CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN SUPERIOR DE ENSENADA, BAJA CALIFORNIA

Sciences Graduate Program in
Electronics and Telecommunications

Modeling of Microwave Transistors Using Artificial Neural Networks

Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Science

Presents:

Andrés Zárate de Landa

Ensenada, Baja California, Mexico
2014

Resumen de la tesis de Andrés Zárate de Landa**,** presentada como requisito parcial para la obtención del grado de Doctor en Ciencias en Electrónica y Telecomunicaciones con orientación en Altas Frecuencias.

Modelado de Transistores de Microondas Utilizando Redes Neuronales

Resumen aprobado por:

Dr. J. Apolinar Reynoso Hernández Dr. Patrick Roblin

Existe un voraz interés en modelos precisos de transistores para el diseño de amplificadores de potencia de RF y microondas. Básicamente los modelos de transistores pueden ser divididos en modelos físicos y empíricos. Los modelos físicos son muy precisos, y como lo indica su nombre, están basados en las características físicas del dispositivo tales como longitud de compuerta, ancho del canal, campo eléctrico, carga y movilidad del electrón, etc. Por otro lado, los modelos empíricos están basados en mediciones, son más sencillos de desarrollar y son computacionalmente eficientes.

El HEMT de AlGaN/GaN es un excelente candidato para ser usado como dispositivo activo en amplificadores de potencia (AP). Tiene una alta densidad de carga y alta velocidad de saturación, lo cual produce altos niveles de potencia de salida. Asimismo, tiene alta movilidad del electrón originando una baja resistividad de encendido, por consiguiente, se pueden obtener altos niveles de eficiencia de potencia agregada. Además, se pueden alcanzar altos voltajes de ruptura como resultado de su amplia banda prohibida, con una alta densidad de carga y un muy amplio rango de temperatura de operación. Todos estos factores indirectamente mejoran la linealidad del AlGaN/GaN. Sin embargo, la tecnología HEMT de AlGaN/GaN aún está bajo desarrollo, por lo tanto, se están investigando modelos confiables de pequeña y gran señal.

En años recientes, las redes neuronales artificiales (RNA) han probado ser una excelente herramienta en la mejora de la precisión de los modelos basados en mediciones debido a sus excelentes propiedades para la aproximación de funciones, teniendo estos modelos mejor desempeño que los basados en tablas los cuales usan funciones spline para interpolación de datos. Por consiguiente, existe una nueva tendencia por utilizar modelos no lineales de RNA basados en, ya sea mediciones lineales, o no lineales del transistor. Al usar una aproximación basada en redes neuronales la cual toma en cuenta la función original y sus derivadas en el proceso de entrenamiento, un modelo no lineal cuasi-estático del FET de GaN basado en mediciones I-V pulsadas y parámetros S pulsados es presentado. El modelo cuasi-estático basado en mediciones es implementado en ADS™ de Agilent y consiste de tres redes neuronales utilizadas para representar las características de corriente y almacenamiento de carga del transistor.

Palabras clave: Modelado no lineal, redes neuronales artificiales, AlGaN/GaN HEMT.

Abstract of the thesis presented by Andrés Zárate de Landa as a partial requirement to obtain the Doctor in Science degree in Electronics and Telecommunications with orientation in High Frequency Electronics.

Modeling of Microwave Transistors Using Artificial Neural Networks

Abstract approved by:

Dr. J. Apolinar Reynoso Hernández   Dr. Patrick Roblin

There is a keen interest in accurate transistor models for the design of RF and microwave power amplifiers, oscillators and mixers. Basically transistor models can be divided in physical and empirical models. Physical models are very accurate and, as it is stated in their name, are based on the physical characteristics of the device such as gate length, channel width, electric field, electron charge, electron mobility, etc. On the other hand empirical models are based on measurements of the device, simpler to develop and computationally efficient.

The AlGaN/GaN HEMT is an excellent candidate to be used as an active device for power amplifiers (PAs). It has high charge density and a high saturation velocity, which produces high output power levels. Besides, it has high electron mobility, originating a low turn on resistivity, and therefore, high power added efficiency levels can be reached. Furthermore, high breakdown voltage levels can be reached as a result of its prohibited band width, with a high charge density, and a very wide range of high temperature operation. All of these factors indirectly improve AlGaN/GaN linearity. However, AlGaN/GaN HEMT technology is still under development, and therefore, reliable small and large signal models are currently being investigated.

In recent years, artificial neural networks (ANNs) have proved to be an excellent aid in improving the accuracy of measurement based models due to their excellent function approximation properties, having these models better performance than table based models that use spline functions for interpolation. Hence, there is a newly born tendency for ANN nonlinear models based on either linear or nonlinear measurements of the transistor. By using a neural network approach that takes into account the original function and its derivatives in the training process, a nonlinear quasi-static model of a GaN FET based on measured pulsed I/V, pulsed S-parameters is introduced. The measurement based quasi-static model is implemented in Agilent's ADS™ and consists of three different neural networks used to represent the current and charge storage characteristics of the transistor.

Keywords: Nonlinear modeling, artificial neural networks, GaN HEMT.

Dedicated to my wife, my mother, my family

and to the loving memory of my father

Acknowledgments and Special Thanks

First of all I have to thank my parents, because I am sure that without their enormous love, support and sacrifice I would have never achieved all of this. My wife Priscilla, for all her support, love, patience and specially, for always believing in me. Your positivism and smile are a ray of sunshine in my life. Of course, I can never forget the rest of my family, my sister Alejandra, Joey, Gael, Gavin, and my new parents Mr. and Mrs. Tentori, thank you for being there every time I needed you, for your love, and specially for putting up with me, I know it is not an easy thing to do.

It is impossible for me to put into words how grateful I am to Prof. J. Apolinar Reynoso Hernández; you have my admiration, respect and deepest affection. You are not only my advisor, but also I consider you now as part of my family. Thank you for everything!

I must express my deepest gratitude to Prof. Patrick Roblin for receiving me as part of his group at The Ohio State University, and for all of his advice and help throughout this thesis. I really admire not only how much knowledgeable you are but also your great kindness. I hope this is not the end of a joint work, but the marking stone for future collaboration.

I would like to thank CINVESTAV Guadalajara especially to Dr. Loo Yau's group for all his collaboration in the measurements needed in this work. To my colleagues Manuel Pulido and Jonathan Monjardín thank you for helping me in the lab and for the enriching discussions. Additionally, I deeply appreciate the time, availability and insights of the committee members, Dr. Raúl Loo, Dr. Arturo Velázquez, Dr. María del Carmen Maya, and Dr. Francisco Hirata. Your thoughts and comments helped making this a quality work.

And finally, but not less important, I would like to thank CONACYT and CICESE for the financial support throughout these years.

Contents

Resumen ii
Abstract iii
Acknowledgments and Special Thanks v
List of Figures x
List of Tables xiii
Chapter 1 - Introduction 1
1.1 Transistor Modeling Process 6
1.2 Objective 7
1.3 Methodology and Dissertation Organization 8
Chapter 2 - AlGaN/GaN HEMT Technology 10
2.1 Introduction 10
2.2 AlGaN/GaN HEMT Material 11
2.3 AlGaN/GaN HEMT Material 13
2.3.1 Polarization Effects in AlGaN/GaN HEMT 14
2.3.2 Trapping in GaN HEMTs 15
2.3.2.1 Surface States (Traps) 16
2.3.2.2 Buffer Traps 17
Chapter 3 - Compact Modeling Basics for FETs 19
3.1 Introduction 19
3.2 Physical Models 21
3.3 Compact Models 22
3.3.1 Measurement-Based Equivalent Circuit Models 23
3.3.2 Physically-Based Equivalent Circuit Models 25
3.3.3 Modeling Approach Implemented 28
3.4 Memory Effects 29
3.4.1 Short-Term Memory Effects 31

3.4.2 Long-Term Memory Effects 32
3.4.2.1 Thermal Effects 32
3.4.2.2 Charge Trapping 33
3.4.2.3 DC Bias Network 34
Chapter 4 - Artificial Neural Networks 36
4.1 Introduction 36
4.1.1 Benefits of Neural Networks 37
4.1.2 Human Brain 40
4.1.3 Artificial Neural Networks 42
4.1.4 Models of a Neuron. 44
4.1.4.1 Types of Activation Function. 46
4.2 Feedforward Networks 49
4.2.1 Multilayer Perceptrons. 49
4.2.1.1 Preface. 51
4.2.1.2 Forward Propagation. 54
4.2.1.3 Update Rule 55
4.2.1.4 Back-Propagation Algorithm. 55
4.3 Training Algorithms. 58
4.3.1 Steepest Descent Method. 59
4.3.2 Steepest Descent Method with Momentum Learning. 61
4.3.3 Improved Resilient Back-propagation (iRprop) Algorithm. 62
4.3.4 Levenberg-Marquardt Algorithm. 64
4.3.5 Optimized Levenberg-Marquardt with Adaptive Momentum (OLMAM) Algorithm. 64
4.4 Approximation of Functions. 66
4.5 Modified Backpropagation Algorithm. 66
Chapter 5 - Small Signal Modeling 71
5.1 Introduction 71
5.2 Package Parasitics in Power FETs 72

5.2.1 Analysis of an Empty Package. 73
5.3 The Forward Cold-FET Model..... 75
5.3.1 Open Drain Condition..... 77
5.3.2 Drain and Source Parasitic Resistances and Inductances Calculation. 78
5.3.3 Gate Parasitic Resistance and Inductance Calculation. 80
5.3.3.1 Frequency Points of Interest 80
5.3.3.2 ω_0 Calculation 82
5.3.3.3 R_0 and C_0 Calculation 84
5.3.3.4 R_g Calculation 85
5.3.3.5 L_g Calculation 86
5.3.4 Disadvantages of the Method. 87
5.3.4.1 Smoothing Derivative Data Using Artificial Neural Networks. 87
5.3.4.2 Low Frequency Dependence. 89
5.4 The Pinched-off Cold-FET Model 91
5.5 De-embedding Process 93
5.6 Intrinsic Elements Extraction 94
Chapter 6 - Large Signal Modeling 98
6.1 Introduction 98
6.2 The Intrinsic Nonlinear Model 98
6.2.1 Measurement Approach to Nonlinear Modeling 100
6.3 Neural Network Model of the Drain Current 102
6.4 Charge Modeling 109
6.4.1 Charge Conservation Law 109
6.4.2 Neural Network Model of the Gate and Drain Charge Functions 111
Chapter 7 - Model Validation 116
7.1 Introduction 116
7.2 Measurement Techniques for Model Extraction 116
7.2.1 Pulsed I-V, Pulsed S-parameter Measurement System. 117

7.3 Implementation in Agilent ADS® 120
7.4 Model Simulation and Results 123
7.4.1 I-V Characteristics. 123
7.4.2 RF Characteristics. 123
Conclusions and Future Work 127
Contributions 130
Journal Publication and Proceeding Articles 130
Bibliographic References 131

List of Figures

Figure 1. Flow chart illustrating the distinct processes required to generate a model. 7
Figure 2. Investigation methodology block diagram. 9
Figure 3. (a) Simplified AlGaAs/GaAs HEMT structure, (b) corresponding band diagram. (Jarndal, 2006 p. 10) 10
Figure 4. (a) Simplified AlGaN/GaN HEMT structure, (b) corresponding band diagram. (Jarndal, 2006 p. 11) 11
Figure 5. Electronic properties of AlGaN/GaN HEMT structure. 12
Figure 6. AlGaN/GaN HEMT structure, showing polarization induced and 2DEG charges. (Jarndal, 2006, p. 15) 15
Figure 7. AlGaN/GaN HEMT structure, showing surface and buffer trapps. (Roblin,2011 p. 112)... 15
Figure 8. Kink effect in DC characteristics. 18
Figure 9. A general modeling hierarchy applied to the level of abstraction of the design. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 52) 20
Figure 10. Small-signal equivalent circuit model of the intrinsic transistor. 24
Figure 11. Large-signal equivalent circuit (compact) model from Curtice et al 1999. 24
Figure 12. Two-zone electron velocity vs electric field relationship used in physically-based models for short gate-length FETs. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 58) 26
Figure 13. Physical origins of the components of the equivalent circuit model of a MESFET shown in Fig. 10. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 60) 28
Figure 14. Two-port representation adopted for the large-signal FET model used in this thesis. ... 29
Figure 15. Origins of short-term and long-term memory effects in a transistor circuit. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 63) 30
Figure 16. Long term memory effect on the compression characteristic of a power amplifier. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 67) 33
Figure 17. Block diagram representation of the nervous system. 41
Figure 18. Types of artificial neural networks. (a) Single layer perceptron (b) Linear neuron (c) Multilayer perceptron (d) Competitive network (e) self-organizing map (f) Recurrent network (g). 43
Figure 19. Simplified diagram of ANN's training procedure. 44
Figure 20. Nonlinear model of a neuron 45
Figure 21. Threshold function 46
Figure 22. Piecewise-linear function 47
Figure 23. Examples of sigmoid functions 48
Figure 24. Architectural graph of a multilayer perceptron with two hidden layers. 51
Figure 25. Directions of two basic signal flows in an MLP. 52
Figure 26. Three layer neural network architecture. 53
Figure 27. Search using gradient information (Principe et al, 2000). 60

Figure 28. Momentum learning (figure extracted from Principe et al, 2000) 62
Figure 29. Three layer artificial neural network architecture with derivatives information. (Zárate-de Landa et al, 2012) 67
Figure 30. Comparison between classical backpropagation and proposed backpropagation training. 70
Figure 31. Illustrations of a power transistor using ceramic and plastic packages. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 124) 73
Figure 32. CREE® AlGaN/GaN HEMT CGH35015 on 440166 package. (Monjardín, 2014) 74
Figure 33. Transistor package equivalent circuit model. 74
Figure 34. Comparison between measured (circles) and modeled (lines) of CREE 440166 package. 75
Figure 35. FET Small-signal equivalent circuit model. 76
Figure 36. Cold-FET Small-signal equivalent circuit under low DC gate forward current with floating drain. (Zárate de Landa et al, 2007, 2009) 77
Figure 37. Cold-FET Small-signal equivalent circuit under low DC gate forward current with zero drain-source condition. 77
Figure 38. Comparison between measured and simulated real part of Z_{12} and Z_{22}. 79
Figure 39. Comparison between measured and simulated imaginary part of Z_{12} and Z_{22}. 79
Figure 40. Real part of Z_{11}, of the Cold-FET and its derivative with respect to the frequency. 81
Figure 41. Imaginary part of Z_{11}.of the Cold-FET and its derivative with respect to the frequency. 81
Figure 42. Intersection between the second derivative of the real and imaginary part of Z_{11}. 84
Figure 43. Real part of Z_{11} of the Cold-FET and its first derivative with respect to the frequency. Measured data (dotted line) and Neural Network model (solid line). 88
Figure 44. Imaginary part of Z_{11} of the Cold-FET and its first derivative with respect to the frequency. Measured data (dotted line) and Neural Network model (solid line). 88
Figure 45. Comparison between the measured and simulated real part of Z_{11}. 90
Figure 46. Comparison between the measured and simulated imaginary part of Z_{11}. 90
Figure 47. Equivalent Circuit Model of the transistor under pinch-off bias (Vgs << Vp; Vds = 0). ... 91
Figure 48. Transformation from T to π network for the pinch-off equivalent circuit model. 91
Figure 49. Comparison between measured (circles) and modeled (lines) Y-parameters of the pinched-off Cold-FET (CREE CGH35015F AlGaN/GaN HEMT). 93
Figure 50. De-embedding process used for the computation of the intrinsic Y-parameters. 94
Figure 51. Small-signal equivalent circuit model of the intrinsic transistor. 95
Figure 52. Physical origins of the intrinsic elements of a FET. 95
Figure 53. Intrinsic elements calculated at different bias points. 97
Figure 54. FET model schematic diagram showing the package capacitances, parasitic elements and the intrinsic quasi-static transistor model. 100
Figure 55. Extrinsic (gridded) and corresponding (non-gridded) voltage domain of a FET. 103
Figure 56. FET model IV constitutive relation expressed as functions of (a) extrinsic and (b) intrinsic voltages. 104

Figure 57. Four-layer ANN used to model the drain current. 106
Figure 58. Comparison between measured (circles) and ANN modeled (solid lines) drain current. 106
Figure 59. Comparison between measured (circles) and ANN modeled (solid lines) output conductance. 107
Figure 60. Comparison between measured (circles) and ANN modeled (solid lines) transconductance. 107
Figure 61. Comparison between measured (circles) and ANN modeled (solid lines) of the first derivative of g_m. 108
Figure 62. Comparison between measured (circles) and ANN modeled (solid lines) of the second derivative of g_m. 108
Figure 63. Two-dimensional electric field. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 226) 111
Figure 64. ANN structure and cost function used to train the gate charge. 113
Figure 65. ANN structure and cost function used to train the drain charge. 113
Figure 66. Gate charge function obtained from ANN. 114
Figure 67. Drain charge function obtained from ANN. 114
Figure 68. Conservation of charge at the gate. 115
Figure 69. Conservation of charge at the drain. 115
Figure 70. Transistor modeling flow. 117
Figure 71. Measurement setup for pulsed I-V, pulsed S-parameters. (Extracted from Auriga Microwave data sheet). 118
Figure 72. Comparison of static I-V (solid red line) and pulsed I-V for a GaN HEMT. Pulse conditions were 1 µs pulse width and 1 ms separation with quiescent bias points V_{dsq} = 0 V, V_{gsq} = -3 V (blue pointed line) and V_{dsq} = 40 V, V_{gsq} = -3 V (dashed green line). 119
Figure 73. Example of a two-port symbolically-defined device component. 120
Figure 74. Model implementation in Agilent's ADS. 122
Figure 75. I-V Comparison between measurements and simulations for the CGH35015 AlGaN/GaN quasi-static model. 123
Figure 76. AM-AM measurement setup. 124
Figure 77. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -2 V and V_{ds} = 15 V. 125
Figure 78. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -1.6 V and V_{ds} = 15 V. 125
Figure 79. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -1.4 V and V_{ds} = 15 V. 126
Figure 80. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -1.6 V and V_{ds} = 20 V. 126

List of Tables

Table 1. Advantages of GaN over other semiconductors (Jarndal 2006) 13
Table 2. Package parasitic capacitances calculated for CREE CGH35015F AlGaN/GaN HEMT......... 75
Table 3. Parasitic elements of the forward Cold-FET calculated of the CREE CGH35015F AlGaN/GaN HEMT.. 91
Table 4. Parasitic capacitances calculated of the CREE CGH35015F AlGaN/GaN HEMT.................. 93

Chapter 1 - Introduction

Nowadays, the idea of being isolated without television, internet, cell phones, etc, can't be conceived. This is the result of the huge peak the communication systems have had in the last few years. Besides the communication services boost the technological development, they also represent the economic development of a country. Hence, the demand of these services has increased in such an unimaginable way, giving rise to a vertiginous evolution of the communication systems, which every day are more sophisticated.

Modern communication systems of third and fourth generation use different multiple access modulation diagrams such as: FDMA (Frequency Division Multiple Access), TDMA (Time Division Multiple Access), CDMA (Code Division Multiple Access). Thus, the power, linearity and efficiency of the radio frequency power amplifiers, located at the base stations, are essential characteristics required to ensure reliability and the quality of the services provided by modern communication systems. However, it is difficult to have all these characteristics satisfied at the same time. In this sense, LINC (Linear Amplification using Nonlinear Components) transmitters have been proposed, which utilize power amplifiers (nonlinear) and linearization circuits. LINC transmitters can obtain high efficiency and nice linearity. In these transmitters, the power amplifier operates at a saturation output power level and this level is adjusted by the operation voltage control of the active device (Stengel and Eisenstadt, 2000), (Woo Y. Y., 2003).

In particular, wireless communication systems have considerably evolved, in part, due to the interest of designers of wireless communication systems to improve, protocols and communication channels, as well as, the devices that perform this task. The system that performs the communication function, in other words, the system that sends information from one point to another point using free space as media, is the transmitter/receptor.

The power amplifier (PA) is the most important element in transmitters. However, this element consumes a large quantity of energy. Because of this fact, it is desirable to obtain a PA which do not consumes a lot of energy; this is accomplished by improving its efficiency. One important reason to improve PA's efficiency is to lower energy losses by dissipation, provoking at the same time, space reduction and an improvement of the capacity of the dissipation system, and as a consequence, a cost reduction of the PA. Another important point is that wireless communication systems need to administrate the battery energy in order to extend the working time of the device. Nevertheless, by improving the PA efficiency, linearity is diminished as well as the gain. This is why PA design is a big challenge and, hence, it is needed to define the PA requirements in order to establish a commitment between linearity, efficiency and output power.

In order to face this challenge, it is necessary to research and study high efficiency amplifiers, which are classified as class F and class E amplifiers. Because of their operation, these types of amplifiers present efficiencies of 100% and they would seem ideal to be used in applications where efficiency is the parameter to be optimized. However, these amplifiers have the enormous disadvantage of being nonlinear, and in order to improve their linearity, linearization methods are used like EER (Envelope and Restoration) or the LINC.

On the other side, class E amplifiers have limitations to operate at high frequencies. Meanwhile, class F amplifiers can be designed to operate at higher frequencies, because basically, it is a saturated class B amplifier. A class F amplifier can reach a maximum level of efficiency of 100% theoretically. This amplifier was analyzed by Snider in 1967, who defined it as a class B amplifier with optimum efficiency where it is characterized by a square wave shape voltage signal, and by a sinusoidal half wave rectified from the drain current signal. The load network must present a short circuit at the output of the active device for even harmonics and an open circuit for odd harmonics, where the voltage wave is formed by odd harmonics only and the current wave is formed by even harmonics. Next, Raab in 1997 performs a complementary analysis and demonstrates that an efficiency of 77% can be achieved by knowing the load impedances of the first three harmonics. Subsequently, Colantonio (2001) introduces the idea of the voltage or current wave forms

can be formed by even or odd harmonics, that is to say, even and odd harmonics are present on either wave form at the same time, therefore efficiency is improved. Based on Colantonio's work, Raab (2001 carried out a complementary analysis finds the maximum efficiency (81%) and output power that can be obtained by a class F amplifier with the first three harmonics control.

Today, there are several technologies competing in the high efficiency power amplifiers market (Class E and Class F) based on field effect transistors. These technologies are:

- Silicon FET-LDMOS,
- Gallium Nitride (GaN) FET,
- Silicon Carbide (SiC) FET.

AlGaN/GaN HEMT is an excellent candidate to be used as an active device for base station PAs. It has high charge density and a high saturation velocity, which produces high output power levels. It also has high electron mobility, originating a low turn on resistivity, and therefore, high efficiency levels can be reached. Besides, high breakdown voltage levels can be reached as a result of its prohibited band width, with a high charge density, and a very wide range of high temperature operation. All of these factors indirectly improve AlGaN/GaN linearity. In high efficiency power amplifiers, parasitic elements of the transistor have a negative impact on the efficiency due to their contribution in the determination of the load impedance (Loo-Yau, 2007).

Development of high efficiency PAs goes with the development of accurate nonlinear models of the active device. In spite of this, an accurate enough model to predict the behavior of AlGaN/GaN transistors hasn't been developed yet. It is a goal of this project to investigate the nonlinear model of AlGaN/GaN transistors, by means of its equivalent electrical circuit. To evaluate power amplifiers performance equivalent circuit type nonlinear models are used, as well as AM-AM, AM-PM and two tones measures. The equivalent electrical circuit can be divided in three parts; one is formed by parasitic or extrinsic elements, which are bias independent but dependent of the package. The second part is formed by intrinsic elements, which are dependent of the voltage applied to the

transistor terminals and the fabrication technology. Finally, the third part is formed by a I(V) nonlinear current source. At CICESE's electronics and telecommunications department, original methods have been developed to extract parasitic elements of field effect transistors based on technologies such as GaAs and GaN. Besides, methods to improve nonlinear models based on GaAs have also been developed. Since 2007, we have started to study emergent technologies of power FETs based on Gallium Nitride and Silicon Carbide (Zárate-de Landa et al 2007), (Reynoso-Hernández 2008). On the other hand, the most popular method used to determine the intrinsic elements is the method developed by Berroth, and Bosh (1990). This method consists on finding analytical expressions, dependent on the admittance parameters of the intrinsic elements (R_i, R_{gd}, R_{ds}, C_{gs}, C_{gd}, τ, g_{m0}) of the equivalent electrical circuit model. To evaluate the GaN FET an appropriate topology of the electrical circuit is required, in order to obtain the best equivalent electrical circuit (linear or nonlinear). There are several topologies where the location of the parasitic capacitances is the main difference between them. These capacitances depend on the geometry and the package of the transistor.

In the process of modeling the transistor, first the parasitic elements are obtained. Then, after a de-embedding process to eliminate the parasitic elements, the intrinsic elements are extracted. A good extraction of the device extrinsic elements, along with an adequate topology, lead to real values of the intrinsic transistor, and hence, to reliable models.

Nonlinear FET equivalent circuit models are used in computer-aided-design (CAD) simulators for evaluating the performance of mixers, oscillators, power amplifiers, etc. Nonlinear FET models consist of two parts: the first is formed with the static or pulsed I-V characteristics and the second part is formed with the reactive elements, gate-source capacitance $C_{gs}(V_{gs},V_{ds})$, gate-drain capacitance $C_{gd}(V_{gs},V_{ds})$, and drain-source capacitance $C_{ds}(V_{gs},V_{ds})$. In the last 30 years, several models have been proposed with the main goal of describe the nonlinearities present in FET devices such as current dispersion due to superficial and buffer traps, temperature dependence, and intermodulation distortion (IMD). Basically, these models can be divided in two categories: physical and empirical. Physical models are based, as it is stated in their name, in the physical characteristics of the

device such as gate length, channel width, electric field, electron charge, electron mobility, etc (Delagebeaudeuf, 1982), (Statz, 1987), (Roblin, 1990). These models, besides describing the transistor operation in terms of device physics, have the property to get a very good agreement between simulated and measured data. However, sometimes it is difficult to have control on the dimensions and physical properties of the transistor, making this approach not suitable for circuit designers who have no control over the fabrication process.

On the other hand, empirical models eventhough they have no meaning related with the physics of the device, are simple to implement, computationally efficient and based on measured data, being these characteristics very convenient for circuit designers in the industry. Empirical models can be developed using analytical equations that describe measured data, these models are also named "equation-based analytical models." (Curtice et al, 1985), (Materka et al, 1985), (Angelov et al, 1992), (Cabral et al, 2004). Analytical models are less accurate than physics based models and are technology independent. Also, parameter extraction can be complex and the range of validity is limited. Another approach is analytical models based on the use of power, Volterra or Chebyshev series (Närhi T, 1996), although the difficulty in these methodologies is increased, a better prediction of the transistor behavior can be obtained taking advantage of the nonlinearities and memory effects capture of these approximations. Table based models interpolate the extracted small signal parameters using techniques like B-splines. These models require large memory storage and computational time which is less practical (Root et al, 1991), (Kompa, 1994), (Schreurs, 1996).

Recently, a new analytical model approach has emerged which takes advantage of artificial neural networks (ANN) (Shirakawa et al, 1998), (Zhang et al., 2000), (Xu et al, 2003, 2006), (Cao et al, 2009). In this technology, the equivalent circuit model is considered a black box. Then, based on input/output characteristics (measured data) the neural network is trained, giving as a result, a "small brain" which learned to behave like a transistor. Once the ANN is trained, it will be capable of accurately simulate the output for any given input even if it wasn't part of the training. Large signal models using this technology take advantage of the nonlinearities inherent in the ANN structure making them

an excellent tool in the design of nonlinear microwave circuits such as high power amplifiers, oscillators and mixers. These models can learn the behavior of the device which is an advantage over equation based models. However, the main drawback is the need of very good measurements for training (and so the method is sensitive to technological dispersion). This thesis is focused on the design of FET models using artificial neural networks.

1.1 Transistor Modeling Process

In this section the overall processes to follow for model development and provide some insight into the tasks required is presented. It is these tasks that are to be presented in detail in subsequent chapters. In general the process of developing and extracting a model is shown in Fig. 1.

At the beginning of the modeling process it is expected that the application of the model and the ranges within which it is expected to operate are defined. For example, the requirements of model to be used in a high-power Doherty or Class F are very different from those for a low-power Class A design. Although it is our objective to generate a model with a wide range of applicability, the model must function in its intended application. From our experience, this step of defining the model is very time-consuming and is often overlooked. A clear and complete definition of the scope and deliverables of a project should be received before moving into the execution phase; this is fundamental to project management methodology. The time taken to investigate and document the model specifications is well invested, as it forces the modeling engineer and the customer to agree upon a set of objectives and validation criteria for the model. The type of model that is required is determined from these discussions. Without completing this task, which is simple in concept, but difficult to achieve in practice, there is a high probability that a model will not match the desired application or expectations of the customer. Once the model topology has been finalized, data need to be gathered either through measurement or simulation. Typically, the frequency range, impedance ranges, and power levels are used for the specification of nonlinear models. For linear devices the frequency range and parametric variations of the geometry are often specified.

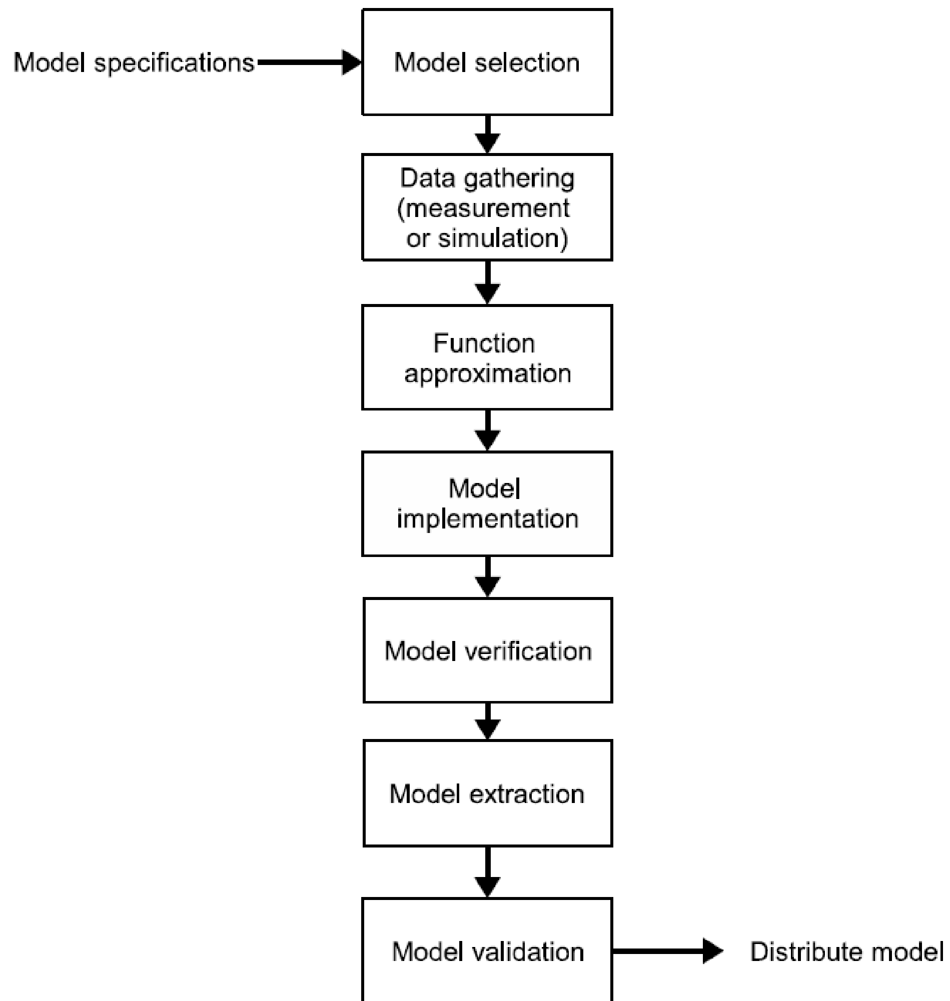


Figure 1. Flow chart illustrating the distinct processes required to generate a model.

1.2 Objective

The main objective of this thesis is to develop a model for AlGaN/GaN FETs based on artificial neural networks which will be capable of accurately reproducing the I/V, Q/V characteristics not just of the fundamental, but also up to the third order harmonics. First a reliable extraction of the parasitic elements will be performed, then, based on the admittance parameters of the intrinsic transistor and pulsed I-V measurements a neural network model will be defined to predict the nonlinearities inherent to the device.

Therefore, it is expected that the model developed in this project can help the circuit designer to develop a high efficiency power amplifier.

1.3 Methodology and Dissertation Organization

Fig. 2 displays a block diagram describing the methodology followed through this research. Every point will be explained thoroughly in the following chapters contained in this thesis.

This thesis is organized as follows: In Chapter II a description of the AlGaN/GaN HEMT technology and physical properties of the device will be presented. A detailed discussion of the model analysis, extraction and construction choices is presented. We shall shift gears in Chapter III where artificial neural networks are studied and a modification of the classical backpropagation algorithm is introduced. In Chapter IV the extraction process of the small-signal model is described. Original methods for the extraction of parasitic elements of transistors are introduced. Chapter V covers the development of a measurement based quasi-static model of the transistor. In Chapter VI the compact model is implemented in Agilent's ADS® and validation results are presented. Finally, Chapter VII is used to present the conclusions and contributions of this investigation.

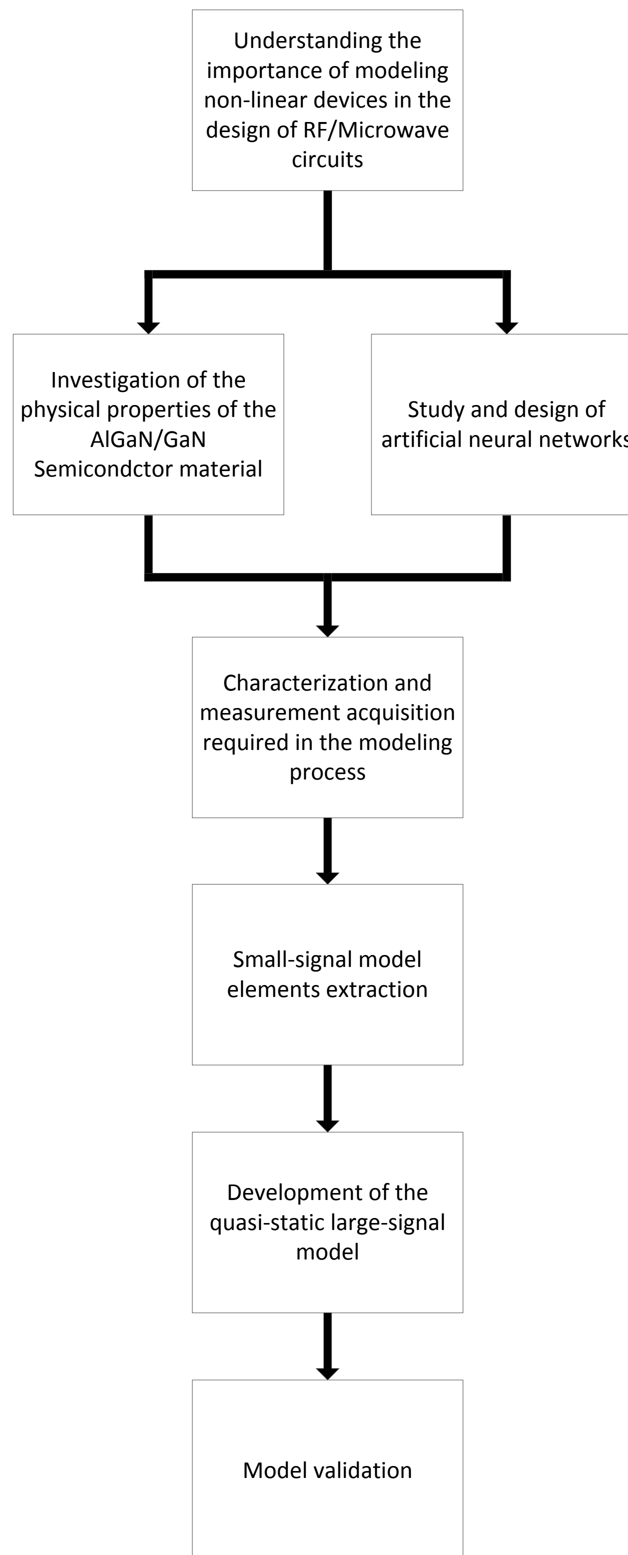


Figure 2. Investigation methodology block diagram.

Chapter 2 - AlGaN/GaN HEMT Technology

2.1 Introduction

The high electron mobility transistor (HEMT) is a heterostructure field effect transistor. The term "HEMT" is applied to the device because the structure takes advantages of superior transport properties of electrons in a potential well of highly doped semiconductor material.

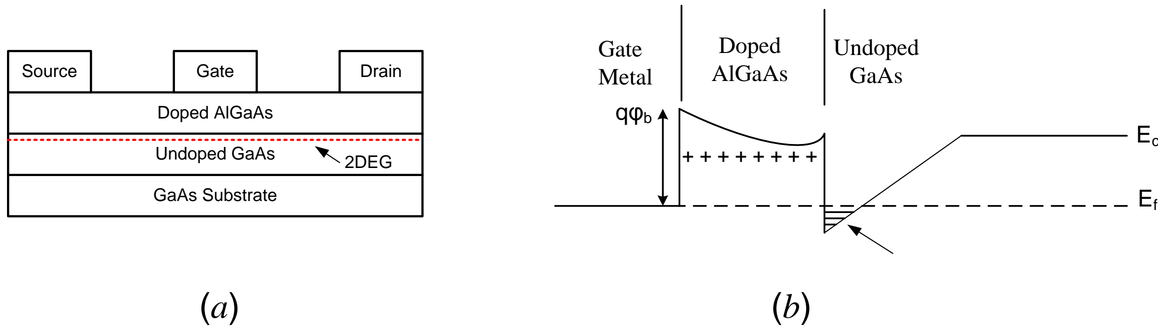


Figure 3. (a) Simplified AlGaAs/GaAs HEMT structure, (b) corresponding band diagram. (Jarndal, 2006 p. 10)

As shown in Fig. 3a, a wide bandgap semiconductor material (doped AlGaAs) lies on a narrow band material (undoped GaAs). The band diagram of correlated structure is shown in Fig. 3b. A sharp dip in the conduction band edge occurs at the AlGaAs/GaAs interface. This results in higher carrier concentration in a narrow region quantum well in the source-drain direction. The distribution of electron in the quantum well is essentially two-dimensional due to the small thickness of the quantum well in comparison to the width and length of the channel. Therefore, the charge density is called a two dimensional electron gas (2DEG) and quantified in terms of sheet carrier density n_s.

AlGaN/GaN HEMT has been fabricated in a similar way using doped or undoped AlGaN layer as shown in Fig 4a. It has been observed that a 2DEG is formed in the AlGaN/GaN interface even when there is no intentional doping of AlGaN layer. It has also

been observed that when the AlGaN layer is intentionally doped, the charge density in the 2DEG is not proportional to the amount of doping.

In AlGaN/GaN HEMT, the formation mechanism of the 2DEG at the heterointerface is different with that in the AlGaAs/GaAs HEMT. Due to the presence of a strong polarization field across the AlGaN/GaN heterojunction, a 2DEG with the sheet carrier density up to $10^{13}cm^{-2}$ can be achieved without any doping (Ambacher et al., 1999). Ibbetson et al (2000) found that surface states act as a source of electron in 2DEG. The built-in static electric field in the AlGaN layer induced by spontaneous and piezoelectric polarization greatly alters the band diagram and the electron distribution of thee AlGaN/GaN heterostructure. Thus, considerable number of electrons transfer from the surface states to the AlGaN/GaN heterointerface, leading to a 2DEG with high density. The band diagram of the structure shown in Fig. 4a is illustrated in Fig. 4b.

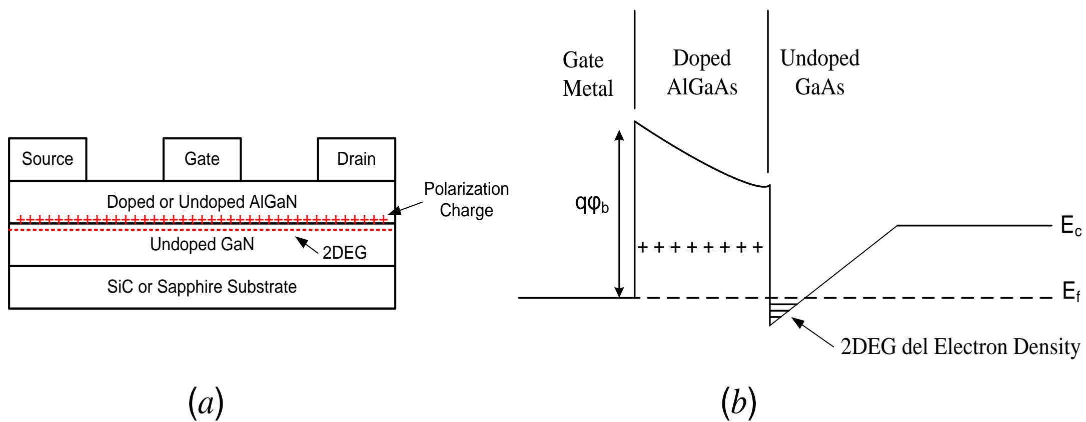


Figure 4. (a) Simplified AlGaN/GaN HEMT structure, (b) corresponding band diagram. (Jarndal, 2006 p. 11)

2.2 AlGaN/GaN HEMT Material

GaN material possesses fundamental electronic properties that make it an ideal candidate for high power microwave devices (Eastman and Mishra, 2002). As a wide bandgap material (E_g = 3.4 eV), GaN has very high electric breakdown field (E_{br} > 2MV/cm). As a result, GaN based devices can be biased at very high drain voltage (V_{br} > 50 V). It can also be operated at higher channel temperature (> 300 ^{o}C). It also possesses

high saturation electron velocity (2×10^7 cm/s), which contributes to higher current density while $I_{max} \approx qn_sv_s$ where q is the electron charge (1.6×10^{-19} Coulomb), n_s is the carrier density, and v_s is the electron saturation velocity. Furthermore, GaN induces high frequency operating while $f_T \approx v_s/L_{eff}$.

AlGaN/GaN heterostructure comprises 1) high sheet carrier density ($n_s \approx 1 \times 10^{13}$ cm^{-2}), which produces high I_{max}, and 2) high electron mobility (μ = 1200 – 1500 cm^2/Vs), which is largely responsible for low on-resistance (low-knee voltage) since the channel resistance is related to $1/(qn_s\mu E)$ at low electric field. Consequently, AlGaN/GaN HEMT can achieve very high breakdown voltage, very high current density and sustain very high channel operating temperature. Furthermore, high operating frequency (f_T) and high drain power added efficiency (PAE) could be achieved. Fig.5 illustrates the relationship of the electronic characteristics mentioned and device properties, meanwhile, the material properties of GaN compared to the competing materials is presented in Table 1. It can be observed that GaN has greater advantage over conventional semiconductors.

