wider dynamic range before significant nonlinear distortion occurs, as illustrated in Figure 3.1.

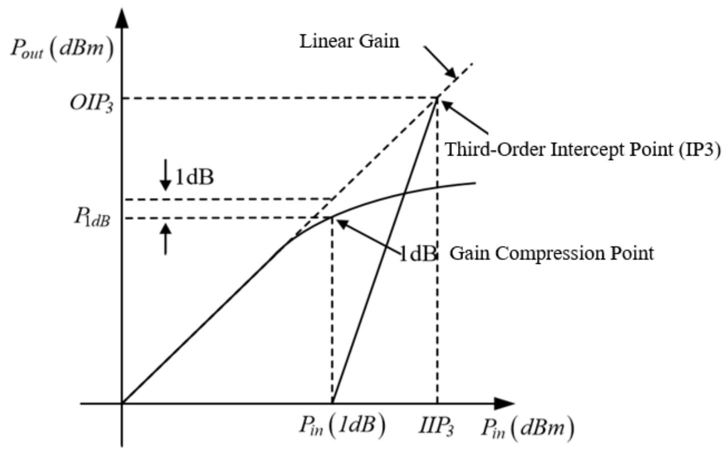

Figure 3.1: Input–Output Characteristic Curve of the Power Amplifier

The **power-added efficiency (PAE)** and **drain efficiency (DE)** are two important metrics used to evaluate the performance of power amplifiers in terms of power conversion. These indicators reflect how effectively a PA transforms the supplied DC power into usable RF output power, considering inevitable losses during the conversion process.

The two efficiency metrics are defined as follows:

$$\mathrm{PAE} = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{DC}}} \times 100\% \tag{3.4}$$

$$\mathrm{DE} = \frac{P_{\text{out}}}{P_{\text{DC}}} \times 100\% \tag{3.5}$$

Here, P_{out} denotes the output RF power, P_{in} is the input RF power, and P_{DC} represents the supplied DC power.

PAE accounts for both the output and input RF powers, offering a broader view of the amplifier's effective power contribution. In contrast, DE only considers the direct conversion from DC to RF output. A higher efficiency in either metric indicates better utilization of the supplied power, helping reduce energy losses and improving the overall energy performance of the amplifier.

3.2 Introduction of X-Parameter Theory

A key challenge in nonlinear systems lies in handling frequency interactions, where the output at any frequency depends on all input frequency components, violating the principle of superposition. To address this complexity, we introduced a separation strategy that decomposes the input into large-signal and small-signal components. The large-signal portion defines the device's dynamic operating point, while small signals are treated as perturbations around this state.

To simplify the characterization of nonlinear mappings, we compared static large-signal linearization with dynamic large-signal linearization. A notable finding was that the small-signal output in dynamic conditions depends not only on the small-signal input but also on its complex conjugate. This insight led to a method for linearizing nonlinear scattering mappings under dynamic operating conditions.

The discussion then expanded on generalizing scattering parameters to accommodate single-tone sinusoidal inputs, culminating in the formulation of single large-tone X-parameters. We introduced the concept of pseudowaves and demonstrated their role in capturing cross-frequency phase and time-invariance properties. By using the phase of the largest input pseudowave as a reference, we simplified the nonlinear mapping process.

To make the X-parameter more practical, we separated the large-signal input pseudowaves from small-signal components and linearized the small-signal behavior around the large-signal operating point. This approach reduced the computational complexity of characterizing the X-parameters, resulting in a manageable single-tone formulation. Finally, we discussed the procedure for extracting X-parameters from experimental measurements, using a canonical nonlinear mapping example to illustrate the process.

3.2.1 Linear Time-Invariant Electrical Networks

Network analysis is a well-established field with abundant references, including [27–29]. Here, we provide a concise discussion on the essential concepts relevant to this study. A two-port network example is depicted in Fig. 3.2. In such systems, a linear time-invariant electrical network is often defined by relating the current and voltage at its ports. If we take current as the independent variable, we can express the system in terms of the impedance matrix **Z**:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}. \tag{3.6}$$

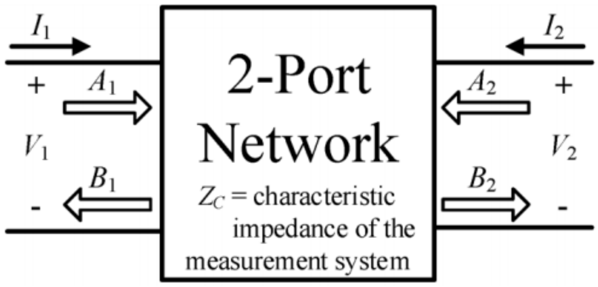

Figure 3.2: Two-port electrical network

The elements Z_{ba} of the impedance matrix are determined as follows:

$$Z_{ba} = \left.\frac{V_b}{I_a}\right|_{I_c=0,\, c\neq a}. \tag{3.7}$$

Here, the current through all ports, except the one being analyzed, is set to zero. This simplification eliminates extraneous terms, reducing the equation to a form where Z_{ba} can be easily determined.

Alternatively, when voltage is treated as the independent variable, the admittance matrix **Y** is defined:

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{3.8}$$

Each admittance term Y_{ba} is calculated by:

$$Y_{ba} = \left.\frac{I_b}{V_a}\right|_{V_c=0,\, c\neq a}. \tag{3.9}$$

Both impedance and admittance matrices face challenges at high microwave frequencies. Specifically, direct measurement becomes impractical due to the inability to terminate ports accurately with shorts or opens. To address this, scattering parameters (**S-parameters**) are introduced, which provide a robust way to describe network behavior at microwave frequencies.

Scattering parameters relate incident and reflected waves at the network ports rather than voltages and currents. These parameters are defined as:

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \tag{3.10}$$

where A represents the incident wave, and B corresponds to the reflected wave.

The scattering matrix provides a complete description of network behavior, similar to the impedance and admittance matrices. However, unlike the latter, scattering parameters are particularly suitable for microwave applications as they work with traveling waves. These parameters can be measured directly using a vector network analyzer (VNA), which simplifies the calibration and analysis process. Further details on VNA operation and calibration are available in [30].

Lastly, the relationships between traveling waves, impedance parameters, and admittance parameters can be understood as linear transformations. This makes scattering parameters an invaluable tool for the characterization of high-frequency networks.

3.2.2 Linearization of Nonlinear Mappings

In this section, we shift our focus to nonlinear scattering mappings and the commonly employed linearization techniques used to simplify their complexity. Nonlinear systems are often linearized around a dynamic operating point rather than a static one. This approach introduces a frequency-dependent interaction between the large-signal and small-signal components. Specifically, positive and negative frequency components of the small signal are handled differently, resulting in a nonlinear mapping that is inherently nonanalytic. Consequently, the output depends on both the small signal's complex phasor and its conjugate component, which we will illustrate further.

Scattering parameters (*S-parameters*) are valid primarily for small-signal, linear, and time-invariant systems, or those that can approximate such conditions. However, in a nonlinear system, the output Y can be expressed as a nonlinear function f of the input X:

$$Y = f(X). \tag{3.11}$$

The relationship described in (2.7) is often complex and challenging to analyze. To simplify, the input X is decomposed into two parts: a static term X_0 and a small time-varying perturbation x, such that $X = X_0 + x$. This decomposition allows the output Y to be separated into:

$$Y = Y_0 + y, \tag{3.12}$$

where $Y_0 = f(X_0)$ represents the response to the static term, and y denotes the perturbation caused by x.

To approximate y, we expand Y using a Taylor series around X_0:

$$Y = f(X_0) + x\frac{df}{dX}\bigg|_{X=X_0} + \frac{1}{2}x^2\frac{d^2f}{dX^2}\bigg|_{X=X_0} + \cdots + \frac{1}{n!}x^n\frac{d^nf}{dX^n}\bigg|_{X=X_0}. \tag{3.13}$$

From this expansion, the perturbation y can be obtained as:

$$y = Y - Y_0 = k_1x + k_2x^2 + k_3x^3 + \ldots, \tag{3.14}$$

where k_i are the coefficients derived as:

$$k_i = \frac{1}{i!}\frac{d^if}{dx^i}\bigg|_{X=X_0}. \tag{3.15}$$

This linearization simplifies the characterization of nonlinear mappings by expressing the system's response as a summation of perturbation terms. Each term in the series corresponds to a specific order of nonlinearity, providing a practical means to analyze and approximate the output behavior caused by small-signal variations.

3.2.3 Properties of Weakly Nonlinear Systems with Large-Signal Bias

In this section, we examine the characteristics of weakly nonlinear time-invariant systems. These devices are defined by their stable, single-valued, and continuous output responses to input signals, specifically when operating around a large-signal bias point. An example of such a device is shown in Equation (3.16), where the output Y is modeled as a cubic polynomial function of the input X:

$$Y = f(X) = k_1 X + k_2 X^2 + k_3 X^3. \tag{3.16}$$

When we apply the input signal $X(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t)$, where $\omega_i = 2\pi f_i$, to Equation (3.16), the output $Y(t)$ consists of spectral components at frequencies that are linear combinations of f_1 and f_2, as shown in Equation (3.17):

$$f_{\text{out}} = n f_1 + m f_2, \tag{3.17}$$

where n and m are integers. These output frequency components include harmonics of f_1 and f_2, as well as intermodulation products where $n \neq 0$ and $m \neq 0$.

To illustrate, consider an input signal composed of a direct current (DC) component $X_0(t)$ and a small signal $x(t)$:

$$X_0(t) = A_0, \tag{3.18}$$

$$x(t) = \frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2} = |\delta| \cos(\omega t + \arg(\delta)), \tag{3.19}$$

where A_0 is a real constant and δ is a small complex phasor. Using this input, we linearize the response $y(t)$ about the operating point $X_0(t) = A_0$.This situation can be seen in Figure 3.3, as shown in Equation (3.20):

$$y(t) = Y(t) - Y_0(t) \approx f'(X_0(t))x(t), \tag{3.20}$$

where $f'(A_0)$, the first derivative of $f(X)$ evaluated at $X = A_0$, is given by:

$$f'(A_0) = k_1 + 2k_2 A_0 + 3k_3 A_0^2. \tag{3.21}$$

Substituting into Equation (3.20), we obtain:

$$y(t) = \left(k_1 + 2k_2 A_0 + 3k_3 A_0^2\right) \frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2}. \tag{3.22}$$

The Fourier coefficient of the output at frequency ω, denoted as $\hat{y}(\omega)$, is derived as:

$$\hat{y}(\omega) = \left(\frac{k_1 + 2k_2 A_0 + 3k_3 A_0^2}{2}\right) \delta. \tag{3.23}$$

Equation (3.23) highlights the linear dependence of $\hat{y}(\omega)$ on the small-signal amplitude and phase (δ) while incorporating the nonlinear relationship with the DC operating point A_0. This demonstrates the interplay between the large-signal bias and small-signal dynamics in weakly nonlinear systems.

