CHAPTER 3

GaN-BASED POWER HIGH ELECTRON MOBILITY TRANSISTORS

Shreepad Karmalkar

EE Department, IIT-Madras, Chennai 600 036, India

Michael S. Shur* and Remis Gaska†

**ECSE Department, RPI, 8th Street Troy, New York 12180, USA*

†Sensor Electronic Technology, Inc., Columbia, SC 29209, USA

Contents

1. Introduction 174
    1.1. Power versus Small-Signal HEMT 176
2. Device Structures and Fabrication 177
    2.1. Basic Device Structure 177
        2.1.1. The substrate 178
        2.1.2. The contacts 178
        2.1.3. Passivation 180
        2.1.4. Packaging 180
    2.2. Improved Device Structures 181
    2.3. Device Fabrication 182
        2.3.1. Material growth 183
        2.3.2. Gate lithography 185
        2.3.3. Annealing 185
3. Characteristics 185
    3.1. DC Characteristics 185
        3.1.1. Drain current 185
        3.1.2. Breakdown voltage 187
        3.1.3. Passivation effects 188
        3.1.4. Geometry scaling effects 188
        3.1.5. Temperature effects 189
        3.1.6. Negative resistance 190

3.2. RF Characteristics 190
3.2.1. Power and cut-off frequency 190
3.2.2. Passivation effects 193
3.2.3. Geometry scaling effects 194
3.3. Reliability 194
3.4. Uniformity 196
3.5. GaN versus GaAs and InP Power HEMTs 198
3.6. HEMTs versus MESFETs and HBTs 200
4. Modeling 200
4.1. Basic Phenomena 201
4.1.1. Gate-controlled 2-DEG 201
4.1.2. Carrier transport 202
4.1.3. Hot electrons 203
4.1.4. Polarization effects 203
4.2. Analytical Modeling 204
4.2.1. Critical DC and RF parameters 204
4.2.2. Drain current and power 206
4.2.3. Gate current 208
4.3. Numerical Modeling 209
4.3.1. Drain current 209
4.3.2. Breakdown voltage 209
4.3.3. Negative resistance 211
5. Circuits 211
6. Summary 212
References 213

1. Introduction

High Electron Mobility Transistors (HEMTs) have emerged as a promising candidate for microwave ($f > 1$ GHz) power amplification,[1–3] with applications ranging from satellite links to wireless communications, from highways to electronic warfare. Also, they have a potential for low frequency ($f < 100$ MHz) high voltage (up to 1 kV) switching power control.[4–6] Until recently, power HEMTs were primarily based on AlGaAs/GaAs, AlGaAs/InGaAs, AlInAs/InGaAs and related epitaxial films grown on GaAs or InP substrates. In late 1990s, AlGaN/GaN, AlGaInN/GaN, and AlGaInN/InGaN power HEMTs grown on sapphire, insulating 4H–SiC, conducting SiC, and even bulk GaN have demonstrated much larger output powers and have become promising contenders for a variety of high power amplification and switching applications. Moreover, the use of wide bandgap semiconductors in power amplifiers not only increases the output power, but also extends the temperature tolerance and the radiation hardness of the circuits. The latter is corroborated by recently demonstrated

operation of GaN-based HFETs at 750°C.[7] In this chapter, we review the various aspects of these devices, namely — history, current status, technology, characteristics, modeling and circuits.

The HEMT is also known as MODFET (Modulation-doped FET), TEGFET (Two-dimensional Electron Gas FET), SDHT (Selectively Doped Heterostructure Transistor) or simply, HFET (Heterojunction FET). The unique feature of the HEMT is channel formation from carriers accumulated along a grossly asymmetric heterojunction,[8] i.e. a junction between a heavily doped high bandgap and a lightly doped low bandgap region. In HEMTs based on GaN substrates, this carrier accumulation is mainly due to polarization charges developed along the heterojunction in the high bandgap AlGaN side.[4] This is in contrast to the situation in other HEMTs, such as those on GaAs or InP substrates. Here, the accumulation is a result of carrier diffusion from the heavily doped to the lightly doped region, the diffusion being enhanced significantly by the bandgap difference between the two regions. Whatever the physical origin of carrier accumulation, the accumulated carriers might have high mobility due to their separation from their heavily doped source region, and their location in the low-doped region where impurity scattering is absent. A discussion of the basic principles of HEMT operation is readily available in several books, e.g. Refs. 9 and 10.

The physics of carrier transport parallel to a heterojunction was first considered in 1969.[11] The development of MOCVD and MBE technologies in the 1970s made heterojunctions practical. Although substantial research on GaN growth was initiated in early 1960s, the technological spin-offs came late because of lack of ideal substrates. The enhanced mobility effect was first demonstrated in AlGaAs/GaAs heterojunctions in 1979,[12] and applied to demonstrate a HEMT in 1980.[13,14] Electron mobility enhancement at AlGaN/GaN heterojunction was first reported in 1991.[15] Later, this enhancement was attributed to the 2D nature of the electrons, based on observations of mobility increase with lowering of the temperature down to 77 K and Shubnikov-de-Haas oscillations. The potential of AlGaN/GaN HEMTs for microwave electronics was demonstrated in 1994.[16] With advancements in material quality, heterostructure design, and ohmic contact formation, the excellent power capability of these HEMTs was established by 1997.[17,18] Recently, some AlGaN/GaN structures useful for several hundred volts power switching have been proposed.[6,19]

Two factors, namely — new markets and technological advances[3,5] have made the HEMT into a viable industrial product. Until 1980, most applications of RF transistors had been military or exotic scientific projects, e.g. electronic warfare, missile guidance, smart ammunition etc. In the 1980s, satellite television using low-noise transistors operating at 12 GHz in the receiver front-ends was the first civil application of RF transistors with a market volume worth mentioning. Currently, we witness far-reaching developments in civil communication technology. Mass consumer markets have been created for RF systems by the advent of mobile phones, whose production surpassed that of personal computers in 1998. There have been several advances in HEMT technology. Electron Beam Lithography (EBL) has enabled realization of a short ($\sim 0.1\ \mu$m) T-shaped gate in a repeatable manner. Improvements in device modeling/simulation have permitted better prediction of experimental characteristics. Progress in MBE and MOCVD growth techniques have made it possible to control the material composition and thickness of the heterojunction layers.

1.1. *Power versus Small-Signal HEMT*

There are some important differences between HEMTs used in power and small-signal applications. A power device has to withstand a larger voltage and current amplitude as compared to a small-signal device, to provide high output power. High drain voltage operation is necessary for two reasons. First, increasing the device output power by increasing $I_{\max}$ alone involves increasing the device area, resulting in the input and output impedances too low for a good match with the surrounding circuitry. Second, to achieve a high power added efficiency, class AB or class B operation, requiring a large quiescent drain voltage, is used, resulting in a high drain voltage when the gate-source voltage is swung below pinch-off.

Consider the role of small-signal and power HEMTs in wireless communication, which occupies a significant RF market.[1] The wireless communication hardware consists of the infrastructure (base station), and the user part (handset) having receiver and transmitter sections. In the handset, supply voltage as low as 3 V is used to reduce power consumption. Small-signal low noise transistors are required to amplify incoming signals in receiver front-ends, and power transistors with low on-resistance and high on-current are required in the transmitter section. Today, the small-signal low-noise applications of a HEMT include collision avoidance radar at 77 GHz and

military radar at 94 GHz. On the other hand, power applications, which are the subject of the present chapter, include base station applications at ~ 40 GHz, mobile communication at $\sim 0.9/1.9$ GHz, and switching power supplies at < 1 MHz.

Small-signal devices typically have a gate length below 0.15 μm and breakdown voltage of $\sim$ 5 V, since they have to be optimized mainly with respect to their RF performance. However, power HEMTs typically have uncritical gate lengths in the range 0.5–1 μm and breakdown voltages > 10 V. Gate lengths > 1 μm may be used for several hundred volts breakdown voltage required in power switching applications. Note that gate definition accounts for a significant fraction of the total fabrication cost. Thus, the use of uncritical gate-lengths in power HEMTs facilitates their cheap large volume production, so that they compete successfully with various other technologies in power applications, such as Si/Ge-HBTs, III-V HBTs, and GaAs-MESFETs.

2. Device Structures and Fabrication

2.1. *Basic Device Structure*

The cross section of a basic AlGaN/GaN power HEMT is given in Fig. 1. We have fabricated AlGaN/GaN HFETs with the source-to-drain spacing from 2 μm to 7 μm, the gate length from 0.25 μm to 5 μm and the total gate width from 50 μm to 150 μm (2×25 μm to 2×75 μm). Although not shown in the figure, plated via holes ($\sim$ 20 μm $\times$ 40 μm) cut through the substrate are used to connect the source metallization to the back

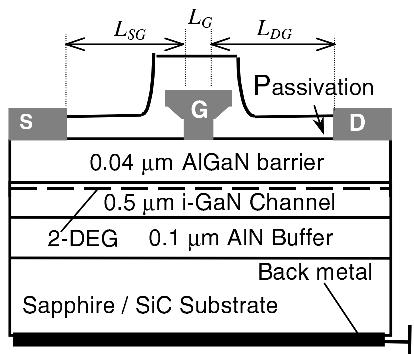

Fig. 1. The basic AlGaN/GaN power HEMT structure.

metal, which is grounded. This reduces the parasitic source inductance and improves thermal dissipation.

2.1.1. *The substrate*

GaN substrates with large enough diameter do not exist for growing GaN channels; the largest GaN substrates obtained so far are 1.7 cm $\times$ 1 cm.[20] Sapphire (Al_2O_3) and SiC are the most popular substrate materials used currently. Sapphire substrates are cheaper for GaN growth than SiC. However, the low thermal conductivity of sapphire presents a serious challenge for packaging of high power devices. Long term reliability may be compromised due to thermally induced stress on the contacting pads, which also serve as the heat conducting path. On the other hand, SiC has lower lattice mismatch with GaN or AlN (3.5% against 13% of sapphire), and 10 times higher thermal conductivity.[21] Other substrates, such as $LiAlO_2$,[22] $LiGaO_2$,[23] Si,[24] and AlN[92,93] are also being investigated. The Si substrates are the cheapest, and have the potential of integrating GaN optoelectronics with silicon devices. However, the growth of GaN on these substrates is still a new area. Properties of various substrate materials appear in Table 1.

2.1.2. *The contacts*

The shape of the gate contact is crucial to device performance. T- or Y-shaped cross-sections are employed (Fig. 1), wherein the stem provides the short channel length, while the wide "hat" reduces the gate resistance of the connection to the bonding pads. Multilayer metallization schemes are used for the gate, source and drain contacts. The initial and barrier layers are usually as thin as 20–50 nm. The final Au layer is thick, with the thickness ranging from ~ 0.15 μm up to several microns. The gate contact is Schottky type, made using Ni/Au, Pt/Au or Pd/Au layers.[25,26] Another material being investigated for Schottky contact is RuO_2.[27] The source and drain contacts are ohmic and employ Ti/Al/Ni/Au or Ti/Al/Ti/Au layers.[25,26,28] Note that the metallization schemes used in GaN devices are different from those in GaAs devices due to two factors.[26] First, in GaN devices, it is difficult to create a sufficient $n{+}{+}$ surface impurity doping during contact annealing. Second, in Schottky contacts on AlGaN, the barrier height increases with increasing metal work function.

Table 1. Properties of some substrate materials for GaN growth.

