Millimeter-wave GaN High Electron Mobility Transistors and Their Integration with Silicon Electronics

by

Jinwook W. Chung

B.S., Electrical Engineering
Korea Advanced Institute of Science and Technology, 2006
S.M., Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2008

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 2011

Author _____
Department of Electrical Engineering and Computer Science
January 28, 2011

Certified by _____
Tomás Palacios
Associate Professor of Electrical Engineering
Thesis Supervisor

Accepted by _____
Terry P. Orlando
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Students

Millimeter-wave GaN High Electron Mobility Transistors and Their Integration with Silicon Electronics

by

Jinwook W. Chung

Submitted to the Department of Electrical Engineering and Computer Science
on January 28, 2011 in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

ABSTRACT

In spite of the great progress in performance achieved during the last few years, GaN high electron mobility transistors (HEMTs) still have several important issues to be solved for millimeter-wave (30 ~ 300 GHz) applications. One of the key challenges is to improve its high frequency characteristics. In this thesis, we particularly focus on f_T and f_{max}, two of the most important figures of merit in frequency performance of GaN HEMTs and investigate them both analytically and experimentally. Based on an improved physical understanding and new process technologies, we aim to demonstrate the state-of-the-art high frequency performance of GaN HEMTs. To maximize f_{max}, parasitic components in the device (R_i, R_s, R_g, C_{gd}, and g_o) are carefully minimized and the optimized 60-nm AlGaN/GaN HEMT shows a very high f_{max} of 300 GHz. The lower-than-expected f_T observed in many AlGaN/GaN HEMTs is attributed to a significant drop of the intrinsic transconductance at high frequency (RF g_m) with respect to the intrinsic DC g_m (called RF g_m-collapse). By suppressing RF g_m-collapse and harmoniously scaling the device, a record f_T of 225 GHz is achieved in the 55-nm AlGaN/GaN HEMT. Another important challenge for the wide adoption of GaN devices is to develop suitable technology to integrate these GaN transistors with Si(100) electronics. In this thesis, we demonstrate a new technology to integrate, for the first time, GaN HEMTs and Si(100) MOSFETs on the same chip. This integration enables the development of hybrid circuits that take advantage of the high-frequency and power capability of GaN and the unsurpassed circuit scalability and complexity of Si electronics.

Thesis supervisor: Tomás Palacios
Title: Associate Professor of Electrical Engineering

Acknowledgements

My successful Ph.D. journey has only been possible because I was surrounded by the right people. I can honestly say that without the inspiration and support of the people around me, I would have never finished this journey. It is my great pleasure to thank the many individuals who have contributed to this work.

First and foremost, I would like to express my sincere gratitude to my research advisor Prof. Tomás Palacios. I still remember the moment when I first decided to join his group in September 2006. Working on nitride semiconductors was never in any plan that I have imagined for myself, however it was without a doubt one of the best decisions that I ever made and I truly appreciate his full support, candid advice, and endless encouragement to make my Ph.D. experience fruitful and exciting.

I gratefully thank my thesis committee members, Prof. Dimitri Antoniadis, Prof. Jesús del Alamo, and Prof. Judy Hoyt. It was my great honor to have them in my thesis committee and their invaluable suggestions and feedbacks largely improved this work.

I also would like to acknowledge the ONR MINE MURI and DURIP programs monitored by Dr. Paul Maki and Dr. Harry Dietrich, and the DARPA NEXT program monitored by Dr. John Albrecht for their continuous support to make GaN electronics a reality. In addition, many thanks go to the KFAS fellowship for the financial support during my entire Ph.D. years.

The work presented in this thesis extensively relies on device fabrication and processing. I would like to thank all the staffs in MTL, NSL, and SEBL at MIT. They are absolutely helpful and I strongly believe that one of the real strengths of MIT EECS is their excellent work and strong support. I have also been lucky to collaborate with Raytheon, Nitronex, TriQuint, Cree, and IQE, and acknowledge them for providing us state-of-the-art GaN samples.

My Ph.D. years have been enriched by many friends and colleagues. In particular, Taegsang Cho, Myungjin Choi, Hwanchul Yoo, Wonyoung Kim, and Jiye Lee made my days happy and enjoyable. Thank you for sharing wonderful years at MIT. I will not forget Taewoo Kim who was my life mentor, and Jaekyu Lee and Jungwoo Joh who inspired me in many ways. I also thank to friendship of TP group members, Bldg 39 friends, and Fab buddies.