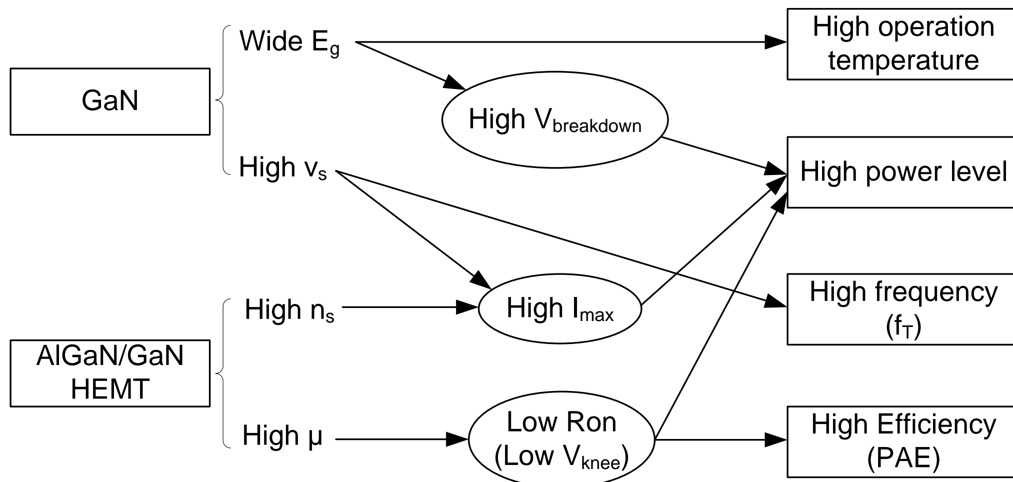


Figure 5. Electronic properties of AlGaN/GaN HEMT structure.

Table 1. Advantages of GaN over other semiconductors (Eastman and Mishra, 2002, p. 31)

Gallium Nitride Advantages Over Other Semiconductor Materials						
Most Common Semiconductor		**Silicon**	**Gallium Arsenide (AlGaAs/ InGaAs)**	**Indium Phosphide (InAlAs/ InGaAs)**	**Silicon Carbide (SiC)**	**Gallium Nitride (AlGaN/ GaN)**
Characteristic	**Units**					
Band Gap	eV	1.11	1.43	1.35	3.2	**3.49**
Electron Mobility @ 300°K	cm^2/Vs	1500	8500	5400	700	**1000 - 2000**
Peak Saturation Velocity	$X10^7$ cm/s	1.0	1.3	1.0	2.0	**2.5**
Breakdown Voltage	MV/cm	0.3	0.4	0.5	3.0	**3.0**
Termal Conductivity	W/cm*K	1.5	0.5	0.7	4.5	**>1.5**
Relative Dielectric Constant	ε_r	11.8	12.8	12.5	10.0	**9.0**

2.3 AlGaN/GaN HEMT Material

As mentioned in section II.1, AlGaN/GaN HEMT exhibits large polarization effects, which origins in the high polarity of the GaN material itself and the larger lattice constant difference between GaN and AlGaN. Nowadays, different researchers have shown that the formation of 2DEG in undoped and doped AlGaN/GaN structure relies on these effects. In this section the relation between polarization effects and the formation of the 2DEG channel will be explained.

2.3.1 Polarization Effects in AlGaN/GaN HEMT

Polarization effects in AlGaN/GaN HEMT include spontaneous and piezoelectric polarization. The spontaneous polarization refers to the built-in polarization field present in an unstrained crystal. This electric field exists because the crystal lacks inversion symmetry, and the bond between the two atoms is not purely covalent. This results in a displacement of the electron charge cloud towards one atom in the bond. In the direction along which the crystal lacks inversion symmetry, the asymmetric electron cloud results in a net of positive charge located at one face of the crystal and a negative charged net at the other face. The electric field and sheet charges present in a Ga-face crystal of GaN and AlGaN grown on c-plane is illustrated in Fig. 6.

On the other hand, piezoelectric polarization is the presence of a polarization field resulting from the distortion of the crystal lattice. Due to the large difference in lattice constant between AlGaN and GaN materials, the AlGaN layer, which is grown on the GaN buffer layer is strained. Due to the large value of the piezoelectric coefficients of these materials, this strain results in a sheet charge at the two faces of AlGaN layer. The total polarization field in the AlGaN layer depends on the orientation of the GaN crystal. MOCVD (Metal Organic Chemical Vapour Deposition) produces GaN crystal orientation that makes the sheet charges caused by spontaneous and piezoelectric polarizations added constructively. Therefore, the polarization field in the AlGaN layer will be higher than that in the buffer layer. Due to this discontinuity of the polarization field, a very high positive charge will be presented at the AlGaN/GaN interface as illustrated in Fig. 6.

As the thickness of the AlGaN layer increases during the growth process, the crystal energy will also increase. Beyond a certain thickness the internal electric field becomes high enough to ionize donor states at the surface and cause electrons to drift toward the AlGaN/GaN interface. As the electrons move from the surface to the interface, the magnitude of the electric field is reduced, thereby acting as a feedback mechanism to diminish the electron transfer process. Under equilibrium condition, a 2DEG charge at the interface will be generated due to the transferred electrons and a positive charge on the surface will be formed from ionized donors as illustrated in Fig. 6.

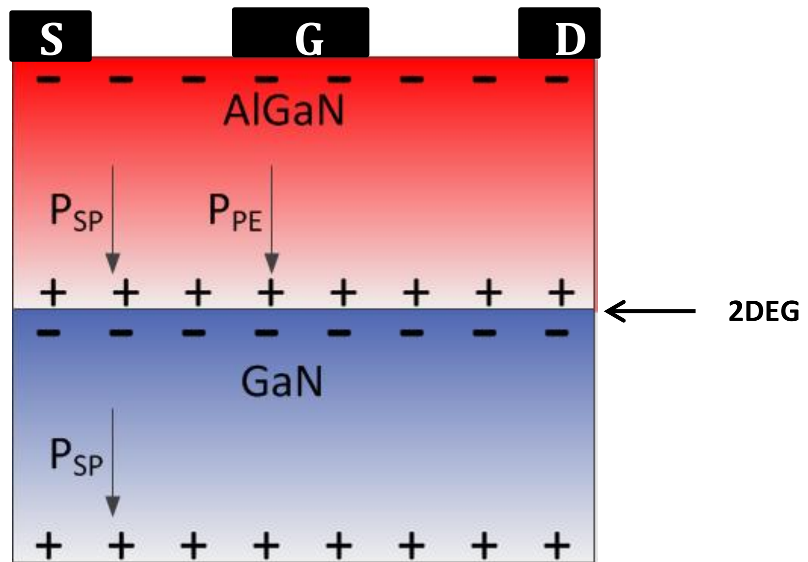


Figure 6. AlGaN/GaN HEMT structure, showing polarization induced and 2DEG charges. (Jarndal, 2006, p. 15)

2.3.2 Trapping in GaN HEMTs

Trapping is a major source of memory effects in some devices. Fig. 7 shows various trapping mechanisms potentially taking place in these devices. They include trapping at the AlGaN surface and trapping in the buffer, which affect, respectively, the source and drain resistance and the threshold voltage (Roblin 2011).

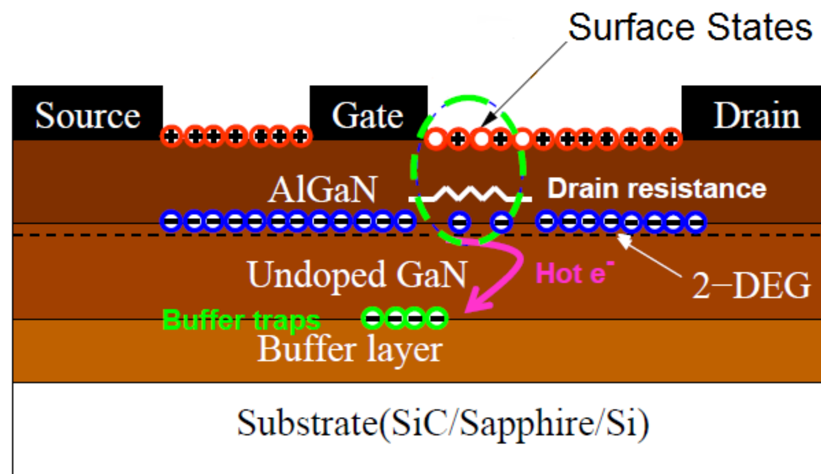


Figure 7. AlGaN/GaN HEMT structure, showing surface and buffer trapps. (Roblin,2011 p. 112).

2.3.2.1 Surface States (Traps)

The mechanism of formation of charged surface states and the importance of these states for generation of the 2DEG channel in AlGaN/GaN HEMT was investigated by Ibbetson et al (1979). For non-ideal surface with available donor-like states, the energy of these states will increase with increasing AlGaN thickness. At certain thickness the energy states reach the Fermi level and electrons are then able to transfer from occupied surface states to empty conduction band states at the interface, creating 2DEG and leaving behind positive surface sheet charge. For ideal surface with no surface states, the only available occupied states are in the valence band. Here, the 2DEG exists as long as the AlGaN layer is thick enough to allow the valence band to reach Fermi level at the surface. Electrons can then transfer from AlGaN valence band to the GaN conduction band, leaving behind surface holes. These accumulated holes produce a surface positive sheet charge. This means that in all cases, a positive sheet charge at the surface must exist in order for the 2DEG to be present in the AlGaN/GaN interface.

The surface states act as electron traps located in the access regions between the metal contacts. Proper surface passivation prevents the surface states from being neutralized by trapped electrons and therefore maintains the positive surface charge. If the passivation process is imperfect, then electrons, leaking from the gate metal under the influence of a large electric field present during high power operation, can get trapped (Vetury et al, 2001). The reduction in the surface charge due to the trapped electrons will produce a corresponding reduction in the 2DEG charge, and therefore reduce the channel current. The amount of trapped electrons and therefore the current reduction depends on the applied bias voltages and the extent to which the device is overdriven beyond the linear gain. The trapped electrons are modulated with the low frequency stimulating voltages and therefore can contribute to the 2DEG channel current. However, they cannot follow the high frequency stimulating voltages producing channel current reduction. This reduction in the current under RF operation is called current dispersion, or more precisely, current dispersion induced by surface traps.

2.3.2.2 Buffer Traps

Buffer traps refer to the deep levels located in the buffer layer or in the interface between the buffer layer and the substrate. Under high electric field condition, due to high drain-source voltage, electrons moving in the 2DEG channel could get injected into buffer traps. Due to the longer trapping time constant (in the order of 0.1ms), the trapped electrons cannot follow the high frequency signal and hence, they are not available for conduction. The trapped electrons produce a negative charge, which depletes the 2DEG, and therefore reduce the channel current. This reduction in the current under RF operation is called current dispersion, or more precisely, current dispersion induced by buffer traps.

Buffer traps are primarily related to the existing large number of threading dislocation in the GaN layer due to the large lattice mismatch between the GaN and the substrate. These threading dislocations manifest themselves as electron traps (Hansen et al, 1998). Therefore, to reduce these generated traps, a relaxation layer is added between the GaN buffer and the substrate. Another source of traps is the buffer compensation process to obtain high insulating material. Availability of background electron concentration in the buffer material due to native shallow donors cannot be avoided. These donors are mostly compensated by adding deep acceptors. If the buffer is not completely compensated, then a leakage current through the buffer will be generated deteriorating the pinch-off characteristic of the device. In the case of over compensation, empty deep acceptors will be generated in the buffer material. These empty acceptors behave as electron traps. The kink effect in the DC characteristic, shown in Fig. 8 can be assumed as a signature of buffer trapping effect, which is attributed to hot electrons injected into the buffer traps under the influence of high drain voltage. These trapped electrons deplete the 2DEG and result in a reduction of the drain current for subsequent V_{ds} traces (Jarndal 2006).

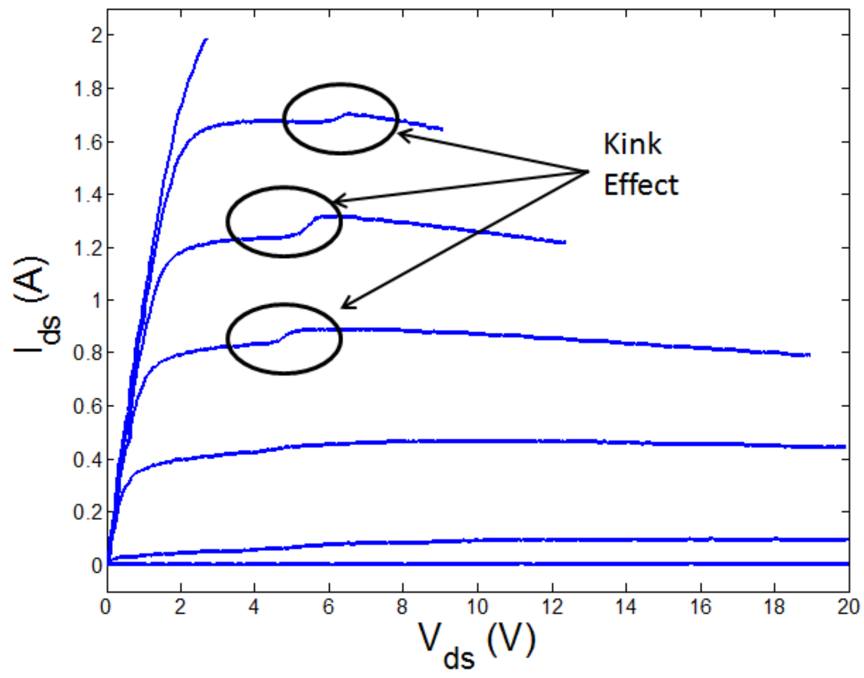


Figure 8. Kink effect in DC characteristics.

Chapter 3 - Compact Modeling Basics for FETs

3.1 Introduction

In Chapter I the importance of developing accurate transistor models, their applications in modern communication systems, different types of models and the modeling extraction process was discussed. With this background in place, we can now discuss in greater detail some of the modeling issues that need to be considered carefully in order to construct an accurate transistor model that can be used in the design of RF power amplifiers. Our aim is to build models for the transistors that can be used in circuit simulators for the design of power amplifiers and power amplifier integrated circuits. These models are known as compact models. To achieve this objective, the models must be able to reproduce with acceptable fidelity the measured electrical and thermal properties of the transistors, and to simulate them quickly, with robust convergence. Another common modeling objective is to be able to inform the physical device design: in other words indicate which of the material and structural properties of a given transistor affect its electrical performance. The accuracy with which any model can achieve this depends on the level of abstraction of the model in the first place, Fig. 9, describes a hierarchy of the modeling approach.

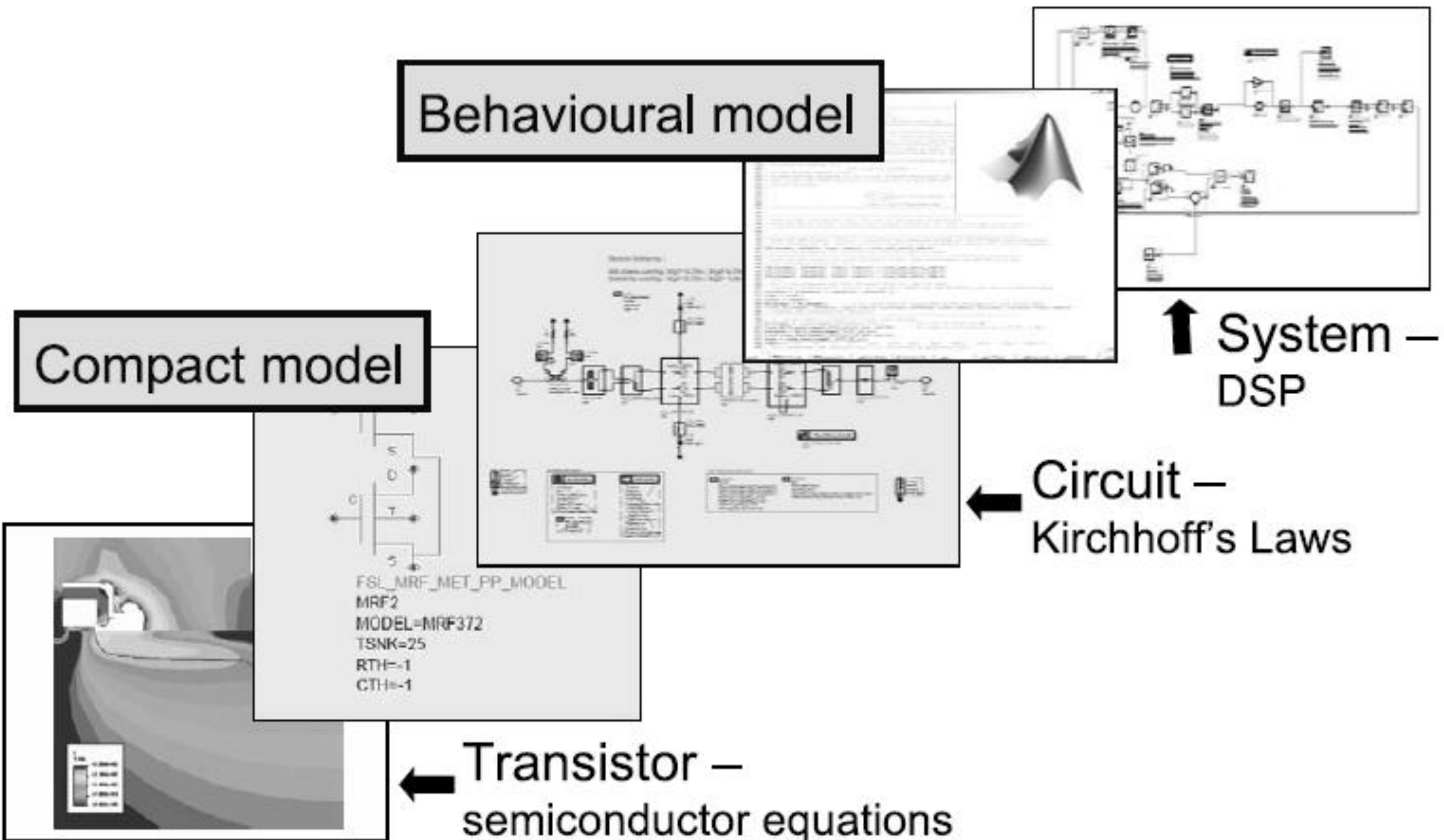


Figure 9. A general modeling hierarchy applied to the level of abstraction of the design. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 52)

3.2 Physical Models

The "physics of operation" of a given transistor are best described using a physical model simulation, in which the geometry, topography, and the material properties of the semiconductors, metals, and insulators that form the transistor are captured in the model description in the simulator. In physical modeling, the nonlinear partial differential equations (PDEs) that describe charge distribution, charge transport, current continuity, and so forth, in the transistor structure are solved in the simulator. As the transistor geometries become smaller, quantum-mechanical (QM) effects also need to be incorporated into this solution. In the simulation, the transistor geometry is discretized in two or three dimensions, and the solution of the PDEs and QM equations is performed for each cell or node in the structure. This requires complex solution techniques to be used, such as finite-difference and finite-element methods. These numerical methods are also found in commercial software for physical transistor modeling, for example: ISE and Silvaco. This class of device modeling is also known as 'TCAD' – Technology CAD.

Despite the many recent mathematical developments in the solution techniques for large systems of nonlinear PDEs, such as parallelization of the problem, using harmonic balance methods instead of traditional time stepping, or modern advanced matrix mathematics, and the advances in computer technology such as increases in processor speed and available memory, this is still a huge problem to solve. The solution of these nonlinear PDEs takes a long time, and the accuracy of the solution depends on how well the physical properties and dimensions of the device are estimated, on the approximations used in the fundamental semiconductor equations, and on the numerical techniques applied in the solution of the system of equations. This means that physical modeling is generally unsuitable for circuit design. However, these modeling techniques have been applied successfully to the technology development cycle for new generations of transistors. Physical model simulations can be used to generate DC I–V characteristics and bias-dependent S-parameter data directly from the simulation. These data can then be used to assess how the physical and geometrical design of the transistor needs to be adjusted to improve the device RF performance. Even "computational loadpull" simulations can be performed using the physical device model loaded by a tunable impedance to assess the

large signal performance of the transistor design and technology, and to inform the device designers of how the nuances in the physical structure of the transistor can influence the RF behavior. Such TCAD simulations are performed at huge computational cost, though it could be argued that this is time better spent, in terms of economy and resources, than building up an array of process lot variations and measuring the electrical performance of each transistor.

A further development of the physical modeling approach is known as global modeling. Here, the physical model simulation describing the semiconductor equations and device geometry is coupled with a thermal model describing the heat transfer in the transistor, and an electromagnetic simulation of the device geometry, metallizations, and substrate. This brings all of the physical principles of operation of the device together in one simulation. Such an approach can provide valuable feedback to the device design team, especially for the physically large structures typical of RF power transistors. For example, the thermal distribution in the transistor under RF drive can be understood and related to the metallization, feeding structures and manifolds, and device structural features such as source-to-drain spacing. Global modeling enables the complete design space of the transistor to be investigated to arrive at the optimal semiconductor and structural arrangement to meet a given design brief.

3.3 Compact Models

While physical modeling provides a viable route for detailed device design and technology development and optimization, it is generally impractical for circuit design. Aside from the computational overhead involved, it is usually expected that the transistor or integrated circuit (IC) process has been determined by the time circuit design begins, and therefore we can devise a more appropriate model that can be used in the circuit simulator for the design of the IC or discrete transistor product. Such models generally fall under the rubric compact. In fact, in many new device technology developments, compact models can be created from the physical model simulations, and hence the circuit or IC design can begin well before the process is finally frozen. The compact models can be

updated as the technology develops, so that the circuit design is always in step with the latest technology improvements, enabling a much reduced new product introduction cycle.

3.3.1 Measurement-Based Equivalent Circuit Models

Generally, compact models are “equivalent circuit” representations of the transistor. The electrical measurements that are performed during the characterization of the transistor can be mapped directly onto a network of circuit components to mimic this electrical behavior. The values of the equivalent circuit parameters are extracted directly from the DC I–V and S-parameter measurements (Dambrine et al 1988). For example, at a given DC bias point and RF frequency, we can measure the S-parameters, which can then be readily converted into generalized two-port Y-parameters using standard conversion relationships. After de-embedding the extrinsic components, this yields an equivalent circuit containing conductance and susceptance components that can be readily incorporated into the circuit simulator. An example is shown in Fig. 10, where the admittance elements in the gate-source and gate-drain branches were transformed into series R–C networks. The values of the equivalent circuit components should, generally speaking, be independent of frequency.

This approach works well for small-signal models; the extraction of the equivalent circuit parameter values can be carried out over a range of bias voltages (V_{gs}, V_{ds}), and the values can be stored in a table indexed by the bias, using interpolation to find the required component values for the given bias voltages, to produce a bias-dependent linear transistor model (Wood and Root, 2000). The equivalent circuit parameter values can be fitted with parameterized functions of the bias voltages, and the model extraction consists of finding these function parameters for each of the circuit elements in the model.

Large-signal models can also be implemented in the simulator using equivalent circuit components, but now the component parameter values are dependent on the large-signal voltages. The example circuit topology shown in Fig. 11 is the Motorola electro-thermal (MET) model (Curtice et al 1999).

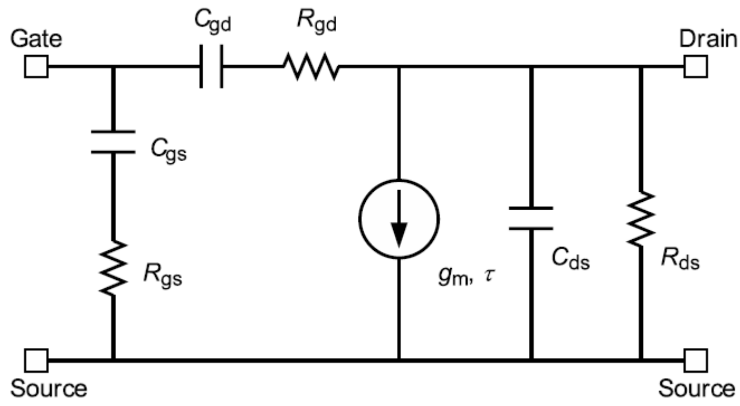


Figure 10. Small-signal equivalent circuit model of the intrinsic transistor

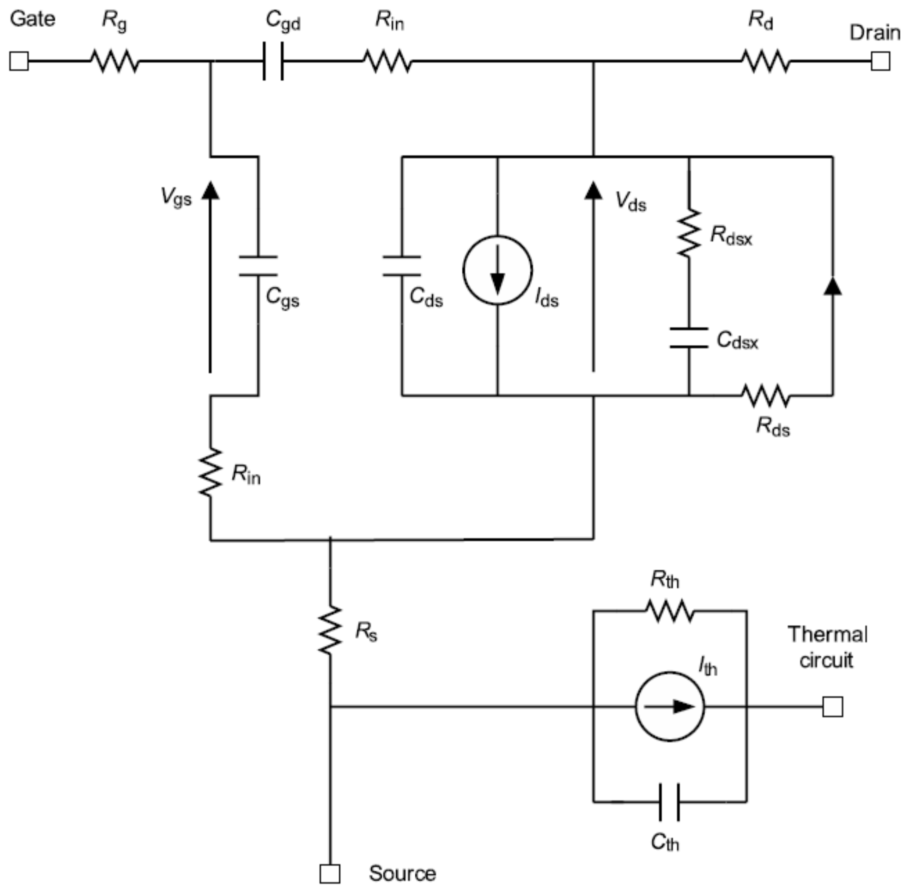


Figure 11. Large-signal equivalent circuit (compact) model from Curtice et al 1999.

For a compact model that can be used for small- and large-signal applications, the functional dependences of the equivalent circuit parameter values on the instantaneous terminal voltages are required. We cannot simply use the DC bias voltages as indices and expect the dynamic behavior of the transistor to be predicted correctly. The small-signal model is a representation of the first moment of the Taylor series expansion of the terminal admittances about the DC bias point, that is, the slopes defined by the infinitesimal voltage swings around the quiescent condition. In the large-signal model, we need to include the responses to large excursions of the terminal voltages. This can be done by integrating the voltage-dependent admittances over the voltage space, subject to some fundamental physical constraints.

We can then apply mathematical function fitting techniques to these derived large-signal equivalent circuit parameters, to obtain a nonlinear functional description for the large-signal model. An example of this mathematical function-fitting approach is the description of the FET I_d–V_{ds} characteristics using a hyperbolic tangent (tanh(x)) curve. The hyperbolic tangent itself has no physical meaning in the context of FET operation, but describes the basic shape of the curve from quasi-linear through saturation regions reasonably well. This function is used in several examples of FET compact models (Curtice et al, 1985, 1999), (Statz et al, 1987), (Parker and Skellern, 1997), and is generally modified with other parameters to describe the sharpness of the 'knee' region, output conductance, and so forth. Instead of fitting these derived parameters with some function approximation, we can simply store them in table form, indexed by the instantaneous terminal voltages ($V_{gs}(t)$, $V_{ds}(t)$) (Root and Meyer, 1991), to produce a "table model".

3.3.2 Physically-Based Equivalent Circuit Models

The main alternative to the measurement-based approach to compact modeling, described above, is a physically-based approach in which the fundamental device physics is used as a basis for a set of "phenomenological" equations that describe the terminal behavior of the transistor in terms of macroscopic physical qualities or parameters, such as the thickness of the active semiconductor layer, gate length, active layer doping, electron mobility, gate oxide thickness, and so forth.

In physically-based compact models, significant simplifications are usually made to the real underlying physics of the FET device operation, in order to construct a model that can run quickly enough in the circuit simulator to be of use in circuit design. Classic examples of such an approach are found in Shockley's original drain current model of the silicon JFET (Shockley, 1952), and in the extension of this for MESFETs by Pucel et al (1975), in which the electron velocity saturation is modeled by a two-zone velocity field relationship, shown in Fig.12; similar long-channel and velocity saturation models exist for MOSFETs (Ytterdahl et al 2003). In these models the transistor I–V characteristics are calculated using the phenomenological equations. Thus, the influence of the material and device parameters on the terminal I–V characteristics can, in principle, be determined. In a similar manner, the gate current and depletion capacitances associated with the Schottky gate contact in MESFETs and HEMTs are often modeled using simple one-dimensional expressions for the rectifying diode. This simple one-dimensional description of the charge storage as a two-terminal capacitance neglects charge conservation principles and results in an incorrect description of the bias-voltage dependence of the measured FET capacitances.

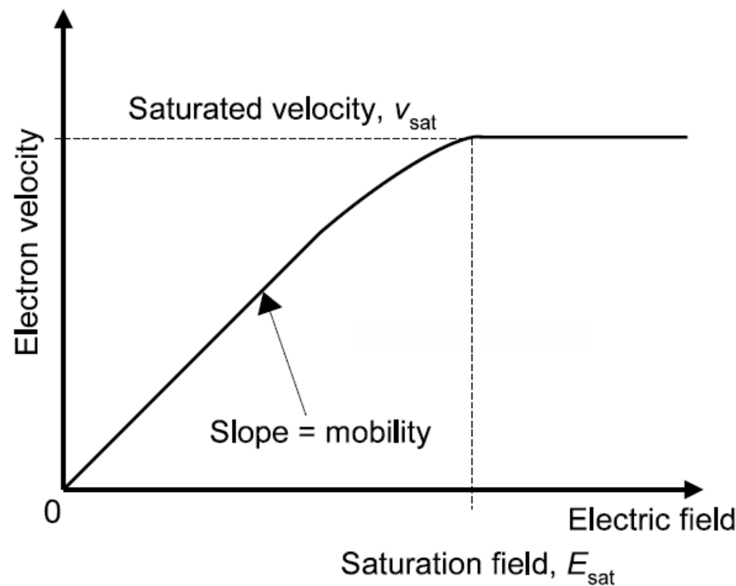


Figure 12. Two-zone electron velocity vs electric field relationship used in physically-based models for short gate-length FETs. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 58)

In contrast, physically-based compact models for bipolar transistors are relatively straightforward: their basic physics of operation can be described by back-to-back diodes and controlled current sources found in Ebers–Moll and Gummel–Poon models. These models are implemented in simulators and can be used successfully in circuit design. Extensions of these basic models to accommodate second-order effects and transport effects in III–V semiconductor heterojunction bipolar transistors (HBT) have been made, resulting in the VBIC model for silicon-based transistors (McAndrew et al, 1996), and the AgilentHBT model for III–V HBTs (Iwamoto et al, 2003).

For any given FET transistor device, we will need to perform measurements to determine the values of the parameters that are used in the equations describing the device physics. This process is known as parameter or model extraction. To extract the parameters for a physically-based compact model, we will generally measure the I–V and C–V or S-parameters of the transistor, and fit the simplified equations by adjusting the parameters in these equations. As we try to improve the model by accounting for higher-order physical phenomena into the equations, we create more parameters that need to be extracted from the measurements. We therefore need to make more measurements to illuminate these higher-order effects. The higher-order parameters are usually determined after the first order model is created, often by using optimization methods. This can lead to a complicated measurement and extraction procedure, and result in a large number of parameters to describe fully the operation of the transistor. An example of this approach is the BSIM MOSFET model (Liu, 2001) with the ‘level 4’ model containing around four hundred extractable parameters to describe the detailed physics of short-gate MOSFET operation. Even the BJT/HBT physically-based compact models contain many parameters; the AgilentHBT model uses about one hundred extractable parameters.

There are several potential problems with a model that comprises a large number of model parameters. The parameter extraction process generally takes a long time, and can be open to interpretation by the modeler. It may also be difficult to know which parameters are the most important and that should be determined accurately for a given device application. The presumed physical origins of the components of the MESFET equivalent circuit model of Fig 10. are illustrated in Fig. 13. The extrinsic or parasitic elements have

also been included. While this can be useful in relating circuit performance to device design, it should be remembered that these circuit elements are derived from two-port small-signal measurements, and not from a physically-based model. This can lead to some confusion between small-signal and large-signal model parameters. For instance, the elements Rgs and R_{gd} (often called R_i and R_j, respectively) are often described as providing a 'charging path' with resulting characteristic time constants for the capacitances C_{gs} and C_{gd}, whereas they are simply small-signal component values determined from a measurement. Further, there is the often irresistible temptation to ascribe poor circuit performance to specific physical properties, leading to 'tweaking' of individual equivalent circuit component values to demonstrate this behavior without paying any attention to how these element values depend on each other in the complete model.

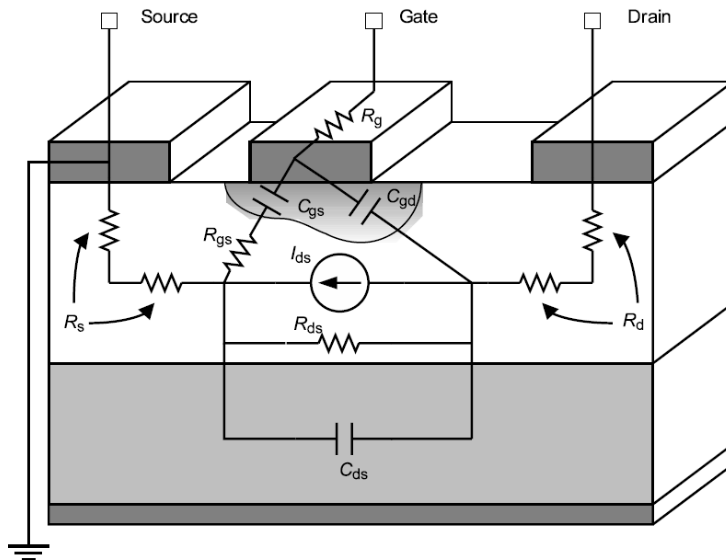


Figure 13. Physical origins of the components of the equivalent circuit model of a MESFET shown in Fig. 10. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 60)

3.3.3 Modeling Approach Implemented

In this thesis we shall focus on creating a compact model of GaN HEMTs. These models will be designed to represent the quasi-static characteristics of the transistor, while being simple to develop and extract based on an artificial neural network approach. We adopt a two-port structure to describe the intrinsic part of the transistor under large-signal

conditions, as shown in Fig. 14. The model consists of controlled current and charge sources that are each dependent on the instantaneous voltages V_{gs} and V_{ds} applied to the transistor (intrinsic) terminals. These current and charge source functions are determined from pulsed bias-dependent S-parameter measurements. The data is transformed, preserving the device dynamics, and is then fitted in the two-dimensional V_{gs}–V_{ds} space using neural network functions. The model structure will preserve small-signal to large-signal consistency.

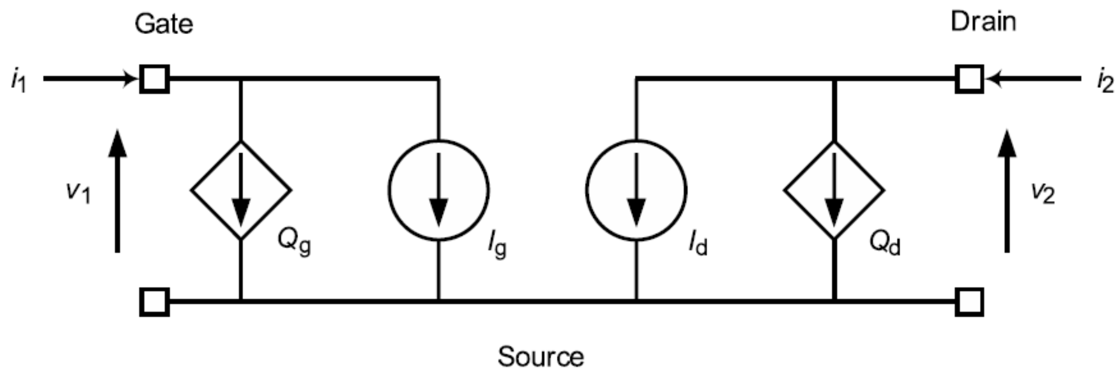


Figure 14. Two-port representation adopted for the large-signal FET model used in this thesis.

3.4 Memory Effects

The expression memory effects has become increasingly common currency in the discussion of the nonlinear behavior of RF power amplifiers. Indeed, it has become something of a catch-all phrase to describe any of the distortions arising from the nonlinearities inherent in the active device – the RF power transistor. In the context of nonlinear systems the term “memory” was proposed by Chua (1987), to describe the influence on the output of a system at a time t of the input signal(s) not only at time t, but also spanning a finite history of the input signal, to some time in the past, t–τ. This is known as a fading memory, as the influence of the input signals deep in the past fades to zero. Essentially, we describe memory effects as the dynamical behavior of the device or system, such dynamics usually being associated with either charge storage or hysteretic phenomena that occur over a wide range of timescales. By this we mean capacitive or inductive behavior with characteristic times that are generally either of the same timescale

as the signal frequency – short-term memory – or at much slower rates – long-term memory. From a modeling point of view, it is important to be able to identify the causes of memory effects in the RF power transistor, and then describe them in the model of the device.

While talking about memory effects in RF power amplifiers, four major sources can be identified; these are shown in Fig. 15. The causes of memory fall into two main categories: those inherent to the transistor itself, and those associated with the external circuitry necessary for the transistor to function. Initially we may be tempted to ignore the external causes, those due to the bias and perhaps any matching circuitry inside the package as being the responsibility of the circuit designer. But in order to this, it would require ignoring how the transistor model works in the circuit, which is a major validation of the model itself. In particular, we would miss the interaction between the long-term memory effects and the short-term (RF) response, which can only occur through a correct description of the nonlinear behavior in our transistor model. Next, the major sources of memory effects in the nonlinear device or system will be outlined, and describe how we might accommodate them in the model of the RF power transistor.

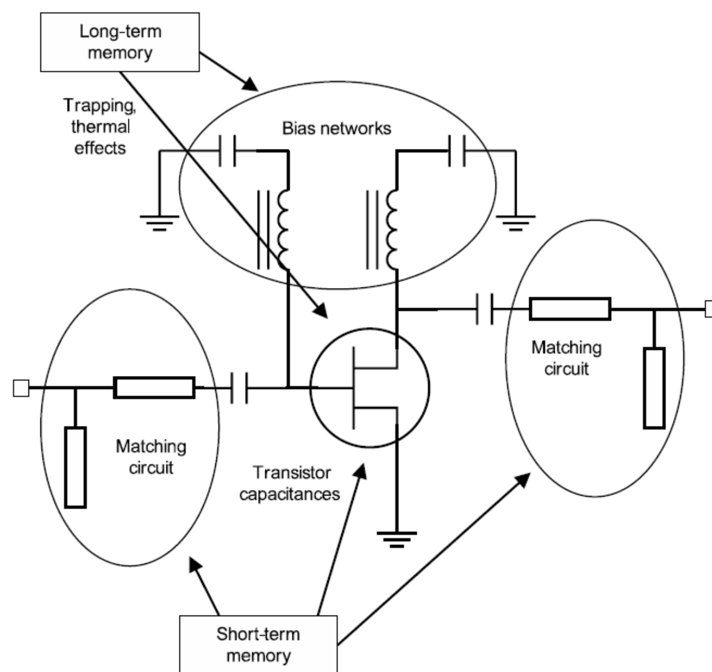


Figure 15. Origins of short-term and long-term memory effects in a transistor circuit. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 63)

3.4.1 Short-Term Memory Effects

The high frequency dynamics of the transistor are determined by the reactances associated with the transistor. In the usual description of a transistor model, these reactances comprise the capacitances and inductances associated with the extrinsic or parasitic elements of the transistor model, and also the charge storage within the transistor's active region, the intrinsic part of the model. The extrinsic components are usually considered to be linear elements, independent of bias, described simply by dv/dt and di/dt expressions for the capacitors and inductors, respectively. The intrinsic charge storage behavior is usually nonlinear, dependent upon the instantaneous voltage or current, and so the dynamic behavior will change with bias or drive signal. When linearized, these nonlinear charge storage components can be described by capacitances in the equivalent circuit representation. For small-signal characterization, the short-term memory effects are simply the frequency response of the transistor. We note that the frequency response is bias-dependent, which requires that the capacitances describing the linearized charge storage behavior in the transistor are also bias dependent: this is a bias-dependent linear transistor model.

Under large-signal conditions, the voltage- or current-dependence of the charge storage functions becomes important. The changing dynamical behavior with signal drive is manifest in measured quantities such as AM-to-AM (gain compression) and AM-to-PM (phase transfer characteristic) of the transistor, and hence the amplifier or system as a whole. The phase transfer characteristic is usually defined as the change in phase of the S_{21} of the transistor from its small-signal value as the signal drive is increased. The AM-to-PM effects are essentially the nonlinear behavior that is often referred to as short-term memory effects. The impact of AM-to-PM on the two-tone response of a transistor in a power amplifier circuit will result in extra components at the intermodulation frequencies that will add in a vector sense to the traditional AM-to-AM IM3 components, but with different phase. The phase of the AM-to-PM IM3 components is signal-dependent, generally resulting in an IM3 response that does not change uniformly with signal drive.

At a system or power amplifier level of description, the matching networks are also a source of short-term memory effects. The matching networks are built from reactive

components – chip capacitors and inductors – or transmission lines. These all have obvious frequency dependence and hence contribute to the short-term dynamics. A less obvious effect is a result of the interaction between the transistor and the matching circuit. The output matching network in a power amplifier is generally designed to present the optimum load resistance to the transistor for maximum power output. This is generally not the conjugate of the small-signal output reflection coefficient, and so there will be reflections at the transistor-circuit interface under large signal conditions: the reflected signals will not see the small-signal reflection coefficient of the transistor, but the hot-S_{22} (Verspecht, 2002), whose dynamical behavior is different from the small-signal case: short-term memory effects. The transistor model must be able to predict both, the small-signal and large-signal behaviors.

