An AlGaN/GaN high-electron-mobility transistor with an AlN sub-buffer layer

2002 J. Phys.: Condens. Matter 14 3499

(http://iopscience.iop.org/0953-8984/14/13/308)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **14** (2002) 3499–3509 PII: S0953-8984(02)29107-9

An AlGaN/GaN high-electron-mobility transistor with an AlN sub-buffer layer

J R Shealy, V Kaper, V Tilak, T Prunty, J A Smart, B Green and L F Eastman

School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA

Received 25 September 2001
Published 22 March 2002
Online at stacks.iop.org/JPhysCM/14/3499

Abstract

The AlGaN/GaN high-electron-mobility transistor requires a thermally conducting, semi-insulating substrate to achieve the best possible microwave performance. The semi-insulating SiC substrate is currently the best choice for this device technology; however, fringing fields which penetrate the GaN buffer layer at pinch-off introduce significant substrate conduction at modest drain bias if channel electrons are not well confined to the nitride structure. The addition of an insulating AlN sub-buffer on the semi-insulating SiC substrate suppresses this parasitic conduction, which results in dramatic improvements in the AlGaN/GaN transistor performance. A pronounced reduction in both the gate-lag and the gate-leakage current are observed for structures with the AlN sub-buffer layer. These structures operate up to 50 V drain bias under drive, corresponding to a peak voltage of 80 V, for a 0.30 μm gate length device. The devices have achieved high-efficiency operation at 10 GHz ($>70\%$ power-added efficiency in class AB mode at 15 V drain bias) and the highest output power density observed thus far (11.2 W mm^{-1}). Large-periphery devices (1.5 mm gate width) deliver 10 W (continuous wave) of maximum saturated output power at 10 GHz. The growth, processing, and performance of these devices are briefly reviewed.

1. Introduction

The undoped AlGaN high-electron-mobility transistor (HEMT) is capable of bandwidths exceeding 75 GHz for a 0.25 μm gate length device [1] and output power densities exceeding 11 W mm^{-1} at 10 GHz [2]. This power density is at least ten times that of GaAs-based FET devices. In order to achieve this level of performance the substrate must be a good thermal conductor and a relatively low-loss insulator. The semi-insulating SiC substrate appears to require more than 200 nm of AlN between the device channel and the substrate to avoid both static leakage and RF conduction in the substrate. In this study a single layer of AlN is placed between the nucleation layer and the GaN buffer layer. Epitaxial structures grown by

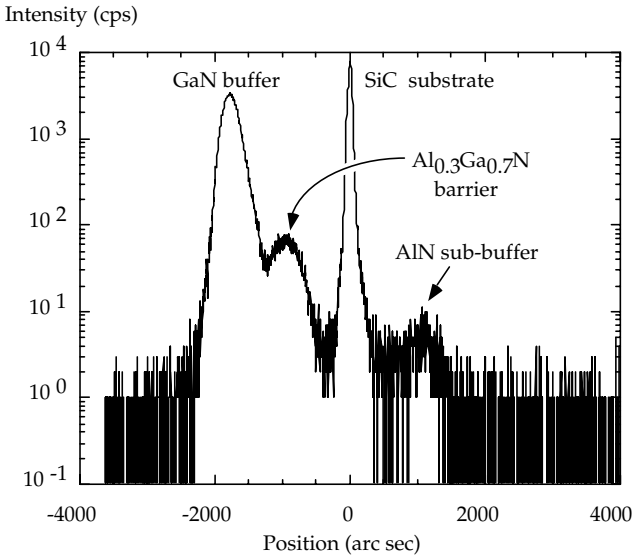

Figure 1. A high-resolution x-ray diffraction spectrum of the AlN/GaN buffered HEMT on semi-insulating SiC. The AlN thickness was 200 nm and the AlGaN barrier layer was ∼30 nm thick.

metal–organic vapour-phase epitaxy (MOVPE) are compared, with different AlN sub-buffer layer thickness. The effects of the AlN layer on device performance, including large-signal operation at 10 GHz, are described.

2. Experimental procedure

In this section the epitaxial growth of AlGaN structures on semi-insulating SiC substrates, followed by the characterization of these layers, is described. Finally, the device fabrication sequence including the passivation with silicon nitride is summarized.