Last but definitely not least, my deepest gratitude goes to my family who allowed me to dream and lifelong brothers who enabled me to move forward. They are simply my reason for living.

Contents

Chapter 1. Introduction

1.1. History of GaN research

During the last 20 years, there has been an explosive evolution of nitride semiconductors (GaN, AlN, InN, and their alloys) in both optoelectronics and electronics. Remarkable breakthroughs achieved by numerous researchers and the unique properties of nitride semiconductors such as a direct bandgap tunable from 6.2 eV (AlN) down to 0.7 eV (InN), piezoelectricity, polarization [1], large breakdown voltage [2], biocompatibility [3], and high chemical and thermal stability [4] enabled this material system to be the prime choice for blue light emitting diodes (LEDs), blue laser diodes, and high-frequency high-power electronics.

Figure 1-1 describes some of the major breakthroughs in nitride growth and technology in the early years [5]. What initially brought the nitrides to the realm of semiconductors for device applications could be the success in growing crystalline GaN on a sapphire substrate by hydride vapour-phase epitaxy (HVPE) in 1969 [6]. This event greatly inspired many researchers and foster the GaN research in its early stage of development. GaN was perceived as the best material for the fabrication of blue LEDs. However, in the late 1970s, the GaN research community faced challenges in growing high quality epitaxial films and, specially, producing p-type GaN. As a result, the activity on GaN research slowly declined in this period. This stagnation, however, did not last long. Two important milestones reignited GaN research: the growth of high-quality single crystal GaN with a specular crack-free surface by Amano *et al.* in 1986 [7] and the discovery of a way to produce p-type GaN by Amano *et al.* in 1989 [8] and Nakamura *et al.* in 1992 [9]. As shown in Figure 1-1, these two breakthroughs were an important inflection point for nitride semiconductor research. The great improvement and superior performance of the devices were instantly followed first in optoelectronics. Amano *et al.* demonstrated the first GaN LED with a p-n junction in 1989 [8]. Then, Nakamura *et al.* built on this work and optimized the design and process which led to a hike in device efficiency and enabled Nichia to launch the first commercial nitride-based LEDs in 1993 [5]. He also