3.4.2 Long-Term Memory Effects

The long-term memory effects are due to dynamics that take place on a timescale that is much longer than the period of the RF signal. In RF power amplifiers, ‘long-term’ is generally considered to be on the order of the timescale of the signal envelope, or even longer. Within the transistor, there are considered to be two main causes of long-term memory: thermal effects, and charge trapping. Additionally, there is a circuit-dominated effect, related to the bandwidth of the DC bias network.

3.4.2.1 Thermal Effects

Under conditions of constant drive, the transistor channel heats up uniformly in cross-section. When driven using a modulated information signal, we may find that at some instant the signal is high amplitude, and hence high energy: the transistor channel heats up a little in response to this signal. A short time later, the signal has returned to a low value, but the channel has not cooled down instantaneously: it is still at a slightly higher temperature. Because of this local change in temperature, some of the transistor’s parameters will be slightly different; for instance, the gain may be slightly reduced from the equilibrium-temperature value. This later signal will see the reduced gain, and therefore the output from the transistor will be slightly reduced from the expected equilibrium-gain

value. The transistor parameters are exhibiting a memory of the previous signal. The time constants associated with thermal transients are generally of the order of milliseconds to microseconds. This timescale is significantly longer than the RF period, and is closer to the timescale of the information signal in the envelope: RF information channel bandwidths are in the range 200 kHz to 20 MHz and beyond. These long-term memory effects can be seen in the AM-to-AM characteristics of RF power amplifiers, as a 'spread' around the mean gain compression curve (de Carvalho and Pedro, 1999). A memoryless device or circuit should have a response that is a single line. However, as shown in Figure 16, the memory effects here are indicated by the spread of the compression characteristic: the actual point response depends on the signal value at some previous instant of time. This compression characteristic is obtained using a modulated signal (Wood et al 2006). The thermal memory effects can be included in the transistor model through a dynamic coupling of the electrical and thermal signals.

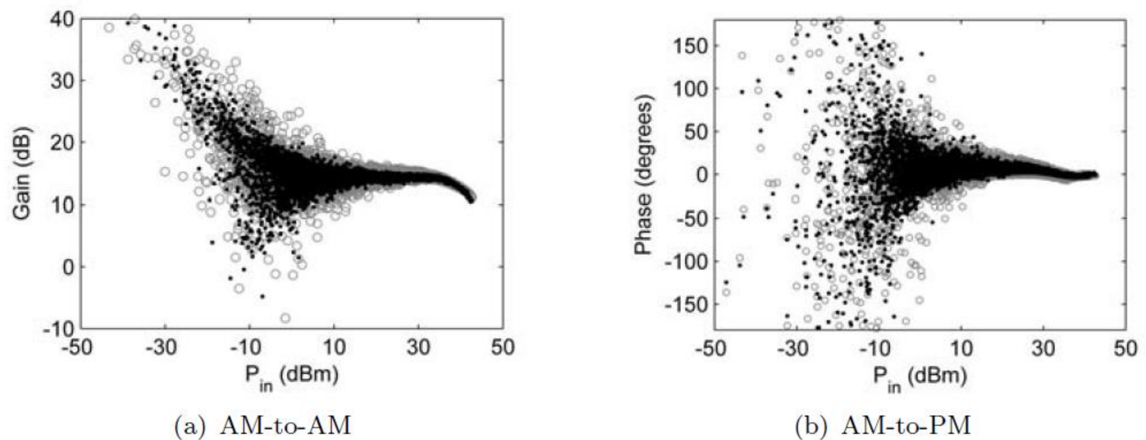


Figure 16. Long term memory effect on the compression characteristic of a power amplifier. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 67)

3.4.2.2 Charge Trapping

Imperfections and defects in the semiconductor occur in the channel itself, at interfaces between the semiconductor and oxide, at the surface of the semiconductor, and at the channel–buffer interface. These imperfections often manifest themselves as available states that can capture and release electrons or holes. The trapping and release are governed

by local potentials and temperature. The action of trapping or releasing an electron is effectively changing the charge density in the channel of the transistor. The rate of trapping and release is on a timescale of kilohertz through megahertz, depending on the nature of the trapping center. Therefore, the trapping mechanism is one that can change the signal current in response to local voltage changes, on a long timescale: a long-term memory effect. Contrary to LDMOS transistor where trapping effects are negligible, gallium arsenide and gallium nitride FETs display several trap related phenomena. One example is the well known dispersion of the small-signal transconductance and output conductance in GaAs FETs: the values of these parameters fall significantly from DC to RF, with the transition occurring between about 10 kHz and 10 MHz. At frequencies above this, the traps are unable to respond to the voltage signal. Other phenomena seen in GaN power transistors include knee collapse and walkout under RF drive, these effects recovering after the drive is removed (Green et al 2003).

3.4.2.3 DC Bias Network

In the strict sense, this may be considered to be a circuit design issue: the DC bias network provides a low impedance path for the DC bias connections that is simultaneously a high impedance to the RF signal. The bandwidth of the low impedance path is known as the video bandwidth, and it is limited by the reactive components that define it. In other words, this path has inductance and capacitance that control the frequency response of the video bandwidth from DC to a few tens of MHz. Any signal components in this frequency range will experience memory effects.

The appearance of signal components in this frequency range is a result of the even-order nonlinearities in the transistor, which should be captured by the model. The finite impedance of the DC bias connections means that these low-frequency components will be impressed upon the transistor, causing small changes in the bias conditions on a timescale appropriate to the signal envelope. The small bias deviations will result in changes in the transistor’s RF behavior, (e.g. its gain) which will affect the output RF signal. These changes will occur at a slow rate compared with the RF signal frequency, and again fall under the rubric long-term memory effects. Essentially, the even-order components in the

video bandwidth are being remixed with the RF signal through this gain modulation effect: they re-appear at the IM3 frequencies, with amplitude and phase that reflect the frequency response of the video bandwidth. This can lead to asymmetry in the IM3 responses, which is often taken to be a characteristic of long-term memory effects (Carvalho and Pedro, 1999).

Chapter 4 - Artificial Neural Networks

4.1 Introduction

Artificial neural networks (ANN) represent a technology that is rooted in many disciplines such as neurosciences, mathematics, statistics, physics, computer science, and engineering. Modeling, time series analysis, pattern recognition, signal processing and control are some examples of the applications of neural networks, all of which take advantage of the most important property of ANNs: the ability to learn from input data.

Work on artificial neural networks, commonly referred to as "neural networks," has been motivated right from its inception by the recognition that the human brain computes in an entirely different way from the conventional digital computer. The brain is a highly complex, nonlinear and parallel computer (information-processing system). It has the capability to organize its structural constituents, known as neurons, so as to perform certain computations (e.g., pattern recognition, perception, and motor control) many times faster than the fastest digital computer in existence today. The brain routinely accomplishes perceptual recognition tasks (e.g., recognizing a familiar face embedded in an unfamiliar scene) in approximately 100-200ms, whereas tasks of much lesser complexity may take days on a conventional computer.

Another illustrative example is the sonar of a bat. Sonar is an active echo-location system. In addition to providing information about how far away a target is, a bat sonar conveys information about the relative velocity of the target, the size of the target, the size of the various features of the target, the azimuth and elevation of the target. The complex neural computations needed to extract all this information from the target echo occur within a brain size of a plum. Indeed, an echo-locating bat can pursue and capture its target with a facility and success rate that would be the envy of a radar or sonar engineer.

How then, does a human brain or the brain of a bat do it? At birth, a brain has great structure and the ability to build up its own rules through what we usually refer to as "experience." Certainly, experience is built up over time, with the most dramatic

development (i.e., hard-wiring) of the human brain taking place during the first two years from birth; but the development continues well beyond that stage.

A "developing" neuron is synonymous with a plastic brain: Plasticity permits the developing nervous system to adapt to its surrounding environment. Just as plasticity appears to be essential to the functioning of neurons as information-processing units in the human brain, so it is with neural networks made up of artificial neurons. In its most general form, a neural network is a machine that is designated to model the way in which the brain performs a particular task or function of interest; the network is usually implemented by using electronic components or is simulated in software on a digital computer. To achieve good performance, neural networks employ a massive interconnection of simple computing cells referred to as "neurons" or "processing units." Thus, we can say that a neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process.
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

The procedure used to perform the learning process is called a learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective.

Neural networks are also referred to in literature as neurocomputers, connectionist networks, parallel distributed processors, etc. Throughout the thesis the term "artificial neural network" or "neural network" is used.

4.1.1 Benefits of Neural Networks

It is apparent that a neural network derives its computing power through, first, its massively parallel distributed structure and, second, its ability to learn and therefore generalize. Generalization refers to the neural network producing reasonable outputs for inputs not encountered during training (learning). These two information-processing

capabilities make it possible for neural networks to solve complex (large-scale) problems that are currently intractable. In practice, however, neural networks cannot provide the solution by working individually. Rather, they need to be integrated into a consistent system engineering approach. Specifically, a complex problem of interest is decomposed into a number of relatively simple tasks that match their inherent capabilities. The use of neural networks offers the following useful properties and capabilities:

1. Nonlinearity. An artificial neuron can be linear or nonlinear. A neural network, made up of an interconnection of nonlinear neurons, is itself nonlinear. Moreover, the nonlinearity is of a special kind in the sense that it is distributed throughout the network. Nonlinearity is a highly important property, particularly if the underlying physical mechanism responsible for generation of the input signal is inherently nonlinear.

2. Input-Output Mapping. A popular paradigm of learning called learning with a teacher or supervised learning involves modification of the synaptic weights of a neural network by applying a set of labeled training samples or task examples. Each example consists of a unique input signal and a corresponding desired response. The network is presented with an example picked at random from the set, and the synaptic weights (free parameters) of the network are modified to minimize the difference between the desired response and the actual response of the network produced by the input signal in accordance with an appropriate statistical criterion. The training of the network is repeated for many examples in the set until the network reaches a steady state where there are no further significant changes in the synaptic weights. The previously applied training examples may be reapplied during the training session but in a different order. Thus, the network learns from the examples by constructing an input-output mapping for the problem at hand.

3. Adaptivity. Neural networks have a built-in capability to adapt their synaptic weights to changes in the surrounding environment. In particular, a neural network trained to operate in a specific environment can be easily retrained to deal with minor changes in the operating environmental conditions. Moreover, when it is operating in a non-stationary environment (i.e., one where statistics change with time), a neural network can be designed to change its synaptic weights in real time. The natural architecture of a neural network for pattern classification, signal processing, and adaptive control applications, coupled with the

adaptive capability of the network, make it a useful tool in adaptive pattern classification, adaptive signal processing and adaptive control. As a general rule, it may be said that the more adaptive we make a system, all the time ensuring that the system remains stable, the more robust its performance will likely be when the system is required to operate in a non-stationary environment. It should be emphasized, however, that adaptivity does not always lead to robustness; indeed, it may do the very opposite. For example, an adaptive system with short time constants may change rapidly and therefore tend to respond to spurious disturbances, causing a drastic degradation in system performance. To realize the full benefits of adaptivity, the principal time constants of the system should be long enough for the system to ignore spurious disturbances and yet short enough to respond to meaningful changes in the environment.

4. Evidential Response**.** In the context of pattern classification, a neural network can be designed to provide information not only about which particular pattern to select, but also about the confidence in the decision made. This latter information may be used to reject ambiguous patterns, should they arise, and thereby improve the classification performance of the network.

5. Contextual Information. Knowledge is represented by the very structure and activation state of a neural network. Every neuron in the network is potentially affected by the global activity of all other neurons in the network.

6. Fault Tolerance. A neural network, implemented in hardware form, has the potential to be inherently fault tolerant, or capable of robust computation, in the sense that its performance degrades gracefully under adverse operating conditions. For example, if a neuron or its connecting links are damaged, recall of a stored pattern is impaired in quality. However, due to the distributed nature of information stored in the network, the damage has to be extensive before the overall response of the network is degraded seriously. Thus, in principle, a neural network exhibits a graceful degradation in performance rather than catastrophic failure.

7. VLSI Implementability. The massively parallel nature of a neural network makes it potentially fast for the computation of certain tasks. This same feature makes a neural network well suited for implementation using very large scale integrated (VLSI)

technology. One particular beneficial virtue of VLSI is that it provides a means of capturing a truly complex behavior in a highly hierarchical fashion.

8. Uniformity of Analysis and Design. Basically, neural networks enjoy universality as information processors. This means that the same notation is used in all domains involving the application of neural networks. This feature manifest itself in different ways:

- Neurons, in one form or another, represent an ingredient common to all neural networks.
- This commonality makes it possible to share theories and learning algorithms in different applications of neural networks.
- Modular networks can be built through a seamless integration of modules.

9. Neurobiological Analogy. The design of a neural network is motivated by analogy with the brain, which is living proof that fault tolerant parallel processing is not only physically possible but also fast and powerful. Neurobiologists look to artificial neural networks as a research tool for the interpretation of neurobiological phenomena. On the other hand, engineers look at neurobiology for new ideas to solve more complex problems than those based on conventional hard-wired design techniques.

4.1.2 Human Brain

The human nervous system may be viewed as a three-stage system, as depicted in the block diagram of Fig. 17. Central to the system is the brain, represented by the neural net, which continually receives information, perceives it and makes appropriate decisions. Two sets of arrows are shown in the figure. Those pointing from left to right indicate the forward transmission of information-bearing signals through the system. The arrows pointing from right to left signify the presence of feedback in the system. The receptors convert stimuli from the human body or thee external environment into electrical impulses that convey information to the neural net (brain). The effectors convert electrical impulses generated by the neural net into discernible responses as system outputs.

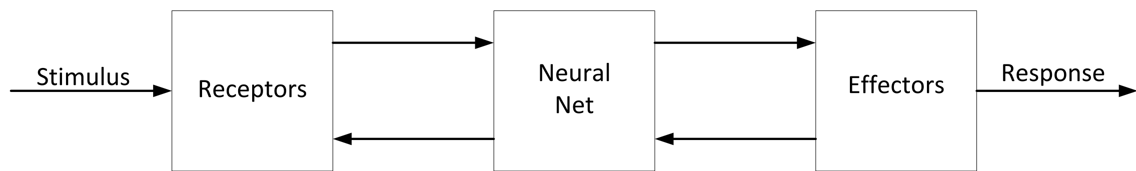


Figure 17. Block diagram representation of the nervous system.

Typically, neurons are five to six orders of magnitude slower than silicon logic gates; events in a silicon chip happen in the nanosecond (10^{-9}s) range, whereas neural events happen in the millisecond (10^{-3}s) range. However, the brain makes up for the relatively slow rate of operation of a neuron by having a truly staggering number of neurons (nerve cells) with massive interconnections between them. It is estimated that there are approximately 10 billion neurons in the human cortex, and 60 trillion synapses or connections. The net result is that the brain is an enormously efficient structure. Specifically, the energetic efficiency of the brain is approximately 10^{-16} joules (J) per operation per second, whereas the corresponding value for the best computers in use today is about 10^{-6} joules per operation per second.

Synapses are elementary structural and functional units that mediate the interactions between neurons. The most common kind of synapse is a chemical synapse, which operates as follows. A presynaptic process liberates a transmitter substance that diffuses across the synaptic junction between neurons and then acts on a postsynaptic process. Thus a synapse converts a presynaptic electrical signal into a chemical signal and then back into a postsynaptic electrical signal. In electrical terminology, such an element is said to be a nonreciprocal two port device. In traditional descriptions of neural organization, it is assumed that a synapse is a simple connection that can impose excitation or inhibition, but not both on the receptive neuron.

Plasticity of the developing nervous system permits it to adapt to its surrounding environment. In an adult brain, plasticity may be accounted for by two mechanisms: the creation of new synaptic connections between neurons, and the modification of existing synapses. Axons, the transmission lines, and dendrites, the receptive zones, constitute two types of cell filaments that are distinguished on morphological grounds; an axon has a

smoother surface, fewer branches, and greater length, whereas dendrite (so called because of its resemblance to a tree) has an irregular surface and more branches. Neurons come in a wide variety of shapes and sizes in different parts of the brain. Like many types of neurons, it receives most of its inputs through dendritic spines. The pyramidal cell can receive 10,000 or more synaptic contacts and it can project onto thousands of target cells.

The majority of neurons encode their outputs as a series of brief voltage pulses. These pulses, commonly known as action potentials or spikes, originate at or close to the cell body of neurons and then propagate across the individual neurons at constant velocity and amplitude. The reasons for the use of action potentials for communication among neurons are based on the physics of axons. The axon of a neuron is very long and thin and is characterized by high electrical resistance and very large capacitance. Both of these elements are distributed across the axon. The axon may therefore be modeled as an RC transmission line, hence the common use of “cable equation” as the terminology for describing signal propagation along an axon. Analysis of this propagation mechanism reveals that when a voltage is applied at one end of the axon it decays exponentially with distance, dropping to an insignificant level by the time it reaches the other end. The action potentials provide a way to circumvent this transmission problem. (Haykin 1999).

4.1.3 Artificial Neural Networks

Artificial neural networks (ANNs) are distributed, adaptive, generally nonlinear learning machines built from different processing elements (PEs). Each PE receives connections from other PEs and or/ itself, hence, the interconnectivity defines the topology. The signals flowing on the connections are scaled by adjustable parameters called weights. The PE sums all these contributions and produces an output that is a nonlinear function of the sum. The PEs outputs become either system outputs or are sent to the same or other PEs.

Processing elements are also called neurons, which can be arranged in different types of configurations in order to perform any given tasks. As shown in Fig. 18 there are a variety of neural networks with special features that have been developed to accomplish these tasks.

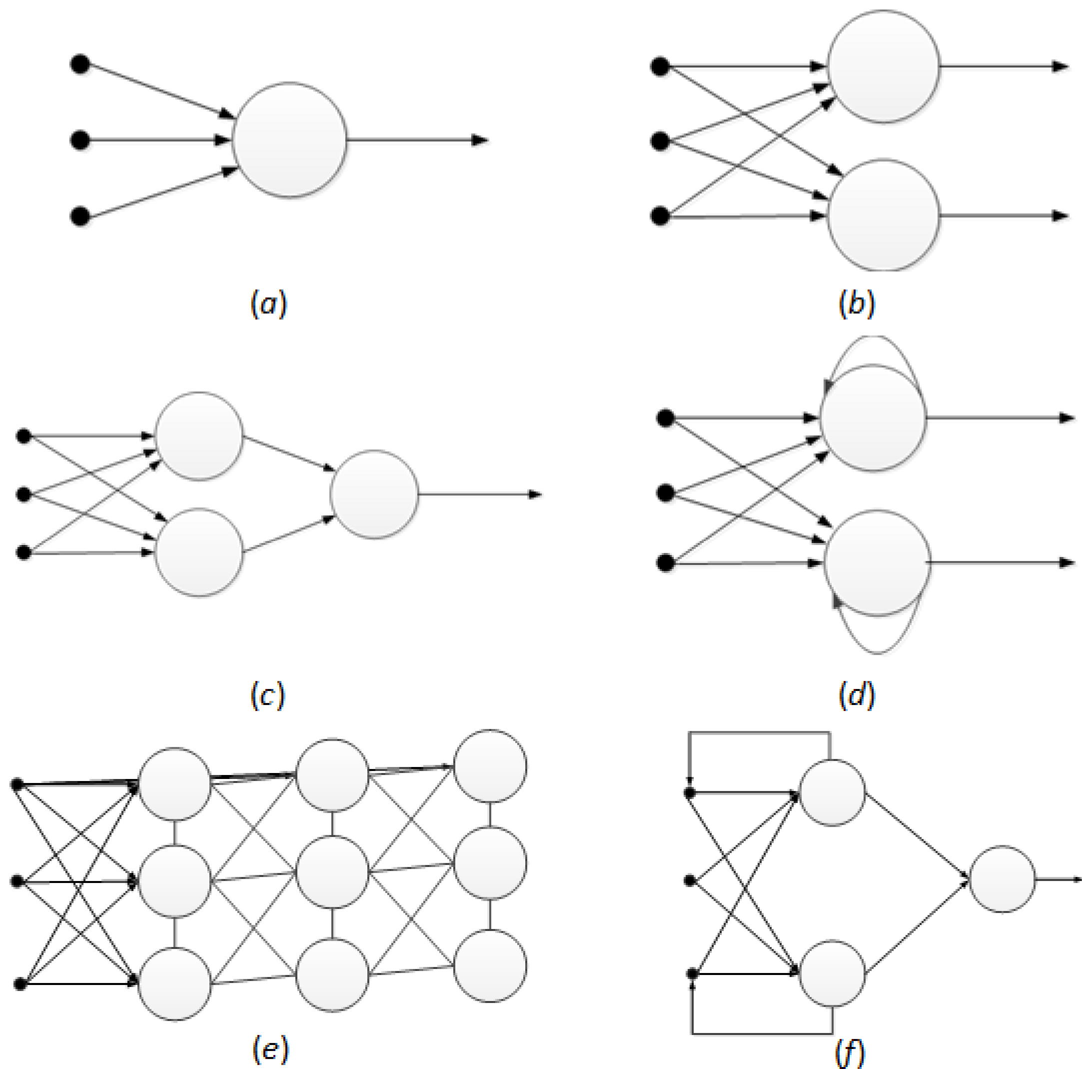


Figure 18. Types of artificial neural networks. (a) Single layer perceptron (b) Linear neuron (c) Multilayer perceptron (d) Competitive network (e) self-organizing map (f) Recurrent network (g).

The common element between all these networks is that each one contains many links connecting inputs to neurons, and neurons to outputs. These links are called weights, and they facilitate a structure for flexible learning that allows a network to freely follow the patterns in the data. The weights are called free parameters, and the neural networks are therefore parametric models involving the estimation of optimum parameters. The flexible structure of these neural networks is what makes them capable of solving such a variety of complex problems.

Following the ideas of nonparametric training, the weights are adjusted directly from the training data without any assumptions about the data's statistical distribution. Hence, one of the central issues in neural network design is to utilize systematic procedures (training algorithms) to modify the weights, so that, as accurate a classification as possible is achieved. The accuracy is quantified by an error criterion.

On the training procedure of a neural network, as depicted in Fig. 19, first data is presented and an output is computed. Second, an error is obtained by comparing output with a desired response, and this error is used to modify the weights using a training algorithm, which usually, is based on the gradient of the error function. This procedure is repeated using all the data in the training set until a convergence criterion is met. Thus, in ANNs (and in adaptive systems in general), the designer does not have to specify the parameters of the system, due to the fact that they are automatically extracted from the input data and the desired response by means of the training algorithm, (Haykin, 1999), (Wang, 2011), (Samarasinghe, 2007).

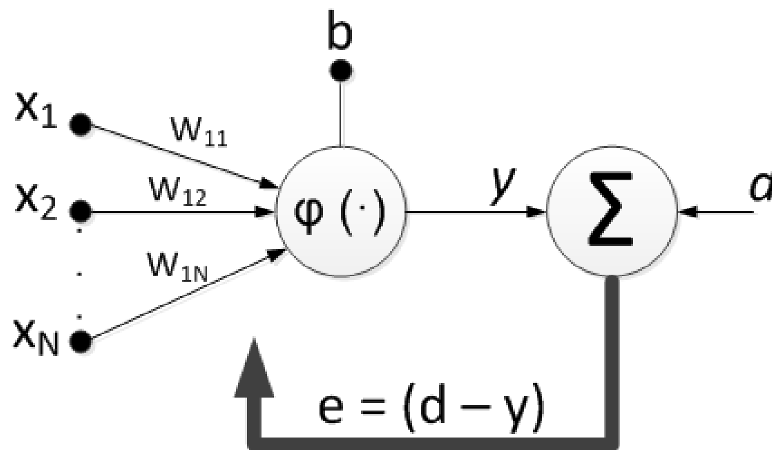


Figure 19. Simplified diagram of ANN's training procedure.

4.1.4 Models of a Neuron

A neuron is an information-processing unit that is fundamental to the operation of a neural network. The block diagram of Fig. 20 shows the model of a neuron, which forms the basis for designing artificial neural networks. Here we identify some basic elements:

1. A set of synapses or connecting links, each of which is characterized by a weight of its own. Specifically, a signal x_i at the input of synapse i connected to neuron j is multiplied by the synaptic weight w_{ji}. It is important to make note of the manner in which the subscripts of the weight w_{ji} are written. The first subscript refers to the neuron in question and the second subscript refers to the input end of the synapse to which the weight refers.
2. An adder for summing the input signals, weighted by the respective synapses of the neuron; the operations described here constitute a linear combiner.
3. An activation function for limiting the amplitude of the output of the neuron. The activation function is also referred to as a squashing function in that it squashes (limits) the permissible amplitude range of the output signal to some finite value.
4. External bias, denoted by b_j, which has the effect of increasing or lowering the net input of the activation function depending on whether it is positive or negative, respectively.

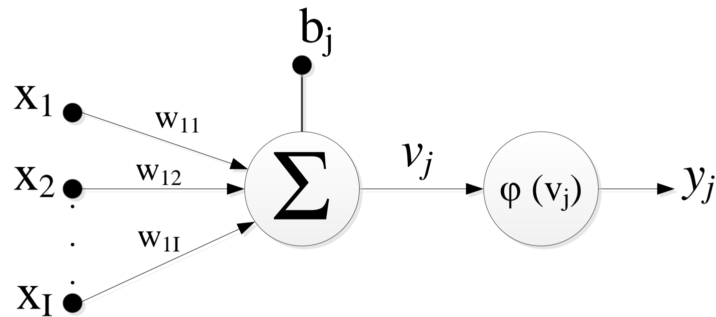


Figure 20. Nonlinear model of a neuron

In mathematical terms, we may describe a neuron j by writing the following pair of equations:

$$v_j = \sum_{i=1}^{I} w_{ji} x_i + b_j \quad (1)$$

$$y_j = \varphi_j(v_j) \quad (2)$$

where x_1, x_2,…,x_I are the input signals; w_{j1}, w_{j2}, …, w_{jI} are the synaptic weights of neuron j; v_j is the linear combiner output for the input signals; b_j is the bias; $\varphi(\cdot)$ is the activation function; and y_j is the output signal of the neuron. The use of bias b_j has the effect of applying an affine transformation to $w_{ji}x_i$ of the linear combiner. The bias b_j is an external parameter of neuron j. Then if an additional input with a value of +1 is added to the model and the bias is considered as a weight, Eq. (1) is simplified as:

$$v_j = \sum_{i=1}^{I} w_{ji} x_i \tag{3}$$

4.1.4.1 Types of Activation Function

The activation function denoted as $\varphi(\cdot)$, defines the output of a neuron in terms of the induced local field v. Here we identify three basic types of activation functions:

1. Threshold Function. For this type of activation function, described in Fig. 21, we have

$$\varphi(v) = \begin{cases} -1, & v < 0 \\ 1, & v \geq 0 \end{cases} \tag{4}$$

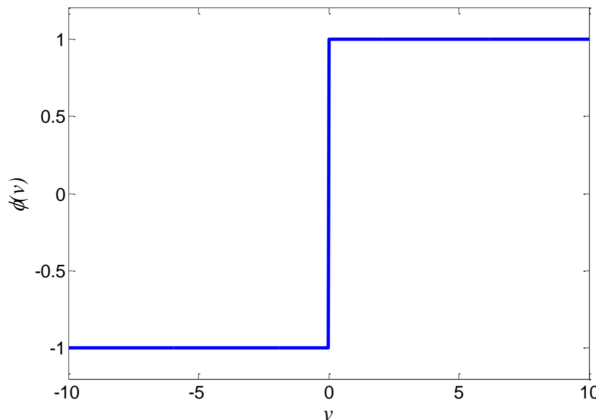


Figure 21. Threshold function

Such a neuron is referred to in the literature as the McCulloch-Pitts model, in recognition of the pioneering work done by McCulloch and Pitts (1943). In this model, the output of a

neuron takes on the value of 1 if the induced local field of that neuron is non-negative, -1 otherwise. Sometimes it is desirable to use an activation function range of 0 to 1, in which case the activation function is considered symmetric with respect to the origin, contrary to the anti-symmetry of the -1 to 1 range where the activation function is an odd function of the induced local field.

2. Piecewise-Linear Function. For the piecewise-linear function described in Fig. 22 we have

$$\varphi(v) = \begin{cases} 0 & v \leq -\frac{1}{2} \\ v + \frac{1}{2} & -\frac{1}{2} < v < \frac{1}{2} \\ 1 & v \leq -\frac{1}{2} \end{cases} \tag{5}$$

where the amplification factor inside the linear region of operation is assumed to be unity. This form of an activation function may be viewed as an approximation to a nonlinear amplifier. The following two situations may be viewed as special forms of the picewise-linear function:

- A linear combiner arises if the linear region of operation is maintained without running into saturation.
- The piecewise-linear function reduces to a threshold function if the amplification factor of the linear region is made infinitely large.

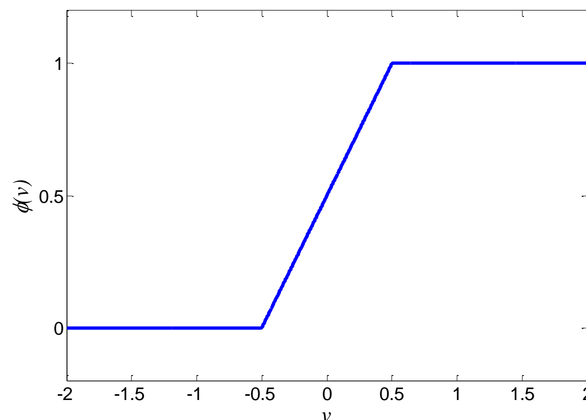


Figure 22. Piecewise-linear function

3. Sigmoid Function. The sigmoid function, which is any S-shaped function, is by far the most common form of activation function used in the construction of ANNs. It is defined as strictly increasing function that exhibits a graceful balance between linear and nonlinear behavior. An example of a sigmoid function is the logistic function defined by

$$\varphi(v) = \frac{1}{1 + e^{-av}} \tag{6}$$

Fig. 23, shows the logistic function and its corresponding anti-symmetric form that is the hyperbolic tangent function, defined by

$$\varphi(v) = tanh(av) \tag{7}$$

where a is the slope parameter of the sigmoid function (usually set to +1). In the limit, as the slope parameter approaches infinity, the sigmoid function becomes simply a threshold function. Whereas a threshold assumes the value -1 or 1, the sigmoid function assumes a continuous range of values from 0 to 1 for the logistic function or -1 to 1 for the hyperbolic tangent function. Note also that a sigmoid function is differentiable, while on the contrary, the threshold function is not. (Differentiability is an important feature of neural network theory as described later in this chapter), (Haykin, 1999).

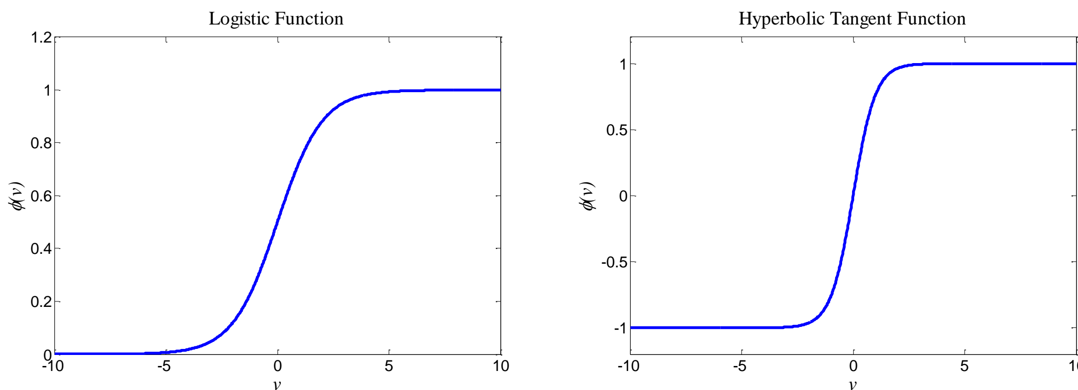


Figure 23. Examples of sigmoid functions

4.2 Feedforward Networks

In this section we study multilayer feedforward networks, an important class of neural networks. Typically, the networks consists of a set of sensory units (source nodes) that constitute the input layer, one or more hidden layers of computational nodes, and an output layer of computation nodes. The input signal propagates through the network in a forward direction, on a layer-by-layer basis. These neural networks are commonly referred to as multilayer perceptrons (MLPs).

4.2.1 Multilayer Perceptrons

The perceptron is the simplest form of a neural network used for the classification of patterns said to be linearly separable (i.e., patterns that lie on opposite sides of a hyperplane). Basically, it consists of a single neuron with adjustable synaptic weights and bias. As introduced in the previous section, the McCulloch-Pitts is the most common processing element based on a single perceptron, which is simply a sum-of-products followed by a threshold nonlinearity.

Multilayer perceptrons have been applied successfully to solve some difficult and diverse problems by training them in a supervised manner with a high popular algorithm known as the error back-propagation algorithm.

Basically, error back-propagation learning consists of two passes through the different layers of the network; a forward pass and a backward pass. In the forward pass, an activity pattern (input vector) is applied to the sensory nodes of the network, and its effect propagates through the network layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the forward pass the synaptic weights of the network are all fixed. On the other hand, during the backward pass, the synaptic weights are all adjusted in accordance with an error-correction rule. Specifically, the actual response of the network is subtracted from a desired (target) response to produce an error signal. This error signal is then propagated backward through the network against the direction of synaptic connections, hence the name “error back-propagation.” The synaptic weights are adjusted to make the actual response of the network move closer to the desired response in a statistical sense. The error back-propagation algorithm is also referred to in

the literature as the back-propagation algorithm, or simply backprop. Henceforth we will refer to it as the back-propagation algorithm. The learning process performed with the algorithm is called back-propagation learning. A multilayer perceptron has three distinctive characteristics:

1. The model of each neuron in the network includes a nonlinear activation function. The important point to emphasize here is that the nonlinearity is smooth (i.e., differentiable everywhere). A commonly used form of nonlinearity that satisfies this requirement is a sigmoidal nonlinearity defined by Eq (6)-(7). The presence of nonlinearities is important because otherwise the input-output relation of the network could be reduced to that of a single layer perceptron. Moreover, the use of the logistic function is biologically motivated, since it attempts to account for the refractory phase of real neurons.
2. The network contains one or more layers of hidden neurons that are not part of the input or output of the network. These hidden neurons enable the network to learn complex tasks by extracting progressively more meaningful features form the input patterns (vectors).
3. The network exhibits a high degree of connectivity, determined by the synapses of the network. A change in the connectivity of the network requires a change in the population of synaptic connection or their weights.

It is through the combination of these characteristics together with the ability to learn from experience through training that the multilayer perceptron derives its computing power. These same characteristics, however, are also responsible for the deficiencies in our present state of knowledge on the behavior of the network. First, the presence of a distributed form of nonlinearity and the high connectivity of the network make the theoretical analysis of a multilayer perceptron difficult to undertake. Second, the use of hidden neurons makes the learning process harder to visualize. In an implicit sense, the learning process must decide which features of the input pattern should be represented by the hidden neurons. The learning process is therefore made more difficult because the

search has to be conducted in a much larger space of possible functions, and a choice has to be made between alternative representations of the input pattern.

The development of the back-propagation algorithm represents a landmark in neural networks in that it provides a computationally efficient method for the training of multilayer perceptrons, (Haykin, 1999).

4.2.1.1 Preface

Fig. 24 shows the architectural graph of a multilayer perceptron with two hidden layers and an output layer. To set the stage for a description of the multilayer perceptron in its general form, the network shown here is fully connected. This means that a neuron in any layer of the network is connected to all the nodes/neurons in the previous layer. Signal flow through the network progresses in a forward direction, from left to right and on a layer-by-layer basis.

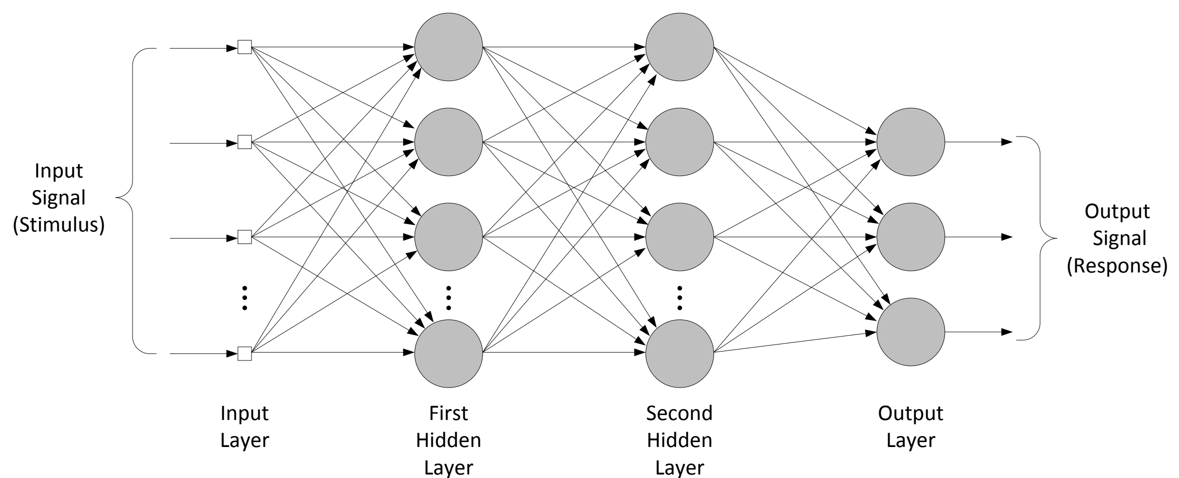


Figure 24. Architectural graph of a multilayer perceptron with two hidden layers.

Figure 25 depicts a portion of the multilayer perceptron. Two kinds of signals are identified in this network.

1. Function Signals. A function signal is an input signal (stimulus) that comes in at the input end of the network, propagates forward (neuron by neuron) through the network, and emerges at the output end of the network as an output signal. We refer to such a

signal as a "function signal" for two reasons. First, it is presumed to perform a useful function at the output of the network. Second, at each neuron of the network through which a function signal passes, the signal is calculated as a function of the inputs and associated weights applied to that neuron. The function signal is also referred to as the input signal.

2. Error Signals. An error signal originates at an output neuron of the network, and propagates backward (layer by layer) through the network. We refer to it as an "error signal" because its computation by every neuron of the network involves an error-dependent function in one form or another.

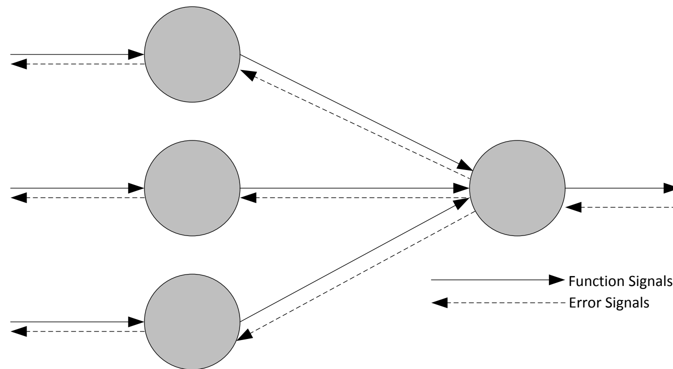


Figure 25. Directions of two basic signal flows in an MLP.

The output neurons constitute the output layers of the network. The remaining neurons constitute hidden layers of the network. Thus the hidden units are not part of the output or input of the network, hence their designation as hidden. The first hidden layer is fed from the input layer made up of sensory units; the resulting outputs of the first hidden layer are in turn applied to the next hidden layer; and so on for the rest of the network.

Each hidden or output neuron of a multilayer perceptron is designed to perform two computations:

1. The computation of the function signal appearing at the output of a neuron, which is expressed as a continuous nonlinear function of the input signal and synaptic weights associated with that neuron.

2. The computation of an estimate of the gradient vector (i.e., the gradients of the error surface with respect to the weights connected to the inputs of a neuron), which is needed for the backward pass through the network.

The derivation of the back-propagation algorithm is rather involved, therefore for simplicity, the back-propagation algorithm of a three layer neural network as shown in Fig. 26 is presented in the next section. To ease the mathematical burden involved in this derivation, it is first presented a summary of the notations used in the derivation (Haykin, 1999).

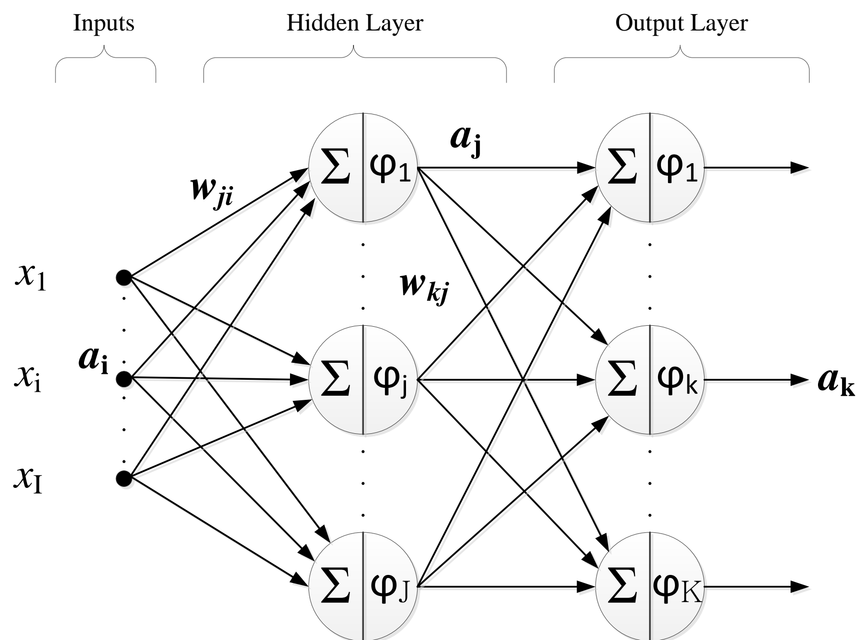


Figure 26. Three layer neural network architecture.

Notation

- The indices i, j and k refer to different neurons in the network. i = 1,2, …I is the number of inputs; j = 1,2, …H is the number of hidden neurons; and k = 1,2, …K is the number of outputs.

- The network is trained from a set of input vectors $\boldsymbol{x}_i^q$ where q = 1, 2, …Q is the number of training samples.
- The bias applied to a neuron is denoted by b_j; its effect is represented by a synapse of weight $w_{j0} = b_j$ connected to a fixed input equal to +1.
- The induced local field (i.e., weighted sum of all synaptic inputs plus bias) of a neuron is denoted by v_j.
- $\varphi(v_j) = \varphi_j$ is the activation function describing the input-output functional relationship of the nonlinearity associated with each neuron.
- The input-output relationship at each layer is denoted by a. In that sense, at the input layer $a_i = p_i$; at the hidden layer $a_j = \varphi_j$; finally, $a_k = \varphi_k$ represents the network output at neuron k.
- d is the desired output (target) of the network.
- e is the error signal at the output neuron.
- The iteration number denoted by n is the time step of the training process where N is the maximum number of epochs.
- E(n) refers to the mean square error function (MSE) at iteration n.
- The symbol $w_{ji}(n)$ denotes the synaptic weight connecting the output of neuron i to the input of neuron j at iteration n. The correction applied to this weight at iteration n is denoted by $\Delta w_{ji}(n)$.