Substrate	E_g (eV)	χ (W/cm°C)	E_c (MV/cm)	ε	μ (cm^2/Vs)	v_s (cm/s)	$\frac{\chi E_c{}^2 \varepsilon \mu \nu_s^*}{(\chi E_c{}^2 \varepsilon \mu \nu_s)_{\mathrm{Si}}}$
Si	1.1	1.31	0.3	11.8	1350	1×10^7	1
GaN	3.39	1.3	3.3	9	1200	2.5×10^7	153
SiC	2.86	4.9	2.0	10	650	2×10^7	136
Sapphire	—	0.25	—	—	—	—	—
AlN	6.2	3.4	—	8.5	300	1.5×10^7	—
GaAs	1.43	0.53	0.4	12.8	8500	2×10^7	10

*Combined figure of merit for high temperature/high power/high frequency applications.

2.1.3. *Passivation*

Passivation of the gate-drain and gate-source surfaces by depositing SiO_2 and Si_3N_4[29,30] arrests the degradation in maximum current and transconductance due to the surface charges and traps present in these regions. These traps are a result of disorder in surface bonds of the III-V compound semiconductor material and cause Fermi-level pinning.

2.1.4. *Packaging*

An AlGaN/GaN power HEMT is encapsulated in packages similar to those developed for silicon power devices. However, the choice of packaging materials plays a more critical role in compound semiconductors than in silicon, due to differences in the coefficient of thermal expansion, and because compound semiconductors are more fragile and may exhibit mechanical stresses causing device degradation and failure.[31] Low power (< 1 W) packages include hermetic metal can packages (TO-39), plastic encapsulations (TO-92), plastic surface mount packages, and metal-ceramic packages for severe environmental conditions. For powers > 1 W, the RF die is mounted in metal ceramic package. Hermeticity is seldom warranted for commercial applications, and is required for special applications such as military and space, where environmental conditions are severe. The difficulty in dissipating heat from conventionally die bonded/wire bonded power HEMT chips forces thinning of finished wafers. Heat dissipation is also a problem with substrates having low thermal conductivity such as sapphire. In such cases, flip-chip mounting of the die on the package base is a solution (see Fig. 2).

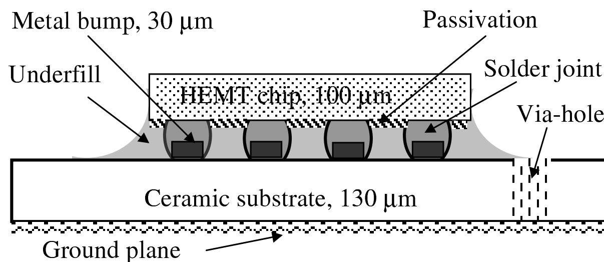

Fig. 2. Cross-section of a flip-chip assembly. The planar dimensions of the chip and the substrate are about 0.5 cm and 1 cm respectively. The diameters of the bump and the plated via hole are is 80 μm, and 200 μm respectively (after Ref. 32).

2.2. *Improved Device Structures*

Some improved device structures proposed by us are shown in Fig. 3. In Fig. 3(a), 50–100 nm doped GaN channel with a donor concentration of 5×10^{17}–1×10^{18} cm^{-3} has been incorporated between the nominally doped GaN and the AlGaN layer.[33] This increases the density of two-dimensional (2D) electron gas near the heterointerface up to $n_s = 2 \times 10^{13}$ cm^{-2}, and thus raises the current carrying capability, $I_{\max}$, of the device. It has also led us to the idea of highly doped channel GaN MESFETs[34] that has many advantages in terms of the ease of fabrication and stability. Note that, the current capability can also be increased simply by increasing the device periphery by laying out multiple gate fingers and interdigitation (see Fig. 4). However, following this approach beyond a point creates problems in matching the device impedance to the surrounding circuitry.[1] Another feature of the improved structure is the use of a conducting 6H–SiC substrate.[35] A high-quality epitaxial insulating AlN buffer layer allows us to use both

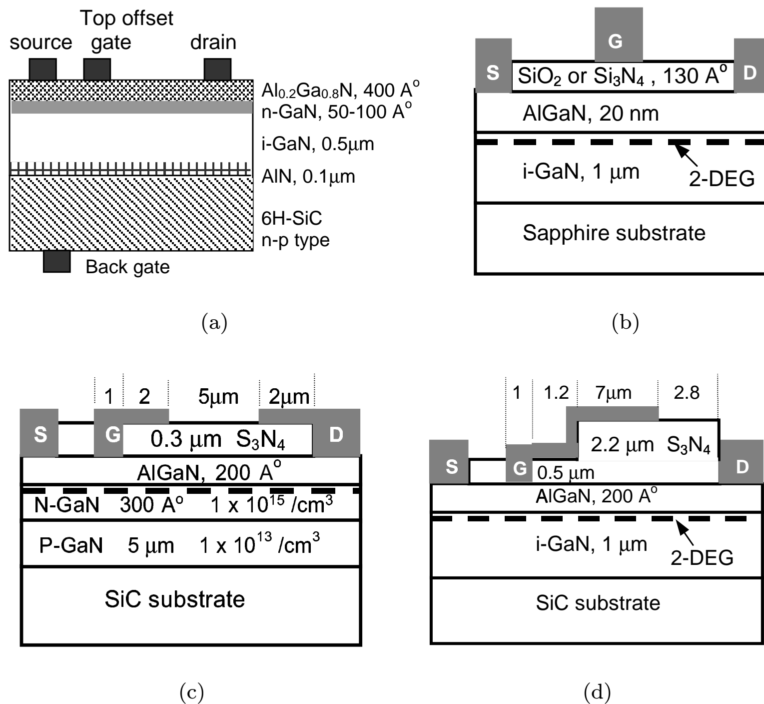

Fig. 3. Improved device structures.

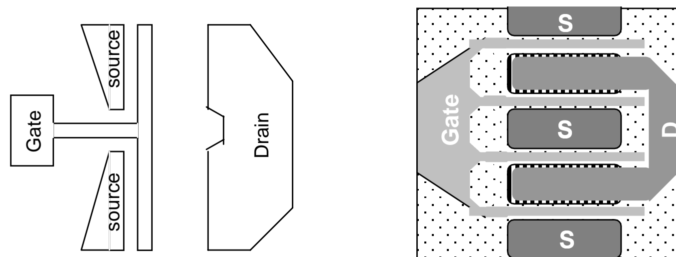

Fig. 4. Top views of (a) offset gate and, (b) multiple finger gate, designs.

n-type and p-type SiC substrates as second (buried) gates, which strongly affect the device performance. Finally, in order to increase the breakdown voltage, $V_{\rm br}$, an offset gate design has been used (also see Fig. 4).[94] We have studied the effect of varying the ratio of the source-to-gate distance, $L_{\rm SG}$, and the gate-to-drain distance, $L_{\rm GD}$, from 0.2 to 5. These separations as well as the gate lengths were determined from the SEM pictures.

The structure shown in Fig. 3(b) is called MOSHFET or MISHFET.[28,36,91] Here, an insulator, such as silicon dioxide or silicon nitride, has been introduced between the gate and the donor layer. It reduces the gate current by 4–6 orders of magnitude, and also permits a large negative to positive gate voltage swing, doubling the maximum current without degrading the frequency response.

Structures shown in Figs. 3(c) and (d), called RESURF HEMT[6] and Field-plate HEMT[19] respectively, can provide a breakdown voltage, $V_{\rm br}$, up to 1 KV. Here, $V_{\rm br}$ is raised by lowering the peak electric field, $E_{\rm peak}$, near the drain edge. This has been achieved in Fig. 3(c), using a novel combination of field plates and a buried p-n junction. Figure 3(d) is a variation of this arrangement, where a gate field plate alone, deposited on a stepped[19] insulator has been employed (uniform insulator[25,37] is also possible). $E_{\rm peak}$ can also be lowered by reducing the 2-DEG concentration in the gate-to-drain separation.[19,37]

Sometimes, a device structure with a thick semiconductor region between the gate and the drain/source, and a recessed gate region is employed, mainly with a view to reduce the source and drain parasitic resistances.[38,38a]

2.3. *Device Fabrication*

AlGaN/GaN HEMTs are fabricated in the following steps:

1. Growth of the semiconductor epitaxial and insulator layers.
2. Photolithography for ohmic contact openings.
3. Ohmic contact metallization.
4. Rapid Thermal Annealing (RTA) of ohmic contacts.
5. Photolithography for device isolation level.
6. Reactive ion etching or ion implantation for device isolation.
7. Contact or *e*-beam lithography for gate openings.
8. Gate metallization.
9. Substrate thinning and slot via formation.
10. Backside metallization.
11. Dicing and packaging.

2.3.1. *Material growth*

To-date, most AlGaN/GaN and AlInGaN/GaN heterostructures are grown by Metal Organic Chemical Vapor Deposition (MOCVD). MOCVD systems are capable of operating in both conventional and atomic layer deposition regimes. The conventional deposition regime is that in which precursors entering the growth chamber are simultaneously used to deposit GaN layers. Triethylgallium and ammonia are used as the precursor gases. $Al_xGa_{1-x}N$ layers are deposited when precursors enter the chamber in a cyclic fashion. Triethylgallium, triethylaluminum and ammonia are used as precursors. The precursors are introduced into the chamber using hydrogen or nitrogen as a carrier gas. Epilayers are deposited on substrates placed on a graphite susceptor, which is heated to the growth temperature by RF-induction or resistive heating. Schematics of AlGaN HFET, MOSHFET, and MISHFET epilayers without and with doped channel are shown in Fig. 5. The epilayers are typically grown on basal (0001) sapphire, conducting 6H–SiC and semi-insulating 4H–SiC substrates. The substrates are slightly off-axis (1–2°) production grade 6H and 4H silicon carbide, available from Cree Research. The doping levels of both *n*- and *p*-type 6H–SiC are on the order of 2×10^{18} cm^{-3}. The micropipe density is approximately 30 cm^{-2}. The deposition of an approximately 50 nm thick AlN buffer is followed by the growth of insulating/modulation doped GaN, which is finally capped with an AlGaN barrier layer. The major difference between the AlGaN/GaN growth on sapphire and SiC substrates is the thickness of the insulating GaN layer. The details of MOCVD epitaxial growth are given in Refs. 39 and 40. An important and somewhat unique feature used in Refs. 39 and 40 and related work show that the use of trace amounts of indium throughout

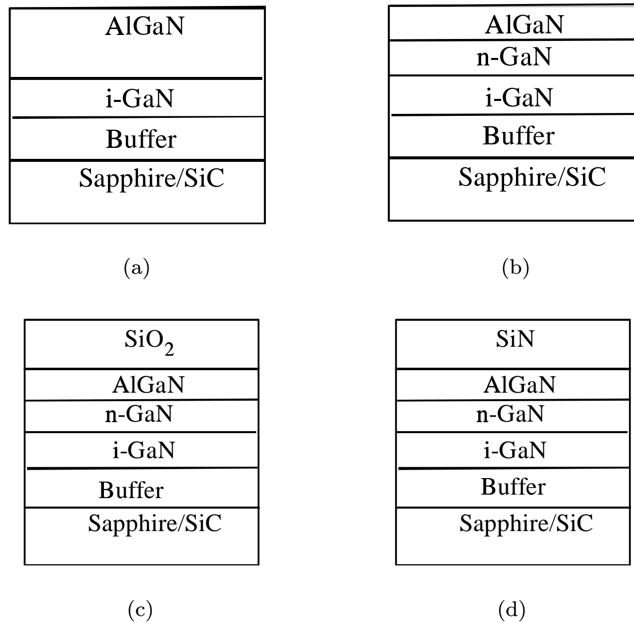

Fig. 5. Schematics of AlGaN HFET, MOSHFET, and MISHFET epilayers without and with doped channel.

the entire epitaxial structure, which dramatically improves materials and surface quality.