Our next example considers a large,see Figure 3.4, periodically time-varying signal added to $X_0(t)$. This

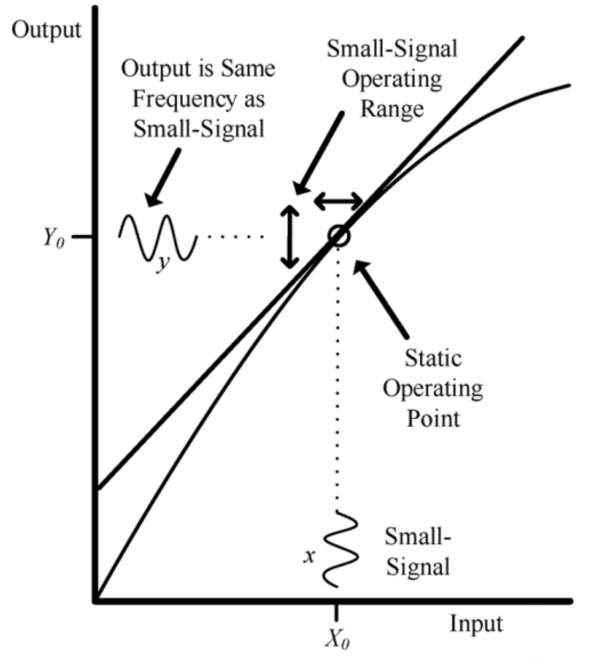

Figure 3.3: There is alinear relationship between the small-signal input and the small-signal output

signal can be expressed as:

$$X_0(t) = A_0 + A_1 \cos(\omega t), \tag{3.24}$$

and the small signal as:

$$x(t) = \frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2} = |\delta| \cos(\omega t + \arg(\delta)). \tag{3.25}$$

At this point, the derivative of $f(X_0(t))$ with respect to X_0 is evaluated as:

$$f'(X_0(t)) = k_1 + 2k_2 \left(A_0 + A_1 \cos(\omega t)\right) + 3k_3 \left(A_0 + A_1 \cos(\omega t)\right)^2. \tag{3.26}$$

Using trigonometric identities, we expand:

$$f'(A_0 + A_1 \cos(\omega t)) = \left(k_1 + 2k_2 A_0 + 3k_3 A_0^2\right) + \left(2k_2 A_1 + 6k_3 A_0 A_1\right) \cos(\omega t) + \frac{3}{2} k_3 A_1^2 \cos(2\omega t). \tag{3.27}$$

Substituting this result into the linearized response, we have:

$$y(t) = \left[\left(k_1 + 2k_2 A_0 + 3k_3 A_0^2\right) + \left(2k_2 A_1 + 6k_3 A_0 A_1\right) \cos(\omega t) + \frac{3}{2} k_3 A_1^2 \cos(2\omega t)\right] \frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2}. \tag{3.28}$$

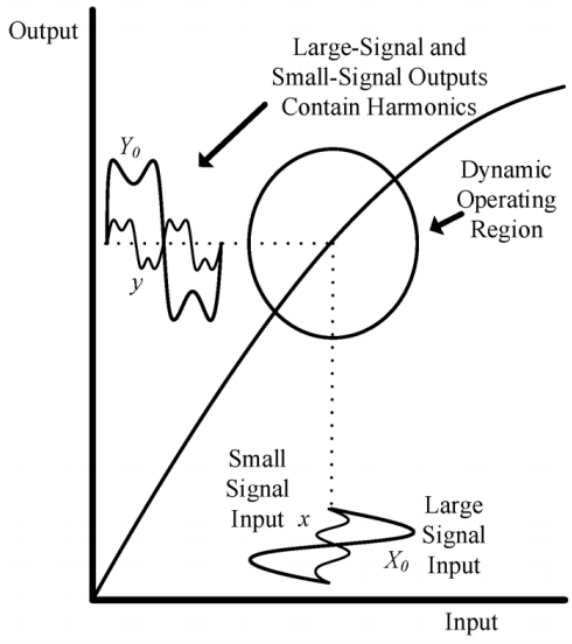

Figure 3.4: The large-signal operating point, the corresponding output of the small-signal and large-signal inputs

Grouping terms by frequency components, we find:

$$\begin{aligned} y(t) = {} & [\beta\delta + \beta^*\delta^*] + [\alpha\delta + \gamma^*\delta^*]\, e^{j\omega t} + [\gamma\delta + \alpha^*\delta^*]\, e^{-j\omega t} \\ & + \beta\delta e^{j2\omega t} + \beta^*\delta^* e^{-j2\omega t} + \gamma\delta e^{j3\omega t} + \gamma^*\delta^* e^{-j3\omega t}, \end{aligned} \tag{3.29}$$

where the coefficients are defined as:

$$\alpha = \frac{1}{2}\left(k_1 + 2k_2A_0 + \frac{3}{2}k_3A_0^2 + 3k_3A_0A_1 + \frac{3}{4}k_3A_1^2\right), \tag{3.30}$$

$$\beta = \frac{1}{2}\left(2k_2A_1 + 6k_3A_0A_1 + 3k_3A_1^2\right), \tag{3.31}$$

$$\gamma = \frac{3}{8}k_3A_1^2. \tag{3.32}$$

The Fourier coefficient of the output at frequency ω is:

$$\hat{y}(\omega) = \left(\frac{1}{2}k_1 + k_2A_0 + \frac{3}{2}k_3A_0^2 + \frac{3}{4}k_3A_1^2\right)\delta + \frac{3}{8}k_3A_1^2\delta^*. \tag{3.33}$$

When $A_1 \to 0$, the periodically time-varying signal vanishes, simplifying the analysis. Contributions from higher harmonics at ω result from nonlinearities in the system. This analysis demonstrates that a linearization around a large-signal dynamic operating point comprises mappings of both the large-signal and small-signal components.

3.2.4 Nonlinear Scattering Mappings

Having established the linearization of a general large-signal scattering mapping, we revisit the scattering parameters introduced in Section 3.1, restating them as:

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{3.34}$$

$$b_2 = S_{21}a_1 + S_{22}a_2. \tag{3.35}$$

Given the linear nature of this system, the frequency components in the incident and scattered waves can be independently analyzed and separated. Consequently, a_1 and a_2 can be assumed to represent spectral content at a single frequency. Similarly, the same holds for b_1 and b_2. Extending this notion to N ports, equations (2.35) and (2.36) become generalized as:

$$b_1 = F_1(a_1, a_2, \ldots, a_N) \tag{3.36}$$

$$b_2 = F_2(a_1, a_2, \ldots, a_N) \tag{3.37}$$

$$\vdots$$

$$b_N = F_N(a_1, a_2, \ldots, a_N). \tag{3.38}$$

Here, $F_1, F_2, \ldots, F_N$ represent nonlinear, time-invariant mappings of the incident waves at each of the N ports. By relaxing the linear constraints, $a_1, a_2, \ldots, a_N, b_1, b_2, b_N$ are no longer modeled solely as single sinusoids, as described in Section 3.3. For simplicity, we assume the incident wave at port 1 is a sinusoidal signal with frequency f_1, expressed as:

$$a_1(t) = |A_{1,1}| \cos(2\pi f_1 t + \arg(A_{1,1})) = \Re\{A_{1,1} \cdot e^{j2\pi f_1 t}\}, \tag{3.39}$$

where $A_{1,1}$ is a complex coefficient [12]. Thus, all frequencies in the system will be represented as $f_k = kf_1$, where k is a non-negative integer. The incident wave at port q and the scattered wave at port p can then be expressed as:

$$a_q(t) = \sum_{l=1}^{K} |A_{q,l}| \cos(2\pi l f_1 t + \arg(A_{q,l})) \tag{3.40}$$

$$b_p(t) = \sum_{k=1}^{K} |B_{p,k}| \cos(2\pi k f_1 t + \arg(B_{p,k})). \tag{3.41}$$

Here, K represents the total number of harmonics considered relevant in the system. The Fourier coefficients $A_{q,l}$ and $B_{p,k}$ correspond to the l-th harmonic of port q and the k-th harmonic of port p, respectively. Substituting (2.37) into its individual Fourier components, we obtain:

$$B_{p,k} = F_{p,k}(A_{1,1}, A_{1,2}, \ldots, A_{1,K}, A_{2,1}, A_{2,2}, \ldots, A_{2,K}, \ldots, A_{N,1}, A_{N,2}, \ldots, A_{N,K}), \tag{3.42}$$

for $p = 1, \dots, N$ and $k = 1, \dots, K$. The coefficient $B_{p,k}$ is often termed the scattered pseudowave at the k-th harmonic of port p, while $A_{q,l}$ represents the incident pseudowave at the l-th harmonic of port q. It is important to note that both the scattering parameters in (3.35) and (3.36) and the nonlinear mapping in (3.37) are influenced by the DC voltage or current bias at all device ports. While incident waves can also include a DC component, DC effects are omitted here for simplicity. For a detailed discussion on DC effects, see [27].

3.2.5 Time-Invariance of Nonlinear Scattering Mappings of Pseudowaves

The nonlinear scattering mapping, as expressed in Equation 3.37, is inherently time-invariant. This implies that if the input signals are uniformly delayed by τ seconds, the resulting outputs will remain identical to those for the original inputs, but shifted by the same time τ.

In the frequency domain, such a delay corresponds to a phase shift. The magnitude of this phase shift depends on the harmonic order. Specifically, a delay of τ seconds results in a phase shift of $k\tau$ for the k-th harmonic, where k indicates the harmonic order. Thus, the nonlinear mapping $F_{p,k}$ satisfies the property:

$$F_{p,k}(A_{1,1}e^{j\theta}, A_{1,2}(e^{j\theta})^2, \dots, A_{1,K}(e^{j\theta})^K, \dots) = F_{p,k}(A_{1,1}, A_{1,2}, \dots, A_{1,K}, \dots)(e^{j\theta})^k, \tag{3.43}$$

where $\theta = 2\pi f\tau$ is the phase shift. Each harmonic's contribution is scaled by a phase term $(e^{j\theta})^k$, proportional to its harmonic order.

Defining $P = e^{j\arg(A_{1,1})}$, the mapping simplifies to:

$$F_{p,k}(A_{1,1}, A_{1,2}, \dots, A_{1,K}, \dots) = F_{p,k}(|A_{1,1}|, A_{1,2}P^{-2}, \dots, A_{1,K}P^{-K}, \dots)P^k. \tag{3.44}$$

This transformation reduces the dimensionality of the parameter space by isolating the phase dependence of $A_{1,1}$ from the mapping.

3.2.6 Single Large-Tone X-Parameter Power Wave Relationship

Using the derived time-invariance property, a generalized X-parameter of type FB can now be defined. The X-parameter describes the relationship between incident and scattered pseudowaves for large-tone excitation. It is given by:

$$X_{p,k}^{(FB)}(|A_{1,1}|, A_{1,2}P^{-2}, \dots, A_{1,K}P^{-K}, \dots) = \frac{F_{p,k}(A_{1,1}, A_{1,2}, \dots)}{P^k}. \tag{3.45}$$

From this definition, the scattered pseudowave $B_{p,k}$ can be expressed as:

$$B_{p,k} = X_{p,k}^{(FB)}(|A_{1,1}|, A_{1,2}P^{-2}, \dots)P^k. \tag{3.46}$$

This formulation emphasizes the modularity of the X-parameter terms for efficient characterization. To approximate nonlinear behavior, linearization techniques are applied as in Section 3.4. The expansion leads to:

$$B_{p,k} \approx F_{p,k}(|A_{1,1}|, 0, \dots)P^k + \sum_{q,l} \frac{\partial F_{p,k}}{\partial A_{q,l}}\Big|_{A_{1,1}} A_{q,l}P^{k-l} + \sum_{q,l} \frac{\partial F_{p,k}}{\partial A^*_{q,l}}\Big|_{A_{1,1}} A^*_{q,l}P^{k+l}. \tag{3.47}$$

To further simplify, partial derivatives are defined as new X-parameters:

$$X^{(S)}_{p,k,q,l} = \frac{\partial F_{p,k}}{\partial A_{q,l}}\Big|_{A_{1,1}}, \quad X^{(T)}_{p,k,q,l} = \frac{\partial F_{p,k}}{\partial A^*_{q,l}}\Big|_{A_{1,1}}. \tag{3.48}$$

Combining these terms, the final power wave relationship is:

$$B_{p,k} = X^{(FB)}_{p,k}(|A_{1,1}|, DC, f)P^k + \sum_{q,l} X^{(S)}_{p,k,q,l}(|A_{1,1}|, DC, f)A_{q,l}P^{k-l} + \sum_{q,l} X^{(T)}_{p,k,q,l}(|A_{1,1}|, DC, f)A^*_{q,l}P^{k+l}. \tag{3.49}$$

This final relationship captures the nonlinear interactions between harmonics and highlights the contribution of each incident wave component.