2.1. Heteroepitaxial material growth and characterization

In our synthesis approach we use a single-flow, single-temperature MOVPE process to realize very high-quality HEMT structures on SiC, sapphire, and Si using AlGaN nucleation layers [3]. The structures compared in this study have AlGaN nucleation layers (∼200 Å thick), followed by the AlN sub-buffer when included, next, the GaN buffer, and finally the $Al_{0.35}Ga_{0.65}N$ top barrier layer (nominally 250 Å). The growth temperature was 1050 °C and the deposition rate was roughly 0.6 μm h^{-1}. With the exception of the nucleation layer, each layer in the structure has a clearly identifiable x-ray signature as shown in the diffraction spectrum given in figure 1.

In this case, the buried AlN layer was roughly 200 nm thick. From the positions of its binary components, the AlGaN barrier layer composition is calculated directly from Vegard's law as 30% for this sample. The AlN layer grown directly on the nucleation layer does not introduce charge from polarization effects in the buffer structure for reasons not entirely understood. The insertion of the AlN sub-buffer layer has little or no effect on the sheet electron density induced through polarization effects at the top heterojunction [4]. However, the carrier mobility is strongly influenced by the thickness of the AlN under the buffer. It turns out that increasing the thickness of the AlN and/or GaN lowers the channel sheet resistance

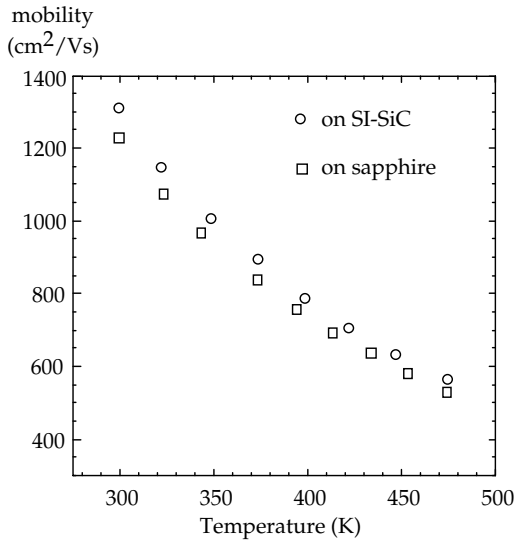

Figure 2. Hall mobility of AlGaN/GaN undoped HEMT structures measured at elevated temperatures.

due to improved electron mobility. As for the AlN layer, the thickness range investigated was from 200 to 500 nm. The thickest AlN layers produced the highest 300 K mobility of 1470 $cm^2\ V^{-1}\ s^{-1}$. In combination with 2 μm thick GaN buffers, we have structures with sheet resistances as low as 330 $\Omega\ \square^{-1}$. This compares to the 470 $\Omega\ \square^{-1}$ observed with a 1 μm GaN buffer which yielded devices operating at over 10 W mm^{-1} at 10 GHz (each structure had a same sheet charge of about $1.3 \times 10^{13}\ cm^{-2}$).

As these devices are expected to operate at high temperatures, the influence of temperature on the electron mobility is of considerable interest. A decrease in mobility at elevated temperature extends the knee of the output curves, reducing the drain efficiency of the transistor. Hall mobility data are given for both sapphire and SiC structures with nominal 1 μm GaN buffers and 200 nm AlN sub-buffers in figure 2.

As seen, increasing the temperature results in a substantial drop in electron mobility. The electron mobility was observed to decrease according to $1/\mu \sim T^{1.8}$ for all samples (substrates). Raising the temperature from room temperature to 200 °C reduces the electron mobility roughly by a factor of 2.3–1. Accordingly, if temperatures of 200 °C are reached during device operation, the corresponding knee voltage would increase from 4 to approximately 10 V in the absence of any dc-to-RF dispersion. The power-added efficiency (PAE) suffers from the reduction in drain efficiency due to the expansion of the knee from heating. As our test data illustrate, this becomes a dominant effect in limiting power and efficiency at large drain bias.