Cross-sectional TEM analysis of GaN grown on sapphire and SiC reveals a strong dependence of growth defect distribution along the growth direction on the substrate material. A significant reduction in the number of threading dislocations in GaN on sapphire is observed for layer thicknesses above 1.5–2 μm. A similar improvement in material quality for GaN grown on SiC was achieved at thicknesses as low as 0.5 μm or even lower. Also, preliminary data obtained at Lawrence Berkeley National Lab point to a significantly lower number of domains with inverted crystal face polarity in GaN layers on SiC. This is in good agreement with our electron transport studies, which show better quality AlGaN/GaN heterointerfaces in the structures grown on SiC, in particular, on 6H–SiC. Thus, the performance of the devices with higher levels of dissipated power can be improved by effective heat sinking through the SiC substrate.

Epitaxial films for GaN-based HEMTs have also been grown by Molecular Beam Epitaxy on 4H–SiC substrates using RF-assisted gas source

molecular beam epitaxy technology. The details of this technique can be found in Refs. 41–44.

2.3.2. *Gate lithography*

Realization of the gate electrode is one of the most crucial steps in device fabrication. Either optical[38,45,46] or Electron Beam Lithography (EBL)[47,48] may be employed for this purpose. Generally, EBL approach is preferred for sub-half micron gate lengths. In the EBL approach, a T-shaped gate (see Fig. 1) is created in a multi-layer *e*-beam resist deposited on the device, by a direct write exposure system. Tri-layer and four-layer resist processes are available.[48] However, devices fabricated using the four-layer resist process show improved values of breakdown voltage and f_{max}, the latter due to lower values of gate-drain feedback capacitance and output conductance, with only slight reduction of drain current and transconductance.[48]

2.3.3. *Annealing*

Rapid Thermal Anneal (RTA) is used for the formation of ohmic contacts to source and drain, for radiation hardening, and for recovering the electrical properties of the devices damaged by plasma-assisted deposition and etching processes. For ohmic-contacts, the annealing is carried out in N_2/H_2 ambient for 30–60 seconds. It reduces the contact resistance to $\sim 10^{-6}\ \Omega\mathrm{cm}^2$ due to alloying. The RTA temperature is $\sim 850^{\circ}$C for Ti/Al/Ni/Au. It has been reported that, an RTA for 40 seconds at 800°C could recover the serious degradation in I_{max} and g_{m} of AlGaN/GaN HEMTs, induced by proton irradiation,[49] and by plasma assisted processes.[50]

3. Characteristics

The power HEMT characteristics can be classified into DC and RF. Their critical parameters, reliability and uniformity are addressed below.

3.1. *DC Characteristics*

3.1.1. *Drain current*

The measured DC characteristics of a power HEMT are shown in Fig. 6. The critical parameters related to these characteristics are the maximum drain current (I_{max}), the threshold voltage (V_{T}), the peak DC transconductance ($g_{\mathrm{m(peak)}}$), the breakdown voltage (V_{br}) and the output conductance

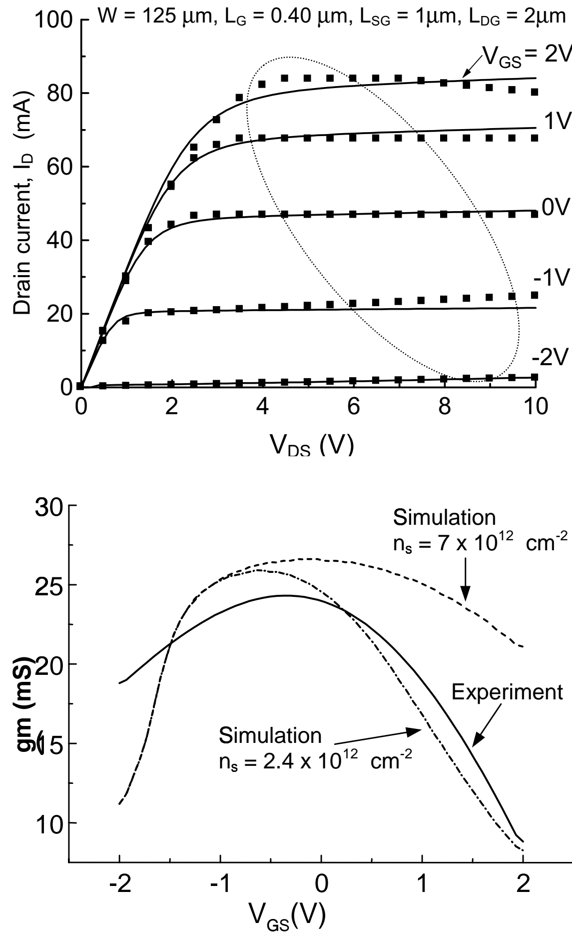

Fig. 6. Measured I_D–V_{DS} and g_m–V_{GS} curves (experimental data from Ref. 51). The dashed oval superimposed over the I_D–V_{DS} curves represents a possible load line at microware frequencies. On the I_D–V_{DS} map, points represent experimental data, and lines represent AIM-spice simulated curves.

(g_o). I_{max} is the maximum current that a device can allow under any bias condition. It is measured for the device biased in saturation, which is the condition used for the measurement of g_m and g_o as well. This is because the active region of a power transistor operation for amplification covers is restricted to the saturation region of the I_D–V_{DS} curves. V_T can be defined in two ways. It may be found as the V_G intercept of the linear extrapolation of the I_D–V_{GS} curve to the V_{GS} axis. Alternately, it can be taken as the

value of V_{GS} corresponding to a given low value of I_D, e.g. 0.01 mA/mm; this is physical pinch-off condition. The value of V_T in these two definitions can differ significantly because the I_D–V_{GS} curve of the power HEMTs approaches cut-off gradually.

3.1.2. *Breakdown voltage*

The device supports the maximum voltage when turned off, i.e. when $V_{GS} < V_T$. Hence, the breakdown voltage for this condition, $V_{br(off)}$, is important. For devices such as the HEMTs, which operate at microwave frequencies, the breakdown in the on-state, i.e. for high V_{GS}, also becomes important for two reasons (see Fig. 6). First, $V_{br(on)}$ can be significantly less than $V_{br(off)}$; e.g. in a study,[51a] the measured values for these were 60 V and 100 V respectively. Second, the shape of the load line for RF applications is a closed curve and not a straight line of the DC case. The curve area increases not only with the applied input voltage swing but also with increasing frequency. Thus, a much larger area of the output characteristics is covered than in the DC case, and higher voltages are reached even at higher currents. It is then possible that the load line may trespass the $V_{br(on)}$ locus. The reason for the change in shape of the load-line with frequency is understood with the help of the basic amplifier circuit shown in Fig. 7. The transistor is biased with a DC voltage V_{DD} and V_{GG} is varied. When the frequency of V_{GG} variation is low, the slope of the load line is defined by the load resistance, so that the load line is straight. However, as the frequency increases, the capacitive and inductive elements of the transistor and the matching networks eventually become significant. In this case, the transistor voltages and currents can be determined by circuit simulation, and these translate to the non-linear load-line.

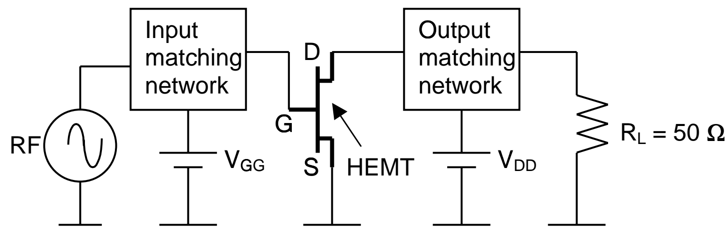

Fig. 7. An amplifier circuit with a HEMT biased at V_{DD} and V_{GG}.

3.1.3. *Passivation effects*

Refer to the measured DC parameters of an AlGaN/GaN device shown in Table 2. Passivation, which suppresses surface traps and charges, improves I_{max} by 24%, V_{br} by 26%, and $g_{\text{m(peak)}}$ by 8%, and shifts V_{T} by -0.25 V. Accordingly, a significant increase in output power capacity and reasonable increase in gain are predicted. Another study of GaN devices[30] supports these trends. The improvement in I_{max} and g_{m} should be attributed partly to the elimination of surface depletion leading to reduction of the sheet resistivity of the recess region, and partly to the V_{T} shift. It is of interest to contrast the effects of passivation in GaN and GaAs devices. Unlike in GaN devices, in GaAs devices,[52] passivation degrades V_{br} and makes it independent of the gate-drain separation, L_{GD}. This points to different origins of breakdown in unpassivated GaAs and GaN HEMTs. In GaN devices, the conduction and ionization processes associated with this breakdown occur in the surface layer, while in GaAs devices, they occur in the channel layer. In the absence of passivation, surface charges deplete the channel over L_{GD} in GaAs HEMTs, spreading the electric field over L_{GD}, and thus reducing the E_{peak} at the drain edge of the gate. This is not possible after passivation. However, the electric field can spread over several microns of L_{GD} in GaN devices even after passivation, mainly due to their much higher breakdown field than GaAs. It is for this reason that V_{br} increases with L_{GD} in GaN devices even after passivation.

Table 2. Measured DC parameters[29]; $V_{\text{br}}I_{\text{max}}$ is a figure of merit for maximum power output capacity. The device has $L_{\text{G}} = 0.5$ μm and $L_{\text{GD}} \sim 1$ μm.

Parameter	Unpassivated	Passivated
I_{max} (mA/mm)	520	640
$V_{\text{br(off)}}$ (V)	42	53
$V_{\text{br}}I_{\text{max}}$ (W/mm)	21.8	33.9
$g_{\text{m(peak)}}$ (mS/mm)	195	210
V_{T} (V)	-4.5	-4.75
n_{so} (/cm^2)	$\sim 10^{13}$	$\sim 10^{13}$

3.1.4. *Geometry scaling effects*

It is important to consider the impact of gate geometry scaling on DC characteristics, since smaller L_{G} and larger width are preferred for better performance. In a study,[53] as L_{G} was reduced in the micrometer range from

6 μm to 2 μm, the extrinsic g_m, measured at 7 V drain bias, rose linearly from 97 mS/mm to 107 mS/mm. However, the g_m reached a maximum of about 130 mS/mm for $L_G = 0.85$ μm, and fell for further reduction of L_G in the sub-micrometer range to 100 mS/mm at $L_G = 0.1$ μm. This could be attributed to short-channel effects such as drain induced barrier lowering and gate-fringing. Other consequences of such short-channel effects observed, such as V_T and V_{br} reduction, are shown in Fig. 8. The effect of increasing the width has also been studied.[28] The g_m increased to 410 mS for a rise in width up to 3 mm, but saturated at 440 mS, on further width increments to $W = 6$ mm. This saturation could be attributed to the random variation in V_T over the large gate area. Thus, the law of diminishing returns starts operating for increase in W and decrease in L_G beyond a certain point.