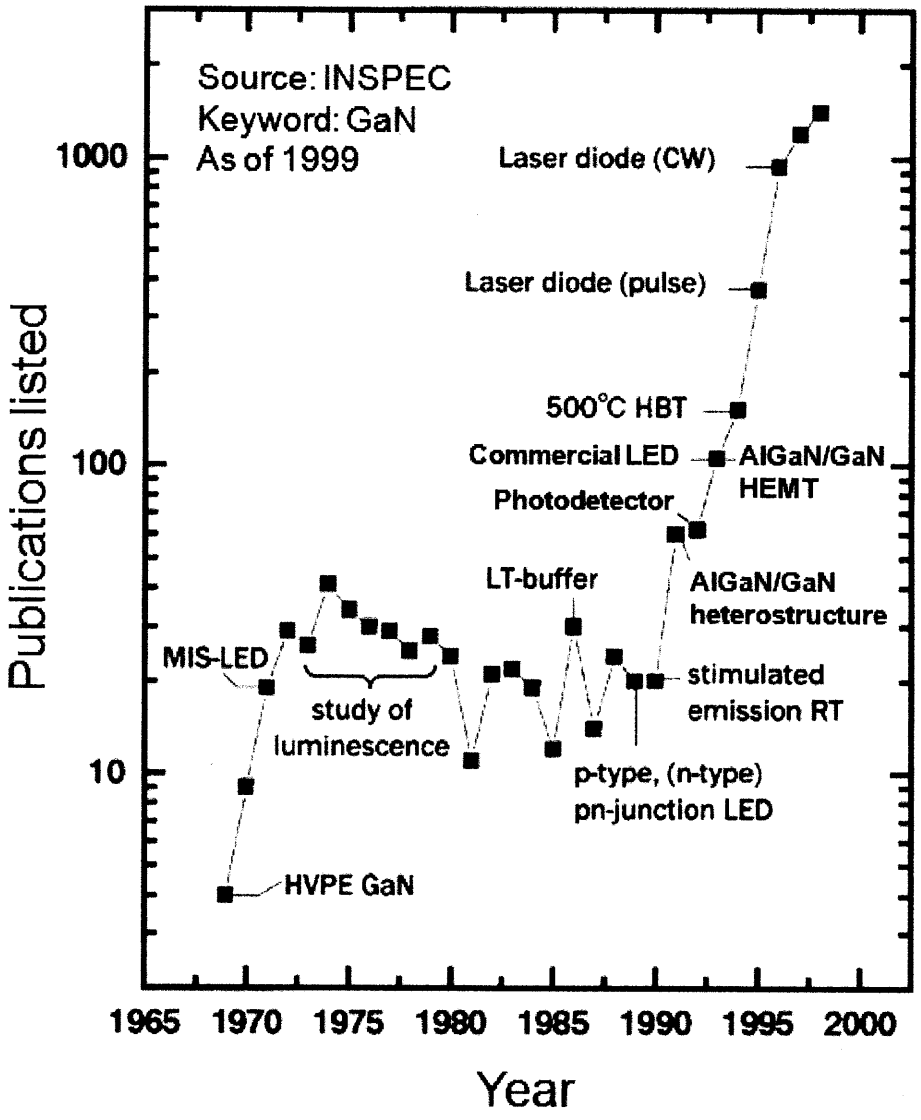


Figure 1-1. Number of publications (based on INSPEC search) and activities of GaN research in the early years (modified from [5]). Marked events indicate when they were first achieved. GaN research in both optoelectronics and electronics progressed rapidly after demonstrating the growth of high quality GaN and conductivity control of both p- and n-type GaN in the early 1990s.

demonstrated the first nitride-based laser diode in 1996 [10]. Continuous achievements in recent years significantly boosted the power and efficiency of GaN optoelectronic devices and finally opened multibillion dollar industry.

In parallel to the revolution in optoelectronics, the high material quality enabled by optimized growth reactors was able to trigger intense research on GaN for electronic applications. One of the most widely studied electronic devices was the AlGaN/GaN high electron mobility transistor

Characteristic	Unit	Material				
		Si	GaAs (AlGaAs/InGaAs)	InP (InAlAs/InGaAs)	SiC	GaN (AlGaN/GaN)
Bandgap (E_g)	eV	1.11	1.42	1.35	3.26	3.39
Critical breakdown field	MV/cm	0.3	0.4	0.5	3.0	3.3
Saturated (peak) electron velocity	$\times 10^7$ cm/s	1.0 (1.0)	1.3 (2.1)	1.0 (2.3)	2.0 (2.0)	1.3 (3.0)
Electron mobility at 300 K	cm^2/Vs	1350	8500	5400	700	2000

Table 1-1. Some of the outstrip material properties of GaN compared to other semiconductor families.

(HEMT). This device attracted considerable attention soon after its excellent transport property was first reported in 1991 [11]. The AlGaN/GaN HEMT has unique and superior properties such as large sheet charge density (~ 1×10^{13} /cm^2), high peak electron velocity (~ 3.0×10^7 cm/s), high breakdown field strength (~ 3.3 MV/cm) and good thermal conductivity (> 1.5 W/cm·K). Since GaN is one of the few materials that can achieve high breakdown voltage and high electron velocity at the same time as shown in Table 1-1, it is particularly promising for RF power amplifiers. In only 20 years, AlGaN/GaN HEMTs have evolved tremendously from the initial devices with less than 50 mA/mm of output current and virtually no high frequency performance [12], to world-wide commercialization as power amplifiers in the microwave spectrum (1 ~ 30 GHz). The excellent intrinsic properties of nitride semiconductors make them one of the best options for high-frequency high-power applications and AlGaN/GaN HEMTs have already shown an output power density of 32.2 W/mm at 4 GHz [13], 13.7 W/mm at 30 GHz [14], and 2.1 W/mm at 80.5 GHz [15] far surpassing other technologies (e.g. Si, GaAs, InP). Recently, Toshiba announced a commercial C-band (4 ~ 8 GHz) GaN HEMT that nearly doubles the gain compared to GaAs power amplifier [16].

1.2. Millimeter-wave GaN HEMTs

The rapidly growing wireless communication market and demanding military applications continuously require high-efficiency high-power GaN devices. Figure 1-2 shows required output power and operating frequency of RF power amplifiers for several military and civil applications. Different material systems (e.g. Si, SiGe, GaAs, InP) allow us to fulfill different requirements.