2.2. Device fabrication

A standard hybrid optical/electron-beam lithographic process is used to fabricate HEMT devices with sub-micron mushroom gate structures. The sheet charge and barrier thickness are first obtained from C–V measurements. A typical HEMT structure on a SiC substrate has a sheet charge of $1.5 \times 10^{13}\ cm^{-2}$ and a barrier thickness of 20 nm. A thinner barrier, although suitable for high frequency, exhibits more dc-to-RF dispersion as the surface is closer to the

two-dimensional electron gas [4]. Hall measurements are made to obtain the mobility and sheet resistance of the sample. The process includes a 200 nm chlorine-based ECR etch to remove the AlGaN barrier layer for mesa isolation. Next, the ohmic pattern is defined using photolithography. The ohmic metal stack of Ti(20 nm)/Al(100 nm)/Ti(45 nm)/Au(55 nm) is evaporated; this is followed by lift-off. Ohmic contacts are made by rapid thermal annealing of the sample at 800 °C for 60 s. Typical contact resistance numbers observed were 0.5 Ω mm. Following the ohmic patterns, the mushroom gate is defined using a three-layer electron-beam resist stack of PMMA/copolymer/PMMA. Evaporation and lift-off of Ni(25 nm)/Au(375 nm) is used to form the gates. Then the contact pads are fabricated using photolithography and evaporation of Ti(20 nm)/Au(350 nm). Devices with a gate footprint of 0.30 μm (nominal) were fabricated into three different configurations: the single centre-fed T-gate, the dual-parallel-channel U-gate, and the manifolded (M), source air-bridged, parallel-finger layout. A gate–drain spacing of 2 μm was chosen for all devices. At this point, 180 nm of PECVD silicon nitride is deposited at 300 °C. The mushroom gate structure is designed with dimensions that prevent the passivation (the step to follow) from reaching the metal edge of the gate footprint on the drain side. This is done to preserve the breakdown voltage and gate leakage of the unpassivated transistor. The quality of the PECVD film is monitored by checking the etch rate in BOE (6:1) [5]. This film is patterned using an optical resist and the dielectric is etched from the contact pads using an RIE etch. The passivation step is required to reduce dc-to-RF dispersion arising from the trapping of electrons on surface states on the drain end of the gate [6]. Electroplating is carried out to fabricate the air bridges to connect isolated drain pads in the multi-finger manifold devices.

3. Results and discussion

The fabricated devices are subjected to several tests which separately identify problems associated with substrate leakage and surface state trapping, both dominant factors in achieving good RF performance. Static testing at pinch-off clearly identifies the contribution of substrate leakage and its suppression with the AlN sub-buffer. The effectiveness of the surface passivation is accessed by means of pulsed I–V (gate-lag) measurements. These same pulsed measurements also identify gate lag associated with the capacitively coupled substrate shunt. Finally, devices with good passivation and AlN sub-buffer isolation are subjected to large-signal testing at 10 GHz.

3.1. Static testing

The influence of substrate conduction is easily observed by examining the static gate-leakage current at pinch-off. A parasitic current flows between the device's probe pads, passing through the undoped GaN buffer via space-charge-limited current flow; this is followed by electrons passing over the relatively small barrier to the SiC substrate where they are trapped and tunnel out of the substrate deep centres (vanadium) at sufficient drain bias. The trapping time for this mechanism, from our pulsed I–V data, is estimated to be in excess of a few milliseconds at room temperature. The insertion of the insulating AlN sub-buffer layer is able to sharply reduce this parasitic current as shown in figure 3. Also observed was a substrate-dependent leakage current for structures without the AlN buried layer. For the structure with the lowest gate current, using the 200 nm AlN sub-buffer, the drain current at pinch-off was reduced to levels as low as 100 nA mm^{-1} for drain bias less than 60 V. The corresponding range of drain current from pinch-off to full-channel current for this device covers seven orders of magnitude. Static testing beyond 60 V to measure the static breakdown voltage on AlN sub-buffered structures was not