3.1.5. *Temperature effects*

Several researchers have studied the effect of temperature variation on device characteristics.[54,55] Results of one such study are given in Table 3. While the device V_T $(= -3.6$ V$)$ showed little change with temperature variation, the g_m and I_{max} fell by 35% over the temperature range of 23–187°C. This was attributed mainly to the fall in v_s and mobility with temperature.

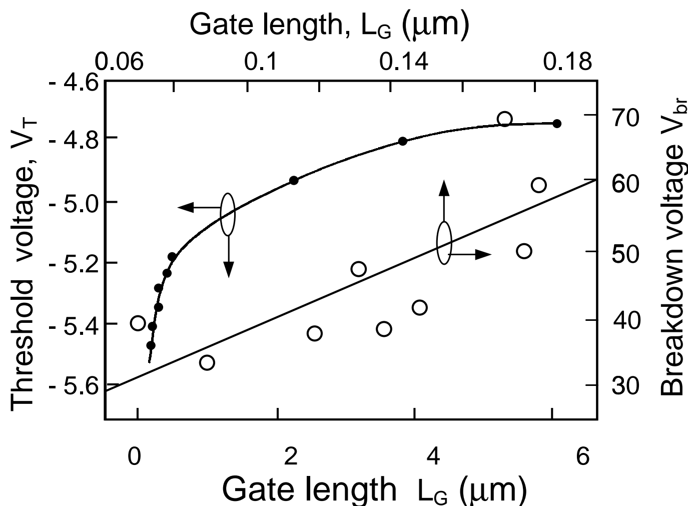

Fig. 8. Short-channel effects, namely V_T and V_{br} reduction, in GaN HEMTs (after Ref. 53).

Table 3. Effect of temperature variation.[54]

T (°C)	$g_{m(peak)}$ (mS/mm)	I_{max} (mA/mm)	f_T (GHz)	v_s (10^7 cm/s)
23	110	480	13.7	1.2
187	72	320	8.7	0.8

3.1.6. *Negative resistance*

An anomaly observed in the I_D–V_{DS} characteristics of GaN HEMTs on sapphire substrates is shown in Fig. 9. These devices show negative resistance for high currents and voltages. This has been attributed to self-heating of the device due to the poor thermal conductivity of sapphire.[56] Detailed explanation for this anomaly will be given in the modeling Sec. 4.3.3. This effect can be minimized by measuring the characteristics under pulsed condition to reduce power consumption, by using higher thermal conductivity substrate, and by proper cooling.

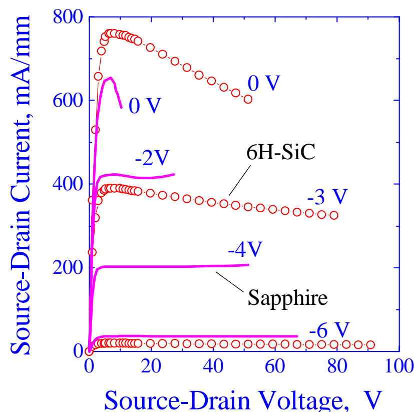

Fig. 9. I-V characteristics for devices grown on sapphire (dots) and 6H–SiC (open circles).[56]

3.2. *RF Characteristics*

3.2.1. *Power and cut-off frequency*

The RF parameters of interest are (see Fig. 10), the unity current gain frequency (f_T), the maximum oscillation frequency (f_{max}), the output power (P_{out}), Maximum Unilateral Gain (MUG) and Power Added Efficiency (PAE).

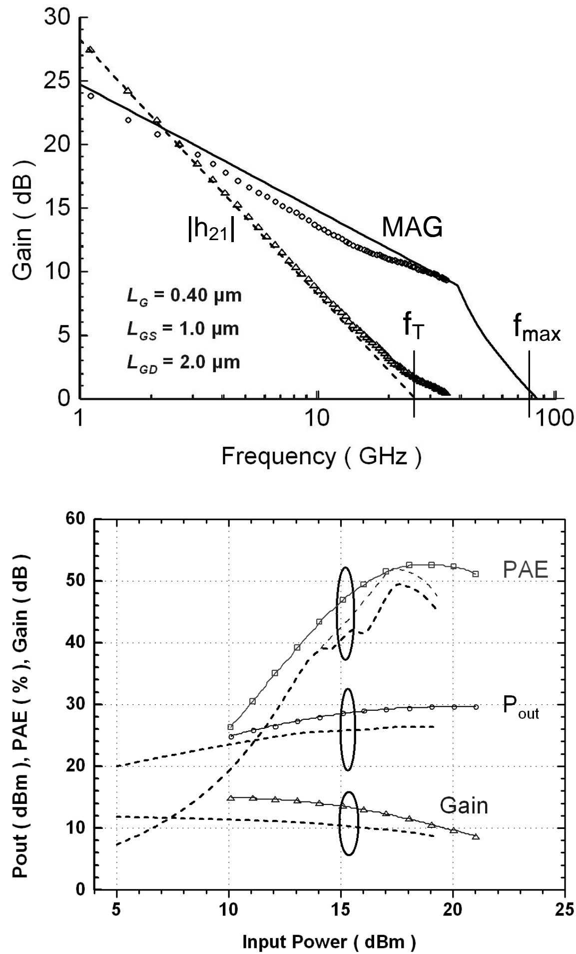

Fig. 10. RF characteristics, namely P_{out}, PAE, gain, f_T and $f_{\max}$ for the device of Fig. 6. The gain is obtained from the P_{out} versus P_{in} curve. Points represent experimental data, while dotted lines represent simulation.[95]

In general, MUG is the power gain of a two-port network having no output-to-input feedback, but with input and output conjugately impedance-matched to signal source and load, respectively. Because any RF transistor has a non-zero feedback from output to input, a lossless network must be added to cancel the feedback. Evidently, the resulting circuit will not oscillate unintentionally, so that MUG is defined over the whole frequency range. It is given by

$$\text{MUG} = \frac{|y_{21} - y_{12}|^2}{2[\text{Re}(y_{11})\text{Re}(y_{22}) - \text{Re}(y_{12})\text{Re}(y_{21})]} \quad (1)$$

or, in terms of the more commonly used S-parameters,

$$\text{MUG} = \frac{|S_{12}|^2}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \,. \quad (2)$$

Note that, MUG (dB) = 10 log MUG. This gain rolls off at −20 dB/dec at high frequencies.

The output power, P_{out}, is the power delivered to the load, and depends upon the amplifier circuit, e.g. class A, AB, or B, for a given device. However, it is common to find P_{out} being expressed as W/mm gate width, which is not exactly the total output power of the device. Since heat dissipation or battery power is of concern, we define the efficiency,

$$\text{PAE} = \frac{P_{\text{out}}(\text{ac}) - P_{\text{in}}(\text{ac})}{P_{\text{in}}(\text{dc})} \,. \quad (3)$$

Note that $P_{\text{in}}(\text{dc})$ is delivered by the DC supply, and is transformed into heat in the device. The variation of the PAE and P_{out} with P_{in} represent the large-signal behavior, of which, the saturated P_{out} is an important figure of merit.

The frequency at which MUG reaches unity (or 0 dB) is called $f_{\max}$. Its somewhat misleading designation, maximum frequency of oscillation, stems from the fact that it is also the highest frequency at which an ideal amplifier would still be expected to operate. On the other hand, f_{T}, also referred to as the gain-bandwidth product, is the frequency at which the small-signal current gain under output short circuit conditions, h_{21}, decreases to unity. In terms of S-parameters, h_{21} is calculated as

$$h_{21} = \frac{2S_{12}}{(1 - S_{11})(1 + S_{22}) + S_{12}S_{21}} \,; \quad (4)$$

$|h_{21}|$ rolls off at −20 dB/dec at high frequencies. Note that transistors with $f_{\max} > f_{\text{T}}$ can have useful power gains for $f > f_{\text{T}}$ and up to $f_{\max}$, since a current gain < 1 may be compensated by a voltage gain > 1 for $f_{\text{T}} < f < f_{\max}$. There is no unequivocal answer to the question, which of the two frequencies, f_{T} and $f_{\max}$, is more important, although generally it is assumed that f_{T} is important for digital and $f_{\max}$ for analog applications. Manufacturers strive to achieve $f_{\text{T}} \approx f_{\max}$ so that the devices have a wide application range. The thumb rules are, $f_{\text{T}} \geq f_{\text{op}}$ (the operating frequency) and $f_{\max} \geq 3f_{\text{op}}$ for power transistors, and $f_{\text{T}} \geq 2f_{\text{op}}$ for small-signal devices. Sometimes in literature, $f_{\max}$ is referred to as the frequency at

which the maximum available power gain (MAG) rather than the gain MUG, decreases to unity. This is not entirely correct, but the results of the two definitions are not significantly different.

The resistivity of the buffer/substrate layer has an important effect on the frequency response. Use of an actively compensated buffer doubled the $f_{\max}$ to 21 GHz from 10 GHz for a conventional buffer.[26] Similarly, the device temperature also affects the frequency behavior. As shown in Table 3, the f_T decreases significantly with temperature, on account of the decrease in saturation velocity. The f_T is a function of V_{DS}, e.g. Ref. 54. It increases rapidly with V_{DS} for small V_{DS} values (below about 10 V), due to increase in the electric field along the channel and hence electron velocity. However, it falls gradually after reaching a peak due to extension of the gate depletion region towards the source.

3.2.2. *Passivation effects*

Measured RF power parameters provided in Table 4 show that passivation improves P_{out} and PAE greatly; 100% increase in P_{out} and 28% increase in PAE are seen. Another study showed these values to be 145% and 11% respectively. (Note that the higher percentage increase in P_{out} is associated with a lower percentage increase in PAE, in keeping with the trade-off between P_{out} and PAE). These improvements should be attributed to the fact that passivation prevents channel depletion due to traps and surface charges, so that the gate can modulate the channel completely. Consequently, $I_{\max}$ increases and source and drain resistances decrease. There is a 30% reduction in gain, however, because of the increase in C_{GD} (gate-drain capacitance) due to two factors. First, the surface dielectric constant increases due to the passivating layer (Si_3N_4 in this example), and second, there is a reduction in gate-drain depletion layer because of the elimination of surface charges/traps which deplete the channel. These effects are also responsible for a reduction in f_T and $f_{\max}$ (of $\sim 9\%$).

Table 4. Measured RF power parameters of the GaN HEMT having $L_G = 0.5$ μm, at $f = 4$ GHz.[30]

Parameter	P_{out} (W/mm)	PAE (%)	Power gain (dB)
Unpassivated	1.0	36	20
Passivated	2.0	46	18.5

3.2.3. *Geometry scaling effects*

It is important to note the effects of scaling the gate geometry, W and L_G. Due to the random variation in parameters, e.g. V_T, over larger device areas, the power tends to saturate with increase in W, particularly at higher frequencies. Decreasing the channel length significantly raises P_{out} and gain, mainly due to an increase in f_T. However, for short channel lengths of $L_G < 0.1\ \mu$m, f_T may saturate, as parasitics from gate fringing and gate bonding pad capacitances become comparable to the gate capacitance.[1]

3.3. ***Reliability***

Since the technology of power HEMTs is less mature than that of silicon power devices, many failure mechanisms limit their useful life, i.e. mean time between failure (MTTF). Some failure mechanisms of silicon devices, such as electromigration, apply to HEMTs as well,[57] but there are many, which are different. Moreover, failure mechanisms of GaAs and InP devices are not necessarily transferable to GaN devices, whose technology is in infancy.

