Characterisation of Gallium Nitride HFETs using Non-Linear Measurement Systems

A thesis submitted to Cardiff University
in candidature for the degree of

Doctor of Philosophy

By

Peter McGovern

Division of Electrical and Electronic Engineering
School of Engineering
Cardiff University
United Kingdom

September 2007

UMI Number: U585151

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U585151
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Table of Contents

Summary v
Acknowledgements vi
List of Publications vii

Chapter 1. Introduction 1
1.1 Development of Wireless Communications and Power Amplifiers 1
1.2 Modern Communications Systems and Power Amplifiers 2
1.3 Thesis Objectives and Structure 4
1.4 References 6

Chapter 2. GaN HFET Overview 7
2.1 Histroy of GaN Semiconductors 7
2.2 GaN Power Amplifiers 9
2.2.1 The Potential of GaN Power Amplifiers 9
2.2.2 GaN Power Amplifier Applications 11
2.3 Device Structure and Operation 12
2.4 Device Issues and Performance 15
2.4.1 Substrates 15
2.4.2 Electron Trapping and Dispersion 16
2.4.3 Device Performance 17
2.5 QinetiQ GaN HFETs 18
2.5.1 DC Characterisation 20
2.5.2 Small-Signal Characterisation 22
2.5.3 RF Power Characterisation 23
2.5 Summary 25
2.6 References 25

Chapter 3. Measurement System Overview 28
3.1 Introduction to Non-Linear RF Measurements 28
3.2 Non-Linear Measurement System at Cardiff University 31
3.2.1 The Microwave Transition Analyzer 31
3.2.2 Measurement System Architecture 34
3.2.3 Calibration 37
3.2.4 Active Loadpull 38
3.2.5 Two-Tone Measurement System 41
3.3 Summary 43
3.4 References 44

Chapter 4. Effects of Dispersion on Power and Efficiency Performance 45
4.1 Introduction to Dispersion 45
4.2 Origins of Dispersion 46

4.2.1 Self-Heating in HFETs 47
4.2.2 Trapping Effects in HFETs 49
4.2.3 Device State 51
4.2.4 Manifestation of Dispersion 51

4.3 RF Waveform Analysis of Dispersion in GaN HFETs 53
4.3.1 Waveform Measurements at Class A Bias Points 54
4.3.2 Waveform Measurements at Class B Bias Points 56
4.3.3 The Importance of Harmonic Loadpull 57

4.4 Pulsed I-V Analysis of Dispersion in GaN HFETs 59
4.4.1 Pulsed I-V Measurements 62
4.4.2 Correlation Between RF and DC Measurements 64

4.5 Design for Maximum Efficiency 68
4.5.1 Efficiency Enhancement Techniques 69
4.5.2 Efficiency Measurements 71
4.5.3 Waveform Engineering for Maximum Efficiency 73

4.6 Summary 75

4.7 References 76

Chapter 5. Linearity Investigations of GaN HFETs 77

5.1 Non-Linear distortion 77

5.2 CW Large-Signal Characterisation 80
5.2.1 CW Theory 80
5.2.2 CW Large-Signal Characterisation 83
5.2.3 Comparison With a GaAs HEMT 89

5.3 Two-Tone Large-Signal Characterisation 92
5.3.1 Two-Tone Measurements 94
5.3.2 IP3 Measurements 98
5.3.3 Comparison of Devices With Different Epitaxial Layer Design 101

5.4 Memory Effects in Power Amplifiers 105
5.4.1 Origins of Memory Effects 106
5.4.2 Envelope Analysis of Two-Tone Results 108
5.4.3 Scope-Based Non-Linear Measurement System 111
5.4.4 Measurement of Memory Effects 116
5.4.5 Effect of Carrier Impedance Terminations 119
5.4.6 Effect of Baseband Impedance Terminations 120
5.4.7 The Importance of Baseband Impedance Terminations 121
5.4.8 Envelope Domain Analysis 124

5.5 Summary 127

5.6 References 128

Chapter 6. Conclusions 130

6.1 Conclusions 130

6.1 Further Work 133

Summary

One of the weakest links in any communications system is the power amplifier, largely due to the necessary trade-off between linearity and efficiency. The challenge of power amplifier design is therefore to provide maximum power and efficiency while maintaining highly linear operation. Gallium Nitride HFET devices are of great interest in power amplifier applications and are potentially the best device technology to provide high power amplification at high frequencies. At the outset of this work, the development of GaN HFETs was in its early stages and the primary aim of this work is to help the development of the QinetiQ GaN program and to investigate the suitability of GaN HFETs for use as power amplifiers. This was achieved through the use of novel measurements using the custom RF non-linear time domain measurement systems at Cardiff University.

These QinetiQ devices exhibit the common problem of DC-RF dispersion, which results in much degraded RF performance relative to that predicted from DC measurements. It has been shown that the devices exhibit a direct correlation between dispersion and channel current, and that it is possible to overcome some of the unwanted side-effects of dispersion through suitable design, so that efficiency can be improved to acceptable levels, while maintaining high power operation. It has also been shown that there is a strong correlation between RF waveform measurements and pulsed I-V measurements, and that through careful choice of bias point, pulsed I-V measurements can be used to accurately predict the RF performance of a device.

Through large-signal CW and two-tone characterisation, it has been determined that these devices operate in a very similar way to that predicted by simple theory and exhibit little AM-PM behaviour. Their CW operation is also found to be less complex than that of GaAs pHEMTs, and their two-tone operation correlates well with the CW analysis. They should therefore be ideally suited to power amplifier applications in terms of linearity as well as power and frequency performance. Finally, after the development of a new two-tone time domain measurement approach, it was determined that the devices suffer from intrinsic electrical memory effects due to their transit delay but are otherwise inherently free from memory within the measurement bandwidth – the range of measurable modulation frequencies is 200KHz-25MHz. All other memory effects detected are not caused by the device but are extrinsic electrical memory effects caused by the impedance environment that is presented to the device.

Acknowledgements

I would like to firstly thank my supervisor Professor Paul Tasker whose immense knowledge, experience and guidance was invaluable in all aspects of my research.

I must also thank the employees of QinetiQ in Malvern who supported this work, both financially and intellectually, particularly Dr. Mike Uren.

I would also like to thank the employees of Mimix Broadband in Belfast (formerly Celeritek UK) who gave me the opportunity to study at Cardiff University, particularly Dr. Andrew Patterson and my good friend Sean McAllister.

I am also grateful to all members of my research group, past and present, who I have worked with and on whose work I have built. In particular I must thank my good friends Dr. Jonathan Lees, Dr. Tudor Williams, and Dr. Scott Lewis.

Finally, I would like to thank all of my family and friends for their support and encouragement throughout.

List of Publications

1. **P. McGovern**, J. Benedikt, P. J. Tasker, J. Powell, K. P. Hilton, J. L. Glasper, R. S. Balmer, T. Martin, and M. J. Uren, "Analysis of DC-RF dispersion in AlGaN/GaN HFETs using pulsed I-V and time-domain waveform measurements", IEEE MTT-S International Microwave Symposium Digest, Long Beach, June 12th-17th, 2005, Vol 2, pp 1411-1414.
   Abstract: AlGaN/GaN HFETs have been analyzed using pulsed I-V measurements and RF time-domain measurements in an attempt to analyze the phenomenon of DC-RF dispersion and achieve maximum RF performance. The pulsed I-V measurements exhibited the common problem of current slump, as did the RF power performance of the device. The RF time-domain waveforms are used to show that the RF knee-walkout increases with drain bias voltage, thus negating any improved power and efficiency performance that a larger voltage swing would achieve. It was also found that the degree of RF knee-walkout changes depending on the class of operation of the device. Although current slump is evident, it is not permanent, and there is negligible degradation of the device after prolonged RF stimulus.

2. **P. McGovern**, P. J. Tasker, J. Powell, K. P. Hilton, J. L. Glasper, R. S. Balmer, T. Martin, and M. J. Uren, "Experimental Investigation of DC-RF Dispersion in AlGaN/GaN HFET's Using Pulsed I-V and Time-Domain Waveform Measurements," presented at The International Conference on Compound Semiconductor Manufacturing Technology, New Orleans, USA., 2005.
   Abstract: AlGaN/GaN HFETs have been analyzed under DC and RF stimulus in an attempt to analyze the phenomenon of DC-RF dispersion. DC pulsed I-V measurements were performed where the pulse "off" state was set to different bias points to simulate a Class A bias condition at various drain voltages. RF time-domain waveform measurements were also performed. The I-V measurements exhibited the common problem of current slump, as did the RF power performance of the device. Interestingly, it was found that there was strong correlation between the pulsed I-V and RF measurements when they are considered as a function of bias point. Although current slump is evident during pulsed I-V measurements, it is not permanent, and there is negligible degradation of the device after prolonged RF stimulus.

3. **P. McGovern**, D. J. Williams, P. J. Tasker, J. Benedikt, J. Powell, K. P. Hilton, R. S. Balmer, T. Martin, and M. J. Uren, "Single-tone and two-tone time-domain large signal characterization of GaN HFETs operated in class A," IEEE MTT-S International Microwave Symposium Digest, Forth Worth, Texas, June 6th-11th, 2004, Vol 2, pp 825 - 828.
   Abstract: GaN HFET's have been analyzed under single-tone and two-tone excitation, using an error corrected time domain measurement system. This approach yields a better understanding of the device's non-linear behavior, particularly with respect to intermodulation distortion. It has been found that under single-tone excitation, the AlGaN/GaN HFET produced fundamental and third harmonic output with virtually constant phase (no AM-PM), even when operated well into compression. This would imply that the device should exhibit symmetric intermodulation products under two-tone excitation, and this was experimentally found to be the case. These results further indicates that AlGaN/GaN HFET technology has the potential for use not only in high power and high efficiency, but also high linearity applications.

4. J. Lees, T Williams, **P McGovern**, S Woodington, S Cripps, J Benedikt, and P Tasker, "Demystifying Device Related Memory Effects using Waveform Engineering and Envelope Domain Analysis" accepted for presentation at European Microwave Conference, Amsterdam, 2008.
Abstract: Although memory-less pre-distortion together with memory mitigation techniques are extensively used to counter the effects of PA non-linearity, the sources of non-quasi-static non-linear behaviour are still not fully understood. In this paper, it is argued that it may also be the case that some observations, generally considered to be evidence of 'mysterious' device memory effects have much more fundamental origins, such as simple linear delay or non-ideal impedance environments that vary over bandwidth. This paper presents detailed modulated measurements that aim to help clarify this problem and to isolate some of the contributors to what has now become a rather general 'catch-all' term of 'memory effects'. The analysis is based on modulated (two-tone) waveform measurements and envelope analysis of initially a simple passive line structure, and later a 2W GaN die operating at 2.1 GHz, and over a wide range of tone separations, up to 80 MHz.

5. M. J. Uren, D. G. Hayes, R. S. Balmer, K. P. Hilton, J. O. Maclean, T. Martin, C. Roff, **P. McGovern**, J. Benedikt, and P. J. Tasker, "Control of Short-Channel Effects in GaN/AlGaN HFETs," presented at European Microwave Conference, Manchester, UK, 2006.
Abstract: GaN/AlGaN HEMTs can suffer from short channel effects as a result of insufficient buffer doping. The paper show that controlled iron doping of the GaN buffer during MOVPE growth can suppress all short-channel effects in 0.25mum gate length devices. The authors show that optimised iron doping has no effect on the RF output power or on the knee walkout (current-slump), but significantly improves the power added efficiency

6. C. Roff, **P. McGovern**, J. Benedikt, P. J. Tasker, R. S. Balmer, D. J. Wallis, K. P. Hilton, J. O. Maclean, D. G. Hayes, M. J. Uren, T. Martin, "Detailed Analysis of DC-RF Dispersion in AlGaN/GaN HFETs using Waveform Measurements", presented at European Microwave Conference, Manchester, UK, 2006.
Abstract: Detailed time-domain IV waveforms at RF frequencies are employed for characterisation of AlGaN/GaN HFETs in order to steer and advance device development. The IV time-domain data is used to isolate the separate effects of pinch-off and knee-walkout behaviour in limiting device performance. Furthermore, the waveform measurements which are obtained with a previously unseen level of detail, allowed the direct extraction of optimum device operating conditions

7. T. Martin, M. J. Uren, R. S. Balmer, D. J. Wallis, K. P. Hilton, A. R. Barnes, D. G. Hayes, **P. McGovern**, P. J. Tasker, " GaN HFET Development at QinetiQ," presented at European Gallium Arsenide and Other Semiconductor Application Symposium (EGAAS), 2005.
Abstract: The AlGaN/GaN HFET project at QinetiQ has the capability to grow high quality layers using MOVPE, fabricate HFETs with 0.8 and 0.25 /spl mu/m gate length and fabricate amplifiers. Here we present recent work on topics as diverse as X-ray determination of aluminium concentration, 0.25 Ohm.mm Ohmic contacts, measurement of saturated velocity, current slump and a 57 W hybrid amplifier.

Chapter 1. Introduction

1.1 Development of Wireless Communications and Power Amplifiers

The first wireless transmissions were performed over 100 years ago after the discovery and application of electromagnetic (EM) wave propagation. The existence of EM waves was predicted by James Maxwell as early as 1867 and they were first detected by Heinrich Hertz in 1887 [1]. Nikola Tesla was the first person to successfully transmit radio waves in 1893 and has been credited as the inventor of radio by the United States Supreme Court, but it was Guglielmo Marconi who was first to commercialise its applications by establishing wireless telegraphy in the mid 1890's and patenting his inventions in 1897. Further developments quickly followed with the development of Amplitude Modulation (AM) and Frequency Modulation (FM) transmission techniques and audio transmissions. These early transmissions were received using a crystal radio receiver, or 'crystal set', which was a very simple and effective device.

The key to wireless communications systems has always been the generation and amplification of the transmitted RF signal. Before the 1920's RF power was generated by spark gap, arc, or alternator transmitters [2], which were able to produce up to 1MW of RF power, but only at frequencies up to about 1MHz [3]. These devices allowed little frequency control so the transmitted signal was broadband and only one transmitter could transmit at a time. During the mid 1920's, amplifying vacuum tubes, or thermionic valves, revolutionised radio transmitter and receiver technology because they enabled the transmission of pure CW signals, which were amplified from more reliable frequency generators, and allowed operation at much higher frequencies. Many of the concepts of modern power amplifier (PA) design were first developed for vacuum tube applications.

The development of the discrete solid state RF power amplifier during the 1960's and 1970's again revolutionised the industry. These devices enabled transmitters to become cheaper, smaller, and more mobile as well as more reliable. Advances in silicon semiconductor technology in the 1980's led to the integration of the discrete PA with other passive components and circuits to form Monolithic Microwave Integrated Circuits (MMICs). This again reduced size, cost, and power consumption of the overall system and made portable communications equipment possible.

Further advances in semiconductor technology during the 1980's and 1990's led to the development of transistors and power amplifiers based on more 'exotic' III-V semiconductor materials such as Gallium Arsenide (GaAs) and Indium Phosphide (InP). The resulting devices are not traditional Field Effect Transistors (FETs) or Bipolar Junction Transistors (BJTs), but rather Hetrostructure Field Effect Transistors (HFETs) and Hetrostructure Bipolar Transistors (HBTs), and their physical properties means that they can operate at much higher frequencies and higher output powers than their silicon-based counterparts. GaAs HBTs in particular have become very successful and are now the device of choice in mobile telephone handsets because of their low operating voltage, high efficiency, and high output power.

1.2 Modern Communications Systems and Power Amplifiers

Modern wireless communications systems are becoming increasingly complex and widespread and include everything from satellite systems and mobile telephones to Bluetooth connections. In all of these systems, the power amplifier is still one of the key components and is very often the limiting factor in its performance capabilities. The main purpose of a PA is to amplify a signal to suitable levels to allow it to be transmitted. The power amplifier is also the major power consumer in any communications system and the overall efficiency of the system is largely dependent on its efficiency. Therefore, it has

always been important to develop PAs that can operate with high efficiency while maintaining high power handling capabilities.

However, in more recent times, because wireless systems have become so abundant and the RF frequency spectrum in which the wireless data is transmitted is finite, there is increased necessity to develop systems with smaller bandwidth requirements and that can operate at increasingly high frequencies. This is in addition to the requirement of continuously delivering higher data transmission rates.

Signals with high data rates and small bandwidth requirements can be achieved, but only through the use of sophisticated digital modulation techniques which require low error rates and linear transmission. Unfortunately, the non-linear nature of semiconductor power amplifiers means that in PA design, there exists a trade-off between efficiency and linearity, so that these advanced modulation schemes cannot be transmitted efficiently [4]. For example, in first generation mobile telephone systems such as the Nordic Mobile Telephone (NMT) system, a constant envelope modulation scheme was employed. This form of modulation has no significant linearity requirements and highly non-linear but power-efficient PAs were used [5]. In third generation systems (3G), complex modulation schemes such as Code Division Multiple Access (CDMA) are used to achieve high data transmission rates, but the efficiencies of PAs that meet the linearity requirements of a CDMA system are very low. Therefore, the operating times of 3G handsets are relatively poor. This continuous trade-off between efficiency and linearity is the biggest challenge facing the power amplifier designer.

The stringent requirements of PA design for the plethora of new and existing wireless communications systems means that new semiconductor material technologies are continuously being examined as candidates for future power amplifier transistors. Different device technologies have different properties and are therefore suitable for different application. For example, silicon-based

devices can only operate at relatively low power densities and low frequencies, but because they are very cheap and reliable they are still commonly found in most mobile phone base stations. GaAs devices are capable of supplying relatively high power densities at frequencies up to around 20GHz [6], while InP devices have proven capable of extremely high frequency operation of over 100GHz, but at relatively low power densities. The latter are also very expensive but are used in very high frequency applications such as military satellite communications.

The ultimate requirement of a power amplifier transistor that can operate simultaneously at very high power densities and very high frequencies, while maintaining highly linear and efficient operation has not yet been realised and may not be possible, but the most interesting device material currently under investigation is Gallium Nitride (GaN). Development of the GaN material system has suffered from many problems, but GaN HFET devices have been shown to be capable of operating simultaneously at very high frequencies and very high power densities [7], and are therefore considered to have the potential to be the most exciting and versatile PA technology available.

1.3 Thesis Objectives and Structure

In theory GaN HFETs should make very good power amplifier devices, but at the outset of this work the technology was largely unproven. QinetiQ is one of many companies that are current developing GaN HFET technology [8], and the overall objective of this work is to help understand, develop, and characterise the QinetiQ GaN devices, as well as determining their overall potential as power amplifiers via the use of the unique capabilities of the non-linear RF measurement systems at Cardiff University.

Chapter 2 reviews the development of gallium nitride materials and the current state of the art in GaN HFET technology. The GaN devices developed

by QinetiQ are also described and an initial assessment of their potential as power amplifiers is achieved through DC, S-parameter and RF power measurements.

Chapter 3 introduces the non-linear RF measurement system at Cardiff University, as much of the data presented in the subsequent chapters is measured using this system and it is further developed during the course of this work. The basic measurement theory is described for CW and two-tone measurements, as well as the system calibration technique and the unique active loadpull setup.

Chapter 4 examines the well known problem of dispersion in GaN HFETs and its effect on power and efficiency performance via RF waveform measurements and pulsed I-V measurements. The advantages of these measurement techniques are illustrated and the information gained from them is used to achieve optimum performance from the GaN devices.

Chapter 5 is an investigation of the linearity of the GaN devices. This is achieved via CW and two-tone power and phase measurements using the non-linear measurement system. A comparison is also made between the GaN device and a typical GaAs device, and between two GaN devices with different epitaxial layer structures. The problem of memory effects is also addressed and this required the development of a new measurement approach in an attempt to measure the memory effects and better understand their origin and how to minimise their effects.

Finally, Chapter 6 contains the conclusions to the work along with discussion and suggestions for further work.

1.4 References

[1] H. Sobol, "Milestones of Microwaves," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 50, pp. 594-611, 2002.

[2] W. J. Bryon, "The arc method of producing continuous waves," *Communications quarterly*, vol. 8, pp. 47-65, Summer 1998.

[3] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "RF and Microwave Power Amplifier and Transmitter Technologies - Part 1," in *High Frequency Design*, 2003, pp. 22–36.

[4] P. Kenington, *High Linearity RF Amplifier Design*: Artech House, 2000.

[5] H. Holma, A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communnications*, 2 ed: John Wiley and Sons Ltd, 2002.

[6] M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur, *Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, and SiGe*: John Wiley and Sons, 2000.

[7] D. Pavlidis, "Recent Advances in III-V Nitride Electronic Devices," presented at IEEE Electron Device meeting digest, 2004.

[8] T. Martin, M. J. Uren, R. S. Balmer, D. J. Wallis, K. P. Hilton, A. R. Barnes, D. G. Hayes, P. McGovern, and P. J. Tasker, "GaN H-FET Development at QinetiQ," presented at European Gallium Arsenide and Other Semiconductor Application Symposium (EGAAS), 2005.

Chapter 2. GaN HFET Overview

2.1 History of GaN Semiconductors

Semiconductor materials have been known since the early twentieth century when the first known report of a light-emitting solid-state diode was made in 1907 by the British engineer H. J. Round [1]. However, no practical use was made of the discovery for several decades. During the early twentieth century, before the development of the semiconductor transistor, vacuum tubes became common in radio electronics as diodes and amplifiers but these were large, inefficient, and unreliable. The first practical semiconductor transistor was built in 1947 by William Shockley, John Bardeen, and Walter Brattain at Bell Labs for which they were later awarded the Nobel Prize in physics.

As development of semiconductors continued through the 1950's and 1960's, the main commercial semiconductor products were diodes and transistors for radio applications. Most of the early work focused on the development of silicon semiconductors and led to the development of integrated circuits (ICs), and microprocessors. Although silicon became the semiconductor material of choice for most applications, largely because it was the most robust and easiest to work with, there has always been much interest in other semiconductor materials and their alloys. The reason for this interest is that different semiconductors have different physical properties, and therefore have potential applications that silicon devices cannot meet. There are now many applications for semiconductor devices, but the three largest markets are microprocessors, optoelectronics, and RF power amplifiers. There are also many emerging markets such as military and medical applications.

Microprocessor applications are by far the biggest semiconductor market, and Silicon has always been the material of choice for this application and continues to be, since device size and performance continue to roughly follow

Moore's law, due to continually improving manufacturing techniques. During the 1980's Silicon Germanium (SiGe) semiconductors were developed for very high speed processing and are now being introduced as continued development of silicon technology becomes limited by the intrinsic material properties. Some of the first semiconductor devices were low power amplifiers for military and commercial communications applications. Originally these PAs were made from silicon but the intrinsic material properties of silicon limit the frequency and power capabilities of this material. This led to the development of Gallium Arsenide (GaAs) PA devices in the 1980's. Other materials that have recently been developed for PA applications include SiGe, GaN, and Indium Phosphide (InP) [2].

Optoelectronics applications including LASERs, Light Emitting Diodes (LEDs), and solar cells have been available for decades, but have recently become the focus of much development. The first LEDs were made in the 1960's and were mainly made from GaAs-based materials, but could only emit visible light in the long-wavelength (red) region of the frequency spectrum [2]. Therefore, a major research aim was to develop materials that could produce light of different colours and greater intensity. Many different materials are available that can meet these criteria, but development of these devices was problematic, and it is only relatively recently that the full spectrum of colours has become available. This is mainly due to the development of GaN-based LEDs, which were the first devices able to produce bright light in the blue part of the visible spectrum [3].

Since the early days of compound semiconductor research it has been known that GaN and its related wide bandgap materials were ideal for the production of optoelectronic devices operating in the visible short-wavelength (blue) region of the frequency spectrum. However, the development of the materials took a long time and the first GaN-based LEDs and LASERs did not appear until the 1990's. This has led to the commercialisation of high-brightness blue LEDs and LASER diodes (LDs) [4]. The high-brightness blue LEDs completed

the range of primary colors, and made possible applications such as full-color LED displays and white LEDs, which are now used in lighting applications.

2.2 GaN Power Amplifiers

2.2.1 The Potential of GaN Power Amplifiers

Because of the fundamental physical properties of GaN, it has also long been recognised that GaN-based semiconductor materials offer great potential for many other applications including high voltage, high power, and high frequency transistor devices such as those required in power amplifier and high voltage switching applications. However, research into these applications has only been possible since the mid 1990's when the multi-million dollar research into GaN optoelectronic devices resulted in material technology that could also be used to make transistor devices [5].

The most desirable qualities of high frequency power amplifier devices are large power handling capability, high frequency operation, large breakdown voltage (for large operating voltage and efficiency, and because most commercial base stations operate at 28V or 48V), and high thermal conductivity (to remove the large amount of heat generated). Table 2.1 summarises the relative merits of the most common PA device technologies in these four categories, while Table 2.2 details the most important material properties of the different semiconductor materials [2]. It can clearly be seen that GaN is the only relatively mature semiconductor material system that currently has the potential to produce both high power and high frequency devices, and as such is an extremely interesting material for use as RF power amplifiers.

Table 2.2 shows that unlike Si and SiC, GaN and GaAs are direct bandgap materials, which is why they make such efficient optoelectronic devices. Also, GaN and SiC, are very wide bandgap materials which means that they have

relatively large breakdown electric fields, and are therefore suitable for high voltage applications. They also both have a large saturated electron velocity which contributes to high current densities and large power handling ability.

	Si	GaAs	SiGe	InP	4H SiC	GaN
Operating Voltage	Medium	Low	Low	Low	High	High
Power	Low	Medium	Low	Low	High	High
Frequency	Low	Medium	High	Very High	Medium	High
Thermal Conductivity	Low	Low	Low	Low	high	high

Table 2.1 – Comparison of power amplifier technologies.

	Si	GaAs	4H SiC	GaN
Bandgap (eV)	1.11	1.43	3.23	3.37
Bandgap Type	Indirect	Direct	Indirect	Direct
Electron Mobility (cm^2/V·s)	1350	8500	900	1350
Lattice Constant (Å)	5.43	5.65	3.07	3.19
Breakdown Field (MV/cm)	0.3	0.4	5	5
Saturated Electron Drift Velocity($x10^6$ cm/s)	6	10	20	25
Thermal Conductivity (W/cm·K)	1.3	0.55	3.7	2
Natural Hetrostructure	no	yes	no	yes
PA Technology	LDMOS	HFET/HBT	MESFET	HFET
JFM (normalised)	1.0	2.2	55.6	69.4

Table 2.2 – Semiconductor material properties

Another important feature of GaN and SiC is their relatively large thermal conductivity which is important to enable the removal of the heat produced by the device when operated at high power densities. However, unlike SiC, GaN forms a natural hetrostructure with AlGaN. This allows the production of GaN HFETs which have the potential to operate at higher frequency than their SiC MESFET counterparts. To emphasise this, the Johnson Figure of Merit (JFM) which defines the frequency-power product of a semiconductor device [6], and is often used to indicate the suitability of a device to power amplifier applications, is also shown in Table 2.2. In this case the JFM of each material is normalised to that of silicon, and clearly shows that of the material systems

currently available GaN has the potential to produce the best high frequency power devices.

2.2.2 GaN Power Amplifier Applications

The one major disadvantage of GaN devices is that wafers of GaN and its alloys are very expensive to produce and therefore GaN power amplifiers will probably never be as cheap to produce as their silicon or GaAs counterparts. For this reason, the deployment of GaN devices is likely to be mainly limited to high power and high frequency transmitter applications, particularly in high power base stations and satellite communications systems, where it has significant advantages over other device technologies.

Other possible applications such as low power amplifiers for mobile telephone handsets are not viable for GaN devices because of its relative expense. Also, the performance advantages of GaN would not be maximised in such applications, and the requirement for device integration such as that seen in GaAs technology is not yet achievable. However, other possible applications are niche markets such as RADAR and low-noise receivers where the large breakdown voltage of GaN means that spurious signals can be absorbed without the receiver being destroyed. Therefore the receiver architecture can be simplified, resulting in reduced overall system noise figure.

GaN power amplifiers also offer several system benefits. Due to their relatively large power density, the output impedance of GaN devices will be larger than that of competing technologies, resulting in reduced matching network complexity and loss, and therefore reduced size and weight of the overall system. This also eases implementation of the multi-octave operating bandwidths that are required for applications such as RADAR. Their large thermal conductivity and ability to operate at relatively high temperatures means that GaN PAs should require less cooling than other devices, and their large breakdown voltages mean than they can operate directly from a 28V

base station supply rail without the need for DC-DC voltage converters. This alone can improve overall system efficiency by 5-10% [7], and coupled with their small size and minimal matching circuitry means that they can be mounted closer to a base station antenna, resulting in a system that is more compact, requires less cooling, and is less lossy than those based on competing technologies such as silicon LDMOS PAs. This could greatly reduce operating costs and therefore offset the expense of the GaN infrastructure. The cost of GaN devices is also expected to fall rapidly due to the continued development of the technology for optoelectronic applications, particularly in the lighting industry.

2.3 Device Structure and Operation

GaN-based transistors are still a relatively new technology but they have undergone much research since the first description of a GaN HFET in 1993 [5]. Figure 2.1 shows the structure of a basic GaN HFET device. The current lack of native GaN substrates of sufficient size and quality means that the GaN layers must be grown via heteroepitaxy on compatible substrates, most commonly sapphire, or SiC. These substrates are severely mismatched to GaN so the nucleation layer is one of the most critical aspects of the growth and typically consists of GaN, AlGaN, or AlN. Then the intrinsic GaN and AlGaN layers can be grown by MBE or MOCVD, and device fabrication is completed with the formation of the gate and ohmic contacts. A Silicon Nitride (SiN) passivation layer is also often added in an attempt to eliminate dispersion between the DC and large-signal AC characteristics [8].

The operation of a HFET device is similar to that of a normal FET except that a HFET has no doped channel in which current flows. Instead, a conductive channel is formed at the hetrointerface between the GaN and AlGaN layers in the form of a two dimensional electron gas (2DEG). In a FET, current flow is controlled via the modulation of the channel cross-section by the gate, but in a

HFET it is the electron concentration in the 2DEG that is modulated. Electrons move from the source ohmic contact into the 2DEG channel formed in the unintentionally doped GaN, and the current is modulated by the intervening Schottky gate.

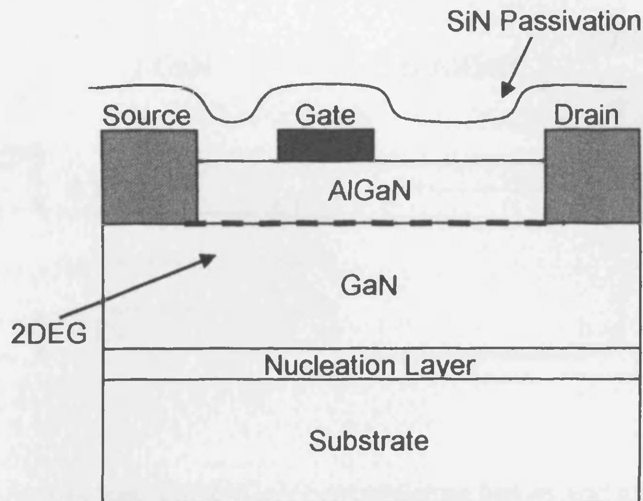


Figure 2.1 – Basic GaN HFET structure

The current carrying ability of FETs is maximised by increasing the channel doping. However, high channel doping causes decreased mobility and effective velocity due to impurity scattering. This compromises the high frequency performance of the device and creates a trade-off between power and frequency performance. The main advantage of having a 2DEG channel is the possibility of increasing the conductivity by increasing the carrier concentration without suffering the mobility degradation effects due to impurity scattering.

In the most common HFET structures the wide bandgap barrier (AlGaN) is doped n-type while the narrow bandgap channel remains undoped. As a result, electrons diffuse from the wide bandgap material into the narrow bandgap material to minimise their energy as illustrated in Figure 2.2. This

process continues until a balanced Fermi level is formed in the two materials and equilibrium is established. Because of the electrostatics, a new triangular well forms on the narrow bandgap side of the hetrojunction. It is here that the 2DEG forms.

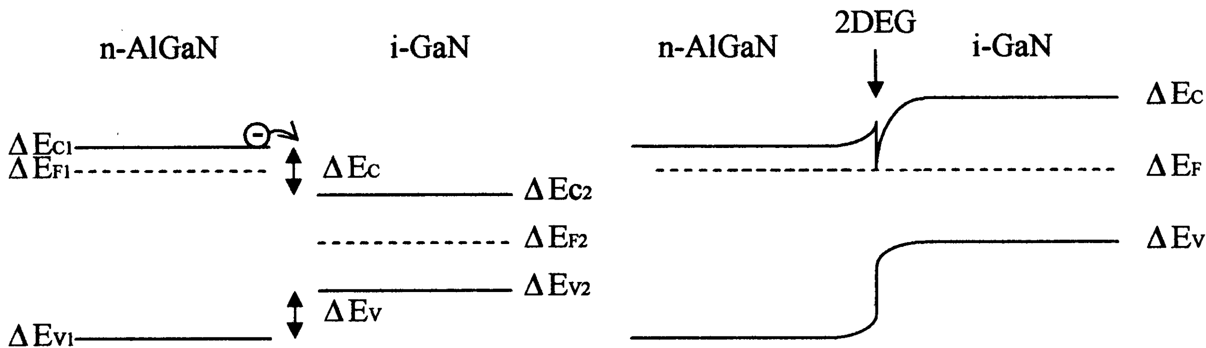


Figure 2.2 – An idealised GaN/AlGaN hetrostructure before and after the junction is complete to form the 2DEG.

In the structure above, a doped barrier provides electrons to the undoped channel, thus spatially separating the channel charge carriers from their ionised donors. In this manner the hetrostructure channel is capable of delivering high carrier concentration with high mobility as impurity scattering is minimised in the undoped channel. As an added advantage, surface scattering is also reduced by moving the current-carrying region below the barrier.

In most HFET devices the carriers originate entirely from the n-type bulk dopants within the barrier layer, but GaN HFETs are different from conventional III-V HFETs because of the way in which the 2DEG is formed. In a typical GaAs HFET the free electrons come from a silicon dopant supply layer in the AlGaAs gate layer, but the GaN/AlGaN 2DEG channel is formed as a result of piezoelectric and spontaneous polarisation effects which arise in the AlGaN layer [9]. The 2DEG concentration is directly related to the

concentration of Al in the AlGaN layer and the polarisation fields are strong enough to allow for the formation of a 2DEG without any barrier doping at all, although n-doping with Si is sometimes employed [10].

2.4 Device Issues and Performance

2.4.1 Substrates

A particular problem for GaN HFETs is the lack of a native GaN substrate. The growth of bulk GaN boules has proven difficult and the largest commercially available GaN substrates are currently just two inches in diameter [11], and are very expensive. Because of this, most GaN devices have been grown on foreign substrates such as sapphire or SiC.