This final relationship captures the nonlinear interactions between harmonics and highlights the contribution of each incident wave component.

3.3 TCAD Modeling and X-Parameter Extraction of GaN HEMT

In the previous two subsections, we introduced the fundamental performance metrics of power amplifiers and the basic principles of X-parameters. These concepts lay the theoretical foundation for analyzing and modeling the nonlinear behavior of RF power devices. Building upon this foundation, the current subsection focuses on the implementation of GaN HEMT device modeling within the TCAD (Technology Computer-Aided Design) environment and the subsequent extraction of X-parameters. This process enables the characterization of large-signal behavior with high accuracy, providing a reliable basis for simulation and circuit-level design optimization.

3.3.1 GaN HEMT Modeling in TCAD

Gallium Nitride (GaN) High Electron Mobility Transistors (HEMTs) are widely recognized as a key technology in the field of high-frequency, high-power, and high-efficiency applications. Thanks to their wide bandgap, high breakdown voltage, and excellent electron transport characteristics, GaN HEMTs significantly outperform traditional silicon-based and GaAs-based devices in terms of power density, frequency handling, and thermal reliability. They have become indispensable in areas such as 5G communications, radar systems, and satellite transmitters.

To analyze the internal physics and predict electrical performance before physical fabrication, Technology Computer-Aided Design (TCAD) offers a simulation-based approach rooted in fundamental semiconductor physics. TCAD enables detailed analysis of carrier transport, electric field distribution, breakdown behavior, and thermal effects, thereby reducing development cycles and supporting design optimization.

Among mainstream TCAD platforms, tools such as **Sentaurus** (by Synopsys), **ATLAS** (by Silvaco), and COMSOL Multiphysics are commonly used. In this study, we utilize Synopsys **Sentaurus**, which provides robust support for complex materials, advanced physical models, and both 2D and 3D device simulations. Its ability to handle phenomena such as self-heating, interface traps, and high-field effects makes it highly suitable for modeling GaN HEMTs.

The following section describes the step-by-step process of constructing a GaN HEMT model using Sentaurus, including structural definition, material and doping setup, physical model selection, meshing, and simulation configuration. The goal is to generate reliable electrical characteristics for subsequent X-parameter extraction.

3.3.2 Modeling Procedure Overview

The TCAD modeling of a GaN HEMT involves several essential steps: defining the device structure, assigning material properties and doping profiles, selecting appropriate physical models, configuring simulation conditions, and analyzing the resulting data.

1. Device Structure Definition

The geometric structure is defined according to the target device specifications. A typical GaN HEMT comprises a substrate (Si, SiC, or sapphire), a buffer layer, a GaN channel, an AlGaN barrier, and metal contacts for the source, drain, and gate. Critical dimensions such as layer thicknesses, gate length, and spacing directly influence device behavior.

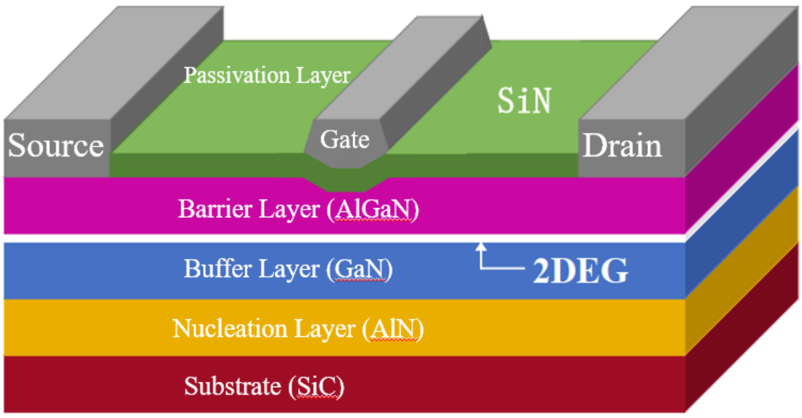

Figure 3.5: Cross-sectional structure of a typical AlGaN/GaN HEMT

2. Material and Doping Profiles

Each layer is assigned its corresponding material parameters, including bandgap energy, mobility, permittivity, and thermal properties. Common materials include GaN, AlGaN, and SiN. Accurate doping profiles—such as background doping in the buffer and high-concentration doping near the source/drain—must also be specified, as these determine carrier concentration and breakdown voltage.

3. Selection of Physical Models

To accurately capture the behavior of GaN HEMTs, several physical models are applied in simulation, including drift–diffusion carrier transport, field-dependent mobility, self-heating (lattice heating) effects, interface trap models, and bandgap narrowing with high-field velocity saturation. These models are selectively enabled based on the simulation focus and the device's intended application.

4. Meshing and Simulation Setup

The simulation mesh is refined in critical areas such as the channel and junction regions to ensure numerical accuracy. Bias conditions are defined, typically with the source grounded, a sweeping voltage applied to the gate, and a fixed drain bias. Environmental conditions such as temperature and convergence criteria are also set at this stage.

5. Solving and Data Analysis

After configuration, the simulation is executed using `Sentaurus Device`. The outputs include distributions of current density, potential, electric field, and carrier concentrations. From these, key I–V characteristics and performance metrics can be extracted to support further behavioral modeling and X-parameter characterization.

3.3.3 X-Parameter Extraction Procedure

In the modeling of RF power devices, X-parameters have emerged as an extended scattering parameter framework capable of capturing nonlinear behavior. They have been widely adopted for large-signal and behavioral modeling in recent years. Traditionally, the extraction of X-parameters relies heavily on the Harmonic Balance (HB) technique. This frequency-domain method interprets nonlinear circuit responses as the superposition of multiple harmonic components, enabling the direct computation of steady-state spectral responses. It is known for its computational efficiency and completeness in frequency information and is well-suited to linear or weakly nonlinear circuits. As a result, HB has been extensively used in commercial microwave CAD tools.

However, when applied in device-level simulations based on the Sentaurus TCAD platform, HB often encounters severe convergence issues. These challenges stem from several factors. First, GaN HEMTs exhibit strong nonlinearities under high voltage bias, including significant self-heating, field-enhanced mobility degradation, and carrier velocity saturation, which make it difficult for frequency-domain solvers to converge to stable solutions. Second, Sentaurus requires fine mesh grids in critical regions such as the junction and channel to maintain numerical accuracy. This dramatically increases the matrix size and computational burden for HB solvers, exacerbating instability [31]. Finally, TCAD simulations frequently involve coupling multiple complex physical models—such as lattice heating, interface traps, and drift-diffusion equations—which leads to highly nonlinear equation systems and complicates convergence due to poor initial approximations.

To address these limitations, this work proposes a time-domain transient simulation combined with multi-harmonic least-squares fitting as an alternative X-parameter extraction method. Specifically, a periodic input signal (e.g., single-tone or two-tone) is applied using `Sentaurus Device`, and the transient response waveform is collected after the system reaches steady periodic operation. Instead of applying a direct Fourier transform, the output waveform is approximated by a composite function composed of multiple harmonic basis terms. The coefficients of this model, including amplitude and phase of each harmonic, are extracted using the least-squares method, allowing the nonlinear mapping between excitation and response to be effectively identified and modeled in the X-parameter framework.

This approach offers several advantages: it avoids dependence on frequency-domain solvers, thus overcoming HB convergence bottlenecks under strong nonlinearity; it preserves harmonic information using time-domain results; and it allows flexible model complexity control to balance accuracy and computational cost. In summary, the transient + least-squares-based X-parameter extraction method introduced in this work provides a viable, robust, and platform-compatible solution for nonlinear modeling of GaN HEMTs, supporting large-signal simulation and behavioral modeling with high reliability.

To provide a rigorous theoretical foundation for the X-parameter extraction method proposed in this work, we present a newly formulated harmonic fitting approach based on transient simulation data. This method is designed to process real-valued waveforms from TCAD simulations by modeling the signal as a sum of harmonic sine and cosine functions, and estimating the amplitude and phase of each component via least squares fitting. It avoids the convergence issues often encountered with frequency-domain harmonic balance (HB) solvers under strong nonlinear conditions.

Unlike the traditional HB method, which solves nonlinear responses directly in the frequency domain using complex exponential basis functions, the approach proposed here is entirely time-domain-based. It relies only on waveform samples and basic trigonometric modeling, thus providing a numerically stable and broadly applicable alternative for large-signal nonlinear characterization.

Assume that the output waveform $f(t)$ of a power amplifier under periodic large-signal excitation can be approximated as:

$$f(t) = a_0 + \sum_{k=1}^{N} a_k \cos(k\omega_0 t + \phi_k) \tag{3.50}$$

where a_0 is the DC offset, a_k and ϕ_k are the amplitude and phase of the k-th harmonic, and ω_0 is the fundamental angular frequency. Since this model contains nonlinear phase terms, we introduce an equivalent linear representation:

$$f(t) = a_0 + \sum_{k=1}^{N} [A_k \cos(k\omega_0 t) + B_k \sin(k\omega_0 t)] \tag{3.51}$$

with coefficients $A_k = a_k \cos(\phi_k)$ and $B_k = -a_k \sin(\phi_k)$. Given M time samples from the transient simulation, i.e., $f(t_1), f(t_2), \ldots, f(t_M)$, the sampled data can be expressed as an overdetermined linear system:

$$\mathbf{f} = \mathbf{X} \cdot \boldsymbol{\theta} + \mathbf{e} \tag{3.52}$$

where $\mathbf{f} \in \mathbb{R}^{M\times 1}$ is the sampled data vector, $\boldsymbol{\theta} \in \mathbb{R}^{(2N+1)\times 1}$ is the parameter vector containing a_0, A_k, B_k, and $\mathbf{X} \in \mathbb{R}^{M\times(2N+1)}$ is the design matrix consisting of basis functions. The optimal solution can be obtained using the least squares criterion:

$$\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{f} \tag{3.53}$$

Once A_k and B_k are obtained, the original harmonic amplitudes and phases are recovered by:

$$a_k = \sqrt{A_k^2 + B_k^2}, \quad \phi_k = -\arctan\left(\frac{B_k}{A_k}\right) \tag{3.54}$$

This method guarantees the uniqueness of the solution and minimizes the squared fitting error. It is both numerically robust and easy to implement. Figure 3.6 illustrates the waveform sampling process and matrix formulation used in this proposed method.

It is worth emphasizing that this modeling and derivation approach was independently developed in this study. The motivation was to overcome the severe convergence limitations observed in TCAD simulations using harmonic balance solvers when modeling strongly nonlinear devices. The proposed approach provides a reliable, frequency-domain-free alternative for extracting harmonic characteristics from transient responses, with strong applicability in nonlinear RF device modeling.