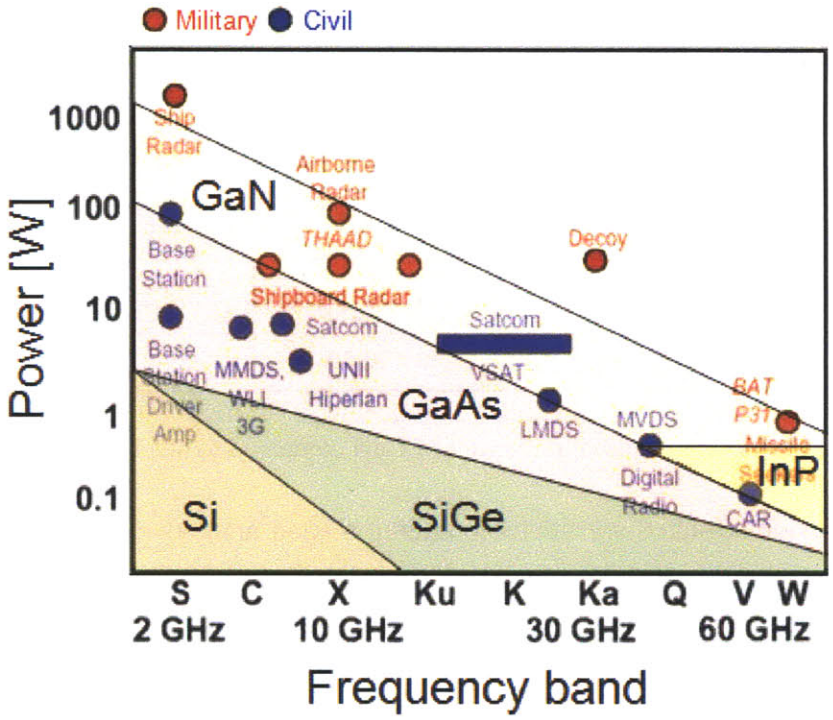


Figure 1-2. Required output power and operating frequency of RF power amplifiers for different military and civil applications. Different semiconductor materials are needed depending on the specific power and frequency requirements.

Due to the excellent properties of GaN, GaN can cover most of the requirements and offer superior performance to any other semiconductor in power amplifiers. For example, next generation cell phones and broadband satellite communications require amplifiers driving high power to reduce their size and operating at high frequencies to increase data transmission rate. From this point of view, wide bandgap material such as GaN is desirable because the ultimate breakdown field is determined by band-to-band impact ionization and its high electron velocity enables high frequency operation of the device. Although SiC shows comparable bandgap and breakdown field as GaN, GaN has an advantage over SiC in that GaN can be used to fabricate high electron mobility transistors (HEMTs) while SiC can only be used to form metal semiconductor field effect transistors (MESFETs). The HEMT structure allows higher electron mobility due to its reduced ionized impurity scattering as will be discussed in Chapter 2.

Figure 1-3 plots two typical figures of merit in high-frequency high-power applications, breakdown voltage and operating frequency (or current-gain cutoff frequency f_T), for various semiconductor material systems [17]. There is a trade-off between breakdown voltage and