Sapphire is very cheap but has a large lattice and thermal mismatch with GaN while SiC is much more expensive but has a smaller thermal and lattice mismatch. The size of both sapphire and SiC substrates is typically four inches (100mm). The lattice mismatch of these substrates is accommodated by strain in the buffer layers, but this strain is often released by the formation of dislocations in the substrate and has been blamed for the large dislocation densities seen in GaN devices which leads to device failure [12]. AlGaN-GaN layers grown on SiC or sapphire substrates typically have threading dislocation densities on the order of $1x10^{9}cm^{-2}$ [13], increasing gate leakage current and reduced device lifetime. The availability of low-dislocation density GaN substrates would help reduce this problem, and such substrates have already been reported with dislocation densities as low as $5x10^{6}cm^{-2}$ [11], although with a diameter of only two inches.

GaN devices based on silicon substrates have also been developed [14] and although they have the same problems as GaN-on-Sapphire and GaN-on-SiC, and silicon has larger thermal and lattice mismatches with GaN, it is inexpensive and is available in wafers of over 12 inches in diameter. This

makes the technology attractive because material and processing costs can be reduced significantly, although at the expense of overall performance.

Currently, due to the absence of a native GaN substrate, the most promising and theoretically best devices are those based on GaN-on-SiC technology. This is reflected in the fact that most of the published record performance figures of GaN-based devices were produced by such devices. This is because of all the possible substrates, SiC has the smallest lattice and thermal mismatch with GaN, and highest thermal conductivity. Large thermal conductivity allows efficient removal of heat from the device which is essential if it is to operate reliably at high power densities. The small lattice mismatch means that the dislocation densities and device failure is generally reduced, and the small thermal mismatch means that as the device changes temperature and expands and contracts during operation it is less likely to create dislocations that result in device failure.

2.4.2 Electron Trapping and Dispersion

Probably the biggest obstacle facing the development of GaN HFET devices is the problem of DC-RF dispersion. This is the difference between the measured DC and RF characteristics whereby the RF performance is much less than would be expected according to the DC performance. Dispersion is a trapping related phenomenon that has been shown to exist in other device technologies [15]. It is known that electron trapping is a significant problem in GaN devices and electron traps have been found to be present in both the buffer and surface regions of GaN HFETs [16]. Some researchers have found that the buffer traps can cause dispersion [17], but it is believed that surface traps are the biggest problem and the main cause of dispersion because it has been reported that proper control of the surface traps can result in significant or even complete removal of dispersion.

Several techniques have been developed to control the surface traps and reduce dispersion with SiN passivation of the device surface being the most common method employed. This passivation layer terminates the surface charge trapping sites and makes them inaccessible to the current carrying electrons [8], [18]. It has also been shown that the addition of a GaN cap layer to the device structure can increase the surface-channel distance to a point where surface potential fluctuations can no longer modulate the channel charge. This results in the reduction or elimination of dispersion effects [19], [20]. Some groups have also successfully reduced dispersion effects through the introduction of field plates above the gate-drain region of the device [21], [22]. The field plate has the effect of controlling the shape of the electric field at the drain side of the gate and reducing the peak field strength. This reduces the high-field trapping effects and also has the useful side-effect of increasing the device breakdown voltage.

2.4.3 Device Performance

Regardless of the problems associated with GaN devices, its performance advantages over other device technologies are significant and many research groups have produced devices with very impressive performance. For example, the record power performance of a GaN HFET is now over 40W/mm [23], while breakdown voltages of over 900V have been recorded by others [24], and frequency performances of f_T=230GHz and f_{MAX}=150GHz have been measured [25]. Although these record performance figures have been achieved on devices that were designed for optimal performance of one specific parameter, many companies are now in the early stages of commercialising GaN power amplifiers for base station applications and these devices also exhibit very good performance. For example, Eudyna Devices now has a commercially available range of GaN high power amplifiers which have breakdown voltages greater than 350V. They are designed for operation in the S-band range at around 2.1GHz and at a drain voltage of 50V. This results in output powers of around 5W/mm [26].

These figures are in contrast to the typical performance of comparable Si LDMOS and GaAs HFET power amplifiers. It is difficult to directly compare different device technologies because they are often optimised for different applications but in general, Silicon LDMOS devices cannot operate at drain voltages much greater than 28V, while their power performance is only about 0.5W/mm and their frequency response is so poor that they are of little use in applications above S-band [27]. Conventional GaAs HFETs typically cannot operate at drain voltages above 10V and their power performance is around 1W/mm. Their frequency performance is much better than that of Si LDMOS but worse than that of GaN devices and they are of limited use in applications above 6 GHz [28].

Overall, Si LDMOS devices can operate at relatively high voltages but have poor power and frequency performance, while GaAs HFETs have better frequency and power performance but lower breakdown voltages. Of course the performance limitations of these devices are continually being improved, but at this moment in time there is no power amplifier technology that has the combined high power and high frequency capabilities of GaN HFETs.

2.5 QinetiQ GaN HFETs

QinetiQ have had a GaN research program since 1998 which is supported largely by the UK MoD. The program capabilities include epitaxy growth, material and electrical characterisation, device processing, and circuit fabrication and modelling [29]. The standard GaN HFET structure currently produced by QinetiQ is shown in Figure 2.4 and consists of a $30nm$ undoped $Al_xGa_{1-x}N$ (x=0.25) layer on a 1.2μm undoped GaN layer. This is grown on a 4H SiC substrate, with an appropriate nucleation layer.

Epitaxial growth is carried out in a Thomas Swan close-coupled 'showerhead' MOVPE reactor which can grow wafers from 2" to 4" with excellent cross-

wafer uniformity. The AlGaN layer is undoped so that the 2DEG charge is entirely generated by spontaneous polarisation and piezoelectric effects without the need for modulation doping. The 2DEG concentration is therefore directly related to the concentration of Al in the AlGaN layer. After formation of the gate and ohmic contacts, the structure is completed with the formation of plated gold metal contacts and airbridges where necessary and the deposition of a SiN passivation layer via PECVD.

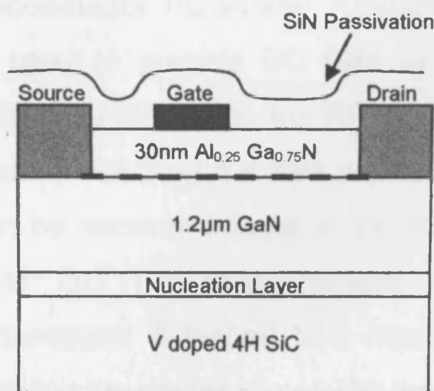


Figure 2.4 – HFET layer structure with an AlN exclusion layer

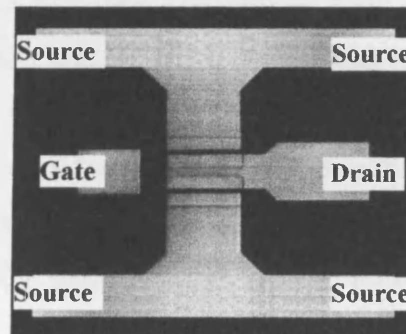


Figure 2.5 – Photograph of HFET on-wafer device

All devices tested and presented in this thesis follow the basic structure described above and are relatively small devices of gate width $2x100\mu m$ or $2x125\mu m$, such as that shown in Figure 2.5. DC and RF testing is achieved via on-wafer probing using two Ground-Signal-Ground (GSG) probes which connect to the gate and source on the input side, and the drain and source on the output side via the gold contact pads shown in Figure 2.5. Measurements are performed on these small devices because they are generally easier to work with than larger devices since they operate at lower RF powers and DC currents, although the results of the measurements are scaleable to larger devices.

2.5.1 DC Characterisation

DC measurements are normally the first characterisation performed on a device and are very important and useful for initial assessment and screening of devices. This is particularly important with new device technologies such as GaN, where cross-wafer device uniformity and yield is often not good.

DC characterisation is performed on-wafer using an Agilent 4155B Semiconductor Parameter Analyzer. This instrument is very accurate and is also used to provide DC bias to the DUT during RF characterisation. The 4155B is connected to the RF probes via a bias tee, and the resistance of the DC path, including the bias tee, is calculated so that the voltage drop through it can be accommodated in the measurement software. Figure 2.6 shows a typical DC I-V measurement for the device described earlier. The measurement is limited to a maximum of 30V and 2W DC power to avoid damaging the device during DC testing.

DC analysis can yield a lot of useful information and from a power amplifier point of view some of the most important figures of merit are breakdown voltage (V_{BD}), saturated drain-source current (I_{DSS}), knee voltage (V_{knee}), transconductance (g_m), and on-resistance (R_{on}). The DC I-V indicates that the device is well behaved and stable, and although it cannot be measured with the 4155B, it has been established that the breakdown voltage of these devices is in excess of 100V. It can also be seen that $V_{knee} \approx 5V$ and $I_{DSS(0)} \approx 160mA$, or $800mA/mm$.

The potential output power of the device can be calculated from Equation 2.1 and indicates that because of its large breakdown voltage and current density, it should be capable of producing an output power density of around $10W/mm$. Transconductance is important because it gives an indication of the small-signal gain of the device, and in this case, $g_m = 20mS$, or $100mS/mm$. Finally, the on-resistance of the device is importance because it determines

slope of the DC I-V up to the knee region, and therefore needs to be as low as possible to enable a low knee voltage which facilitates high output power density and efficiency.

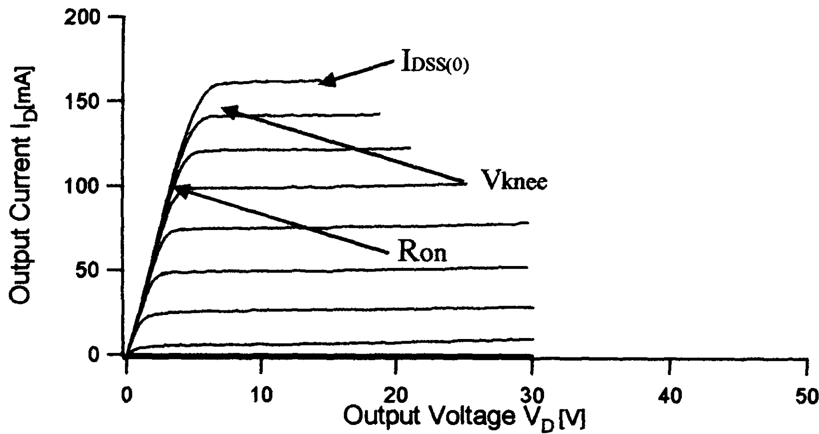


Figure 2.6 – Typical DC I-V measurement

$$P_{out} = \frac{(\Delta I \times \Delta V)}{8} \tag{2.1}$$

Although none of these DC figures of merit are particularly high for GaN devices, they are comparable with other work [10], and are better than comparable power amplifier devices made from other material systems such as Si or GaAs [2]. Additionally, the devices are not optimized for any particular function, although optimisation of each parameter is possible. For example, V_{BD} can be optimized by various methods including the addition of field plates on top of the passivation layer, across the gate-drain region of the device [24], the 2DEG concentration can be increased by increasing the Al fraction in the AlGaN layer, or by Si doping, and R_{on} can be optimized by shortening the drain-source length [30], although a tradeoff must be made between on-resistance and breakdown voltage.

2.5.2 Small-Signal Characterisation

After DC characterisation it is usual to perform S-parameter measurements on the DUT in order to determine its frequency performance. S-parameter measurements are relatively simple to perform and are usually measured using a standard VNA. They can also be measured using the MTA-based non-linear measurement system that is described in chapter 3, where the DUT is connected via an on-wafer probe station and the automated software allows fast measurement of all S-parameters. With respect to frequency performance, the main parameters of interest are the short circuit current gain (h_{21}) and cut-off frequency (f_T). The short circuit current gain is determined from the S-parameter measurements according to Equation 2.2, and f_T is defined to be the frequency at which h_{21} becomes equal to one (0dB).

f_{MAX} is another figure of merit that is often quoted for high frequency devices and is defined as the frequency at which the unilateral power gain (U) of the device becomes equal to one (0dB). f_{MAX} is always higher than f_T and is sometimes referred to as the maximum frequency of oscillation because it defines the boundary where the device no longer works as an amplifier and becomes a passive rather than active device. f_T and f_{MAX} are useful figures of merit for high frequency devices, and while the useful gain from a device can only be obtained at much lower frequencies, the higher f_T and f_{MAX} are, the higher is the gain of the device at any lower frequency.

$$h_{21} = \frac{-S_{21}}{(1-S_{11})(1+S_{22})+S_{12}S_{21}} \tag{2.2}$$

The plot of Figure 2.7 shows h_{21} for the $2x100\mu m$ GaN HFET described earlier and was calculated from bias dependent S-parameters measured using the non-linear measurement system. The measurement indicates that this device has an f_T of approximately $25GHz$. f_T and f_{MAX} are largely

controlled by the device gate length, L_g, which for this device is $0.29\mu m$. Although the devices have not been specifically designed for optimum frequency performance, the measured f_T is similar to that reported by others for devices with similar gate length [16].

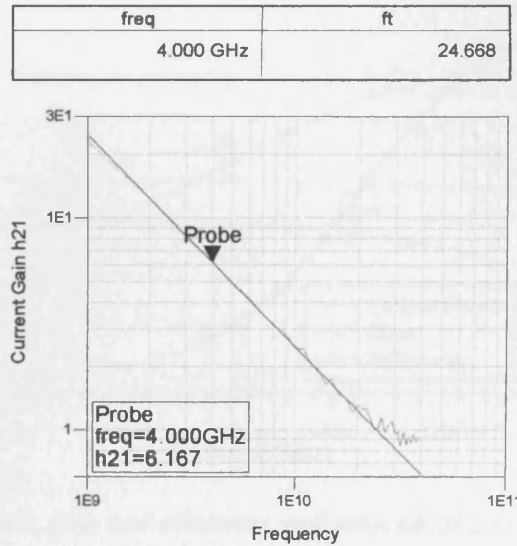


Figure 2.7 – Magnitude of current gain h_{21} versus frequency, and resultant f_T

2.5.3 RF Power Characterisation

After an initial understanding of the DUT is gained from DC and S-parameter analysis, RF power measurements can be performed using standard lab equipment or using the non-linear RF measurement system that is described in chapter 3. Figure 2.8 shows the measured power, efficiency, and gain performance of the $2x100\mu m$ device when a 840MHz CW signal is applied. The device is biased in Class A at a drain voltage of 30V and tuned for optimum power performance.

The theoretical power performance of the device can be determined from the DC I-V of Figure 2.6 using Equation 2.1. If it is assumed that $\Delta I = I_{DSS} = 0.16A$, and since $V_{knee} = 5V$, the voltage swing is 50V. Therefore, when biased at 30V the maximum output power of the device should be: $P_{out} = (\Delta I \times \Delta V)/8 = (0.16 \times 50)/8 = 1W$, or $5W/mm$.

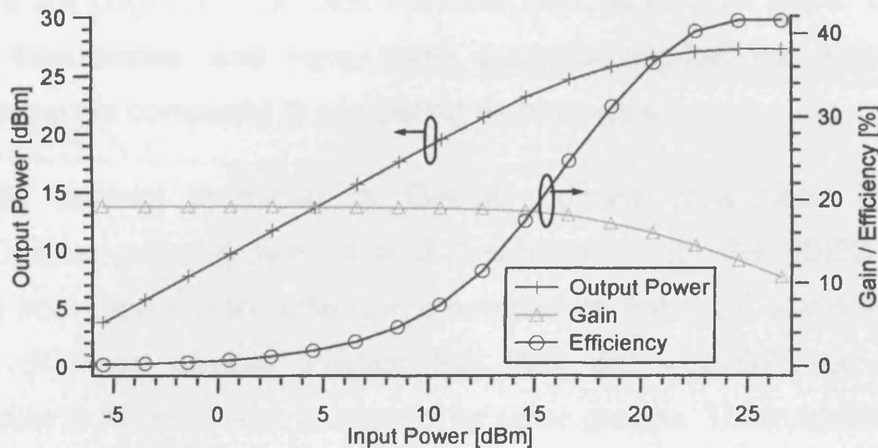


Figure 2.8 – Power, gain and efficiency performance of a 2x100 μm GaN HFET

The measured power sweep of Figure 2.8 shows that even when the device is operated in saturation its output power is only around 0.5W, or 2.5W/mm, which is only 50% of the maximum theoretical power that was expected. It can also be seen that the maximum measured drain efficiency is approximately 42%, and this is also much less than the 50% efficiency that should be achievable from a Class A amplifier. Unexpectedly poor RF performance such as this has been well documented and has been attributed to the problem of DC-RF dispersion, as discussed earlier. At the outset of this work dispersion in GaN HFETs was still not well understood, so dispersion and its effect on the power and efficiency performance of these devices is examined in greater detail in chapter 4.

2.6 Summary

GaN HFETs are a relatively new technology and would not exist but for the technology advancement of GaN optoelectronic devices which have important applications due to their ability to produce UV and blue light. GaN power amplifiers are potentially the best available devices for high power operation at high frequencies, and have many potential applications despite their relative expense compared to competing technologies.

The GaN devices produced by QinetiQ consist of a basic two layer GaN/AlGaN structure grown on a SiC substrate, with the 2DEG channel resulting from piezoelectric effects in the materials. Initial DC and small-signal analysis of these devices indicate that they are well behaved and are comparable to GaN devices produced by other groups. Their relatively large breakdown voltage, current handling ability, and high frequency operation mean that they have the potential to make effective Power Amplifiers. However, initial RF power measurements indicate that dispersion is a significant problem that must be well characterised, understood, and overcome.

2.7 References

[1] H. J. Round, in *Electrical World*, vol. 19, 1907, pp. 309.
[2] M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur, *Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, and SiGe*: John Wiley and Sons, 2000.
[3] S. Nakamura, M. Senoh, N. Iwasa, and S. Nagahama, "High-Brightness InGaN Blue, Green and Yellow Light-Emitting Diodes with Quantum Well Structures," *Japanese Journal of Applied Physics*, vol. 34, pp. 797-799, 1995.
[4] S. Nakamura, "III-V Nitride-Based Blue LDs with Modulation-doped Strained-Layer Superlattices," presented at IEEE Intl. Symp. Compound Semiconductors, 1997.
[5] M. A. Khan, A. Bhattarai, J. N. Kuznia, and D. T. Olson, "High Electron Mobility Transistor Based on a GaN-AlGaN Hetrojunction," *Applied Physics Letters*, vol. 63, pp. 1214, 1993.

[6] E. O. Johnson, "Physical Limitations on Frequency and Power Parameters of Transistors," *IRE International Convention Record*, vol. 13, pp. 27-34, 1965.

[7] M. J. Uren, A. R. Barnes, T. Martin, R. S. Balmer, K. P. Hilton, D. G. Hayes, and M. Kuball, "GaN devices for microwave applications," presented at The 10th IEEE International Symposium on Electron Devices for Microwave and Optoelectronic Applications, 2002, 2002.

[8] A. P. Zhang and L. F. Eastman, "Stability of AlGaN/GaN High-Power HEMTs Under DC and RF Stress," *Electronics Letters*, vol. 40, 2004.

[9] O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, and L. Eastman, "Two-Dimentional Electron Gases Induced by Spontaneous and Pizoelectric Polarization Charges in N- and Ga-faced AlGaN/GaN Hetrostructures," *Journal of Applied Physics*, vol. 85, pp. 3222-3233, 1999.

[10] R. Ventury, "Polarization Induced 2DEG in AlGaN/GaN HEMTs: On the Origin, DC and Transit Characterization," University of California Santa Barbara, 2000.

[11] A. D. Hanser, "An Overview of Gallium Nitride Substrate Materials Developments for Optoelectronic and microelectronic Applications," presented at CS MANTECH Conference, Vancouver, British Columbia, Canada, 2006.

[12] K. K. Chu, "9.4 W/mm Power Density AlGaN-GaN HEMTs on Free-Standing GaN Substrates," *IEEE Electron Device Letters*, vol. 25, 2004.

[13] J. W. P. Hsu, M. J. Manfra, R. J. Molnar, B. Heying, and J. S. Speck, "Direct Imaging of Reverse-Bias Leakage Through Pure Screw Dislocations in GaN Films Grown by Molecular Beam Epitaxy on GaN Templates," *Applied Physics Letters*, vol. 81, pp. 79-81, 2002.

[14] J. W. Johnson, E. L. Piner, and S. Singhal, "12 W/mm AlGaN-GaN HFETs on Silicon Substrates," *IEEE Electron Device Letters*, vol. 25, 2004.

[15] P. Ladbrooke and S. Blight, "Low-Field Low-Frequency Dispersion of Transconductance in GaAs MESFETs with Implications for Other Rate-Dependent Anomalies," *IEEE Transactions on Electron Devices*, vol. 35, pp. 257-267, 1988.

[16] S. C. Binari, "Trapping Effects and Microwave Power Performance in AlGaN/GaN HEMTs," *IEEE Transactions on Electron Devices*, vol. 48, 2001.

[17] H. Takayanagi, H. Nakano, and K. Horio, "Analysis of buffer-trapping effects on current reduction and pulsed I-V curves of GaN FETs," presented at European Gallium Arsenide and Other Semiconductor Application Symposium (EGAAS), 2005.

[18] R. Ventury, N. Q. Zhang, S. Keller, and U. K. Mishra, "The Impact of Surface States on the DC and RF Characteristics of AlGaN/GaN HFETs," *IEEE Trans. Electron Devices*, vol. 48, pp. 560-566, 2001.

[19] R. Coffie, D. Buttari, S. Heikman, A. Chini, L. Shen, and U. K. Mishra, "p-Capped GaN-AlGaN-GaN High-Electron Mobility Transistors (HEMTs)," *IEEE Electron Device Letters*, vol. 23, 2002.

[20] T. Kikkawa, M. Nagahara, T. Kimura, S. Yokokawa, S. Kato, M. Yokoyama, Y. Tateno, K. Horino, Y. Yamaguchi, N. Hara, and K. Joshin, "A 36W CW AlGaN/GaN Power HEMT using Surface-Charge-Controlled Structure," presented at IEEE MTT-S International Microwave Symposium, 2002.

[21] Y. F. Wu, A. Saxler, M. Moore, R. P. Smith, s. sheppard, P. M. Chavarkar, T. Wisleder, U. K. Mishra, and P. Parikh, "30-W/mm GaN HEMTs by Field Plate Optimization," *IEEE Electron Device Letters*, vol. 25, 2004.

[22] A. Koudymov, V. Adivarahan, J. Yang, G. Simin, and M. A. Khan, "Mechanism of Current Collapse Removal in Field-Plated Nitride HFETs," *IEEE Electron Device Letters*, vol. 26, October 2005.

[23] Y. F. Wu, M. Moore, A. Saxler, T. Wisleder, and P. Parikh, "40 W/mm Double Field-Plated GaN HEMTs," presented at 64th Device Research Conference, 2006.

[24] H. Xing, Y. Dora, A. Chini, S. Keller, and U. K. Mishra, "High Breakdown Voltage AlGaN-GaN HEMTs Achieved by Multiple Field Plates," *IEEE Electron Device Letters*, vol. 25, April 2004.

[25] U. K. Mishra, L. Shen, T. E. Kazior, and W. Yi-Feng, *Proceedings of the IEEE*, vol. 96, pp. 287-305.

[26] A. Maekawa, M. Nagahara, T. Yamamoto, and S. Sano, " A 100W high-efficiency GaN HEMT amplifier for S-Band wireless system," presented at European Microwave Conference, October 2005.

[27] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "RF and Microwave Power Amplifier and Transmitter Technologies," in *High Frequency Design*, 2003, pp. 22–36.

[28] D. Fanning, E. Beam, and P. Saunier, "pHEMTs Rise to Wide-Gap Challenge," in *Compound Semiconductor*, September 2007, pp. 25-27.

[29] T. Martin, M. J. Uren, R. S. Balmer, P. McGovern, and P. J. Tasker, "GaN H-FET development at QinetiQ," presented at European Microwave Conference, 2005.

[30] M. Inada, S. Yagi, Y. Yamamoto, G. Piao, M. Shimizu, H. Okumura, K. Arai, Y. Yano, and N. Akutsu, "Low Specific On-Resistance AlGaN/GaN HEMT on Sapphire Substrate," presented at IEEE International Symposium on Power Semiconductor Devices and IC's, 2006.

Chapter 3. Measurement System Overview

The non-linear RF time domain waveform measurement system which has been developed over many years is central to much of the unique work carried out at Cardiff University's Centre For High Frequency Engineering and is described in detail in this section of the thesis.

3.1 Introduction to Non-Linear RF Measurements

Normally, to characterise any electrical device, linear or non-linear, it is desirable to directly measure the voltages and currents at the device ports. At DC and low frequencies these are relatively easy measurements to perform, using high and low impedance probes respectively. However, these measurements become increasingly difficult to perform with increasing frequency, especially across large bandwidths because the high and low impedances can only be maintained across narrow bandwidths. Additionally, active devices may oscillate or self-destruct with the connection of shorts or opens, and the practical limitations of placing the measurement probes at the correct positions must be considered.

It is however relatively easy to maintain a constant 50Ω impedance across a broadband frequency range. This led to the definition of scattering parameters (S-parameters) [1], which treats the propagation of electromagnetic energy through a conducting media in terms of voltage travelling waves instead of terminal voltages. The S-parameters of a network are measured as a ratio of the incident and reflected voltage travelling waves at a given port and once they are known, they can be converted into impedances and admittances using conversion matrices. These S-parameters led to the development of the Vector Network Analyzer (VNA), which allows the automated measurement of both magnitude and phase of all travelling waves at each network port.

However, S-parameters only apply to the measurement of a signal which is suitably small so that there is no harmonic generation and all of its power remains confined to a single frequency. Hence S-parameters are often referred to as small-signal S-parameters, and can be used to fully characterise a linear network, but are of limited use in the characterisation of non-linear networks where, by definition, multiple harmonics and mixing terms are produced. These are the phenomena that must be accurately characterised in order to describe the behaviour of non-linear networks and devices. Also, although the VNA is capable of measuring across a large frequency range while maintaining a broadband 50Ω impedance, it can only measure a single frequency component at a time and is therefore of limited use in the characterisation of non-linear power amplifiers. Some of the most useful measurements that can be performed on power amplifiers using a VNA include frequency sweeps, power sweeps, and AM-PM measurements.

The Spectrum Analyzer is another measurement instrument that is commonly found in RF labs. This instrument is inherently broadband and can simultaneously measure all frequency components in its measurement bandwidth, typically from DC to several GHz, while maintaining a broadband 50Ω impedance environment. The Spectrum Analyzer typically has large measurement bandwidth and is generally used for the measurement of harmonics, mixing products, and spurious signals such as device oscillations. It requires little or no calibration and is a fast and effective evaluation tool. However, the drawback of the Spectrum Analyzer is that it is a scalar instrument only and therefore contains no phase information on the measured frequency components, and is of limited use for more advanced characterisation, and model extraction and verification of non-linear devices.

For complete characterisation of non-linear RF networks it is desirable to be able to perform broadband vector measurements, i.e. to measure all harmonic and mixing terms simultaneously across a broadband frequency range, and to measure the relative phase of each of these frequency components as well as

their absolute magnitudes. One instrument that can perform broadband vector measurements is the high frequency sampling oscilloscope. The obvious solution to the measurement problem is to use this instrument to directly measure the terminal voltages of the non-linear device in a characteristic system impedance of 50Ω.

At low frequencies a non-linear device can be fully characterised using an oscilloscope to measure directly the large-signal time domain voltage waveforms at the device ports, and then determine from these the large-signal current waveforms. Performing a Fast Fourier Transformation (FFT) on the time domain data yields the magnitude and phase of all frequency components present in the bandwidth of the oscilloscope. Also, absolute phase information is available since the phase of each component is measured relative to that of the fundamental frequency, and an inverse FFT on the Fourier coefficients yields the original time-domain waveforms. At higher frequencies this measurement approach is not so easy because of the limited functionality of sampling oscilloscopes. They generally have poor sampling rates, external triggering and only 8-bit analog-to-digital converters (ADCs). This results in poor resolution and dynamic range, and therefore a poor quality representation of the original signal [2]. Hence, there were no commercially available systems that could perform the desired measurement of non-linear devices at RF frequencies.

However, since the late 1980's many research groups have developed custom measurement systems with the aim of characterising non-linear RF devices via the measurement of large-signal time domain waveforms. Some early systems have used VNA samplers as the receiver [3], while others have been oscilloscope-based systems [4]. The introduction of the Hewlett Packard Microwave Transition Analyzer (MTA) [5] saw many new systems based around this instrument because of its unique capabilities and improved performance. The MTA consists of two 40GHz receivers and can operate in the time domain as a sampling oscilloscope or in the frequency domain as a

traditional VNA. In contrast to VNAs or oscilloscopes, the two channels of the MTA are triggered internally from a common time base, enabling the relative phase of the RF signals to be measured very accurately. The MTA also has much improved measurement performance due to its 10-bit ADC resolution, resulting in approximately 60dB dynamic range.

This improved performance led to the development of many experimental systems based around the MTA [6], [7], [8] since the mid 1990's. The system demonstrated by Demmler, Tasker, and Schlechtweg [8] is the basis of the system developed and currently used at Cardiff University. These systems are generally referred to as Non-Linear Vector Network Analyzers (NVNA), Large-Signal Network Analyzers (LSNA), or RF Time Domain Waveform Measurement Systems. Finally, in 2003, Maury Microwave made commercially available a LSNA measurement system [9] based around four MTA receivers, one for the measurement of each travelling wave in a 2-port network.

3.2 Non-Linear Measurement System at Cardiff University

3.2.1 The Microwave Transition Analyzer

Many of the non-linear measurement systems that have been developed, including those at Cardiff University, have been based around the Hewlett Packard 70820A Microwave Transition Analyzer (MTA) [5]. Figure 3.1 shows a simplified block diagram of the MTA. It consists of two identical signal processing channels, each of which samples and digitizes signals from DC to 40 GHz. The signals received on both channels are down-converted to an intermediate frequency (IF). This IF signal is then sampled and digitized using a 10-bit sample-and-hold ADC, and converted to the frequency domain using an FFT. The digital data is then stored in memory, and error corrections and

averaging are also applied in the frequency domain. Finally, an inverse FFT is performed to display the averaged waveform in the time domain.

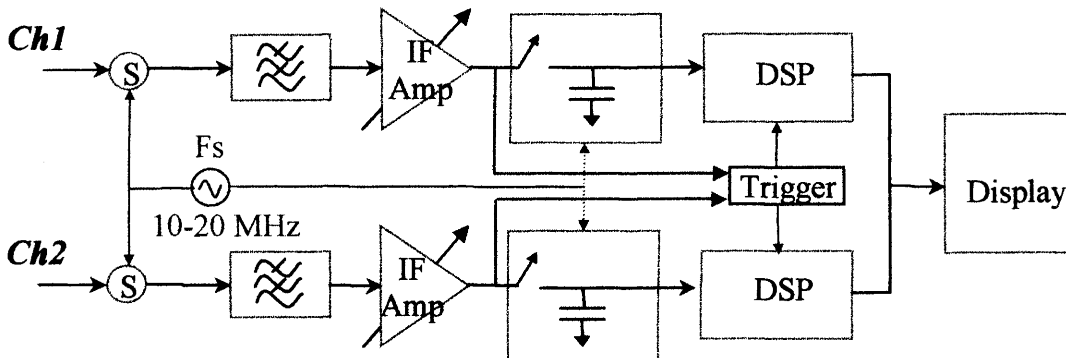


Figure 3.1 – Simplified Block Diagram of MTA

The DC components of the measured signal are separated before reaching the microwave sampler and are re-added after the signal has been sampled. A single local oscillator (LO) which has a frequency range of 10 MHz to 20 MHz drives the two identical microwave samplers and the step recovery diode. The signal produced at the output of the sampler is processed by additional analog circuitry including switchable low-pass filtering to remove sampler feedthrough and improve the signal-to-noise ratio (SNR), and a programmable shaping amplifier to compensate for the sampler IF response roll-off. The signal then passes through the step recovery diode, which runs at the same frequency as the microwave sampler, and is digitized by a 10-bit ADC. The triggering circuitry is also connected to the IF signal path and initiates storage of the data. Once stored, the measured data is converted to the frequency domain using FFT algorithms, and additional error corrections are applied, as determined by a calibration that is performed intermittently on the IF signal path. The stored data is now fully calibrated to the ends of the MTA sampler ports and at this point additional averaging measurements can be performed as required. The averaged data is converted back to the time domain using an inverse FFT algorithm, and is ready for display and download to a PC.

The MTA has become the measurement instrument of choice for use in non-linear measurement systems, because of its superior performance and better functionality compared with VNAs and sampling oscilloscopes. It is a multi-functional instrument and can be used as a sampling oscilloscope in its time domain mode, or as a VNA or Spectrum Analyzer in its frequency sweep mode. Additionally, it can measure absolute powers and so can perform the function of a calibrated power meter. In terms of performance, it is not quite as good as the best available sampling oscilloscope, VNA, Spectrum Analyzer, or power meter, but it is more than adequate for most requirements.

The main advantage of MTA-based systems over VNA-based systems is:

- A VNA must measure one frequency at a time and a very complex measurement setup is required to measure over a wide bandwidth and maintain a good phase reference, whereas the MTA can measure in the time domain and so can easily measure over a wide bandwidth while maintaining an accurate phase reference. Therefore, by definition, an MTA-based or oscilloscope-based system should be a better solution.

The main advantages of MTA-based systems over sampling oscilloscope-based systems are:

- The measurement sensitivity of the MTA is typically much better than that of high frequency sampling oscilloscopes because of its internal trigger and its 10-bit ADC compared to the usual external triggering and 8-bit ADC of most sampling oscilloscopes. This results in approximately 60dB dynamic range.

- The MTA is internally triggered up to 40GHz, with no additional external trigger circuitry required, while most high frequency sampling scopes have limited internally triggered frequency range. The frequency range can be extended using external triggering, but this complicates the measurement

setup, and the frequency range and sensitivity is still less than that of the MTA.

- Both MTA channels are driven by the same LO and the two channels are sampled simultaneously (within 10ps of each other). This allows for accurate phase measurements between the two channels and also allows the MTA to be used as a VNA which greatly simplifies system calibration.

- The MTA enables faster measurements due to the fast sampling rate of between 10MHz–20MHz, compared to a typical value of 2KHz for sampling oscilloscopes. Lower phase jitter also allows for faster measurements as less averaging is required.

- The MTA can also act as a calibrated power meter, whereas most other systems require an external power meter for absolute power measurement and power calibration.