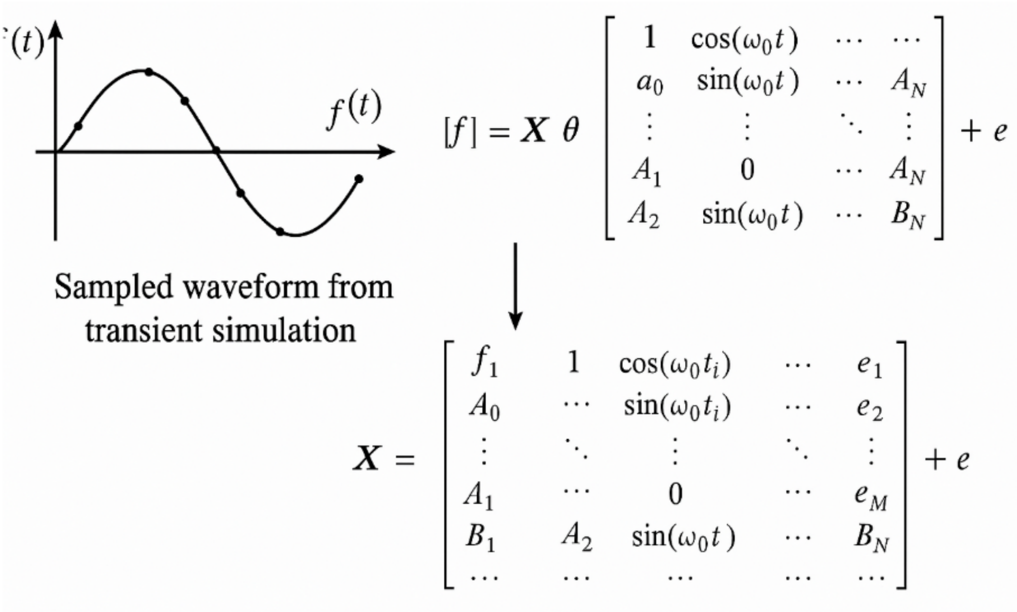

Figure 3.6: Proposed method: transient waveform samples are fitted using harmonic basis via least squares

3.3.4 Development of an Artificial Neural Network Model for TCAD-Based GaN HEMT

Artificial neural networks are mathematical models capable of approximating complex nonlinear functions. Their advantage lies in their ability to learn and extract features from large-scale, high-dimensional data. Compared with traditional linear models, neural networks use multiple interconnected neurons to fit the relationship between input and output. Each neuron applies a nonlinear activation function to its inputs, enabling the network to capture complex patterns.

Let the input vector be $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and the output vector be $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$, where the neural network model can be expressed as:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}, \mathbf{b}) \tag{3.55}$$

To train the neural network parameters $\mathbf{w}$ and $\mathbf{b}$, a dataset containing input features and corresponding output labels is required. The network minimizes the error (loss) between predicted and actual values by updating weights iteratively using optimization algorithms such as gradient descent. The following sections introduce three key components: activation functions, loss functions, and backpropagation.

3.3.4.1Nonlinear Activation Functions

3.3.4.0.1 Tanh Function The Tanh (hyperbolic tangent) activation function maps input values to the range $[-1, 1]$, centered at zero. It is defined as:

$$y = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.56}$$

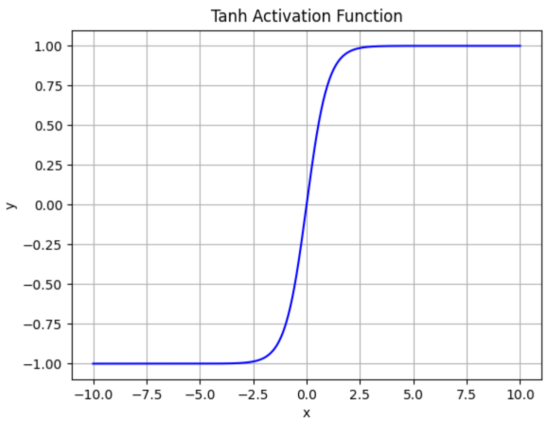

Figure 3.7: Tanh Activation Function

Although smooth, the Tanh function can lead to vanishing gradients for large or small inputs, which may hinder training convergence.

3.3.4.0.2 ReLU Function The Rectified Linear Unit (ReLU) is widely used due to its simplicity and efficiency:

$$y = \text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.57}$$

ReLU allows positive signals to pass unchanged while setting negatives to zero, which may result in the ”dead neuron” problem.

3.3.4.0.3 Leaky ReLU Function Leaky ReLU addresses the dead neuron issue by allowing a small negative slope α:

$$y = \text{Leaky ReLU}(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases} \tag{3.58}$$

The typical value of α is 0.01. Leaky ReLU retains some gradient in the negative region, thus improving robustness.

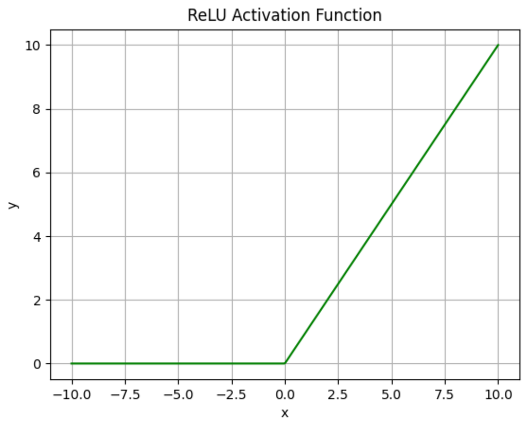

Figure 3.8: ReLU Activation Function

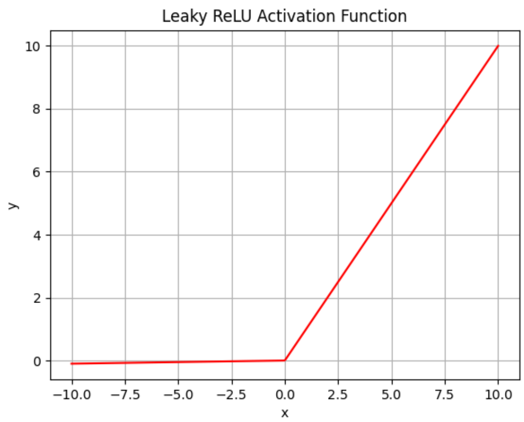

Figure 3.9: Leaky ReLU Activation Function

3.3.4.2Optimization Criteria

The objective of neural network training is to minimize the loss function $\text{Loss}(\mathbf{w}, \mathbf{b})$:

$$\mathbf{w}, \mathbf{b} = \arg\min_{\mathbf{w}, \mathbf{b}} \text{Loss}(\mathbf{w}, \mathbf{b}) \tag{3.59}$$

3.3.4.0.4 Mean Squared Error (MSE) MSE is a standard loss for regression problems, sensitive to large errors:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3.60}$$

3.3.4.0.5 Mean Absolute Error (MAE) MAE computes the average absolute difference, more robust to outliers:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{3.61}$$

3.3.4.3Backpropagation Algorithm

Backpropagation computes gradients of the loss function with respect to each parameter and updates weights using gradient descent. For a network with L layers and activation functions $\sigma^{(l)}$, the forward pass is:

$$\hat{y} = \sigma^{(L)}(w^{(L)}\sigma^{(L-1)}(\ldots \sigma^{(1)}(w^{(1)}x))) \tag{3.62}$$

The loss gradient with respect to the output is:

$$\frac{\partial \text{Loss}}{\partial \hat{y}} = \hat{y} - y \tag{3.63}$$

Gradient at the output layer:

$$\delta^{(L)} = \frac{\partial \text{Loss}}{\partial \hat{y}} \cdot \sigma^{(L)\prime}(z^{(L)}) \tag{3.64}$$

Gradient at hidden layers ($1 \leq l \leq L - 1$):

$$\delta^{(l)} = (w^{(l+1)T}\delta^{(l+1)}) \cdot \sigma^{(l)\prime}(z^{(l)}) \tag{3.65}$$

Gradient of weights:

$$\frac{\partial \text{Loss}}{\partial w^{(l)}} = \delta^{(l)}(a^{(l-1)})^T \tag{3.66}$$

Weight update rule:

$$w_{\text{new}}^{(l)} = w^{(l)} - \eta \frac{\partial \text{Loss}}{\partial w^{(l)}} \tag{3.67}$$

where η is the learning rate controlling update step size.

3.3.5 Incremental Learning-Based X-Parameter Prediction

As modeling precision demands increase for GaN HEMT devices, X-parameters have emerged as an effective method for capturing the nonlinear behavior of RF transistors under large-signal excitation. However, traditional offline training approaches require collecting all data beforehand, which limits adaptability to changing device conditions. To overcome this limitation, we propose an **incremental learning** framework combined with neural networks to support continuous and efficient X-parameter modeling.

Incremental learning allows a model to dynamically update its parameters while continuously receiving new data, without retraining from scratch. The goal is to acquire new knowledge θ_{new} from new data D_{new} while retaining prior knowledge θ_{old}. This learning objective can be mathematically expressed as:

$$\theta_{\text{new}} = \arg\min_{\theta} \left(\mathcal{L}_{\text{new}}(\theta; D_{\text{new}}) + \lambda \cdot \Omega(\theta, \theta_{\text{old}}) \right) \tag{3.68}$$

where:

- $\mathcal{L}_{\text{new}}$ is the loss function on the new dataset, such as MSE or Huber Loss;
- $\Omega(\theta, \theta_{\text{old}})$ is a regularization term penalizing deviation from old knowledge;
- λ is a trade-off coefficient balancing knowledge stability and adaptability.

This formulation reflects the stability–plasticity trade-off essential to incremental learning, allowing the model to generalize to new conditions without catastrophic forgetting.

In this work, we fine-tune a pretrained neural network to adapt to new conditions. Initial training is based on simulated data D_0 generated by TCAD, yielding a base model θ_0. Upon arrival of new data, the model is updated using a few epochs or low learning rate fine-tuning:

$$\theta_t \leftarrow \theta_{t-1} - \eta \cdot \nabla_\theta \mathcal{L}(f(\mathbf{x}; \theta_{t-1}), \mathbf{y}) \tag{3.69}$$

where η is the learning rate, $f(\cdot)$ is the network, and $\mathcal{L}$ is the loss on the new data.

To further improve model stability, we integrate the **Elastic Weight Consolidation (EWC)** method. EWC introduces a regularization term that preserves important parameters from previous tasks by penalizing deviations weighted by the Fisher Information Matrix F_i:

$$\mathcal{L}_{\text{EWC}}(\theta) = \mathcal{L}_{\text{new}}(\theta) + \frac{\lambda}{2} \sum_i F_i (\theta_i - \theta_i^*)^2 \tag{3.70}$$

where θ_i^* are parameters from the previous model and F_i reflects their importance. This strategy mitigates catastrophic forgetting by preserving critical weights.

Compared to retraining from scratch, our method significantly reduces training time and memory usage while adapting to changing operating conditions. It enables real-time or near real-time behavioral modeling of GaN HEMTs, and lays a practical foundation for future intelligent modeling in RF systems.