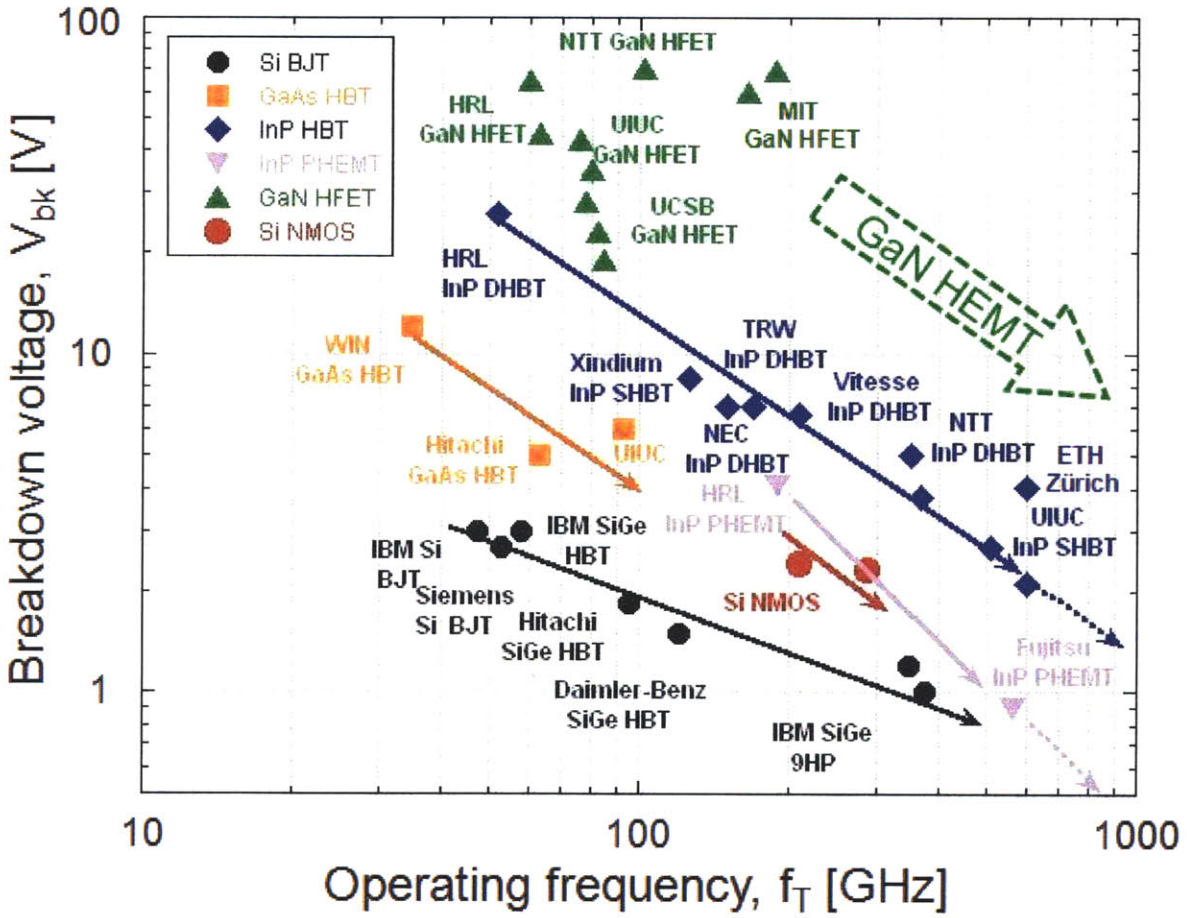


Figure 1-3. Breakdown voltage (V_{bk}) and operating frequency (f_T) of different material systems (modified from [17]). GaN HEMTs have a tremendous potential to outperform other competing materials for high-frequency high-power applications.

operating frequency, and in each material system this trade-off is characterized by its Johnson figure of merit [18]. The Johnson figure of merit is the product of the saturation velocity and critical electric field by the impact ionization initiated breakdown and it is a measure of suitability for high frequency power device applications. Among the various material systems typically used in power amplifiers, GaN HEMTs have already demonstrated the highest breakdown characteristics at frequencies up to 100 GHz and we can also expect higher breakdown behaviors at millimeter-wave (mm-wave) frequencies (30 ~ 300 GHz) by following the trade-off line (Johnson figure of merit) of GaN as shown in Figure 1-3. To understand what limits the high frequency performance of GaN transistors and to demonstrate state-of-the-art f_T-vs.-V_{bk} results are some of the main motivations of this project.