The technique used to determine MTTF of HEMTs is also a subject of discussion. A measurement based rather than theoretical technique is preferred.[57] The usual extrapolation of the MTTF corresponding to the required temperature (usually 125°C) from high temperature accelerated stress experiments may yield unrealistically high MTTF estimates, because the failure mechanisms of compound semiconductor devices frequently show high values of activation energy at temperatures > 200°C.[58] Hence, electrically accelerated stress methods are also being considered.[59]

Failures are classified into catastrophic and degradation failures. Their exact mechanism depends on device bias, resultant channel temperature, passivation, and interaction among materials including those used for packaging. Below, we provide an account of the various failure mechanisms. Some of these are unique to power HEMTs, and may not affect small-signal HEMTs. High power and efficiency come at the cost of reduced reliability due to several reasons. First, the close proximity of the channel to the gate, essential for a high transconductance, makes the device susceptible to surface effects. Second, the plasma treatments and/or wet etching associated with gate recessing, employed for reducing parasitic source/drain resistances, can cause surface damage. Third, high efficiency power amplification involves large quiescent drain voltage condition (see Sec. 1.2), resulting in high electric fields at the drain end of the gate.

The reliability considerations of GaN devices revolve around epitaxial growth, contact technology, implantation and dry etching.[26] Presently, the epitaxial buffer layer has high defect concentration and low resistivity. Refined epi-growth using Mg-compensation and LEO may provide defect free, high resistivity buffer layers. Ga-faced rather than N-faced buffers should be deposited to provide higher channel 2-DEG concentration. The standard Ti/Al ohmic contact system shows weak edge definition and needs to be modified to avoid occurrence of liquid phases during annealing. Pt-based Schottky contacts pose problems at temperatures $> 300°C$ attained by power devices. Their replacement by Re-based or WSiN-barrier metal system is being considered. Another problem is related to the implantation process used for device isolation. The high activation temperature (1100°C) of implanted dopants affects the integrity of semiconductor layers because of self-diffusion. Finally, dry-etching processes induce nitrogen vacancies enhancing surface conductivity. The hydrogen used in RIE gas compositions may compensate semiconductor layers.

Other failure mechanisms are related to trapping of hot electrons in the buffer and donor layers adjacent to the channel, or on the surface. Such trapping has been shown[60,61] to be the origin of transconductance frequency dispersion, drain current collapse, gate and drain lag transients, and restricted microwave P_{out}. Current collapse is the reduction in I_{D} at low V_{DS} after the application of a high V_{DS} (see Fig. 11). This has a characteristic time dependence of minutes, after which the current is restored through release of trapped charge by thermal emission or illumination. This phenomenon is due to hot electron injection and trapping in the buffer layer. The trapped charge reduces 2-DEG concentration in the active channel, thereby reducing I_{D}. Since surface is not involved, this effect cannot be controlled by passivation. But lower resistivity, i.e. lesser deep level concentration, in the buffer layer lowers this effect. On the other hand, gate lag, which is the delayed response of I_{D} to a V_{GS} change (see Fig. 11), is sensitive to surface states in the gate-to-drain/source access regions. The trapped charge in these states reduces 2-DEG concentration in the access regions, raising parasitic source and drain resistances. Thus passivation can control this effect. Both buffer and surface layer trapping effects restrict the practical microwave P_{out} to a value significantly below the estimations from the DC drain current and operating point. This is because the gate lag due to surface traps enhances large signal response time, while buffer layer traps reduce I_{max} and increase the knee voltage.

Recently, Simin *et al.* reported on a novel, current collapse free, Double Heterostructure AlGaN/InGaN/GaN Field Effect Transistors (DHFETs)

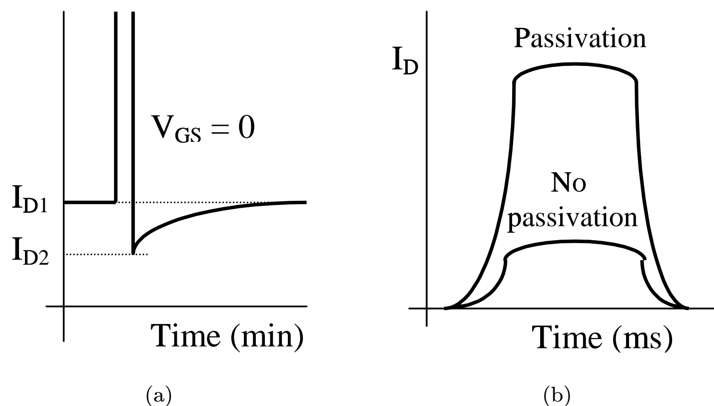

Fig. 11. Schematics illustrating drain current response to short voltage pulses, due to trapping effects. (a) Drain collapse: V_{DS} is pulsed between a low value (10–100 mV) to a high value (15–20 V). (b) Gate lag: V_{GS} pulsed from V_T to 0 V, keeping V_{DS} low (10–100 mV).

fabricated on the insulating SiC substrates.[90] Their simulations showed that a combined effect of the bandgap offsets and polarization charges provides an excellent 2D carrier confinement. These devices demonstrated output RF powers as high as 4.3 W/mm in CW mode and 7.5 W/mm in the pulsed mode, with the gain compression as low as 4 dB.

3.4. *Uniformity*

Manufacturing uniformity control is a critical issue, because high cost represents a significant road block in the development of RF systems. For cost-effective manufacturing, continuous process yield diagnostics and improvement are essential. Understanding of this issue requires a statistical study of characteristics of a large number of devices on an experimental wafer. For this purpose, simple DC measurements on actual transistors without any specialized test structures may suffice. Uniformity of device parameters is strongly process related.

Reproducibility and yield can be enhanced using C-doping for the GaN insulating buffer, and precision temperature control during growth of undoped GaN channel by reactive ammonia MBE. This is confirmed by an analysis of 2-DEG mobility values obtained in 50 experiments on 2 inch diameter substrates.[62] About 75% of the experiments yielded values higher than 800 cm^2/Vs. Reproducibility measurements from five consecutive growth runs showed that the variation in the electron density values

was < 7.5%, and in electron mobility values was < 5.7%; the average values were $1.54 \times 10^{13}/\mathrm{cm}^2$ and 987.6 $\mathrm{cm}^2/\mathrm{Vs}$ respectively. *C*-doping provides both a high resistivity buffer and a low defect density channel. Note that, a thick insulating buffer is essential for device isolation and good RF performance.

Figure 12 gives the distributions of $I_{\max}$, f_{T} and $f_{\max}$, and Fig. 13 gives the variation of the latter two, across a 2 inch diameter wafer.[63] The f_{T} and $f_{\max}$ have maximum values of 8.3 and 10.2 GHz, and average values of 7.7

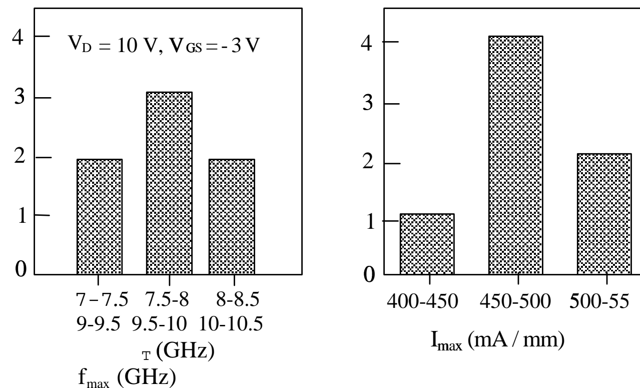

Fig. 12. The spread in critical frequencies and $I_{\max}$ distributions of AlGaN/GaN HEMTs on 2 in diameter sapphire at $T = 23°\mathrm{C}$ (after Ref. 63).

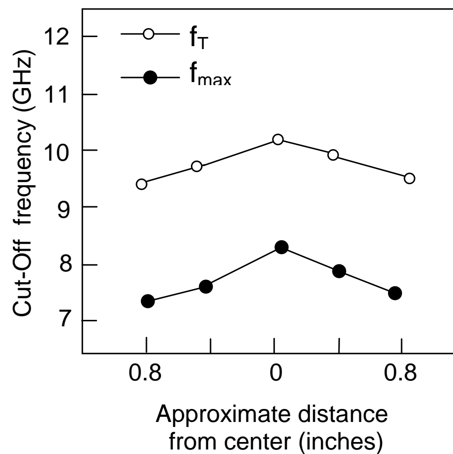

Fig. 13. Spatial distribution across 2 inch wafer of cut-off frequencies for AlGaN/GaN HEMTs at 23°C and $V_{\mathrm{DS}} = 10$ V, $V_{\mathrm{G}} = -3$ V (after Ref. 63).

and 9.8 GHz. The $I_{\max}$ has minimum and maximum values of 473 mA/mm and 505 mA/mm, and an average of 473 mA/mm.

3.5. *GaN versus GaAs and InP Power HEMTs*

We now compare GaN HEMTs with GaAs and InP HEMTs, based on their characteristics. Some important details appear in Table 5. GaAs pHEMTs have been a commercial reality for a few years, and cover an extremely wide frequency range. On the other hand, GaN and InP HEMTs are just more than a laboratory product. However, they have important niche markets and are striving to fully emerge from technological infancy. Table 5 shows that InP devices are the best bet for higher frequencies, since they have the highest f_{T} and $f_{\max}$ due to the high v_s of InP (note f_{T} is proportional to v_s). Useful amplification up to 200 GHz has been reported using InP MMIC amplifiers.[64] GaN devices have the highest V_{br} due to the high channel bandgap and associated critical field E_{c}. Also, their $I_{\max}$ and g_{m} are higher than other devices of the same gate-length, on account of the high n_{so} due to polarization effects and the high v_s. Therefore, the highest $P_{\mathrm{out}} =$ 9.8 W/mm at 8 GHz (PAE = 47%) has been obtained from GaN HEMTs.[65] This is $\sim$ 6 times the highest $P_{\mathrm{out}} = 1.6$ W/mm at 2 GHz (PAE = 62%) of the widely prevalent GaAs pHEMT.[5] With continued improvements in material quality and device design, a power density of 12 W/mm or higher appears practical. Recently, it has also been shown[66] that these devices can achieve low microwave added Noise Figures (NF = 0.6 dB at 10 GHz) while maintaining a large breakdown voltage (> 60 V) and hence a large dynamic range. These results imply that AlGaN HEMTs can be used to perform the active transmit and receive functions in more robust, higher dynamic range modules. Thus for X-band applications, GaN HEMTs are a powerful alternative to GaAs pHEMTs. As mentioned in Sec. 2.2, very high voltage low frequency (f < 100 MHz) GaN HEMTs for switching power control are also possible.

To make full use of the high voltage capability, the transistor must not be thermally limited. GaN also has an advantage over GaAs in this regard, with thermal conductivities up to 2.0 W/cmK for GaN versus 0.46 W/cmK for GaAs. Furthermore, GaN HEMTs can be grown on semi-insulating SiC with a thermal conductivity of 3.3 W/cmK. The high breakdown field and good thermal conductivity allow these devices to be used in high efficiency (theoretical efficiency of 78.5%) class B push/pull amplifiers at full power

Table 5. Typical values of parameters in various power HEMTs.