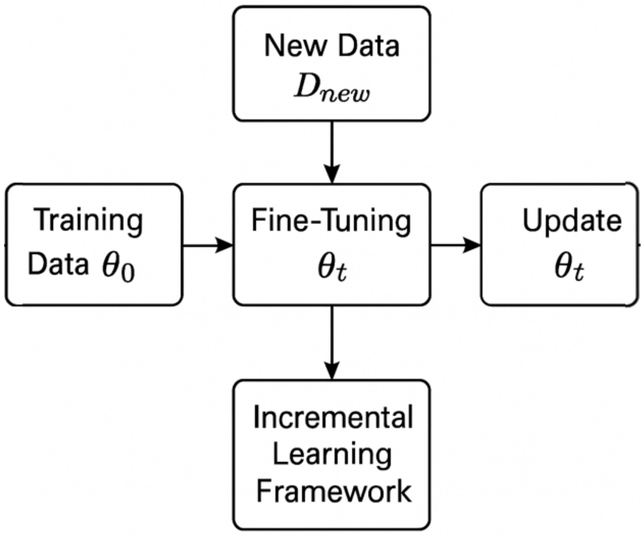

Figure 3.10: Incremental Learning Framework for X-Parameter Prediction

3.4 Summary

This chapter systematically addressed the behavioral modeling of RF power amplifiers (PAs), combining theoretical foundations, device-level simulation, data extraction, and machine learning–based prediction. It lays a comprehensive modeling framework that integrates physical insight with data-driven approaches, forming the technical basis for the subsequent analysis and design chapters.

The chapter began by introducing fundamental performance metrics of PAs, including power gain, 1 dB compression point, third-order intermodulation intercept point (IP3), and power-added efficiency (PAE). These metrics serve not only as critical indicators of amplifier performance but also provide quantitative references for non-linear modeling and verification. We further reviewed the theoretical background of X-parameters, which extend traditional S-parameters into the large-signal nonlinear domain and are widely applied for high-frequency behavioral modeling.

For device-level modeling, this chapter utilized Synopsys Sentaurus TCAD tools to simulate a GaN HEMT structure under multiple bias conditions. Key modeling steps—including geometry construction, material configuration, doping setup, meshing, and physical model selection (e.g., mobility degradation, thermal effects, trap states)—were described in detail. Due to the limitations of harmonic balance (HB) methods in handling strong nonlinearity within Sentaurus (such as poor convergence), a novel time-domain fitting approach was proposed. The method reconstructs the output signal using cosine-based harmonic superposition and least-squares fitting, effectively extracting amplitude and phase components of the X-parameters without relying on frequency-domain solvers.

At the data-driven modeling level, an artificial neural network (ANN)–based behavioral model was proposed to predict X-parameter behavior. Various activation functions (Tanh, ReLU, Leaky ReLU) and loss functions (MSE, MAE, Huber Loss) were incorporated to enhance learning flexibility and generalization. The backpropagation algorithm was mathematically described, showing how model parameters are optimized iteratively to reduce the error between predicted and target outputs.

To address the need for continuous learning and adaptability in real-world deployment, this chapter introduced an incremental learning framework for updating the ANN model with newly generated data. Specifically, the Elastic Weight Consolidation (EWC) method was adopted to prevent catastrophic forgetting by penalizing parameter deviations based on their importance. A regularized objective function incorporating the Fisher information matrix was derived to balance knowledge retention and new data adaptation.

Finally, a complete flowchart of the incremental learning–based X-parameter prediction system was presented to illustrate the continuous learning process, from initial model training to dynamic fine-tuning and model refinement.

In summary, this chapter establishes a unified, multi-level modeling strategy for GaN HEMT–based PAs that integrates physics-based simulation, harmonic fitting, deep learning prediction, and continual learning techniques. The framework provides a robust foundation for advanced modeling, error analysis, and circuit-level implementation, as explored in the following chapters.

4 Experimental Results and Discussion

Gallium Nitride (GaN) HEMT devices are prone to defect formation due to varying levels of impurity incorporation during the fabrication process [32]. These defects can introduce significant variability in device behavior, particularly under large-signal operating conditions, affecting both DC and AC performance. As a result, large-signal modeling of GaN HEMTs has become a crucial bridge between semiconductor process technology and circuit-level design. Accurate modeling is essential for the successful implementation of GaN-based power amplifier (PA) circuits.

However, conventional physical models often suffer from limitations in practicality. Their structural complexity, rooted in nonlinear partial differential equations (PDEs), leads to high computational overhead and extended simulation times. Moreover, analytical models typically require intensive process-dependent calibration and iterative correction, making them both time-consuming and difficult to generalize across technologies.

To overcome these challenges, this work adopts a TCAD-based modeling approach. Technology Computer-Aided Design (TCAD) enables precise characterization of GaN HEMTs at the physical level by incorporating material, geometry, and field interaction effects. By using TCAD-generated data as the foundation, this chapter focuses on developing and validating behavioral models that more accurately capture the device's nonlinear characteristics, offering a scalable and efficient alternative for large-signal modeling.

4.1 GaN HEMT Device Modeling Using TCAD

In order to accurately simulate the electrical characteristics of GaN HEMT devices, a comprehensive TCAD model was constructed that includes multiple layers and considers thermal effects. Figure 4.1 presents the cross-sectional view of the GaN HEMT structure, which includes an AlGaN barrier, GaN channel, AlGaN buffer, AlN seed, and oxide layers. The thermal conductivities (*kth*) of each material are defined to enable self-heating simulation.

To investigate the influence of different physical models on the device's behavior, the output I_D-V_{DS} characteristics were simulated under three models: Drift-Diffusion (DD), Thermodynamic (Thermo), and Hydrodynamic (Hydro). As shown in Figure 4.2, the DD model exhibits the highest drain current, while Thermo and Hydro models demonstrate the reduction due to thermal feedback and carrier energy effects.

Subsequently, the I_D-V_G transfer curves were extracted for varying drain voltages (V_{DS} = 0 V, 5 V, 10 V, 15 V, 20 V). Figure 4.3 shows the consistency of threshold voltage and peak transconductance across voltage levels, demonstrating stable channel formation. The curves remain consistent across thermal models, indicating that the DD model is sufficiently accurate for small-signal simulations.

Finally, the output I_D-V_{DS} curves were obtained for a set of gate voltages ranging from -4 V to +4 V. Figure 4.4 reveals that with increasing gate bias, the device gradually enters saturation, and the current level increases linearly with V_{DS} before compressing, which aligns with the expected large-signal performance.

Based on the above results, the Drift-Diffusion (DD) model is selected for subsequent behavioral modeling, balancing computational efficiency and predictive accuracy.

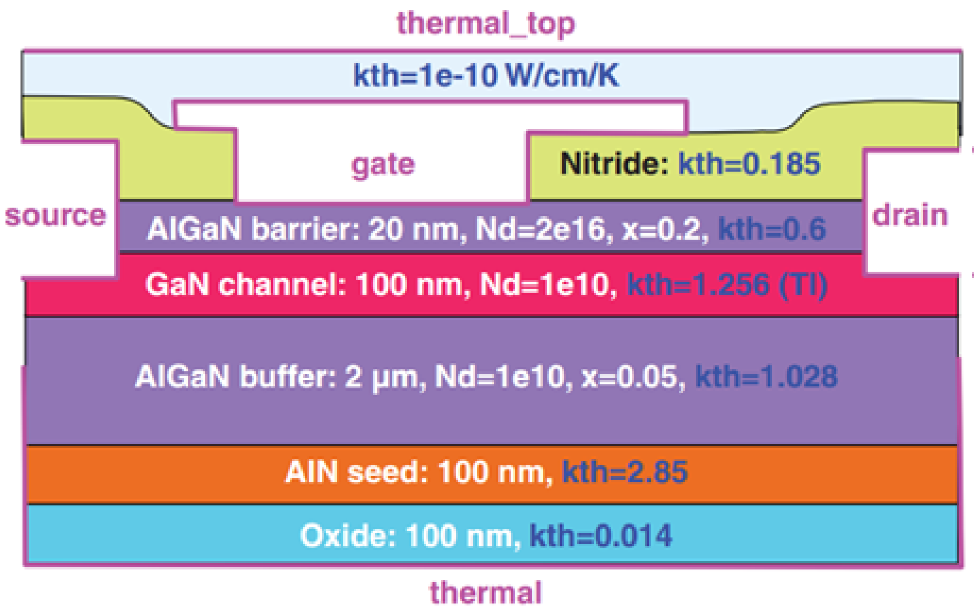

Figure 4.1: Cross-sectional structure of the GaN HEMT model with defined thermal conductivities.

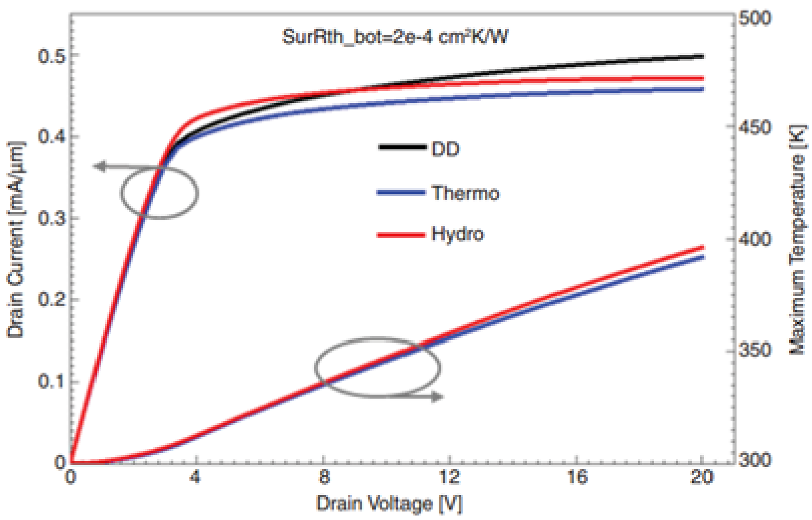

Figure 4.2: Output characteristics under DD, Thermo, and Hydro models. DD model shows higher drain current due to the absence of thermal effects.

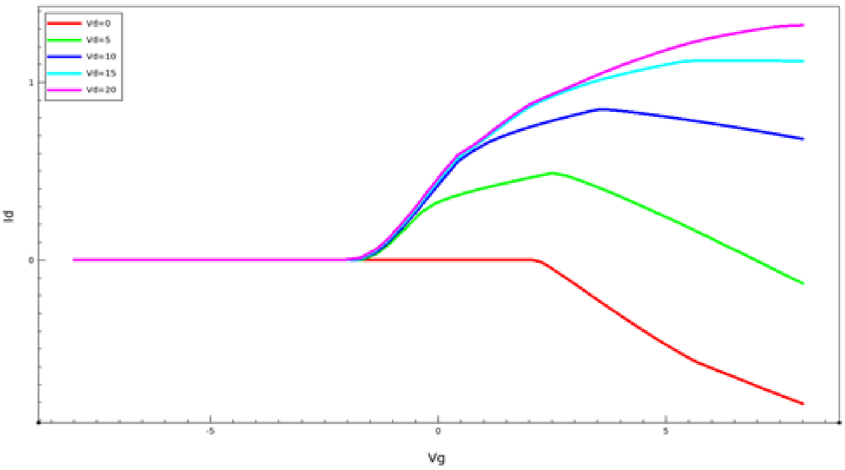

Figure 4.3: Transfer characteristics (I_D-V_G) under different drain voltages. Minimal variation confirms the robustness of the channel behavior.