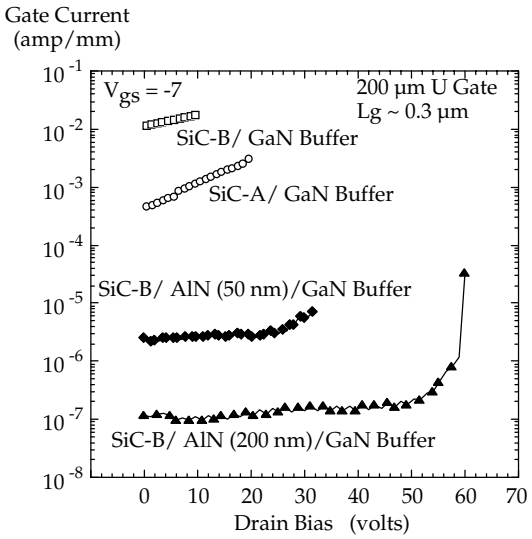

Figure 3. Gate current at pinch-off for several AlGaN HEMT structures on the buffer structure indicated. The substrates are designated SiC-A and SiC-B, SiC-B having the larger defect density.

generally done, as devices and probes were often damaged. However, these structures were able to operate at class B bias voltages up to 50 V under full RF drive without breakdown occurring.

3.2. Pulsed I–V testing

The pulsed I–V curves are taken under several conditions to examine both gate and drain lag. The gate-lag data are generally the more useful, as we have separately observed surface-related and substrate-related carrier trapping—they occur on vastly different timescales and the substrate-related gate lag has a different dependence on the drain bias. The gate-lag data were taken in the following manner: a fixed drain bias was maintained and the gate was kept below pinch-off; the gate was then pulsed up from below pinch-off in steps to eventually full-channel conditions (gate–source bias of +1 V) for a duration of 100 ns at a low-duty cycle (10^{-6}). The drain current was sampled midway into the 100 ns pulse. Both the input and output of the device are coplanar-waveguide probed and terminated into 50 Ω. In this fashion, the thermal effects from self-heating can be ignored. The drain current response is measured with a current probe with 5 ns transient response. The drain current pulse has an exponential rise on the front end of the pulse with a time constant which scales linearly with the device periphery. At sufficient drain bias, the exponential response can vanish on AlN sub-buffered devices. Small device peripheries are used (100 μm or less) which results in a time constant of less than 10 ns on a good device with little or no gate lag.

Two devices with nearly the same pinch-off voltage, one with and one without the AlN sub-buffer, are compared using the pulsed I–V measurements described above. In figure 4, the resulting transfer characteristics are plotted for each type of device at a fixed drain bias of 6 V (beyond the 4 V static knee). Each device had nearly the same full-channel current measured under static conditions (1.2 A mm^{-1}), but as shown, the pulsed current response of the device without the AlN sub-buffer layer is reduced by a factor of two. Examination of these structures at longer gate pulses indicates that the drain current rise to the static value takes over 2 ms. This represents strong circumstantial evidence that electron trapping in the substrate produces

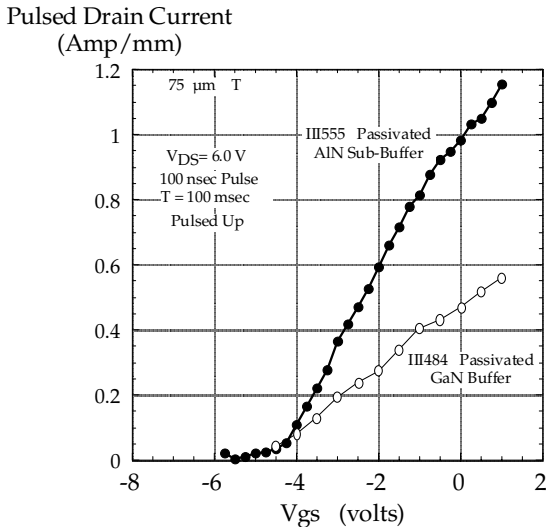

Figure 4. Pulse gate-lag measurement comparison of nearly identical structures with and without the AlN sub-buffer layer. The AlGaN barrier was approximately 20 nm in both cases.