4.2.1.2 Forward Propagation

Considering a three layer neural network as shown in Fig. 26., and by using the notation previously presented, the output at each neuron on the hidden layer can be expressed as

$$v_j = \sum_{i=1}^{I} w_{ji} a_i \tag{8}$$

$$a_j = \varphi_j(v_j) \tag{9}$$

where a_j is the hidden layer output, and at the same time, the output layer input. Finally, the network output is defined as

$$v_k = \sum_{j=1}^{H} w_{kj} a_j \quad (10)$$

$$a_k = \varphi_k(v_k) \quad (11)$$

The network output can be also expressed as

$$a_k = \varphi_k\left(\sum_{j=1}^{H} w_{kj} \varphi_j \left(\sum_{i=1}^{I} w_{ji} a_i\right)\right) \quad (12)$$

4.2.1.3 Update Rule

The whole idea behind gradient descent is to gradually, but consistently, decrease the output error by adjusting the weights; strengthening an active synaptic weight if the postsynaptic neuron fails to fire when it should have fired; weaken an active synaptic weight if the neuron fires when it should not have fired. Intuitively, we know that if a change in a weight will increase (decrease) the error, then we want to decrease (increase) that weight, (Wang, 2011).

$$w(n+1) = w(n) + \Delta w(n) \quad (13)$$

4.2.1.4 Back-Propagation Algorithm

The error signal at the output of neuron k at iteration q (i.e., presentation of the q_{th} training sample) is defined by

$$e_k(q) = d_k(q) - a_k(q) \quad (14)$$

The instantaneous value of the error energy for neuron k is defined as $½e_k^2(q)$. Correspondingly, the instantaneous value ε(q) of the total error energy is obtained by summing $½e_k^2(q)$ over all neurons in the output layer; these are the only "visible" neurons for which error signals can be calculated directly. We may thus write

$$\varepsilon(q) = \frac{1}{2} \sum_{k \in K} e_k^2(q) \quad (15)$$

where the set K includes all the neurons in the output layer of the network. Let Q denote the total number of pattern samples contained in a training set. The average squared error

energy at iteration (n) is obtained by summing ε(q) over all q and then normalizing with respect to the set size Q, as shown by

$$E(n) = \frac{1}{Q}\sum_{q=1}^{Q} \varepsilon(q) \quad (16)$$

Since the application of neural networks in this thesis is focused on function approximation, a single neuron will be defined at the output of the network architecture, simplifying then, the mean square error function (MSE) of Eq. (17) as

$$E(n) = \frac{1}{2Q}\sum_{q=1}^{Q} e_q^2 \quad (17)$$

Eq. (17) can be more clearly expressed as

$$E(n) = \frac{1}{2Q}\sum_{q=1}^{Q} \left(d_q - a_{k,q}\right)^2 \quad (18)$$

where for simplification, $e_q = e(q)$, $d_q = d(q)$ and $a_{k,q} = a_k(q)$.

The MSE is a function of all the free parameters of the network (i.e., synaptic weights and bias). For a given training set, E(n) represents the cost function as a measure of learning performance. The objective of the learning process is to adjust the free parameters of the network to minimize E(n). In order to do this minimization, an approximation similar to the LMS algorithm is used. Specifically, it is considered a simple method of training in which the weights are updated on a pattern-by-pattern basis until one epoch, that is, one complete presentation of the entire training set has been dealt with. The adjustments to the weights are made in accordance with the respective errors computed for each pattern presented to the network. The MSE computed at each iteration, is an estimate of the change that results from modifying the weights based on minimizing the cost function E over the entire training set.

In a similar manner to the LMS, the back-propagation algorithm applies a correction Δw(n) to the synaptic weights, which is proportional to the gradient of the error function ∇E with respect to the network weights as defined in Eq. (14). Thus, in order to

update the synaptic weights of the network a minimization of the cost function must be performed. Hence, the gradient of the error function can be calculated by the following expression

$$\nabla E(n) = -\frac{1}{Q}\sum_{q=1}^{Q}\left(d_q - a_k\right)\frac{\partial a}{\partial w} \tag{19}$$

Eq. (19) can be written in matrix form as

$$\nabla E(n) = -\frac{1}{Q}J^T e(n) \tag{20}$$

where J is q by m Jacobian matrix of the error vector e(n), and m is the number of inputs at the operating neuron.

$$J = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \cdots & \frac{\partial e_1}{\partial w_m} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \cdots & \frac{\partial e_2}{\partial w_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_Q}{\partial w_1} & \frac{\partial e_Q}{\partial w_2} & \cdots & \frac{\partial e_Q}{\partial w_m} \end{bmatrix} \tag{21}$$

The Jacobian is calculated depending on its location in the network architecture as follows

- At the output layer. When a neuron is located in the output layer of the network, it is supplied with a desired response of its own. Then, the Jacobian can be determined in a straightforward manner as

$$J_k = \delta_k a_j \tag{22}$$

$$\delta_k = \varphi'_k \tag{23}$$

where δ_k represents the local gradient at neuron k and is defined as the derivative of the activation function with respect to the induced local field $\varphi'_k = \frac{\partial \varphi_k(v_k)}{\partial v_k}$.

- At the hidden layer. When a neuron is located in the hidden layer of the network, there is no specified desired response for that neuron. Accordingly, the error signal for a hidden neuron would have to be determined recursively in terms of the error signals of all the neurons to which that hidden neuron is directly connected. For this

case, the Jacobian calculation is not straightforward. Therefore, at the hidden layer is Jacobian is expressed as

$$J_j = \delta_j a_i \tag{24}$$

$$\delta_j = \varphi'_j \sum_k \delta_k w_{kj} \tag{25}$$

where δ_j represents the local gradient at neuron j and is defined as the derivative of the activation function with respect to the induced local field $\varphi'_j = \frac{\partial \varphi_j(v_j)}{\partial v_j}$.

The procedure presented above to calculate δ is called the delta rule. The delta rule applies to the output layer and the generalized delta rule applies to hidden layers, layer by layer from the output end.

For a neural network consisting of S layers the back-propagation algorithm can be easily extended following the same procedure. The Jacobian expression is the same for every layer and δ at each hidden layer is computed as

$$\delta_j^s = \varphi_j^{s\prime} \sum_k \delta_k^{s+1} w_{kj}^{s+1} \tag{26}$$

Where j is the current neuron at layer s and k represents the neurons at layer s+1 (Haykin, 1999), (Principe et al, 2000), (Wang, 2011).

4.3 Training Algorithms

Consider a cost function E(**w**) that is a continuously differentiable function of some unknown weight vector **w**. The function E(**w**) maps the elements of w into real numbers. It is a measure of how to choose the weight parameter **w** so that it behaves in an optimum manner. We want to find an optimal solution **w** that satisfies the condition.

$$E(w*) \leq E(w) \tag{27}$$

That is, we need to solve an unconstrained optimization problem, stated as follows: “Minimizing the cost function E(**w**) with respect to the weight vector **w**”. The necessary condition for optimality is

$$\nabla E(w*) = 0 \quad (28)$$

where $\nabla E(w)$ is the gradient vector of the cost function:

$$\nabla E(\mathrm{w}) = \left[\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \cdots, \frac{\partial E}{\partial w_m},\right]^T \quad (29)$$

A class of unconstrained optimization algorithms that is particularly well suited for the design of adaptive filters is based on the idea of local iterative descent. In this section we describe first and second order optimization methods (Haykin, 1999).

4.3.1 Steepest Descent Method

Since the performance surface is a paraboloid, which has a single minimum, a procedure to find the best value of the weight parameter **w** is to search the performance surface instead of computing the best coefficient analytically. The search for the minimum of a function can be done efficiently using a broad class of methods based on gradient information. The gradient has two main advantages for the search:

- The gradient can be computed locally.
- The gradient always points in the direction of maximum change.

If the goal is to reach the minimum, the search must be in the direction opposite to the gradient. Thus, the overall method of searching can be stated in the following way.

Start the search with an arbitrary initial weight w(0), where the iteration number is denoted by the index in parentheses. Then compute the gradient of the performance surface at w(0), and modify the initial weight proportionally to the negative of the gradient at w(0). This changes the operating point to w(1). Then compute the gradient at the new position w(1), and apply the same procedure again, that is,

$$w(n+1) = w(n) - \eta \nabla E(n) \quad (30)$$

where η is positive constant called the step size or learning rate parameter. In going from iteration n to n+1 the algorithm applies the correction

$$\Delta w(n) = w(n+1) - w(n)$$

$$\Delta w(n) = -\eta \nabla E(n) \quad (31)$$

The constant is used to maintain stability in the search by ensuring that the operating point does not move too far along the performance surface. This search procedure is called the steepest descent method. Fig. 27 illustrates the search procedure (Principe et al, 2000).

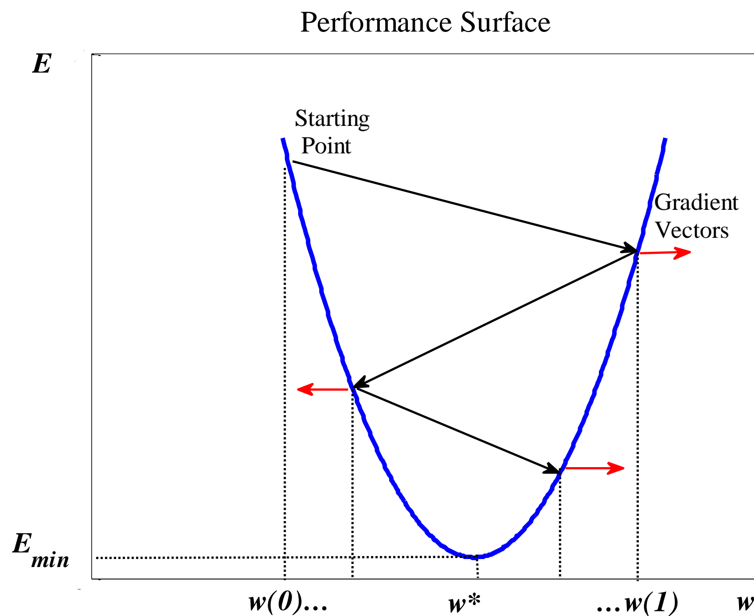


Figure 27. Search using gradient information (Principe et al, 2000).

For a given training set, back-propagation learning based on the steepest descent algorithm may proceed in one of two basic ways.

1. Sequential Mode. The sequential mode of steepest descent learning is also referred to as online, pattern, or stochastic mode. In this mode of operation weight updating is performed after the presentation of each training example.
2. Batch Mode. In the batch mode, weight updating is performed after the presentation of all the training examples that constitute an epoch.

From an "online" operational point of view, the sequential mode of training is preferred over the batch model because it requires less local storage for each synaptic connection. Moreover, given that the patterns are presented to the network in a random manner, the use of pattern-by-pattern updating of weights makes the search in weight space stochastic in

nature. This in turn makes it less likely for the back-propagation algorithm to be trapped in a local minimum.

In the same way, the stochastic nature of the sequential mode makes it difficult to establish theoretical conditions for convergence of the algorithm. In contrast, the use of batch mode of training provides an accurate estimate of the gradient vector; convergence to a local minimum is thereby guaranteed under simple conditions. Also, the composition of the batch mode makes it easier to parallelize than the sequential mode.

When the training data is redundant (i.e., the data set contains several copies of exactly the same pattern), we find that unlike the batch mode, the sequential mode is able to take advantage of this redundancy because the examples are presented one at a time. This is particularly so when the data set is large and highly redundant.

4.3.2 Steepest Descent Method with Momentum Learning

Momentum learning is an improvement to the straight gradient-descent search in the sense that a memory term (the past increment to the weight) is used to speed up and stabilize convergence. In momentum learning the equation to update the weights becomes

$$\Delta \mathrm{w}(n) = -\eta \nabla E(n) + \alpha \Delta w(n-1) \qquad (32)$$

where α is the momentum constant. Normally α should be set between 0.5 and 0.9. This is called momentum learning due to the form of the last term in equation (32), which resembles the momentum in mechanics. Note that the weights are changed proportionally to how much they were updated in the last iteration. Thus if the search is going down the hill and finds a flat region, the weights are still changed, not because of the gradient (which is practically zero in a flat spot), but because of the rate of change in the weights. Likewise, in a narrow valley, where the gradient tends to bounce back and forth between hillsides, the momentum stabilizes the search because it tends to make the weights follow a smoother path. Fig. 28 summarizes the advantage of momentum learning. Imagine a ball (weight vector position) rolling down a hill (performance surface). If the ball reaches a small flat part of the hill, it will continue past this local minimum because of its momentum. A ball without momentum, however, will get stuck in this valley (Principe et al, 2000).

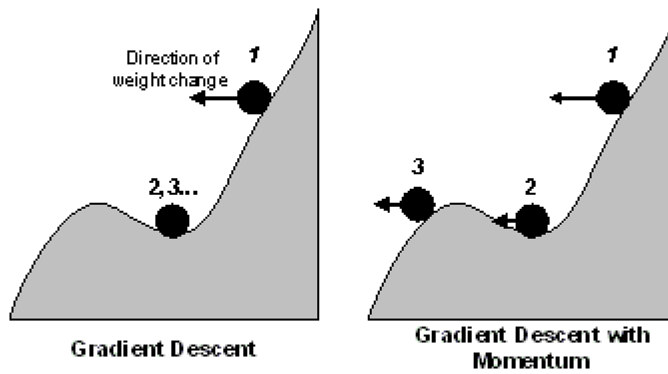


Figure 28. Momentum learning (figure extracted from Principe et al, 2000)

4.3.3 Improved Resilient Back-propagation (iRprop) Algorithm

Let w_{ij} denote the weight in a neural network from neuron j to neuron i and E(n) be an arbitrary error measure that is differentiable with respect to the weights. Bias parameters are regarded as being weights from an extra input; n indicates the learning epoch (iteration). In the Rprop learning algorithm the direction of each weight update is based on the sign of the partial derivative $\partial E/\partial w_{ij}$. A step-size, i.e. the update amount of a weight, is adapted for each weight individually. The main difference to other techniques is that the step-sizes are independent of the absolute value of the partial derivative.

One iteration of the original Rprop algorithm can be divided into two parts. The first part is related to the adjustment of the step-sizes. For each weight w_{ij} an individual step-size Δ_{ij} is adjusted using the following rule:

$$\Delta_{ij}(n) = \begin{cases} \eta^{+}\Delta_{ij}(n-1), & if \ \dfrac{\partial E(n-1)}{\partial w_{ij}} \cdot \dfrac{\partial E(n)}{\partial w_{ij}} > 0 \\ \eta^{-}\Delta_{ij}(n-1), & if \ \dfrac{\partial E(n-1)}{\partial w_{ij}} \cdot \dfrac{\partial E(n)}{\partial w_{ij}} < 0 \\ \Delta_{ij}(n-1), & if \ \dfrac{\partial E(n-1)}{\partial w_{ij}} \cdot \dfrac{\partial E(n)}{\partial w_{ij}} = 0 \end{cases} \qquad (33)$$

where $0 < \eta- < 1 < \eta +$. If the partial derivative $\partial E/\partial w_{ij}$ possesses the same sign for consecutive steps, the step-size is increased, whereas if it changes sign, the step-size is decreased. The step-sizes are bounded by the parameters Δ_{min} and Δ_{max}.

After adjusting the step-sizes, the weight updates Δw_{ij} are determined by applying the following algorithm:

If $\nabla E(n-1) * \nabla E(n) > 0$

$$\Delta_{ij}(n) = min(\eta^{+}\Delta_{ij}(n-1), \Delta_{max})$$

$$\Delta w_{ij}(n) = sign(\nabla E(n))\Delta_{ij}(n)$$

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n)$$

elseif $\nabla E(n-1) * \nabla E(n) < 0$

$$\Delta_{ij}(n) = max(\eta^{-}\Delta_{ij}(n-1), \Delta_{min})$$

if $\nabla E(n) < E(n-1)$

$$w_{ij}(n+1) = w_{ij}(n) - \Delta w_{ij}(n)$$

end

$\nabla E(n) = 0$

elseif $\nabla E(n-1) * \nabla \mathrm{E}(n) = 0$

$$\Delta_{ij}(n) = \Delta_{ij}(n-1)$$

$$\Delta w_{ij}(n) = sign(\nabla E(n))\Delta_{ij}(n)$$

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n)$$

end

In the previous algorithm it can be noticed that if the sign of the partial derivative of the error function has not changed, a regular weigh update is executed. Meanwhile, in the case of a change of sign of the partial derivative, the previous weight update is reverted, (Igel and Husken, 2000).

4.3.4 Levenberg-Marquardt Algorithm

While backpropagation is a steepest descent algorithm, the Levenberg-Marquardt algorithm is an approximation to Newton's method. Assuming a function E(w) which is minimized with respect to the parameter vector w, then Newton's method would be

$$\Delta w = -[\nabla^2 E(w)]^{-1} \nabla E(w) \tag{34}$$

where $\nabla^2 E(\boldsymbol{w})$ is the Hessian matrix.

For the Gauss-Newton method Eq(33) can be written as

$$\Delta w = [J^T(w)J(w)]^{-1} J^T(w)e(w) \tag{35}$$

The Levenberg-Marquardt modification to the Gauss-Newton method is

$$\Delta w = [H(\mathrm{w}) + \mu I]^{-1} J^T(w)e(w) \tag{36}$$

where $H = J^T(w)J(w)$.

The parameter μ is multiplied by some factor β whenever a step would result in an increased E(w). When a step reduces E(w), μ is divided by β. Usually μ is set to 0.01 with β = 10. Notice that when μ is large the algorithm becomes steepest descent (with step 1/ μ), while for small μ the algorithm becomes Gauss-Newton. The Levenberg-Marquardt algorithm can be considered a trust region modification to Gauss-Newton (Hagan and Menhaj, 1994).

4.3.5 Optimized Levenberg-Marquardt with Adaptive Momentum (OLMAM) Algorithm

In the OLMAM method introduces two Lagrange multipliers λ_1 and λ_2, yielding a weight update rule equivalent to the Levenberg-Marquardt algorithm with an additional term of adaptive momentum.

$$\Delta w(n) = -\frac{\lambda_1}{2\lambda_2} [\nabla E(w)]^{-1} \nabla E(w) + \frac{1}{2\lambda_2} \Delta w(n-1) \tag{37}$$

where

$$\lambda_1 = \frac{I_{GF} + \xi\sqrt{\frac{A}{(1-\xi^2)}}}{I_{GG}} \tag{38}$$

$$\lambda_2 = \frac{1}{2}\sqrt{\frac{\mathrm{A}}{I_{GG}\delta P^2(1-\xi^2)}} \tag{39}$$

$$I_{GG} = \nabla E(w(n))^T[\nabla^2 E(w(n))]^{-1}\nabla E(w(n)) \tag{40}$$

$$I_{GF} = \nabla E(w(n))^T \Delta w(n-1) \tag{41}$$

$$I_{FF} = \Delta w(n-1)^T \nabla^2 E(w(n))\Delta w(n-1) \tag{42}$$

$$A = I_{GG}I_{FF} - {I_{GF}}^2 \tag{43}$$

$$\delta P = \sqrt{I_{GG}} \tag{44}$$

$$\xi = \sqrt{1 - \frac{{I_{GF}}^2}{I_{GG}I_{FF}}} \tag{45}$$

It should be noted that this approximation ensures the positive definiteness of the Hessian. The update of μ is performed in the same manner as in the Levenberg-Marquardt algorithm with the only consideration that a step is successful only when

$$E(w(n) + \Delta w(n)) < E\big(w(n)\big) + \sigma_1 \nabla E\big(w(n)\big)^T \Delta w(n) \tag{46}$$

with σ1 = 0.1. The above inequality is known as the Wolfe Condition which states that the cost function should be sufficiently decreased (Ampazis and Perantonis, 2002).

4.4 Approximation of Functions

The universal approximation theorem (Barron, 1993) is important from a theoretical point of view, because it provides the necessary mathematical tool for the viability of feedforward networks with a single hidden layer as a class of approximate solutions. Without such a theorem we could conceivably be searching for a solution that cannot exist. However, the theorem is not constructive, that is, it does not actually specify how to determine a multilayer perceptron with the stated approximation properties. The universal approximation theorem assumes that the continuous function to be approximated is given and that a hidden layer of unlimited size is available for the approximation. Both of these assumptions are violated in most practical applications of multilayer perceptrons.

The problem with multilayer perceptrons using a single hidden layer is that the neurons therein tend to interact with each other globally. In complex situations this interaction makes it difficult to improve the approximation at one point without worsening it at some other point. On the other hand, with two hidden layers the approximation (curve-fitting) process becomes more manageable. In particular, the procedure may be described as follows:

1. Local features are extracted in the first hidden layer. Specifically, some neurons in the first hidden layer are used to partition the input space into regions, and other neurons in that layer learn the local features characterizing those regions.
2. Global features are extracted in the second hidden layer. Specifically, a neuron in the second hidden layer combines outputs of neurons in the first hidden layer operating on a particular region of the input space, and thereby learns the global features for that region and outputs zero elsewhere.

4.5 Modified Backpropagation Algorithm

In this section, we present a simple training algorithm for ANN which uses not only the desired function, but also, data of its first order derivatives with respect to the inputs for training the network (Zárate-de Landa et. al 2012). The ANN model is capable of accurately predict the behavior of the device up to the third derivative.

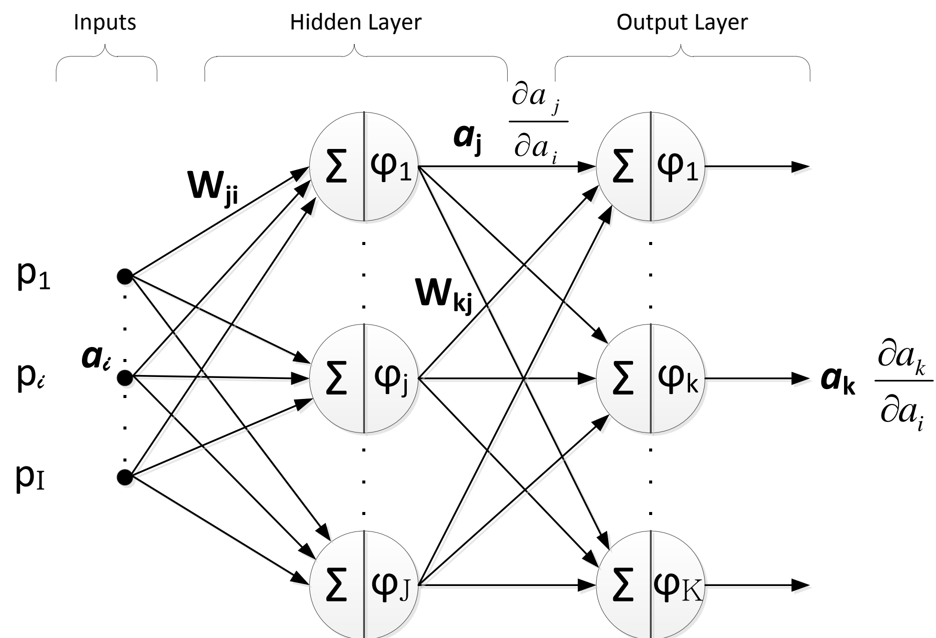


Figure 29. Three layer artificial neural network architecture with derivatives information. (Zárate-de Landa et al, 2012)

A three layer neural network is shown in Fig. 29. The network notation is the same as previously presented, where the network is trained to approximate a function **D** which is the desired output of the network. However, now our objective is to also fit the first derivatives of **D** with respect to the network inputs. We will use the notation $\mathbf{G}_i$ to denote $\frac{\partial \boldsymbol{D}}{\partial a_i}$. As a result, information related to the derivatives $\frac{\partial a_j}{\partial a_i}$, and $\frac{\partial a_k}{\partial a_i}$ will be used during the training process, expanding then, the already well known learning rule algorithm. The proposed mean square error to be minimized is defined as:

$$E = \rho_0 E_0 + \rho_1 E_d \tag{47}$$

where

$$E_0 = \frac{1}{2Q}\sum_{q=1}^{Q}\left[D_q - a_{k,q}\right]^2 \tag{48}$$

$$E_d = \frac{1}{2IQ}\sum_{i=1}^{I}\sum_{q=1}^{Q}\left[G_{i,q} - \frac{\partial a_{k,q}}{\partial a_{i,q}}\right]^2 \tag{49}$$

In (47) ρ_0 and ρ_1 are weight parameters, which values are heuristically selected between 0 and 1, these parameters determine the importance given to E_0 and E_d for the function approximation. Hence, we use the information contained in the original function and its derivatives during the training process in order to improve the generalization of the network.

In order to minimize (47) the calculation of $\partial\mathbf{E}/\partial w_{ji}$ and $\partial\mathbf{E}/\partial w_{kj}$ must be performed. Since $\partial\mathbf{E_0}/\partial w$ can be computed using the standard backpropagation algorithm we are going to focus on the determination of $\partial\mathbf{E_d}/\partial w$.

1) At the Output Layer

The derivative of E_d with respect to w_{kj} is

$$\frac{\partial E_d}{\partial w_{kj}} = -\frac{1}{IQ}\sum_{i=1}^{I}\sum_{q=1}^{Q}\left[G_{i,q} - \frac{\partial a_{k,q}}{\partial a_{i,q}}\right]\left[\frac{\partial}{\partial w_{kj}}\left(\frac{\partial a_{k,q}}{\partial a_{i,q}}\right)\right] \tag{50}$$

Then, the gradient of the error function at the output layer is,

$$\frac{\partial E_d}{\partial w_{kj}} = -\frac{1}{IQ}\sum_{i=1}^{I}\left(J_{d,i}{}^{k}\right)^T e_{d,i} \tag{51}$$

where

$$J_{d,i}{}^{k} = \delta_k \frac{\partial a_k}{\partial a_i} + \delta'_k a_k \tag{52}$$

$$\delta'_k = \varphi''_k w_{kj} \frac{\partial a_k}{\partial a_i} \tag{53}$$

$\mathbf{e_d}$ is the error vector at each epoch, and φ''_k is the second derivative of the processing function at the output neurons.

2) At the Hidden Layer

For the hidden layer the derivative of E_d with respect to w_{ji} is

$$\frac{\partial E_d}{\partial w_{kj}} = -\frac{1}{IQ}\sum_{i=1}^{I}\left(J_{d,i}{}^{j}\right)^{T} e_{d,i} \tag{54}$$

where

$$J_{d,i}{}^{j} = \delta_j \frac{\partial a_j}{\partial a_i} + \left(\delta'_j + \delta''_j\right)a_j \tag{55}$$

$$\delta'_j = \varphi'_j \sum_k \delta'_k w_{kj} \tag{56}$$

$$\delta''_j = \left(w_{ji}\varphi''_j \sum_k \delta_k w_{kj}\right)\frac{\partial a_j}{\partial a_i} \tag{57}$$

Finally, the new weights are calculated following the update rule,

$$w(n+1) = w(n) + \beta(n) \tag{58}$$

where $\beta(n)$ is the search direction, at iteration n, containing information of the error gradient. The search direction can be computed by first or second order training methods.

In order to demonstrate the feasibility of the proposed method, a function f(x) = sin(2x) was created and approximated by using both the original backpropagation algorithm and the backpropagation with derivatives proposed in this thesis. For this example a 1-10-1 ANN architecture was implemented and trained using 100 samples and the Levenberg-Marquardt method for minimum search. As can be observed in Fig. 30, the proposed backpropagation can accurately predict the behavior of the original function up to the third derivative contrary to the classical backpropagation which can only match the first order derivative.

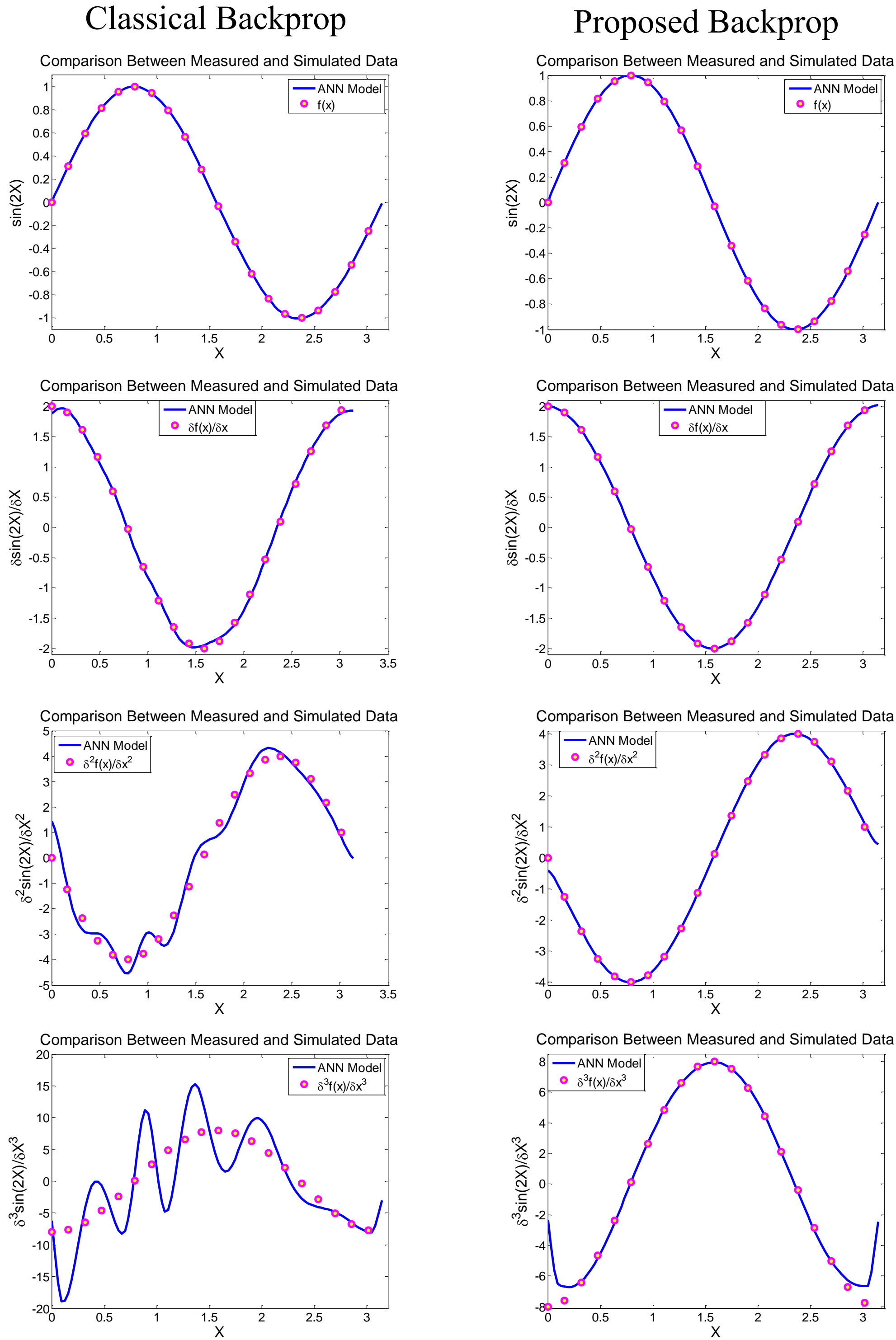


Figure 30. Comparison between classical backpropagation and proposed backpropagation training.

Chapter 5 - Small Signal Modeling

5.1 Introduction

Acurate parasitic elements extraction is the foundation of every linear and nonlinear transistor model used in the design of microwave amplifiers, oscillators and mixers. Since Diamond and Laviron introduced the Cold-FET in the early 80's, the technique has been extensively used in the extraction procedure of parasitic elements for microwave transistors (Curtice and Caamisa, 1984), (Dambrine et al, 1998), (Reynoso-Hernández et al, 1996), due to the simplification of the transistor equivalent circuit model under such bias conditions.

The new trend in the design of high power amplifiers is the use of transistors based on AlGaN/GaN as a consequence of its extraordinary qualities of high voltage, current and frequency operation as explained in chapter II. However, since AlGaN/GaN is a relatively new technology, more accurate models are still under development. One of the main difficulties in the model extraction for this kind of transistor is calculating reliable values of the gate parasitic resistance and inductance (Jarndal and Kompa, 2005), (Crupi et al, 2006), (Zárate-de Landa et al, 2007, 2010), (Reynoso-Hernández et al, 2008).

Reynoso-Hernández et al (2008) presented a reliable method for calculating R_g, L_g, the dynamic Schottky resistance and capacitance R_0 and C_0, based on the extrema points of Z_{11} of the cold-FET model was presented. The difficulty of this method is that even with a very good calibration of the network analyzer a significant amount of noise is observed on the derivatives of the measured Z-parameters, which precludes the algorithm from accurately calculating the extrema points of the real and imaginary parts of Z_{11}, respectively, which are needed for the extraction of R_g and L_g. Generally, the extrema points are tuned with human assistance by observing the plots and selecting the best minimum and maximum points that better fit the measured data, making the procedure time consuming and tedious, especially, when several transistors are being characterized. Although, noise can be lessen by using polynomial fitting, even with a very high

polynomial degree the noise is still present and the problem is not completely solved.

This chapter extends the theory already published by CICESE's RF/Microwave group, but also, makes use of the artificial neural networks (ANN) to mitigate the noise observed in the derivatives of the measured Z_{11} parameter of the Cold-FET. The extraction of R_g and L_g is improved significantly by precisely determining the angular frequency at which the extrema points occur, simplifying then, the accuracy and automation of the parasitic extraction procedure.

5.2 Package Parasitics in Power FETs

In the design and modeling of a microwave packaged power transistor, linear models of the package and matching networks are combined with a nonlinear model of the transistor. The resulting performance is dictated by the impedances presented by the matching networks to the transistor at the fundamental, harmonic and low-frequency terminations. These matching networks are often composed of arrays of small-diameter bondwires, metal-oxide semiconductor (MOS) capacitors, and packages. The passive components provide the necessary low-loss impedance transformation essential for the successful operation of the RF power amplifier.

Typically, the packages that are used for wireless infrastructure RF power amplifier applications are constructed from high-conductivity metals and low-loss dielectrics. The package is one component of the low-loss matching network located between the transistor die and the microstrip matching circuitry on the printed circuit board of the amplifier. Stringent thermal-mechanical design practices are required to ensure that the package can dissipate the substantial heat-flux generated by the transistor. Low thermal resistance packages are required (typically, around 0.5 W/◦C), since all of the energy not converted into RF power is dissipated as heat. For example, a 200 W amplifier operating at 50% efficiency dissipates 100 W of power as heat through the package. The packages, as illustrated in Fig. 31, are designed for the leads to rest on top of the microstrip transmission lines on the printed circuit board (PCB). The back side of the flange contacts the heatsink of the power amplifier forming a conductive electrical connection to the bottom conductor

of the microstrip and a conductive thermal connection to the heatsink, which permits heat to flow away from the packaged transistor (Aaen et al. 2007).

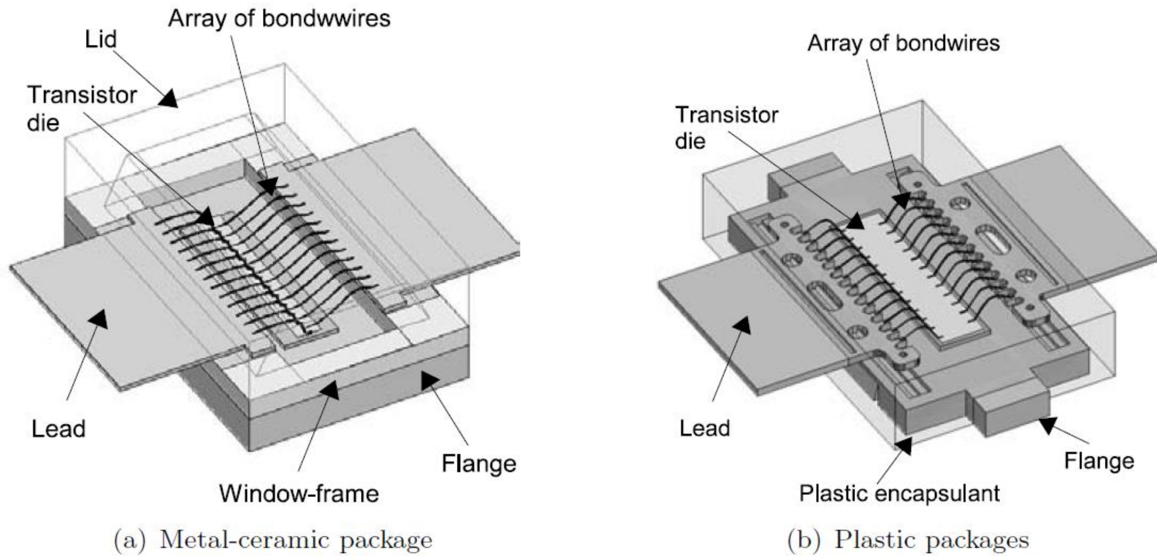


Figure 31. Illustrations of a power transistor using ceramic and plastic packages. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 124)

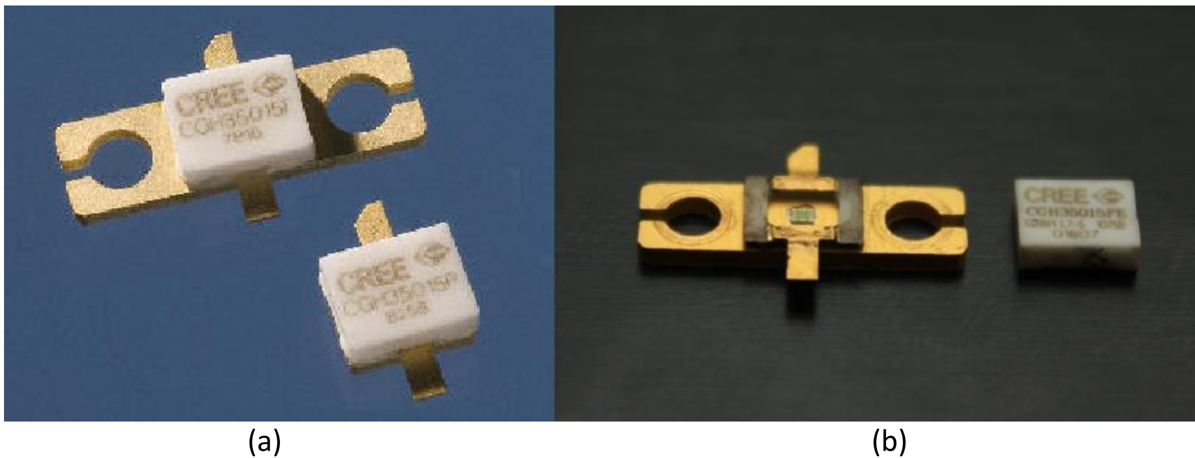
5.2.1 Analysis of an Empty Package

In this thesis 15 Watts AlGaN/GaN CGH35015 HEMTs provided by CREE® were studied, in Figure 32 (a) can be observed the package type (440166) where the transistors are mounted. By using a hot plate, a damaged transistor was opened (Fig. 32b), the bondwires connecting the package to the active device were cut off and its S-parameters measured in order to characterize the transistor package (Monjardín, 2014). Then, the package capacitance can be extracted directly from the measured Y-parameters of the package. Assuming a Π-network of capacitors as illustrated in Figure 23, the package capacitances can be expressed as,

$$C_{11} = \frac{Im(Y_{11}) + Im(Y_{12})}{\omega} \tag{59}$$

$$C_{12} = -\frac{Im(Y_{12})}{\omega} \tag{60}$$

$$C_{22} = \frac{Im(Y_{22}) + Im(Y_{12})}{\omega} \tag{61}$$



(a) (b)

Figure 32. CREE® AlGaN/GaN HEMT CGH35015 on 440166 package. (Monjardín, 2014)

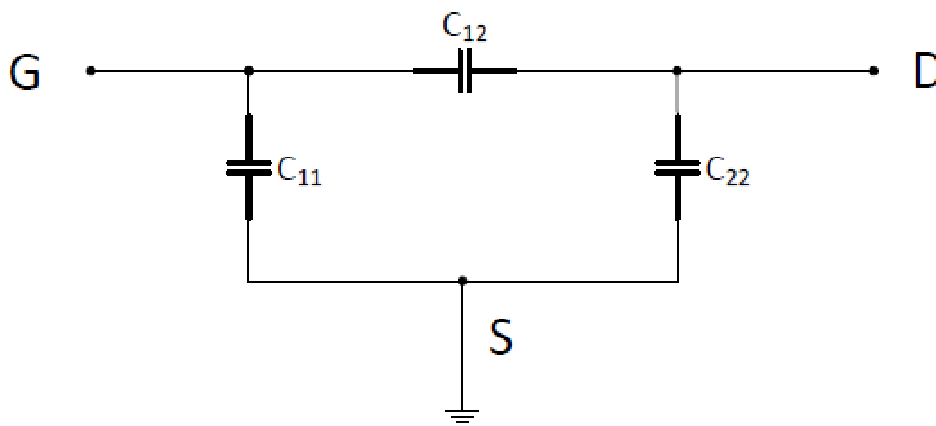


Figure 33. Transistor package equivalent circuit model.

After the package capacitances were extracted the Y-parameters of the structure depicted in Fig. 33 where calculated. It can be observed in Fig. 34 that the model accurately predicts the behavior of the package. The calculated values of the package parasitic capacitances is reported in Table

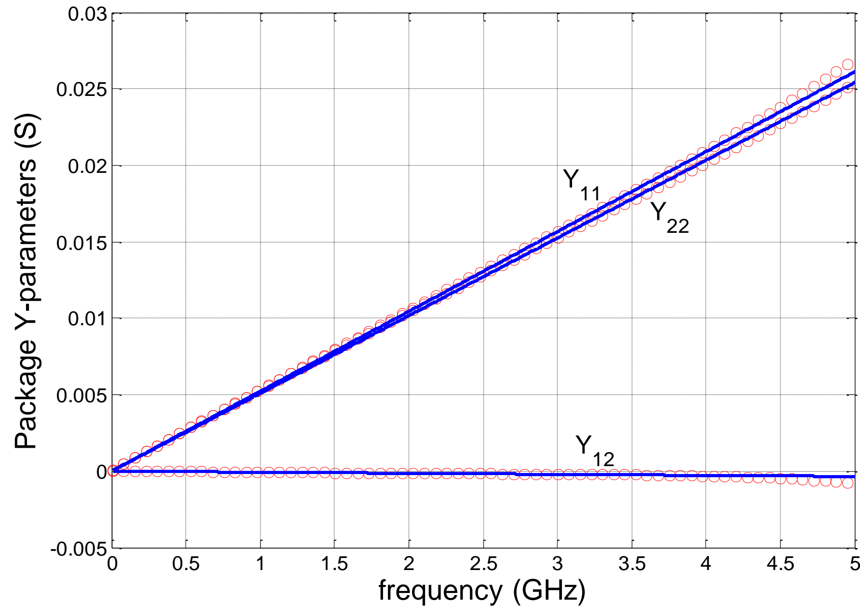


Figure 34. Comparison between measured (circles) and modeled (lines) of CREE 440166 package.