3.2.2 Measurement System Architecture

There are currently two MTA-based measurement systems at Cardiff University. The first of these systems was developed by Tasker et al. [10], and was further refined and automated, resulting in the current system [11], which is largely used for the characterisation of on-wafer low power amplifiers (up to approximately 1W CW), and has a measurement bandwidth of approximately 0.5GHz to 40GHz. The second system [12] was developed to characterise in-fixture high power amplifiers (up to 100W CW), and has a measurement bandwidth of approximately 1GHz to 12GHz. The high power measurement system was not used to perform any measurements described in this thesis, and so only the low power system is described here, although the two systems are in principle the same, except for some of the hardware components. The system is illustrated in Figure 3.2.

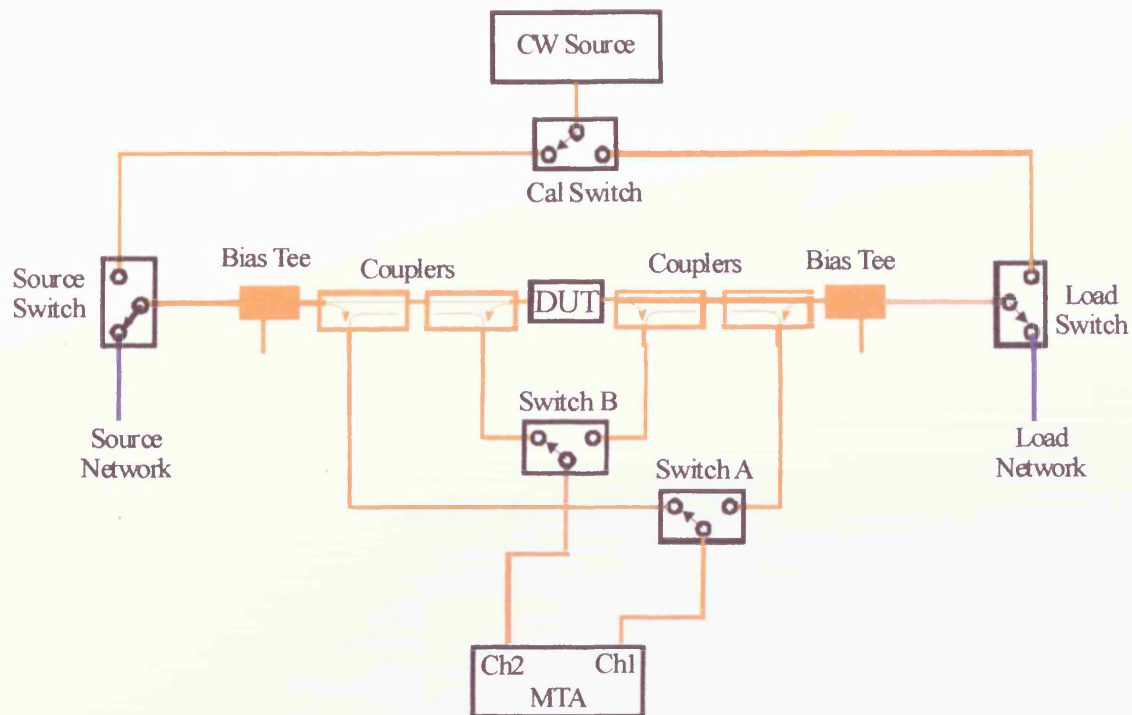


Figure 3.2 – Measurement System Schematic

The overall objective of the measurement system is to measure the error corrected RF voltage and current time domain waveforms that are present at the device plane. To do this, the MTA is operated in the time domain mode and used to measure all four uncorrected a and b voltage travelling waves of the two port network (power amplifier) which are coupled through the broadband couplers. Averaging is also performed by the MTA before the data is downloaded to a PC where the a and b waves are error corrected in the frequency domain via the application of a previously measured S-parameter calibration.

The voltage and current waveforms for each harmonic component are then calculated using Equations 3.1 and 3.2, where Z_0, the characteristic impedance of the system, is 50Ω. They are then transformed back into the time domain via an inverse FFT and added together to produce the complete time domain waveforms. The resulting voltage and current waveforms now

constitute a complete description of the DUT and its measurement environment. The incident and reflected powers at each device port can easily be calculated along with many other parameters such as gain and efficiency. The DC operating points are also measured and from the calibrated incident and reflected travelling waves, all harmonic source and load impedances can be calculated.

$$V_n = \sqrt{Z_0}\left(a_n + b_n\right) \qquad (3.1)$$

$$I_n = \left(a_n - b_n\right)/\sqrt{Z_0} \qquad (3.2)$$

The DUT is probed on-wafer although it may also be connected in-fixture. DC bias is applied via the bias tees and supplied by a computer controlled DC source such as the HP6629 or the more accurate HP4155 semiconductor parameter analyser. The DC cannot pass through the couplers to the MTA so it is recorded separately by the measurement software and later added to the RF waveforms. The CW input signal is applied via the CW source but a different source can be connected via the source switch if a higher power source, or input tuning is required, while the output signal can be terminated into a passive load or tuning network via the load switch.

Since there are four travelling waves to be measured and the MTA has only two channels, the signals must be multiplexed through switches A and B so that the input waves a_1 and b_1 are measured first, and then the output waves a_2 and b_2 are measured. The a waves pass through switch A and are measured on channel 1 of the MTA while the b waves pass through switch B and are measured on channel 2 of the MTA. However, with the measurements performed in this way, there is no phase reference between the input and output signals. To maintain a phase reference, a phase handover measurement must also be performed. After the input waves are measured, switch B is switched to measure the output wave, b_2, while switch

A is unchanged measuring the input wave, a_1. This measurement gives the phase offset between input and output, and after the final measurement of the a_2 and b_2 output waves, the phase offset is applied to the them in the measurement software before the error correction is applied. The whole measurement process is automated by the measurement software which is connected to and controls all of the measurement instruments and switches.

3.2.3 Calibration

All RF measurement systems require calibration to account for the phase delay and signal attenuation that is introduced by the system. S-parameter calibrations as used in VNAs are well established, but because S-parameters are relative measurements, the absolute magnitude of the traveling waves is not determined. For complete large-signal characterisation, the absolute magnitude and phase of the traveling waves must be determined so that the voltage and current waveforms at the device plane can be reconstructed. Therefore, the calibration of this system consists of a small-signal S-parameter calibration, and a large-signal absolute calibration to determine the absolute magnitude and phase of the traveling waves at the RF probe tips. The full calibration is described in detail elsewhere [12], but is described in brief here.

The small-signal S-parameter calibration is a TRM (Thru, Reflect, Match) two port on-wafer calibration. This is a common calibration technique and employs known calibration standards that are supplied on a calibration substrate. During calibration the MTA is used in frequency sweep mode so that it acts like a VNA and the calibration is effectively the same as that used for standard two port VNA measurements. The calibration procedure is fully automated in the measurement software so that it is relatively fast and the operator is needed only to probe the different calibration standards. The resulting error coefficients are stored in a calibration file ready to be applied to the uncorrected measured data.

The second part of the calibration is the absolute calibration which is a more unique measurement and is required to determine the absolute magnitude of the traveling waves by determining the correct scaling factor that must be applied to the S-parameters. This can be achieved by directly measuring the traveling waves at port 1 of the device plane using one of the MTA channels. However, it is not physically possible to directly connect this port to the MTA so the solution is to shift the reference plane at port 1 to the end of a coaxial cable that is connected to the source switch. A short, open, load (SOL) calibration is then performed at the end of the cable while the wafer probes are on a thru line and energy is injected through the port 2 side of the system. When the cable is then connected to the MTA, the absolute magnitude and phase of the system can be calculated and the resulting scaling factor is applied to the S-parameter calibration error coefficients.

3.2.4 Active Loadpull

Passive tuners and loadpull systems have long been used to tune the fundamental and harmonic impedance terminations to optimise device performance. Their operation is relatively simple and is illustrated in Figure 3.3, where the b_2 output wave from the DUT is phase shifted and/or attenuated resulting in a modified a_2 wave which is reflected back to the device. The resulting load impedance presented to the device Γ_L, is determined by Equation 3.3.

Passive loadpull works very well but inevitably losses occur between the output of the device and the tuner. This affects the magnitude of the reflected signal and means that impedances near the outside of the Smith chart, with large reflection coefficients, are not possible. This is a major problem when the optimum fundamental impedance of a DUT is very small, or if an open or short harmonic termination is required. Another problem with older manual passive tuners is that the measurements are very time consuming, although

automated mechanical tuners are now commercially available. These can operate quite fast but are generally even more lossy than the manual versions.

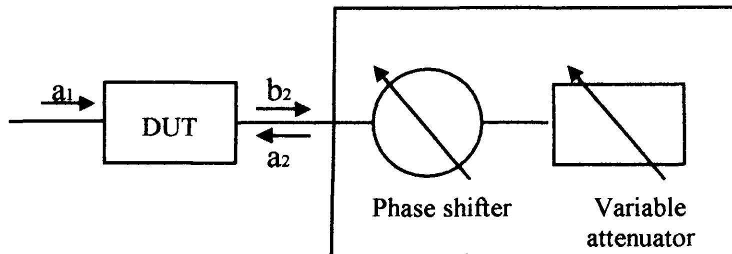


Figure 3.3 – Passive loadpull system

$$\Gamma_L = \frac{b_2}{a_2} \qquad (3.3)$$

The desire to perform loadpull measurements with reflection coefficients close to the edge of the Smith chart led to the development of active loadpull systems, whereby extra energy is injected into the system along with or in place of the reflected energy from the DUT. Many different active loadpull architectures have been developed but the type used with this measurement system is open-loop active loadpull, and the basic operation is illustrated in Figure 3.4.

The basic premise is that the b_2 wave, which contains all output energy from the device, passes through the circulator and is terminated into a 50Ω load. The device is therefore presented with a load impedance of 50Ω until the Agilent Electronic Signal Generator (ESG) is switched on. The ESG is a very phase-stable frequency source and the signal that it produces passes back through the circulator and effectively creates an 'artificial' a_2 wave which in turn changes the load impedance that is presented to the DUT. The ESG is controlled by the measurement software and when a desired load impedance is set, the software varies the magnitude and phase of the ESG signal and

then measures the new load in an iterative process until the desired load is achieved.

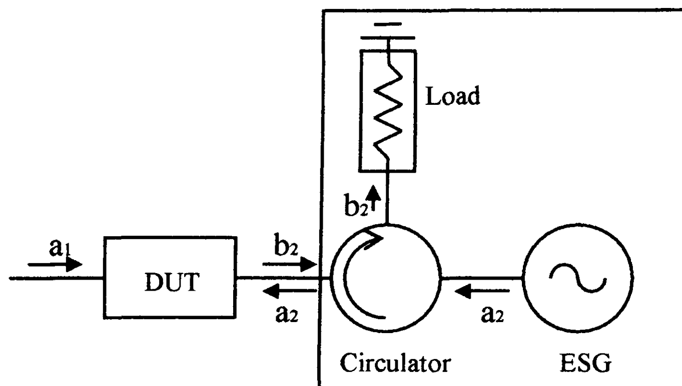


Figure 3.4 – Open-loop active loadpull system

This open-loop loadpull architecture is unconditionally stable unlike some other active loadpull architectures, but one disadvantage of active loadpull is that when the input power of the device changes, the magnitude of the b_2 wave also changes, but the magnitude and phase of the a_2 wave remains unchanged. Therefore the load presented to the DUT also changes and the iterative loadpull process must be repeated to recreate the desired load impedance. In contrast, when passive loadpull is used to set the load impedance it is constant and independent of the output power of the device.

The active loadpull system at Cardiff University is set up to allow the simultaneous loadpull of up to three harmonics as illustrated in Figure 3.5. A triplexer is attached to the load switch so that three isolators and three ESGs can be attached. The triplexer and circulators are narrowband components and here they are selected to work at a fundamental frequency of approximately 840MHz and the corresponding harmonic frequencies of 1.68GHz and 2.52GHz. An identical loadpull setup is also available on the input side of the system so that harmonic sourcepull can be performed.

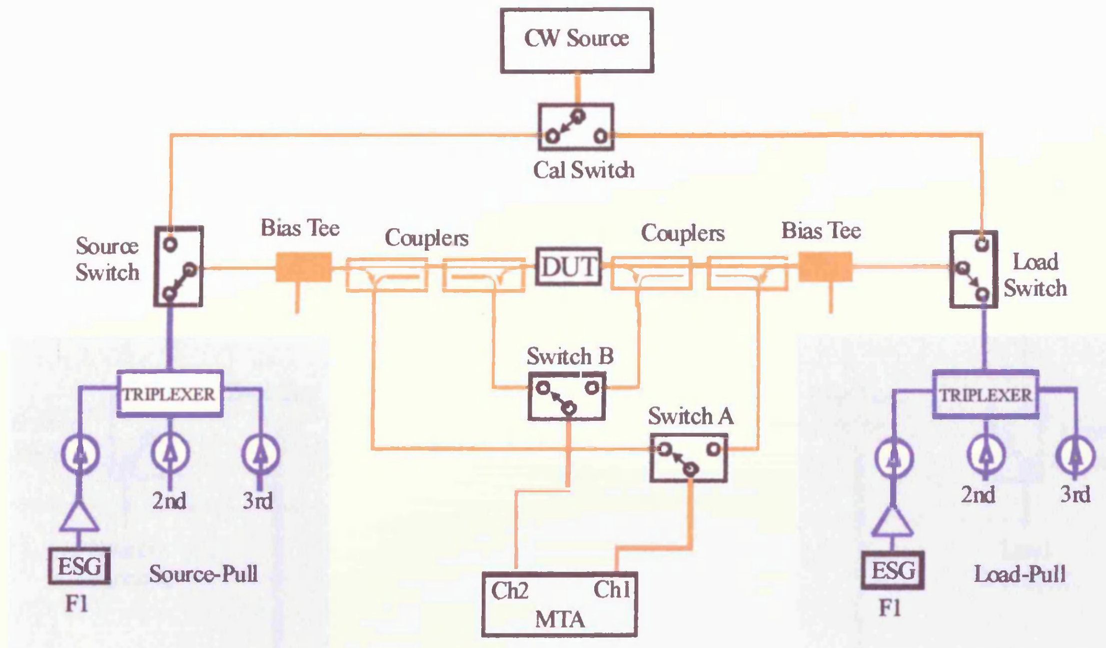


Figure 3.5 – Measurement system with active loadpull

3.2.5 Two-Tone Measurement System

The measurement capabilities of the system were further developed to enable two-tone measurements to be performed and measured [11]. The MTA is operated in its standard sampling mode and the IF difference frequency, $\omega_2 - \omega_1$, is treated as the fundamental signal. Therefore the MTA measures on a frequency grid that is multiples of the two-tone tone spacing, $\omega_2 - \omega_1$, and this allows the measurement of all signals that are created by the two-tone stimulus.

Unfortunately, the minimum sampling frequency of the MTA is 10MHz and therefore the minimum tone spacing of the two-tone signal is also 10MHz. Since the maximum number of measurement points is 512, the maximum measurable frequency is 5.12GHz. While a tone spacing of 10MHz is sufficient to allow detailed two-tone characterisation, ideally the tone spacing could be set much lower than this so that the bandwidth of the two-tone signal

is closer to that of a real modulated signal, allowing linearity and memory effects to be examined in more detail.

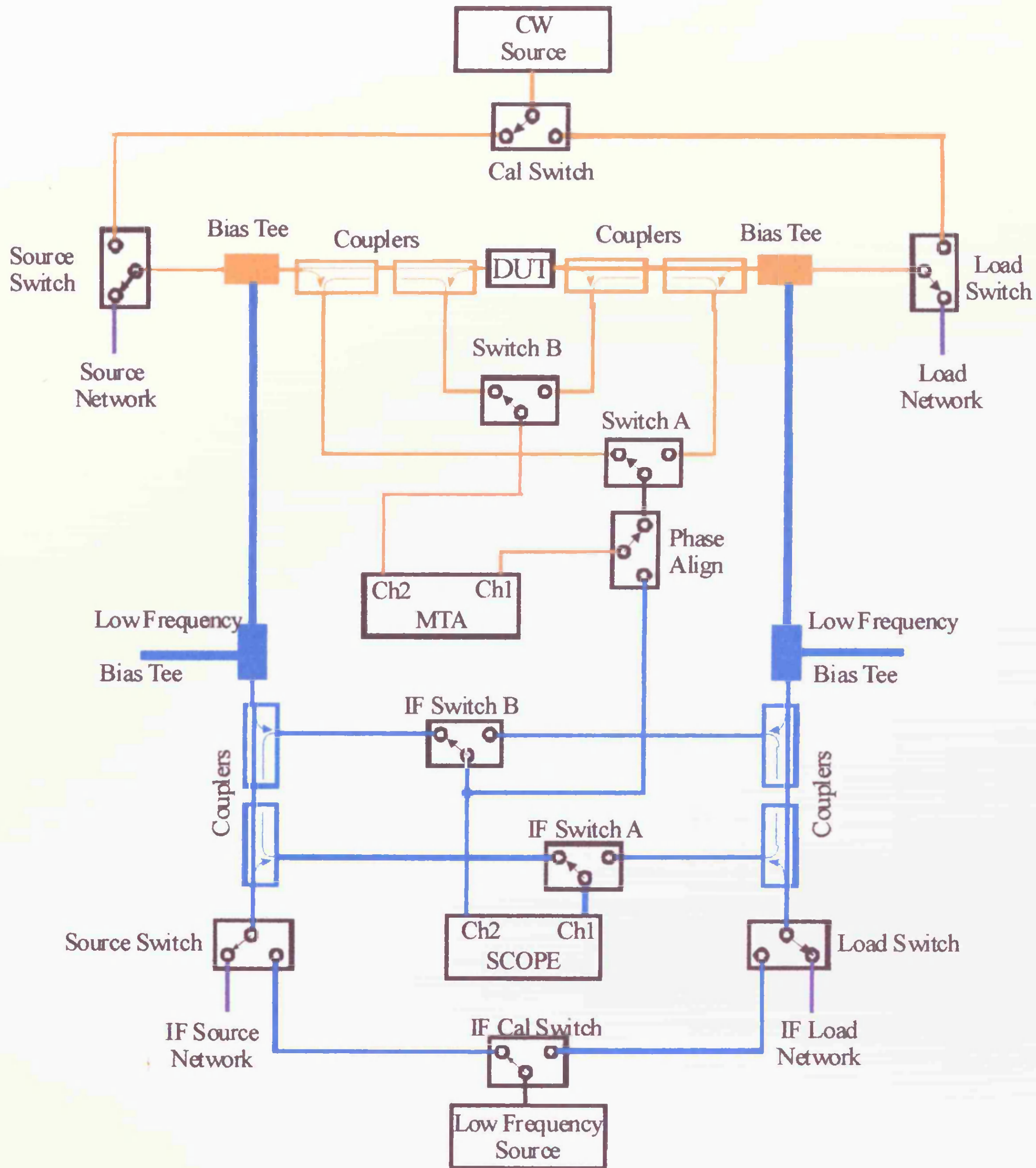


Figure 3.6 – Two-tone measurement system

The two-tone measurement system architecture is shown in Figure 3.6. The high frequency signals are measured by the MTA in the normal way, but

because the couplers of the CW system have a low frequency cutoff of approximately 500MHz, the lower frequency components pass through the coupler along with the DC signal and cannot be measured by the MTA. All couplers work in the same way so this cannot be avoided. Therefore, a new measurement approach is required to measure the low frequency signals. This is achieved by adding a new IF testset to the system which is capable of measuring the IF signal components. The IF and RF signal components are then added together in the measurement software, resulting in the complete two-tone signal including DC, IF, and RF frequency components.

The lower cutoff frequencies of the RF couplers and bias tees are approximately 500MHz and 50MHz respectively. Therefore any signal of lower frequency will pass through these components on the DC line. The DC is separated from this IF signal by the low frequency bias tees and then measured in the normal way. The IF signal is coupled by low frequency couplers and the traveling waves are measured by an oscilloscope. A normal oscilloscope can be used since the frequencies are below 50MHz. The operation of the IF testset is identical to that of the MTA RF system and is also fully automated by the measurement software. Loadpull of the IF frequency components is also possible via the IF load network switch, and calibration of the IF system is also almost identical to that of the RF system.

3.3 Summary

The non-linear measurement system which has been developed over many years at Cardiff University is a unique and powerful solution for the characterisation of non-linear devices because it is capable of measuring fully error corrected time domain voltage and current waveforms at the device plane. The reflection coefficients present in the system are also measured so that the load and source impedances presented to the device are also known, and the open-loop active loadpull system enables complete control of the impedance environment. The CW system can measure RF signals of up to

1W from approximately 0.5GHz to 40GHz while the two-tone system can measure two-tone signals with tones spacings as low as 10MHz.

3.4 References

[1] D. M. Pozar, *Microwave Engineering*: John Wiley & Sons, Inc., 1998.
[2] M. Golio, *The RF and Microwave Handbook*: CRC, 2000.
[3] U. Lott, "A Method for Measuring Magnitude and Phase of Harmonics Generated in Nonlinear Microwave Two-Ports," *MTT-S International Microwave Symposium Digest*, 1988.
[4] M. Sipila, K. Lehtinen, and V. Porra, "High-Frequency Periodic Time-Domain Waveform Measurement System," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, pp. 1397-1402, 1988.
[5] "The Microwave Transition Analyser: Measure 25ps Transitions in Switched and Pulsed Microwave Components," in *Product Note: 70820-2*: Hewlett Packard, 1991.
[6] F. v. Raay and G. Kompa, "A 40GHz large-Signal Double-Reflectometer Waveform Measurement System Designed for Laod-Pull Applications," *26th European Microwave Conference*, pp. 657-661, 1996.
[7] J. Verspecht, P. Debie, A. Barel, and L. Martens, "Accurate On Waver Measurement Of Phase And Amplitude Of The Spectral Components Of Incident And Scattered Voltage Waves At The Signal Ports Of A Non-linear Microwave Device," presented at IEEE-S International Microwave Symposium, 1995.
[8] M. Demmler, P. J. Tasker, and M. Schlechtweg, "A Vector Corrected High Power On-Wafer Measurement System with a frequency Range for the Higher Harmonics up to 40 GHz," presented at 24th EuMC, 1994.
[9] "Large-Signal Network Analyzer, Bringing Reality to Waveform Engineering," in *Technical Dat Sheet 4T-090*: Maury Microwave, 2003.
[10] P. J. Tasker, S. S. O'Keefe, and M. Demmler, "Vector Corrected Non-Linear Transistor Characterisation," *European Gallium Arsenide and Related III-V Compounds Applications Symposium*, 1997.
[11] D. J. Williams, "Non-Linear Measurement System and Techniques for RF Power Amplifier Design." PhD Thesis: Cardiff University, 2003.
[12] J. Benedikt, R. Gaddi, P. J. Tasker, M. Goss, and M. Zadeh, "High power time domain measurement system with active harmonic load-pull for high efficiency base station amplifier design," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 2617 –2624, 2000.

Chapter 4. Effects of Dispersion on Power and Efficiency Performance

The initial DC and S-parameter investigations of the QinetiQ GaN devices in Chapter 2 showed that they have the potential to make good power amplifiers. However, the RF power measurements indicate that they suffer from the well documented problem of dispersion. The aim of this chapter is to investigate the effects of dispersion on the power and efficiency performance of the devices.

4.1 Introduction to Dispersion

In the context of semiconductor power amplifiers, dispersion is a frequency dependent phenomenon which is used to indicate that the dynamic (RF) characteristics of a device are different from its static (DC) characteristics, and is often called DC-RF dispersion. This means that the measured DC I-V characteristics of a dispersive semiconductor device may be completely different from the measured large-signal RF I-V characteristics along which the device will work at RF and microwave frequencies. A non-dispersive semiconductor device will therefore have exactly the same DC I-V, and RF I-V characteristics.

The manifestation of DC-RF dispersion can be best illustrated by considering Figure 4.1 which shows a typical DC I-V characteristic for a FET or HFET power amplifier. It illustrates how the large-signal RF parameters of a power amplifier may be obtained from simple DC measurements using classical loadline theory. According to this simple theory, the maximum output power of a device is given by Equation 4.1 and its optimum load impedance is given by Equation 4.2. If these measured DC characteristics do not accurately reflect the RF characteristics of a given device, its maximum output power and optimum load impedance may be very different from the predicted values.

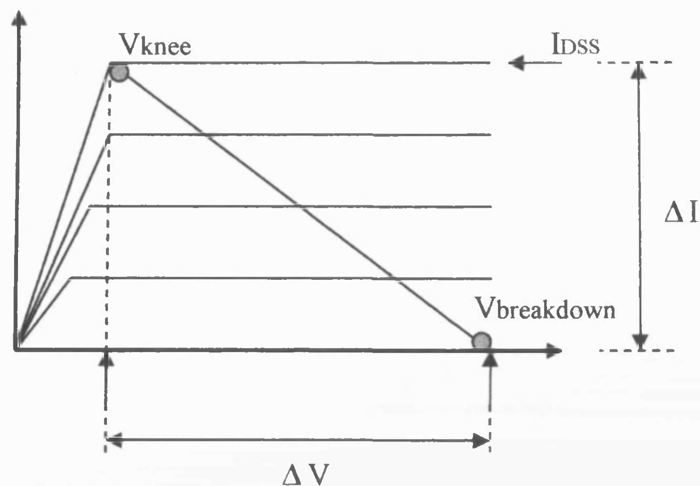


Figure 4.1 – Typical DC I-V Characteristics of a Power Amplifier

$$P_{out} = \frac{\Delta I \times \Delta V}{8} \quad (4.1)$$

$$R_{opt} = \frac{\Delta V}{\Delta I} \quad (4.2)$$

This is just a simple but important example of the problem of dispersion, since DC I-V measurements are perhaps the most basic of all measurements that can be performed on a power amplifier device, and are often used as the basis of large-signal models in PA design. In some semiconductor devices this dispersion can be significant enough to invalidate the DC data for the purpose of device modeling and circuit design from classical loadline theory. Dispersion is a well documented problem in semiconductor devices, and it is important to note that all semiconductor transistor devices are subject to dispersion to some extent, although in many cases it is insignificant enough to be neglected.

4.2 Origins of Dispersion

It is generally acknowledged [1], [2], [3], that dispersion in semiconductor devices has two distinct physical origins:

1. Electrical self-heating of the device, when DC power is dissipated during normal operation.

2. Changes in charge held in 'traps':

   - Deep levels – charge carriers (electrons or holes) move into or out of sites, located in the bulk of the semiconductor, where they become trapped for times of typically microseconds to milliseconds. These sites are called deep level traps.

   - Surface interface states – charge carriers become trapped in sites similar to deep levels, but located on or close to the semiconductor surface. These sites are called surface states, or surface traps, and the charge which is trapped in them is called surface charge.

4.2.1 Self-Heating in HFETs

Dispersion caused by self-heating occurs in all semiconductor devices due to the dissipation of electrical energy during operation, and is a significant problem in PA devices, especially high power amplifiers such as those used in base stations, where the devices may operate at only 20-30% efficiency [4]. Under DC excitation the electrical energy dissipated is the product of the drain voltage and current. This electrical energy causes a temperature rise in the device, which is a function of many variables including the physical properties of the device material, device geometry, and thermal resistance of the device and its packaging [5]. The device will heat up until a stable temperature is reached when the rate of heat energy removed equals the rate of electrical energy dissipated.

When an external electric field is applied to a device, the current carrying electrons in the device acquire energy from it. Over time, energy is transferred from these electrons to the crystal lattice of the device, making the device itself hot. Electron current flow is impeded as the temperature of the device

rises because of greater electron interaction with the lattice due to greater lattice vibrations. Therefore, the current carrying capability of the device is inversely proportional to temperature, and current gain will decrease with increasing temperature. Figure 4.2 illustrates how temperature and self-heating can affect the DC I-V characteristics of a FET or HFET device. When the device is hot the dashed I-V curves will be produced but when the device is cooler, the solid I-V curves will be produced.

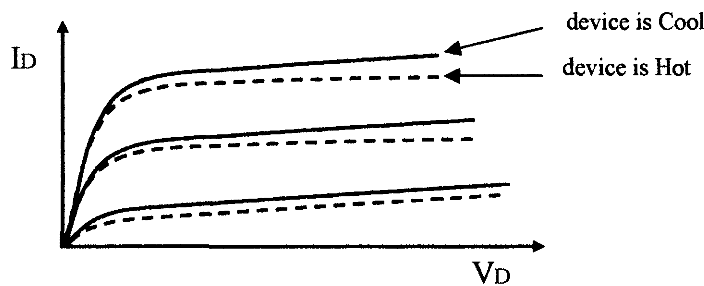


Figure 4.2 – Effect of self-heating on a devices DC I-V

This problem occurs in all power amplifier devices and means that their performance is significantly temperature dependent. Therefore, any large-signal models based on these measurements may be invalid because the temperature of the device may be significantly different during RF operation than during DC measurement. This is because of the concept of 'RF cooling' whereby the device operates more efficiently during RF operation since some of the DC electrical power is converted into RF output power instead of heat in the device. Since less DC power is dissipated in the device during RF operation, it is also cooler. This dispersion due to self-heating can be analysed using a hot chuck to control the temperature of the DUT during operation, or by measuring I-V curves with fast sweeps of drain voltage and then with slow sweeps, assuming that little or no self-heating occurs during very fast sweeps and that slow sweeps are slow enough for heating to occur.

4.2.2 Trapping Effects in HFETs

In semiconductor transistors, impurity atoms provide the electrons and holes that are necessary for device operation. Donor impurities give up electrons to the conduction band and acceptor impurities capture electrons and create holes in the valence band. Energy band diagrams such as that shown in Figure 4.3 are used to illustrate the relative energies of the conduction and valence bands. As illustrated, the energy level of impurity atoms which give rise to donor electrons is generally just below the energy of the conduction band edge, and the energy level of atoms which give rise to acceptors is just above the valence band edge.

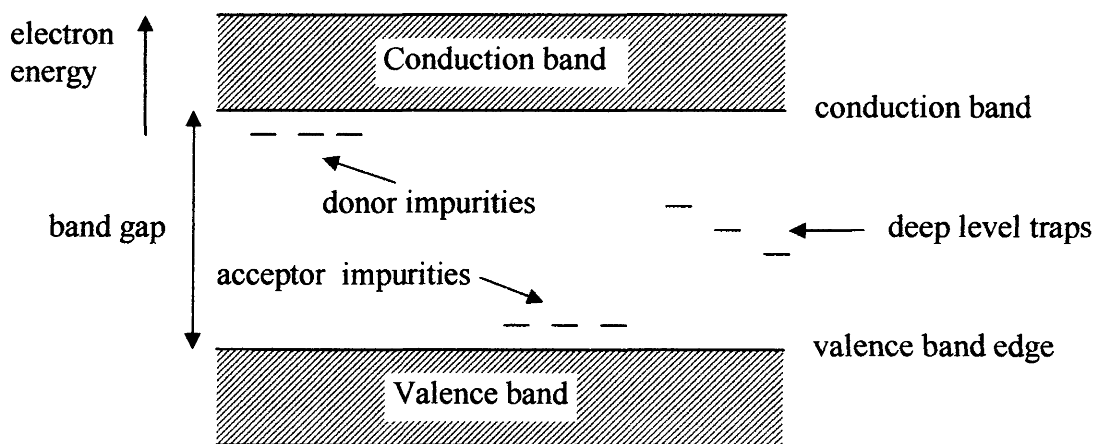


Figure 4.3 – Energy band diagram with traps

If however, the impurity atom which provides the mobile electrons and holes is much further removed from the band edges in energy terms, it may still give up an electron to the conduction band, or accept an electron from the valence band, but the time it takes for this to occur is much longer than normal, typically between microseconds and milliseconds [1].

These deep levels trap and hold electrons for relatively long periods of time because an electron can only move out of an impurity site when it is dislodged by a phonon (lattice vibration). However, phonons do not all contain the same energy, and the probability of finding a phonon of sufficient energy to excite an

electron held in a deep level is much less than the probability of finding one with sufficient energy to excite an electron held in a site of lower energy. Therefore, on average it will take much longer to excite and move an electron held in a deep level, than one held in a site close to the valence band edge.

Unfortunately, this also means that devices based on wider bandgap materials such as Gallium Nitride (Eg=3.4 eV) and Silicon Carbide (Eg=2.9 eV) are likely to contain a much higher density of deep levels and trapping related dispersion problems than narrower bandgap devices such as Silicon (Eg=1.1 eV) or Gallium Arsenide (Eg=1.4 eV).

Therefore, the explanation for DC-RF dispersion is that under DC and low frequency excitation, the current carrying electrons have sufficient time to move into and out of these deep level trapping sites and the device performance is not affected by them. However, at RF frequencies, the time required for electrons to be emitted from the deep level sites is very long compared with the period of the RF cycle, and so the electrons appear to be 'trapped' in the deep level impurity site. Thus, there are fewer electrons available for current transport in the device, which affects device performance.

In mature device technologies, the concentration of bulk deep level traps can often be reduced to such low concentrations that they have negligible effect on device performance. Instead it is the deep levels located at the device surface which are often the major cause of dispersion, as well as self-heating effects which are impossible to remove. It is also worth noting that in general, bipolar devices are affected much less by DC-RF dispersion originating from surface states than FETs and HFETs. This is simply because bipolar devices are vertical structures, where the direction of current flow through the device is perpendicular to the surface, and FETs and HFETs are lateral structures where the current flows parallel to the surface, so that the electrons are likely to be exposed to more surface trapping sites.

4.2.3 Device State

The problem of DC-RF dispersion is inextricably linked to the 'state' of the device. As discussed, dispersion is caused by trapping effects and self heating effects in the device, but it is important to note that the occupied trap density and location, and lattice temperature of the device are determined by the DC bias point, (V_{DS}, V_{GS}) [5], [6]. This pre-set condition is often referred to as the instantaneous 'state' of the device. More generally, the device state refers to the totality of physical variables which together determine how much current the device will pass. This includes the variables of trapped charge held in bulk deep levels or surface states, and the lattice temperature, as well as more obvious variables such as depletion region size and shape, and electron drift velocity in the various layers of the device.

It is also important to note that the self-heating and lattice temperature of a device are largely inseparable. For example, the rate at which electrons are emitted from a deep level is dependent on the lattice temperature, which in turn is set by self-heating. Self-heating can also be affected by an applied RF large-signal since it can affect the efficiency of the device.

4.2.4 Manifestation of Dispersion

The overall manifestation of dispersion is illustrated in Figure 4.5, which is a frequency domain current response of a typical device from DC to cut-off, (f_T). The plot consists of three distinct phases, and would be created by biasing the device at a steady state and applying a small-signal. During phase 1 between DC and low frequencies, typically a few KHz, the measured drain current remains constant. This is the static or quasi-static condition. Above a few KHz, the current starts to fall until it flattens off, typically at a few MHz [1]. This is the dispersion transition of phase 2. Finally, above this frequency, the output current response remains constant (phase 3), until the cut-off frequency of the device is approached.