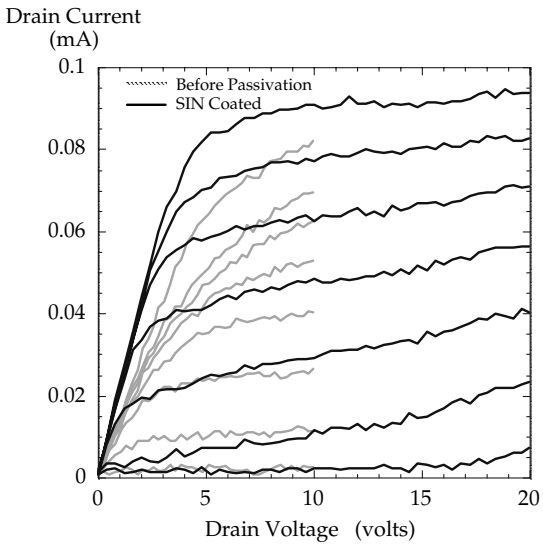

Figure 5. Pulsed output curves of a 75 μm SiC-based AlGaN HEMT with the AlN sub-buffer before and after the application of the SiN passivation over the metal gate.

a parallel conduction path which effectively shunts at least one half of the device current on this timescale. Accordingly, the RF current swing is reduced by this same factor. Devices without the AlN layer produce saturated output powers no greater than 4 W mm^{-1}.

The pulsed output curves are reconstructed by sweeping the drain bias up to 20 V for a fixed pulse amplitude on the gate (up from pinch-off). This measurement has proven to be useful in monitoring several device performance issues including: the output conductance related to space-charge-limited current flow in the GaN buffer; the substrate shunt and the impact of the AlN sub-buffer; and the effectiveness of the passivation step in the fabrication process.

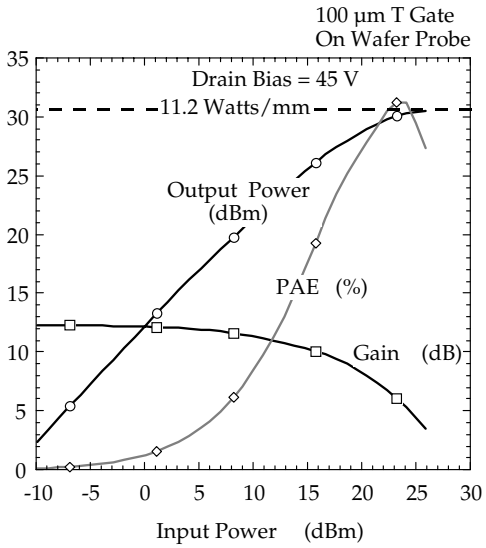

Figure 6. CW output power, PAE, and gain versus microwave input drive at 10 GHz for a 100 μm periphery device with the AlN sub-buffer layer.

There is also a rough correlation with the pulse 'knee' voltage and the low-voltage excursion of the microwave load line at X band. The pulsed output curves for a 75 μm periphery transistor on an AlN sub-buffered structure are shown in figure 5. In this case the same device data are shown before and after the passivation step. Even with drain biases up to 10 V, the full-channel current is not reached for the unpassivated device. If loaded for optimal RF power, this unpassivated transistor would exhibit a RF 'knee' (from load-line mapping) exceeding 15 V. Now considering the passivated device, the full-channel current (1.2 A mm^{-1}) is reached at 5–6 V. The static knee voltage was roughly 4 V, so the passivation step, which leaves an uncoated gate extension towards the drain, is not completely effective. Because device self-heating is avoided, the output conductance from GaN buffer conduction (space-charge limited) is observed. Note that increasing the reverse bias on the gate beyond pinch-off suppresses this conduction as the drain bias is increased.

3.3. Large-signal testing

These devices were tested at 10 GHz, with input drive up to the saturation level, while optimally loaded using a Maury^{TM} automatic tuning set up. The gate bias is set such that the dc drain current is approximately 10% of the full-channel current. This bias was experimentally found to be optimal for several reasons:

(1) it is near the maximum transition frequency ($f_{t\,\max} = 70$ GHz) point, extracted from small-signal S-parameter measurements;
(2) dc power and heat dissipation are minimized;
(3) device linearity is acceptable—transducer gain expansion is avoided at low input power levels, which happens when the device was biased near or beyond pinch-off.