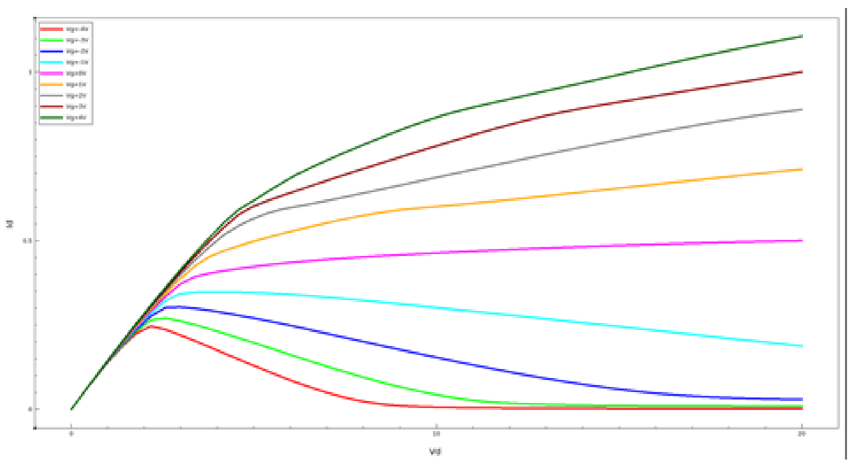

Figure 4.4: Output characteristics (I_D-V_{DS}) at different gate voltages (V_G = -4 V to +4 V).

4.2 X-Parameter Extraction for GaN HEMT

4.2.1 X-Parameter Extraction Circuit for GaN HEMT

To accurately characterize the nonlinear behavior of GaN HEMT devices under large-signal conditions, a dedicated simulation circuit for X-parameter extraction is designed, as illustrated in Fig. 4.5. This dual-path structure separately applies large-signal excitation and small-signal probing, aligning with the requirements of harmonic excitation and scattering response modeling, and serves as a crucial foundation for X-parameter-based behavioral modeling.

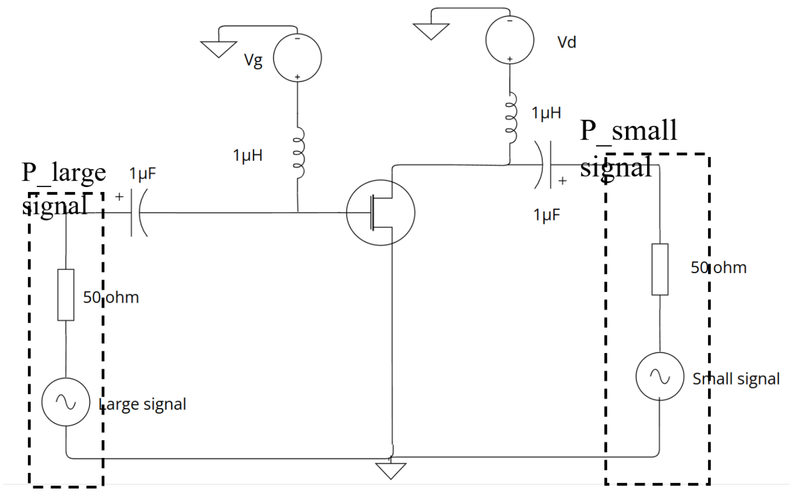

Figure 4.5: Schematic diagram of X-parameter extraction circuit for GaN HEMT

In this testbench, both large and small signal sources are connected through $50\,\Omega$ matching resistors to ensure impedance matching and suppress signal reflection. Inductors of $1\,\mu H$ are used to decouple DC bias from the RF paths at both the gate and drain terminals. Capacitors of $1\,\mu F$ are employed to block DC while allowing high-frequency signals to pass through.

The simulation conditions are configured as follows:

- **Gate bias (V_g)**: 2V and 4V;
- **Drain bias (V_d)**: 10V and 20V;
- **Excitation frequency range**: 1 GHz to 6 GHz, with a step size of 0.2 GHz;
- **Large-signal amplitude sweep**: 0 V to 8 V, in 0.2 V steps;
- **Total simulation sets**: 4800 data points.

This configuration ensures comprehensive coverage of the device's dynamic response under various biasing, frequency, and excitation conditions. The extracted X-parameters contain rich complex-domain information linking incident and reflected waves at multiple harmonics. They capture key nonlinear phenomena such as harmonic interaction, gain compression, and memory effects, making them an effective bridge between physical-level TCAD simulations and system-level behavioral modeling.

4.2.2 X-Parameter Extraction Process

To assess the nonlinear behavior of the GaN HEMT device under large-signal excitation, we selected a representative bias condition with gate voltage $V_g = -4\,\mathrm{V}$ and drain voltage $V_d = 15\,\mathrm{V}$. A large-signal sinusoidal excitation with a frequency of 2 GHz and an amplitude of 4 V was applied. This setup provides a suitable environment for examining the harmonic content and convergence behavior of the device under strong nonlinearity.

Figure 4.6 shows the gate voltage (V_{gs}) and gate current (I_g) waveforms. It can be observed that the voltage waveform is approximately sinusoidal, while the gate current exhibits a slight asymmetry due to nonlinear effects. Figure 4.7 displays the corresponding drain voltage (V_{ds}) and drain current (I_{ds}) waveforms, where the nonlinear nature of I_{ds} is clearly reflected in its waveform distortion.

To extract the harmonic content of V_{ds} and I_{ds}, we applied a least-squares fitting method to the transient waveform data. The fitting results are shown in Figure 4.8 and Figure 4.9. The fitted curves align closely with the original data, demonstrating the validity and accuracy of the harmonic decomposition method used in this work.

Based on the time-domain data and fitted harmonics, we further derived the amplitude and phase spectra of the fundamental and higher-order harmonics. Figure 4.10 and Figure 4.11 present the amplitude and phase spectra of V_{ds}, respectively, while Figure 4.12 and Figure 4.13 show the corresponding spectra for I_{ds}. These spectral results are essential inputs for constructing the X-parameters that capture the nonlinear large-signal behavior of the device.

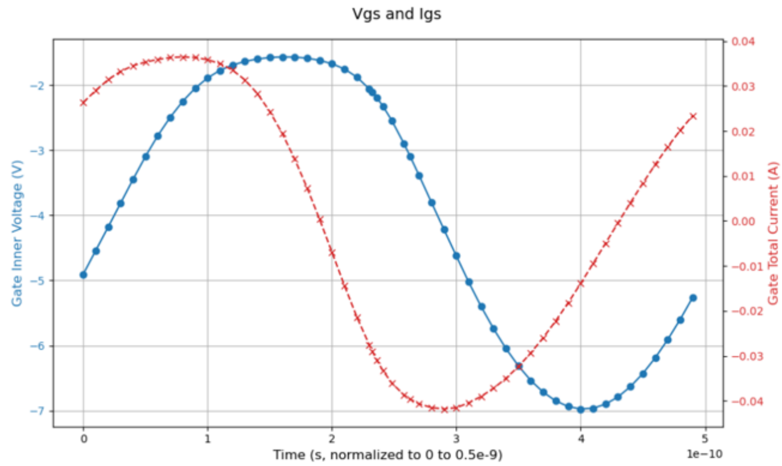

Figure 4.6: Gate voltage (V_{gs}) and gate current (I_g) waveform under large-signal excitation.

4.2.3 Optimized Harmonic Parameters Extraction

To evaluate the effectiveness of the cosine-based least squares fitting method for modeling nonlinear responses under large-signal excitation, we extracted the DC component and the first seven harmonics (amplitude and phase) for both V_{ds} and I_{ds} from the fitted waveform data. The extracted harmonic coefficients are summarized in Tables 4.1 and 4.2.

As shown in the results, all fitting errors are below 0.001, which indicates that the fitting method achieves high accuracy. The close match between the fitted waveform and the original signal demonstrates the

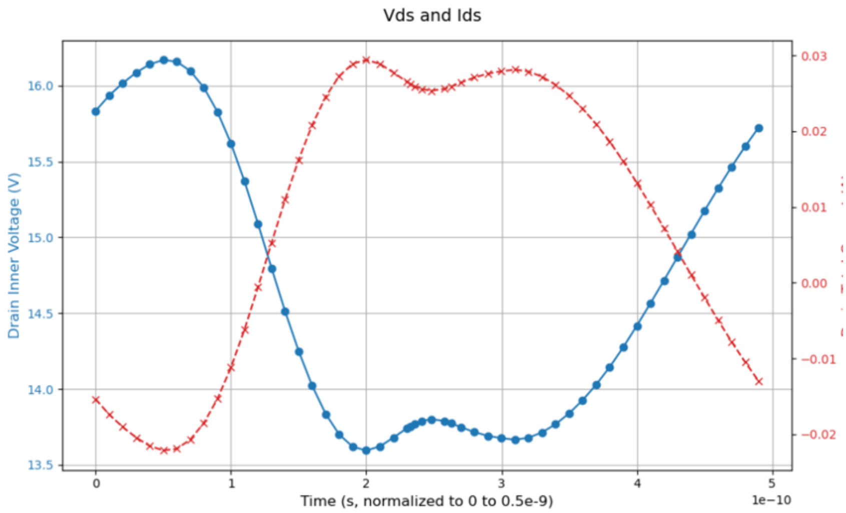

Figure 4.7: Drain voltage (V_{ds}) and drain current (I_{ds}) waveform under large-signal excitation.

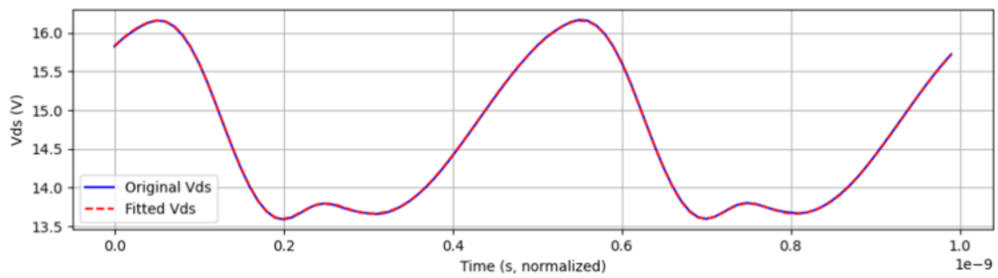

Figure 4.8: Fitted V_{ds} waveform using the least-squares method.

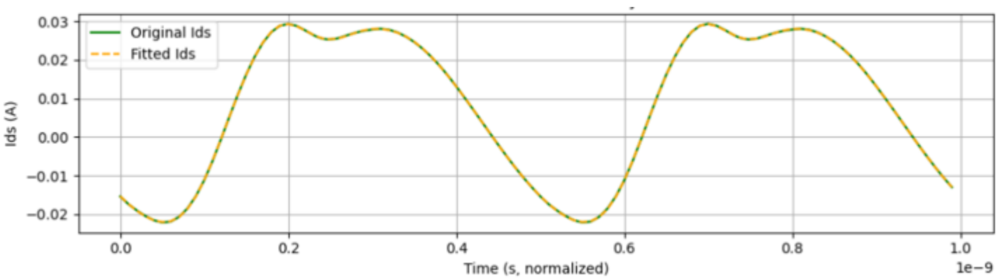

Figure 4.9: Fitted I_{ds} waveform using the least-squares method.

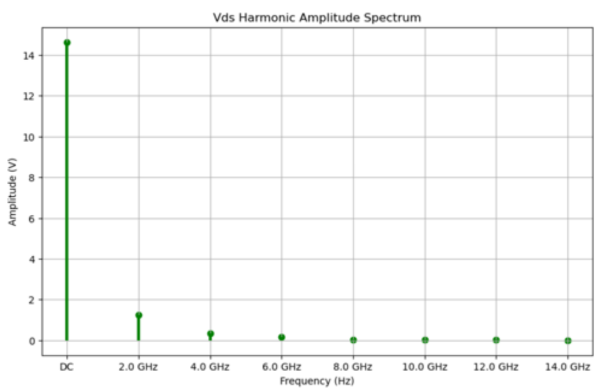

Figure 4.10: V_{ds} harmonic amplitude spectrum.