This plot would be typical of that created by a device which suffers from trapping related dispersion. At DC and low frequencies (phase 1), the current carrying electrons have time to move into and out of the trap sites without performance degradation, while at the higher frequencies of phase 3, the rate of electron emission from the traps (the trap time constant) is much larger than the period of the AC cycle. Therefore, the electrons are trapped and no longer available for current transport, and the output current performance is reduced. In a dispersive device there may be many different trap locations and trap time constants. This causes the dispersion transition of phase 2 where some traps are active while others are not, dependent on the period of the AC cycle relative to the trap time constant.

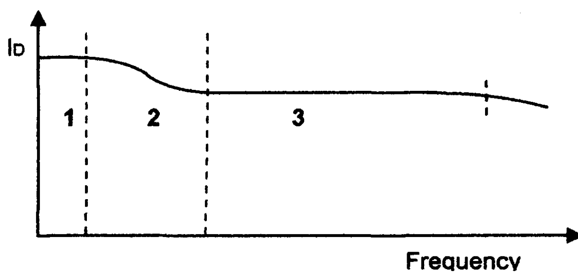


Figure 4.5 – Frequency domain representation of DC-RF dispersion

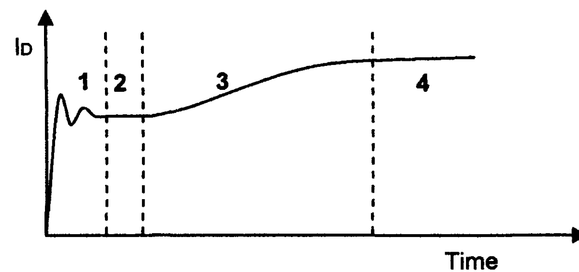


Figure 4.6 – Time domain current transient representation of DC-RF dispersion

This frequency domain representation of dispersion can also be translated into the time domain current transient representation of Figure 4.6. This plot consists of four distinct sections and is typical of a plot created by performing a current transient measurement on the device using a pulsed I-V measurement system. Phase 1 is the turn-on transient, and includes some ringing. The next three phases correspond to the three phases of the frequency domain plot of Figure 4.5. Phase 2 is a current plateau which is reached after the turn-on transient, but before the slow dispersion transients of phase 3 begin. In phase 3 the output current increases steadily until it reaches the steady-state condition of phase 4.

During phase 2 the output current is below its maximum because all trap sites are inactive and no electrons can move through them since the duration of the measurement is shorter than the trap time constants. During phase 3 the duration of the measurement becomes equal to the trap time constants and more trap sites become active, until phase 4 is reached and they are all active and current flow reaches saturation.

4.3 RF Waveform Analysis of Dispersion in GaN HFETs

The device under test is a basic *2x100 μm* GaN HFET on a SiC substrate as described in chapter 2. The device, shown in Figure 4.7, is grown by MOVPE and consists of 30nm AlGaN, and a *1.2 μm* insulating GaN on Vanadium doped SiC, with no intentional doping included in the structure. A Silicon Nitride passivation layer has been added in an attempt to reduce current slump and dispersion. DC characterisation of the device shows that it has a breakdown drain voltage in excess of 80V, and pinches-off at approximately V_G=-7V. The DC I-V characteristic is shown in Figure 4.8.

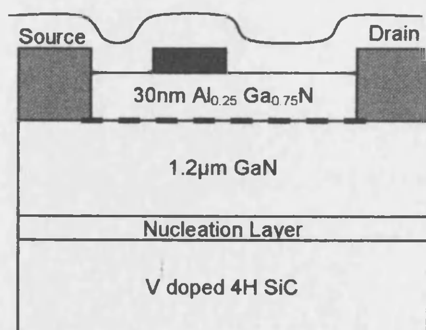


Figure 4.7 – 2x100 μm GaN HEMT

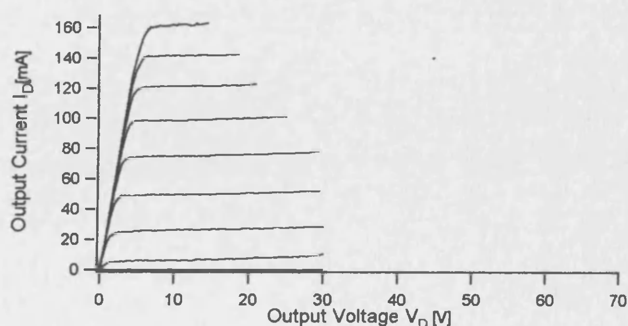


Figure 4.8 – DC I-V Characteristic of DUT

If an assumption is made from this DC I-V measurement that the 'knee' of the device is at approximately 5V and 150mA, its maximum theoretical output power can be calculated based on Equation 4.1. Figure 4.9 shows the

theoretical Class A loadlines at various drain voltages while Figure 4.10 shows a plot of the resulting normalised output power as a function of drain voltage bias, and indicates that a maximum output power of approximately 7W/mm should be attainable at a 40V drain bias.

However, simple Class A power measurements at 840MHz show that the real maximum output power is much less than predicted. This is also illustrated in Figure 4.10. To ensure that maximum output power is achieved, the device is biased so that the DC current is equal to $I_{max}/2$, as shown in Figure in 4.9, and active loadpull is employed so that the fundamental frequency component is presented with an appropriate impedance to achieve the theoretically predicted maximum voltage and current swing, and therefore achieve maximum power performance.

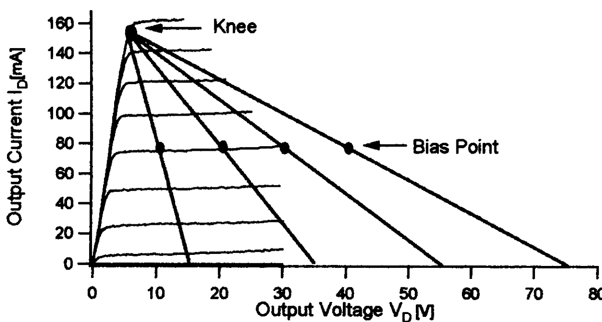


Figure 4.9 – DC I-V characteristic of the DUT and theoretical Class A loadlines

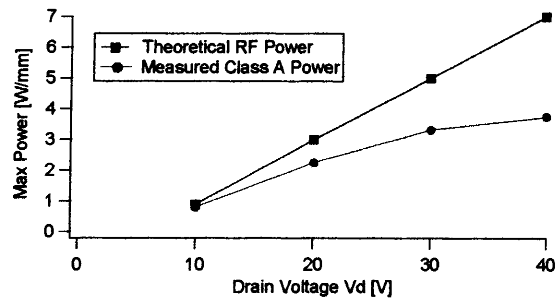


Figure 4.10 – Maximum theoretical and measured output power as a function of Vd bias point

4.3.1 Waveform Measurements at Class A Bias Points

The simple power measurements described above are performed using the non-linear RF measurement system described in Chapter 3, but the system can also measure the RF current and voltage waveforms present at the device plane and this allows the actual RF I-V boundaries of the device to be examined under real operating conditions.

Figure 4.11 shows the DC I-V characteristic of the DUT along with the measured RF dynamic loadlines that correspond to the power measurements of Figure 4.10. The DC bias points confirm that the device is biased in Class A during the measurements and the RF loadlines, which are created by plotting the output RF current waveform versus the output RF voltage waveform, confirm that the DC I-V boundaries of the device are not being reached by the RF waveforms during the power measurements. It should be noted that these measurements are achieved through the use of fundamental loadpull, but no harmonic loadpull is employed, so that the harmonics remain terminated into a broadband 50Ω load.

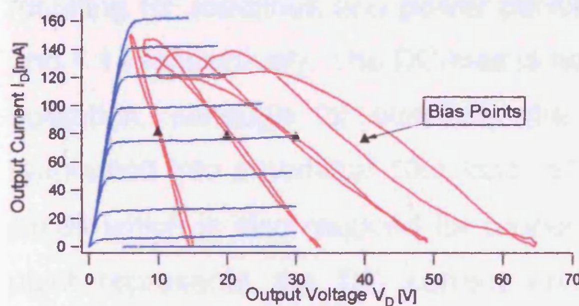


Figure 4.11 – RF dynamic loadlines illustrating the effects of dispersion

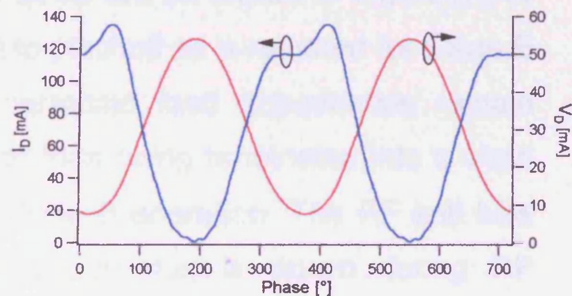


Figure 4.12 – Compressed RF output current and voltage waveforms

One of the most useful aspects of these waveform measurements is that the dynamic loadlines can be used to quickly confirm that the correct impedance has been presented to the device to achieve optimum performance. In this case it can be seen that the selected fundamental impedances are appropriate to allow penetration of the knee region of the device, where the knee is expected to be at approximately 5V and 150mA. This should result in maximum voltage and current swing and hence maximum output power performance. By viewing the output voltage and current waveforms individually it can be seen that the device is indeed operating in current saturation and no more power can be derived from the device. As an example, the waveforms of the measurement at 30V drain bias are shown in Figure 4.12.

These measurements clearly show the effects of dispersion since the RF I-V boundaries of the device are different to the DC I-V boundaries. It is also apparent that the dispersion is bias dependent such that the degree of dispersion worsens with increasing drain voltage, and operation at relatively low drain voltages results in little or no dispersion. This results in the discrepancy between theoretical and measured output power increasing with increasing drain voltage bias, as illustrated in Figure 4.10.

4.3.2 Waveform Measurements at Class B Bias Points

When the experiment above is repeated using a Class B bias point, the resulting RF loadlines and power performance are as shown in Figures 4.13 and 4.14 respectively. The DC bias is set to pinchoff as is required for Class B operation, although for simplicity the harmonic load impedances remain terminated into a nominal 50Ω load rather than being terminated into a short circuit which is also required for proper Class B operation. The RF self bias point represents the DC current and voltage that is drawn during RF operation.

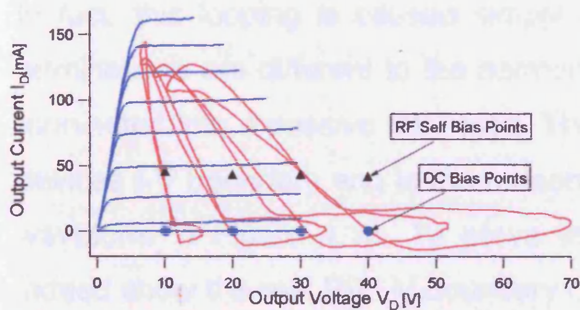


Figure 4.13 – Class B RF dynamic loadlines illustrating the effects of dispersion

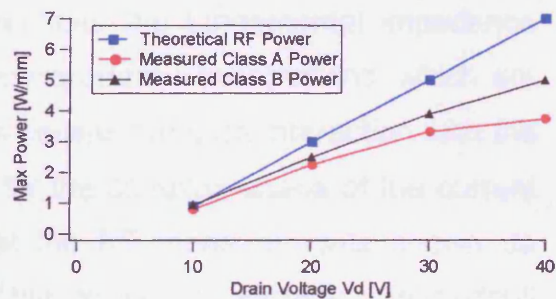


Figure 4.14 – Theoretical and measured output power in Class A and Class B

It can readily be seen that when biased at pinchoff, the maximum output power of the device is significantly improved, especially when biased at higher drain voltages. The waveform measurements indicate that the reason for this improved power performance is that the degree of dispersion is reduced when

the device is biased at pinchoff, which results in increased voltage and current swing compared to Class A operation. This results in RF power performance of approximately 5W/mm when operated at 40V in a Class B bias, compared with approximately 3.5W/mm when operated in Class A. If the device was non-dispersive, the power performance should be the same in both modes of operation.

4.3.3 The Importance of Harmonic Loadpull

The previous measurements indicate that the device is dispersive and that a new artificial I-V boundary is created for each measurement, depending on the bias point chosen. For example, the new boundary for the Class A measurement at 30V may look something like that shown in Figure 4.15. However, the looping at the top of the RF waveform is a cause of confusion. It is possible that the actual RF boundary of the device is different to that shown here, and that the difference in relative impedance between the fundamental and harmonic RF components is causing some sort of artificial boundary to be presented to the device.

In fact, this looping is caused simply because the fundamental impedance terminations are different to the harmonic impedance terminations, which are terminated into a passive 50Ω load. This causes complex interaction with the devices I-V boundary and is the reason for the complex shape of the current waveform in Figure 4.12. To prove that the RF measurements shown do indeed show the real RF I-V boundary of the device it is necessary to loadpull the harmonic impedances so that they are the same as that presented to the fundamental. Since the device is biased in class A, there is relatively little harmonic content above the third harmonic and it is therefore sufficient to loadpull only the second and third harmonics. This loadpull is performed using the active loadpull part of the measurement system, and illustrates the importance of harmonic loadpull in RF waveform measurements.

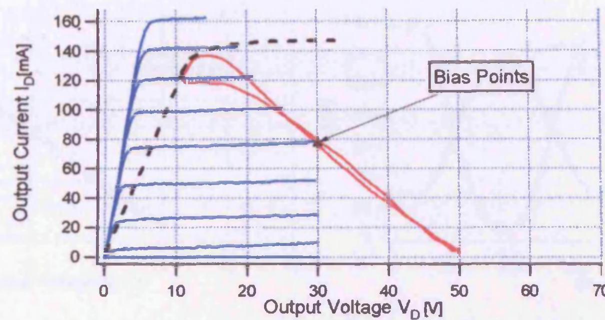


Figure 4.15 – Theoretical RF I-V boundary of the device when biased in Class A at 30V

Figure 4.15 shows the RF dynamic loadline for a Class A measurement at 30V, and Figure 4.12 shows the corresponding output RF current waveform. This measurement was made with the fundamental RF component loadpulled into a real impedance of approximately 100Ω and all other harmonics terminated into a passive 50Ω load. It can be seen that there is a large dip in the output current waveform which may lead to some confusion over the correct value of the saturated current.

If the measurement is repeated but with the fundamental, second, and third harmonics all loadpulled into the same load impedance of 100Ω, the resulting RF loadline becomes a straight line, as shown in Figure 4.16, where it is superimposed on the previously looping measurement when the impedances were different. Figure 4.17 shows the corresponding output current and voltage waveforms of the device before and after the harmonics are loadpulled to the same impedances. It can be seen that after loadpull, the total voltage and current swing are reduced slightly and the waveforms are more compressed and regular.

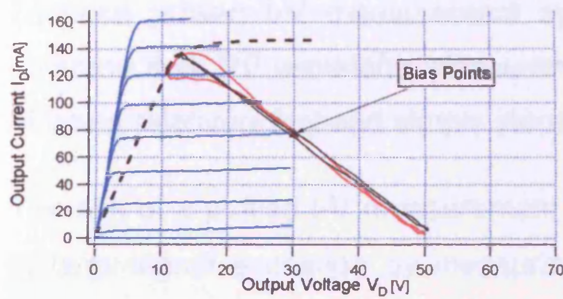


Figure 4.16 – RF dynamic loadline created when the fundamental and harmonic impedances are the same

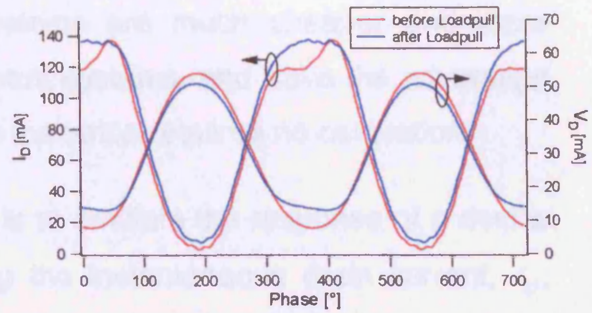


Figure 4.17 – Corresponding saturated output current and voltage waveforms

This result shows that the looping loadline waveforms do indeed show the actual RF boundary of the device, even though the harmonics are terminated into different impedances to the fundamental component. The complete set of repeated measurements is shown in Figure 4.18.

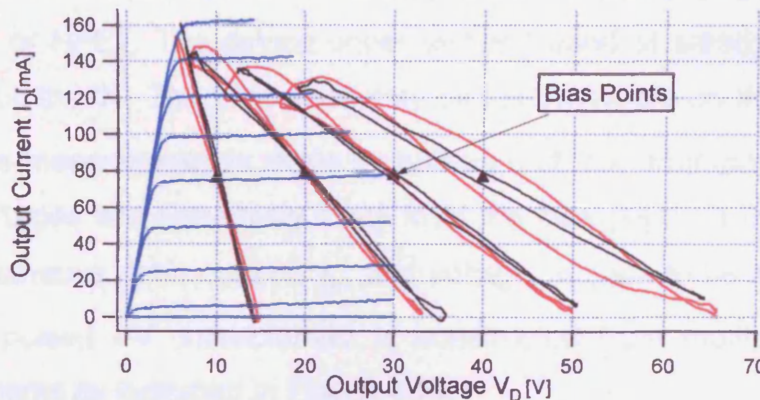


Figure 4.18 – RF dynamic loadlines created before and after harmonic loadpull

4.4 Pulsed I-V Analysis of Dispersion in GaN HFETs

The measurements of the previous section show how useful RF I-V measurements are in quantifying and understanding the effects of dispersion on RF performance. To date however, pulsed I-V measurements have been the main tool used in the analysis of dispersion and are more often used

because pulsed I-V measurement systems are much cheaper and more common than RF waveform measurement systems, and have the advantage of being relatively fast and simple since the setup requires no calibration.

The aim of a pulsed I-V measurement is to emulate the response of a device to large-signal excitation by measuring the instantaneous drain current, i_D, during phase 2 of the current transient plot of Figure 4.6. This means that the measurement is unaffected by ringing, but is made while the state of the device is unchanged from that set by the DC bias, as would be the case when it is operated under RF excitation. At this point the device is also unaffected by thermal dispersion mechanisms other than those set by the DC bias since thermal time constants are also greater than the length of this pulsed measurement [7].

Figure 4.19 illustrates how a complete pulsed I-V characteristic is produced for a FET or HFET. The device under test is biased at steady DC voltages, indicated by the 'X'. This bias point may be set anywhere on the $I_D(V_{DS}, V_{GS})$ plane. One measurement is made by pulsing both the drain-source and gate-source voltages simultaneously away from the bias point, to the point where the instantaneous drain current, i_D, and voltage, v_D, are to be measured. The complete pulsed I-V characteristic is constructed from multiple pulsed I-V measurements as indicated in Figure 4.19.

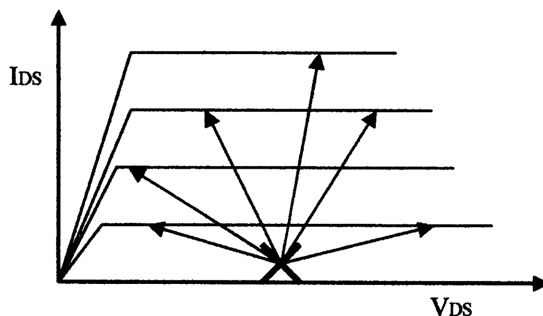


Figure 4.19 – Diagrammatic representation of a pulsed I-V measurement

Pulsed I-V characteristics are often used to quickly determine the degree of dispersion present in a device by comparing them to DC I-V data, but the pulsed I-V measurement system can also be used to create current transient plots such as the one shown in Figure 4.6. These measurements can be used to determine trap time constants and help identify trap locations [6], [8].

The pulsed I-V measurement instrument used at Cardiff University is the Accent DiVA D225 [7], and the system specifications are shown in Table 4.1. The minimum pulse length of the system is 100ns, which equates to a frequency of 10MHz, and is sufficient to allow pulsed measurements to be performed in the current plateau of Figure 4.6. The pulse separation can also be varied greatly so that self-heating in the DUT can be varied, and current transients can be measured up to 1ms. Also, because the DC quiescent bias point is important in determining the state of the device, the DiVA is capable of setting this parameter accurately. This means that the DiVA can be used to simulate large-signal measurements over a range of quiescent bias points, while accounting for the dispersive effects caused by self-heating and trapping.

	Min.	Max.
V_{GS}, v_{GS}	-15V	+10V
V_{DS}, v_{DS}	0V	25V
I_D, i_D	0A	2A
Maximum Power	0W	2W
V_{DS} sweep rate	0.1V/s	10V/s
v_{DS} sweep rate	0.0125V	1V
Pulse length	100ns	1ms
Pulse Separation	0.1ms	1s
Averages	16	2048

Table 4.1 – Accent DiVA D225 specifications

4.4.1 Pulsed I-V Measurements

The pulsed I-V characteristics of the device when biased at V_G=0V, V_D=0V is shown in Figure 4.20. This measurement shows the best-case I-V boundary of the device, where it is free from trapping and self-heating effects, and correlates well with the DC I-Vs shown earlier.

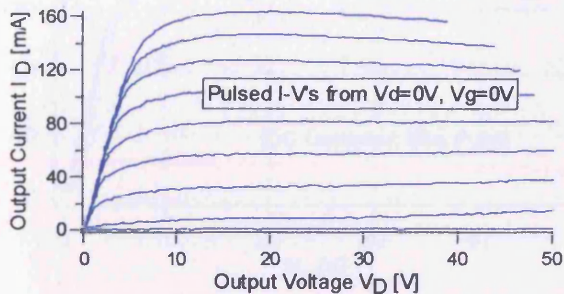


Figure 4.20 – Pulsed I-V measurements when biased at V_G =0V, V_D =0V

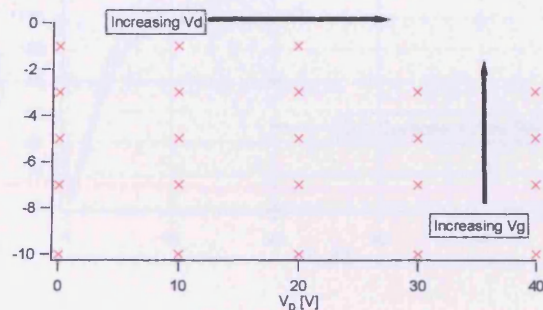


Figure 4.21 – Pulsed I-V bias points

As mentioned previously, the 'state' of the DUT will change depending on its DC bias point. This means that the RF boundary of the device is also bias dependent. To examine this, multiple pulsed I-V measurements can be made about different bias points as shown in the grid in Figure 4.21. The DC quiescent bias point is dependent on the two variables V_G and V_D, so the simplest way to analyse the problem of dispersion is to hold one variable constant and change the other. Such measurements are presented in Figures 4.22 and 4.23.

To compare directly the resulting pulsed I-Vs for each bias point, only one I-V trace need be measured for each bias point. In this case, and to maintain measurement consistency, each measurement is performed with a 'pulse-to' value of v_G=0V. Figure 4.22 shows the resulting pulsed I-Vs when the gate voltage is held constant at V_G=-3V, and the drain voltage is varied. This indicates that dispersion increases as a function of drain bias voltage, when

the gate voltage is held constant. Figure 4.23 shows the resulting pulsed I-Vs when the drain voltage is held constant at V_D=20V and the gate voltage is varied, and indicates that decreasing the gate voltage bias towards pinchoff helps reduce dispersion.

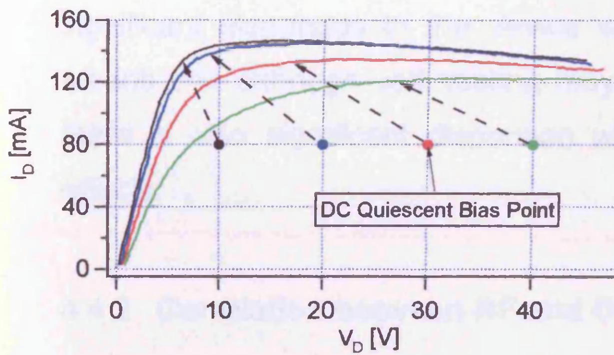


Figure 4.22 – Pulsed I-Vs as a function of drain voltage bias

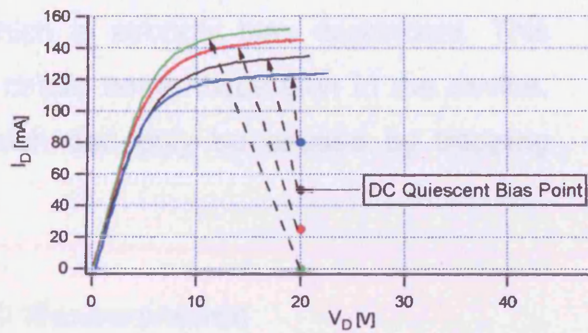


Figure 4.23 – Pulsed I-Vs as a function of gate voltage bias

These measurements again indicate that this device is dispersive, with the dispersive mechanism likely to be trapping effects, either bulk traps or surface states. However, it is possible that the dispersion may be due to self-heating effects in the DUT. In order to examine this possibility an isothermal measurement can be performed using the pulsed I-V measurement system.

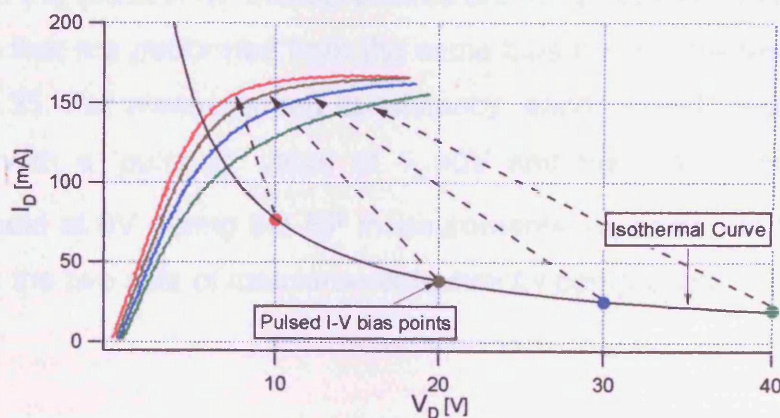


Figure 4.24 – Isothermal pulsed I-V measurement

The isothermal curve in Figure 4.24 is a curve of constant dissipated power, and several pulsed I-V measurements are performed where the DC bias for each measurement is at a different point on this isothermal curve. The results show that although the dissipated DC power, and therefore the temperature of the device is constant during each measurement, there is evidence of significant dispersion in the device which is strongly bias dependent. This means that although self-heating may cause some dispersion in the device, there is also significant dispersion which can only be caused by trapping effects.

4.4.2 Correlation between RF and DC Measurements

It has now been shown that both RF time domain waveform measurements and pulsed I-V measurements are suitable measurements to view and analyse the problem of dispersion. However, one major problem with pulsed I-V measurements is the difficulty in choosing the correct bias point to accurately represent a specific RF operating condition.

The obvious solution is to choose a pulsed I-V bias point that is the same as the DC quiescent bias point of the RF measurements, although there are many factors that determine the state of the device as well as DC quiescent bias point. If the Class A RF measurements presented earlier are overlaid with pulsed I-Vs that are performed from the same bias points, the result is shown in Figure 4.25. For measurement consistency, each pulsed measurement is performed with a 'pulse-to' value of v_G=0V and the maximum input gate voltage is held at 0V during the RF measurements, as shown in Figure 4.26. This makes the two sets of measurements directly comparable.

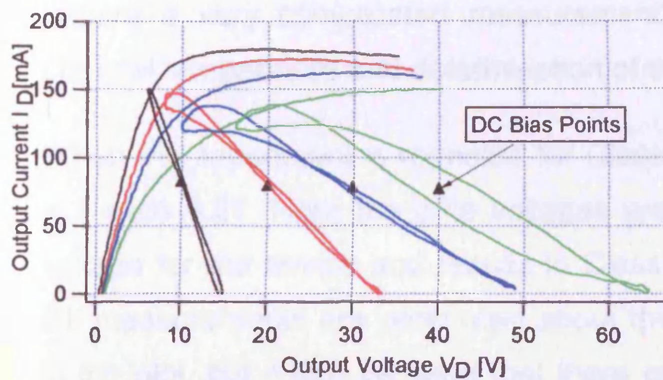


Figure 4.25 – Correlation between pulsed I-V and RF time domain waveform measurements at various class A bias points

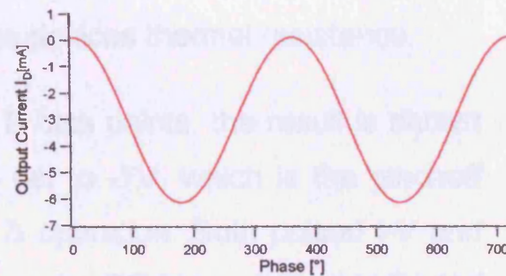


Figure 4.26 – RF maximum input gate voltage is held constant at 0V for measurement consistency

The measurements show a strong correlation between pulsed I-V and RF waveform measurements at class A bias points. However, this may have been expected since the DC quiescent bias point is the same during both measurements, and the DC voltage and current is the same as the average RF voltage and current, which means that the state of the device is largely the same during both measurements.

Therefore, the temperature of the device is the only significant variable that is not constant during both types of measurement. During pulsed I-V measurements the DC bias is held constant resulting in the DUT being 0% efficient since all of the DC power is converted into heat. During Class A RF operation the drain efficiency of the DUT can reach 50%, so that up to 50% of the DC power is converted into RF power.

The result is 'RF cooling' whereby the temperature of the device is cooler during RF operation than during pulsed I-V operation. The temperature of the DUT during the measurements is an important factor since self heating effects can also cause dispersion, but it seems that this temperature difference has little effect on the correlation of the measurements as trapping effects are undoubtedly the major mechanism of dispersion. The self heating effects in the DUT could be taken into account by the use of a hot chuck, but this would

require a very complicated measurement setup and measurement of the channel temperature and determination of the devices thermal resistance.

When the experiment is repeated for Class B bias points, the result is shown in Figure 4.27. Here the gate voltages are set to -7V, which is the pinchoff voltage for the device and results in Class B operation. Both pulsed I-V and RF measurements are performed about the same DC bias point as indicated in the plot, but it can be seen that there is no particular correlation between these measurements.

However, on closer inspection it is observed that during the pulsed I-V measurements the device is biased at pinchoff and no current flows. During RF measurements, where the device is also biased at pinchoff, current does flow in the device. This is due to the self-biasing effect of class B operation and is indicated on the plot. Importantly, these self-biasing points do not exist on the DC I-V plane because no current should flow when the device is biased at pinchoff.

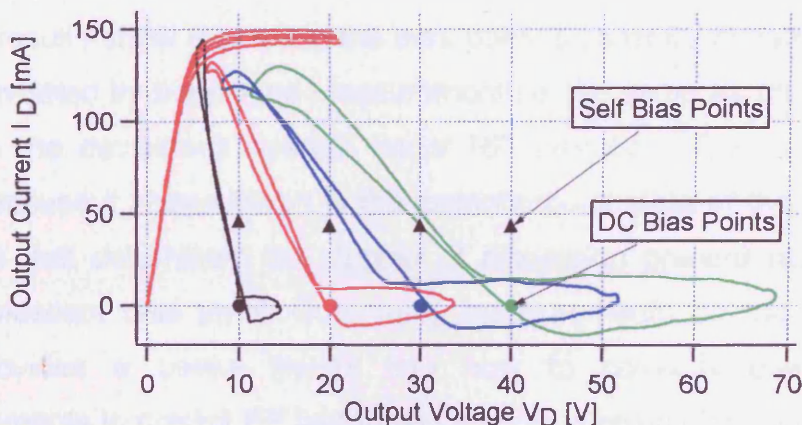


Figure 4.27 – Correlation between RF waveform measurements and pulsed I-V measurements biased at pinchoff

This complicates the bias point selection for the pulse I-V measurement, but it is found that when the pulsed I-Vs are repeated at suitable gate bias voltages which result in drain current flow that is the same as that present during RF

measurements, the plot of Figure 4.28 is obtained. Now when the measurements are considered as a function of DC drain current, good correlation between the DC and RF measurements is apparent. RF cooling will again be a factor during these RF measurements, but it seems to have little effect on the correlation between the two types of measurement.

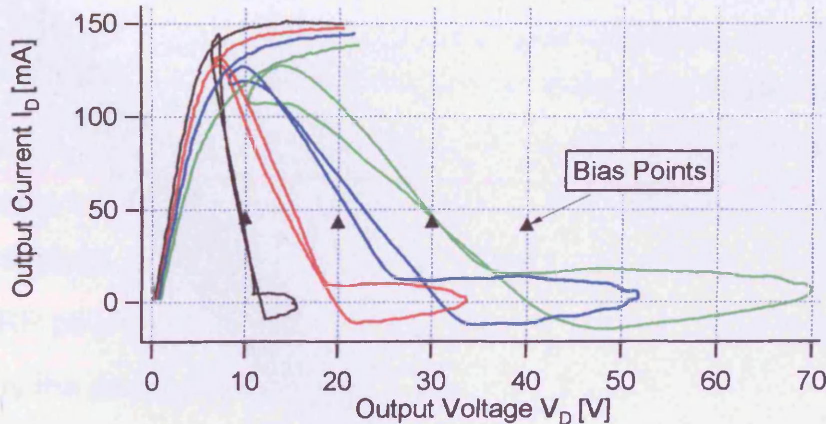


Figure 4.28 – Correlation between RF waveform measurements and pulsed I-V measurements biased with the same quiescent drain current.

Theses results show that when the bias point is carefully chosen, the DC I-V plane emulated by the pulsed measurements is the same as the RF I-V plane in which the device will operate under RF excitation. This is an important result because it shows that it is the instantaneous state of the device under RF drive that determines the degree of dispersion present rather than the initial quiescent bias point. Comparing measurements on the two systems thus provides a useful insight into how to correctly use pulsed I-V measurements to predict RF performance under a given bias condition.

Another important consideration is that since the pulse length can be varied significantly, it is possible that the pulsed measurements are not always performed at the correct point on the current transient – the correct point being the current plateau (phase 2) of Figure 4.6. Therefore, by referencing

the pulsed I-V measurements to the RF waveform measurements in this way, it is possible to confirm that the pulsed I-V system is being used correctly.