The impedance at device's input plane is set close to the complex conjugate of the device's input impedance to maximize power transfer between RF signal source and the device. The impedance selected for the device loading was obtained by maximizing the PAE at low drain

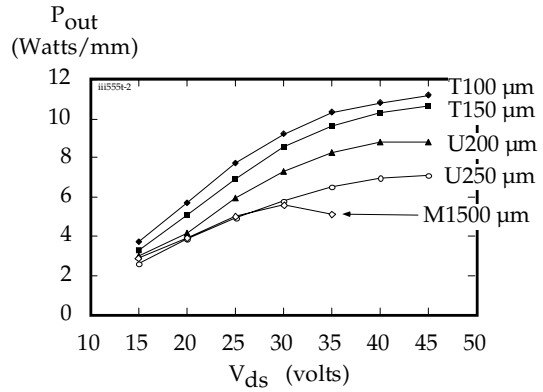

Figure 7. Saturated output power density for class B measurements at 10 GHz. The designation T, U, M preceding the device periphery refers to a T-gate, a U-gate, and a manifolded air-bridged device, respectively.

bias voltage (typically 10–15 V). This results in a load line which does not penetrate deep into the device knee until the drain bias is raised beyond typically 35 V. The drain bias is then raised in 5 V steps and successive power sweeps are taken without adjusting the gate bias or the load impedance. This procedure has yielded higher output powers than performing load pull under compression at large drain bias, as devices are often damaged (breakdown) during such load-pull measurements. Figure 6 presents the CW output normalized power, PAE, and gain, as a function of drive power, in class AB operation. Output power reaches a new state-of-the-art value of 11.2 W mm^{-1} for a device with 100 μm total gate width at a drain–source bias of 45 V with a load resistance of $\sim$500 Ω, while dissipating 24 W mm^{-1} of heat.

Also examined were the CW normalized output powers for five different periphery devices using on-wafer probing. For multi-finger devices, the 50 μm pitch between fingers causes an increased temperature rise in devices with a U layout (dual finger), and especially in the 12-channel (1.5 mm total periphery) manifold devices (designated M). The results of these measurements are summarized in figure 7. As shown, the maximum saturated output power obtained exceeded 10 W mm^{-1} only for the single-finger T-device layouts. The increase in device heating as the periphery is increased causes the output power to saturate at a lower drain bias voltage in these on wafer-probed measurements. The heating lowers the drain efficiency due to the rise in the knee voltage in proportion to $(T/300)^{1.8}$, with T being the channel temperature in kelvins due to the drop in electron mobility with temperature (see figure 2).

Figure 8 shows the power-added efficiencies of these HEMTs with different gate widths as a function of drain bias voltage. It may be seen that all devices have dropping efficiency with increased drain voltage from increased self-heating. The device with 200 μm periphery gives the highest PAE, due to the best combination of input and output impedance matching for that device. Both smaller- and larger-periphery devices are less easily impedance matched. As can be seen from figure 8, the maximum measured PAE is 73% for 200 μm U-gate device. This number corresponds to 81% drain efficiency, which exceeds the theoretical limit of 78.5% for ideal class B operation. It should be noted that this record efficiency is obtained at an input drive corresponding to 6 dB gain compression. However, in such overdrive conditions, Cripps [7] has shown that the conduction angle reduces, and, as a result, the maximum attainable drain efficiency increases. In a case where the output current conduction occurs for less than half a cycle, the drain efficiency is not limited by the class B limit of 78.5%.

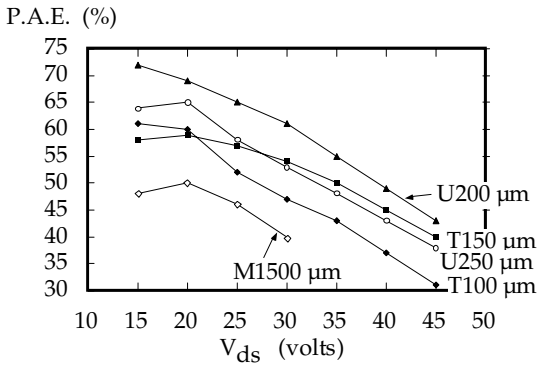

Figure 8. PAE for CW class B measurements at 10 GHz using on wafer coplanar-waveguide probing.