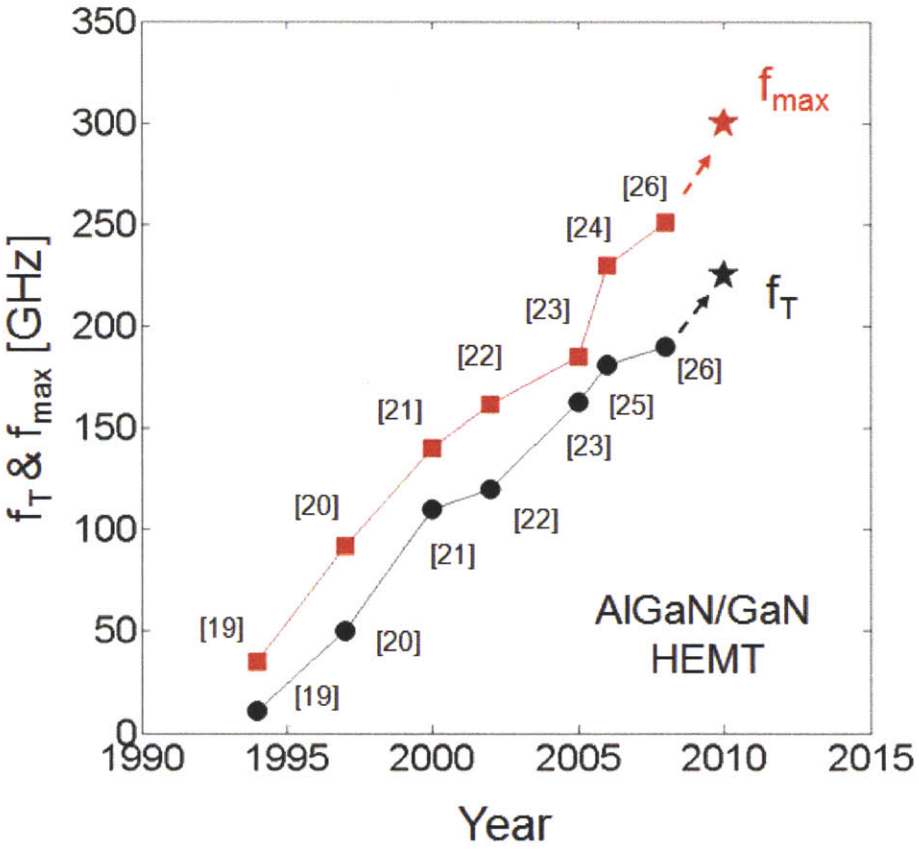


Figure 1-4. Evolution of the f_T and f_{max} in AlGaN/GaN HEMTs. Only the highest values reported in each year were plotted. In this work, we demonstrate state-of-the-art f_T of 225 GHz and f_{max} of 300 GHz in AlGaN/GaN HEMTs.

The excellent material properties of nitride semiconductors in conjunction with novel fabrication technologies have allowed very fast progress in the high frequency performance of GaN HEMTs. Figure 1-4 shows the evolution of frequency performance in AlGaN/GaN HEMTs over the last 15 years. In 1994, Khan *et al.* reported the first small-signal high frequency performance of an AlGaN/GaN transistor with 0.25 μm gate length [19]. This transistor showed a current-gain cutoff frequency (f_T) of 11 GHz and a power-gain cutoff frequency (f_{max}) of 35 GHz. In 1997, Wu *et al.* demonstrated f_T of 50 GHz and an f_{max} of 92 GHz in a device with a gate length of 0.2 μm [20]. In 2000, Micovic *et al.* improved frequency performance of AlGaN/GaN HEMTs to an extrinsic f_T of 110 GHz and f_{max} over 140 GHz with 0.05 μm gate length [21]. In 2002, Kumar *et al.* obtained f_T of 121 GHz and f_{max} of 162 GHz with the gate length of 0.12 μm [22]. In 2005, Palacios *et al.* reported f_T of 163 GHz and f_{max} of 185 GHz with a gate length of 90 nm using Ge-spacer technology [23]. In 2006, Palacios *et al.* further increased f_{max} to 230 GHz