4.5 Design for Maximum Efficiency

It has been shown that due to dispersion, the maximum output power of these GaN devices is not as high as would have been expected, but dispersion also has a significant effect on the maximum achievable drain efficiency. For example, Figure 4.29 shows the maximum measured Class A drain efficiency as a function of drain voltage bias, along with the corresponding output power. Drain efficiency, μ, is calculated from Equation 4.2 where P_{RF} is the average output RF power produced by the device, and P_{DC} is the average DC power drawn by the device during RF operation.

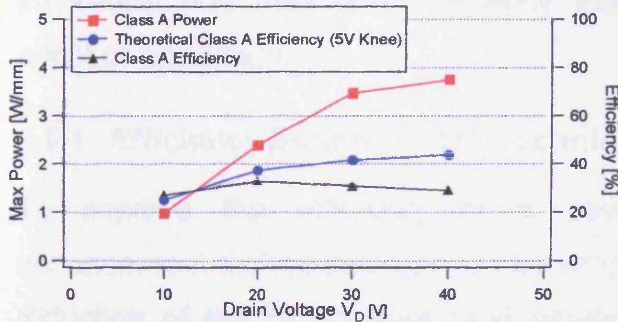


Figure 4.29 – Maximum Class A efficiency as a function of drain bias

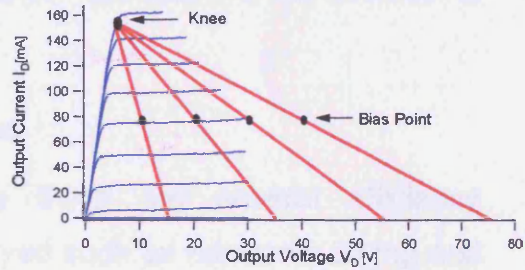


Figure 4.30 – DC I-V characteristic of the DUT and theoretical Class A loadlines

$$\mu_{Drain} = \frac{P_{RF}}{P_{DC}} \quad (4.2)$$

The theoretical efficiency of the device is also shown in Figure 4.29. In theory the maximum Class A efficiency is 50% but this figure is calculated based on an ideal device which has a knee voltage of 0V. The knee voltage is the key parameter that determines efficiency, and maximum efficiency can only be achieved when the knee voltage becomes small relative to the total voltage

swing. This is illustrated in Figure 4.29 where the theoretical efficiency increases towards 50% as the drain voltage bias increases. This theoretical efficiency is based on the assumption of a 5V knee voltage and 150mA knee current, as illustrated in Figure 4.30.

Despite the presence of a non-zero knee voltage, efficiencies of around 50% are still achievable for most devices operated in Class A, when the device is overdriven and the waveforms become clipped rather than sinusoidal [9]. However, the plot of Figure 4.29 shows that when this GaN device is driven into saturation the maximum measured efficiency is nowhere near 50%. The reason is of course dispersion, which increases the knee voltage significantly and reduces output power and efficiency. Interestingly, Figure 4.29 also shows that when the device is biased at 10V, where there is little or no evidence of dispersion, the measured efficiency is actually more than had been predicted. As the drain bias increases, the discrepancy between the theoretical and measured efficiency increases because of the increasing effect of dispersion.

4.5.1 Efficiency Enhancement Techniques

To improve the efficiency of a device there are several efficiency enhancement techniques that can be employed such as harmonic tuning and reduction of the DC current (and power) by biasing closer to pinchoff. In classical power amplifier design several modes of operation have been specifically defined in terms of bias conditions and harmonic tuning.

In Class A operation the quiescent DC voltage and current are chosen so that they are half of the total RF voltage and current swings. This is shown in Figure 4.31a and means that when RF power is applied, the DC current drawn by the device remains the same as the quiescent current, and the resulting efficiency is no greater than 50%. Class A operation has no specifically defined harmonic tuning so the harmonics are normally terminated into the same impedance as the fundamental or a nominal 50Ω load, and the

resulting RF voltage and current waveforms are sinusoidal, as shown in Figures 4.32a and 4.32b.

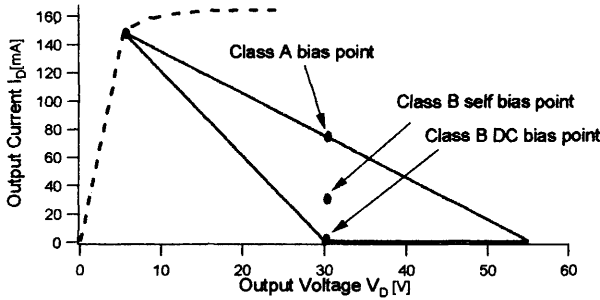


Figure 4.31a – Loadline representation of Class A and Class B operation

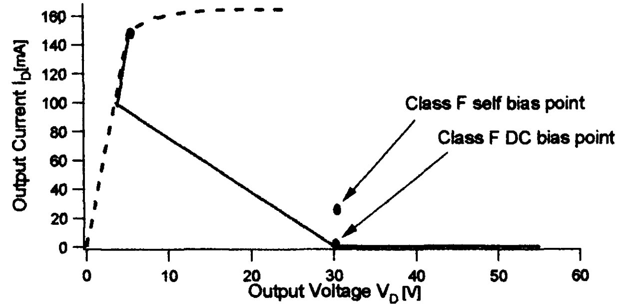


Figure 4.31b – Loadline representation of Class F operation

In Class B operation the DC quiescent current is set to pinchoff so that no current flows in the device, and the second harmonic is terminated with a short circuit. When RF is applied no current flows for half of the conduction cycle so that the device is 'off' half of the time. This results in a half rectified current waveform as shown in Figure 4.33b, while the voltage waveform of Figure 4.33a remains sinusoidal.

The resulting RF Class B loadline is also shown in Figure 4.31 and shows that the device is biased at pinchoff and is off for half of the conduction cycle. When RF is applied, DC current must flow to allow the device to operate. This 'self bias point' is determined by the average DC current that is drawn during RF operation and is considerably less than the current drawn during Class A operation. The total RF voltage and current swing (and RF power) is the same as in Class A so the overall result is significantly improved efficiency. In an ideal device with no knee voltage the theoretical maximum Class B efficiency is 78.5% [9].

Class F operation is similar to Class B but with the addition of harmonic tuning to present the odd order harmonics with an open circuit. This has the effect of altering the RF voltage waveform from sinusoidal to a square wave, as shown

in Figure 4.34a. In theory this means that the RF voltage and current are never non-zero at the same time and results in 100% efficiency. However, the creation of a perfect square wave means presenting an open circuit to an infinite number of odd order harmonics, which is not realisable. In practice, the greatest improvement in efficiency is achieved by tuning the lower order harmonics and it is found that by tuning only the third harmonic to an open circuit results in a theoretical efficiency of 90.7% [10]. The resulting band-limited Class F loadlines look like that shown in Figure 4.31b.

Figure 4.32a –Class A voltage waveform

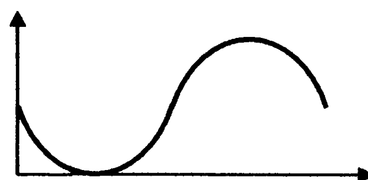


Figure 4.32b –Class A current waveform

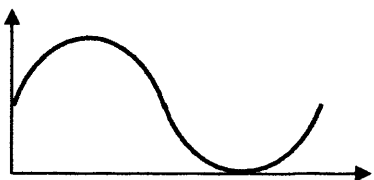


Figure 4.33a –Class B voltage waveform

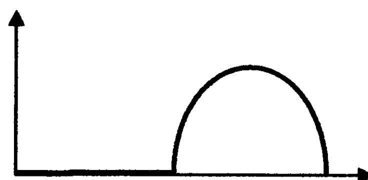


Figure 4.33b –Class B current waveform

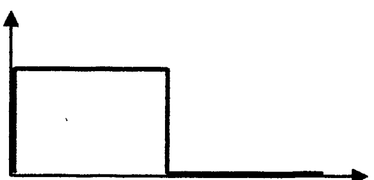


Figure 4.34a –Class F voltage waveform

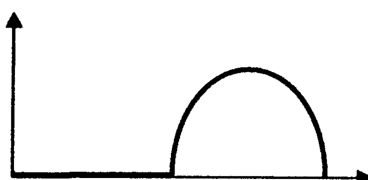


Figure 4.34b –Class F current waveform

4.5.2 Efficiency Measurements

To determine the maximum achievable drain efficiency of the device under test, it was measured in Class A, Class B, and Class F operation at various input drive levels and at several drain voltages. Active loadpull was used to loadpull the fundamental impedance and to present precise short circuit and open circuit terminations to the harmonics where necessary. The results are shown in Figure 4.35 as a function of drain voltage bias.

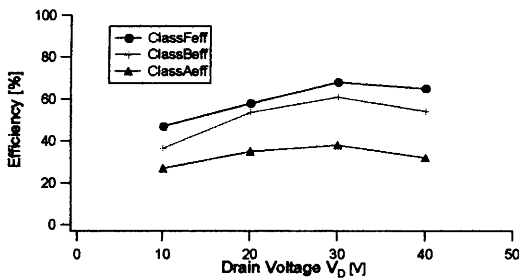


Figure 4.35 – Maximum drain efficiency

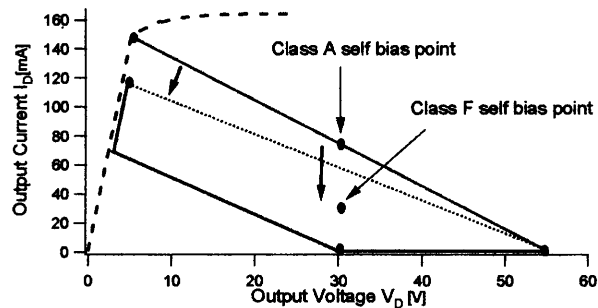


Figure 4.36 – Mechanism of increased efficiency performance

The best performance is achieved in Class F operation into a 380Ω fundamental load at 30V drain bias. This results in 68% efficiency and output power of approximately 3.5W/mm. In theory, the larger 40V drain bias should result in better efficiency but in this case it is reduced because of increased dispersion at the higher voltage.

To be certain of achieving the maximum possible efficiency, the fundamental load presented to the device must also be varied and interestingly, the best efficiencies are achieved when the fundamental is presented with larger impedances than expected. This is illustrated in Figure 4.36 and has the effect of reducing the total RF current swing but increasing the RF voltage swing. It also reduces the DC current drawn by the device so that DC power consumption is reduced, and the overall effect is increased efficiency.

An unwanted side effect of this reduced RF current swing is reduced RF output power, but at 3.5W/mm the power density is still much higher than that achievable from comparable technologies such as Si LDMOS and GaAs HBTs or HFETs [4].

4.5.3 Waveform Engineering for Maximum Efficiency

The RF voltage and current waveforms shown in Figure 4.37a and 4.37b correspond to the measurement of maximum efficiency, and it is obvious that these waveforms are far from the ideal Class F waveforms of Figure 4.34. The voltage waveform is possibly as close to a square wave as is achievable with only three harmonic terminations, but the current waveform is not even close to an ideal half rectified waveform.

There are two main reasons for the non-ideal shape of the waveforms. Firstly, the output capacitance of the device causes the non-flat shape at the bottom of the waveform, and secondly, the non-linear transfer characteristic of the device means that not all of the harmonic components line up in-sync to create the ideal waveform shape. However, the full capability of the waveform measurement system can be used to systematically tune the shape of the waveforms via fundamental and harmonic loadpull so that they become closer to the ideal case. This type of device optimisation has been termed 'waveform engineering' and the result is a better design solution with better efficiency performance.

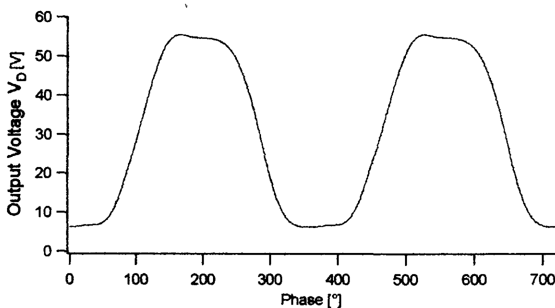


Figure 4.37a –Class F voltage waveform

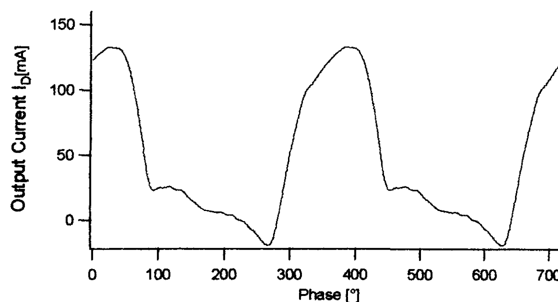


Figure 4.37b –Class F current waveform

The fundamental load impedance of the device is inductively tuned to compensate for the effect of the output capacitance. This has the effect of flattening out the bottom of the current waveform and is achieved by loadpulling the load away from the real axis as illustrated in the Smith chart of Figure 4. 39. The polar coordinates of the resulting optimal fundamental load impedance is found to be: $\Gamma_{f_0} = 0.8\angle 5^0$. Similarly, the second and third harmonic load impedances are loadpulled to impedances that result in optimal efficiency performance (and more ideal waveform shape). The second harmonic impedance, $\Gamma_{2f_0} = 0.98\angle 177^0$, remains close to the ideal short circuit termination of Class F operation, while the third harmonic impedance, $\Gamma_{3f_0} = 0.98\angle 12^0$, is loadpulled significantly from the ideal open circuit condition.

This waveform engineering results in the RF voltage and current waveforms of Figure 4.38a and 4.38b, which are closer to ideal Class F waveforms. It also results in an improved performance of 73% efficiency and 3.5W/mm. The RF dynamic loadline is also shown in figure 4.40 and shows that dispersion is still apparent. This efficiency performance is still far from what would be expected of a well behaved non-dispersive device, but is a significant improvement on the 68% efficiency that is achieved from the original Class F design and shows the advantages of the waveform measurement system and the insight it provides.

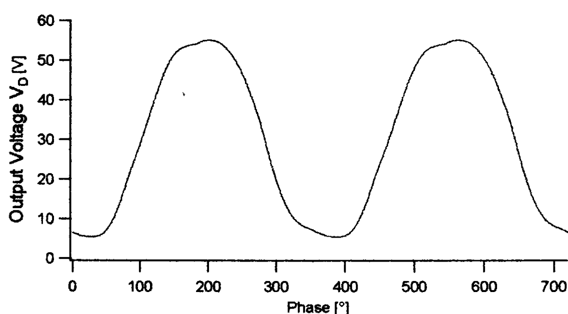


Figure 4.38a – Tuned Class F voltage waveform

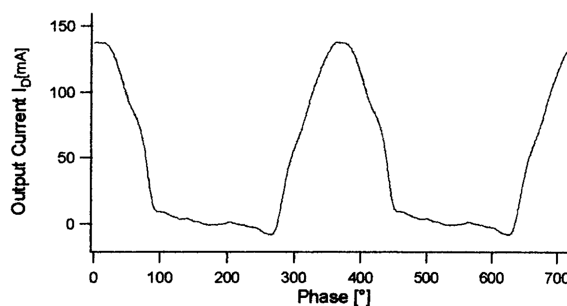


Figure 4.38b – Tuned Class F current waveform

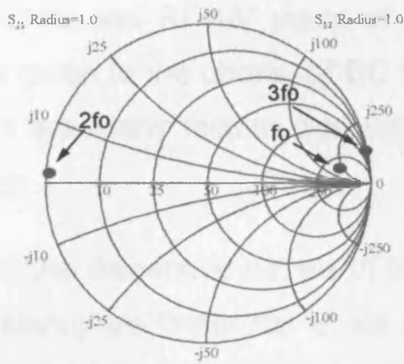


Figure 4.39 – Smith chart showing optimal harmonic impedances

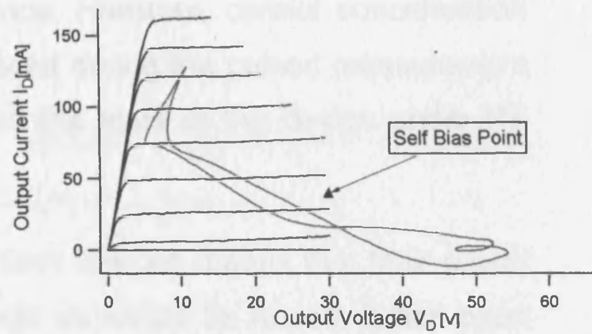


Figure 4.40 – RF dynamic loadline of device at optimum efficiency

4.7 Summary

A detailed RF waveform analysis of GaN HFETs has been performed for the first time in an attempt to better understand the performance limitations introduced by the presence of DC-RF dispersion. This waveform analysis confirms that the unexpectedly poor power and efficiency performance is due to dispersion, which causes the RF I-V boundary to be different to the DC I-V boundary. It has also been shown that the effects of dispersion can be reduced by biasing the device closer to pinchoff, which improves the power and efficiency performance.

The non-linear measurement system can measure the RF voltage and current waveforms at the device plane. This has allowed a direct comparison to be made between the real large-signal RF I-V plane in which the device operates and the pulsed DC I-V plane as measured by a pulsed I-V measurement system. It has long been known that because pulsed I-V measurements can be performed very fast, they can emulate the effect of an RF signal, and it has been shown for the first time that there is very strong correlation between the pulsed I-V measurements and the RF waveform measurements, and that the pulsed I-V measurements can quite accurately predict the true RF

performance and RF I-V plane of a device. However, careful consideration must be given to the choice of DC bias point during the pulsed measurement as there are many factors that determine the state of the device under RF excitation.

Although the dispersive nature of these GaN devices means that their power and efficiency performance is not as good as would be hoped, it has been shown that it is possible to produce a power amplifier solution that has very good power performance relative to other device technologies, and relatively good efficiency. This is achieved through use of the knowledge gained from the pulsed I-V and RF waveform measurements, as well as standard tuning techniques and waveform engineering techniques that are unique capabilities of the non-linear measurement system at Cardiff University.

4.7 References

[1] P. H. Ladbrooke, *Pulsed I(V) Measurement of Semiconductor Devices*: Accent Optical Technologies Inc., 2004.

[2] S. C. Binari, "Trapping Effects and Microwave Power Performance in AlGaN/GaN HEMTs," *IEEE Transactions on Electron Devices*, vol. 48, 2001.

[3] A. E. Parker and J. G. Rathmell, "Measurement and Characterization of HEMT Dynamics," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, November 2001.

[4] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "RF and Microwave Power Amplifier and Transmitter Technologies - Part 1," in *High Frequency Design*, 2003, pp. 22–36.

[5] S. Nuttinck, "Thermal Analysis of AlGaN-GaN Power HFETs," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 2445-2452, 2003.

[6] S. C. Binari, "Trapping Effects in GaN and SiC Microwave FETs," *Proceedings of the IEEE*, vol. 90, No. 6, 2002.

[7] *Accent DIVA Models D210, D225, D225HBT, D265 Dynamic I-V Analyzer, User Manual*, vol. 1.0: Accent Optical Technologies, 2001.

[8] P. Ladbrooke and S. Blight, "Low-Field Low-Frequency Dispersion of Transconductance in GaAs MESFETs with Implications for Other Rate-Dependent Anomalies," *IEEE Transactions on Electron Devices*, vol. 35, pp. 257-267, 1988.

[9] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*: Artech House, 1999.

[10] F. H. Raab, "Maximum Efficiencyand Output of Class-F Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, 2001.

Chapter 5. Linearity Investigations of GaN HFETs

It has been shown in chapter 4 that despite suffering from dispersion, QinetiQs GaN devices can operate effectively as power amplifiers in terms of power and efficiency performance. Additionally, advances in the technology since the start of this work has resulted in some groups producing GaN devices which they claim are dispersion free [1], [2]. It would therefore seem likely that advances in material technology will eventually enable the mass production of dispersion free devices for power amplifier and other applications.

However, if these devices are to be successfully commercialised, it is necessary to also examine their linearity performance in more detail as this is one of the most important characteristics of a power amplifier, along with power and efficiency. At the time of this work, little had been published on the linearity of GaN HFETs since the technology was in the early stages of development, but the non-linear measurement system is a useful tool to gain an insight into their linearity performance through CW, two-tone, and memory effect analysis.

5.1 Nonlinear Distortion

The power amplifier is a central component of any communications system and its purpose is to amplify the input signal to an appropriate level while maintaining high efficiency and linearity, where linearity relates how accurately the output signal correlates to the input signal. In order to achieve maximum efficiency, the PA must be driven close to saturation [3]. However, all active semiconductor devices including HFETs are non-linear in nature. This non-linear nature can be described by a non-linear transfer function, and as the device is operated closer to saturation, the output becomes more non-linear.

The result is a trade-off between efficiency and linearity and is one of the biggest challenges facing a PA designer.

When a complex multi-carrier modulated signal such as a Code Division Multiple Access (CDMA) signal, which is used in 3G mobile phones, is transmitted through a power amplifier, the non-linear nature of the device results in distortion of both the amplitude and phase of the signal. This distortion causes the generation of unwanted frequency components inside and outside the intended frequency band via harmonic distortion and intermodulation distortion (IMD). Many of these out-of-band frequency components can be filtered but the in-band frequency components cannot.

This IMD also results in spectral regrowth whereby unwanted IMD components appear in adjacent transmission channels around the main transmitted signal and may cause interference with the signal that is allocated to that adjacent channel. This is called adjacent channel interference and is measured as an adjacent channel power ratio (ACPR). A typical plot of the input and output signal from a power amplifier driven with a modulated signal is shown in Figure 5.1(a) and illustrates the resultant spectral regrowth sidebands.

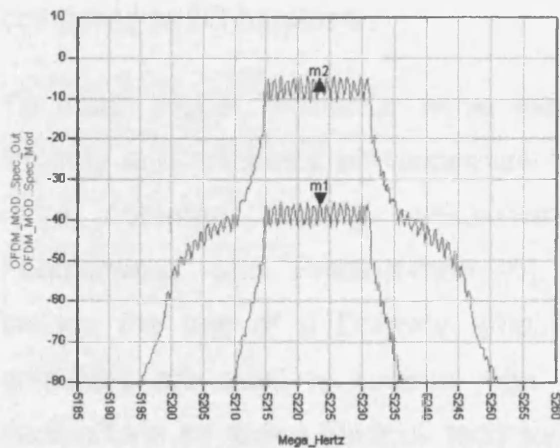


Figure 5.1(a) – Simulated input and output modulated signal illustrating spectral regrowth

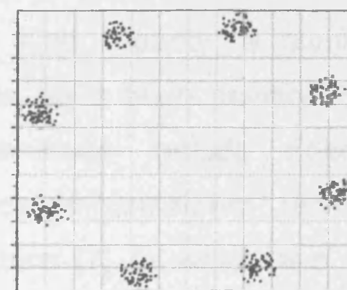


Figure 5.1(b) – Errors in the demodulated I-Q constellation diagram.

From a system point of view the ultimate effect of the non-linear power amplifier is the introduction of bit errors in the digital data. The effect of non-linear distortion on a digitally modulated signal can be seen in the demodulated I-Q constellation diagram illustrated in Figure 5.1(b). When the signal is demodulated, errors in the output I-Q constellation occur at the sample points, possibly resulting in bit errors and ultimately a corrupted data stream.

A particular problem in modern communications systems is that in order to transmit higher data rates and reduce frequency bandwidth usage, the transmission formats that they employ, such as CDMA and Wideband CDMA (WCDMA), have stringent linearity specifications. For example, the maximum ACPR of a WCDMA transmitter should be $\geq 45dBc$ for bandwidth up to *5 MHz* and ≥ 50 *dBc* for *10 MHz* [4]. Additionally, the CDMA and WCDMA formats also have higher peak-to-average power ratios than previous transmission formats such as those employed in GSM for 2G mobile phones. This means that the amplifier has to be operated significantly backed-off from saturation so that linearity specifications are met [5], and therefore the power amplifier cannot maintain a highly efficient mode of operation. The practical result of this can be seen in the reduced battery life of 3G mobile phone handsets compared to 2G handsets.

To attain higher efficiency while maintaining high linearity, a number of linearity and efficiency enhancement techniques have been developed. The more common linearity enhancement techniques include Feedback, Feedforward, and Predistortion [6]. Efficiency enhancement techniques include the use of a Doherty amplifier structure [7] in which two power amplifiers are used to sustain high efficiencies while operating at power back-off via an active loadpull technique, and harmonic manipulation at both input and output of the power amplifier [8], [9].

5.2 CW Large-Signal Characterisation

In communications systems the power amplifier is the main cause of distortion to the transmitted signal, and therefore its linearity must be accurately characterised. There are many ways to measure and characterise the linearity of a PA, such as the measurement of ACPR, constellation diagrams, and bit error rate, and figures of merit have been developed to compare different devices. However, some of these characterisation techniques require expensive measurement systems and are time and labour intensive, and ultimately are only applicable to complete systems, not individual power amplifier modules.

When designing, characterising, and testing PAs it is not necessary or even advantageous to test using a realistic multi-tone modulated signal. The simplest and default characterisation tool is CW analysis. CW analysis can be used to determine amplitude and phase distortion introduced by a PA and simple figures of merit are used to compare different devices.

5.2.1 CW Theory

When a power amplifier is presented with a CW sinusoidal signal at the input, its non-linear transfer characteristic results in harmonic distortion, gain compression (AM-AM), and phase distortion (AM-PM). Harmonic distortion is the generation of frequency components that are not present in the input signal, while gain compression describes how the gain of the device becomes a function of the input amplitude, and phase distortion describes how the phase of the output frequency components changes relative to that of the input, as a function of input amplitude.

The Volterra series analysis [10] is the most general representation of a non-linear system but it can be very complicated. However, by assuming that the DUT is memoryless, the non-linear transfer characteristic can be described by the relatively simple power series of Equation 5.1. If a CW sinusoid is applied

to a non-linear power amplifier, the output contains frequency components that are integer multiples of the input frequency, and if $v_i(t) = A\cos(\omega t)$, the power series is expanded to that of Equation 5.2.

$$i_o(t) = a_0 + a_1 v_i(t) + a_2 v_i^2(t) + a_3 v_i^3(t) + ... \qquad (5.1)$$

$$i_o(t) = a_0 + a_1 A\cos(\omega t) + a_2 A^2 \cos^2(\omega t) + a_3 A^3 \cos^3(\omega t) + ... \qquad (5.2)$$

When expanded, the equation reveals the magnitude of the DC and harmonic components that constitute the output signal:

$$i_o(t) = \left(a_0 + \frac{a_2 A^2}{2} + \frac{3a_4 A^4}{8} + ... \right) + \left(a_1 A + \frac{3a_3 A^3}{4} + \frac{5a_5 A^5}{8} + ... \right)\cos(\omega t)$$
$$+ \left(\frac{a_2 A^2}{2} + \frac{a_4 A^4}{8} + ... \right)\cos(2\omega t) + \left(\frac{a_3 A^3}{4} + \frac{5a_5 A^5}{6} + ... \right)\cos(3\omega t) + ... \qquad (5.3)$$

The terms in this power series have been evaluated up to the fifth order and clearly show how harmonic distortion arises. It is also important to note that the magnitude of each odd harmonic component is dependent on the higher odd order components, and the magnitude of each even harmonic component is dependent on the higher even order components, although this is insignificant until A, the magnitude of the input signal, becomes relatively large. Therefore, at low drive levels, the magnitudes of the nth harmonic components are proportional to A^n and the slope of each harmonic component on the output power versus input power curve is n:1.

Taking the fundamental output component as an example, at low input drive levels, when A is small, the $a_1 A$ term is dominant and the output signal increase linearly with the input. At higher drive levels the $\frac{3a_3 A^3}{4}$ term becomes significant and makes the output deviate from the linear 1:1 ratio. In power amplifiers, at high drive levels, the output is always a compressive, or saturating, function of the input so that the gain approaches zero for

sufficiently large input levels. This occurs when $a_3 < 0$ and is the mechanism behind gain compression, or AM-AM. It is also important to note that all harmonic components are subject to gain compression in the same way, as the higher order components that comprise them become significant.

The non-linear response of the device can also be viewed through its I-V transfer characteristic which will look something like the plot of Figure 5.2(a). The dotted line shows an ideal transfer characteristic, which is linear between pinchoff and current saturation, while the solid line is the more realistic case which is a softer, compressive function of input amplitude. In the example shown, this softer transfer characteristic is still symmetrical about the midpoint, $I_{max}/2$. The ideal, linear transfer characteristic would require a polynomial power series with an infinite number of terms, while the more realistic transfer characteristic can be created by considering a much lower number of terms in the power series [13], with a simple third order polynomial often being used.

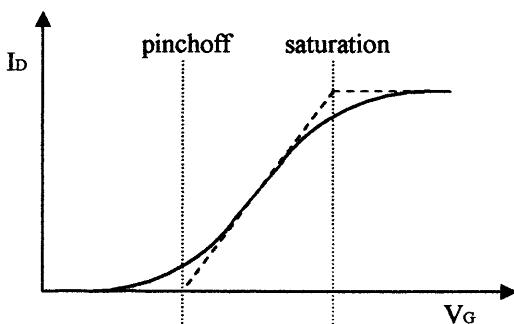


Figure 5.2(a) – DC transfer characteristic

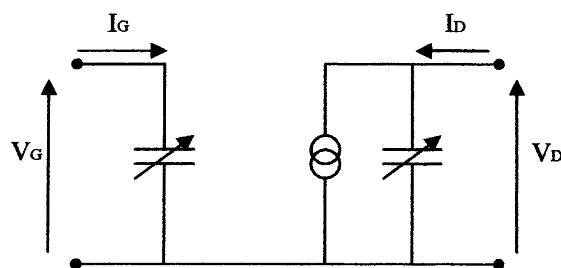


Figure 5.2(b) – Generic device model

Figure 5.2(b) shows a simple generic device model that is used to represent the sources of the non-linear output of the device for modelling purposes. This model consists of a non-linear input capacitor, or charge generator, and a non-linear output current generator and charge generator. These reactive elements can have the effect of phase shifting the signal and are the cause of AM-PM. Equations 5.1–5.3 only describe AM-AM behaviour, but one way of

including AM-PM behaviour in the description is to add individual time delay components into Equation 5.1, which results in Equation 5.4 [13].

$$i_0(t) = a_0 + a_1 v_i(t+\tau_1) + a_2 v_i^2(t+\tau_2) + a_3 v_i^3(t+\tau_3) + \ldots \qquad (5.4)$$

5.2.2 CW Large-Signal Characterisation

The non-linear waveform measurement system described in chapter 3 is a suitable instrument to measure and characterise the non-linear effects of a power amplifier via CW power and phase measurements. The device under test is a *2x125* μm AlGaN/GaN HFET on a SiC substrate with gate length of 0.29μm and source-drain gap of *3.4* μm, resulting in small-signal performance of $f_T = 35GHz$, $f_{MAX} = 80GHz$. The HFET layer structure, as illustrated in Figure 5.3, consists of 30nm AlGaN and 1.2μm insulating GaN on Vanadium doped SiC, with no intentional doping included in the structure. A silicon nitride passivation layer is added in an attempt to reduce dispersion.

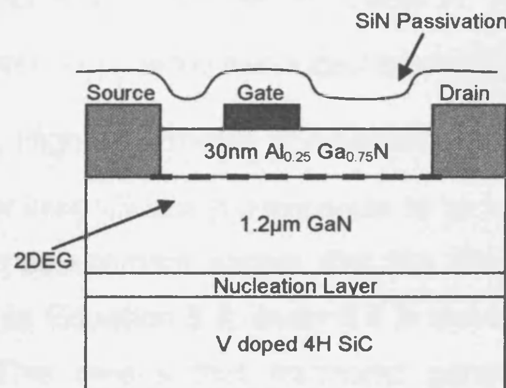


Figure 5.3 – HFET layer structure

CW analysis is performed on-wafer at a frequency of $840MHz$ and all input and output harmonic components are passively terminated into a passive 50Ω load impedance. The device is operated in Class A at a drain voltage of 10V and quiescent drain current of 125mA, resulting in a maximum output

power of approximately 2W/mm and maximum current density of approximately 1A/mm. The 10V drain bias is chosen because it allows excitation of the devices knee region using a 50Ω load termination, without the need for loadpull. At this relatively low drain voltage, the stress on the device is also minimized, as is dispersion.

Unfortunately at the time of this work these devices were in the early stages of development and were not very robust. This means that they could only be operated in Class A bias and could not be operated reliably and repeatedly in other more stressful modes of operation without degradation occurring. However, Class A operation is the most basic amplifier mode of operation and many high power amplifier applications utilize Class A or Class AB operation, so a linearity investigation of Class A operation is relevant and useful.

5.2.2.1 CW Power Measurements

Figure 5.4 shows the measured output fundamental and harmonic power of the device under test when operated in Class A. The input power is the available source power, $|a_1|^2$, while the output power is the power absorbed by the load, $|b_2|^2 - |a_2|^2$. Higher harmonic components can also be measured but are often too noisy or insignificant in magnitude to be included.

The power sweep measurement shows that the device behaves largely as expected according to Equation 5.3, even if it is truncated to a relatively low order polynomial. This means that harmonic generation occurs and the amplitudes of the fundamental and harmonic components increase linearly until saturation is approached, when gain compression and AM-AM interaction starts to occur. At low drive levels the fundamental output power sweep has a gradient of 1:1, while the second and third harmonics have gradients of 2:1 and 3:1 respectively. This is confirmed by comparison with the gradient lines that have been superimposed on Figure 5.4.

The fourth harmonic does not however follow a gradient of 4:1 as expected, although the points shown in the plot are at relatively high drive levels where the fourth harmonic is affected by higher order term. It is therefore likely that the fourth harmonic sweep does have a gradient of 4:1 at lower drive levels, but the dynamic range of the measurement system is not large enough to measure accurately at these low drive levels.