Table 2. Package parasitic capacitances calculated for CREE CGH35015F AlGaN/GaN HEMT

C_{11} (fF)	C_{12} (fF)	C_{22} (fF)
820.43	10.92	798.08

5.3 The Forward Cold-FET Model

The accustomed small signal equivalent circuit model (ECM) of a FET transistor is depicted in Figure 35. In this circuit, the elements are divided in two categories: the extrinsic or parasitic elements (L_g, R_g, L_s, R_s, L_d, R_d, C_{pg} and C_{pd}) which are bias independent and represent the effects of the pads, probe tips and metal contacts of the device; and the intrinsic elements (C_{gs}, R_i, C_{gd}, R_{gd}, R_{ds}, C_{ds}, g_m and τ) which represent the actual device and depend of the control voltages V_{gs} and V_{ds}. Depending on the technology, if the transistor is asymmetric the resistance R_{gd} can be disregarded from the intrinsic

model. When the FET is biased under zero drain to source bias, the kinetic energy of the electrons in the channel is cold with respect to the typical operating condition (Crupi et al 2006). Therefore, for the Cold-FET configuration under low DC forward gate bias current ($0 < V_{gs} < V_{bi}$; $I_{gs} > 0$) the small signal model is simplified as shown in Figure 36 (Zárate-de Landa et al, 2007).The Z-parameters of the forward Cold-FET are expressed as:

$$Z_{11} = R_g^* + R_s^* + \frac{R_0}{1+\omega^2 C_0^2 R_0^2} + j\omega\left[L_g + L_s - \frac{C_0 R_0^2}{1+\omega^2 C_0^2 R_0^2}\right] \quad (62)$$

$$Z_{12} = R_s^* + j\omega L_s \quad (63)$$

$$Z_{22} = R_d^* + R_s^* + j\omega\left(L_g + L_s\right) \quad (64)$$

where, $R_g^* = R_g - \frac{R_{ch}}{6}$; $R_s^* = R_s + \frac{R_{ch}}{2}$; $R_d^* = R_d + \frac{R_{ch}}{2}$.

R_{ch} is the channel resistance which is disregarded in the extraction procedure due to its small value compared to the parasitic resistances. Also, in this configuration, parasitic capacitances have no influence in the Z-parameters of the forward Cold-FET, thus, are not taken into account in the model.

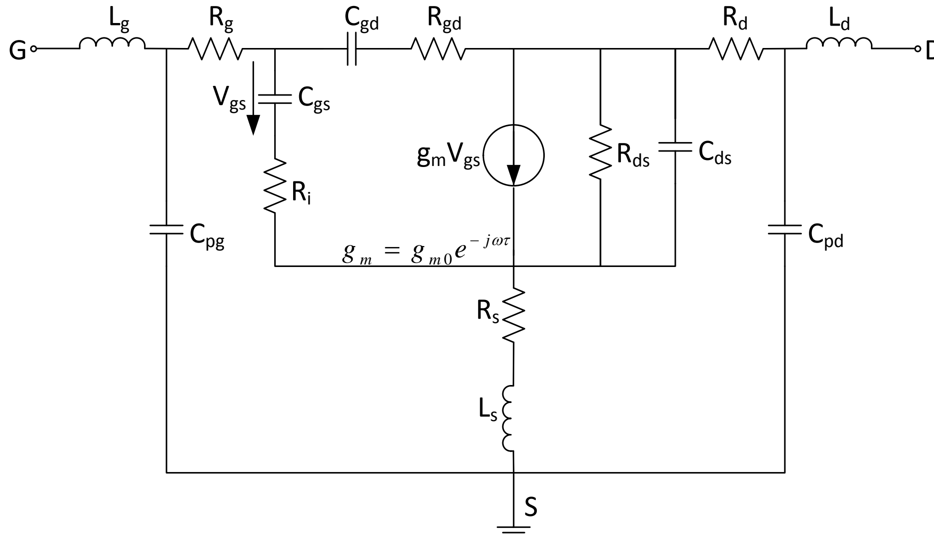


Figure 35. FET Small-signal equivalent circuit model.

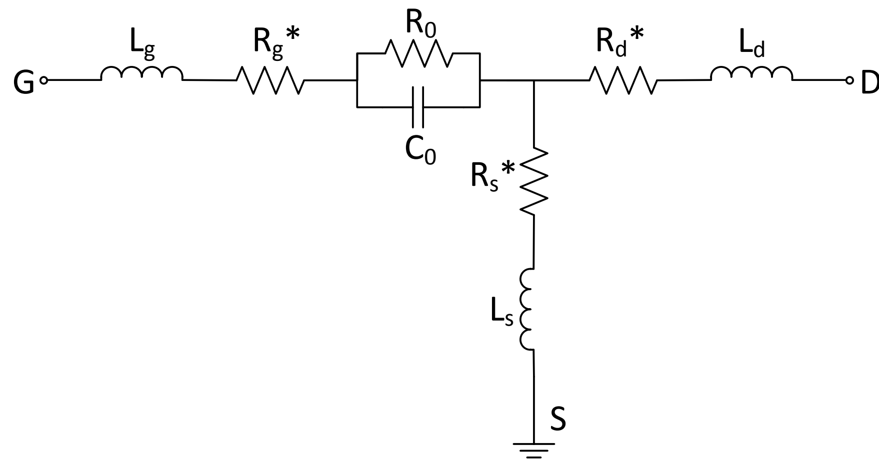


Figure 36. Cold-FET Small-signal equivalent circuit under low DC gate forward current with floating drain. (Zárate de Landa et al, 2007, 2009)

5.3.1 Open Drain Condition

In the forward Cold-FET configuration there is a debate between the differences between applying a zero drain-source voltage ($V_{ds} = 0$) and leaving the drain open. Unlike most authors, an open drain configuration is used in this work as suggested Reynoso-Hernández (1996). This is due to the fact that when the drain is shorted ($V_{ds} = 0$) the gate current is divided into two trajectories, one flowing in the gate-source direction and the other in the drain-source path. In Fig. 37 it can be observed how the Cold-FET ECM changes when $V_{ds} = 0$ is applied.

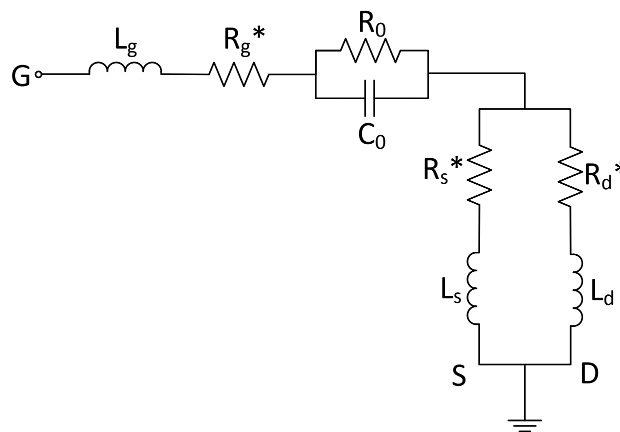


Figure 37. Cold-FET Small-signal equivalent circuit under low DC gate forward current with zero drain-source condition.

Consequently, the Z-parameters of the forward Cold-FET with $V_{ds} = 0$ would be defined as

$$Z_{11} = R_g^* + \frac{R_s R_d^2 + R_s^2 R_d + \omega^2 (L_s^2 R_d + L_d^2 R_s)}{(R_s + R_d)^2 + \omega^2 (L_d + L_s)^2} + \frac{R_0}{1 + \omega^2 C_0^2 R_0^2} + j\omega \left[L_g + \frac{L_s R_d^2 + L_d R_s^2 + \omega^2 (L_d^2 L_s + L_s^2 L_d)}{(R_s + R_d)^2 + \omega^2 (L_d + L_s)^2} - \frac{C_0 R_0^2}{1 + \omega^2 C_0^2 R_0^2} \right] \quad (65)$$

$$Z_{12} = Z_{22} = \frac{R_s R_d^2 + R_s^2 R_d + \omega^2 (L_s^2 R_d + L_d^2 R_s)}{(R_s + R_d)^2 + \omega^2 (L_d + L_s)^2} + j\omega \frac{L_s R_d^2 + L_d R_s^2 + \omega^2 (L_d^2 L_s + L_s^2 L_d)}{(R_s + R_d)^2 + \omega^2 (L_d + L_s)^2} \quad (66)$$

Hence, in order to make use of equation (62)-(64) a forward Cold-FET configuration with open drain must be used.

5.3.2 Drain and Source Parasitic Resistances and Inductances Calculation

Drain and source parasitic resistances can be determined by DC measurements of the Schottky diode model of the transistor (Cheung, 1986), (White and Healey, 1993), (Reynoso-Hernández, 1996). This is a very accurate and easy approach for the calculation of the drain and source parasitic resistances, however, it is necessary to perform extra measurements to the FET which could be time consuming. Therefore, these parasitic resistances can be directly extracted from the real part of Z_{12} and Z_{22} parameters of the Cold-FET measurement. Equations for the calculation of R_s and R_d can be expressed as

$$R_s^* = Re(Z_{12}) \quad (67)$$

$$R_d^* = Re(Z_{22} - Z_{12}) \quad (68)$$

L_s and L_d elements are directly calculated from the slope of the imaginary parts of Z_{12} and Z_{22} expressed by equation (63) and (64).

$$L_s = \frac{Im(Z_{12})}{\omega} \quad (69)$$

$$L_d = \frac{Im(Z_{22} - Z_{12})}{\omega} \quad (70)$$

Once the source and drain parasitic resistances and inductances are calculated a comparison between measured and simulated data is performed as depicted in Figs. 38 and 39.

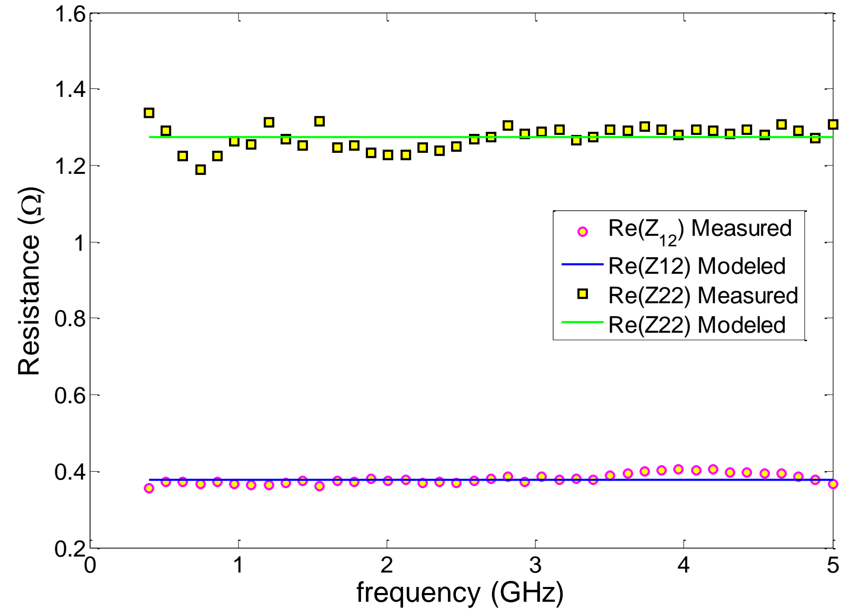


Figure 38. Comparison between measured and simulated real part of Z_{12} and Z_{22}.

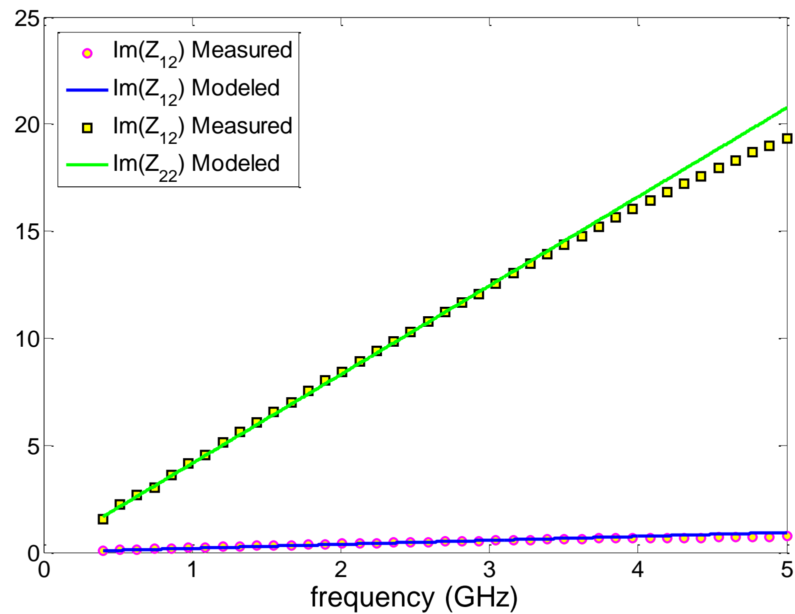


Figure 39. Comparison between measured and simulated imaginary part of Z_{12} and Z_{22}.

It is worth mentioning that for transistors with small gate width, the calculated source inductance L_s might be a negative value. This is a consequence of ignoring the parasitic capacitances in the forward Cold-FET model, which is not a precise assumption for small sized FETs. When negative values of L_s are computed, the equations published in section IV-B of the article of Reynoso-Hernández et al, (1996) should be applied, in order to take into account the parasitic capacitances in the forward Cold-FET model.

5.3.3 Gate Parasitic Resistance and Inductance Calculation

Due to the dynamic resistance and capacitance elements R_0 and C_0 used to model the gate Schottky diode, expressed by equation (62), the direct calculation of R_g and L_g is not straight forward. R_g and L_g could be calculated from the real and imaginary part of Z_{11}, respectively, provided R_s, L_s, R_0, and C_0 are known. Reynoso-Hernández et al, in 2008, proposed a procedure for the calculation of R_0, C_0, R_g and L_g from expression (62). In this section an extension of such method will be deeply explained.

5.3.3.1 Frequency Points of Interest

An important condition to apply the proposed method is to measure the Cold-FET under a low gate-source current, contrary to the classical method proposed by Dambrine et al (1988) where, in order to calculate the parasitic gate resistance R_g, a linear regression of several measurements must be acquired. These measurements are performed under a high current gate-source current in order to suppress the capacitive effect of the Schottky diode. It has been found that for high power transistors such as GaN HEMTs, the gate is irreversible damaged long before the capacitive effect of the Schottky diode is suppressed. The proposed method not only protects the device under test by not stressing it under large gate currents, but also, saves time by only measuring the Z-parameters of the Cold-FET once. Nevertheless, due to the dynamic resistance and capacitance of the Schottky diode, the real and imaginary parts of Z_{11} are nonlinear functions which make the extraction procedure complex. By analyzing the Z_{11} parameter and its derivative, some particular frequency points can be used as aid to ease our way through the calculation of the gate

elements. Figures 40 and 41 depict the real and imaginary part of Z_{11} with respect to the frequency and the points of interest that will be used in the extraction procedure.

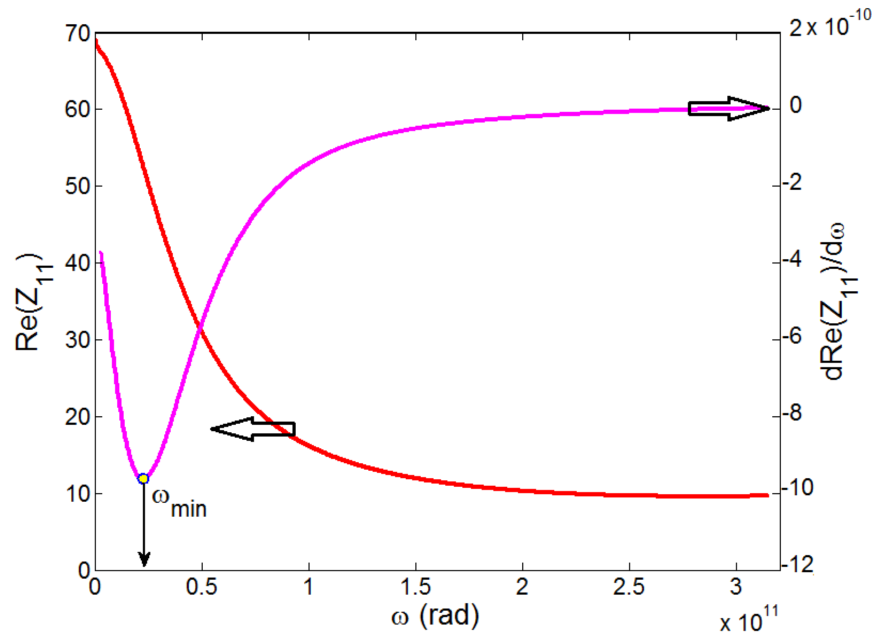


Figure 40. Real part of Z_{11}, of the Cold-FET and its derivative with respect to the frequency.

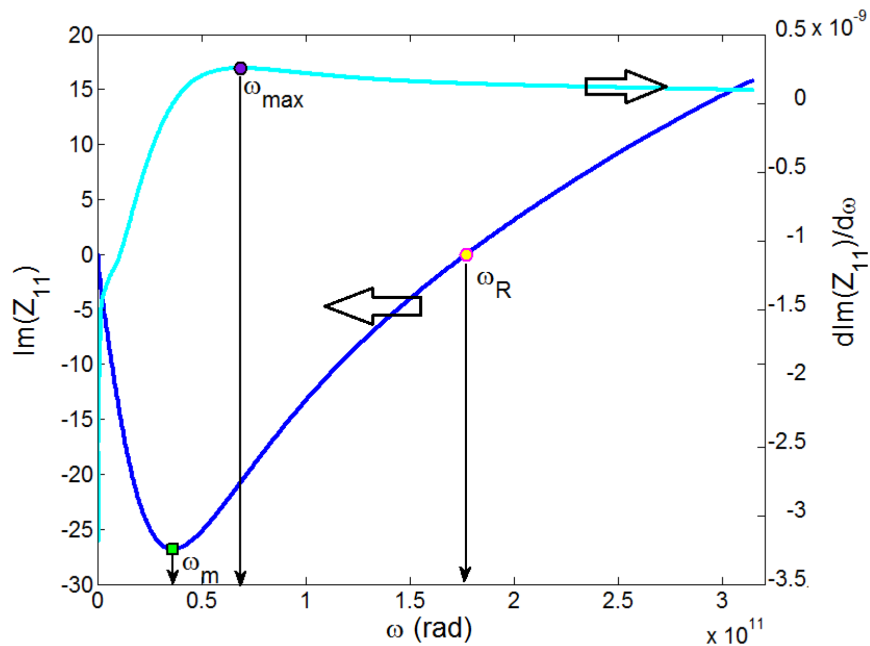


Figure 41. Imaginary part of Z_{11}.of the Cold-FET and its derivative with respect to the frequency.

There are important premises that can be concluded from Figures 40 and 41 shown above:

- The imaginary part of Z_{11} has also a minimum point at $\omega = \omega_m$. Besides, it can be observed that a resonance frequency, which we will call ω_R, occurs when $Im(Z_{11}) = 0$. (Fig. 41).
- The first derivative of the real part of Z_{11} has a minimum at $\omega = \omega_{min}$. (Fig. 40)
- The first derivative of the imaginary part of Z_{11} has a maximum point at $\omega = \omega_{max}$. (Fig. 41)
- Evaluating equation (62) at the resonance frequency ω_R where $Im(Z_{11}) = 0$ it can be defined that

$$\omega_R^2 = \omega_x^2 - \omega_0^2 \tag{71}$$

where,

$$\omega_x = \frac{1}{\sqrt{LC_0}} \tag{72}$$

$$\omega_0 = \frac{1}{\tau_0} = \frac{1}{R_0 C_0} \tag{73}$$

$$L = L_g + L_s \tag{74}$$

So far, ω_{max}, ω_{min}, ω_m and ω_R can be computed from measured data. However, ω_x and ω_0 which are indispensable in the parasitic extraction procedure are still unknown. Next, a procedure for the calculation of ω_0 is explained.

5.3.3.2 ω_0 Calculation

ω_0 is a significant parameter in the parasitic elements calculation. It can be obtained by three deferent ways:

- From the first derivative of $Im(Z_{11})$

The imaginary part of Z_{11} has a minimum point at ω_m which occurs when $\frac{dIm(Z_{11})}{d\omega} = 0$. Therefore, by evaluating the first derivative of $Im(Z_{11})$ at $\omega = \omega_m$ one finds

$$\omega_0 = \omega_m \sqrt{\frac{\omega_R^2 + \omega_m^2}{\omega_R^2 - 3\omega_m^2}} \tag{75}$$

- From the second derivative of $Im(Z_{11})$

The first derivative of the imaginary part of Z_{11} with respect to ω has a maximum at the frequency point ω_{max} we use the criterion of the second derivative, which states that the second derivative of $Im(Z_{11})$ with respect to ω, evaluated at ω_{max} is zero $\frac{d^2Im(Z_{11})}{d\omega^2} = 0$. Then, from the imaginary part of Z_{11}, ω_0 is given by

$$\omega_0 = \frac{\omega_{max}}{\sqrt{3}} \tag{76}$$

- From the second derivative of $Re(Z_{11})$

The first derivative of the real part of Z_{11} with respect to ω has a minimum. At ω_{min}, where the minimum occurs, the second derivative of $Re(Z_{11})$ with respect to ω is zero. Then, another expression for the computation of ω_0 is derived from $\frac{d^2Re(Z_{11})}{d\omega^2} = 0$, and given by

$$\omega_0 = \sqrt{3}\omega_{min} \tag{77}$$

- From the intersection of the second derivative of $Re(Z_{11})$ with the second derivative of $Im(Z_{11})$

The second derivative of Z_{11} with respect to ω is defined as

$$\frac{d^2Z_{11}}{d\omega^2} = 2R_0\tau_0^2\left[\frac{3\omega^2\tau_0^2-1}{\left(1+\omega^2\tau_0^2\right)^3}\right] + j2\omega R_0\tau_0^3\left[\frac{3-\omega^2\tau_0^2}{\left(1+\omega^2\tau_0^2\right)^3}\right] \tag{78}$$

At one point $\frac{d^2Re(Z_{11})}{d\omega^2} = \frac{d^2Im(Z_{11})}{d\omega^2}$, then it is found that

$$\omega^3\tau_0^3 + 3\omega^2\tau_0^2 - 3\omega\tau_0 - 1 = 0 \tag{79}$$

It is obvious that the only frequency point where equation (79) is fulfilled is when $\omega = \omega_0$. Then, ω_0 can also be found at the frequency point where $\frac{d^2Re(Z_{11})}{d\omega^2} - \frac{d^2Im(Z_{11})}{d\omega^2} = 0$. It can be observed in Fig. 42 that at high frequency, the curves are asymptotic to each other, meanwhile, at low frequency where ω_0 occurs, an intersection between the second derivative of the real and imaginary part of Z_{11} is found.

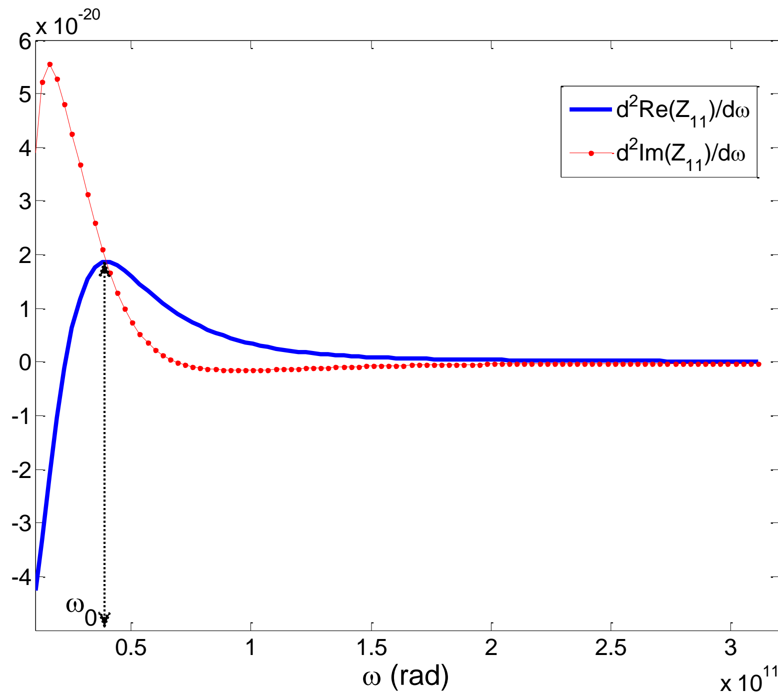


Figure 42. Intersection between the second derivative of the real and imaginary part of Z_{11}.

It is important to notice that the exactitude to calculate the gate parasitic elements depends on the computation of the different values of ω. Since ω is in the range of 10^8 to 10^{11}, a small variation will have a huge impact on R_0, C_0, R_g and L_g. Consequently, accuracy in the numerical differentiation of Z_{11}, as well as, in the interpolation performed to find the frequency points of interest is crucial for a successful extraction. This issue is deeply addressed latter in this chapter.

5.3.3.3 R_0 and C_0 Calculation

So far, the procedure to find essential frequency points has been explained. Next, by taking advantage of such frequencies and with the information provided by the real and imaginary parts of Z_{11} and its derivatives with respect to the ω, equations for the calculation of the gate parasitic elements are given.

The dynamic resistance and capacitance R_0 and C_0 are used to model the gate Schottky diode for the Cold-FET configuration. Since these elements are bias dependent their calculation can be intricate. With the information acquired from Z_{11} of the Cold-FET,

R_0 can be computed from any of the following equations:

$$R_0 = -\frac{8\omega_0}{3\sqrt{3}} \frac{dRe(Z_{11})}{d\omega}\bigg|_{\omega=\omega_{min}} \tag{80}$$

$$R_0 = -2\omega_0 \frac{dRe(Z_{11})}{d\omega}\bigg|_{\omega=\omega_0} \tag{81}$$

$$R_0 = 2\omega_0^2 \frac{d^2Re(Z_{11})}{d\omega^2}\bigg|_{\omega=\omega_0} \tag{82}$$

$$R_0 = 2\omega_0^2 \frac{d^2Im(Z_{11})}{d\omega^2}\bigg|_{\omega=\omega_0} \tag{83}$$

$$R_0 = 8\omega_0 \left[\frac{dIm(Z_{11})}{d\omega}\bigg|_{\omega=\omega_{max}} - L \right] \tag{84}$$

It can be noted that equations (82) and (83) are found from (79). Also, eq (84) requires previous knowledge of $L = L_g + L_s$ which is a constraint.

C_0 can be now calculated from (72)

$$C_0 = \frac{1}{\omega_0 \mathrm{R}_0} \tag{85}$$

5.3.3.4 R_g Calculation

With the previous knowledge of R_0, different expressions to calculate the gate parasitic resistance can be defined, such as

$$R_g^* = Re(Z_{11})|_{\omega=\omega_0} - R_s - \frac{R_0}{2} \tag{86}$$

$$R_g^* = Re(Z_{11})|_{\omega=\omega_{min}} - R_s - \frac{3R_0}{4} \tag{87}$$

$$R_g^* = Re(Z_{11})|_{\omega=\omega_R} - R_s - \frac{R_0}{1 + \omega_R^2 \tau_0^2} \tag{88}$$

5.3.3.5 L_g Calculation

In the same way as R_g, expressions for the calculation of the gate inductance can be found.

$$L = R_0 \frac{\tau_0}{1 + \omega_R^2 \tau_0^2} \tag{89}$$

$$L = R_0 \frac{\omega_0}{\omega_x^2} \tag{90}$$

$$L = \tau_0 \left[Im(Z_{11})|_{\omega=\omega_0} + \frac{R_0}{2} \right] \tag{91}$$

$$L = \tau_0 \left[\frac{1}{\sqrt{3}} Im(Z_{11})|_{\omega=\omega_{max}} + \frac{R_0}{4} \right] \tag{92}$$

$$L = \left. \frac{dIm(Z_{11})}{d\omega} \right|_{\omega=\omega_0} \tag{93}$$

$$\mathrm{L} = \left. \frac{dIm(Z_{11})}{d\omega} \right|_{\omega=\omega_{max}} - \frac{R_0}{8\omega_0} \tag{94}$$

$$L = \frac{1}{C_0 \omega_x^2} \tag{95}$$

Once L is known L_g can be calculated from (74) as $L_g = L - L_s$.

The following algorithm will clarify the steps to select the best values for each gate element:

1. By using interpolation tools, from the forward Cold-FET measurement, find the frequency points of interest ω_R, ω_m, ω_{min} and ω_{max}.
2. Determine ω_0 from (75)-(79).
3. For each value of ω_0, compute ω_x, C_0, R_0, R_g and L_g.
4. Create an error function between the modeled and measured Z_{11} parameter of the forward Cold-FET, where the best values for the gate elements are selected.

5.3.4 Disadvantages of the Method

Although, the methodology to calculate the parasitic elements presented in the previous section is very reliable it has some limitations. Next, a deep explanation of these disadvantages as well as approaches to overcome them is presented.

5.3.4.1 Smoothing Derivative Data Using Artificial Neural Networks

The method presented in this chapter for calculating the parasitic elements of the equivalent circuit model of FETs is quite simple to understand and implement due to the simplistic mathematics involved in it. Also, it can be implemented in different technologies, and due to the inherent low bias requirements of the method, the device is never at risk of irreversible damage. However, one of the drawbacks of the method is the sensitivity in the calculation of particular frequency points of ω. Therefore, measurement accuracy of the Z-parameters of the Cold-FET is required. Besides, first and second order numerical differentiation are needed, hence, the slightest variation can greatly affect the performance of the Cold-FET model. In order to overcome this constraint, a neural network approach is used (Zárate-de Landa et al, 2012). The advantage of such algorithm, as explained in chapter III, is that it takes into account for training the neural network not only measured data, but also, its first derivatives making it an excellent tool for this application.

It has been observed that measured data that is corrupted by noise or errors can be smoothed by using a neural network (Zárate de Landa et al., 2012). Figures 43 and 44 show how measured data is smoothed by using a neural network, this is more noticeable in the first derivative of both the real and imaginary part of Z_{11}. It is interesting to notice that it would be very difficult to accurately determine the exact frequency point in which the extrema points of Z_{11} occur from the measured data. Consequently, using the data provided by the neural network significantly improves the exactitude in which the gate parasitic elements are calculated.

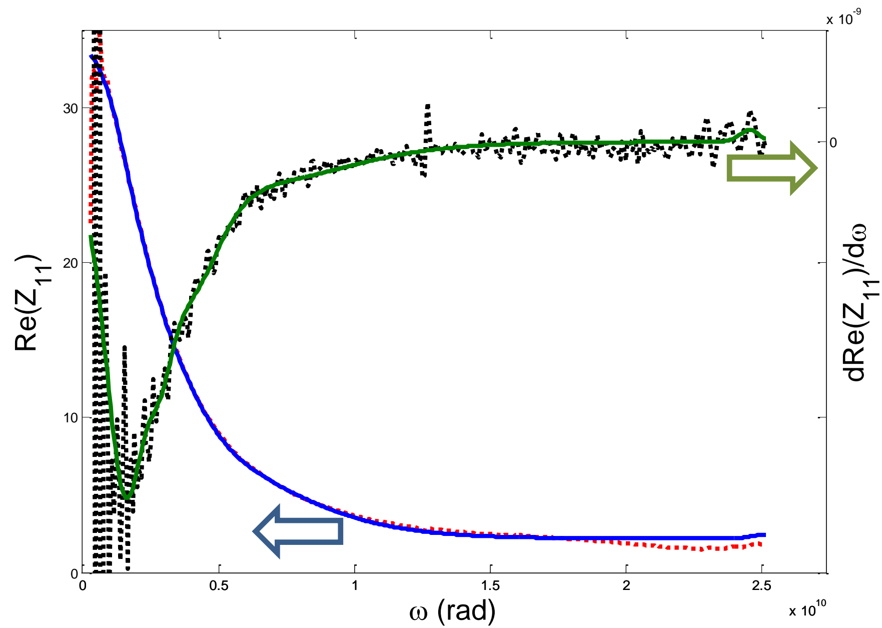


Figure 43. Real part of Z_{11} of the Cold-FET and its first derivative with respect to the frequency. Measured data (dotted line) and Neural Network model (solid line).

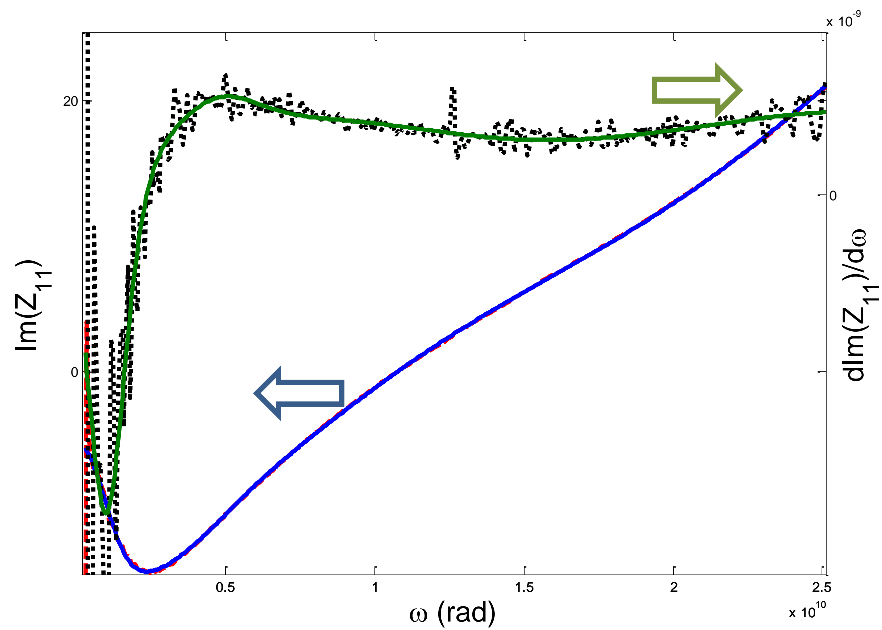


Figure 44. Imaginary part of Z_{11} of the Cold-FET and its first derivative with respect to the frequency. Measured data (dotted line) and Neural Network model (solid line).

5.3.4.2 Low Frequency Dependence

It has been observed that the location of the extrema points is dependent on the gate length of the device, the larger the gate, the smallest is the frequency where the extrema points are located. Hence for high power devices, the extrema points can be positioned in the range of 10-100 MHz, making the measurement dependable on the capabilities of the network analyzer used. Next, a method introduced by Reynoso-Hernández et al. (2009). Thus, if the real part of Z_{12} is subtracted from the real part of Z_{11}, one obtains an expression that can be used to find R_g.

$$Re(Z_{11} - Z_{12}) = R_g + R_0 \frac{{\omega_0}^2}{{\omega_0}^2 + \omega^2} \tag{96}$$

This procedure for calculating R_g assumes that at measurement frequencies $\omega >> \omega_0$ equation (96), can be written as

$$Re(Z_{11} - Z_{12}) = R_g + R_0 \frac{{\omega_0}^2}{\omega^2} \tag{97}$$

Thus, using (97), R_g and ${\omega_0}^2 R_0$ can be determined by applying a linear regression. This method for calculating R_g is very simple since it only requires the knowledge of an approximate value of ω_0 and then the frequency range were $\omega >> \omega_0$ is fulfilled is very easy to find. The main advantage of this method is that the R_g calculation does not depend on R_0, ω_R or ω_0. In a similar way, by subtracting the imaginary part of Z_{12} from the imaginary part of Z_{11} an expression to determine L_g directly is obtained

$$Im(Z_{11} - Z_{12}) = \omega L_g - \frac{1}{C_0} \frac{\omega}{{\omega_0}^2 + \omega^2} \tag{98}$$

Assuming that at measurement frequencies $\omega >> \omega_0$ equation (98), can be written as

$$\omega Im(Z_{11} - Z_{12}) = \omega^2 L_g - \frac{1}{C_0} \tag{99}$$

Then, by using simple linear regression, L_g can be calculated from the slope of $\omega Im(Z_{11} - Z_{12})$ vs ω^2.

Table 3 reports the parasitic resistances and inductances of the transistor studied in this investigation (CREE CGH35015F) measured in Cold-FET with open drain configuration from 1 to 5 GHZ with I_g = 4 mA. For this frequency range the minimum point of $Im(Z_{11})$ cannot be observed. Therefore, the method described in this section for the

calculation of the gate resistance and inductance is used. In Figures 45 and 46 the measured real and imaginary part of Z_{11} is compared to the model obtained from the extraction of R_g and L_g.

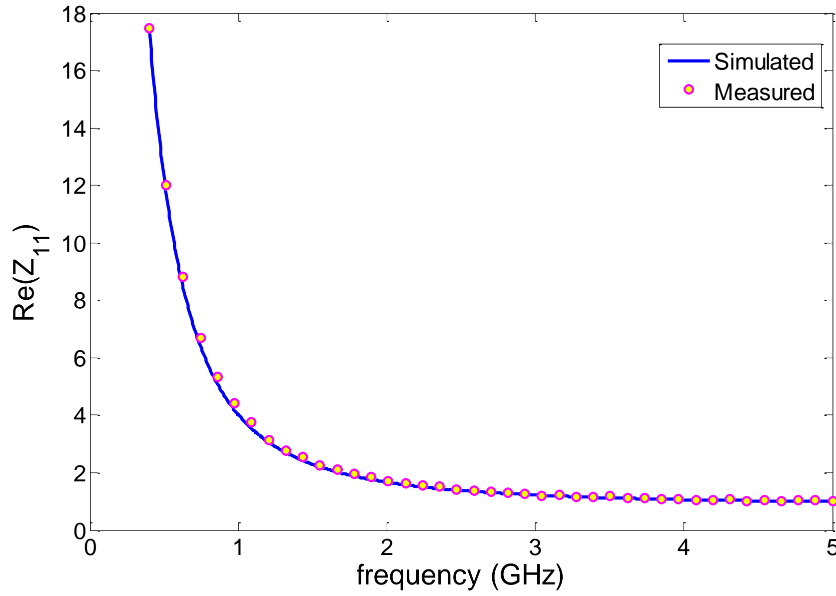


Figure 45. Comparison between the measured and simulated real part of Z_{11}.

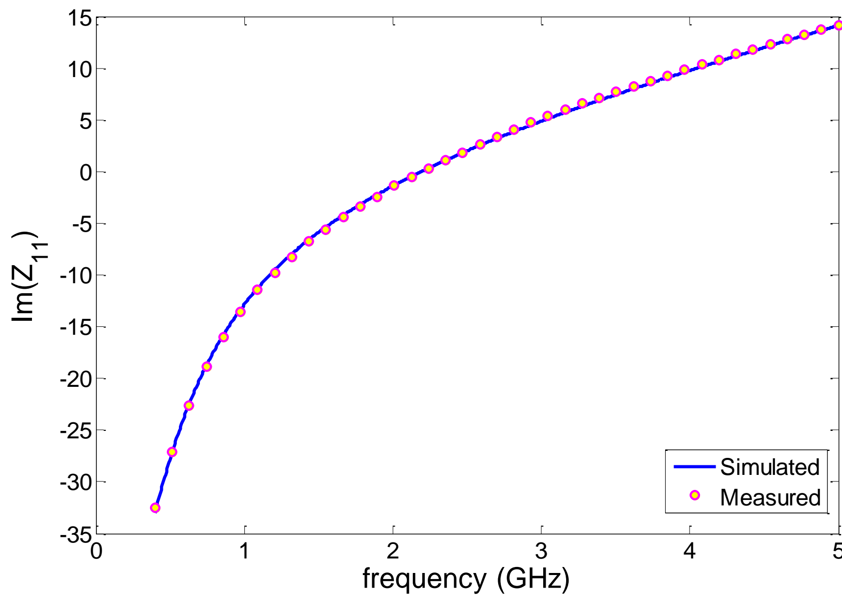


Figure 46. Comparison between the measured and simulated imaginary part of Z_{11}.

Table 3. Parasitic elements of the forward Cold-FET calculated of the CREE CGH35015F AlGaN/GaN HEMT

R_0 (Ω)	Co (pF)	R_g (Ω)	R_s (Ω)	R_d (Ω)	L_g (pH)	L_s (pH)	L_d (pH)
85.45	9.24	0.48	0.38	0.90	536.06	29.61	630.92

5.4 The Pinched-off Cold-FET Model

Parasitic capacitances extraction is based on the Y-parameters of the blocked transistor at DC gate bias beyond pinch-off voltage V_P, and with zero drain –source voltage ($V_{gs} < V_P < 0$; $V_{ds} = 0V$). Indeed, for this condition $V_{ds} = 0$, contrary to the forward Cold-FET, this is due to the fact that drain and source electrodes must be at the same potential, in such a way that the depletion region under the gate is uniform and symmetrical.

Fig. 47 shows the equivalent circuit model pinched-off Cold-FET.

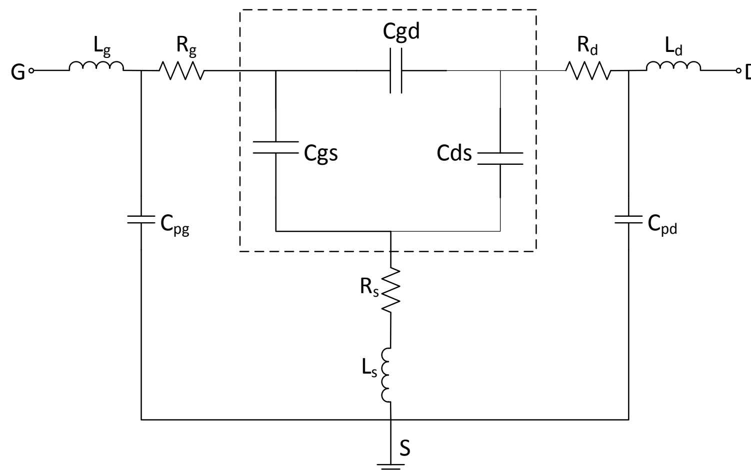


Figure 47. Equivalent Circuit Model of the transistor under pinch-off bias (Vgs << Vp; Vds = 0).

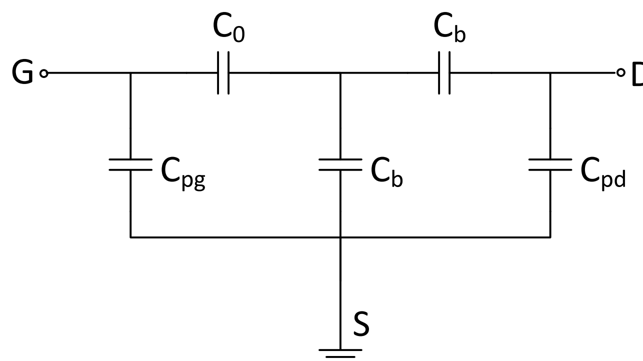


Figure 48. Transformation from T to π network for the pinch-off equivalent circuit model.