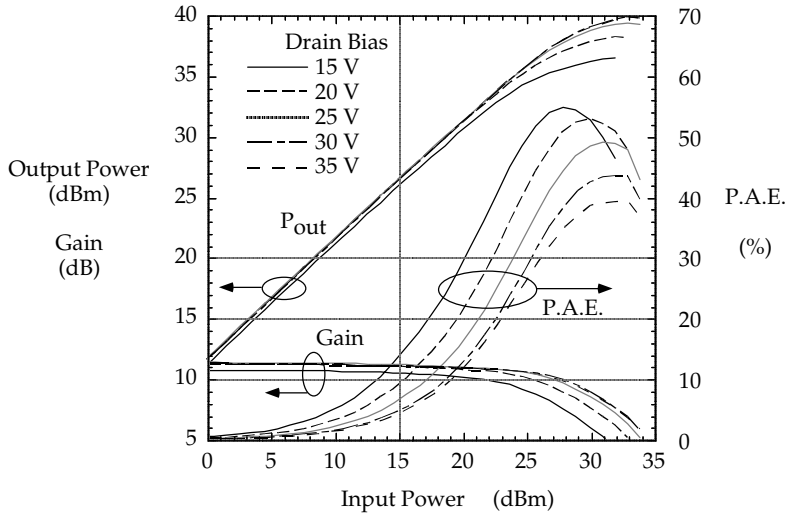

Figure 9. CW performance at 10 GHz of a 1.5 mm manifold device mounted on a Cu heat spreader and coplanar-waveguide probed.

In order to reduce the device self-heating during operation, improved thermal management over on-wafer probing was necessary. The processed wafers (300 μm in thickness) were sawn into transistor dice (0.65 by 0.75 mm), each containing a single 1.5 mm manifold device. These dice were soldered to a Cu heat spreader using In. The Cu sub-mount was then placed on a 5 °C Cu chuck on a coplanar-waveguide probe station. The device dice were coplanar-waveguide probed as before. In this fashion, during maximum power dissipation, the Cu heat spreader was maintained below 20 °C. Devices were biased in class AB mode as before and a series of 10 GHz power sweeps were taken as the drain bias was raised in 5 V steps. The results of these measurements are shown in figure 9 where the CW output power, the gain, and the PAE are plotted versus input RF drive. Several observations can be made:

(1) the drain bias where the saturated output power peaked was increased from 30 (on-wafer) to roughly 40 V;
(2) the PAE remained over 40% up to 40 V bias on the drain;

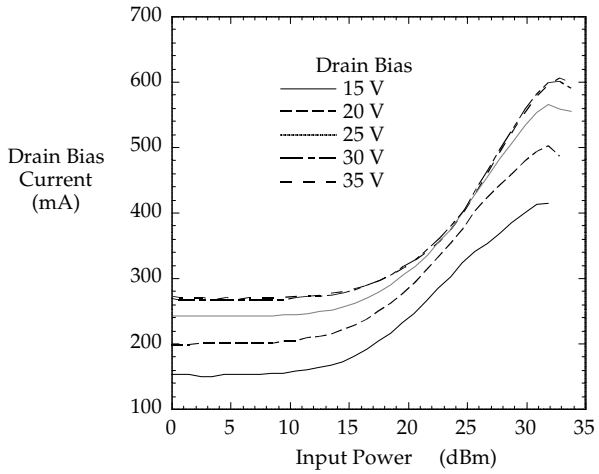

Figure 10. Drain bias current versus input drive at 10 GHz for the 1.5 mm periphery device.

(3) the saturated output power achieved was 10 W CW or 6.7 W mm^{-1}, a 22% improvement over the on-wafer measurement result; and
(4) the linear power (1 dB compression) was over 5 W.

The gain is also seen to increase slightly with drain bias due to the loading conditions chosen. Although this is not shown, if the load impedance is increased to lower the drain bias current under full drive, then 7 W of linear power was obtained. Finally, this same device under pulsed RF operation (1% duty cycle, 500 ns pulses) operated at 14.5 W (9.7 W mm^{-1}) at 2 dB compression. This suggest that the device self-heating (or lack of thermal management) is a principal cause of gain compression in these HEMTs under increasing RF drive.