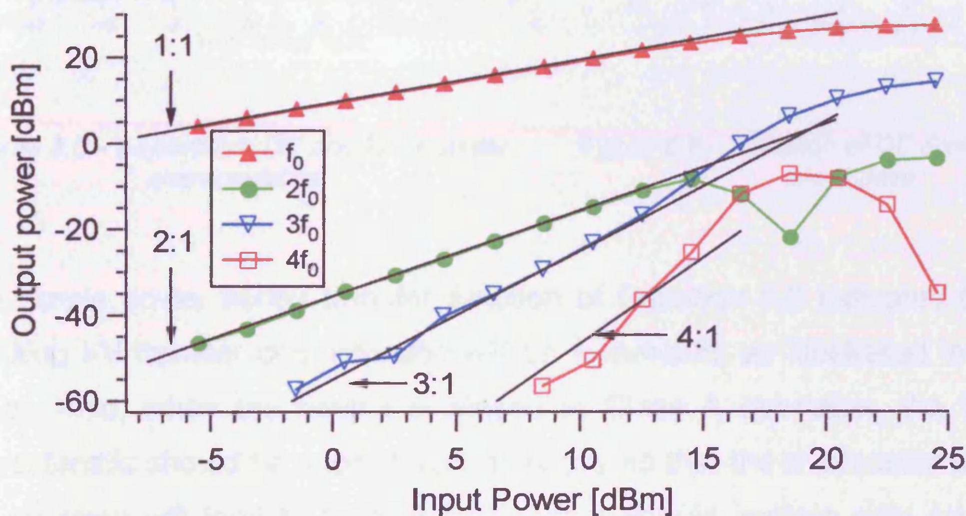


Figure 5.4 – CW power sweep

At higher drive levels the gradients of the harmonic components change due to the vector addition and/or subtraction of higher order components as described in Equation 5.3. This can cause complex interaction between the harmonic components, as is evident. The fundamental component experiences gain compression because of the rapidly rising third harmonic which is in anti-phase to it and has a significant contribution to the overall magnitude of it. Also, at higher drive levels the third harmonic starts to rise at a rate greater than 3:1. This is due to a significant fifth harmonic term which adds to it and causes its magnitude to increase. Similar interactions can be seen in the other harmonic components. For example, as compression is approached, the magnitude of the second harmonic starts to fall rapidly to a

minimum point before rising again, due to the presence of significant higher even order harmonics.

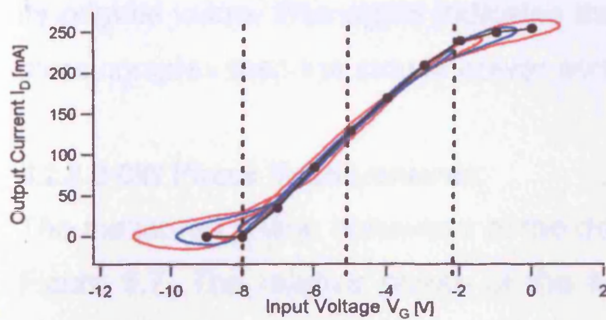


Figure 5.5 – Asymmetric DC and RF transfer characteristics

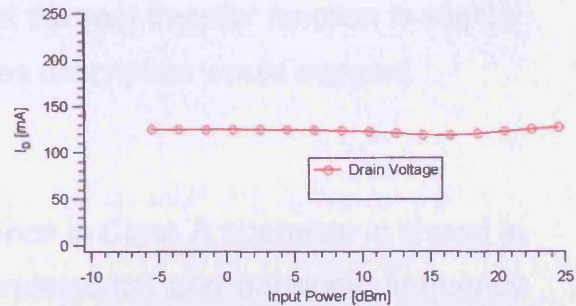


Figure 5.6 – Variation of DC drain current with input drive

The simple power series transfer function of Equation 5.2 indicates that the resulting I-V transfer characteristic will be symmetric, as illustrated in Figure 5.2(a). Also, when the device is biased in Class A operation, the transfer characteristic should be excited symmetrically so that the even order terms of the equation will tend to zero and its output should contain only odd order harmonics [11]. In practice however, there are always significant even order harmonics present in all amplifier modes of operation, as is observed in Figure 5.4. This is consistent with other work on HFET structures [12] and indicates that the real transfer function of the device is not exactly symmetric. This is confirmed by the measured DC and RF transfer characteristics which are shown in Figure 5.5. For Class A operation, the device is biased at $V_G = -5.2V$ and the DC transfer characteristic shows that pinchoff is at $V_G = -8V$.

If the device was biased in Class AB rather than Class A, more significant even harmonics might be expected, because the transfer characteristic would be excited asymmetrically. However, Figure 5.6 shows the variation of drain current, I_D, with drive level and confirms that the device is very close to Class A operation. Interestingly, when the device approaches saturation, it starts to self-bias downwards, indicating that the top end of the transfer characteristic

is reaching saturation first. Then when the bottom end of the transfer characteristic starts to saturate also, the self-bias point moves back towards its original value. This again indicates that the real transfer function is slightly more complex than the simple power series description would suggest.

5.2.2.2 CW Phase Measurements

The measured phase behaviour of the device in Class A operation is shown in Figure 5.7. The relative phase of the fundamental and harmonic frequency components is calculated using the output large signal travelling voltage wave of each frequency component, b_2, referenced to the fundamental input travelling voltage waveform, a_1. Higher harmonic components can also be measured but are too noisy or insignificant in magnitude to be included.

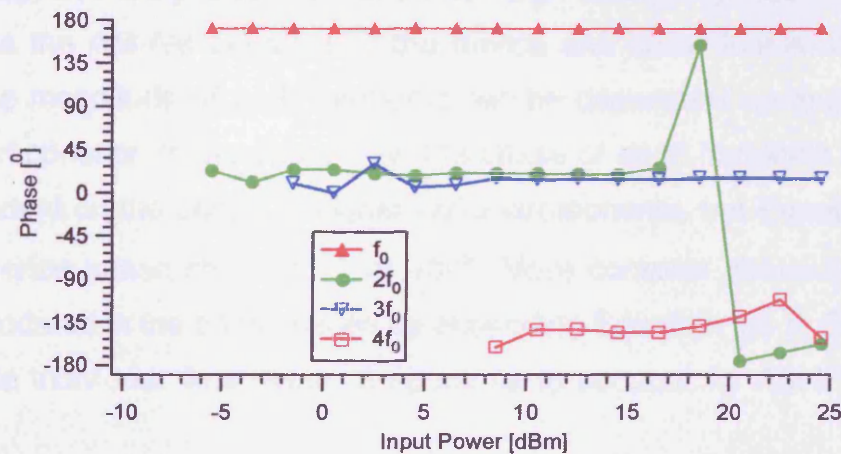


Figure 5.7 – Output phase behavior as a function of input power

According to Equation 5.3, the phase of the output frequency components should be either 0^0 or $\pm 180^0$ relative to the input. At first glance this appears to be largely correct, with the phase of the fundamental being in anti-phase to the input. This is expected since the device is a voltage controlled current generator and the input voltage and output current waveforms should be in-phase, with the output voltage in anti-phase to these.

At low drive levels the exact phase of the fundamental, second harmonic, and third harmonics are almost constant at 171^0, 20^0, and 15^0 respectively, while the phase of the fourth harmonic is not constant but is approximately -145^0. The fundamental is offset from the expected 180^0 phase shift by 9^0 and this constant offset is due to the transit delay of the device. The transit delay will also affect the relative phases of the other harmonics and explains why the harmonics do not line up exactly in-phase or anti-phase with each other. However, this alone cannot explain all of the observed phase behaviour, such as the non-constant phase of the even harmonics.

At higher drive levels the even harmonics exhibit significant phase rotation and the phase shift of the second harmonic coincides with a minimum point in the power sweep, as would be expected, indicating that the higher order harmonics have a significant effect on its magnitude and phase. Equation 5.3 describes the AM-AM behavior of the device and show that at higher drive levels the magnitude of each harmonic can be dependent on multiple higher order components. In the same way, the phase of each harmonic component is dependent on the phase of higher order components, but Equation 5.3 can only describe phase shifts of 0^0 or 180^0. More complex phase behaviour is accommodated in the power series by expanding Equation 5.3 to Equation 5.4 to include individual time delay components to account for AM-PM behavior [13].

The odd harmonics experience little or no phase deviation, or AM-PM, so it can be assumed that their phase offsets are due mainly to the transit delay through the device. Therefore, $\tau_1 = \tau_3 = \tau_5$. This implies that for the phase of the fundamental to be constant, the phase of the third and other significant odd harmonics must be either in phase or anti-phase with the fundamental and also exhibit no phase rotation. This is clearly seen to be the case.

The phase behaviour of the even harmonics is more complex and they are not as constant as the phases of the odd harmonics. The second harmonic undergoes a phase shift of approximately 180^0 as the device approaches

saturation and then continues to phase rotate. A significant fourth order component must be included in the power series to account for this behavior, but since the phase is non-constant, $\tau_2 \neq \tau_4$. The fourth harmonic is also non-constant, indicating that $\tau_4 \neq \tau_6$. It is likely that this non-constant phase behaviour is due to the non-linear reactive elements of the device model, as illustrated in Figure 5.2(b), and these results suggest that these reactive elements are largely even order.

Therefore, the overall power and phase behaviour of the GaN device is close to ideal, as predicted by the simple low order power series of Equation 5.3, although minor alterations are required to fully describe it, as the presence of even harmonics and an asymmetric transfer characteristic, along with non-constant phase components suggests that the real description is slightly more complex.

5.2.3 Comparison with a GaAs HEMT

To better understand the performance of the GaN HFET devices, it is instructive to compare it to other, similar device technologies such as a GaAs HEMT. In theory the two device technologies are very similar, and GaAs pHEMTs have been shown to make very good power amplifiers with relatively linear performance [15], [16]. The measurements below are performed on a *2x100μm* Filtronic GaAs pHEMT and are directly comparable to the earlier measurements performed on the GaN HFETs. The device is again operated in Class A at *840MHz* and biased in the same way, to produce maximum output power by penetrating the knee region of the device.

The power and phase measurements of Figures 5.8(a) and 5.8(b) show that the GaAs device does not quite follow the predicted behaviour according to the power series transfer function of Equation 5.3. The gradient of the output fundamental power sweep is 1:1 as expected, but the gradients of the harmonic power sweeps, particularly that of the third harmonic, are not very close to what is expected. Overall, both the power and phase plots exhibit

much more complex behaviour than that of the GaN HFET, which indicates that the GaAs device would require a more complicated, higher order polynomial transfer function to accurately describe it.

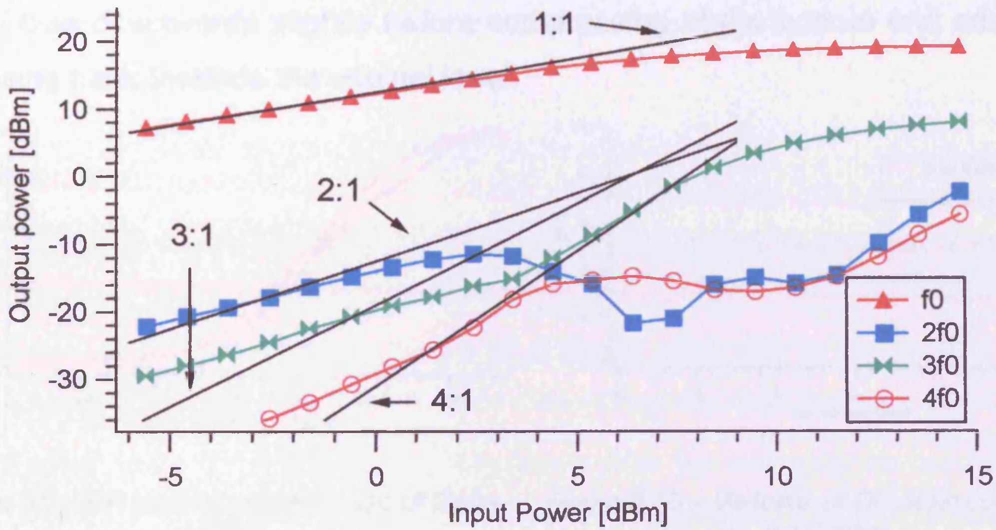


Figure 5.8(a) – Power sweep of a Filtronic GaAs pHEMT

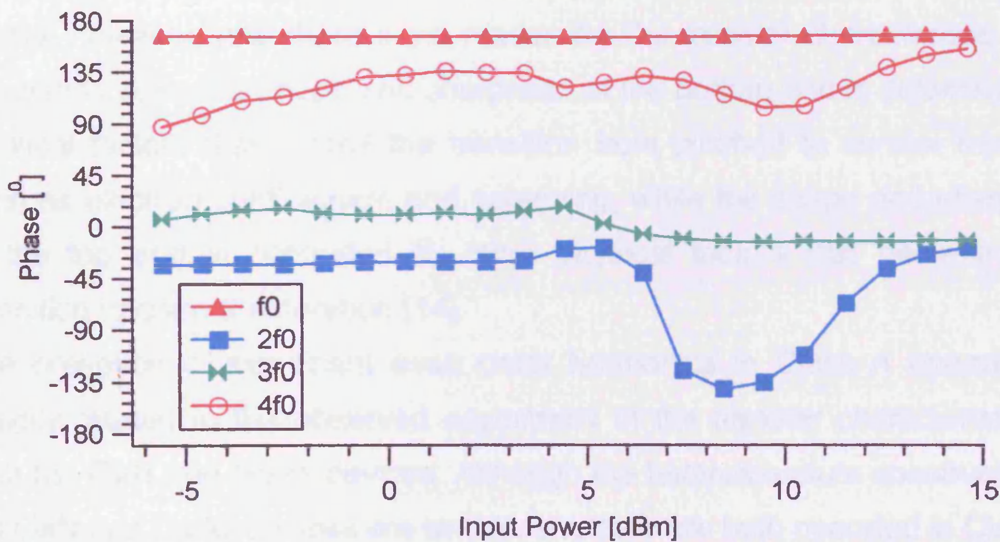


Figure 5.8(b) – Phase behaviour of a Filtronic GaAs pHEMT

Figure 5.9 shows the measured RF I-V transfer characteristic while Figure 5.10 shows the variation of drain current with input drive level. The asymmetry

of the RF I-V transfer characteristic is interesting and indicates that this is a characteristic common to HFET structures, while the plot of drain current is also very similar to that of the GaN device and shows that this device is also saturating at the top end of the transfer characteristic first. This causes it to self-bias downwards slightly before compressing at the bottom end and self-biasing back towards the original level.

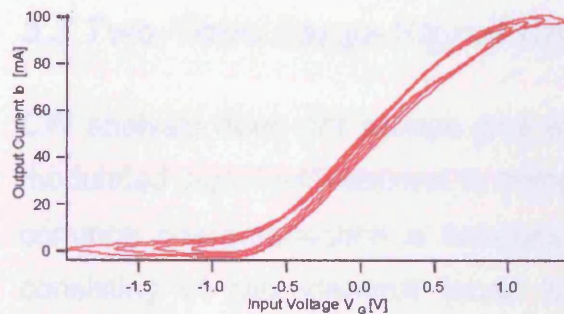


Figure 5.9 – RF transfer characteristic of GaAs pHEMT

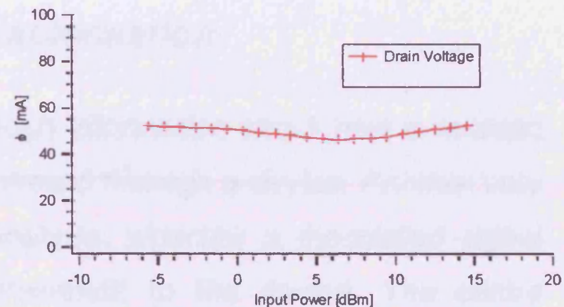


Figure 5.10 – Variation of DC drain current with input drive

The RF transfer characteristic is clearly not symmetric and the transition part of the plot at the top end is not as 'sharp' as that at the bottom end. This is normal however, and there is no reason for the transfer characteristic to be symmetric since the shape and sharpness of the bottom end is determined by physical factors that control the transition from pinchoff to current transport such as electron confinement and screening, while the shape and sharpness of the top end is controlled by other physical factors that determine the transition to current saturation [14].

The presence of significant even order harmonics in Class A operation is directly related to the observed asymmetry in the transfer characteristics of both the GaN and GaAs devices. Although the heterostructure construction of the GaN and GaAs devices are similar, and they are both operated in Class A, it is clear that their harmonic power and phase behaviour is not very similar. In particular, the behaviour of the GaAs device is more complex and would require a more complex, higher order polynomial function to fully describe it. The relative absence of AM-PM in the third harmonic of the GaN device is

interesting because it is the odd order terms in the power series that create both the odd harmonics and the intermodulation distortion products that affect system linearity. Therefore, a lack of AM-PM in the odd harmonics should also translate into a lack of AM-PM in the IM3 products and result in a device that is relatively easy to linearize using predistortion techniques.

5.3 Two-Tone Large-Signal Characterisation

CW analysis does not always give enough information about how a realistic modulated signal will respond to being passed through a device. Another very common characterisation is two-tone analysis, whereby a modulated signal consisting of two identical tones is presented to the device. The centre frequency and tone spacing of the tones can be selected to emulate the bandwidth of a more complex modulated signal. While this is not a complete representation of a multi-tone signal, it provides a much simpler analysis and is good enough to provide some important figures of merit to compare different devices.

If the power series of Equation 5.1 is again used to represent the transfer function of a device, and $v_i(t) = A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t)$, the output becomes:

$$i_o(t) = a_0 + a_1[A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t)] + a_2[A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t)]^2$$
$$+ a_3[A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t)]^3 + \ldots \qquad (5.5)$$

When expanded to the third order, Equation 5.5 reveals the multitude of frequency components and their magnitudes, which are produced by the two-tone excitation of a nonlinear device:

$$\omega_1, \omega_2 : \; (a_1A_1 + \frac{3}{4}a_3A_1^3 + \frac{3}{2}a_3A_1A_2^2)\cos(\omega_1 t) + (a_1A_2 + \frac{3}{4}a_3A_2^3 + \frac{3}{2}a_3A_1^2A_2)\cos(\omega_2 t) \qquad (5.6)$$

$$\omega_1 \pm \omega_2 : \; a_2A_1A_2\cos(\omega_1 + \omega_2)t + a_2A_1A_2\cos(\omega_1 - \omega_2)t \qquad (5.7)$$

$$2\omega_1 \pm \omega_2: \quad \frac{3a_3A_1^2A_2}{4}\cos(2\omega_1+\omega_2)t + \frac{3a_3A_1^2A_2}{4}\cos(2\omega_1-\omega_2)t \qquad (5.8)$$

$$2\omega_2 \pm \omega_1: \quad \frac{3a_3A_1A_2^2}{4}\cos(2\omega_2+\omega_1)t + \frac{3a_3A_1A_2^2}{4}\cos(2\omega_2-\omega_1)t \qquad (5.9)$$

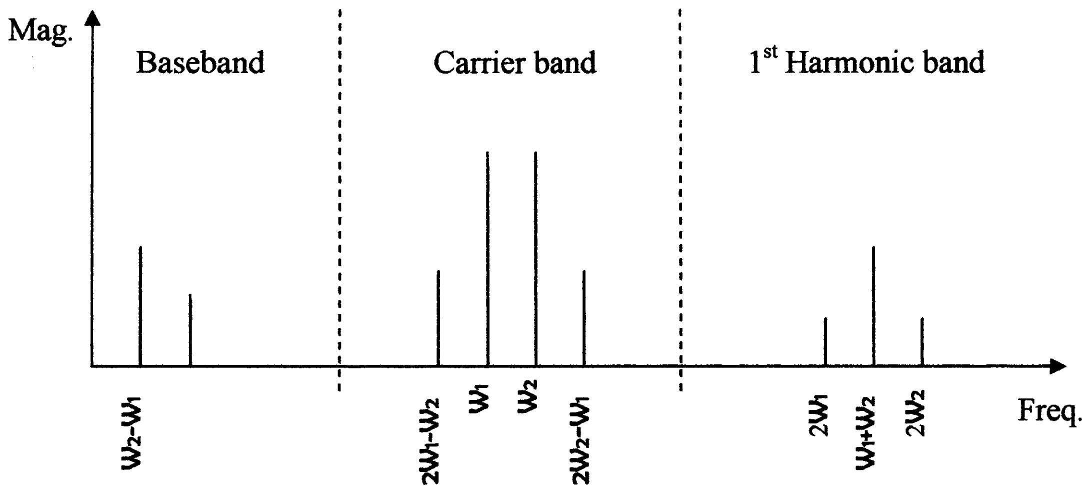


Figure 5.11 – Two-tone output frequency components

The output contains intermodulation products that are not harmonics of the input signals, as illustrated graphically in Figure 5.11. In terms of linearity, the most important frequency components are those at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$. These are the third order intermodulation products (IM3s) and result in intermodulation distortion that cannot be filtered. It is therefore important that these IM3s are as small as possible to reduce adjacent channel interference and increase linearity.

This led to the definition of the third order intercept point (IP3), which is the theoretical point at which the IM3 distortion products are equal in power to the carrier tones. The input power at this point is called the input IP3 (IIP3) and the output power is called the output IP3 (OIP3). The amplitude of both the carriers and IMD products saturate with increasing input power, so in most circuits the IM3s can never reach the same magnitude as the carriers. Therefore the third order intercept point is an extrapolated point.

5.3.1 Two-Tone Measurements

The two-tone measurement setup described in chapter 3 was used to analyze the GaN devices under two-tone excitation. In an attempt to directly compare the CW and two-tone measurements, all bias conditions are identical to those for the CW measurements described above. All harmonic and high frequency mixing components are again passively terminated into a broadband 50Ω load impedance, but the baseband components are terminated with a short circuit since they do not exist under CW excitation and because it is general practice in circuit design to short these components. The measured output intermodulation product power, and phase, versus input power are shown in Figures 5.12(a) and 5.12(b) respectively. The two carrier tones, ω_1 and ω_2, are *830MHz* and *840MHz* respectively, and are supplied by the same signal generator. The input power is the combined available source power from both tones, with each tone having equal power, and the phase components are referenced to the phase of the input carrier ω_1.

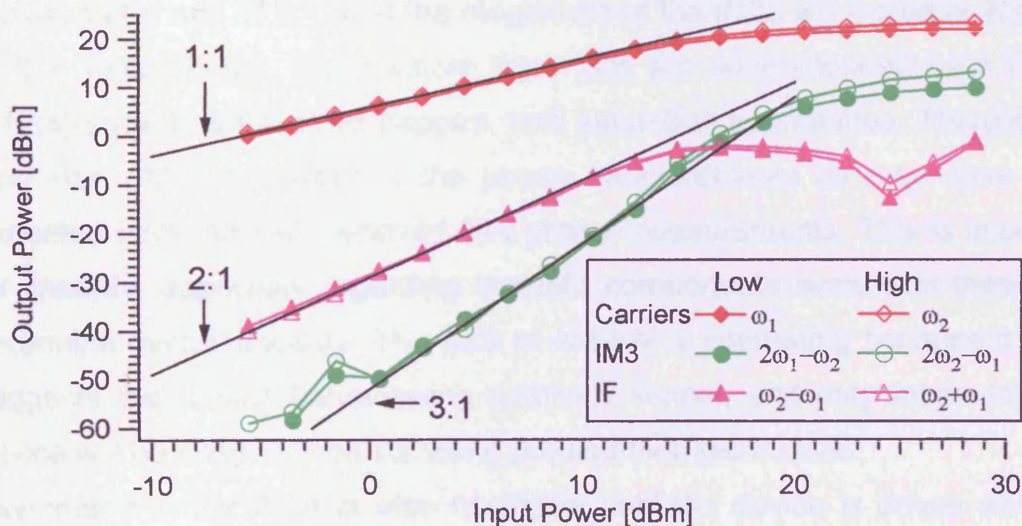


Figure 5.12(a) – Output power of carrier tones and IMD products due to two-tone excitation

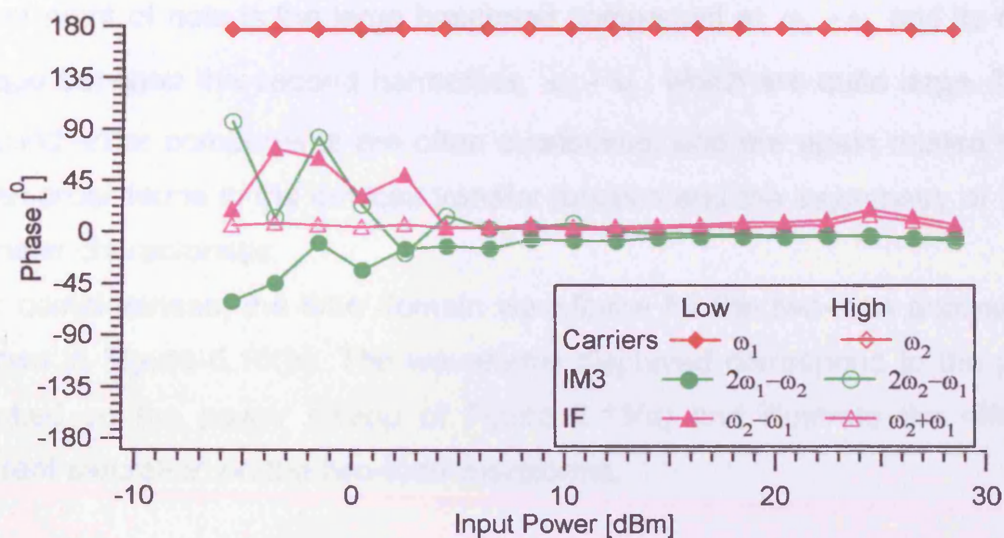


Figure 5.12(b) – Relative phase of output carrier tones and IMD products

Equations 5.6-5.9 indicate that if the two input tones are equal, the magnitude of the output carrier tones will increase linearly at a rate of 1:1, while the magnitude of the baseband (IF) and second harmonic components will increase at a rate of 2:1, and the magnitude of the IM3s will increase at a rate of 3:1. Interestingly, the gradient lines that are superimposed onto Figure 5.12(a) show that this does happen, until saturation approaches. There is also very little AM-PM present in the phase measurements as may have been expected from the well behaved CW phase measurements. This is important for linearity, especially regarding the IM3 components since it is these that determine overall linearity. This lack of AM-PM is interesting because it again suggests the device behaviour is relatively simple, and may mean that the device will be easy to linearize using predistortion techniques.

Asymmetry in the IM3s is also negligible until the device is driven well into compression, at which point other factors such as mixing terms start to affect the IM3s. For example, IM5 products affect IM3s significantly in compression and saturation regimes [13]. This symmetry of the IM3s is also important for linearity because if they are asymmetric, linearization is more difficult [17]. The

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

other point of note is the large baseband component at $\omega_2 - \omega_1$ and its mirror image between the second harmonics, $\omega_2 + \omega_1$, which are quite large. These second order components are often overlooked, and are again related to the even order terms in the devices transfer function and the asymmetry of its I-V transfer characteristic.

For completeness, the time domain waveforms for the two-tone analysis are shown in Figure 5.13(b). The waveforms displayed correspond to the points marked on the power sweep of Figure 5.13(a) and illustrate the effect of current saturation on the two-tone waveforms.

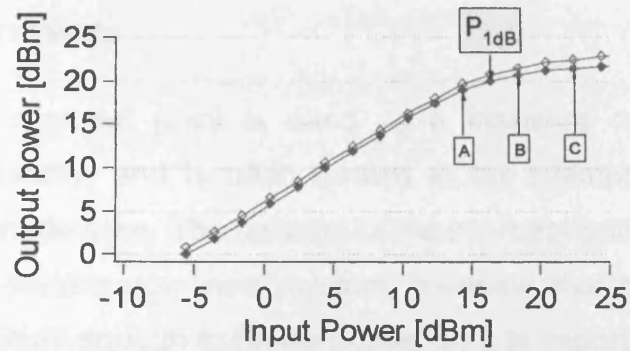


Figure 5.13(a) – Two-tone power sweep

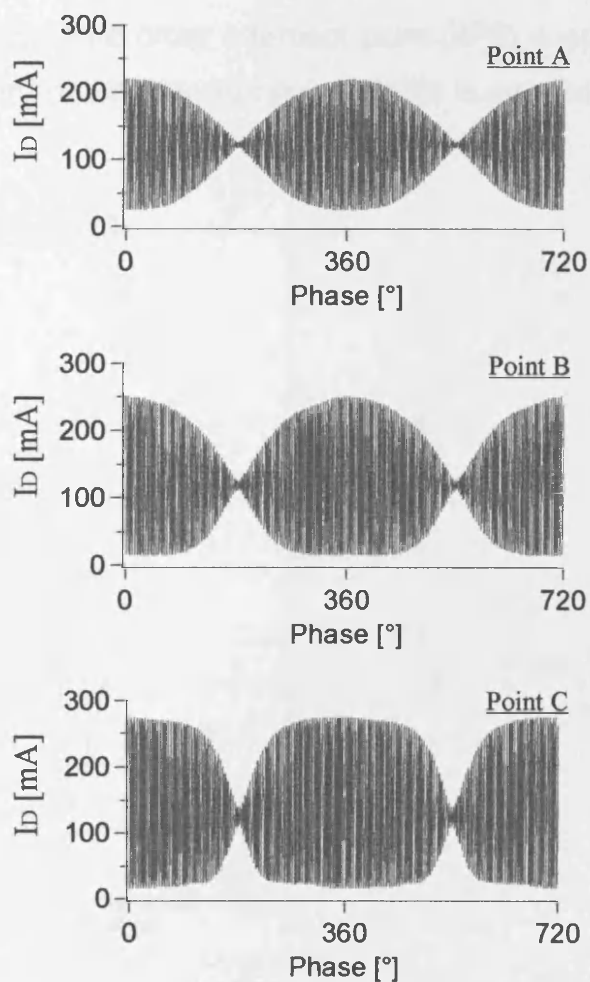


Figure 5.13(b) – Output current RF waveform measurements corresponding to different points on the power sweep

5.3.2 IP3 Measurements

The third order intercept point is used as a measure of intermodulation distortion and linearity, and is often quoted in an attempt to compare the linearity of different devices. The concept of the intercept point is based on the assumption of a weakly non-linear system, meaning that higher order non-linear terms are small enough to be negligible, so it is important to extrapolate from points at low drive levels where higher order effects are not significant. This extrapolation is illustrated in the IP3 measurement of Figure 5.14. This shows that the input third order intercept point (IIP3) is approximately 26dBm, and the output third order intercept point (OIP3) is approximately 34.5dBm.

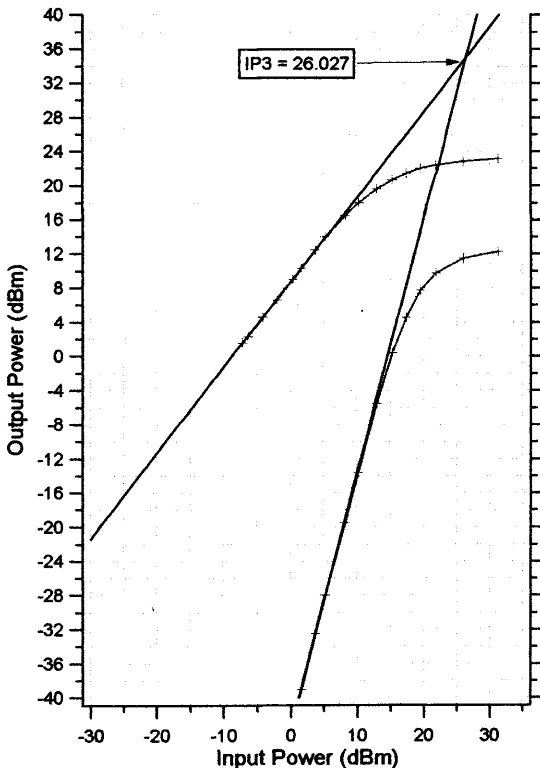


Figure 5.14 – Two-tone IP3 measurement

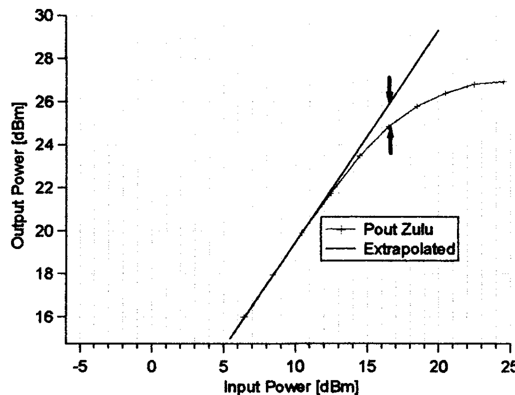


Figure 5.15 – CW P_{1dB} measurement

These figures in themselves are not much use until they are compared to the 1dB compression point of the CW measurement under identical bias

conditions. A general rule of thumb is that the IIP3 should be approximately 9.6dB higher than the 1dB compression point (P_{1dB}) under CW excitation [13]. A high IP3 is desirable because it indicates that the third order products that create intermodulation distortion are more suppressed relative to the carrier.

It has already been determined that the 1dB compression point of the CW power sweep occurs at an input power of 16.5dB. This is confirmed by the close-up view of the CW power sweep in Figure 5.15. The difference between the measured P_{1dB} and the IIP3 is therefore approximately 9.5dB. This is close enough to the desired figure of 9.6dB to be within the tolerance for measurement and extrapolation errors and the obvious conclusion is therefore that the device is well behaved and suitable for power amplifier applications.

A simple mathematical analysis can show that any weakly non-linear device that is third order dominant should conform to the same 9.6dB relationship. The relationship between the magnitude of the P_{1dB} and the *IIP3* can be derived as follows:

The P_{1dB} is defined as the input signal that causes the small-signal gain to drop by 1dB and the input amplitude, A_{1dB}, at that point can be calculated from the magnitude of the fundamental component of Equation 5.3:

$$20\log\left|a_1 + \frac{3}{4}a_3 A_{1dB}^2\right| = 20\log|a_1| - 1dB \qquad (5.10)$$

$$A_{1dB} = \sqrt{0.145\left|\frac{a_1}{a_3}\right|} \qquad (5.11)$$

If $A_1 = A_2 = A$, the two-tone equation of 5.5 can be simplified to:

$$i_o(t) = \left(a_1 + \frac{9}{4}a_3 A^2\right)A\cos\omega_1 t + \left(a_1 + \frac{9}{4}a_3 A^2\right)A\cos\omega_2 t$$
$$+\frac{3}{4}a_3 A^3\cos(2\omega_1 - \omega_2)t + \frac{3}{4}a_3 A^3\cos(2\omega_2 - \omega_1)t + \ldots \qquad (5.12)$$

From this, a simple expression can be derived for the magnitude of the *IIP3*. Assuming that $a_1 >> 9a_3A^2$, the input amplitude, A_{IIP3}, for which the output components at ω_1 and ω_2 have the same amplitude as those at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ is given by:

$$|a_1|A_{IIP3} = \frac{3}{4}|a_3|A_{IIP3}^3 \quad (5.13)$$

Therefore the *IIP3* is:

$$A_{IIP3} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|} \quad (5.14)$$

From Equations 5.11 and 5.14, the relationship between the P_{1dB} and *IIP3* can be calculated:

$$\frac{A_{IIP3}}{A_{1dB}} = \frac{\sqrt{4/3}}{\sqrt{0.145}} \approx 9.6dB \quad (5.15)$$

Most semiconductor transistors can be described as weakly non-linear and conform roughly to the behaviour of Equation 5.2 [13], so that this 9.6dB relationship is generally quite accurate and is an intrinsic property of the device. Unfortunately, at the time of this work it was not possible to perform two-tone measurements on the GaAs pHEMT device for comparison, but the overall linearity performance of the GaN devices as determined by the IP3 measurement, is likely to be no better or worse than any other device technology that is weakly non-linear.