Device	Hetero-junction	$\Delta\phi_c$ (eV)	n_{so} (cm^{-2})	μ (cm^2/Vs)	V_{br} (V)	f_T (GHz)	f_{max} (GHz)
GaN HEMT	$Al_{0.3}Ga_{0.7}N/GaN$	0.45	1.5×10^{13}	1180	40	65	180
GaAs PHEMT	$Al_{0.3}Ga_{0.7}As/GaAs$	0.24	1.0×10^{12}	8500	7	130	250
InP HEMT	$Al_{0.5}In_{0.5}As/In_{0.6}Ga_{0.4}As$	0.64	3.4×10^{12}	13000	5	290	480

rating. GaAs microwave devices, on the other hand, are usually implemented in a class B push/pull amplifier by backing off the voltage bias (and hence the power level) to accommodate the higher voltage swing in this configuration as compared to class A, single-ended operation.

Other advantages of GaN are that the intrinsic temperature is much higher (due to wide bandgap), and the v_s less temperature sensitive, than in GaAs. So, GaN power devices are more suitable for high temperature applications than GaAs devices, and can operate with less cooling and with fewer high cost processing steps associated with complicated structures designed to maximize heat extraction. Also, electrons in GaN exhibit much more pronounced overshoot effects than in GaAs but at much higher electric fields. This result illustrates the potential of GaN for ballistic power devices.[4]

3.6. *HEMTs versus MESFETs and HBTs*

It is of interest to compare HEMTs *vis-à-vis* their other compound semiconductor counterparts, namely MESFETs and HBTs. The advantages of HBTs are a higher current and power density, due to vertical current transport, which offers better utilization of wafer area.[67] This allows compact devices with smaller periphery, and therefore, easy matching with the surrounding circuit. Also, the device dimensions critical for HBT speed are not planar but vertical and thus independent of the photolithography process. These devices also have better linearity and more uniform threshold voltage. However, the HEMT has lower noise and better performance at high frequencies. Further, as compared to MESFET it has a higher transconductance, stemming from two factors. First is the close confinement of channel carriers to the gate, and second is the high mobility of the carriers due to diminished impurity scattering. The carrier confinement close to the gate also results in lower short-channel effects.

4. Modeling

The various models for calculating the DC and RF characteristics are now described. These may be employed to design a power HEMT, or an amplifier circuit based on this device. The model equations are presented following a discussion of the basic phenomena, namely gate-controlled charges, carrier transport, hot electrons, and polarization effects.

4.1. *Basic Phenomena*

4.1.1. *Gate-controlled 2-DEG*

The DC and RF characteristics, i.e. I_D or g_m and the capacitance, are controlled by the behavior of 2-DEG (n_s) as a function of the V_GS (see Fig. 14). The n_s–V_GS behavior is non-linear at low V_GS due to 2-DEG movement. A unified expression for the subthreshold and above threshold behavior has been given as[68]

$$V_\mathrm{GS} - V_\mathrm{T} = \left(1 + \frac{d}{W_d}\frac{\varepsilon_d}{\varepsilon_b}\right)\frac{kT}{q}\ln\left(\frac{n_s}{n_0}\right) + q\left(\frac{d}{\varepsilon_d} + \frac{\Delta d}{\varepsilon_b}\right)n_s\,, \tag{5}$$

and

$$W_d = \sqrt{\frac{\varepsilon_b \phi_\mathrm{G}}{(qN_b)}}\,. \tag{6}$$

Here, d is the depth of the heterojunction from the gate, Δd ($\sim$ 4.5 nm) is the location of the 2-DEG from the heterojunction, $n_0 = 5 \times 10^{10}/\mathrm{cm}^2$, W_d is the channel depletion width, N_b is the channel doping.

On the other hand, following approach Refs. 69 and 70 gives a simple and explicit expression for the nonlinear behavior, including the gradual saturation at high V_GS (see Fig. 14), over the complete bias range of interest. First, a simple function is chosen to represent the detailed shape of the charge–voltage curve. For example, above threshold, a function that represents the n_s–V_GS behavior including gradual saturation and gradual

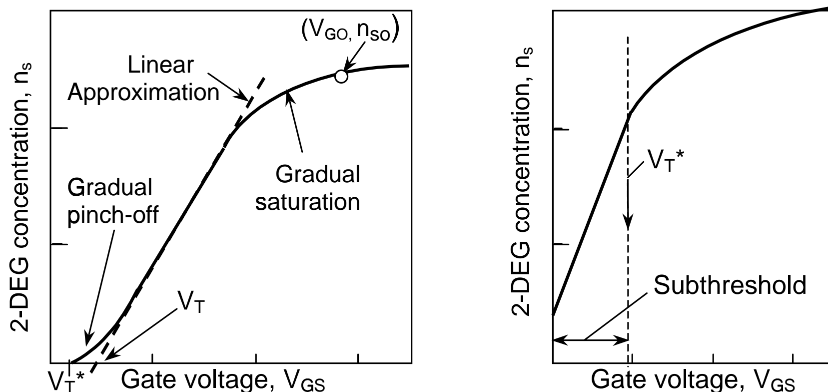

Fig. 14. Gate controlled charges in a HEMT. (a) Above threshold behavior; (b) Subthreshold behavior of the 2-DEG charge, (after Ref. 69).

pinch-off can be written as[69,71]

$$n_s = n_c\left[\alpha + (1-\alpha)\tanh\left(\frac{V_{\mathrm{GS}} - V_{\mathrm{GM}}}{V_1}\right)\right] \quad V_{\mathrm{T}}{}^* < V_{\mathrm{GS}} < V_{\mathrm{GO}}\,. \tag{7}$$

Here, $V_{\mathrm{T}}{}^*$ is the threshold voltage representing the pinch-off condition and is less than the linearly extrapolated V_{T}, due to the gradual pinch-off nature of the I_{DS}–V_{GS} curve. V_{GO} is a critical voltage[8] representing the onset of gate control on the 2-DEG, and n_c, α, V_{GM} and V_1 are the constants to be determined in terms of device parameters. Next, the constants are expressed in terms of device parameters, by forcing the function to pass through (intersect) readily derivable critical points (lines) on the accurate curve. These provide the required number of simultaneous equations to solve for the constants. For example, we need four such equations for the constants in n_s–V_{GS} function (Eq. (7)). Two of these are obtained by forcing the function to pass through points ($V_{\mathrm{T}}{}^*$, 0) and (V_{GO}, n_{so}); the other two are obtained by forcing the function to intersect the well known linear n_s–V_{GS} approximation,

$$n_s = \theta(V_{\mathrm{GS}} - V_{\mathrm{T}}) \quad V_{\mathrm{T}} \le V_{\mathrm{GS}},\, n_s \le n_{\mathrm{so}}\,. \tag{8}$$

Note that $V_{\mathrm{T}}{}^*$, V_{T}, V_{GO}, n_{so}, and θ are readily derived in terms of device parameters.[69] Current (capacitance) expressions may be derived by integrating (differentiating) Eq. (7) respectively.

4.1.2. *Carrier transport*

Since power HEMTs are compound semiconductor devices, the carriers flowing in the bulk between the source and drain contacts through high field regions, experience not only velocity saturation, but also velocity overshoot preceding velocity saturation. Therefore for accurate modeling of the current, a simple Drift-Diffusion (DD) transport model, which can simulate velocity saturation alone, is not adequate. A Hydrodynamic (HD) model is required for capturing velocity overshoot effects on the current.[72]

Current through the contacts requires accurate modeling of the transport across interfaces. The Schottky contact governs I_{G}. The metal-semiconductor interface associated with this contact is clean and sharp, so that its modeling is straightforward. However, the ohmic source/drain contact, which governs the all-important I_{D}, is a concern. This contact interface is rough due to metal alloy penetration into the semiconductor. There are two possible models for transport across contact interfaces: Thermionic Emission (TE) without tunneling, and Thermionic Field Emission (TFE),

which includes tunneling. For Schottky contacts, these apply for low and high reverse bias, respectively. For ohmic contacts, the latter has been found to be more appropriate.[72]

It is also necessary to include effects of temperature. This is because, power HEMTs have to dissipate large amount of power, so that heat sinking is a significant matter. A substrate with low thermal conductivity, such as sapphire, which is, presently, the preferred choice for growing AlGaN/GaN power HEMT, may accentuate self-heating effects, and these need to be modeled.

So far, analytical modeling of the velocity overshoot, TFE and temperature effects has been difficult, and numerical simulation techniques have been employed.[56,72,73]

4.1.3. *Hot electrons*

High-mobility electrons are easily "heated" up by the channel field. This effect is exacerbated by the use of high V_{DS} for higher P_{out}, and is intense near the gate edge close to the drain where the electrical field is the highest. A few electrons can accumulate sufficient energy to overcome the heterojunction barrier, resulting in the reversible charging/discharging of deep levels in the donor and/or interface layers. Similarly, a few hot electrons may overcome the channel potential well, and get injected and trapped in the buffer layer beneath. In principle, Monte Carlo simulation approach employed to model hot electron effects in GaAs and InP HEMTs,[74,75] may be employed for GaN HEMTs as well. In GaN HEMTs, these phenomena may affect the DC characteristics through transconductance frequency dispersion, current collapse, and the RF performance through gate and drain lags.[60,61] However, so far, modeling of these phenomena has mostly remained qualitative.[60,61]

4.1.4. *Polarization effects*

Group-III nitrides such as AlN, GaN and InN, exhibit large polarization effects at heterointerfaces.[4] The origin of polarization is two-fold: piezoelectric effects due to stress along the *c*-direction of these ionic and non-centrosymmetric materials, and the difference in spontaneous polarization between these. The polarization dipole in the AlGaN layer is responsible for the high $\sim 10^{13}/\mathrm{cm}^2$ 2-DEG concentration in the AlGaN/GaN HEMTs. Although this effect is complex, and numerical simulation has been employed for accurate results,[76] in cases such as

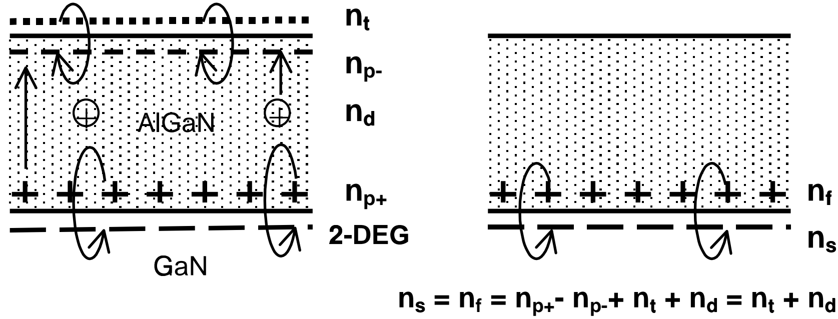

Fig. 15. Polarization effects in AlGaN/GaN heterojunction. Actual picture (left); simplified model (right), (after Ref. 37).

simulation of V_{br} enhancement using field plate techniques, the simple model shown in Fig. 15 has worked well.[6,19,37]

4.2. *Analytical Modeling*

4.2.1. *Critical DC and RF parameters*

The maximum 2-DEG concentration, n_{so}, equals the charge transferred across the heterojunction under equilibrium. In the case of AlGaN/GaN heterojunctions, $n_{\text{so}} \sim 10^{13}/\text{cm}^2$ is decided primarily by polarization effects in AlGaN, and is almost independent of donor doping.