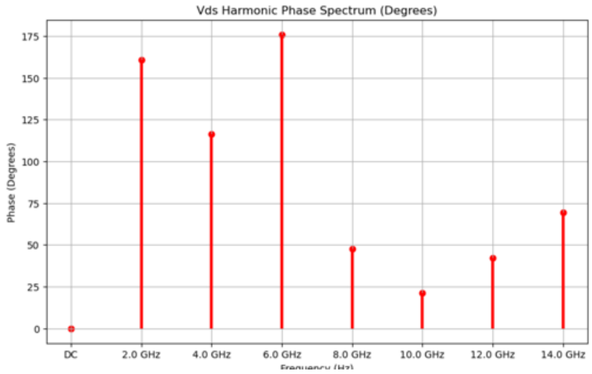

Figure 4.11: V_{ds} harmonic phase spectrum.

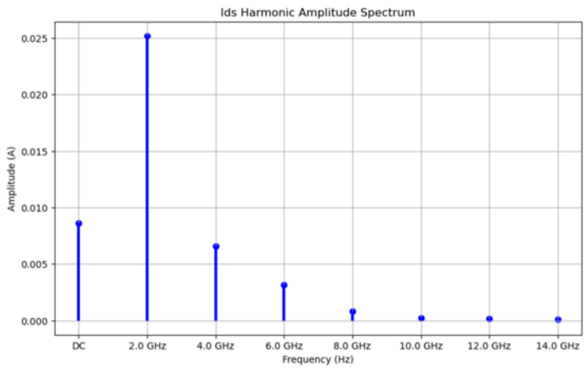

Figure 4.12: I_{ds} harmonic amplitude spectrum.

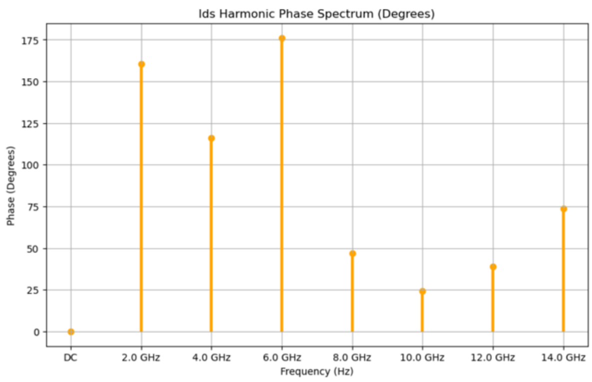

Figure 4.13: I_{ds} harmonic phase spectrum.

method's reliability in representing nonlinear behavior and transient responses of GaN HEMT devices under large-signal conditions.

Table 4.1: Optimized Harmonic Parameters for V_{ds}

Harmonic	Amplitude	Phase (radians)	Fitting Error
DC	14.6327	—	—
1st	1.2574	-188.8308	0.0004
2nd	0.3622	-133.0594	0.0006
3rd	-0.1592	6.2116	0.0002
4th	-0.0244	13.3970	0.0003
5th	0.0103	-2.7735	0.0005
6th	0.0096	0.7328	0.0004
7th	0.0036	4.3499	0.0003

Table 4.2: Optimized Harmonic Parameters for I_{ds}

Harmonic	Amplitude	Phase (radians)	Fitting Error
DC	0.0086	—	—
1st	0.0252	9.0835	0.0005
2nd	-0.0067	5.1672	0.0003
3rd	0.0032	-0.0670	0.0004
4th	-0.0008	0.8148	0.0002
5th	0.0002	0.4252	0.0003
6th	-0.00002	0.6816	0.0004
7th	0.00008	1.2860	0.0002

To extract complete X-parameters for GaN HEMT under large signal excitation, this study adopts a three-tone superposition method, which is theoretically sufficient to separate the linear and nonlinear response terms. The general expression for the X-parameter expansion is given by:

$$B_{ik} = X_{ik}^{(F)}(A_{11})P^k + X_{ik}^{(S)}(d|A_{11}|)P^k a_{ji} + X_{ik}^{(T)}(d|A_{11}|)P^{k+1} a_{ji}$$

Here, P^k represents the k-th harmonic of the large signal at port 1, while a_{ji} denotes the small signal stimulus injected during the superposition process. The three key terms represent: the large-signal-only contribution $X^{(F)}$, the linear response to the small signal $X^{(S)}$, and the second-order modulation component $X^{(T)}$.

To extract these components, three independent simulations are conducted:

1. Only the large signal is applied.
2. A large signal is applied along with a small signal in-phase (0° phase difference).
3. A large signal is applied along with a small signal in quadrature (90° phase difference).

Figure 4.14 and Figure 4.15 show the output voltage and current waveforms under these three conditions. It is evident that the introduction of small signals induces subtle differences in the waveforms. These variations are the key to extracting the small-signal response embedded within the large-signal excitation.

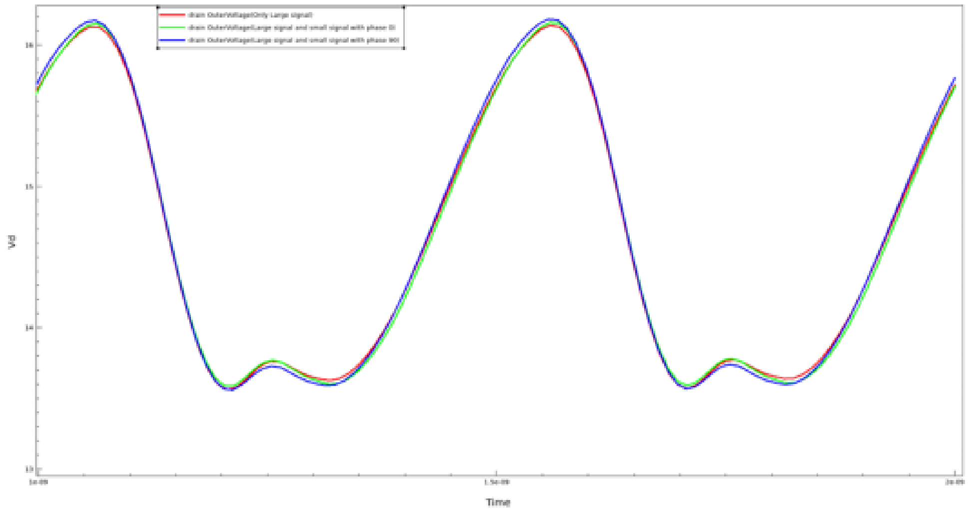

Figure 4.14: Drain Voltage under Large Signal Only and Combined Signals (Phase 0° and 90°)

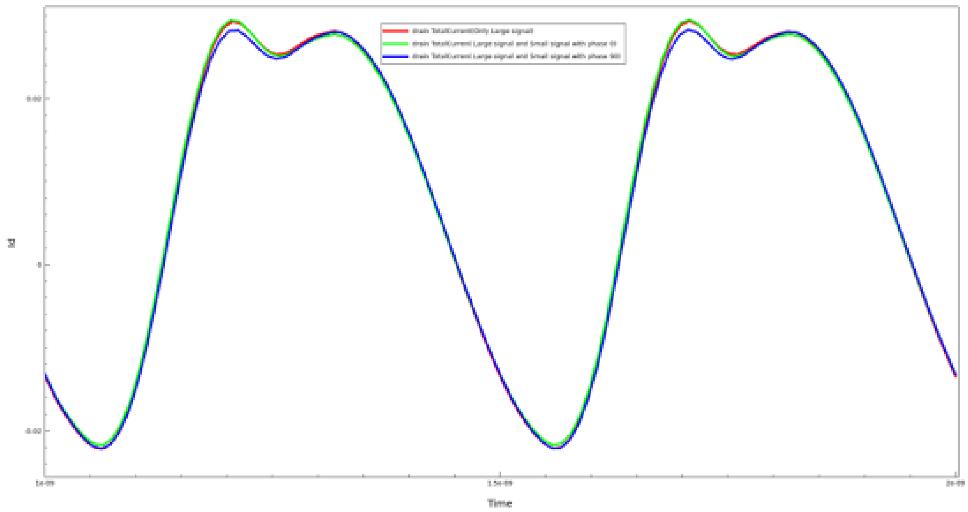

Figure 4.15: Drain Current under Large Signal Only and Combined Signals (Phase 0° and 90°)

By applying least-squares fitting across these three simulations, the system of equations is solvable, yielding the three X-parameter components: $X^{(F)}$, $X^{(S)}$, and $X^{(T)}$. The obtained results demonstrate excellent agreement, with back-calculated output signals exhibiting minimal errors (on the order of 10^{-17}) compared to the original simulated values. This confirms the accuracy of the coefficient extraction method and validates the feasibility of using transient simulations to derive X-parameters in a TCAD-based modeling framework.

4.3 GaN HEMT Artificial Neural Network Model Development result

4.3.1 Incremental Learning for X-Parameter Prediction

To enhance the flexibility and accuracy of X-parameter modeling under diverse operating conditions, we developed an artificial neural network (ANN) prediction framework using incremental learning techniques. This approach allows the model to be updated with new data batches over time without retraining from scratch, effectively addressing the problem of catastrophic forgetting common in standard neural networks.

The overall modeling and training workflow is shown in Figure 4.16. Initially, the neural network is trained on a portion of the dataset. As new data becomes available, the model is updated incrementally using Elastic Weight Consolidation (EWC) to retain important parameters from previous tasks, while learning new X-parameter mappings.

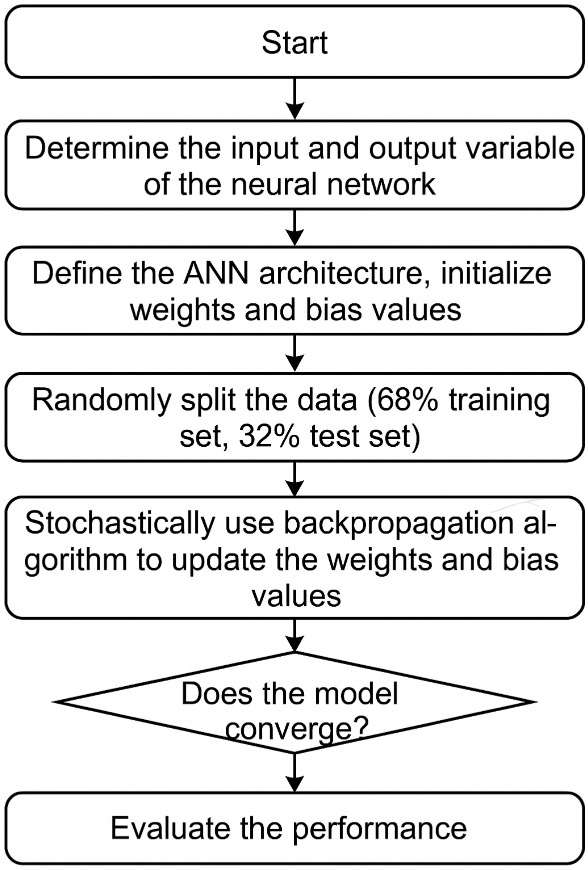

Figure 4.16: Workflow of X-parameter prediction using incremental learning

In total, 4800 sets of simulation data were generated using a TCAD-based GaN HEMT model across varying gate and drain voltages and signal amplitudes. The model was trained incrementally and evaluated on unseen samples for predictive performance.

Table 4.3 presents a portion of the prediction results for both V_{ds} and I_{ds}. The predicted values show excellent agreement with the ground truth, and the corresponding errors remain very low—demonstrating the reliability of our approach in capturing nonlinear X-parameter behavior.