The derivation of Equation 5.15 also suggests that where a weakly non-linear device has a relationship between its measured *IIP3* and P_{1dB} that is greatly different than 9.6dB, it is either because of the effects of the bias/matching circuitry, or due to incorrect extrapolation - not extrapolating from low enough drive levels. For example, the ratio will normally be larger than 9.6dB where

linearization techniques are applied to improve linearity. Since these measurements are performed on-wafer at the device plane it should be expected that the result would be very close to theory.

Overall, the two-tone power and phase behaviour of the GaN device confirms the observations from the CW measurements that its behaviour is close to ideal and can be described by a relatively simple low order transfer function. There is a high degree of consistency between the two-tone and CW results, and the lack of AM-PM in the third harmonic translates into a lack of AM-PM in the IM3 components of the two-tone results. This may be expected since it is the odd order terms in the power series that create both the odd harmonics and the intermodulation distortion products that affect overall linearity. This combined with the fact that the device can be described by a low order power series, means that it may be relatively easy to linearize.

5.3.3 Comparison of Devices with Different Epitaxial Layer Design

As shown previously, the structure of the QinetiQ GaN/AlGaN HFET is very simple and can consist of only two or three epitaxial layers since no doping layer is necessary to create the 2DEG. However, a common addition to the basic structure is a thin Aluminium Nitride (AlN) layer between the AlGaN barrier layer and the GaN buffer layer [18], [19]. This structure is shown in Figure 5.16. Figure 5.17 illustrates that the AlN bandgap is much larger than either AlGaN or GaN so that it forms a large barrier which has been shown to help exclude carriers from the AlGaN layer and reduce alloy scattering in the 2DEG channel and therefore improve electron mobility by up to 10%. It was also hoped that a side-effect of this is a possible improvement in linearity.

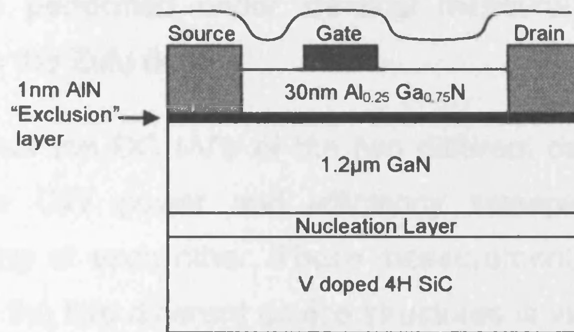


Figure 5.16 – Device structure with additional AlN layer

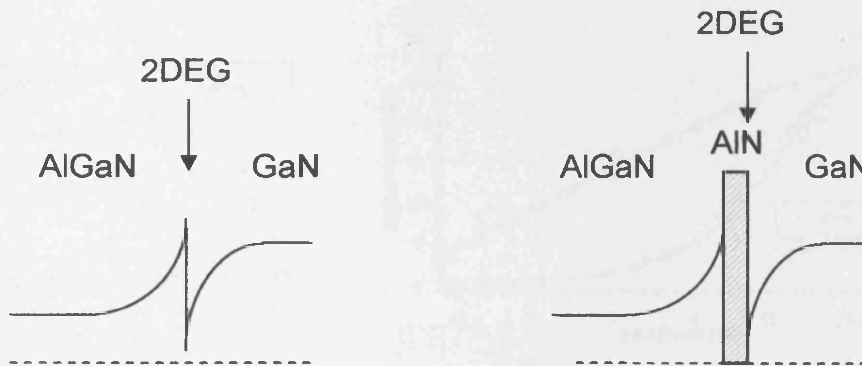


Figure 5.17(a) – GaN HFET without AlN barrier layer

Figure 5.17(b) – GaN HFET with AlN barrier layer

To test the effect of the AlN layer on device performance, devices were produced by QinetiQ with and without the added layer. To maintain uniformity between both types of device they are produced side by side on the same wafer run. The devices with the AlN layer have been designated the name 'Yankee', while the devices with no AlN layer are called 'Zulu'. The devices are in theory identical apart from the additional *1nm* AlN layer. All measurements presented in this chapter thus far have been performed on the Zulu devices (no AlN layer). To compare the performance of the two device structures, an analysis of the Yankee device is presented, where all

measurements are performed under identical measurement conditions to those performed on the Zulu device.

Figure 5.18(a) shows the DC I-V's of the two different devices while Figure 5.18(b) shows the CW power and efficiency sweeps of both devices superimposed on top of each other. These measurements clearly show that the performance of the two different device structures is virtually identical and indicates that the AlN layer has no noticeable effect on device performance. The slight difference in the two measurement sets is no greater than the difference observed from device to device, which is due to the non-uniformity of the wafers, and is not necessarily due to the presence of the AlN layer.

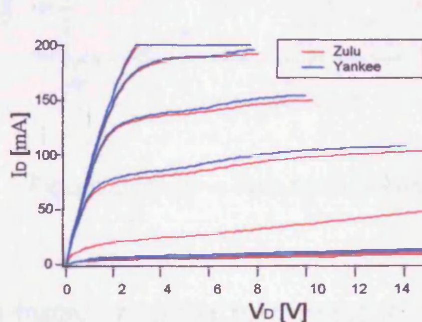


Figure 5.18(a) –DC I-V's of Device with and without AlN Layer

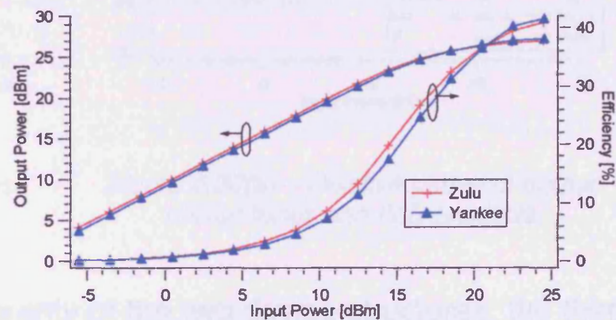


Figure 5.18(b) – CW Power, Gain and Efficiency performance of both devices

The CW harmonic power and phase plots of Figure 5.19(a) and (b) are also presented for the Yankee device, as are the two-tone power and phase plots of Figure 5.20(a) and (b). These measurements are directly comparable to those presented earlier for the Zulu device, and again verify that there is virtually no measurable difference between the two device structures.

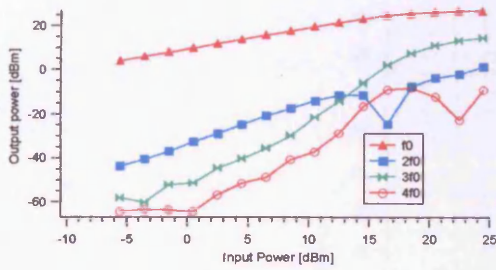


Figure 5.19(a) – CW power sweep

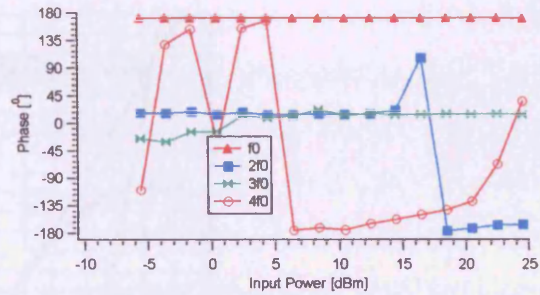


Figure 5.19(b) – Output phase behavior

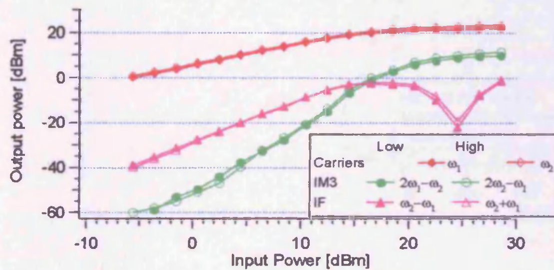


Figure 5.20(a) – Two-tone power sweep

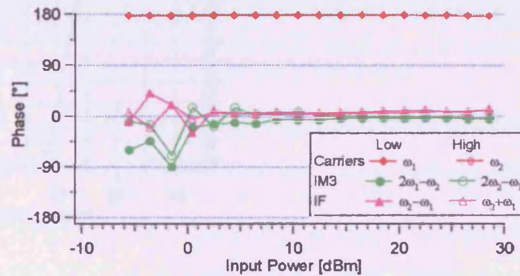


Figure 5.20(b) – Relative phase of output carrier tones and IMD products

Finally, in order to compare the linearity of the two device structures, the third order intercept points were extrapolated and compared and the results can be seen in Figure 5.21. There is a very small difference between the IP3 points of the two different devices, but this is within measurement tolerance limits.

The power sweeps of the two devices are virtually identical, so the IP3 and 1dB compression points are also almost identical. Therefore, the addition of the AlN layer provides little or no measurable improvement in linearity. Any Improved linearity would have to be a side effect of an altered device transfer function. It is possible that the extra current confinement and carrier mobility provided by the AlN layer sharpens up the bottom end of the transfer characteristic by improving the transition from pinchoff to current conduction. However it is likely that any linearity improvement created by this is swamped by the other, more dominant non-linear mechanisms.

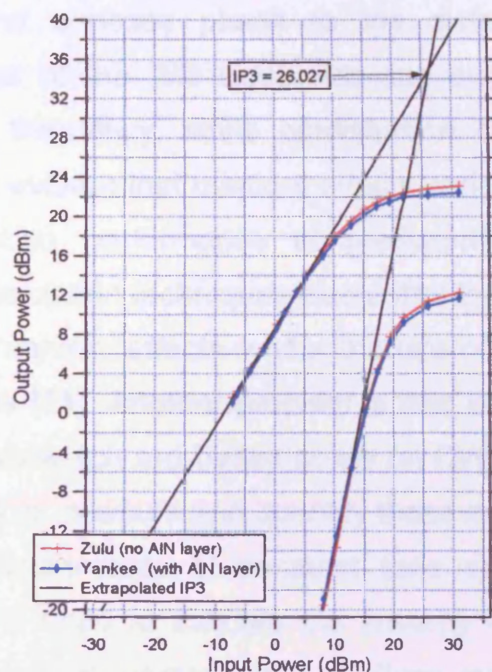


Figure 5.21 – Comparison of IP3 points of devices with and without AlN layer

5.4 Memory Effects in Power Amplifiers

As mentioned earlier, the non-linear nature of the device introduces distortion products such as IM3s that cannot be easily filtered. These distortion products need to be minimised to avoid interference with the transmitted signal and the signals in the adjacent channels. The distortion components determine the linearity of the device and are easily measured via ACPR or two-tone measurements. However, it is known that the magnitude and phase of these distortion components can be significantly affected by varying modulation frequency, or tone spacing, of the transmitted signal [20], [17]. This frequency dependent behaviour is known as memory effects.

These memory effects reduce overall linearity and more importantly, can make linearization much more difficult and seriously limit the maximum achievable cancellation performance of linearization techniques. For example, predistortion linearization works by producing signal components that are of

equal amplitude and opposite phase to the distortion products. If the amplitude and phase of the IM3 sidebands can change continuously with varying modulation frequency, while predistortion signals usually have a constant phase, it is evident that memory effects seriously limit the maximum achievable cancellation performance of predistortion. Unfortunately, the majority of other linearization techniques also suffer from similar problems and as a consequence, memory effects render the use of linearization ineffective for many applications [21]. Another problem is that any memory effects may also invalidate models which are based solely on CW or two-tone analysis. It is therefore desirable to measure and quantify these memory effects.

Basic two-tone measurements of constant tone spacing, such as those presented earlier, are used to indicate the linearity of a device but cannot determine the presence of memory effects. Since memory effects vary with modulation frequency, they are generally measured and characterised via two-tone measurements of varying tone spacing, and are observed as asymmetry in the magnitude and/or phase of the intermodulation products [17], [22].

5.4.1 Origins of Memory Effects

Although a full and detailed treatment of the origins of memory effects is beyond the scope of this work, it is generally recognised that memory effects in power amplifiers can be split into two main categories: electrical memory, and thermal memory. In highly dispersive devices such as GaN HFETs it is also likely that there are some dispersion-related memory effects present, due to electron trapping [23].

It is thought that electrical memory effects are caused by baseband, fundamental, or second harmonic impedance variations at different modulation frequencies [17], [24]. The power amplifier polynomial transfer function of Equation 5.1 indicates that the distortion products are not a function of tone spacing. However, Vuolevi [17] suggests that because there is always more than one non-linear mechanism present in a power amplifier,

Equation 5.1 does not hold exactly and the PA must be treated as a cascade of at least two polynomials. Because of this there is significant mixing between the spectral components that combine to create additional distortion in the IM3s, especially when the device is operated in or near compression. The greatest source of additional distortion is third order mechanisms which are affected by carrier impedances. Second order mechanisms generated by the baseband and second harmonic components, and which are affected by their impedances, also have a significant effect on IM3 distortion. It is therefore important to analyse the effect of the baseband and carrier impedances on the IM3 distortion.

Thermal memory effects are caused by electro-thermal coupling, which is affected by low modulation frequencies of less than 1MHz [22], [25]. The temperature of a device is determined by the electrical power that is dissipated in it and since the state of the device is continuously changing due to the modulated signal, temperature variations arise. The heat flow from the device is described by the thermal impedance which is affected by the device packaging and heat sink as well as the transistor itself. The thermal impedance of the device is not purely resistive but instead it forms a distributed low-pass filter with a wide range of time constants [21]. This means that the temperature changes caused by dissipated power do not occur instantaneously, and frequency-dependent phase shifts always exist.

Trap-related memory effects are similar to thermal memory effects in that they appear at low modulation frequencies which have a period similar to the trapping time constants present in the device. These devices are likely to have trap time constants of around 1ms and higher, meaning that they will be activated at modulation frequencies below 1KHz [26]. Because these devices have been shown to suffer significantly from trap related dispersion it is reasonable to assume that they may also suffer from trap related memory effects at very low modulation frequencies [23].

5.4.2 Envelope Analysis of Two-Tone Results

Since the presence of memory effects can cause significant problems in power amplifiers, it is important to analyse how the output of the device changes with modulation frequency. Ideally, the device would be tested across all modulation frequencies from zero modulation (CW) to several MHz, since this is the range of modulation frequencies that can exist in many modern modulation schemes. Unfortunately, the MTA-based measurement system described in chapter 3 is only capable of measuring two-tone signals with tone spacings that are multiples of 10MHz. Therefore, this two-tone system is not ideal for the analysis of memory effects.

However, the CW and two-tone measurement results that were presented earlier in this chapter were performed under conditions that are almost identical. The two-tone measurement had a tone spacing of 10MHz which is equivalent to a modulation frequency of 5MHz, so the two measurement points are at the extremes of the desired measurement bandwidth, but it is interesting to compare the two measurements to see if there are any memory effects evident in the device. One way of comparing the measurements is by doing an envelope analysis of the two-tone data. This is an alternative method of data display and interpretation which utilizes a combination of time domain and frequency domain techniques, and is a useful way of viewing the data to qualitatively compare the CW and two-tone behaviour of the device.

Using envelope domain analysis a modulated signal is described as a complex time varying envelope containing both magnitude and phase information of each Fourier component, and the distortion in the output spectrum of the device is quantified in terms of the generated 'envelope harmonic components' around the carrier and its harmonics [10]. The envelope extraction process involves extracting time varying I and Q components that are associated with each envelope harmonic and is briefly described below.

A Fourier transform is first performed on the time domain multi-tone waveform to transform it to the frequency domain. A new frequency domain representation of each harmonic 'band' is then created that contains only the spectral information relating to that particular band, and the spectrum is shifted so that the band is centred around DC. The I and Q envelope components are generated by the summation of the magnitude (ρ) and phase (ϕ) of the distortion products about each harmonic component using Equations 5.16 and 5.17. Each distortion term is calculated between *–m* and *m* about the centre point of the carrier or harmonic frequency component *n*.

$$I_n(t) = \sum_{i=-m}^{m} \rho_{n+1} \cos(i2\Pi F_{\text{mod}} - \phi_{n+1}) \quad (5.16)$$

$$Q_n(t) = \sum_{i=-m}^{m} -\rho_{n+1} \sin(i2\Pi F_{\text{mod}} - \phi_{n+1}) \quad (5.17)$$

These I and Q signal can then be converted into time varying magnitude and phase envelopes using Equations 5.18 and 5.19.

$$A_n(t) = \sqrt{I_n(t)^2 + Q_n(t)^2} \quad (5.18)$$

$$\phi_n(t) = \tan^{-1}\left(\frac{Q_n(t)}{I_n(t)}\right) \quad (5.19)$$

If the envelope extraction is performed on the frequency components of the carrier band, which are shown in Figure 5.11, the time domain a and b waves can be turned into envelope power waveforms. Figure 5.22(a) shows an example of the resulting output power envelope waveform of the carrier band when the device is operated in compression. When the output power envelope is plotted against the input power envelope, it results in an envelope power sweep, as shown in Figure 5.22(b). Since the modulation depth of the two-tone signal is 100%, this envelope power sweep is effectively the same as a CW power sweep except that it is dynamic. It is therefore a very useful way of viewing distortion in the carrier signals.

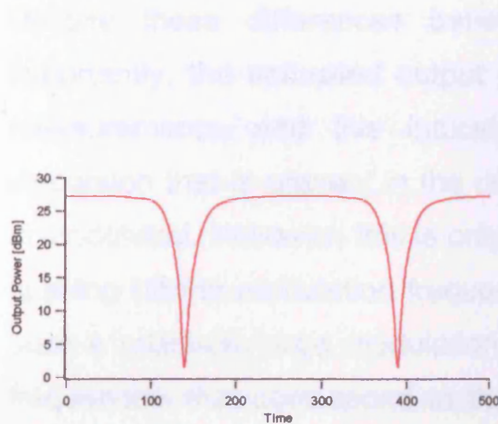


Figure 5.22(a) – Two-tone envelope power sweep

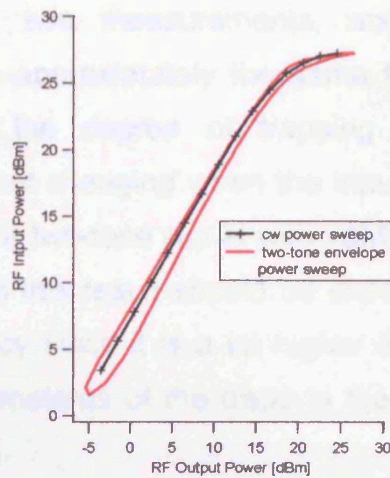


Figure 5.22(b) – Comparison of CW power sweep and two-tone envelope power sweep

Figure 5.22(b) also shows the CW power sweep that was presented earlier in this chapter and it can be seen that there is strong correlation between the two plots, although they are not an exact match. If the device was in an identical state during both measurements, it would be expected that the two plots would line up exactly. However, this is unlikely to ever happen because the device will always be in a slightly different state during the two measurements. In particular, the thermal state of the device will be very different, and the system impedances are not exactly the same during each measurement since additional measurement equipment is required for the two-tone measurement.

However, the impedance terminations of the carrier tones is likely to be the single biggest cause of error. The two-tone plot has some looping in it and this is due largely to the fact that the impedance termination of each carrier tone is slightly different. This is because the passive 50Ω termination is not constant across the relatively large 10MHz bandwidth of the two tones. This is generally considered to be a memory effect and is discussed in more detail later. Additionally, the impedance termination of the CW signal is also different to that of the two carrier tones.

Despite these differences between the two measurements, and most importantly, the saturated output power is approximately the same for both measurements, and this indicates that the degree of trapping related dispersion that is present in the device is not changing when the input signal is modulated. However, this is only true for a two-tone signal with 10MHz tone spacing (5MHz modulation frequency), and this result should be expected at such a relatively large modulation frequency since it is a lot higher than the frequencies that correspond to the time constants of the traps in the device [26].

It is therefore desirable to test the device at multiple modulation frequencies to determine if memory effects, including trapping related memory effects, are present in the device. Unfortunately, the measurement systems described in chapter 3, and used to perform the measurements presented in this thesis thus far, are not capable of measuring at any modulation frequency other than 5MHz. However, recent development of the measurement capabilities at Cardiff University have enable such measurements to be performed, and this new measurement approach is discussed in the next section. Detection of the presence of memory effects is then achieved through analysis of the magnitude of the IM3 components of the two-tone signal. This is generally considered to be the simplest and easiest way to detect the presence of memory effects.

5.4.3 Scope-Based Non-Linear Measurement System

Many measurement systems have been developed to measure the magnitude and phase of two-tone signals of varying modulation frequency [21], [25], [27], and most of these systems use a combination of power meters, phase detectors, spectrum analyzers, or network analyzers to measure the asymmetry created in the IM3 components. As discussed in chapter 3, the time domain waveform measurement approach has established itself as the optimum measurement solution for the analysis of power amplifier

performance because it allows the measurement of error-corrected voltage and current waveforms at the device plane which can then be analyzed in either the time domain or frequency domain.

However, most time domain measurement systems currently developed are only capable of measuring CW signals, and have very limited capability in the measurement of modulated signals. For example, because of the sampling method employed by the MTA-based system described previously, it measures on a fixed frequency grid, where the minimum tone spacing is 10MHz. Therefore it is only possible to measure a two-tone signal with tone spacing that is a multiple of 10MHz, and it is not possible to measure signals with lower tone spacing. Unfortunately, any memory effects present in the device are likely to appear at much lower modulation frequencies than this, and it is not possible to emulate the behaviour of a realistic signal such as WCDMA which has a carrier bandwidth of 3.84MHz, and up to four carriers spaced 5MHz apart.

Therefore there is a necessity for a new method of measuring modulated signals of lower modulation frequency, in the time domain. The solution to the problem is a new time domain measurement system which is based around a sampling oscilloscope and has been in development at Cardiff University for several years. The measurement system architecture is shown in Figure 5.23. This new system works in a similar way to the MTA-based system described earlier and was intended as a replacement for it because the MTA is no longer in production or supported by Agilent, and no replacement instrument has been introduced.

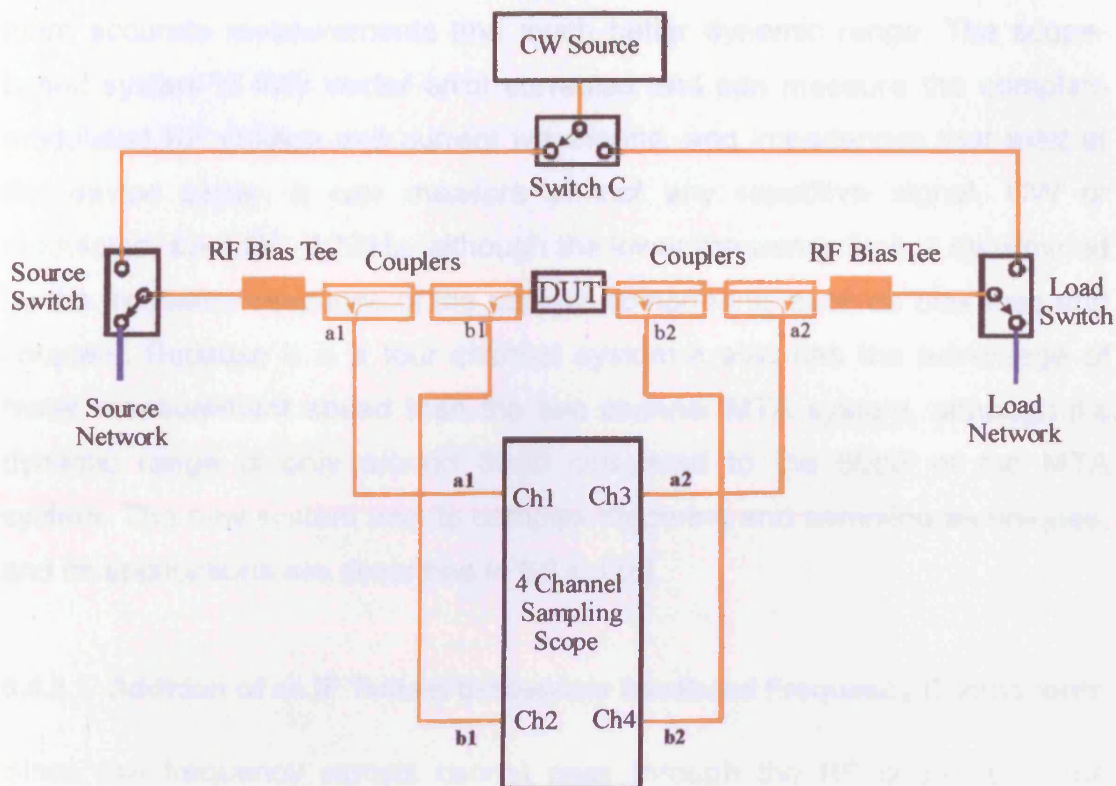


Figure 5.23 – Scope based measurement system

In effect, the architecture and operation of the new scope-based system is identical to the MTA-based system, but it is the way in which the sampling scope makes its measurements that is unique. Like the MTA, the sampling scope can only measure repetitive signals, but unlike the MTA, it does not have internal triggering, and so the triggering is controlled externally. As a useful side effect of this, it becomes possible to take control of the measurement grid on which the scope samples, and engineer the grid so that almost any repetitive signal can be measured, including modulated signals. The new system is very unique and is the first scope-based system of its kind, and the first time domain measurement system capable of measuring modulated signals.

Previously, sampling scopes were not as accurate as the MTA instrument and were not generally used in waveform measurements systems. However, new sampling scopes now have better analog-to-digital convertors which results in

more accurate measurements and much better dynamic range. The scope-based system is fully vector error corrected and can measure the complete modulated RF voltage and current waveforms, and impedances that exist at the device plane. It can measure almost any repetitive signal, CW or modulated, from DC–12GHz, although the lower frequency limit is determined by the frequency response of the passive components such as bias tees and couplers. Because it is a four channel system it also has the advantage of faster measurement speed than the two channel MTA system, although it's dynamic range is only around 50dB compared to the 60dB of the MTA system. The new system and its complex triggering and sampling techniques, and its applications are described in full in [28].

5.4.3.1 Addition of an IF Testset to Measure Baseband Frequency Components

Since low frequency signals cannot pass through the RF couplers of the system, an additional low frequency (IF) test set was developed by the author and added to the system to measure the baseband frequency components created by a two-tone input signal. This is illustrated in Figure 5.24 below. This part of the system differs to that of the MTA-based two-tone system described in chapter 3 because the new architecture splits the RF and IF signals through the RF bias tee before recombining them through a diplexer, allowing the simultaneous measurement of all four travelling waves at both RF and baseband frequencies. Diplexing the coupled RF and IF signals prior to measurement ensures phase coherence is maintained between the measured RF and IF signals, and simplifies the measurement process significantly. The calibration procedure is also simplified as both the IF and RF sections of the system can be calibrated at once, in a similar way to a CW calibration.

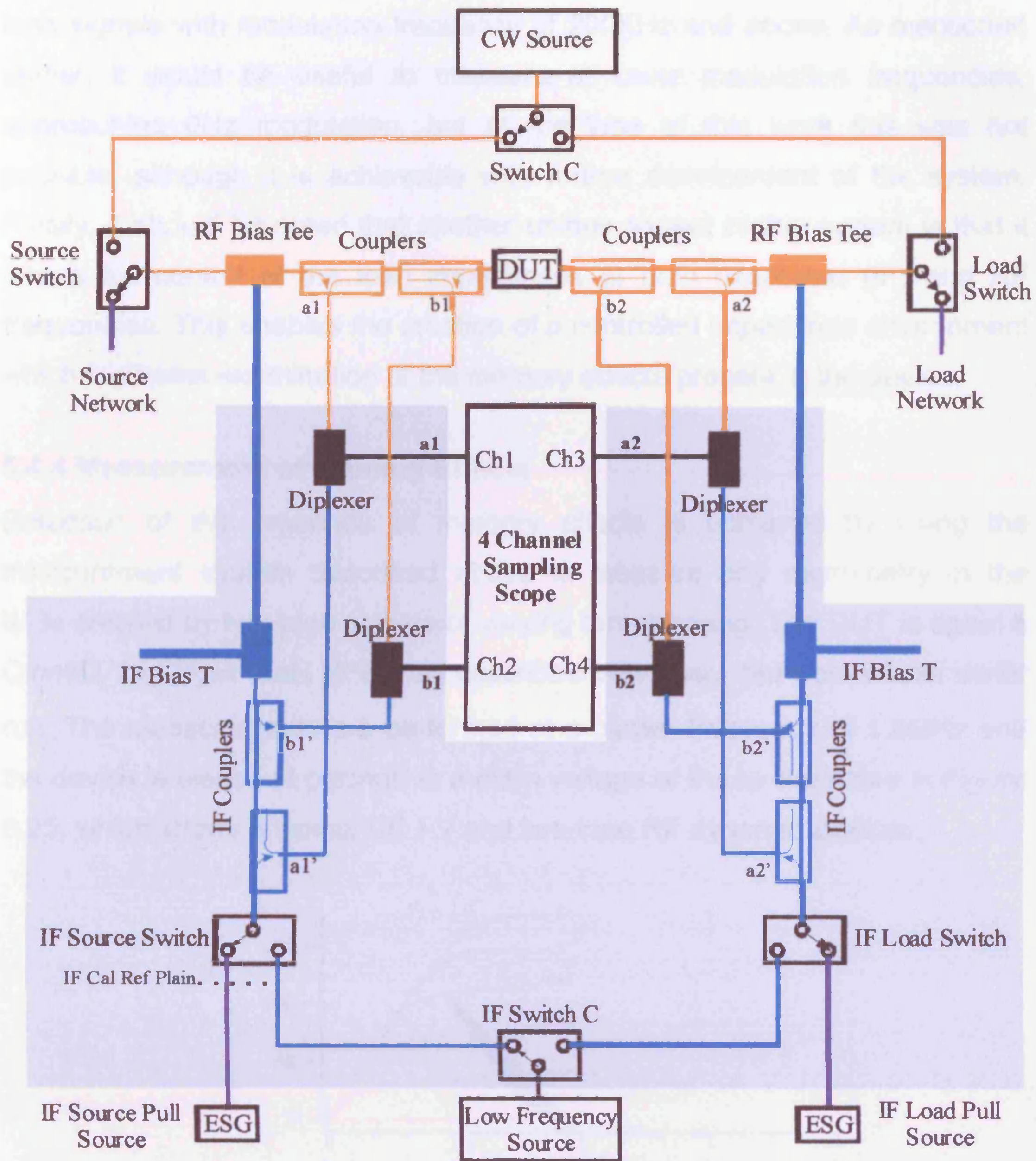


Figure 5.24 – Scope-based modulated measurement system with additional IF test set for simultaneous measurement of baseband components

The RF bandwidth of the system is approximately 500MHz to 12GHz, while the frequency limitations of the IF bias tees and couplers mean that the IF bandwidth is approximately 400KHz to 50MHz. Therefore, in its current guise the lowest measurable baseband frequency is 400KHz. This equates to two-

tone signals with modulation frequency of 200KHz and above. As mentioned earlier, it would be useful to measure at lower modulation frequencies, approaching 0Hz modulation, but at the time of this work this was not possible, although it is achievable with further development of the system. Finally, it should be noted that another unique aspect of this system is that it allows full control of the load impedances at both baseband (IF) and RF frequencies. This enables the creation of a controlled impedance environment which facilitates examination of the memory effects present in the device.

5.4.4 Measurement of Memory Effects

Detection of the presence of memory effects is achieved by using the measurement system described above to measure any asymmetry in the IM3s created by two-tone signals of varying tone spacing. The DUT is again a QinetiQ $2\times125\mu m$ GaN HFET, as described previously, but from a later wafer run. The measurements are performed at a carrier frequency of 1.8GHz and the device is biased at pinchoff at a drain voltage of 8V as illustrated in Figure 5.25, which shows a typical DC I-V and two-tone RF dynamic loadline.

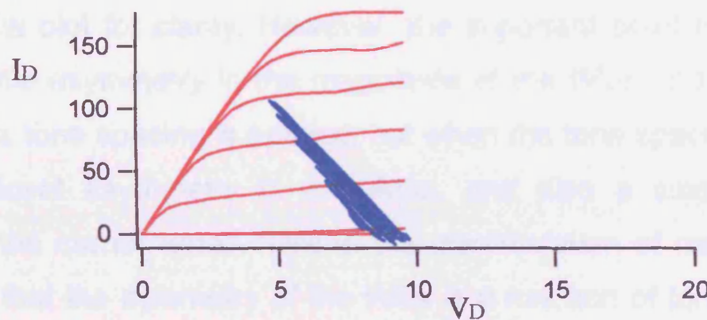


Figure 5.25 – DC I-Vs and RF dynamic loadlines

Figures 5.26(a) shows a two-tone power sweep illustrating the output carrier powers and IM3 powers for a tone spacing of 400KHz, while Figure 5.26(b) shows the same measurements when the tone-spacing is 20MHz. It should be noted that for measurement consistency, all frequency components are

terminated into passive 50Ω loads, and the magnitude of the two input carrier tones is monitored and adjusted by the controlling software so that any asymmetry is automatically corrected. Such errors are created by the passive components that the signal must pass through, as well as the source amplifier.

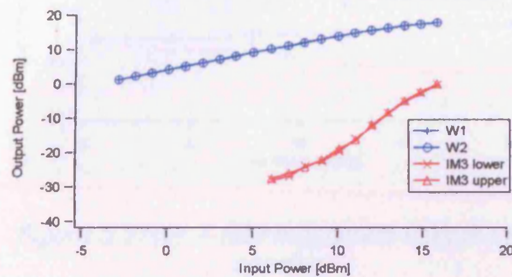


Figure 5.26(a) – Two-tone power sweep at 400KHz tone spacing

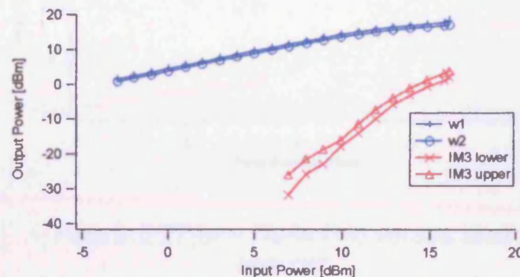


Figure 5.26(b) – Two-tone power sweep at 20MHz tone spacing

These results correlate very well with the measurements presented earlier that were measured using the MTA-based two-tone measurement system but the first several points in the IM3 measurements are very noisy due to them being near the bottom end of the dynamic range of the system and have been omitted from the plot for clarity. However, the important point to note is that there is very little asymmetry in the magnitude of the IM3s and carrier tones when a 400KHz tone spacing is applied, but when the tone spacing is 20MHz, there is significant asymmetry in the IM3s, and also a small amount of asymmetry in the carrier tones. This is the manifestation of memory effects and illustrates that the symmetry of the IM3s is a function of tone spacing, or modulation frequency.