Based on the threshold voltage expressions given in Refs. 10 and 68, the threshold voltage corresponding to the linearly extrapolated I_{D}–V_{GS} characteristics to the V_{GS} axis, at low V_{DS}, can be written as[10,68]

$$V_{\text{T}} = \phi_b - \Delta\phi_c - \frac{q}{\varepsilon}\left[\frac{N_d {d_d}^2}{2} + n_f d_d - N_b W_d\right], \tag{9}$$

where ϕ_b is the gate Schottky-barrier height, n_f is the effective positive polarization charge in AlGaN (can be regarded equal to n_{so}) defined in Fig. 14, and d_d is the distance between the gate and the donor/spacer interface in the gate region. The expression for the threshold voltage representing pinch-off condition, ${V_{\text{T}}}^*$, can be written as,[69]

$${V_{\text{T}}}^* = \phi_b - \Delta\phi_c + \delta_3 - \frac{q}{\varepsilon}\left[\frac{N_d {d_d}^2}{2} + n_f d_d - \left(N_b W_d + \frac{n_{\text{so}}}{100}\right)(d_d + d_i)\right], \tag{10}$$

where δ_3 is a doping dependent potential. The difference $(V_{\text{T}} - {V_{\text{T}}}^*)$ can exceed 100 mV.

In short-channel devices ($L_G < 0.25\ \mu$m), V_T (or $V_T{}^*$) will reduce (see Fig. 8) below the above value of the long channel HEMTs, and this reduction is a function of V_{DS}. The movement of the 2-DEG location accentuates this V_T reduction. To obtain short-channel V_T, numerical simulation involving a simultaneous 2D solution of Schrödinger's and Poisson's equations is necessary. This procedure may be simplified using an analogy between the HEMT and a Buried-Channel (BC) MOSFET.[77] Here, the HEMT is replaced by an equivalent BC-MOSFET, whose oxide thickness and BC doping profile are adjusted so that the BC depletion edge simulates the movement of the 2-DEG location in the HEMT. The V_T of the BC-MOSFET is obtained by solving the Poisson's equation alone, and corresponds to that of the HEMT. For circuit simulation purposes, the following simple equation is adequate

$$V_T(\text{or } V_T{}^*) = V_{T0}(\text{or } V_{T0}{}^*) - \sigma V_{DS}\,. \tag{11}$$

The parameter σ may be extracted experimentally.

Simple and approximate expressions for I_{max} and intrinsic $g_{m(peak)}$ (per unit width) are given by[4,78]

$$I_{max} \approx q n_{so} \nu_s \tag{12}$$

$$g_{m(peak)} \approx \frac{q n_{so}\mu / L_G}{\sqrt{1 + \left(\frac{\theta n_{so}\mu(d+\Delta d)}{\varepsilon_d \nu_s L_G}\right)^2}}\,. \tag{13}$$

The complete I_D–V_{DS} characteristics including all effects can only be derived numerically. For purposes of hand calculation and circuit simulation, however, analytical models are necessary. Below, we discuss analytical models followed by numerical models.

The frequencies f_T and f_{max} can be expressed in terms of the small-signal device equivalent circuit parameters (see Fig. 16 later) as[79]

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GD})} \tag{14}$$

$$f_{max} = \frac{f_T}{\sqrt{4\{1 + [R_S + R_G]g_o\} + 2\left(\frac{C_{GD}}{C_{GS}}\right)\left[\frac{C_{GD}}{C_{GS}} + g_m\left(R_S + \frac{1}{g_o}\right)\right]}}\,. \tag{15}$$

Extrinsic values of g_m and C_{GS} are given by $g_m{}^* = g_m/(1 + g_m R_S)$ and $C_{GS}{}^* = C_{GS}(1 + g_m R_S)$. In the active bias region of a HEMT ($V_{DS} > 0.5$ V and $V_{GS} > V_T$) C_{GD} is only about 15% of C_{GS}. Therefore f_T can also be

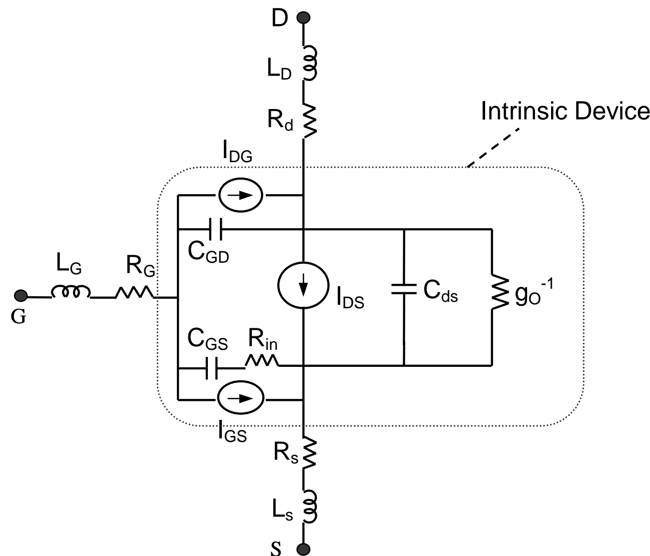

Fig. 16. RF large signal equivalent circuit model. In the small-signal model, the current sources I_{DG} and I_{GS} are removed (after Ref. 87).

approximated to be $\approx g_m{}^*/2\pi C_G$, where $C_G = C_{GS}{}^* + C_{GD}$ is the extrinsic total gate capacitance.

4.2.2. *Drain current and power*

An I_D–V_{DS} expression can be derived by integrating analytical n_s–V_{GS} and n_f–V_{GS} expressions, e.g. those given in Sec. 4.1.1, assuming gradual channel approximation and simple saturating velocity-field curve ignoring velocity overshoot. Such expressions have been shown to represent the experimental results adequately (e.g. see Ref. 69). We shall discuss detailed simulations of an experimental GaN HEMT using AIM-Spice HFET model.[80–82] The goal of our simulation is to establish a device model, valid at both DC and microwave frequencies. To this end, we compare the results of such a simulation with the experimental data from GaN/AlGaN HFETs from Cree Research, Inc.[83]

The HFET model implemented in AIM-Spice uses a relatively small number of scalable parameters and allows for easy parameter extraction from the experimental data.

From DC measurements[83] on the 0.4 and 0.45 μm gated channel GaN/AlGaN HFETs, we have extracted the following parameter

values: gate-to-channel separation 27 nm, output conductance parameter 0.005 V^{-1}, low-field mobility 1900 cm^2 $V^{-1}s^{-1}$, maximum sheet charge density in the channel 2.4×10^{12} cm^{-2}, Schottky barrier height 1.1 eV, drain series resistance 13 Ω, gate series resistance 12 Ω, source series resistance 13 Ω, saturation velocity 1.9×10^7 cm/s, knee voltage parameter 4, threshold voltage −2.1 V, charge partitioning parameter 1.2. The automatic extraction program described in Ref. 84 and parameter adjustment procedure were employed for this purpose. The agreement between measured and simulated results is quite reasonable (see Fig. 6). However, at $V_{GS} = 2$ V, the measured I_D decreases with increased V_{DS} due to self-heating. The extracted value of the saturation drift velocity, $v_s = 1.9 \times 10^7$ cm/s, is higher than the values extracted from the I-V characteristics and microwave data for other GaN-based HEMTs ($v_s = 1$–1.5×10^7 cm/s).[85]

The extracted value of the maximum 2D-electron density, $n_{max} = 2.4 \times 10^{12}$ cm^{-2} is lower than the Hall measurement value of over 7×10^{12} cm^{-2}.[83] This might reflect a considerable transfer of hot electrons into the three dimensional states in GaN or into the AlGaN barrier layer at high V_{GS} as evidenced by a drop in g_m at high V_{GS} (see Fig. 6). The thermal impedance is $\sim$ 2.5 Kmm/W,[85] and the power dissipation is around 5–15 W/mm. Hence, the operating temperature might $\sim$ 380–390 K, and the deduced parameters might be somewhat different from those expected at $T = 300$ K.

Microwave simulations of the device should allow us to better understand microwave device performance, and to establish the expected performance limitations of such devices. So far, simulations of GaN/AlGaN HFETs have been mostly limited to DC fitting, although Trew *et al.* reported on microwave simulations of GaN/AlGaN HFETs using the harmonic balance method.[86] The parameters P_{out}, gain and PAE depend on the entire circuit configuration including the device equivalent circuit, and are obtained by circuit simulation. An equivalent circuit model useful for this purpose is given in Fig. 16. Here, the current I_{DS} is the large signal drain-to-source current. I_{DG} and I_{GS} represent the drain to gate to source currents, respectively. At large signal operation if V_{DG} exceeds the breakdown voltage (V_B) or/and V_{GS} exceeds the Schottky cut-in voltage (V_{bi}), respectively; simple linear relationships, $I_{DG} = (V_{DG} - V_B)/R_1$ and $I_{GS} = (V_{GS} - V_{bi})/R_F$, may be used.[87] The currents I_{DG} and I_{GS} may be represented by the two diodes used to model the gate current, as shown in Fig. 17 and discussed later.

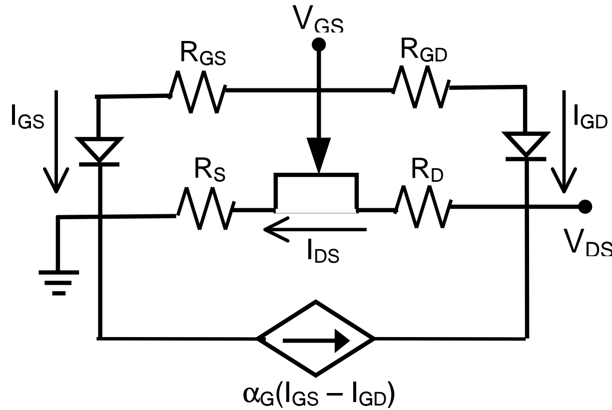

Fig. 17. Power HEMT equivalent circuit accounting for gate leakage (after Ref. 81).

The equivalent circuit parameters may be calculated using formulae based on device analysis. Alternately, they may be extracted from measured S-parameters. For this purpose, initially, a reasonable value is assumed for each parameter, and for this assumption, the S-parameters of the circuit are calculated. This procedure is repeated by modifying the assumption until the differences between the calculated and measured S-parameters is minimum.

We implemented the AIM-Spice small-signal model in the microwave simulator Libra (see Fig. 10). The device is biased under class-A operation in the large signal simulation (V_{gs} = 0.7 V and = 21.2 V). The largest simulated output power is close to 4.5 W/mm, while the experimental value is 6.8 W/mm. Some of this discrepancy can be attributed to self-heating, which might play a smaller role at microwave frequencies than at DC, since power is low during a substantial part of the microwave cycle.

4.2.3. *Gate current*

For values of V_{DS} low enough so that breakdown does not occur, an equivalent circuit showing the effect of I_{G} is given in Fig. 17.[81] Each of the two diodes in this figure obeys the conventional Schottky diode law

$$I_{\text{G}} \approx J_{\text{SS}} L_g W \left[\exp\left(\frac{qV}{mkT} \right) - 1 \right] . \tag{16}$$

Here, J_{SS} is the reverse saturation current density calculated using the TE or TFE model. V and m for the diodes may differ. Note that I_{G} affects I_{D}.

This is because I_G is distributed along the channel, with the largest density near the source side of the channel. This redistributes the channel electric field, increasing it near the source, and decreasing I_D.