Prior to parasitic capacitance calculation, a de-embedding of the parasitic gate and drain inductances L_g and L_d must be performed in order to eliminate their effect on the imaginary part of the inverse Cold FET Y-parameters.

Models presented by Dambrine et al (1988) and White and Healy (1993) where developed for estimating C_{pg} and C_{pd} capacitances. It was demonstrated by Caddemi et al (2006) that the depletion region under the gate of the pinched-off Cold-FET can be modeled using a T-network. Transforming from a Π to a T circuit, the ECM of the pinched-off Cold-FET becomes as depicted in Fig. 48, where parasitic resistances and inductances are neglected since they have no influence on the Y-parameters of the pinched-off Cold-FET. Parasitic capacitances C_{pg}, C_{pd}, and C_b are then calculated from the Y-parameters of the network shown in Fig. 48. The expressions for the parasitic capacitances are:

$$C_{pg} = \frac{Im(Y_{11}) + 2Im(Y_{12})}{\omega} \tag{100}$$

$$C_{pd} = \frac{Im(Y_{22}) + Im(Y_{12})}{\omega} + \frac{C_b}{C_0}\frac{Im(Y_{12})}{\omega} \tag{101}$$

where,

$$C_b = \frac{-C_0 Im(Y_{12})}{\omega C_0 + 2Im(Y_{12})} \tag{102}$$

It should be noted that, for calculating C_{pg} and C_{pd} with equations (100) and (102), it is required the knowledge of C_b and C_0 which was previously determined according to the procedure explained in the previous section. Besides, it is very important to comment that the expression to calculate C_{pg} is exactly the same as published by Dambrine (1988). Nevertheless, the novelty of this method is the C_b/C_0 term in the C_{pd} expression because if $C_b << C_0$ the calculation of C_{pd} would be the same as the one introduced by Dambrine. Nevertheless, if $C_b = C_0$, then C_{pd} would be the same as the equation presented by White (1993). However, we have found that for GaN transistors $C_b \neq C_0$, so that neither Dambrine nor White methods can precisely extract the value of C_{pd} for this kind of transistor. The modeled Y-parameters computed from the extracted parasitic capacitances are displayed in Fig. 49. Table 3

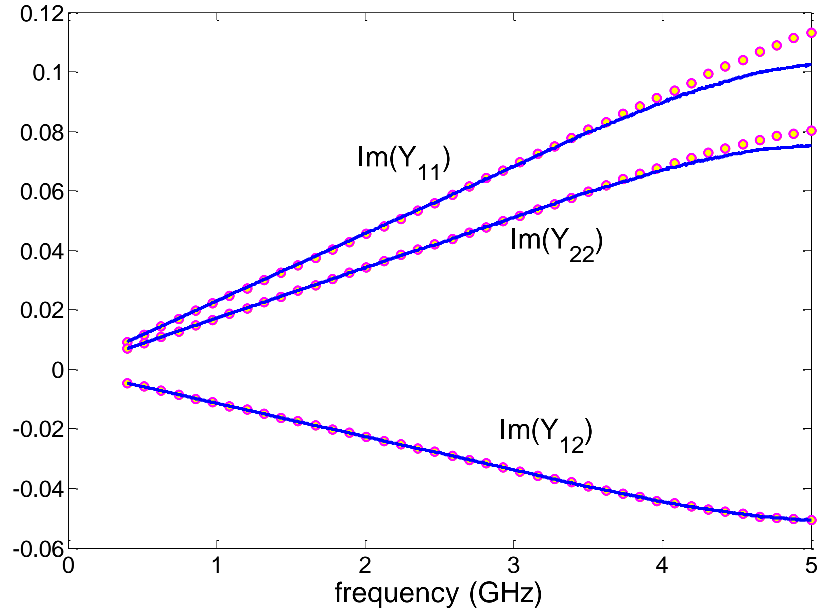


Figure 49. Comparison between measured (circles) and modeled (lines) Y-parameters of the pinched-off Cold-FET (CREE CGH35015F AlGaN/GaN HEMT).

Table 4. Parasitic capacitances calculated of the CREE CGH35015F AlGaN/GaN HEMT

C_b (pF)	C_{pg} (fF)	C_{pd} (fF)
2.88	34.21	347.28

5.5 De-embedding Process

Once the extrinsic elements are calculated, a de-embedding process is performed on the measured S-parameters of the device under different bias conditions (Dambrine et al, 1989). With the knowledge of the intrinsic Y-parameters, the intrinsic elements (C_{gs}, C_{gd}, C_{ds}, R_i, R_{ds}, g_{m0} y τ) can be determined as explained in the next section. In order to obtain the intrinsic Y-parameters the de-embedding process depicted in Fig. 50 must be performed.

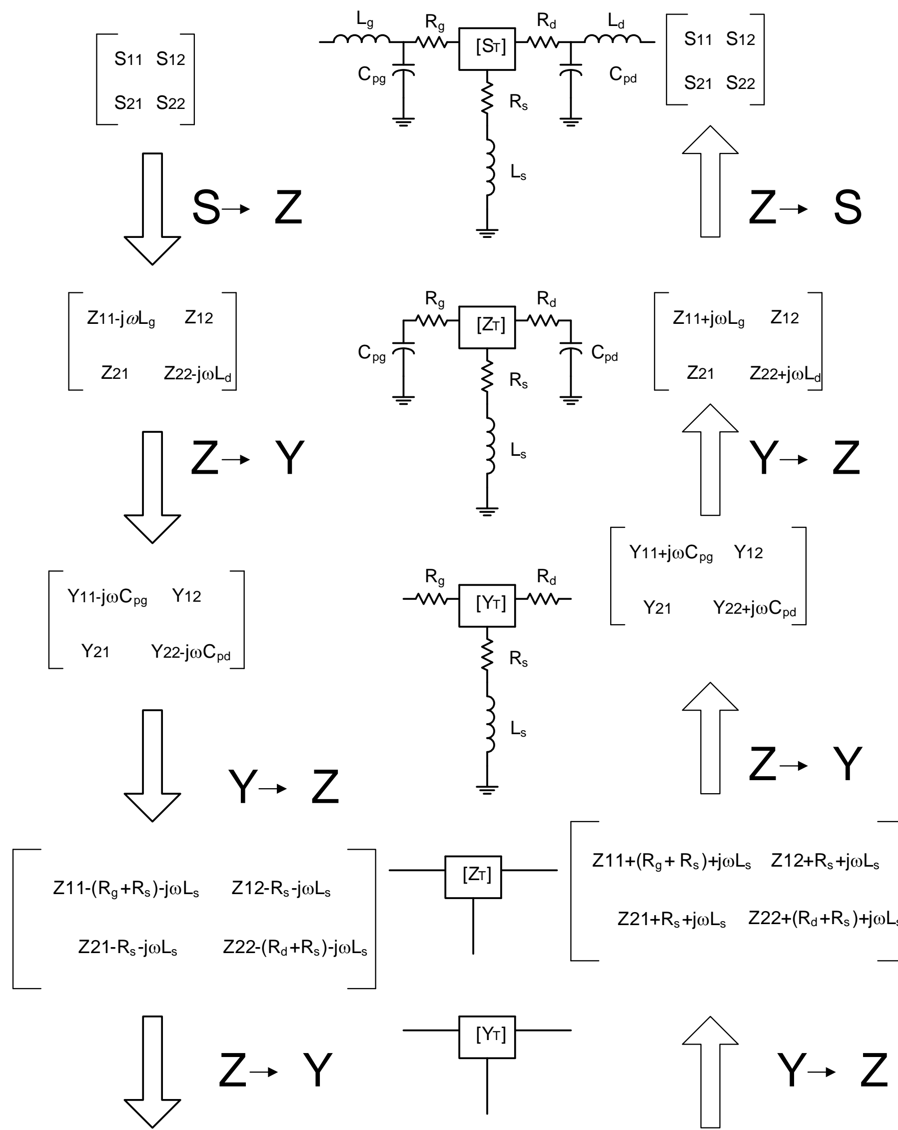


Figure 50. De-embedding process used for the computation of the intrinsic Y-parameters.

5.6 Intrinsic Elements Extraction

In 1990 Berroth and Bosh proposed an equivalent circuit model of the intrinsic expressed in the form shown in Fig. 51. In this network the intrinsic transistor is represented by electrical elements such as resistors and capacitors that are related to the physical properties of the device, as illustrated in Fig 52. Furthermore, a controlled current source is added to represent the active device as a transadmittance.

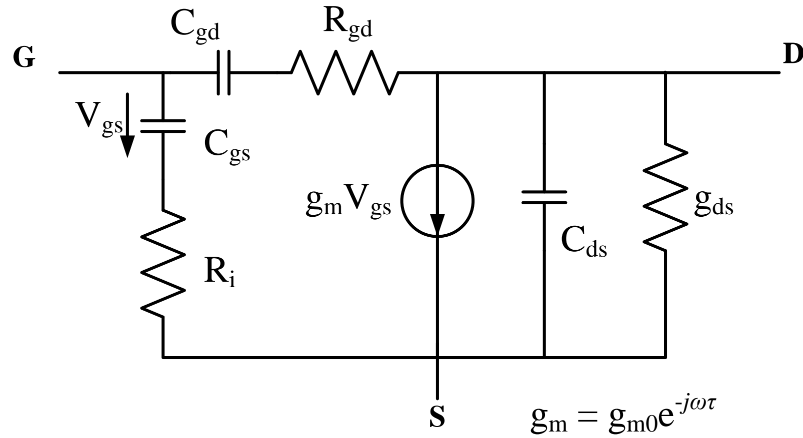


Figure 51. Small-signal equivalent circuit model of the intrinsic transistor.

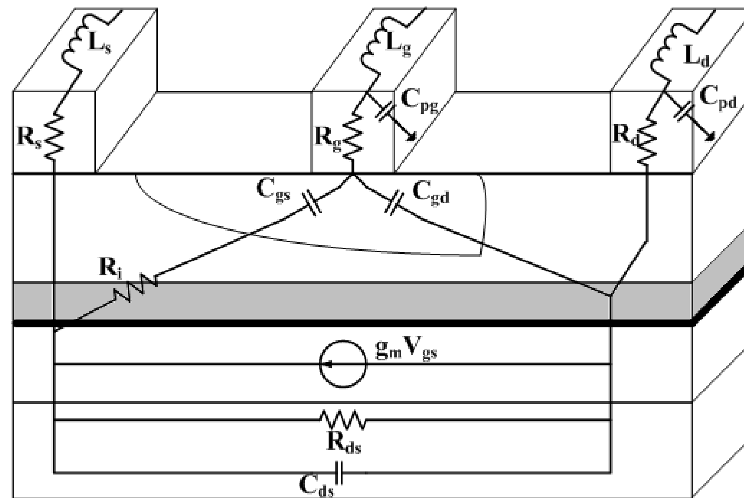


Figure 52. Physical origins of the intrinsic elements of a FET.

The Y-parameters corresponding to the equivalent circuit model of Fig. 51 are defined as

$$[Y_T]=\begin{bmatrix} \omega^2\left(\frac{R_iC_{gs}^2}{1+\omega^2C_{gs}^2R_i^2}+\frac{R_{gd}C_{gd}^2}{1+\omega^2C_{gd}^2R_{gd}^2}\right)+j\omega\left(\frac{C_{gs}}{1+\omega^2C_{gs}^2R_i^2}+\frac{Cgd}{1+\omega^2C_{gd}^2R_{gd}^2}\right) & -\frac{\omega^2C_{gd}^2R_{gd}}{1+\omega^2C_{gd}^2R_{gd}^2}-j\omega\frac{C_{gd}}{1+\omega^2C_{gd}^2R_{gd}^2} \\ \frac{g_m}{1+\omega^2C_{gs}^2R_i^2}-\frac{\omega^2C_{gd}^2R_{gd}}{1+\omega^2C_{gd}^2R_{gd}^2}-j\omega\left(\frac{g_mC_{gs}R_i}{1+\omega^2C_{gs}^2R_i^2}+\frac{C_{gd}}{1+\omega^2C_{gd}^2R_{gd}^2}\right) & g_{ds}+\frac{\omega^2C_{gd}^2R_{gd}}{1+\omega^2C_{gd}^2R_{gd}^2}+j\omega\left(C_{ds}+\frac{C_{gd}}{1+\omega^2C_{gd}^2R_{gd}^2}\right) \end{bmatrix} \quad (103)$$

Following the extraction method proposed by Berroth and Bosh (1990) the values of the intrinsic elements can be obtained as

$$g_{ds} = \mathrm{Re}(Y_{22}) + \mathrm{Re}(Y_{12}) \tag{106}$$

$$C_{ds} = \frac{\mathrm{Im}(Y_{22}) + \mathrm{Im}(Y_{12})}{\omega} \tag{107}$$

$$C_{gs} = \frac{[\mathrm{Im}(Y_{11}) + \mathrm{Im}(Y_{12})]^2 + [\mathrm{Re}(Y_{11}) + \mathrm{Re}(Y_{12})]^2}{\omega[\mathrm{Im}(Y_{11}) + \mathrm{Im}(Y_{12})]} \tag{108}$$

$$R_i = \frac{\mathrm{Re}(Y_{11}) + \mathrm{Re}(Y_{12})}{[\mathrm{Im}(Y_{11}) + \mathrm{Im}(Y_{12})]^2 + [\mathrm{Re}(Y_{11}) + \mathrm{Re}(Y_{12})]^2} \tag{109}$$

$$C_{gd} = -\frac{\mathrm{Im}(Y_{12})}{\omega}\left[1 + \frac{\mathrm{Re}(Y_{12})^2}{\mathrm{Im}(Y_{12})^2}\right] \tag{110}$$

$$R_{gd} = -\frac{\mathrm{Re}(Y_{12})}{\mathrm{Im}(Y_{12})^2 + \mathrm{Re}(Y_{12})^2} \tag{111}$$

$$g_{m0} = \sqrt{\left\{[\mathrm{Re}(Y_{21}) - \mathrm{Re}(Y_{12})]^2 + [\mathrm{Im}(Y_{21}) - \mathrm{Im}(Y_{12})]^2\right\}\left(1 + \omega^2 R_i^2 C_{gs}^2\right)} \tag{112}$$

$$\tau = -\frac{1}{\omega}\arctan\left[\frac{Y + \omega X R_i C_{gs}}{X - \omega Y R_i C_{gs}}\right] \tag{113}$$

where

$$X = \mathrm{Re}(Y_{21}) - \mathrm{Re}(Y_{12}) \tag{114}$$

$$Y = \mathrm{Im}(Y_{21}) - \mathrm{Im}(Y_{12}) \tag{115}$$

As mentioned previously, for this thesis a 15W AlGaN/GaN (CGH35015) HEMT developed by CREE® with V_{th} = -3 V was used. Pulsed IV/RF measurements were performed using an Auriga AU4750 Pulsed IV/RF System along with pulse I/V head, AU4750-0004 and AU4750-0007A, at the gate and drain of the transistor respectively. The

pulse width was configured to 10 µs with 1% duty cicle and quiescent point V_{gsq} = -3 V and V_{dsq} = 40 V. At each bias point, the S-parameters of the device were measured using Agilent's PNA-N5242. Fig. 53 depicts the intrinsic elements calculated for V_{gs} measured in the range of -3.8 V to 0 V with 0.25 V step and V_{ds} from 0 V to 45 V with steps of 0.5 V.

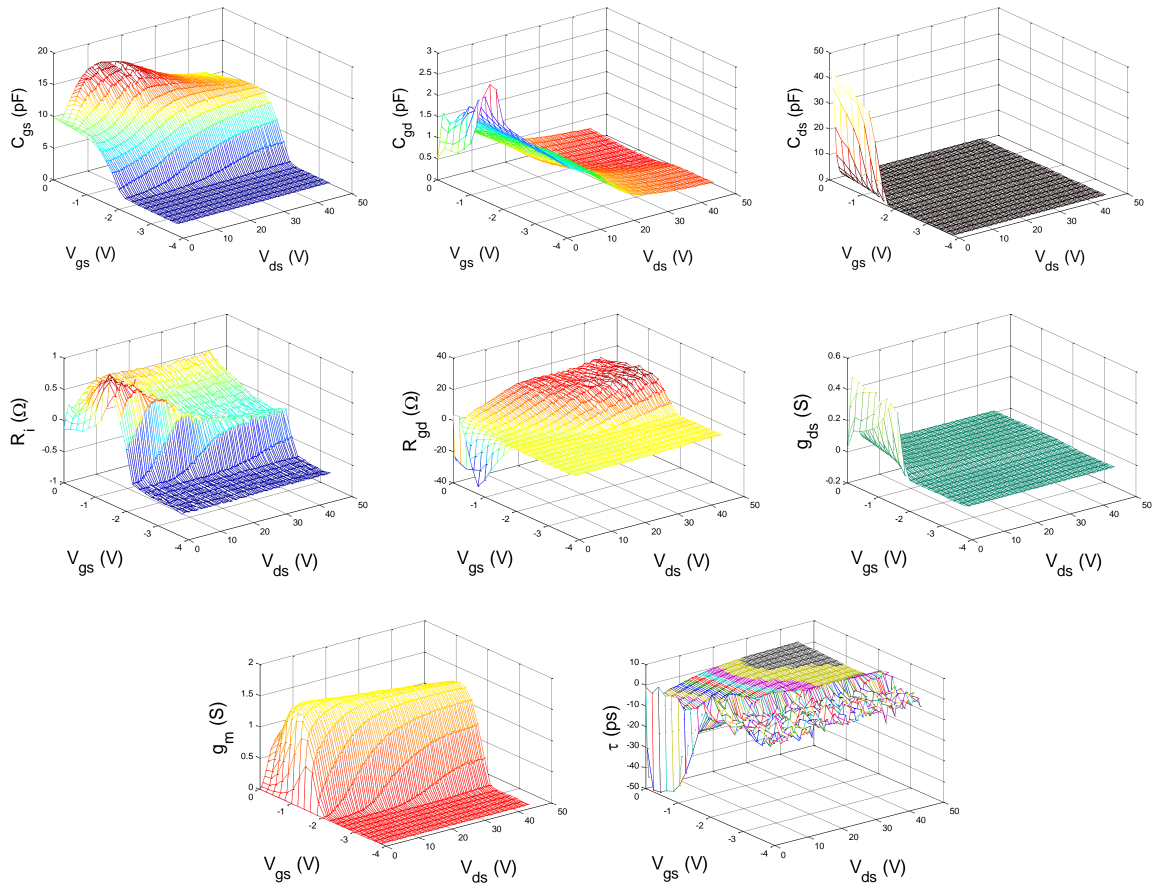


Figure 53. Intrinsic elements calculated at different bias points.

Chapter 6 - Large Signal Modeling

6.1 Introduction

So far, the physics and principles of operation of AlGaN/GaN HEMTs have been described, and how such principles need to be considered in the context of developing a compact model of the device for use in a circuit simulator to aid power amplifier design. Also, Chapter V thoroughly described how to characterize and model the transistor package and its parasitic elements, how to de-embed these extrinsic elements in order to reach the intrinsic transistor. Furthermore, based on the Y-parameters obtained from measurements at different bias conditions, the computation of the intrinsic elements of the small-signal model was explained. Now the attention will be placed on the intrinsic transistor itself: the active semiconductor channel.

The "transistor action" in the FET occurs in the active channel that lies under the gate: this is the place where the current amplification happens. This part of the transistor is called the intrinsic device. The intrinsic large-signal model that is developed in this work is quasi-static. In other words, the model can describe the transistor's RF frequency dependence through the various reactive components in the model, such as the intrinsic capacitances at the gate and drain of the device, and the extrinsic inductance and capacitance associated with the manifold and extrinsic parts of the transistor.

This chapter presents the theoretical foundations of large-signal device modeling for nonlinear circuit simulation. Application of artificial neural networks in the modeling of the drain current, and terminal charges is going to be explained, as well as, a comprehensive discussion on terminal charge conservation.

6.2 The Intrinsic Nonlinear Model

The separation of a circuit-level transistor model into intrinsic and extrinsic parts is an idealization that simplifies the treatment of an otherwise very complicated device.

Conceptually, the intrinsic model describes the dominant nonlinearities of the transistor that occur in the active region, inside the feed networks, manifolds, and other parasitic particularities of the layout. For FETs, the intrinsic model includes that part of the active drain-source channel controlled by the gate and modulated by gate–source and drain–source voltages. The channel current from drain to source and charge storage between the gate and channel are the dominant phenomena, represented in the intrinsic model by nonlinear circuit elements, I_{ds}, Q_{gs}, and Q_{ds}, respectively, within the dashed box of Fig. 54. Other elements can be added to account for gate leakage and breakdown in reverse bias and forward gate conduction at large forward bias conditions. In FET models, parasitic elements are usually modeled with simple circuit elements whose parameter values do not change with bias. The parasitic elements that make up the extrinsic model are usually associated with capacitive coupling between the electrodes, and inductance and resistance of feed structures and manifold metallization, dependent on the device layout (Wood and Root, 2006). Even the FET access resistances, those parts of the semiconducting channel outside the control of the gate, are usually modeled by simple resistors with fixed values, R_s and R_d, independent of the current or the voltage applied.

Procedurally, the intrinsic model can also be defined as that part of the model that remains after the parasitic elements have been identified and removed. Failure to account properly for parasitics can therefore negatively impact the modeling of the intrinsic device. Both perspectives on the intrinsic model should be consistent for a reliable and useful overall model (Rudolph, Fager, and Root, 2012).

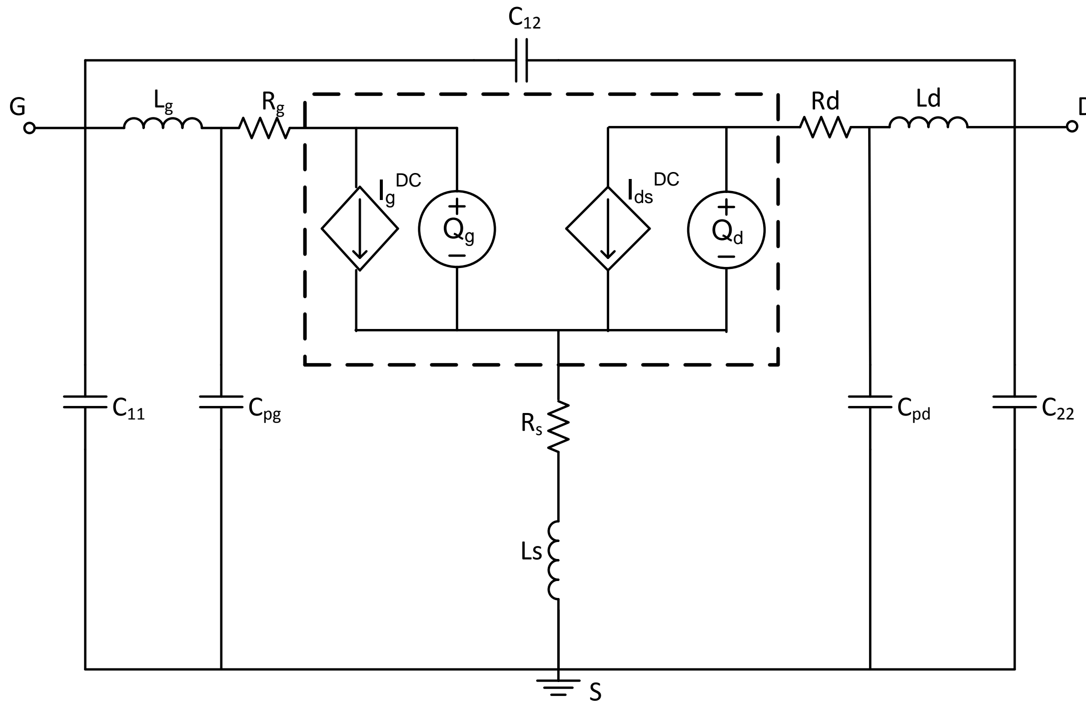


Figure 54. FET model schematic diagram showing the package capacitances, parasitic elements and the intrinsic quasi-static transistor model.

6.2.1 Measurement Approach to Nonlinear Modeling

In this section it is explained how measured small-signal data is used directly to obtain the nonlinear model. This procedure was first explained by Närhi in 1996 where it is assumed that the currents at the two terminals of the intrinsic transistor as illustrated in Fig. 54, driven with large-signal voltages $v_1(t)$, $v_2(t)$ can be written as.

$$i_i(\mathrm{t}) = g_i^{(0)}(v_1, v_2) + \frac{d}{dt} q_i^{(1)}(v_1, v_2) + \frac{d^2}{dt^2} q_i^{(2)}(v_1, v_2) + \frac{d^3}{dt^3} q_i^{(3)}(v_1, v_2) \qquad (116)$$

These is the conventional quasi-static formulation where only the first two terms of the series expansion are included, namely the static current through a nonlinear conductance $g_i^{(0)}$ and the first order dynamic current through a nonlinear capacitance, $dq_i^{(1)}/dt$.

For the quasi-static model it is considered that the model is time-invariant (i.e., $g_i^{(0)}$ and $q_i^{(1)}$ are not explicit function of time), but their time dependence is solely through the dependence on the two controlling voltages. In addition, it is assumed that the partial derivatives of this functions

$$h_{ij}^{(0)}(v_1, v_2) = \frac{\partial g_i^{(0)}}{\partial v_j} \qquad h_{ij}^{(1)}(v_1, v_2) = \frac{\partial q_i^{(1)}}{\partial v_j} \qquad i, j = 1,2 \tag{117}$$

depend only on the instantaneous voltages $v_1(t)$, $v_2(t)$, and not on their time derivatives. Therefore, the small-signal response of the device at port i, at DC bias point V_1, V_2, to a small variation in the voltages $dv_1(t)$, $dv_2(t)$ in the following form

$$di_i(t) = h_{i1}^{(0)} dv_1 + h_{i2}^{(0)} dv_2 + h_{i1}^{(1)} d\dot{v}_1 + h_{i2}^{(1)} d\dot{v}_2, \tag{118}$$

where $d\dot{v}_i$ is related to the port derivative with respect to time. Moving to the frequency domain, the response to a small sinusoidal excitation $dv_1(f)$, $dv_2(f)$, at frequency f.

$$di_i(f) = \left[h_{i1}^{(0)} + j\omega h_{i1}^{(1)}\right] dv_1(f) + \left[h_{i2}^{(0)} + j\omega h_{i2}^{(1)}\right] dv_2(f) \tag{119}$$

By observing eq (119) one can realize that this expression can be compared to the measured Y-parameters of the device at a DC bias point.

$$di_i(f) = Y_{i1}(V_1, V_2, \omega) dv_1(\omega) + Y_{i2}(V_1, V_2, \omega) dv_2(\omega) \tag{120}$$

Then, the large signal functions in eq. (116) can be calculated from the path-independent line integrals

$$g_i^{(0)}(v_1, v_2) = I_{i0}(V_1, V_2) + \int_{V_1}^{v_1(t)} h_{i1}^{(0)}(v_1, V_2) dv_1 + \int_{V_2}^{v_2(t)} h_{i2}^{(0)}(v_1(t), v_2) dv_2 \tag{121}$$

$$q_i^{(1)}(v_1, v_2) = \int_{V_1}^{v_1(t)} h_{i1}^{(1)}(v_1, V_2) dv_1 + \int_{V_2}^{v_2(t)} h_{i2}^{(1)}(v_1(t), v_2) dv_2 \tag{122}$$

Since it is considered a quasi-static model, it is assumed that the DC characteristics do not change with frequency (which is just an initial approximation dispersion effects should also be considered), the nonlinear conductance can be described by the measured I-

V characteristics of the transistor. Further, from the FET equivalent circuit model of Fig. 54, the intrinsic model dynamic equations (121) and (122) can be simplified by:

$$i_g(t) = I_{gs}\left(V_{gs}(t), V_{ds}(t)\right) + \frac{\partial Q_g\left(V_{gs}(t), V_{ds}(t)\right)}{\partial t} \tag{123}$$

$$i_d(t) = I_{ds}\left(V_{gs}(t), V_{ds}(t)\right) + \frac{\partial Q_d\left(V_{gs}(t), V_{ds}(t)\right)}{\partial t} \tag{124}$$

Due to the inherent reverse-bias junction in HEMTs, the gate current is several orders of magnitude smaller than the drain current making it negligible, simplifying then, the nonlinear model. Thus, the intrinsic nonlinear model that will be developed in this chapter will consist on three neural networks describing the drain current and gate and drain charges.

6.3 Neural Network Model of the Drain Current

The intrinsic model constitutive relations are defined on the set of intrinsic voltages, V_{gs}^{int} and V_{ds}^{int} after accounting for the voltage drop across the parasitic resistances. Measured I–V data, on the other hand, are defined on the applied (extrinsic) voltages. The relationship between extrinsic and intrinsic voltages is simple, given the resistive parasitic element values, previously extracted, and the simple equivalent circuit topology. The equations are given in eq (118) (Root, 1999). An important issue for table models is that the extrinsic voltages at which the measurements are taken are usually defined on a grid, but the resulting intrinsic voltages, explicitly computed by substitution using eq. (118), do not fall on a grid, as shown in Fig. 55, and therefore cannot be directly tabulated. Solving (118) in this sense enables the data to be re-gridded on the intrinsic space so that the terminal currents can be tabulated as functions of the intrinsic voltages.

$$\begin{bmatrix} V_{gs}^{int} \\ V_{ds}^{int} \end{bmatrix} = \begin{bmatrix} V_{gs}^{ext} \\ V_{ds}^{ext} \end{bmatrix} - \begin{bmatrix} R_g + R_s & R_s \\ R_s & R_d + R_s \end{bmatrix} \begin{bmatrix} I_{gs}^{DC} \\ I_{ds}^{DC} \end{bmatrix} \tag{125}$$

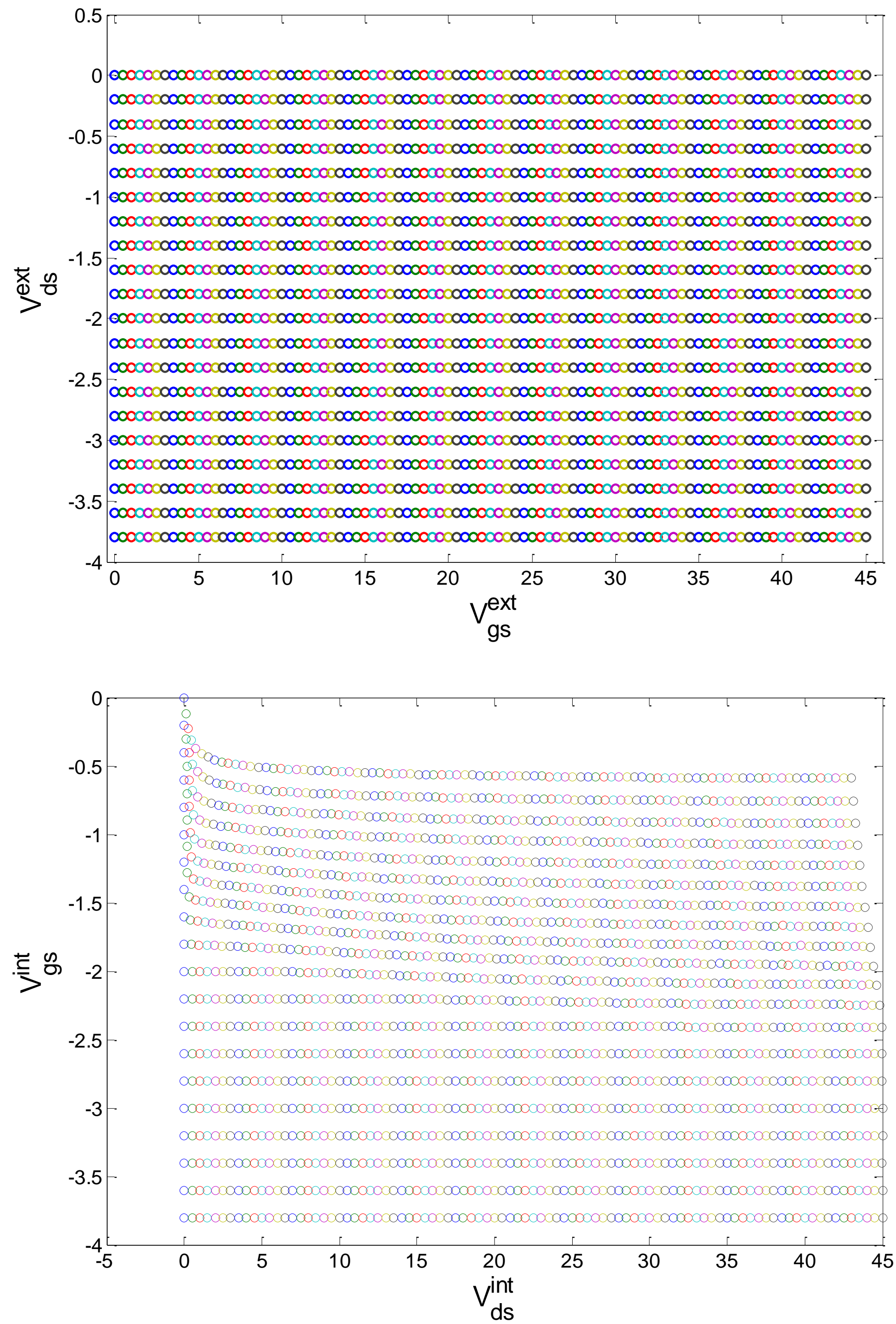
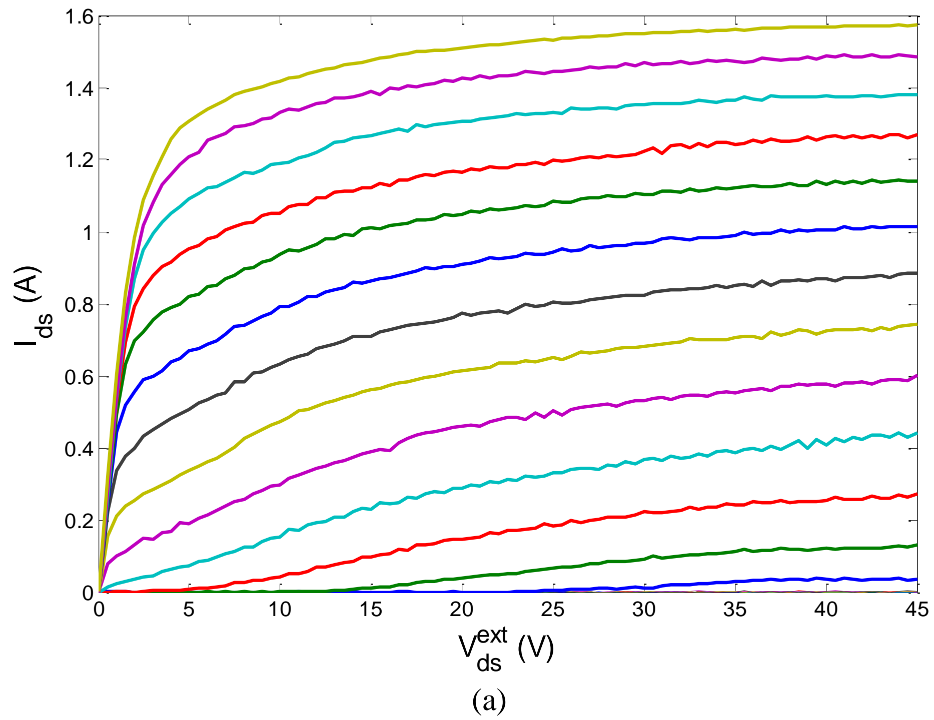
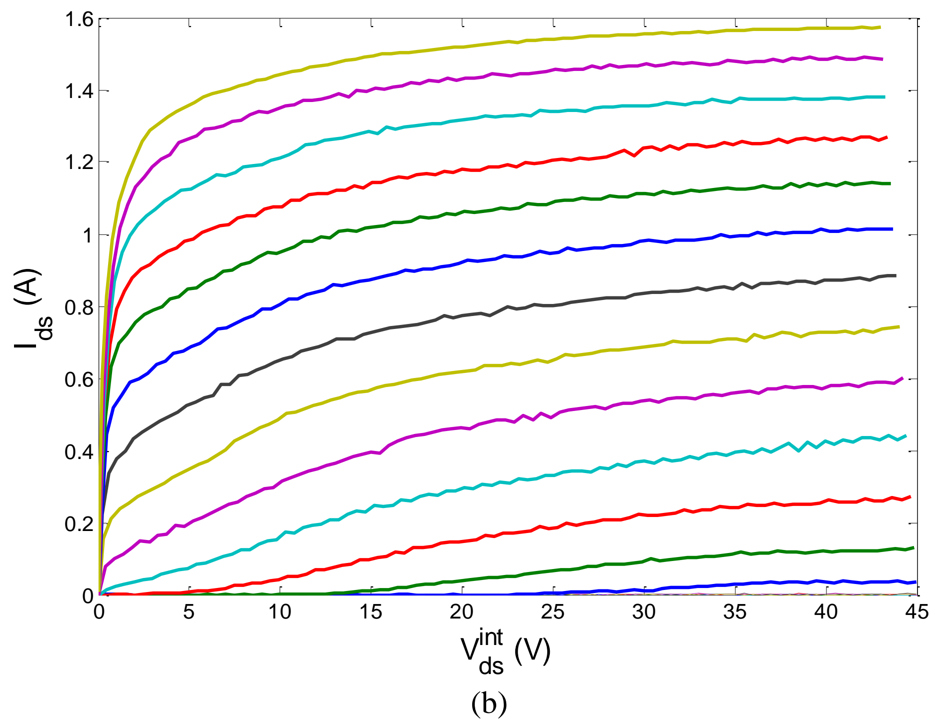


Figure 55. Extrinsic (gridded) and corresponding (non-gridded) voltage domain of a FET.



(a)



(b)

Figure 56. FET model IV constitutive relation expressed as functions of (a) extrinsic and (b) intrinsic voltages.

Modeling the measured I–V data as functions of the intrinsic voltages reveals characteristics quite different from the model expressed in terms of extrinsic data. This is shown in Fig. 56. In Figure 56a, the modeled I–V curves as functions of the applied (extrinsic) voltages V_{gs}^{ext} and V_{ds}^{ext} are plotted. In Figure 56b, intrinsic I-V modeled constitutive relations, defined on V_{gs}^{int} and V_{ds}^{int}, are plotted. There is a big difference between Figure 5.6a and 5.6b, especially around the knee of the curves. This process also makes clear that errors in parasitic extraction can distort the characteristics that we would otherwise attribute to the intrinsic model. In addition to poor interpolation properties of table based models, this gridding constraint can be circumvented by using artificial neural networks as will be demonstrated next.

With the intrinsic voltages computed, the next step is to determine the ANN architecture that will be used to approximate the drain current. For the CGH35015 transistor studied in this thesis a [2-25-25-1] ANN was trained using the logistic function as nonlinearity in the hidden layers. Also, the modified backpropagation introduced in section IV.5 (Zárate-de Landa et al, 2012) was applied as training method in conjunction to the iRprop (Igel and Husken, 2000) as minimum search method. Fig. 57 illustrates the neural network architecture used to approximate the drain current. Since the higher order derivatives of I_{ds} are important in the prediction of harmonic signals in the nonlinear simulation, it is important to accurately approximate not only I_{ds} but also g_{ds}, g_m and its higher order derivatives. Hence, both ρ_0 and ρ_1 were set to 1 in the modified backpropagation algorithm to take into account in training information of the function to be approximated (I_{ds}) and its derivatives with respect to the inputs (g_m and g_{ds}). The error function is propagated backwards in order to update the network weights. Results obtained from the trained ANN are shown in Figs 58-62.

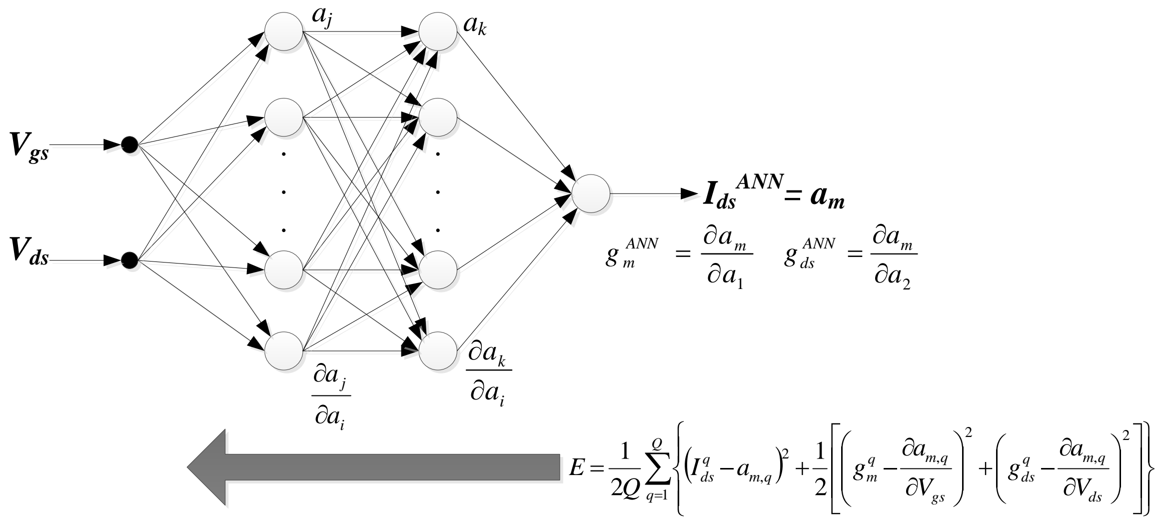


Figure 57. Four-layer ANN used to model the drain current.

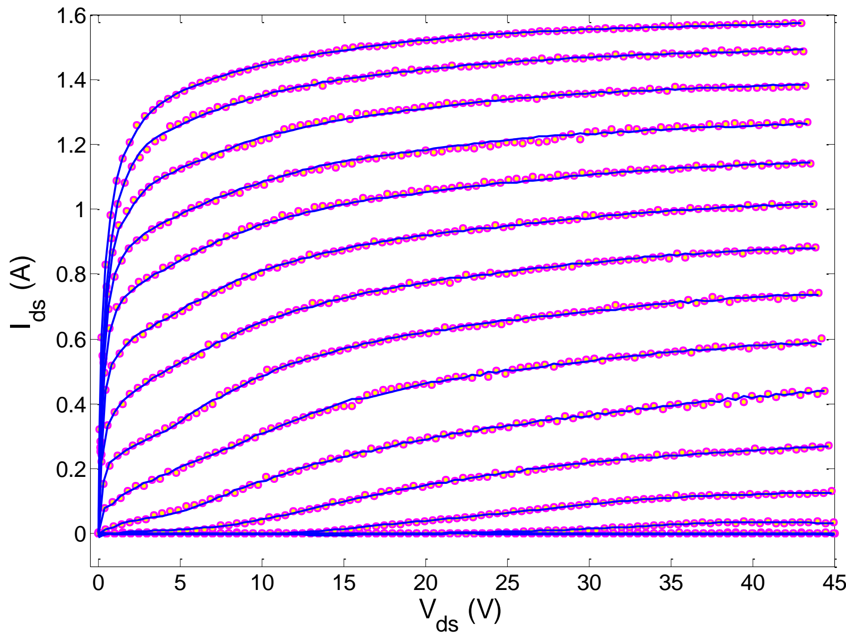


Figure 58. Comparison between measured (circles) and ANN modeled (solid lines) drain current.