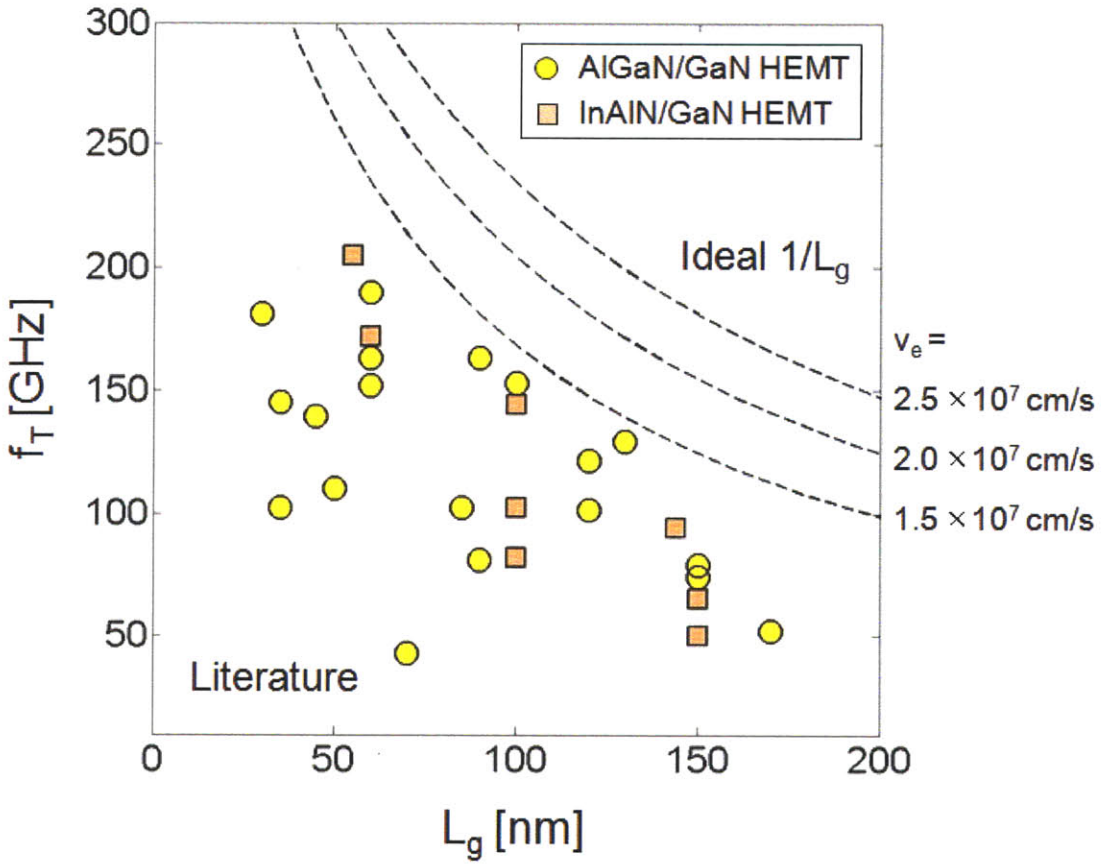


Figure 1-5. High frequency performance of deep-submicron GaN transistors reported in the literature. There is a very large variation of reported f_T in the literature even with the same L_g, as well as a considerable deviation of f_T from its ideal 1 / L_g scaling behavior.

with a gate length of 100 nm using InGaN back-barrier [24]. In the same year, Higashiwaki *et al.* achieved f_T of 181 GHz by scaling the gate length to 30 nm [25]. Finally in 2008, Higashiwaki *et al.* fabricated thin barrier AlGaN/GaN HEMTs with 60 nm gate length and demonstrated f_T of 190 and f_{max} of 251 GHz [26]. Throughout this thesis work, we demonstrate new record f_T and f_{max} of 225 GHz and 300 GHz, respectively in 2010 and add new points to Figure 1-4.

Although these results are excellent, there are still important issues that need to be overcome to further increase the performance of GaN HEMTs at mm-wave frequencies. One of the biggest challenges in mm-wave GaN HEMTs is that the typical frequency performance of these devices, as measured by the current-gain cutoff frequency (f_T) and the power-gain cutoff frequency (f_{max}), is lower than what the intrinsic material properties predict. Figure 1-5 shows f_T values reported in the literature as a function of gate length (L_g), compared to the calculated f_T behavior from ideal gate length scaling. There is a large variation of reported f_T in the literature even with the same L_g

and the reported f_T values are significantly lower than their expected values, especially for the shorter gate lengths. Since f_{max} is a function of f_T, we also observe a large variation of the f_{max} values reported by different groups and a considerable discrepancy between measured and expected f_{max}. Several hypotheses have been proposed to explain the lower-than-expected frequency performance. Some of them include short-channel effects in scaled devices [27], parasitic charging delay associated with the access regions [28], much larger effective gate length than the lithographic gate length [29], and degradation of the gate modulation efficiency by interface defects and traps [30]. However, it is still not clear which one of these effects is primarily limiting high frequency performance of GaN HEMTs and more detailed and systematic understanding is needed to identify solutions to this problem.