Table 4.3: Sample prediction results for V_{ds} and I_{ds}

Vds (True)	Vds (Pred)	Error	Ids (True)	Ids (Pred)	Error
0.58962	0.57269	0.01694	0.08809	0.08821	0.00012
0.00527	-0.02502	0.03030	0.09079	0.08850	0.00229
0.15381	0.14488	0.00893	0.08297	0.08369	0.00072
-0.01496	0.00216	0.01713	-0.02597	-0.02485	0.00112
-0.60951	-0.60523	0.00428	-0.09691	-0.09474	0.00217

4.4 Summary

Chapter Summary

This chapter presents a comprehensive study on GaN HEMT modeling and X-parameter extraction under large-signal conditions. Starting with TCAD simulation, we explored various physical models and selected the Drift-Diffusion approach as a balance between accuracy and computational efficiency. A dedicated extraction circuit was designed, and a time-domain fitting method based on least squares was used to replace traditional harmonic balance techniques, thus overcoming convergence issues and improving the stability of parameter extraction.

Following this, a three-tone excitation scheme was employed to fully extract the X-parameter coefficients. The reconstructed waveforms showed negligible error compared to the simulation data, validating the reliability of the method.

Finally, we introduced artificial neural networks and incremental learning to predict X-parameters based on 4800 simulation samples. The proposed model demonstrated high prediction accuracy for both V_{ds} and I_{ds}, showing promise for future behavioral modeling and device-level prediction tasks.

5 Conclusion and Recommendations

This thesis presented a comprehensive investigation into the modeling and prediction of large-signal behavior in GaN HEMT devices, using a hybrid approach that combines physics-based TCAD simulations with data-driven artificial neural network (ANN) techniques.

Firstly, the study addressed the limitations of traditional physical models—particularly those relying on harmonic balance (HB) methods—when applied within the TCAD environment. These limitations include poor convergence due to high nonlinearity, complex mesh refinement, and coupling of multiple physical models. To overcome this, a novel time-domain method based on transient simulation and harmonic fitting using the least-squares technique was introduced. This approach avoids frequency-domain solvers altogether, enabling numerically stable and accurate X-parameter extraction even under strong nonlinear conditions:contentReference[oaicite:0]index=0.

Secondly, we developed a dedicated extraction circuit and employed three-tone simulations to decouple linear and nonlinear response components. The reconstructed waveforms showed excellent agreement with the simulated data, validating the proposed X-parameter extraction technique.

To extend the modeling capabilities, we introduced an artificial neural network framework trained on 4800 sets of TCAD-generated data. The ANN was designed to predict X-parameters under varying bias and signal conditions. To ensure scalability and adaptability, incremental learning via Elastic Weight Consolidation (EWC) was integrated, allowing the model to retain previously learned knowledge while adapting to new data. The predicted results of Vds and Ids were consistent with the reference values, with minimal errors, demonstrating strong generalization and prediction capabilities:contentReference[oaicite:1]index=1.

The major contributions of this thesis can be summarized as follows:

- A novel X-parameter extraction method based on time-domain simulations and least-squares harmonic fitting, which overcomes the convergence bottlenecks of HB solvers in TCAD.
- Integration of TCAD simulation with behavioral modeling, enabling accurate device-level prediction while preserving physical insight.
- The use of incremental learning to construct a continually improving ANN framework, which supports dynamic updates and avoids catastrophic forgetting.
- Demonstration of high accuracy and reliability in both model fitting and ANN prediction, supported by extensive numerical validation.

Recommendations for Future Work:

- Explore the integration of more advanced deep learning architectures (e.g., transformers or graph neural networks) to improve nonlinear mapping accuracy in multi-port or broadband scenarios.
- Extend the modeling framework to include memory effects and thermal feedback mechanisms, further enhancing the fidelity of GaN HEMT behavioral models.
- Investigate real-time model adaptation strategies in hardware-in-the-loop environments, accelerating the deployment of data-driven models in RF system design.

- Apply the developed methods to other compound semiconductors such as GaAs and InP, validating the generality of the proposed approach across device technologies.
- Incorporate fabrication-induced variability and process uncertainty into the model, potentially using probabilistic neural networks to predict distributions rather than single deterministic outputs.

In conclusion, this work provides a robust, scalable, and practical modeling strategy that bridges the gap between physics-based device modeling and efficient behavioral prediction. It lays a strong foundation for future research in intelligent semiconductor modeling, system-level co-design, and accelerated RF design workflows.

BIBLIOGRAPHY

[1] M. Chen, W. Sutton, I. Smorchkova, B. Heying, W.-B. Luo, V. Gambin, F. Oshita, R. Tsai, M. Wojtowicz, R. Kagiwada *et al.*, “A 1–25 ghz gan hemt mmic low-noise amplifier,” *IEEE Microwave and Wireless Components Letters*, vol. 20, no. 10, pp. 563–565, 2010.

[2] A. H. Jarndal, *Large signal modeling of GaN device for high power amplifier design*. kassel university press GmbH, 2006.

[3] T. Gasseling, “Compact transistor models: The roadmap to first-pass amplifier design success,” *Microwave Journal*, vol. 55, no. 3, pp. 74–86, 2012.

[4] M. R. Moure, M. Casbon, N. Ladero, M. Fernandez-Barciela, and P. J. Tasker, “A systematic investigation of admittance domain behavioral model complexity requirements,” in *2018 IEEE MTT-S Latin America Microwave Conference (LAMC 2018)*. IEEE, 2018, pp. 1–4.

[5] J. Xu, S. Halder, F. Kharabi, J. McMacken, J. Gering, and D. E. Root, “Global dynamic fet model for gan transistors: Dynafet model validation and comparison to locally tuned models,” in *83rd ARFTG Microwave Measurement Conference*. IEEE, 2014, pp. 1–6.

[6] D. E. Root, “Future device modeling trends,” *IEEE Microwave Magazine*, vol. 13, no. 7, pp. 45–59, 2012.

[7] H. Ryssel and P. Pichler, *Simulation of Semiconductor Devices and Processes*. Springer Science & Business Media, 2012.

[8] F. Lafon, A. Ramanujan, and P. Fernandez-Lopez, “Black box model of integrated circuits for esd behavioral simulation and industrial application case,” *Advanced Electromagnetics*, vol. 4, no. 2, pp. 26–37, 2015.

[9] R. S. Rao, *Microwave engineering*. PHI Learning Pvt. Ltd., 2015.

[10] P. Roblin, D. E. Root, J. Verspecht, Y. Ko, and J. P. Teyssier, “New trends for the nonlinear measurement and modeling of high-power rf transistors and amplifiers with memory effects,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 6, pp. 1964–1978, 2012.

[11] T. Gasseling, D. Barataud, S. Mons, J.-M. Nebus, J.-P. Villotte, J. J. Obregon, and R. Quéré, “Hot small-signal s-parameter measurements of power transistors operating under large-signal conditions in a load-pull environment for the study of nonlinear parametric interactions,” *IEEE Transactions on microwave Theory and Techniques*, vol. 52, no. 3, pp. 805–812, 2004.

[12] A. Elgamal and H. Heuermann, “Design and development of a hot s-parameter measurement system for plasma and magnetron applications,” in *2020 German Microwave Conference (GeMiC)*. IEEE, 2020, pp. 124–127.

[13] D. E. Root, J. Verspecht, D. Sharrit, J. Wood, and A. Cognata, “Broad-band poly-harmonic 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, 2005.

[14] D. Root *et al.*, "Polyharmonic distortion modeling," *IEEE microwave magazine*, vol. 7, no. 3, pp. 44–57, 2006.

[15] P. J. Tasker and J. Benedikt, "Waveform inspired models and the harmonic balance emulator," *IEEE Microwave Magazine*, vol. 12, no. 2, pp. 38–54, 2011.

[16] H. Qi, J. Benedikt, and P. Tasker, "A novel approach for effective import of nonlinear device characteristics into cad for large signal power amplifier design," in *2006 IEEE MTT-S International Microwave Symposium Digest*. IEEE, 2006, pp. 477–480.

[17] R. Saini, J. J. Bell, T. Williams, J. Lees, J. Benedikt, and P. J. Tasker, "Interpolation and extrapolation capabilities of non-linear behavioural models," in *78th ARFTG Microwave Measurement Conference*. IEEE, 2011, pp. 1–4.

[18] T. Husseini, A. Al-Rawachy, J. Benedikt, J. Bell, and P. Tasker, "Global behavioural model generation using coefficients interpolation," in *2019 IEEE MTT-S International Microwave Symposium (IMS)*. IEEE, 2019, pp. 200–203.

[19] E. M. Azad, J. J. Bell, R. Quaglia, J. J. M. Rubio, and P. J. Tasker, "New formulation of cardiff behavioral model including dc bias voltage dependence," *IEEE Microwave and Wireless Components Letters*, vol. 32, no. 6, pp. 607–610, 2022.

[20] H. Wu, "Modeling of power transistors and design of rf and microwave power amplifiers," Ph.D. dissertation, Tianjin University, 2015.

[21] F. Zeng, J. An, G. Zhou, W. Li, H. Wang, T. Duan, L. Jiang, and H. Yu, "A comprehensive review of recent progress on gan high electron mobility transistors: Devices, fabrication and reliability," *Electronics*, vol. 7, no. 12, 2018.

[22] F. Danneville, "Microwave noise and fet devices," *Microwave Magazine IEEE*, vol. 11, no. 6, pp. 53–60, 2010.

[23] M. A. Khan, J. N. Kuznia, A. R. Bhattarai, and D. T. Olson, "Metal semiconductor field effect transistor based on single crystal gan," *Applied Physics Letters*, vol. 62, no. 15, pp. 1786–1787, 1993.

[24] C. H. Wann and K. Noda, "A comparative study of advanced mosfet concepts," *Electron Devices IEEE Transactions on*, vol. 43, no. 10, pp. 1742–1753, 1996.

[25] I. Angelov, H. Zirath, and N. Rosman, "New empirical nonlinear moden for hemt and mesfet devices," *IEEE Trans Microwave Theory Tech*, vol. 40, no. 12, pp. 2258–2266, 1992.

[26] D. Delagebeaudeuf and N. T. Linh, "Metal-(n) algaas-gaas two-dimensional electron gas fet," *IEEE Transactions on Electron Devices*, vol. 29, no. 6, pp. 955–960, 1982.

[27] D. E. Root, J. Verspecht, J. Horn, and M. Marcu, *X-parameters: characterization, modeling, and design of nonlinear RF and microwave components*. Cambridge University Press, 2013.

[28] Z. Mousavirazi, "Advanced component and antenna designs for wireless communication systems." Ph.D. dissertation, Université du Québec, Institut national de la recherche scientifique, 2024.

[29] J. P. Dunsmore, *Handbook of microwave component measurements: with advanced VNA techniques*. John Wiley & Sons, 2020.

[30] J. E. Schutt-Ainé, “Latency insertion method (lim) for the fast transient simulation of large networks,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 1, pp. 81–89, 2001.

[31] P. Beleniotis, C. Zervos, S. Krause, S. Chevtchenko, D. Ritter, and M. Rudolph, “Investigation of traps impact on pae and linearity of algan/gan hemts relying on a combined tcad–compact model approach,” *IEEE Transactions on Electron Devices*, no. 6, p. 71, 2024.

[32] C. Wang, Y. C. Wei, X. Tan, L. Ali, and C. Q. Jin, “Multilayer sinx passivated al2o3 gate dielectric featuring a robust interface for ultralong-lifetime algan/gan hemt,” *Materials Science in Semiconductor Processing*, vol. 135, p. 106038, 2021.

APPENDICES

APPENDIX I

Title of Appendix