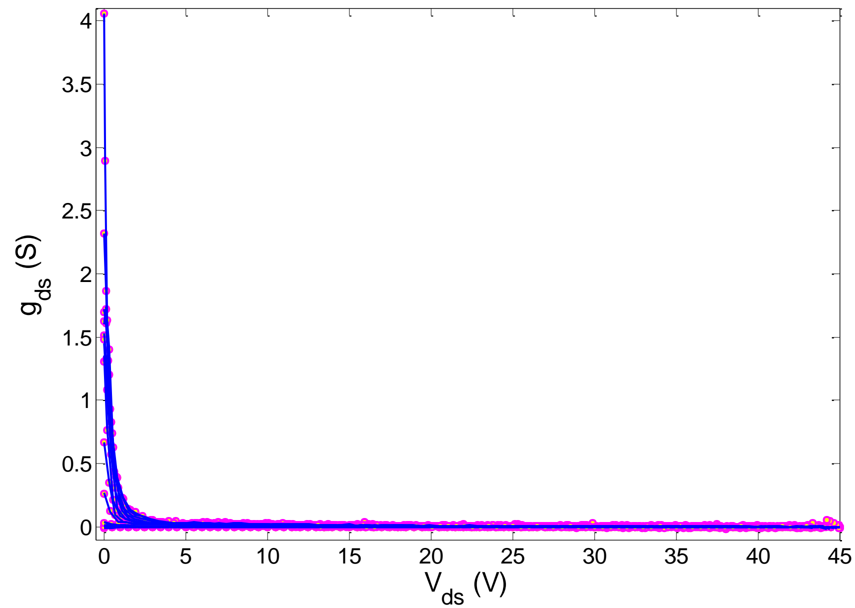


Figure 59. Comparison between measured (circles) and ANN modeled (solid lines) output conductance.

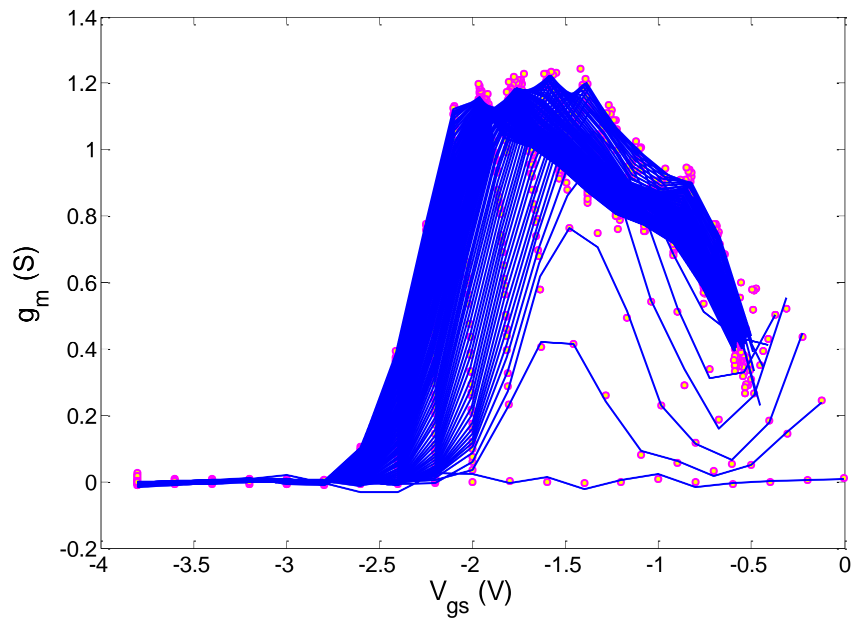


Figure 60. Comparison between measured (circles) and ANN modeled (solid lines) transconductance.

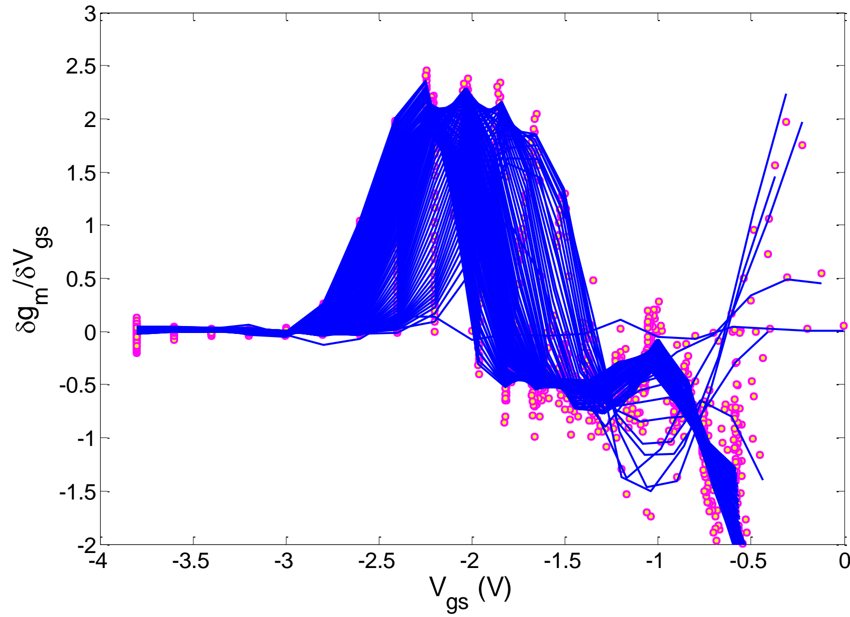


Figure 61. Comparison between measured (circles) and ANN modeled (solid lines) of the first derivative of g_m.

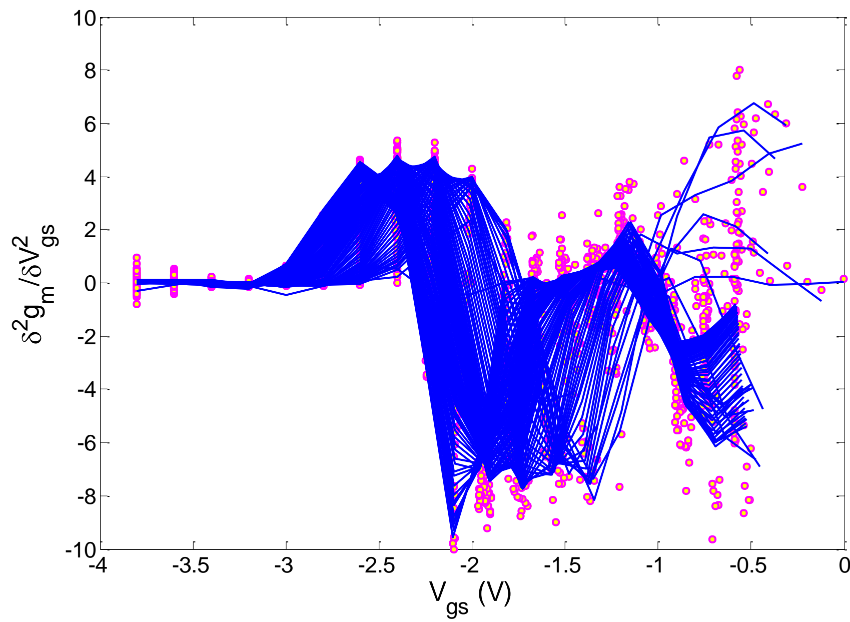


Figure 62. Comparison between measured (circles) and ANN modeled (solid lines) of the second derivative of g_m.

6.4 Charge Modeling

Determination of charge state functions is a critical step in the nonlinear modeling process due to its decisive influence in the prediction of bias-dependent high frequency S-parameters, intermodulation distortion and ACPR in FETs (Rudolph et al, 2012).

The charge modeling problem can be simply stated as the specification of the nonlinear constitutive relations defining the independent terminal charges at the (intrinsic) nodes of the circuit model, as functions of the relevant independent controlling variables, usually voltages. The charge-based contribution to the current at the ith terminal is then the total time derivative of the charge function, Qi. This is expressed in equation (126).

$$I_i(t) = \frac{dQ_i\left(V_{gs}(t), V_{ds}(t)\right)}{dt}, \qquad (126)$$

where, i = 1,2 represent the intrinsic ports. According to eq. (126) the charge constitutive relations contribute to the current model through the time operator. Therefore, it can be said that the charge plays a crucial part in the nonlinear model as the frequency increases.

6.4.1 Charge Conservation Law

The conservation of charge states that charge can be neither created nor destroyed, and this concept is described by the continuity equation from electromagnetic theory

$$\nabla \cdot \boldsymbol{J} = -\frac{\partial \rho}{\partial t} \qquad (127)$$

where **J** is the current density across a surface, and ρ is the charge density in the volume enclosed by the surface. In the derivation of the state variable for the terminal charges, Q_i, above, it is used the notion of conservation in terms of a conservative field of capacitance. Using electromagnetic (or electrostatic) field theory to illustrate this concept, the most familiar example of a conservative field is the electric field. A two-dimensional electric field in x and y is shown in Fig. 63; there are two arbitrary locations A and B in this field. If we integrate along any path from A to B, we obtain the same potential difference:

$$\underset{contour1}{\int_A^B} \vec{E} \cdot \overrightarrow{dl} = \underset{contour1}{\int_A^B} \vec{E} \cdot \overrightarrow{dl} = V_{BA} \qquad (128)$$

Or, if the integration was performed around the closed contour, from A to B along contour 1 back to A along contour 2, then we end up at the same potential:

$$\oint_C \vec{E} \cdot \overrightarrow{dl} = 0 \tag{129}$$

and the electric field conserves potential energy. Also, a conservative field is an irrotational field: the curl of the field vector is zero. For the two-dimensional electric field the curl is written as:

$$curl(\vec{E}) = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \tag{130}$$

By replacing the electric field by a capacitance field, and the x-y coordinates become vectors in the directions of V_{gs} and V_{ds} then

$$curl_{\vec{V}}(\overrightarrow{C_\iota}) = \frac{\partial C_{i1}(V_{gs,}V_{ds})}{\partial V_{ds}} - \frac{\partial C_{i2}(V_{gs,}V_{ds})}{\partial V_{gs}} = 0 \tag{131}$$

where

$$\vec{V} = V_{gs}\vec{u}v_{gs} + V_{gs}\vec{u}v_{ds} \tag{132}$$

From equations (121)-(122) it can be concluded that the capacitive field in the transistor terminals is defined by the imaginary part of the Y-parameters of the device. Hence, eq. (131) can be written as

$$curl_{\vec{V}}(\overrightarrow{C_\iota}) = \frac{\partial Im[Y_{i1}(V_{gs,}V_{ds})]}{\partial V_{ds}} - \frac{\partial Im[Y_{i2}(V_{gs,}V_{ds})]}{\partial V_{gs}} = 0 \tag{133}$$

Then, in order to consider that in a closed contour in this field would result in no change in the charge, the following expression must be fulfilled

$$\frac{\partial Im[Y_{i1}(V_{gs,}V_{ds})]}{\partial V_{ds}} = \frac{\partial Im[Y_{i2}(V_{gs,}V_{ds})]}{\partial V_{gs}}, \tag{134}$$

meaning that charge is not being created nor destroyed, it is conserved.

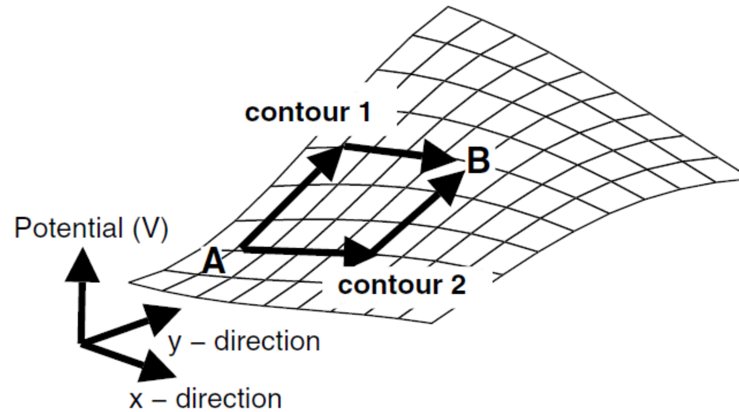


Figure 63. Two-dimensional electric field. (Aaen Peter H., Plá Jaime A. and Wood John, 2007, p. 226)

6.4.2 Neural Network Model of the Gate and Drain Charge Functions

For the gate and drain charges model a similar ANN approach as for the drain current is performed. By taking advantage of the modified backpropagation explained in section IV.5, a neural network model of the charge functions can be obtained straightforward from the bias dependent Y-parameters of the device computed from measurements.

Let's take a pause to point an important property of the modified backpropagation algorithm that will be of significant assistance in the determination of the ANN model for the charge functions. In summary, the backpropagation algorithm consist in taking the input/output relationship (**X**,**D**) of the ANN and computing the mean square error which will be sent back through the network as a feedback, which will be used to update the network weights at each iteration. For the modified backpropagation (Zárate-de Landa et al, 2012), the cost function (mean square error) that will be used in training was modified as explained in equations (47) - (49). For convenience these equations can be rewritten as

$$E = \frac{\rho_0}{2Q}\sum_{q=1}^{Q}\left[D_q - a_{k,q}\right]^2 + \frac{\rho_1}{2IQ}\sum_{i=1}^{I}\sum_{q=1}^{Q}\left[G_{i,q} - \frac{\partial a_{k,q}}{\partial a_{i,q}}\right]^2 \tag{135}$$

It can be seen in equation (135) that output data (**D**), as well as, derivative information (**G**) is used for training the network. ρ_0 and ρ_1 are weight parameters that will balance the importance to either output or derivative data for training. Therefore, it can be observed three different cases:

1. By setting $\rho_0 = 1$ and $\rho_1 = 0$, the ANN will learn the input/output relationship from the (**X**,**D**) data. This is the same as using the classical backpropagation algorithm.
2. By setting $\rho_0 = 1$ and $\rho_1 = 1$, the ANN will learn the input/output relationship from the (**X**,**D**) data and also using derivative information (**X**,**G**). This will improve accuracy and robustness of the network.
3. Finally, by setting $\rho_0 = 0$ and $\rho_1 = 1$, the input/output (**X**,**D**) relationship is learned from the derivative data (**X**,**G**). This is a very interesting and important property of the method because it gives the integral solution to of the derivative data (**X**,**G**).

Since there are not direct measurements of the device charges, exploiting the integration properties of the algorithm, neural networks can be trained to predict the charge functions from the Y-parameters of the device as illustrated in Figure 64 and 65. As a result, Figure 66 and 67 show the predicted gate and drain charges by the trained ANNs. The major advantage of this neural network approximation is that it avoids the difficulties and inconsistencies from numerical integration. From eq. (122) expressions for the gate and drain charges are found

$$Q_g(V_{gs}, V_{ds}) = \int_{V_{gso}}^{Vgs} \mathrm{Im}[Y_{11}(v_{gs}, V_{dso})]dv_{gs} + \int_{V_{dso}}^{Vds} \mathrm{Im}[Y_{12}(V_{gs}, v_{ds})]dv_{ds} \qquad (136)$$

$$Q_d(V_{gs}, V_{ds}) = \int_{V_{gso}}^{Vgs} \mathrm{Im}[Y_{21}(v_{gs}, V_{dso})]dv_{gs} + \int_{V_{dso}}^{Vds} \mathrm{Im}[Y_{22}(V_{gs}, v_{ds})]dv_{ds} \qquad (137)$$

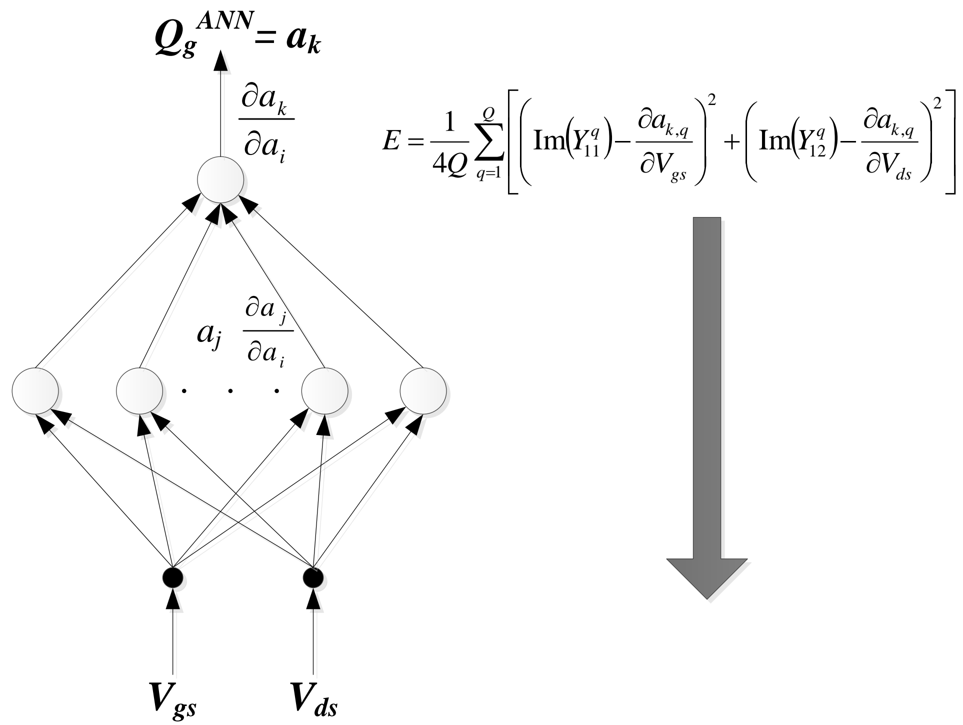


Figure 64. ANN structure and cost function used to train the gate charge.

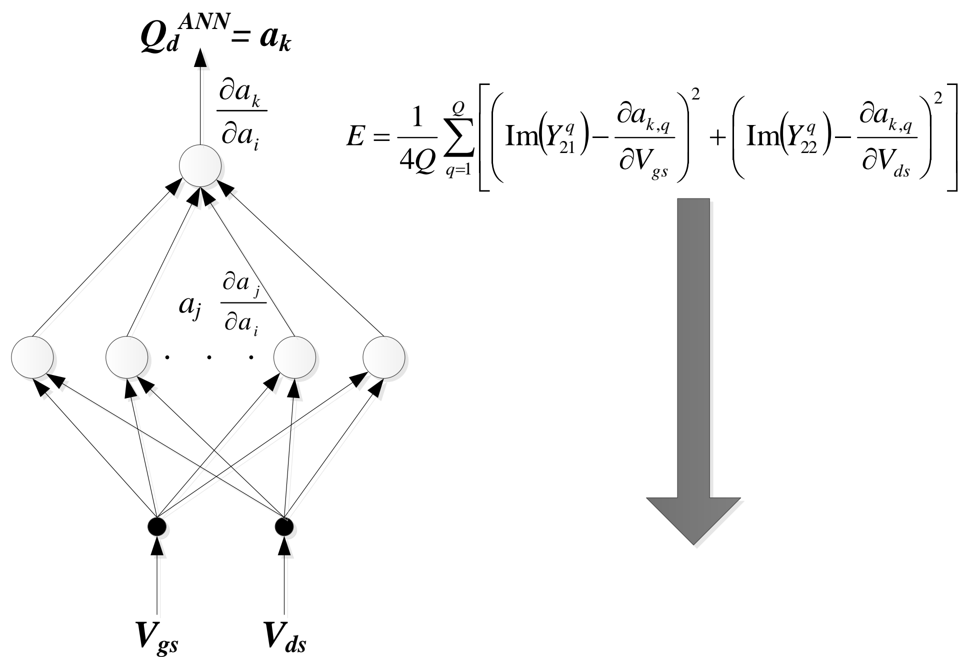


Figure 65. ANN structure and cost function used to train the drain charge.

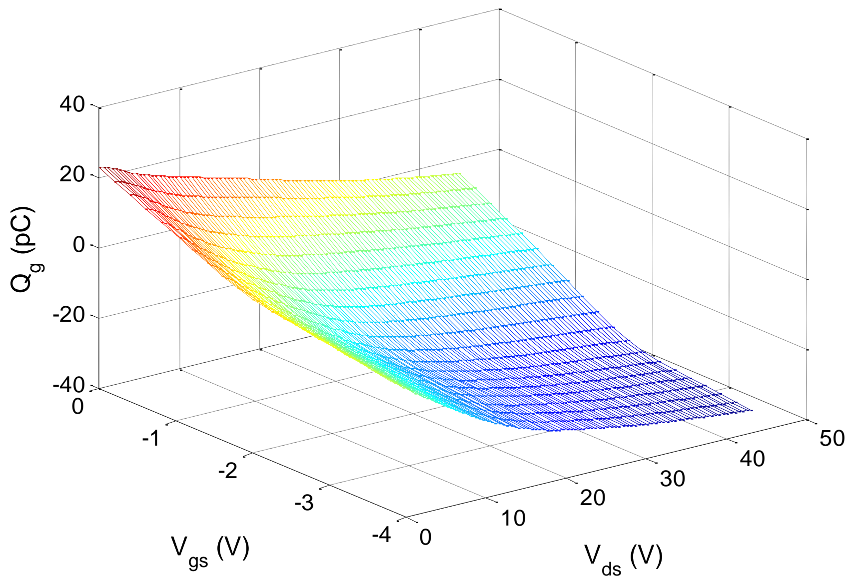


Figure 66. Gate charge function obtained from ANN.

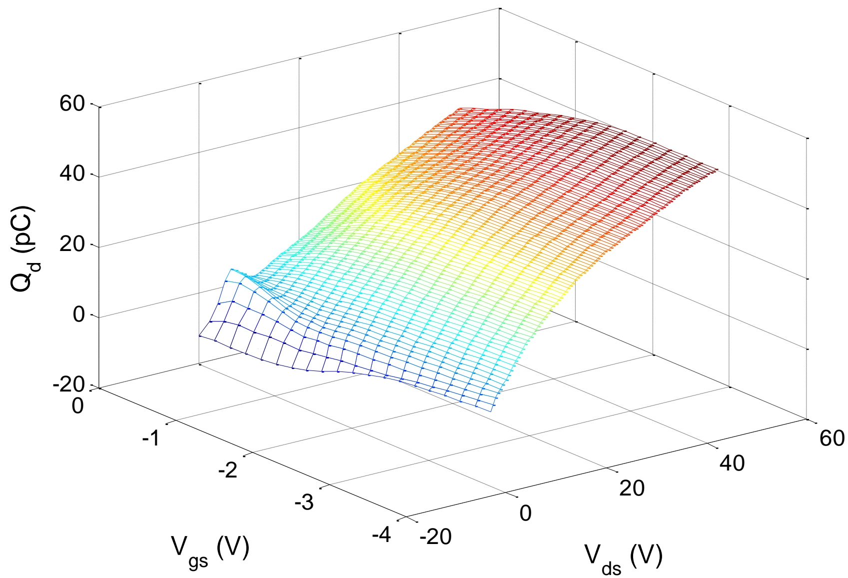


Figure 67. Drain charge function obtained from ANN.

Once the ANN functions for the gate and drain charges are obtained, equation (134) is applied in or to prove charge conservation. Figures 67 and 68 depicts how charge is conserving for the gate and drain terminals of the ANN model.

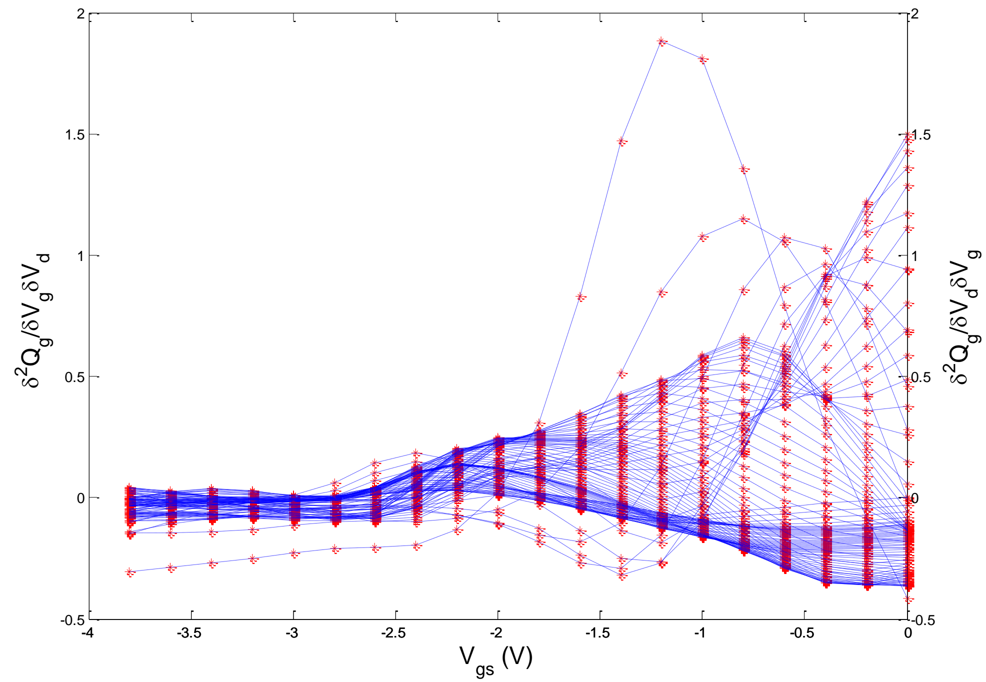


Figure 68. Conservation of charge at the gate.

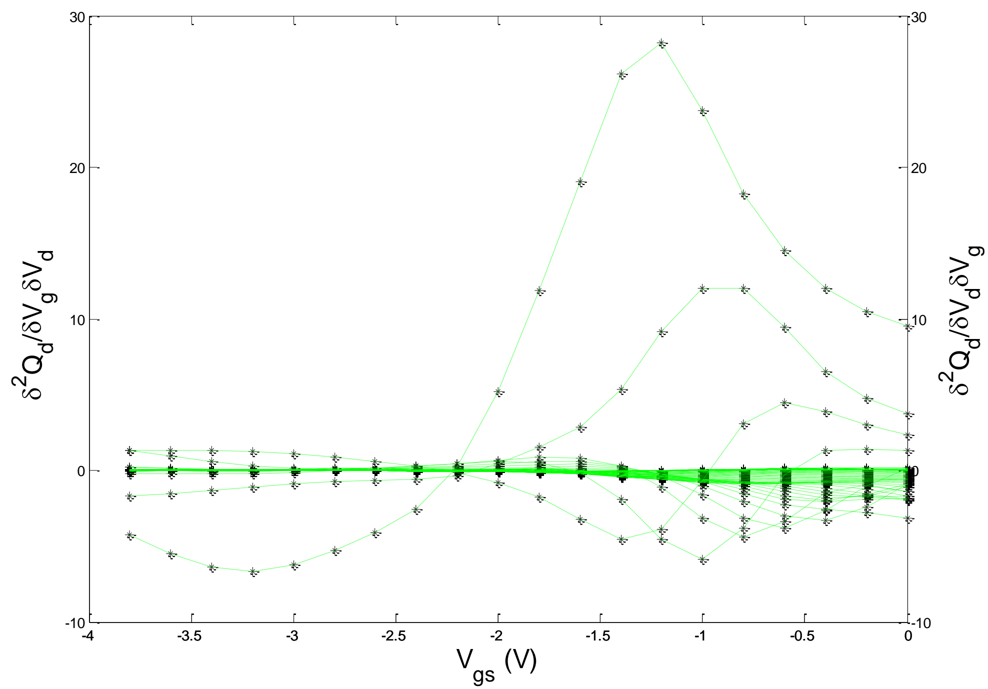


Figure 69. Conservation of charge at the drain.

Chapter 7 - Model Validation

7.1 Introduction

Model validation is the process of determining if the model is an accurate description of the real device from the perspective of the intended use of the model. Model validation has different purposes depending on the user perspective. For example, for the modeling engineer the main purpose of model validation is to guide the development and refinement of a model, while for the circuit designer, its validation provides a confidence level for the accuracy and limitations of the model. A solid and comprehensive model validation exercise results in increased confidence in assumptions behind the construction of the model and a higher level of assurance of its predictive capabilities outside the validation domain.

There is a subtle but important difference between the validation and verification of a model. The verification of the model is the process by which the implementation of the model in the CAD package is demonstrated to be consistent with the equations and topology of the model, and to ensure that the model produces the expected results. In other words, the verification ensures that the model was properly implemented in the circuit simulator. On the other hand, model validation is the process by which the model simulation results are compared with an independent set of data not used during the model extraction. In essence, the model validation provides confidence and guidance on the predictive qualities of the model. An example of model validation is the comparison of model versus measured load-pull contours for a model in which only DC–IV characteristics and small-signal S-parameters were used during the model extraction.

7.2 Measurement Techniques for Model Extraction

As mentioned in Chapter III, in this thesis we are focused in developing a neural network measurement based compact model of the transistor. In this section the techniques

used to collect the necessary data for the extraction of the transistor model will be explained. Basically the device must be characterized in three different forms for model extraction: First, the S-parameters of the device measured in Cold-FET configuration, in both, forward and pinch-off conditions must be measured, this data is used to extract the parasitic elements of the transistor; pulsed I-V measurements are used to train the drain current source neural network model; pulsed S-parameter measurements under different bias conditions will provide the data needed to train the gate and drain neural network charge models. The modeling flow followed in this thesis is illustrated in Figure 70.

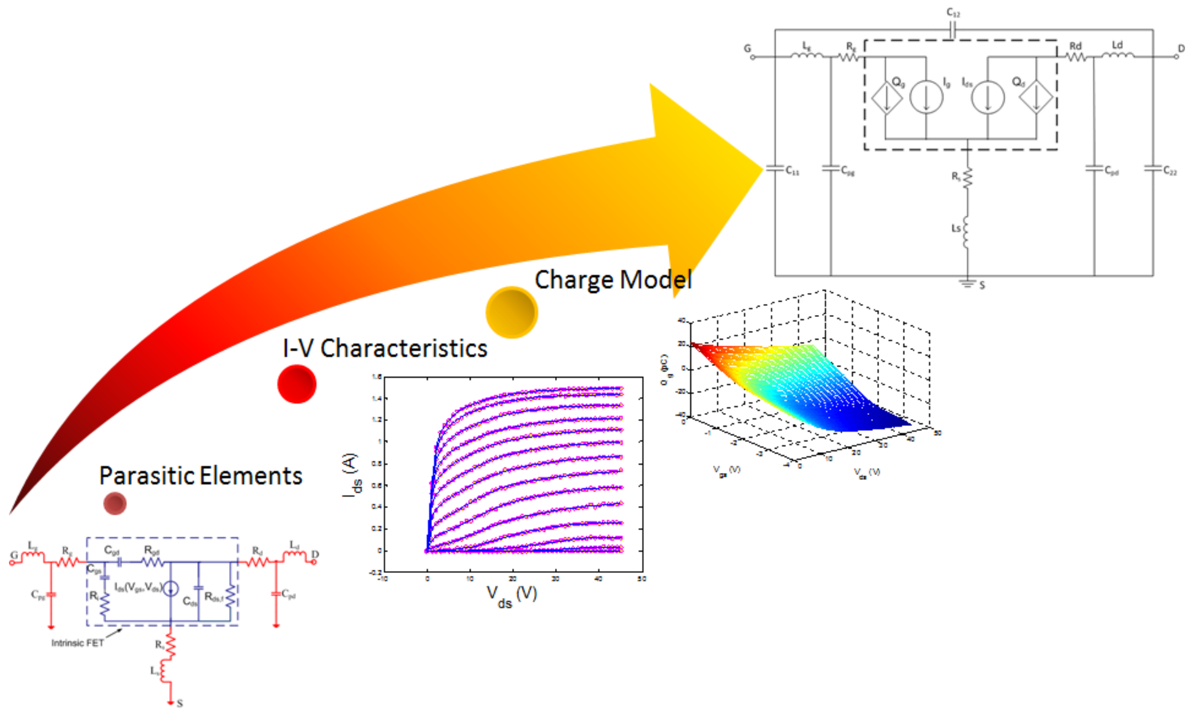


Figure 70. Transistor modeling flow.

7.2.1 Pulsed I-V, Pulsed S-parameter Measurement System

The Auriga AU4750 illustrated in Figure 71 is an example of a commercially available pulsed I–V test system. Pulsed I–V analysis has been applied by many to understand trapping effects in III-V devices, including the original GaN trapping origins

and its characterization (Trassaert et al, 1999), (Siriex et al, 2000), (Meneghesso et al, 2004). For the case of GaN HEMT modeling, pulsed I–V measurements are effective for extracting an isodynamic equation for Ids. The term isodynamic is used when thermal and trap conditions are held at values corresponding to the specified quiescent bias condition (Dunleavy et al, 2010). This is in comparison to a static I–V measurement, representative of a traditional curve tracer or DC parameter analyzer, where acquisition of each data point is slow enough that traps, if present, and thermal conditions can be different at each of the acquired data points. Figure 72 shows a comparison of pulsed and static I–V measurements for a GaN transistor, where the thermal heating effects in the high Vds* Ids power dissipation region are quite clear in contrast to the isodynamic pulsed I–V data. Typical pulsed I–V conditions might consist of 0.1 to 0.5 ms pulse widths, separated by time intervals on the order of 1 ms. These data were acquired using an Accent Optoelectronics DiVA instrument.

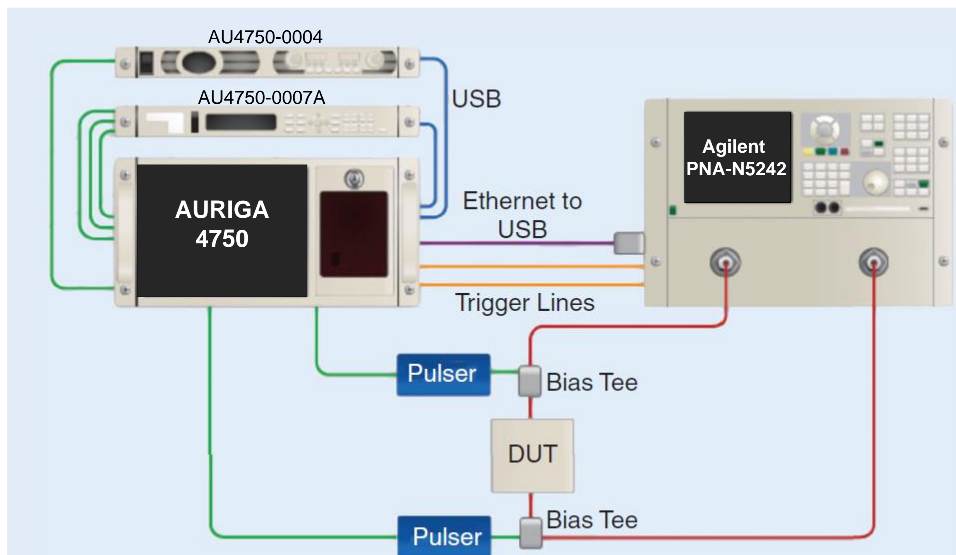


Figure 71. Measurement setup for pulsed I-V, pulsed S-parameters. (Extracted from Auriga Microwave data sheet).

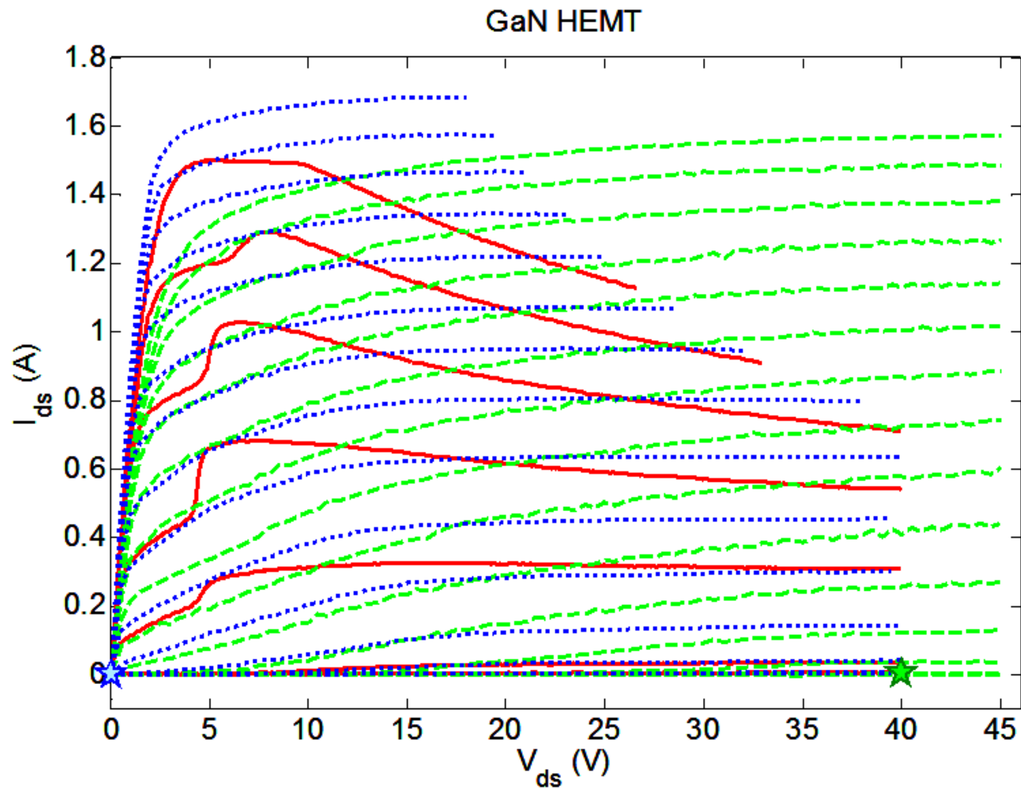


Figure 72. Comparison of static I-V (solid red line) and pulsed I-V for a GaN HEMT. Pulse conditions were 1 µs pulse width and 1 ms separation with quiescent bias points V_{dsq} = 0 V, V_{gsq} = -3 V (blue pointed line) and V_{dsq} = 40 V, V_{gsq} = -3 V (dashed green line).

For this thesis a 15W GaN HEMT (CGH35015F) developed by CREE® for WiMAX applications with V_{th} = -3 V was studied. Measurements were performed using an Auriga AU4750 Pulsed IV/RF system along with pulse I/V head, AU4750-0004 and AU4750-0007A. The pulse width was configured to 10 µs with 1% duty cycle and quiescent point V_{gsq} = -3 V and V_{dsq} = 40 V since the model will be used in the future in the design of a Class E amplifier. V_{gs} was measured in the range of -3.8 V to 0 V with 0.2 V step and V_{ds} from 0 V to 45 V with steps of 0.5 V. At each bias point, pulsed S-parameters of the device were measured using Agilent's PNA-N5242. It is important to comment that pulsed I-V, pulsed S-parameter measurements were performed in order to mitigate the self-heating and trapping effects of the device since the quasi-static model constitutive relationships are considered to be dependent only of the control voltages V_{gs} and V_{ds}. Also,

by performing pulsed measurements, high bias points which could not be reached by static DC conditions are measured. In total 1820 bias points where measured which were used as data to train the neural network models described in Chapter VI. The sample points were divided 80% for training and 20% for validating the network. The logistic function was selected as the nonlinear processing function at the hidden layers. Also, the iRprop was the training method selected as the search direction applied in the modified backpropagation algorithm

7.3 Implementation in Agilent ADS®

The main objective of developing accurate transistor models is being able to implement it in CAD simulators, such as ADS, by circuit designers. Transistor models are of great aid in the design of RF/microwave circuits such as amplifiers, oscillators and mixers. Therefore, a model is successful if it is used in the design of successful products.

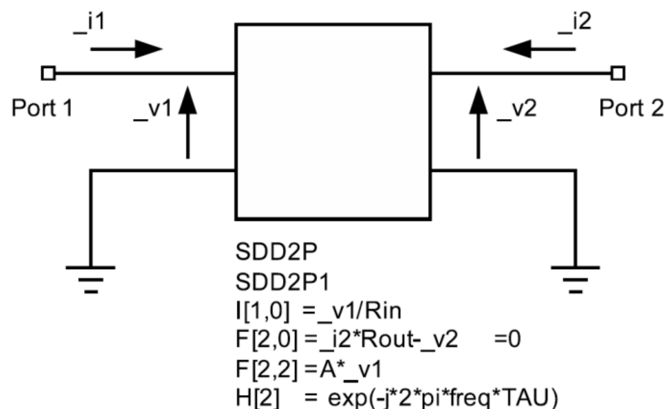


Figure 73. Example of a two-port symbolically-defined device component.

The symbolically-defined device (SDD) is an equation-based component that is well suited to the rapid prototyping of nonlinear device models. It is a component available in Agilent ADS. The SDD is a multi-port component that can be placed directly into the circuit schematic in the simulator. A schematic symbol for an SDD is shown in Figure 73. The current at a given port is defined using an equation, or constitutive relationship, that is

expressed in the time domain. These equations can be functions of the port voltages or currents, their time derivatives or delayed versions of the port voltages or currents, or of external signals that can be referenced. They can be explicit or implicit expressions. For example, in an n-port SDD, the explicit representation of the current at the kth port is given by:

$$i_k = f(v_1, v_2, \dots, v_n) \quad (138)$$

And the current defined by an implicit expression at the kth port is

$$f_k(v_1, v_2, \dots, v_n, i_1, i_2, \dots, i_n) = 0 \quad (139)$$

The explicit equation is a voltage-controlled expression, which is how model equations are often written. The explicit form is used in standard nodal analysis, and is very efficient to execute in the simulator, as the expression can be solved directly, and no new variables are created. In contrast, the implicit representation requires modified nodal analysis for its solution, adding an extra branch variable, i_k, and current equation to the set of algebraic equations that needs to be solved.

The weighting functions H[m] are used to define time derivatives or time delays. The weighting function H[1] is the first time derivative, and it is a built-in function. A time delay can be constructed by defining the following weighting function:

$$H[2] = e^{-j\omega\tau} \quad (140)$$

where ω is the frequency and τ is the delay.

The equations used in the SDD can be the function approximations for the model functions described in earlier chapters, using the device gate and drain voltages as the controlling signals. The expressions should be constructed with the usual considerations for simulator convergence in mind: the functions must be continuous in the voltages and currents, and should be differentiable with respect to the voltage and current. Additionally, the derivatives should be continuous, to aid convergence. An advantage of using the SDD as a vehicle for nonlinear model implementation is that the Jacobian is calculated within the component itself, during simulation. This avoids the lengthy time involved and

potential for errors that arise when the modeling engineer has to create the Jacobian for the model by hand. The development and prototyping of the model is therefore significantly faster when using the SDD; often, this is also the final form of the model for distribution. The numerical efficiency of SDD devices is just slightly worse than similar models implemented in a high-level computer programming language.