Figure 10 shows the rise in measured dc drain current with drive power for these different drain bias voltages. This is a manifestation of a fact that the device is being self-biased by the input RF drive, which is expected for near-pinch-off gate biases (deep class AB and class B). The bias current rise, at low drive power, for increased drain bias voltage, is due to significant output conductance at near-pinch-off gate biases, caused by short-channel effects, as was observed in the pulsed I–V data (figure 5). As the input power increases, so does the amplitude of RF drain current. At this point, the RF current waveform is only limited on low side and the dc value increases with the amplitude. When the current swing at high RF input drive reaches the knee voltage (for optimal load resistance) or the linear portion of the I–V curve (for higher-than-optimal load resistance) or the full-channel current (for lower-than-optimal load resistance), the RF current waveform is being limited on both sides, and the dc current under drive saturates at 600 mA as shown. Note that this saturated value corresponds reasonably well to the average current of a half-wave-rectified RF current waveform whose amplitude is the full-channel current of 1.8 A for a 1.5 mm periphery.

4. Summary and conclusions

The AlGaN/GaN HEMT structure on SiC requires an isolation layer to avoid a pronounced substrate parasitic current. Undoped structures with the AlN sub-buffer suppress virtually all carrier trapping in the substrate if its thickness exceeds 200 nm. On these structures, transistors are made which deliver over 10 W CW at 10 GHz (1.5 mm periphery). The power density

on small devices exceeds 11 W mm^{-1} (CW) and approaches 10 W mm^{-1} (pulsed) on large-periphery devices. With appropriate loading the PAE exceeds 70% on devices with the best impedance match (200 μm periphery). Although this level of performance is impressive, there are still a few remaining issues which limit the output power and linearity of these devices. First, a residual surface related dc-to-RF dispersion remains from the uncoated gate extension towards the drain. This extends the RF 'knee' voltage to roughly 6 V, reducing the drain efficiency (especially at low drain bias voltage). Second, the undoped GaN buffer layer introduces a significant output conductance from the short-channel effect (space-charge-limited conduction in the buffer). This results in an increase in drain bias current with increasing drain bias voltage. This introduces additional device heating during operation at large drain bias. Currently, alternative gate structures with a drain extension are under investigation to address the remaining dc-to-RF dispersion. To address the short-channel effects, an undoped AlGaN layer is placed under the channel to introduce a barrier at the back. With these materials and process improvements, it appears that the undoped AlGaN/GaN HEMT will offer significantly higher-performance microwave power devices than has currently been demonstrated.

Acknowledgments

This study was supported in large part by the ONR-MURI program and by the BMDO. In addition, we acknowledge the support of several US industries including Raytheon, Northrup Grumman, Welch Allyn, and RF Nitro Communications.

References

[1] Eastman L F *et al* 2001 Undoped AlGaN/GaN HEMTs for microwave power amplification *IEEE Trans. Electron Devices* **48** 479–85

[2] Shealy J R, Smart J A, Kaper V and Prunty T R 2001 The effects of an AlN sub-buffer structure on the performance of OMVPE grown AlGaN/GaN HEMTs on SiC substrates *2nd Workshop on GaN Electronics (May 2001)*

[3] Smart J A, Schremer A T, Weimann N G, Ambacher O, Eastman L F and Shealy J R 1999 AlGaN/GaN heterostructures on insulating AlGaN nucleation layers *Appl. Phys. Lett.* **75** 388–90

[4] Tilak V, Green B, Kaper V, Kim H, Prunty T, Smart J, Shealy J and Eastman L 2001 *Electron Device Lett.* **22** 504

[5] Green B M, Chu K K, Chumbes E M, Smart J A, Shealy J R and Eastman L F 2000 *IEEE Electron Device Lett.* **21** 268–70

[6] Tilak V, Green B, Dimitrov R, Smart J, Kim H, Schaff W and Eastman L 2000 Effect of passivation on HEMT device performance *Proc. Int. Symp. on Compound Semiconductors (Monterey, CA, Oct. 2000)* pp 357–63

[7] Cripps S C 1999 *RF Power Amplifiers for Wireless Communications* (Boston, MA: Artech House Publishers)