4.3. *Numerical Modeling*

4.3.1. *Drain current*

Several reports discussing numerical I_D–V_{DS} models are available.[6,19,37,56,72,73,76] These show that device geometry used for simulation is extremely important. The ohmic source/drain contacts should be assumed to be on the top of the cap layer. Results of any other assumption, such as contacts directly on the channel, or on the cap with heavy doping between the contacts and the channel, deviate significantly from experiment. Furthermore, the choice of transport models is very important. TFE model should be assumed for the transport across the ohmic contact interface; a TE model gives currents much lower than measured. Assumption of a DD model for transport parallel to the hetero-interface grossly underestimates the I_D and $g_{m(peak)}$, and overestimates the width of the flat portion of g_m–V_{GS} curve. An HD model in the channel (and to accurately describe the parasitic MESFET behavior occurring for high V_{GS}, a HD model in the donor layer as well) is required. A simultaneous fitting of the currents and capacitances is necessary to separate the effects of the carrier concentration and the velocity. The carrier concentration and surface potential within the ungated source-gate and gate-drain regions vary widely, so modeling of these regions as linear resistors is inaccurate. Some other issues will be considered in the context of breakdown simulation below.

4.3.2. *Breakdown voltage*

The breakdown voltage in AlGaN/GaN HEMTs is at least several tens of volts, so that $V_{br(on)}$ measurement is difficult due to excessive self-heating, and $V_{br(off)}$ is the relevant parameter. So far, impact ionization has been assumed to be the cause of breakdown. Modeling of polarization charge is an important issue. Generally, numerical simulation has been employed to predict $V_{br(off)}$.

Because the intrinsic V_{br} and I_{max} of GaN devices can be > 50 V and > 600 mA/mm [see Table 2, Eq. (12)], respectively, some numerical simulation studies have explored the possibility of raising the V_{br} to as high as 600–1000 V using field plate structures V (see Figs. 3(c), (d), and Refs. 6,

19 and 37). This was with a view to consider high voltage power switching applications for this device. In these simulations, DD transport with impact ionization was used. The V_{br} was directly extracted from the simulated I_{D}–V_{DS} curve for $V_{\text{GS}} = V_{\text{T}}$. This approach is better than the indirect inference of V_{br} from the peak electric field condition. The peak field is sensitive to the simulation mesh while the area under the field at breakdown is not, and the impact ionization depends not only on the peak but also on the shape of the field distribution near the peak. The simplified polarization charge model of Fig. 15 was employed, as the spatial separation of the polarization and other charges associated with the AlGaN layer was irrelevant due to the presence of a thick insulator between the AlGaN layer and the field plate. A dielectric medium above the device was found to raise the V_{br} by 6–20%. These simulations showed that an optimized gate field plate over uniform insulator configuration can enhance the V_{br} by as much as 5 times, depending upon the dielectric constant of the insulator below the field plate, and the value of n_s.[37] By way of example, the simulated reduction in peak electric field for this configuration is illustrated in Fig. 18. More than 10 times enhancement in V_{br} is predicted, either by stepping the insulator thickness along the channel (see Fig. 3(d) and Ref. 19), or by introducing a *p*-*n* junction below the channel along with a drain field plate (see Fig. 3(c) and Ref. 6).

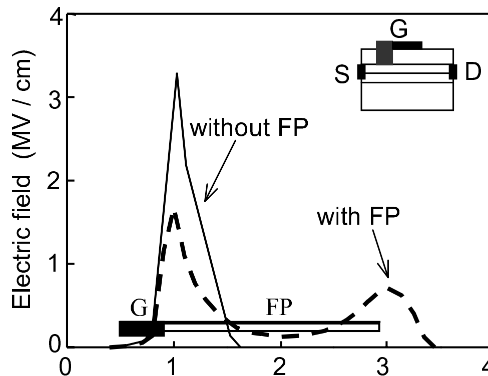

Fig. 18. The simulated distribution of the electric field along the 2-DEG, directed from drain to source, with and without a field plate (FP).[29] The FP extends 2 μm from the gate ($L_{\text{G}} = 0.4$ μm), and is deposited on a 0.3 μm silicon nitride insulator; $n_s = 1 \times 10^{13}/\text{cm}^2$; V_{DS}, $V_{\text{GS}} = 123$ V, -2.8 V, respectively. The device without FP is on the verge of breakdown.

4.3.3. *Negative resistance*

Mobility reduction with rise in temperature is the main cause of the downward sloping current curves in Fig. 9. The device temperature increases at high I_D — high V_{DS} due to enhanced power dissipation. Proper modeling of this effect requires accurate determination of channel temperature distribution from the Laplace Heat Conduction Equation, which is possible by numerical simulation. Here, definition of the heat generation source is a crucial issue. Single[73] and dual[56] heat source models have been employed. Alternative assumptions for the heat source length are: the gate, the gate-drain and the source-drain regions. Predictions for these assumptions can vary by as much as 30%. The dual heat source model assumes the gate length plus the drain depletion depth as one source, and the ohmic contacts together with the source/drain access regions as the other source. The total power is divided between the two sources in the ratio of their voltage drops. Temperature rises caused by the two heat sources are superimposed to get the overall temperature distribution. In all the models, the strong temperature dependence of thermal conductivity of substrate is accounted for. Simulations with the dual heat source model showed a channel temperature of 360°C for 4 W/mm dissipation on sapphire substrates.[56] For such high temperatures, the mobility can degrade by a factor of three as compared to room temperature.

5. Circuits

Here, we shall discuss the circuit configuration of a broadband power amplifier for X-band (0.2–6 GHz) phased array-radar and instrumentation. Simple lumped broadband amplifiers have gain-bandwidth products limited by f_T. Distributed or Traveling Wave Amplifiers (TWAs) can provide gain-bandwidth products up to f_{max}. However, a lumped amplifier based on f_T-doubler topology[88] can provide higher PAE and occupy smaller die area than the distributed amplifiers.

The basic f_T-doubler configuration is shown in Fig. 19(a). Essentially, it is a modified Darlington stage where suitable source loading of Q_1 splits the input voltage equally between Q_1 and Q_2, so that $I_{D1} = I_{D2}$ at all frequencies. Note that the two devices have half the width of a common-source stage of equal P_{out}. The current gain of the f_T-doubler stage becomes unity at $2f_T$. Thus, the bandwidth limitation due to C_{GS} is alleviated. To absorb the effect of C_{DS} and C_{GD}, broadband L-C π-sections may be used between the amplifier output and the load. The f_T-doubler has been implemented

(a) (b)

Fig. 19. f_T-doubler resistive feed back amplifier. (a) Basic configuration; (b) Complete circuit diagram.[89]

using 0.7 μm gate-length AlGaN/GaN HEMTs.[89] The complete circuit diagram is given in Fig. 19(b). Resistive feedback has been used to match the input and output to 50 Ω. The die size is 1.1×1.45 mm^2. Following parameters were measured: gain = 11 dB over 0.2–7.5 GHz, $P_{\text{out}} > 1$ W and PAE > 10% over $1 - 7$ GHz.

6. Summary

In this chapter, we have presented an account of the evolution, technology, DC and RF characteristics, modeling and circuits of GaN based power HEMTs. These devices are the focus of intense research, as they are a powerful alternative to GaAs pHEMTs for use in transmitter and receiver sections of mobile communication sets operating in the X-band. A power density of 12 W/mm or higher appears practical for these devices, which can also achieve low noise figures (< 0.6 dB at 10 GHz) while maintaining a large breakdown voltage (> 60 V) and hence a large dynamic range. Very high voltage (∼ 1 KV) low frequency ($f < 100$ MHz) switching power control is also possible using this device. However, the technology of these devices is still in infancy, so that a number of processes and hot electron effect related reliability issues afflict this device. These issues have been pointed out. The discussion on technology considered various device structures, and critical fabrication steps including the all important material growth. The effects

of passivation, device geometry scaling and temperature on drain current and RF power were presented. It was pointed out that some of these effects, particularly those caused by thermal and electron heating, need numerical techniques for accurate modeling. However, simple analytical models were discussed, which are useful for preliminary design calculations of the device current and frequency parameters, and for simulating microwave power performance in circuit applications. Finally, a circuit configuration of this device for X-band amplification was considered by way of example.

References

1. F. Schwierz and J. J. Liou, *Microelectron. Reliab.* **41**, 145–168 (2001).
2. S. Arai and H. Tokuda, *Solid-State Electron.* **41**, 1575–1579 (1997).
3. S. Takamiya *et al.*, *Solid-State Electron.* **38**, 1581–1588 (1995).
4. M. S. Shur, *Solid-state Electron.* **42**, 2131–2138 (1998).
5. S. J. Pearton *et al.*, *Materials Science and Engineering*: *Reports* **30**, 55–212 (2000).
6. S. Karmalkar *et al.*, *IEEE Electron Device Lett.* **22**, 373–375 (2001).
7. I. Daumiller, C. Kirchner, M. Kamp, K. J. Ebeling, L. Pond, D. E. Weitzel and E. Kohn (Device Research Conference Digest, 1998) pp. 114–115.
8. S. Karmalkar, *IEEE Trans. Electron. Devices* **45**, 2187–2195 (1998).
9. H. Morkoc *et al.*, *Principles and Technology of MODFETs* Vols. 1 and 2 (Wiley, New York, 1991).
10. M. Shur, *GaAs Devices and Circuits* (Plenum press, New York, 1997).
11. L. Esaki and R. Tsu, *IBM Research*, RC 2418 (1969).
12. R. Dingle *et al.*, *Appl. Phys. Lett.* **33**, 665 (1978).
13. T. Mimura *et al.*, *Jap. J. Appl. Phys.* **19**, L225–L227 (1980).
14. D. Delagebeaudeuf *et al.*, *Electron Lett.* **16**, 667–668 (1980).
15. M. A. Khan *et al.*, *Appl. Phys. Lett.* **58**, 2408–2410 (1991).
16. M. A. Khan *et al.*, *Appl. Phys. Lett.* **65**, 1121–1123 (1994).
17. Q. Chen *et al.*, *IEEE Electron Device Lett.* **19**, 44–46 (1998).
18. Y.-F. Wu *et al.*, *Electron. Lett.* **33**, 1742–1743 (1997).
19. S. Karmalkar and U. K. Mishra, *Solid-State Electron.* **45**, 1645–1652 (2001).
20. Akira Usui *et al.*, *Jap. J. Appl. Phys.* Part 2 **36**, L899–L901 (1997).
21. Q. Chen, *Materials Science and Engineering* **B59**, 395–400 (1999).
22. S. Fischer *et al.*, *Symp. Proc.* EGN-1, 1996.
23. P. Kung *et al.*, *Appl. Phys. Lett.* **69**, 2116–2118, (1996).
24. F. Semond *et al.*, *Appl. Phys. Lett.* **78**, 335–337 (2001).
25. N. Q. Zhang *et al.*, *IEEE Electron Device Lett.* **21**, 421–423 (2000).
26. J. Würfl, *Microelectron. Reliab.* **39**, 1737–1757 (1999).
27. S.-H. Lee *et al.*, *IEEE Electron Device Lett.* **21**, 261–263 (2000).
28. G. Simin *et al.*, *IEEE Electron Device Lett.* **22**, 53–55 (2001).
29. M. Bruce Green *et al.*, *IEEE Electron Device Lett.* **21**, 268–270 (2000).
30. J.-S. Lee *et al.*, *Electron. Lett.* **37**, 130–132 (2001).
31. S. Kayali, *Microelectron. Reliab.* **39**, 1723–1736 (1999).