One of the advantages of neural networks is that once the ANN is trained it is represented by an equation which is easily implemented in a circuit simulator such as ADS. In this work three neural networks that describe the constitutive relations of an AlGaN/GaN HEMT (CGH35015F) where developed and trained as described in the last chapter. After training, the neural networks that model de drain current source I_{ds}, and the gate and drain charge function Q_g and Q_d where implemented, in equation form, on a SDD nonlinear component as illustrated in Figure 74. The package and parasitic elements were determined following the procedures described in Chapter V. The intrinsic resistance R_i and time delay τ are defined as bias-dependent neural networks trained with information obtained from the calculated intrinsic elements.

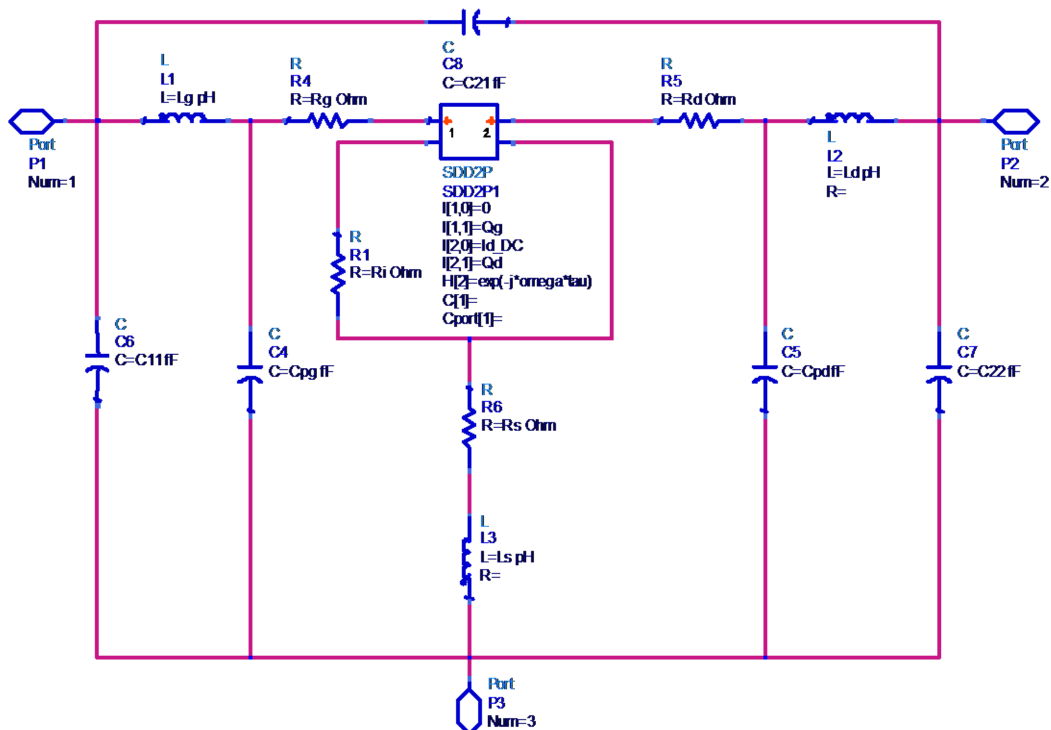


Figure 74. Model implementation in Agilent's ADS.

7.4 Model Simulation and Results

The developed quasi-static model was verified by small and large signal measurements. First, the model is checked whether it is consistent with the pulsed I-V and pulsed S-parameter measurements. Finally large signal single tone simulations are compared to measurements.

7.4.1 I-V Characteristics

The first step in model validation is to compare the simulated to measured I-V characteristics of the device as depicted in Figure 75. The good agreement between measured and simulated data is proof of the excellent function approximation properties of neural networks. It also demonstrates that the intrinsic voltages where accurately calculated in the de-embedding process.

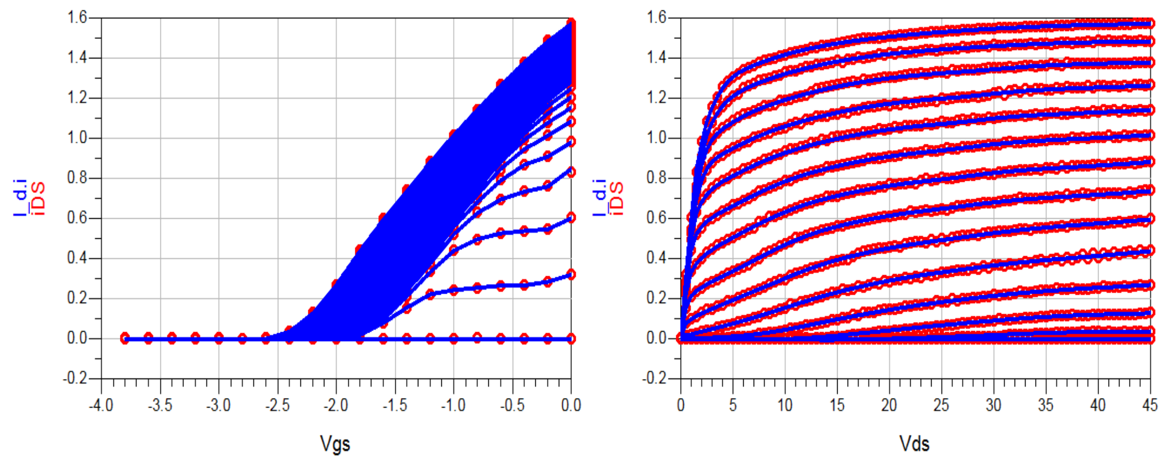
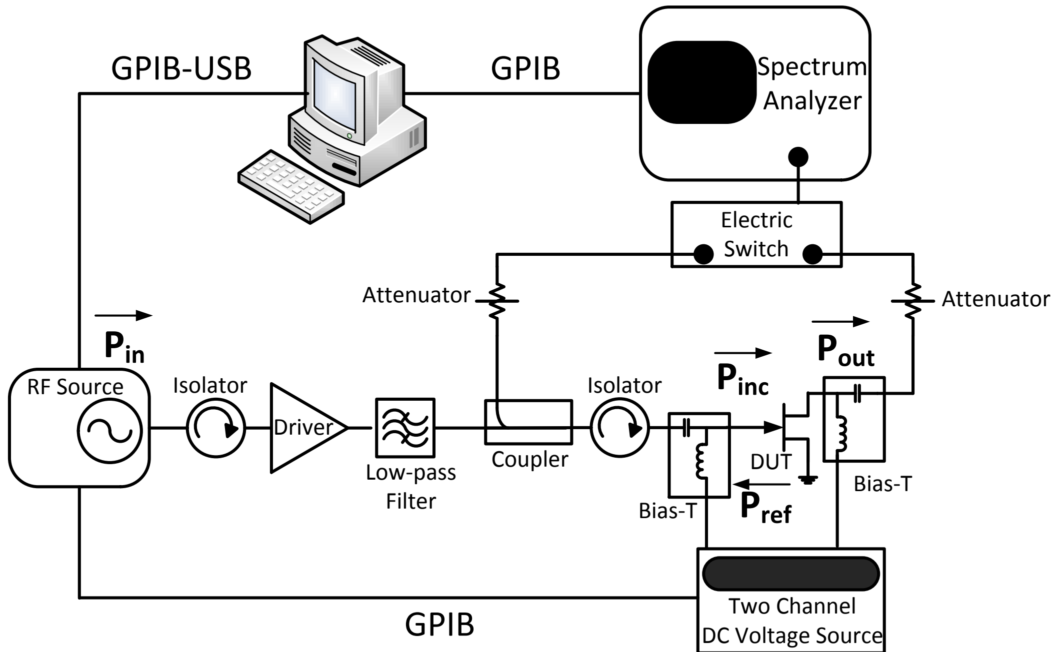


Figure 75. I-V Comparison between measurements and simulations for the CGH35015 AlGaN/GaN quasi-static model.

7.4.2 RF Characteristics

The large-signal model was validated by comparing harmonic balance simulation performed on Agilent ADS™ to measurements. The AM-AM measurement setup used in this work is illustrated in Figure. 76. A good agreement of the Pout vs Pin relationship for

the fundamental, 2nd and 3rd harmonics, as well as the S-parameters of the device is obtained and displayed in Figures 77, 78, 79 and 80. Results were obtained at 3.5 GHz and four different bias conditions which are low due to the absence of long term memory effects in the model.



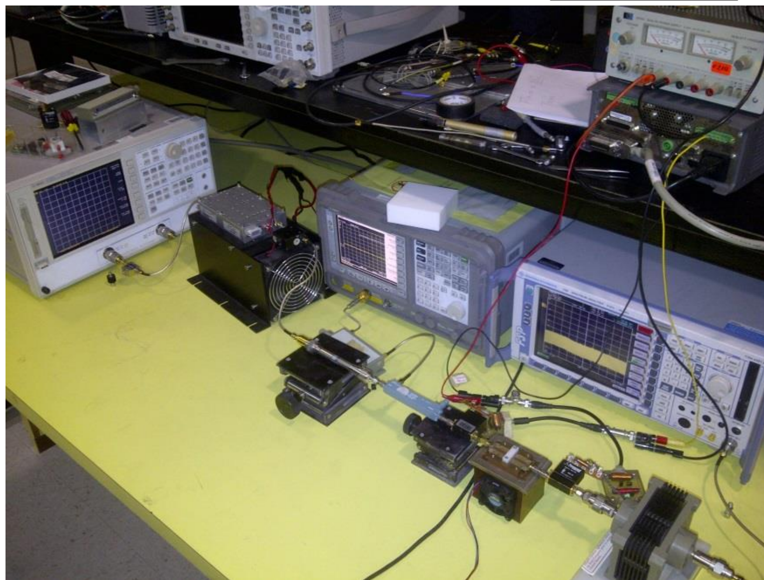


Figure 76. AM-AM measurement setup.

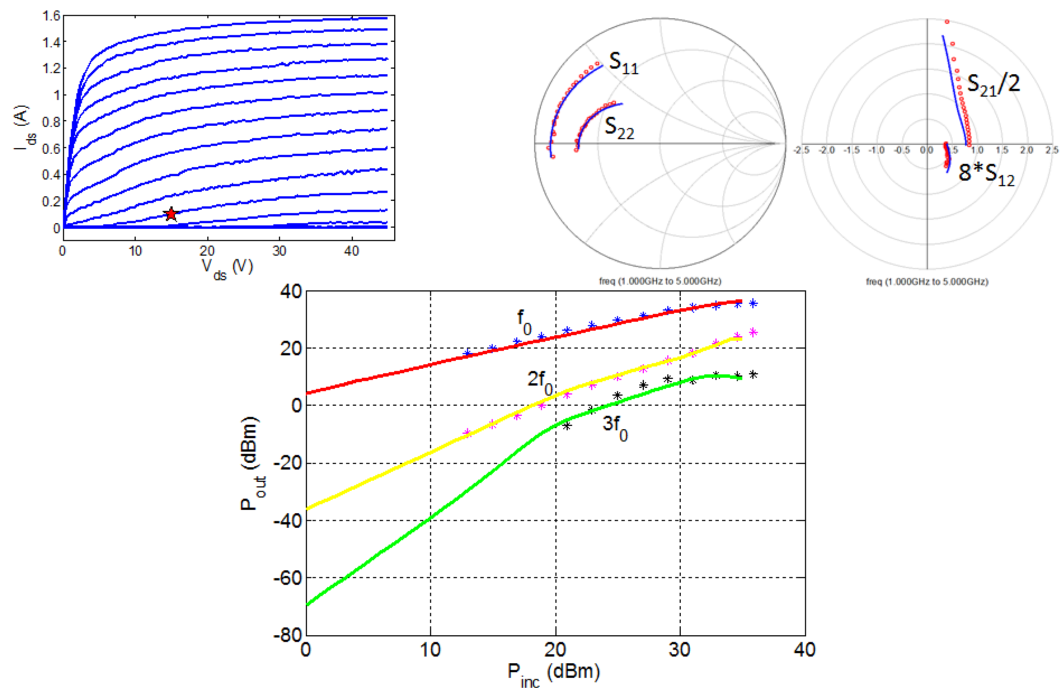


Figure 77. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -2 V and V_{ds} = 15 V.

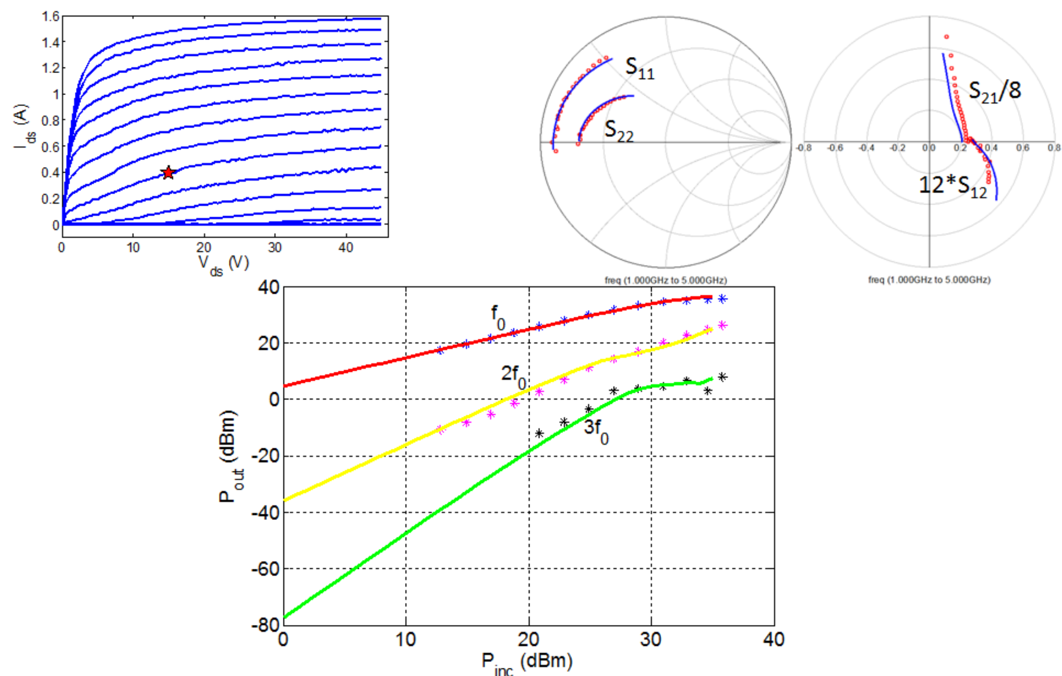


Figure 78. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -1.6 V and V_{ds} = 15 V.

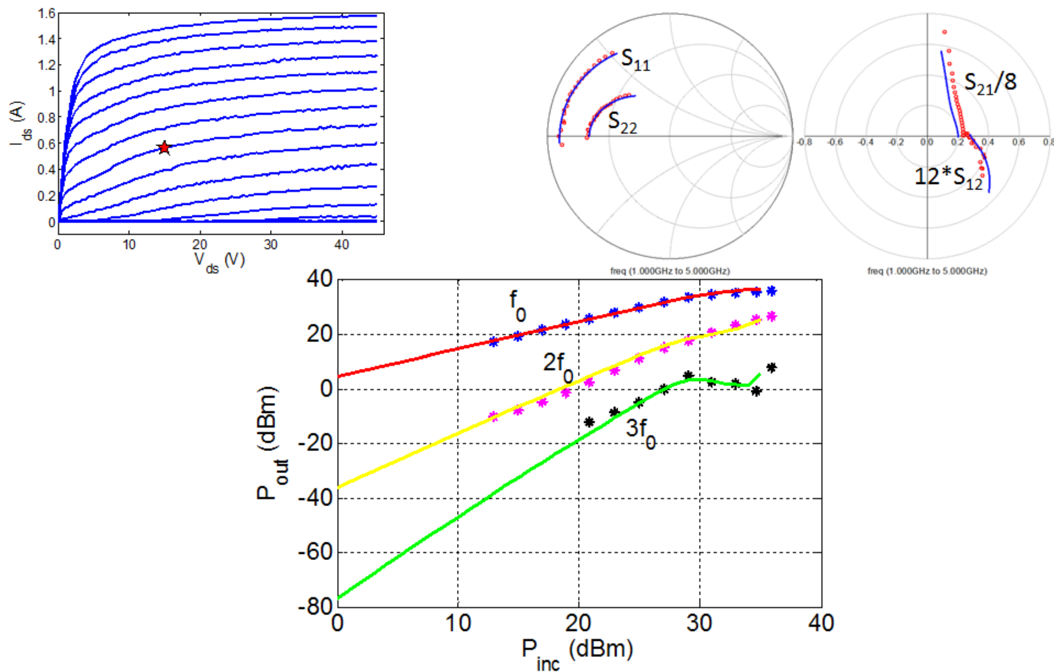


Figure 79. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -1.4 V and V_{ds} = 15 V.

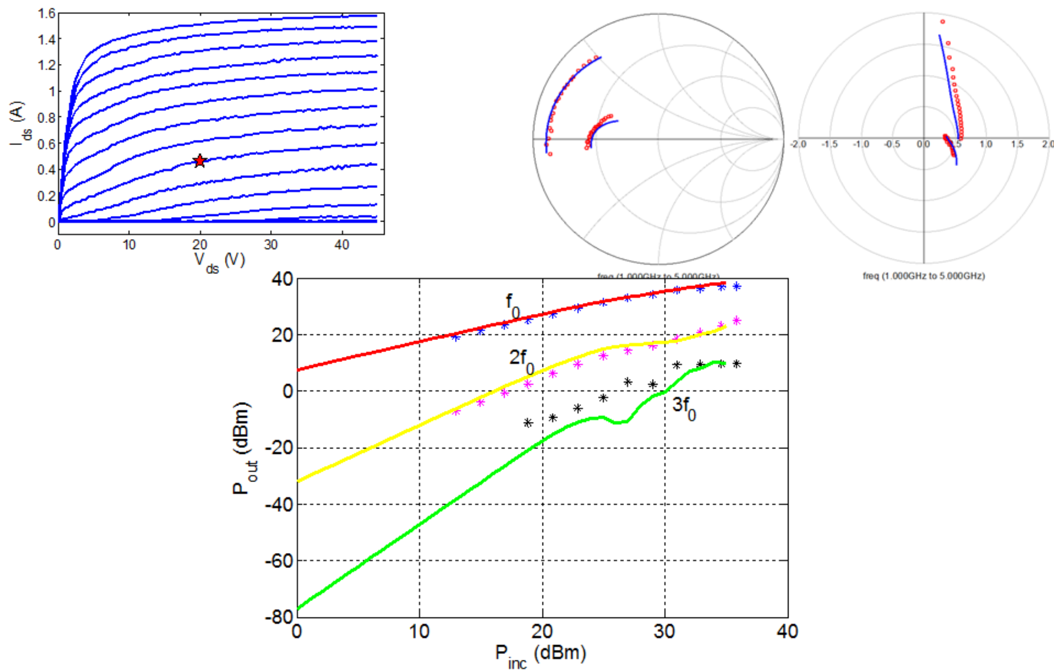


Figure 80. Comparison between measured and simulated S-parameters and output power of the transistor at f_0 = 3.5 GHz, V_{gs} = -1.6 V and V_{ds} = 20 V.

Conclusions and Future Work

In this thesis a neural network based quasi-static model of a commercial AlGaN/GaN HEMT, which accurately reproduces the current and charge storage characteristics of the device was developed. This measurement based compact model was derived from pulsed I-V, pulsed S-parameter measurements since no long term memory effects were taken into account. The model is constructed to account for the main physical and electrical characteristics of the power device. A detailed discussion of efficient and reliable techniques for parasitic elements extraction was presented. Package parasitics were extracted from the measured admittance parameters of a dummy structure where there was not active device. Then from the measurement of the transistor in forward cold-FET configuration the parasitic resistances and inductances were determined. The main two advantages of the method presented in this thesis is that only one measurement is needed, and also, since the proposed forward cold-FET measurement uses a small forward current applied to the gate Schottky barrier, the device is never at risk of suffering irreversible damage. Finally the parasitic capacitances, the last of the extrinsic elements, are calculated from the pinched-off cold-FET measurement. It is important to mention that a good parasitic extraction is the foundation of a good model, there is no way to stress this enough, since these elements are used in de-embedding processes to find the intrinsic voltages and intrinsic admittance parameters of the device as well.

In Chapter IV it was introduced a modification to the classical backpropagation algorithm for artificial neural networks which uses derivative information in the training process, this is the main contribution of this research work. The modified backpropagation permits a straightforward methodology to develop a neural network based model directly after device characterization. By using derivative information in the training process it is possible to develop a robust neural network based drain current model since information of the output conductance and transconductance can be used in the training process. Also, this technique enables the direct computation of smooth ANN-based constitutive relations of

the transistor terminal charge functions from bias-dependent S-parameter data. After de-embedding the parasitic elements, the bias-dependent intrinsic admittance parameters of the device are calculated and used as training data samples for the ANN-based terminal charge functions. By taking advantage of the integration properties of the algorithm, the gate and drain charge ANN models can be directly developed from the imaginary part of the intrinsic admittance parameters of the device, improving the numerical path-integration technique used in analytical or table-based models. Another advantage of neural networks is that they can be trained from training scattered data contrary to table based models which need data that fall in a grid to tabulate the model parameters. In addition to these advantages, neural network models are represented as a set of simple equations that can be easily implemented in circuit simulators, which is the objective of compact models, but obviating difficult parameter extraction procedures found in analytical, table-based and physics-based models. All the parameters needed in the model are obtained during the training process which just requires the input/output relationship extracted from measured data. Once the ANNs are trained, the model exhibits excellent accuracy and simulation speed properties when it is implemented in a commercial circuit simulator such as Agilent's ADS®.

Finally the quasi-static model was validated by comparing simulations to small and large signal measurements. The model is able to accurately predict the S-parameters of the device and the output power for the fundamental, second and third harmonics at different bias conditions, verifying then the small-large signal consistency of the model.

Artificial neural networks are starting to play a crucial role for nonlinear device modeling and the modified backpropagation added a significant value to this field. ANN-based nonlinear transistor models have demonstrated superior capabilities compared to table-based models in most aspects. From the same set of data from which the table-based model is constructed, the ANN model is more uniformly accurate and smoother. Besides, due to their nature, ANN-based models can be used in different technologies such as FETs, LDMOS and HBTs as long as measurements are available since this approach has proven versatility to be implemented in any equivalent circuit topology. In addition, the information used from training can be from any source: I-V and C-V curve tracer along

with vector network analyzers (VNA) measurements; pulsed I-V and pulsed S-parameters; passive or active load pull systems; nonlinear vector network analyzers (NVNA), large signal network analyzers (LSNA) or X-parameter measurements, etc. There is a huge universe of possibilities to continue this research, it is recommended that future work focuses in the inclusion of long term memory effects such charge trapping and a thermal effects to improve the reliability and robustness of the model. The neural network methodology presented in this thesis can be extended or adapted to use data measured from nonlinear network analyzers. Therefore it would be interesting to compare the model presented in this work with an electrothermal and trap-dependent model constructed directly from nonlinear measurements and using neural networks as a fitting tool. Another research branch can be found in the use of the model to design a power amplifier.

Final thoughts regarding this topic fall in the idea that distinct measurement approaches can be combined with ANN models in order to create end-to-end methodologies that suit a variety of needs. A nonlinear model can be generated from linear measurements as demonstrated in this work, but also, the recent commercial availability of NVNA/LSNA instruments enables many new and powerful transistor modeling flows. NVNA/LSNA or X-parameter waveform data combined with advanced ANN modeling technology can produce a large-signal electrothermal and trap-dependent nonlinear time-domain model for III-V FETs, HBTs or LDMOS devices with advanced memory effects, high accuracy and considerable generality.

Contributions

- A thorough methodology to obtain the compact model of microwave transistors using artificial neural networks is explained in this thesis.
- The model surpasses the accuracy of table and equation based models.
- The model is fast and easy to extract.
- The model of a 15W GaN HEMT was developed. It can be used in the future for the design of power amplifiers.
- A neural network toolbox specifically developed to be used as a tool in the modeling of transistors was created. The toolbox will be available for the Microwave/RF Group at CICESE and the Nonlinear RF Laboratory at The Ohio State University.

Journal Publication and Proceeding Articles

Andrés Zárate-de Landa, Patrick Roblin, J.A. Reynoso-Hernandez and J.R. Loo-Yau, "Modeling the I-V Curves and its Derivatives of Microwave Transistors Using Neural Networks," IEEE Microwave and Wireless Components Letters, Vol, 22, Issue 9, pp 468-470, Sept, 2012.

Andrés Zárate-de Landa, J.A. Reynoso-Hernández, Patrick Roblin, M.A. Pulido-Gaytán, J. R. Monjardín López, J.R. Loo-Yau, "On the Determination of Neural Network Based Nonlinear Constitutive Relations for Quasi-Static GaN FET Models," 82th ARFTG Microwave Measurement Conference (ARFTG), Columbus Ohio, Nov 20th 2013.

Andrés Zárate-de Landa, M.A. Pulido-Gaytán, J.A. Reynoso-Hernández, Patrick Roblin, J.R. Loo-Yau, "A Neural Network Approach to Smooth Calibrated Data From Switching Errors," Microwave Measurement Symposium (ARFTG), 80th ARFTG Microwave Measurement Conference (ARFTG), San Diego California, November 2012.

Bibliographic References

Aaen Peter H., Plá Jaime A. and Wood John, “Modeling and Characterization of RF and Microwave Power FETs,” The Cambridge RF and Microwave Engineering Series, Cambridge University Press, 2007.

Ambacher O., et al., “Two-dimensional electron gases induced by spontaneous and piezoelectric polarization in undoped and doped AlGaN/GaN heterostructures,” J. Appl. Phys., Vol. 85, pp. 3222-3232, March 1999.

Ampazis N. and Perantonis S.J., “Two Highly Efficient Second-Order Algorithms for Training Feedforward Networks,” IEEE Trans. On Neural Networks, Vol. 13, No. 5, pp 1064-1074, September 2002.

Angelov I., Zirath H., and Rorsman N, “A new empirical nonlinear model for HEMT and MESFET devices,” IEEE Trans. Microwave Theory Tech., Vol. 40, No. 12, pp. 2258–2266, December 1992.

Barron Andrew R., “Universal Approximation Bounds for Superposition of a Sigmoidal Function,” IEEE Transactions on Information Theory, Vol. 39, No. 3, pp 930-945. May 1993,

Berroth M., and Bosh R., “Broad-band determination of the FET small-signal equivalent circuit,”IEEE Trans. Microwave Theory and Tech., vol.38.no.7, pp 891-895,july 1990.

Cabral P.M., Pedro J.C. and Carvalho N., “Nonlinear Device Model of Micrwave Power GaN HEMTs for High Power-Amplifier Design,” IEEE Transactions on Microwave Theory and Techniques, Vol. 52, No. 11, pp 2585-2592, November 2004.

Caddemi A., Crupi G. and Donato N., “Microwave Characterization and Modelling of Packaged HEMTs by a Direct Extraction Procedure Down to 30K”, IEEE Trans. On Instrumentation and Measurement. Vol. 55, No. 2, pp.465-470, April 2006.

Cao Y., Chen X., Wang G., “Dynamic Behavioral Modeling of Nonlinear Microwave Devices Using Real-Time Recurrent Neural Network” IEEE Trans. on Electron Devices, Vol 56, No 5, pp 1020-1026, May 2009.

de Carvalho N. B. and Pedro J. C., “Large- and small-signal IMD behavior of microwave power amplifiers,” IEEE Trans. Microwave Theory Tech., Vol. 47, No. 12, pp. 2364–74, December 1999.

Cheung S. K. and Cheung N. W., "Extraction of Schottky diode parameters from forward current-voltage characteristics," Applied Physics Letters., Vol. 49, No. 2, pp. 85–87, July 1986.

Chua L.O., Desoer C. A., and Kuh E. S., Linear and Nonlinear Circuits. New York, NY: McGraw-Hill, 1987.

Colantonio Paolo, "Multiharmonic Manipulation for Highly Efficient Microwave Power Amplifiers", International Journal of RF and Microwave CAE, Vol. 11, No. 6, pp. 366-385, October 2001.

Crupi G., Xiao D., Schreurs D, Limiti E., Caddemi A., Raedt W., and Germain M., "Accurate multibias equivalent-circuit extraction for GaN HEMTs," IEEE Trans. Microwave Theory Tech., Vol. 54, No. 10, pp.3616-3622, October 2006.

Curtice W.R. and Caamisa R., "Self-cosistent GaAsFET models for amplifier design and device diagnostic," IEEE Trans. Microwave Theory Tech., Vol. 32, No. 12, pp. 1573–1578, July 1984.

Curtice W.R. and Ettenberg M., "A nonlinear GaAs FET model for use in the design of output circuits for power amplifiers," IEEE Trans. Microwave Theory Tech., Vol. 33, No. 12, pp. 1383-1394, December 1985.

Curtice W.R., Plá J. A., Bridges D., Liang T., and Shumate E. E., "A new dynamic electro-thermal nonlinear model for silicon RF LDMOS FETs," in IEEE MTT-S Int. Microwave Symp. Dig., pp. 419–422, Anaheim, CA, June 1999.

Dambrine G., Cappy A, Heliodore F., and Playez E., "A new method for determining the FET small-signal equivalent circuit," IEEE Trans. Microwave Theory Tech., Vol. 36, no. 7, pp, 1151-1159, July 1988.

Delagebeaudeuf D. and Linh N.T., "Metal-(n) AIGaAs-GaAs twodimensional electron gas FET," IEEE Trans. Electron Devices, Vol. 29, pp. 955-960, June 1982.

Diamond and Laviron M., "Measurement of the extrensic series elements of a microwave MESFET under zero current conditions," in Proc. 12th European Microwave Conf., (Finland), pp. 451-456, September 1982.

Dunleavy Lawrence, Baylis Charles, Curtice Walter and Connick Rick, "Modeling GaN: Powerful But Challenging," IEEE Microwave Magazine, Vol. 11, No. 6, pp 82-96, October 2010.

Eastman Lester F. y Mishra Umesh K., "The Toughest Transistor", IEEE Spectrum Magazine. Vol. 39, No. 5, pp. 28-33, May 2002

Gao J., Zhang L., Xu Q.J. and Zhang Q.J., “Nonlinear HEMT Modeling Using Artificial Neural Network Technique,” IEEE MTT-S International Microwave Symposium Digest, June 2005.

Green B. M., Tilak V., Kaper V. S., Smart J. A., Shealy J. R., and Eastman L. F., “Microwave power limits of AlGaN/GaN HEMTs under pulsed bias conditions,” IEEE Trans. Microwave Theory Tech., 52, no. 2, pp. 618–623, February 2003.

Hagan M.T. and Menhaj M., “Training feedforward networks with the Marquardt algorithm,” IEEE Trans. Neural Networks, Vol. 5, pp. 989–993, November 1994.

Hansen P.J., Strausser Y.E., Erickson A.N., Tarsa E.J., Kozodoy P., Brazel E.G., Ibbetson P., Mishra U., Narayanamurti V., DenBaars S.P., and Speck J.S., “Scanninf capacitance microscopy imaging of threading dislocations in GaN films grown on (0001) sapphire by metalorganic chemical vapor deposition,” Appl. Phys. Letters,. Vol. 72, pp. 2247-2249, May 1998.

Haykin Simon, 1999, “Neural Networks: A Comprehensive Foundation,” Pearson Prentice Hall International Edition Series. July 1998.

Ibbetson J.P., Fini P.T., Ness K.D., DenBars S.P., Speck J.S. and Mishra U.K., “Polarization effects, surface states and the source of electrons in AlGaN/GaN heterostructure field effect transistors,” Appl. Phys. Letters, Vol. 77, pp 250-252, July 2000.

Igel Christian and Husken Michael, “Improving the Rprop Learning Algorithm”. In Proceedings of the Second International Symposium on Neural Computation, NC’2000, pages 115–121. Academic Press, 2000.

Iwamoto M, Root D. E., Scott J. B., Cognata A., Asbeck P. M., Hughes B., and D’Avanzo D. C., “Large-signal HBT model with improved collector transit time formulation for GaAs and InP technologies,” in IEEE MTT-S Int. Microwave Symp. Dig., Philadelphia, PA, pp. 635–638, June 2003.

Jarndal A.H., “Large-Signal Modeling of GaN Device for High Power Amplifier Design,” PhD. Dissertation, University of Kassel, November 2006.

Jarndal A., and Kompa G., “A new small signal model parameter extraction method applied to GaN devices,” IEEE MTT-S Int. Microwave Symposium, Long Beach,CA, June 2005.

Kompa G., “Modeling of Dispersive Microwave FET Devices Using a Quasi-Static Approach,” International Journal of Microwave and Milimiter-Wave Computer Aided Engineering, Vol. 5 No 3, pp 173-194 December 1994.

Liu W., MOSFET Models for SPICE Simulation, including BSIM3v3 and BSIM4. New York, NY: Wiley-IEEE Press, February 2001.

Loo-Yau J.R., "Desarrollo de modelos no lineales de transistores GaAs para el diseño de amplificadores de potencia de alta eficiencia" CICESE, División de Física Aplicada, Departamento de Electrónica y Telecomunicaciones. PhD Dissertation, July 2006.

Materka A. and Kacprzak T., "Computer calculation of large-signal GaAs FET amplifier characteristics," IEEE Trans. Microwave Theory Tech., Vol. 33, No. 2, pp. 129-134, February 1985.

McAndrew C. C., Seitchik J. A., Bowers D. F., et al., "VBIC95, the vertical bipolar inter-company model," IEEE J. Solid State Circuits, Vol. 31, No. 10, pp. 1476–83, October 1996.

Meneghesso G., Verzellesi G., Pierobon R., Rampazzo F., Chini A., Mishra U., Canali C., and Zanoni E., "Surface-related drain current dispersion effects in AlGaN-GaN HEMTs," IEEE Trans. Electron Devices, Vol. 51, No. 10, pp. 1554–1561, October 2004.

Monjardín López Jonathan Raúl, "Comparación de modelos lineales en modo pulsado y no pulsado de transistores GaN de potencia. MSc. Dissertation, February, 2014.

Närhi T, "Frequency Domain Analysis of Strongly Nonlinear Circuits Using a Consistent Large Signal Model," IEEE Transactions on Microwave Theory and techniques, Vol. 44, No 2, pp 182-192, February 1996.

Parker A. E. and Skellern D. J., "A realistic large-signal MESFET model for SPICE," IEEE Trans. Microwave Theory Tech., Vol. 45, No. 9, pp. 1563–71, September 1997.

Principe J.C., Euliano N.R., Lefebvre W.C., "Neural and Adaptive Systems Fundamentals Through Simulations", Wiley, December 2001.

Pucel R. A., Haus H. A., and Statz H, "Signal and noise properties of Gallium Arsenide microwave field effect transistors," In Advances in Electronics and Electron Physics. New York, NY: Academic Press, Vol. 38, pp. 195–265, 1975,

Raab F.H., "Class-F Power Amplifiers with Maximally Flat Waveforms", IEEE Transactions on Microwave Theory and Techniques, Vol. 45, pp. 2007-2011, No. 11, November 1997.

Raab F.H., "Maximum Efficiency and Output of Class-F Power Amplifiers", IEEE Transactions on Microwave Theory and Techniques, Vol. 49, pp. 1162-1165, No. 6, June 2001.

Reynoso-Hernández J.A., Loo-Yau J.R., Zúñiga-Juárez J.E. and del Valle-Padilla J.L., "A Straightforward Method to Determine the Parasitic Gate Resistance of GaN FET," IEEE Microwave Symposium Digest, pp. 877-880 Boston, MA, June 2009.

Reynoso-Hernández J.A., Rangel-Patiño F.E., and Perdomo J., "Full RF Characterization for extracting de small-signal equivalent circuit in microwave FET's," IEEE Trans. Microwave Theory Tech., Vol. 44, No.12, pp. 2625-2633, December 1996.

Reynoso-Hernández J.A., Zúñiga-Juárez J. E., Zárate-de Landa A., "A New method for Determining the Gate Resistance and Inductance of GaN HEMTs Based on the Extrema Points of Z_{11} Curves," presented at the IEEE MTT-s Int. Microwave Syp, Atlanta, Georgia, USA, June 2008.

Roblin P., Kang S.C. and Morkoc H., "Analytic Solution of the Velocity-Saturated MOSFET/MODFET Wave Equation and Its Application to the Prediction of the Microwave Characteristics of MODFET's," IEEE Transactions on Electron Devices, Vol. 37 No 7, pp 1608-1622, July 1990.

Roblin Patrick, "Nonlinear RF Circuits and Nonlinear Vector Network Analyzers," Cambridge University Press, July 2011.

Root D.E., Fan S., and Meyer J., "Technology-Independent Large Signal FET Models; A Measurement Based Approach to Active Device Modeling," Proceedings of 15th ARMMS Conf. pp 1-21, September 1991.

Root D.E., "Measurement-based mathematical active device modeling for high frequency circuit simulation," IEICE Trans. Electron. Vol. 82, pp. 924–936, June 1999.

Rudolph M., Fager C. and Root D., "Nonlinear Transistor Model Parameter Extraction Techniques," The Cambridge RF and Microwave Engineering Series, Cambridge University Press, December 2011.

Samarasinghe Sandhya, "Neural Networks for Applied Sciences and Engineering From Fundamentals to Complex Pattern Recognition," Aurebach Publications. September 2006.

Schreurs D., "Table-Based Large Signal Models Based on Large Signal Measurements," IEEE MTT-S Workshop: New Directions in Nonlinear RF and Microwave Characterization, USA, paper 7, June 1996.

Shirakawa K., Shimizu M., Okubo N. and Daido Y., " Structural Determination of Multilayered Large Signal Neural Network HEMT Model", IEEE Trans. Microwave Theory Techniques, Vol. 46, No 10, pp. 1367 – 1375, October 1998.

Shockley W., "A unipolar field-effect transistor," Proc. IRE, 40, pp. 1365–76, November 1952.

Siriex D., Barataud D., Sommet P., Noblanc O., Quarch Z., Brylinski C., Teyssier J., and Quere R., "Characterization and modeling of nonlinear trapping effects in power SiC MESFETs," in IEEE MTT-S Int. Microwave Symp. Dig., pp. 765–768, June 2000.

Snider D. M., "A Theoretical Analysis and Experimental Confirmation of the Optimally Loaded and Overdriven RF Power Amplifier", IEEE Transactions on Electron Devices, Vol. 14, pp. 851-857, June 1967

Statz H, Newman P., Smith W., Pucel R. A. and Haus H. A., " GaAs FET Device and Circuit Simulation in SPICE," IEEE Transactions on Electron Devices, Vol. 34, No 2, pp 160- 169, February 1987.

Stengel B. and Eisenstadt W.R., "LINC power amplifier combiner method efficiency optimization," IEEE Transactions on Vehicular Technology, Vol. 49, No. 1, pp. 229-234, January 2000.

Trassaert S., Budart B., Gaquière C., Théron D., Crosnier Y., Huet F., and Poisson M. A., "Trap effects studies in GaN MESFETs by pulsed measurements," Electron. Lett., Vol. 35, No. 16, pp. 1386–1388, August 1999.

Verspecht J., "Everything you've always wanted to know about Hot-S22 (but were afraid to ask," in IEEE MTT-S Int. Microwave Workshop 'Introducing New Concepts in Nonlinear Network Design', Seattle, WA, June 2002.

Vetury R., Zhang N.Q., Keller S., and Mishra U.K., "The impact of surface states on the DC and RF characteristics of AlGaN/GaN HFETs," IEEE Trans. On Electron Devices, Vol. 48, pp. 569-566, March 2001.

Wang DeLiang, "Artificial Neural Networks," Class Notes, Department of Computer Science and Engineering, The Ohio State University, January-April 2011.

White P.M. and Healy R. M., "Improved equivalent circuit for determining of MESFET and HEMT parasitic capacitances from "Cold-FET" measurements" IEEE Microwave and guided wave letters, Vol. 3, No. 12, pp. 453-454, December 1993.

Woo Y.Y., Yi J., Yang Y, and Kim B., "SDR transmitter based on LINC amplifier with bias control," IEEE MTT-S Digest, International Microwave Symposium (IMS), pp. 1703-1706, June 2003.

Wood J., LeFevre M., Runton D., Nanan J. C., Noori B. H., and Aaen P. H., "Envelope-domain time series (ET) behavioral model of a Doherty RF power amplifier for system design," IEEE Trans. Microwave Theory Tech., Vol 54, No. 8, pp. 3163–72, August 2006.

Wood J. and Root D. E., “Bias-dependent linear scalable millimeter-wave FET model,” IEEE Trans. Microwave Theory Tech., Vol. 48, No. 12, pp. 2352–2360, December 2000.

Xu J., Gunyan D, Iwamoto, M., Cognata, A., Root, D.E., “Measurement-Based Non-Quasi-Static Large-Signal FET Model Using Artificial Neural Networks”, IEEE MTT-S Int. Microwave Symposium Digest, June 2006.

Xu J., Yagoub M.C.E., Ding R., and Zhang Q-J, “Exact adjoint sensitivity analysis for neural-based microwave modeling and design,” IEEE Trans. Microwave Theory Techniques, Vol 51, pp. 226 – 237, January 2003.

Ytterdahl T, Cheng Y., and Fjeldly T. A., “Device Modeling for Analog and RF CMOS Circuit Design,” Chichester, UK: Wiley, May 2003.

Zárate-de Landa A, Zuniga-Juarez J. E., Loo-Yau J. R., Reynoso-Hernandez J. A., Maya-Sanchez M. C., and del Valle- Padilla J. L., “Advances in linear modeling of microwave transistors,” IEEE Microwave Magazine, Vol. 10, pp. 100-111, 146, April 2009.

Zárate-de Landa A. Zúñiga-Juárez J.E, Reynoso-Hernández J.A., Maya-Sánchez M.C., Piner E.L., and Linthicum K.J., “A New and Better method for extracting the parasitic elements of the on-wafer GaN transistors,” in IEEE MTT-S Symposium Digest., Honolulu, Hawaii, pp. 791-794, June 2007.

Zárate-de Landa A., Zúñiga-Juárez J.E, Reynoso-Hernández J.A., Maya-Sánchez M.C., del Valle-Padilla J.L. and Loo-Yau J.R., “An Investigation on the Modified cold-FET Method for Determining the Gate Resistance and Inductance of the Packaged GaN and SiC Transistors,” 70th ARFTG Conference Digest, November 2007.

Zárate-de Landa A., Pulido-Gaytan M.A., Reynoso-Hernandez J.A., Roblin P. and Loo-Yau J.R., “A Neural Network Approach to Smooth Calibrated Data Corrupted From Switching Errors,” 80th ARTFG Conference Digest, November 2012.

Zárate-de Landa A.,Roblin P., Reynoso-Hernández J.A. and Loo-Yau J.R., “Modeling the IV Curves and Its Derivatives of Microwave Transistors Using Neural Networks,” IEEE Microwave and Wireless Components Letters, Vol 22., No. 9, pp 468-470, September 2012.

Zhang Q. J., Gupta K. C., “Neural Networks for RF and Microwave Design,” Boston, MA: Artech House, July 2000.