To further examine the problem it is useful to measure the asymmetry of the IM3 components as a function of tone spacing. This is illustrated in Figures 5.27(a) and 5.27(b). For measurement consistency, all measurements are performed at a constant input power which is in the linear region, approximately 3dB below the 1dB compression point. These plots again show

that the asymmetry of the IM3 components increases with increasing tone spacing. This result is well documented [21], [25], but the exact origins of the effect is not so well understood.

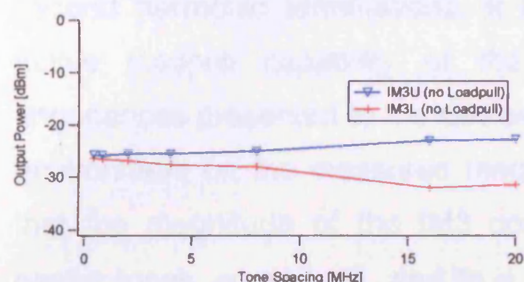


Figure 5.27(a) – IM3 magnitude versus tone spacing

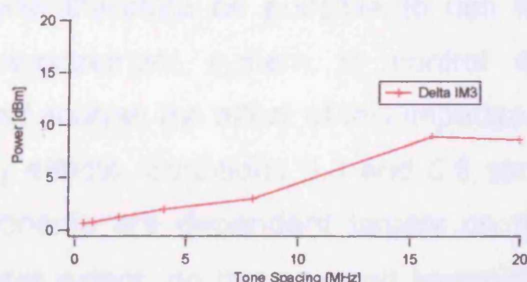


Figure 5.27(b) – Delta IM3 versus tone spacing

Because of the lack of asymmetry at the lower tone spacings it is reasonable to assume that there is little or no thermal or trapping related memory effects present within the measurement bandwidth. It has been shown that thermal memory effects become apparent at tone spacings lower than 1MHz [25]. Additionally, because the device is very small and is on a SiC substrate of high thermal conductivity, any thermal memory effects may be expected to be negligible compared to those experienced in high power amplifiers such as Si LDMOS where packaging also has a significant impact on memory effects.

The device also contains electron traps so it is possible that it would be susceptible to memory effects caused by these traps. However, the trapping time constants in these devices is expected to be around 1ms and higher [26], meaning that they will be activated at modulation frequencies below 1KHz. For these reasons it is reasonable to assume that the memory effects evident here are due solely to electrical memory, and although other sources of memory do undoubtedly exist in the device, they cannot be measured within the bandwidth of the system.

5.4.5 Effect of Carrier Impedance Terminations

Vuolevi [17] indicates that electrical memory effects are due to the nodal impedances of the device and circuit, particularly at the baseband, carrier and second harmonic terminations. It should therefore be possible to use the active loadpull capability of the measurement system to control the impedances presented to the device and analyse the effect of this impedance environment on the measured memory effects. Equations 5.8 and 5.9 show that the magnitude of the IM3 components are dependent largely on the carrier tones ω_1 and ω_2, and to a lesser extent, on their second harmonics $2\omega_1$ and $2\omega_2$. Therefore, since the carrier tones will always have the largest magnitude, it is reasonable to assume that if electrical memory effects are caused by the impedance variations across the bandwidth of these frequency components, the most significant contribution to the memory effects will come from the carrier tones.

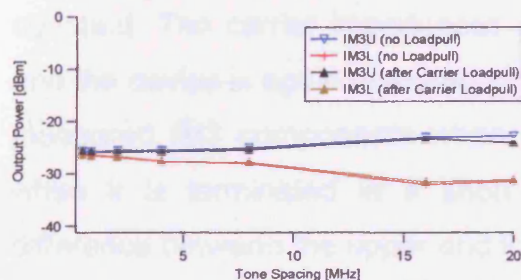


Figure 5.28(a) – IM3 asymmetry before and after the two carrier tones are loadpulled to the same impedance

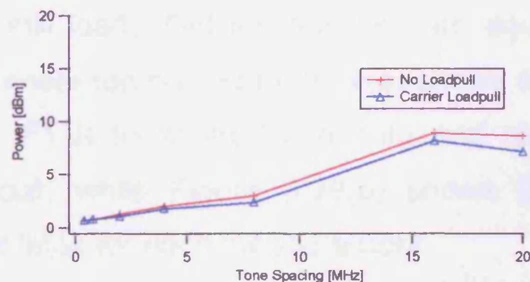


Figure 5.28(b) – Delta IM3 before and after the two carrier tones are loadpulled to the same impedance

To analyse the effect of the relative carrier impedances on the memory effects present in the device, the measurement above is repeated but with the ω_1 carrier tone loadpulled to exactly the same impedance as the ω_2 carrier tone, while all other parameters are held constant. The device is again operated in the linear region, and all other frequency components are terminated into passive 50Ω loads. Figures 5.28(a) and 5.28(b) show the resultant asymmetry of the IM3s before and after the two carrier tones are loadpulled to

the same impedance, and it is evident that there is a small improvement in IM3 asymmetry.

5.4.6 Effect of Baseband Impedance Terminations

In the previous measurements all baseband frequency components were passively terminated into a 50Ω load, but it has been shown that for optimum device operation, the $\omega_2 - \omega_1$ baseband frequency component (IF1) should be presented with a short circuit termination [29]. It is standard practice in circuit design to present a short circuit to all baseband components but here we will focus on the IF1 component since it is by far the most significant baseband component.

To analyse the effect of the IF1 impedance on the memory effects present in the device, the measurement above is repeated but with the IF1 impedance termination loadpulled to a short circuit, while all other variables are held constant. The carrier impedances are still loadpulled so that they are equal and the device is again operated in the linear region. Figure 5.29(a) shows the measured IM3 components when the IF1 is terminated into 50Ω and also when it is terminated in a short circuit, while Figure 5.29(b) shows the difference between the upper and lower IM3s for each measurement.

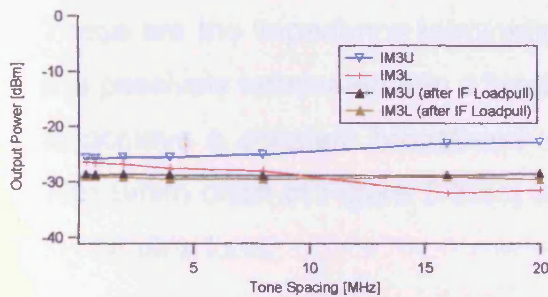


Figure 5.29(a) – Comparison of IM3 magnitude versus tone spacing, when IF impedance is 50Ω and a short circuit

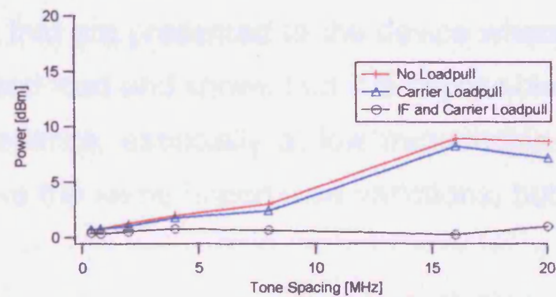


Figure 5.29(b) – Delta IM3 versus tone spacing when IF impedance is 50Ω and a short circuit

These results show that IM3 asymmetry is improved significantly by terminating the IF1 component into a short circuit, and it can also be seen that the absolute magnitude of the IM3s are suppressed by a couple of dBs. After termination of the IF1 component into a short circuit there is still a small amount of asymmetry present in the IM3s, which is likely to be due to other mechanisms such as impedance variations across the second harmonic and other baseband frequencies.

5.4.7 The Importance of Baseband Impedance Terminations

The results above indicate that the memory effects present over the measurement bandwidth are due to non-constant impedance terminations (extrinsic electrical memory) and that the baseband impedance termination is the most important factor in creating these memory effects, particularly the IF1 impedance, since a significant improvement in IM3 asymmetry is observed when it is properly controlled.

This can be explained by the fact that the impedance variation across the bandwidth of the two carrier tones is quite small relative to the impedance variation across the bandwidth of the IF1 termination. This is illustrated in Figures 5.30(a) and 5.30(b) which show the measured system impedance across the carrier bandwidth and across the IF1 bandwidth respectively. These are the impedance terminations that are presented to the device when it is passively terminated into a broadband load and shows that it is impossible to achieve a constant broadband impedance, especially at low frequencies. The Smith chart of Figure 5.30(c) shows the same impedance variations, but in complex form.

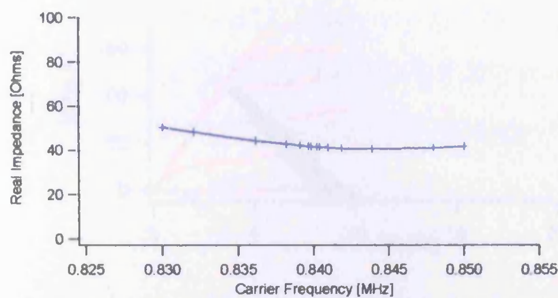


Figure 5.30(a) – Impedance across the carrier bandwidth

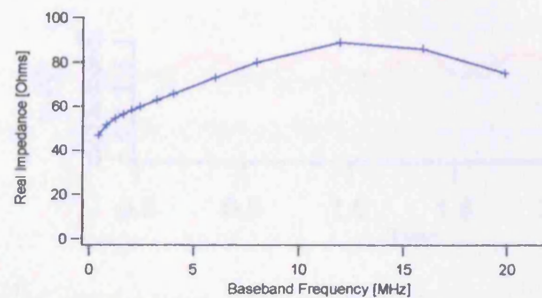


Figure 5.30(b) – Impedance across the IF1 baseband bandwidth

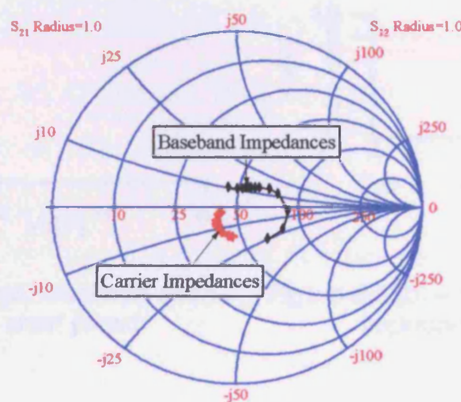
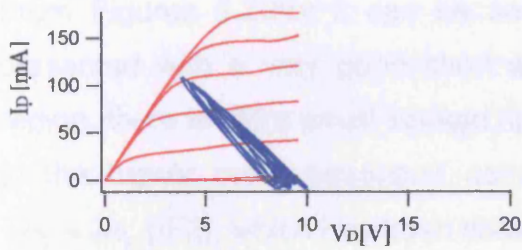


Figure 5.30(c) – Impedance variation across the IF baseband and carrier frequency bandwidths

It is also known that the baseband frequency components can have the effect of modulating the DC supply voltage [20]. This is a common problem and because the baseband frequencies can range from several multiples of the maximum modulation frequency to close to DC, it is normally solved by the addition of large capacitances in an attempt to short all baseband frequency components. The problem is illustrated in Figures 5.31(a) and 5.31(b) which show the measured RF dynamic loadline and baseband voltage when the device is operated in the linear region and the baseband impedances are 50Ω. Figures 5.32(a) and 5.32(b) show the same measurements after the IF1 impedance is loadpulled to a short circuit.



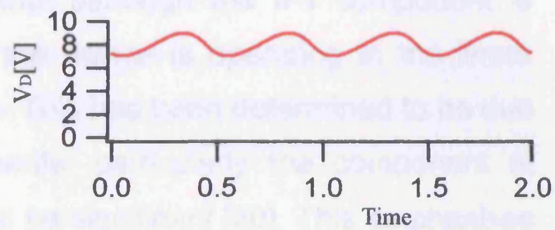
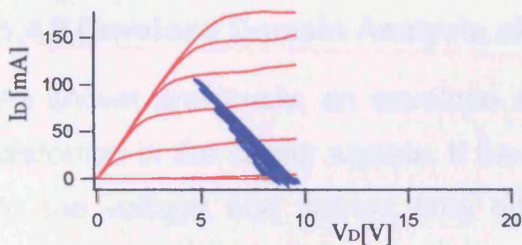


Figure 5.31(a) – RF dynamic loadlines when baseband impedances are 50Ω

Figure 5.31(b) – Envelope voltage when baseband impedances are 50Ω



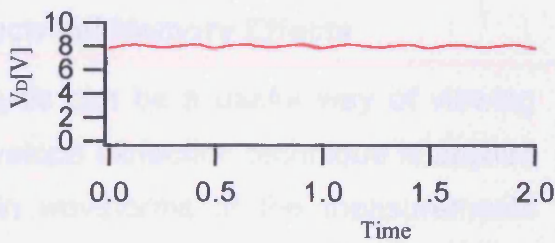


Figure 5.32(a) – RF dynamic loadlines when IF1 impedance is a short circuit

Figure 5.32b) – Envelope voltage when IF1 impedance is a short circuit

The baseband voltage is created in the software by simply filtering out all RF signal components, leaving just the DC and baseband components. These measurements graphically illustrate the effect of the IF1 termination on the overall device performance. When the IF1 impedance is 50Ω there is a significant IF1 component present along with the DC voltage. This has the effect of creating more asymmetry in the IM3s than is the case when the IF1 impedance is a short circuit. The reason for this is that when the DC voltage becomes modulated, the RF signal is free to move around in the voltage plane. This is seen in Figures 5.31(a) where the RF dynamic loadline is more spread out than that in Figures 5.32(a), and can be seen to be closer to the DC I-V boundary, even though the input power is the same for both measurements. The resulting interaction with the device boundary causes the device to compress sooner than expected and results in reduced linearity and greater IM3 asymmetry.

From Figures 5.32(b) it can be seen that although the IF1 component is presented with a very good short and the device is operating in the linear region, there is still a small voltage ripple. This has been determined to be due to the higher order baseband components, particularly the component at $2\omega_2 - 2\omega_1$ (IF2), which has been shown to be significant [30]. This emphasises the importance of correctly terminating all frequency components across the entire baseband bandwidth to minimise electrical memory effects.

5.4.8 Envelope Domain Analysis of Electrical Memory Effects

As shown previously, an envelope analysis can be a useful way of viewing distortion in the carrier signals. If the envelope extraction technique is applied to the voltage and current time domain waveforms of the measurements above it can be used to view the effects of electrical memory. Figure 5.33(a) shows an example of the resulting input voltage carrier envelope and the corresponding output current carrier envelope when the device is operated in the linear region. The two-tone tone spacing is 16MHz and all frequency components are passively terminated into 50Ω loads, except the IF1 component, which is presented with a short circuit. When the output current envelope is plotted against the input voltage envelope it results in an envelope transfer characteristic, as shown in Figure 5.33(b). The hystersis looping in this dynamic transfer characteristic is indicative of memory effects because the 'sweep' is different on the way up than on the way down. If no memory effects were present, there would be no looping at all.

The output voltage carrier envelope can also be plotted against the input voltage carrier envelope as shown in Figure 5.33(c) and shows much more looping compared to the output current transfer characteristic of Figure 5.33(b). This is because much of the looping present is due to electrical memory effects caused by impedance variations across the measurement bandwidth. The device is a voltage controlled current source, and the impedance of the circuit will not affect the output current, but it will affect the

output voltage since the voltage is a function of impedance. The looping in the output voltage transfer waveform therefore indicates that there is significant memory in the system, but the fact that there is very little looping in the output current transfer waveform indicates that the memory is caused by the circuit impedances.

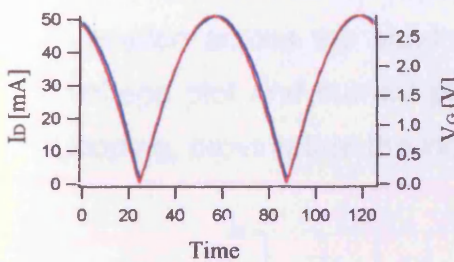


Figure 5.33(a) – Input voltage and output current envelope waveforms when device is in the linear region

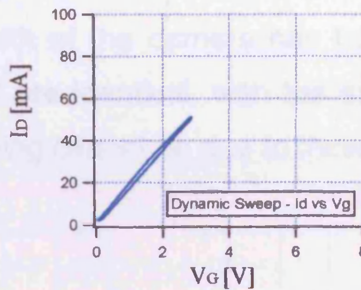


Figure 5.33(b) – Envelope current transfer characteristic

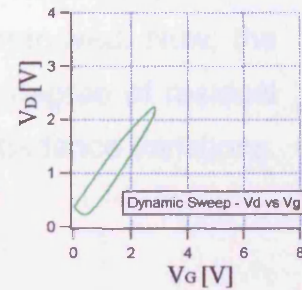


Figure 5.33 (c) – Envelope voltage transfer characteristic

The measurements of Figure 5.33 can be repeated at higher drive levels, and Figures 5.34(a)-(c) shows the equivalent measurements when the device is operated close to the 1dB compression point. These measurements confirm that there is little looping in the current transfer waveforms while there is much more looping in the voltage transfer waveforms, and that the memory effects are therefore largely due to circuit impedance variations.

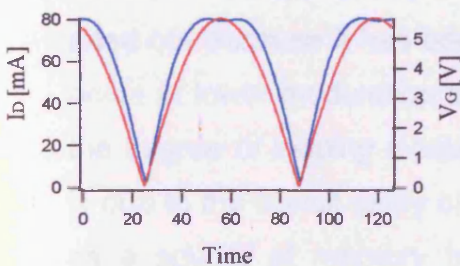


Figure 5.34(a) – Input voltage and output current envelope waveforms when device is close to saturation

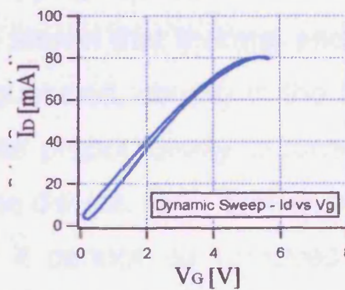


Figure 5.34(b) – Envelope current transfer characteristic

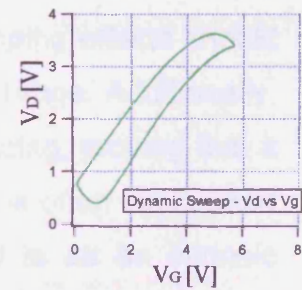


Figure 5.34(c) – Envelope voltage transfer characteristic

Although the memory effects appear to be due largely to circuit impedance variations, there is still a small amount of looping in the current transfer plots. This can be attributed to transit delay through the device and not the circuit impedances, and is proven by the plots of Figure 5.35. These plots are similar to the plots of Figure 5.33(b) and 5.33(c), but in this case the two carrier tones are loadpulled to be exactly the same impedance so that the impedance variation across the bandwidth of the carriers has been removed. Now, the voltage plot and current plot are identical, with the same degree of residual looping, proving that the looping cannot be due to those impedance variations.

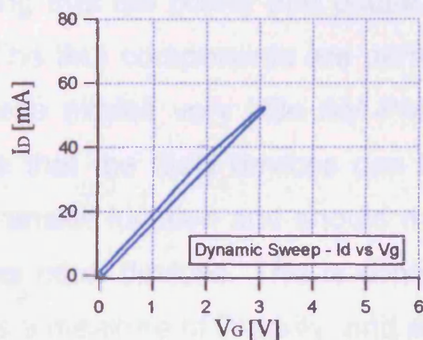


Figure 5.35(a) – Envelope current transfer characteristic with carriers loadpulled to same impedance

Figure 5.35(b) – Envelope voltage transfer characteristic with carriers loadpulled to same impedance

It could be argued that the residual looping is due to other memory effects such as thermal memory or trapping related memory in the device, but this is ruled out because it has been shown that thermal and trapping effects should occur at lower modulation frequencies, usually in the KHz range. Additionally, the degree of looping increases proportionally to tone spacing, proving that it is due to the transit delay of the device, [31]. Transit delay is often overlooked as a source of memory but it cannot be removed and is as an intrinsic electrical memory effect which must be accounted for.

5.5 Summary

A CW analysis of QinetiQs GaN HFET devices has shown that they are well behaved and their operation is very similar to that predicted by simple theory. A comparison with a GaAs HEMT device shows that although the two devices are based on heterostructure technology and are operated in Class A, their harmonic power and phase performance is quite different. The GaN devices exhibit more ideal behaviour than the GaAs device, and the harmonic components exhibit less complex phase behaviour.

The two-tone measurement results correlate well with the CW analysis, indicating that the power and phase behaviour of the GaN device is close to ideal. The IM3 components are particularly important with respect to linearity and these exhibit very little AM-PM. Overall, the CW and two-tone results indicate that the GaN devices can be described by a relatively simple, low order transfer function and should make good power amplifiers which are as linear as other devices. This is confirmed by the IP3 measurement, which is used as a measure of linearity, and also indicates that the GaN devices are as linear as any other weakly non-linear device technology. An additional two-tone experiment to compare the linearity of two devices with and without an AlN confinement layer resulted in a null result, indicating that although the AlN layer improved electron confinement, it had no apparent effect on RF performance.

Memory effects are a generic term that describes the fact that the response of an amplifier to a two-tone signal varies with varying modulation frequency. An envelope analysis of the two-tone data allows a direct comparison to be made between the CW measurement (no modulation) and the two-tone measurement (5MHz modulation). This shows that the device behaves in a very similar way during both measurements and indicates that the degree of trapping related dispersion that is present in the device is not changing when the input signal is modulated. This may be expected at a modulation frequency of 5MHz, but the development of a new measurement approach

meant that it was possible to perform a memory effect analysis of the device to examine the effect of modulation frequencies of varying tone spacing.
It was found that there was no evidence of thermal or trap related memory effects within the measurement bandwidth. Significant electrical memory was observed and it has been determined that most of these electrical memory effects were due to the impedance environment presented to the DUT, particularly the impedance variations across the bandwidth of the baseband frequencies. It was found that proper control of these impedance terminations can significantly reduce the measured memory effects, and that the device itself it inherently memoryless apart from the memory due to its transit delay.

5.6 References

[1] T. Kikkawa, M. Nagahara, T. Kimura, S. Yokokawa, S. Kato, M. Yokoyama, Y. Tateno, K. Horino, Y. Yamaguchi, N. Hara, and K. Joshin, "A 36W CW AlGaN/GaN Power HEMT using Surface-Charge-Controlled Structure," presented at IEEE MTT-S International Microwave Symposium, 2002.

[2] Y. F. Wu, M. Moore, A. Saxler, T. Wisleder, and P. Parikh, "40 W/mm Double Field-Plated GaN HEMTs," presented at 64th Device Research Conference, 2006.

[3] P. Kenington, *High Linearity RF Amplifier Design*: Artech House, 2000.

[4] H. Holma, A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communnications*, 2 ed: John Wiley and Sons Ltd, 2002.

[5] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Pothecary, J. Sevic, and N. Sokal, "RF and Microwave Power Amplifier and Transmitter Technologies - Part 1," in *High Frequency Design*, 2003, pp. 22–36.

[6] M. Johansson and L. Sundstrom, "Linearisation of RF Multicarrier Amplifiers using Cartesian Feedback," *IEEE Electron Device Letters*, vol. 30, pp. 1110-1112, 2004.

[7] R. J. McMorrow, D. M. Upton, and P. R. Maloney, "The Microwave Doherty Amplifier," presented at IEEE MTT-S International Microwave Symposium Digest, 1994.

[8] F. H. Raab, "Maximum Efficiencyand Output of Class-F Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, 2001.

[9] D. J. Williams, "Non-Linear Measurement System and Techniques for RF Power Amplifier Design." PhD Thesis: Cardiff University, 2003.

[10] S. A. Maas, *Nonlinear Microwave Circuits*: Artech House, 1988.

[11] R. Behzad, *RF Microelectronics*: Prentice Hall, 1998.

[12] J. G. Leckey, J. A. C. Stewart, and A. D. Patterson, "Analysis of HEMT Harmonic Generation Using a Vector Nonlinear Measurement System," presented at IEEE MTT-S International Microwave Symposium, 1996.

[13] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*: Artech House, 1999.

[14] M. Shur, *GaAs Devices and Circuits*: Plenum Publishing Corporation, 1986.
[15] D. Fanning, E. Beam, and P. Saunier, "pHEMTs Rise to Wide-Gap Challenge," in *Compound Semiconductor*, September 2007, pp. 25-27.
[16] H. Z. Liu, H. K. Huang, C. C. Wang, Y. H. Wang, C. H. Chang, W. Wu, C. L. Wu, and C. S. Chang, "High linearity power amplifier for PHS base station using a 50 mm AlGaAs/InGaAs/GaAs PHEMT," presented at Asia-Pacific Conference on Circuits and Systems, 2005.
[17] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*: Artech House, 2003.
[18] L. Shen, S. Heikman, B. Moran, R. Coffie, N.-Q. Zhang, D. Buttari, I. P. Smorchkova, S. Keller, S. P. DenBaars, and U. K. Mishra, " AlGaN/AlN/GaN high-power microwave HEMT," *IEEE Electron Device Letters*, 2001.
[19] R. S. Balmer, K. P. Hilton, K. J. Nash, M. J. Uren, D. J. Wallis, D. Lee, T. Martin, A. Wells, and M. Missous, "Analysis of thin AlN carrier exclusion layers in AlGaN/GaN microwave heterojunction field effect transistors," *Semicond. Sci. Tech*, vol. 19, pp. L65-67, 2004.
[20] S. C. Cripps, *Advanced Techniques in Power Amplifier Design*: Artech House, 2002.
[21] J. Vuolevi, T. Rahkonen, and J. Manninen, "Measurement Technique for Characterizing Memory Effects in RF Power Amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 49, 2001.
[22] N. Carvalho and J. Pedro, "A Comprehensive Explanation of Distortion Sideband Asymmetries," *IEEE Trans. Microwave Theory Tech.*, vol. 50, 2002.
[23] A. E. Parker and J. G. Rathmell, "Bias and Frequency Dependence of FET Characteristics," *IEEE Trans. Microwave Theory Tech.*, vol. 51, 2003.
[24] M. J. Madero-Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Phase Characterization of Intermodulation Products Including Electrothermal Memory Effects," presented at 2006 International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits, Aveiro, 2006.
[25] J. Kenney and P. Fedorenko, "Identification of RF Power Amplifier Memory Effect Origins unsing Third-Order Imtermodulation Distortion Amplitude and Phase Asymmetry," *IEEE Electronic Letters*, 2006.
[26] P. H. Ladbrooke, *Pulsed I(V) Measurement of Semiconductor Devices*: Accent Optical Technologies Inc., 2004.
[27] S.-Y. Lee, Y.-S. Lee, and Y.-H. Jeong, "A Novel Phase Measurement Technique for IM3 Components in RF Power Amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 54, 2006.
[28] T. Williams, "A Large-Signal Multi-Tone Time-Domain Waveform Measurement System with Broadband Active Load Impedance Control," Cardiff University, 2008.
[29] D. J. Williams, J. Leckey, and P. J. Tasker, "A study of the effect of envelope impedance on inter-modulation asymmetry using a two-tone time domain measurement system," presented at IEEE MTT-S, 2003.
[30] A. Alghanim, J. Lees, T. Williams, J. Benedikt, and P. J. Tasker, "Investigation of Electrical Base-Band Memory Effects in High-Power 20W LDMOS Power Amplifiers," presented at European Microwave Conference, 2007.
[31] J. Lees, T. Williams, S. Woodington, P. McGovern, S. Cripps, J. Benedikt, and P. Tasker, "Demystifying Device related Memory Effects using Waveform Engineering and Envelope Domain Analysis," presented at 38th EuMC, 2008.

Chapter 6. Conclusions and Further Work

6.1 Conclusions

One of the weakest links of any communications system is the power amplifier, largely due to the necessary trade-off between linearity and efficiency. The challenge of power amplifier design is therefore to provide maximum power and efficiency while maintaining highly linear operation. The GaN material system is of great interest in power amplifier applications and is potentially the best device technology to provide high power amplification at high frequencies. In particular, it has intrinsic properties that allow it to operate at very high frequencies, power densities, current densities, and drain voltages, relative to traditional power amplifier devices such as Si LDMOS, or even GaAs HBTs and HFETs.

The chief aim of this work was to investigate the suitability of GaN HFETs for use as power amplifiers and to help the development of the QinetiQ GaN program, largely through the use of novel measurements using custom RF non-linear time domain measurement systems. At the outset of this work these QinetiQ devices were in the early stages of development and it has been found that they exhibit the common problem of DC-RF dispersion, which results in much degraded RF performance relative to that predicted from DC measurements. Apart from this however, their fundamental characteristics such as breakdown voltage and frequency performance indicate that they should perform very well as power amplifiers.

DC analysis is the most basic form of characterisation performed on a power amplifier and is customarily used to predict the power performance and optimum load impedance of a device. However, DC-RF dispersion means that under RF excitation, the actual maximum output power and optimum load impedance may be very different from the predicted values. Large-signal RF waveform measurements provide a unique insight into the problem of dispersion and confirm that the reduced power and efficiency performance of the devices is due to dispersion, which is caused by trapping problems in the

semiconductor material. A comparison of the RF waveform data with pulsed I-V measurements shows that there is a direct correlation between the two types of measurement. Detailed analysis also shows, for the first time, that through careful choice of bias point, pulsed I-V measurements can be used to accurately predict the RF performance of a device.

The degree of dispersion present in the device is found to correlate directly with the device channel current drawn during operation. This current is largely controlled by the DC bias point and it has been shown that dispersion can be reduced by biasing the device closer to pinchoff. Using the full capabilities of the non-linear waveform measurement system, including harmonic loadpull, it was possible to overcome some of the unwanted side-effects of dispersion through suitable design, so that efficiency can be improved to acceptable levels, while maintaining high power operation.

Through large-signal CW and two-tone characterisation, it has been determined that these GaN devices operate in a very similar way to that predicted by simple theory and should therefore work well as power amplifiers in terms of linearity performance. The CW measurements show that both the power and phase behaviour of the device can be largely described by a simple power series transfer function, and the relative absence of AM-PM in the phase measurements indicates that the devices are very well behaved. A comparison with GaAs pHEMT devices is interesting because they are a relatively well developed technology and should in theory operate in a very similar way to the GaN devices since they are both based on heterostructure technology. CW measurements have shown that the power and phase behaviour of the GaAs devices is more complex than that of the GaN HFETs and again suggests that the GaN devices should work well as power amplifiers.

When the device is measured under two-tone excitation its behaviour shows a high level of consistency with the CW results and with the simple power series theory of operation. The lack of AM-PM in the CW measurements translates to a lack of AM-PM in the IM3s of the two-tone measurements, and suggests

that the devices may be relatively easy to linearize. IP3 measurements are also consistent with theory, indicating that the device is no more or less linear than any other third order dominant, weakly non-linear device. The effect of the addition of a thin AlN exclusion layer in the device structure is also investigated but the results show no measurable difference in linearity performance although the AlN layer has the effect of reducing alloy scattering and increasing carrier mobility in the 2DEG by up to 10%.

Memory effects are a generic term that describes the fact that the response of an amplifier to a two-tone signal varies with varying modulation frequency and are a very important factor when considering linearity. The origins of these memory effects are not well understood or documented but they are generally split into different groups: electrical memory, thermal memory, and trap-related memory. It is known that these GaN devices suffer greatly from trapping effects, so it is interesting to analyze the device under two-tone excitation across a variety of modulation frequencies ranging from several MHz to no modulation (CW). To this end, an envelope analysis of the earlier two-tone results, which had a modulation frequency of 5MHz, enabled a direct comparison to be made with the CW measurements. This shows strong correlation between the two measurements and indicates that the degree of trapping related dispersion that is present in the device does not change when the input signal is modulated at 5MHz.

However, this result should be expected since the trapping time constants in these devices is expected to be around 1ms and higher, meaning that they will be activated at modulation frequencies below 1KHz. Therefore it is desirable to measure two-tone signals with a wider range of modulation frequencies, and following the development of a new scope-based multi-tone waveform measurement system it is possible to measure two-tone signals of varying tone spacing in the time domain. Along with the addition of a new IF measurement testset it is now possible to analyse memory effects in the device through the measurement and control of all DC, baseband, and RF frequency components produced by the device when it is presented with a

two-tone signal. Unfortunately, the bandwidth limitations of the IF testset means that it can only measure baseband frequency components of 400KHz and greater.

It was found that these devices do not suffer from thermal or trapping-related memory effects within the measurement bandwidth of the system. Additionally, it was found that most memory effects evident are a result of the impedance environment presented to the device (extrinsic electrical memory), and are not caused by the device itself. The most significant cause of memory is the impedance variation across the baseband frequencies. The devices also suffer from intrinsic electrical memory effects due to their transit delay, but are otherwise inherently free from memory, indicating that most memory effects present in any system are caused by the impedance environment created by the matching and bias circuitry, rather than by the device itself.

6.2 Further Work

The measurements performed during the course of this work were severely limited by the robustness and reliability of the available devices. The robustness of the QinetiQ GaN devices has improved greatly so that it should now be possible to perform a more complete analysis, particularly by allowing the devices to be tested repeatedly in different modes of operation.

It would also be of great interest to measure GaN devices from other manufacturers for comparison purposes, and a further and more complete comparison between different device types such as Si LDMOS and GaAs HEMTs is also desirable. The linearity of the GaN HFETs was analysed using CW and two-tone analysis, but measurement and analysis of the effect of presenting the device with a more realistic multi-tone, modulated signal such as a CDMA signal would be very interesting, especially when compared directly with different device technologies to compare linearity.

Although the robustness of QinetiQs GaN devices has improved greatly since the start of this work, device robustness and reliability continues to be of great concern for all GaN manufacturers. It would therefore be very useful to measure the channel temperature during DC and RF operation. This could be achieved via infra-red microscopy and would provide a valuable insight into device operation, especially when combined with RF time-domain measurements. These devices operate at very high powers and thermal management is therefore very important, especially regarding device reliability.

During the course of this work a new two-tone measurement approach was developed that enabled the analysis of memory effects in the device via the measurement of two-tone signals of varying modulation frequency. Unfortunately, the baseband bandwidth of the system was restricted since frequencies below 400KHz could not be measured, and this limited the frequency range of the two-tone measurements performed. However, practical modulated signals can have very low modulation frequencies that approach DC. It is therefore desirable to measure the two-tone signal and resulting baseband components at all modulation frequencies. To this end, a useful extension of this work would be to extend the measurement capabilities of the system to baseband frequencies below 400KHz, all the way to DC. This could be achieved through the integration of low frequency voltage and current probes. The measured low frequency voltage and current components would then be added to the other frequency components in the control software. It is crucial to maintain phase coherence, and this is now possible with voltage and current probes that can be synchronised to the other measurement instruments via a standard 10MHz phase reference.