1.3. Integration of GaN and Si electronics

Another important challenge (and opportunity) in GaN HEMTs is their integration with silicon electronics. Over the last 50 years, Si electronics has revolutionized our world and one of the main drivers for this revolution has been Moore's law. One of the several possible interpretations of Moore's law states that the number of transistors on an integrated circuit (IC) doubles about every two years and this trend has hold for the last half a century [31][32]. However, as the device size gets smaller and smaller, it becomes more and more difficult and expensive to keep up with Moore's law. Physical limits of atomic structures and power density could limit device scaling by 2020 [33]. Thus, we need to find out alternative ways to keep increasing the circuit performance and functionality besides simply scaling down the device size. One way around the size limitation is to introduce new semiconductor materials. There are several material systems that offer better performance than silicon and by integrating them into Si ICs on an as-needed basis, one can expect higher performance as well as new functionalities and flexibilities. The compound semiconductor materials on silicon (COSMOS) program of the U.S. defense advanced research projects agency (DARPA) represents an excellent example of efforts to integrate InP HBTs and Si CMOS electronics [34]. However, in spite of the great commercial and military interest in nitride electronics, no work has been done on the integration of nitride semiconductors and Si CMOS electronics.

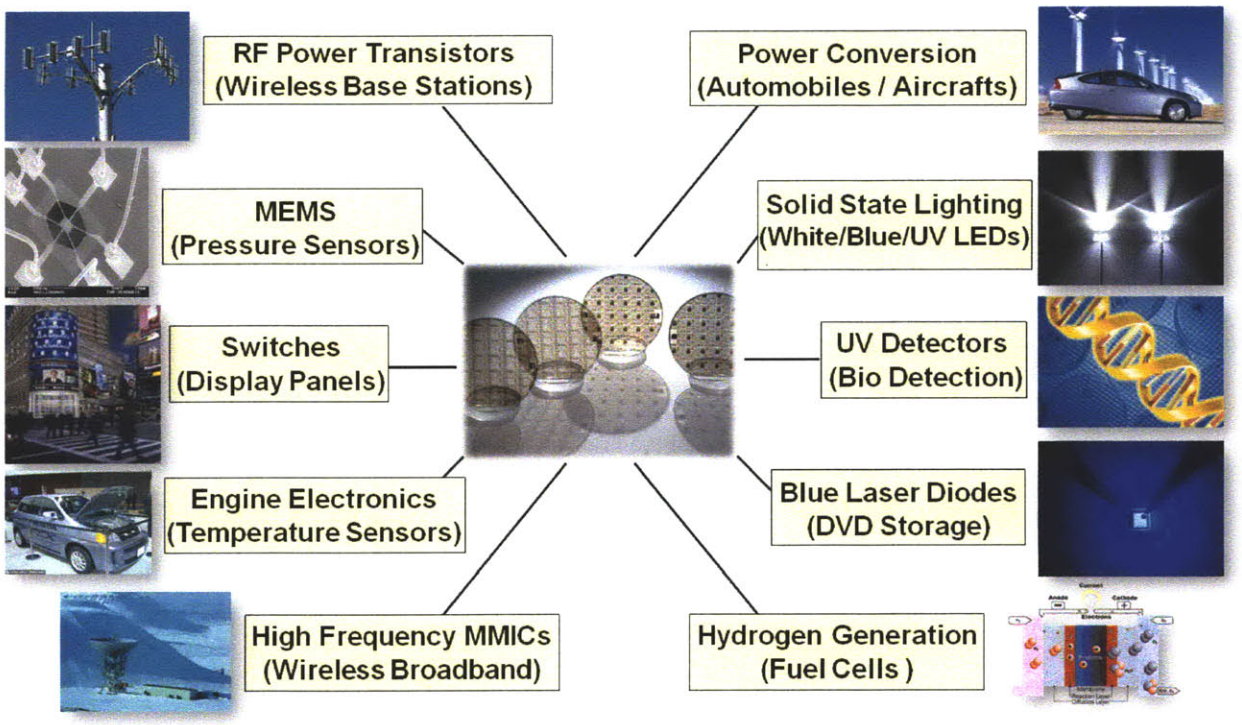


Figure 1-6. Summary of some of the many applications of nitride semiconductors.

GaN is one of the most promising candidates for the heterogeneous integration with Si. GaN has unique properties in both electronics and optoelectronics that Si cannot achieve. While Si electronics has shown unsurpassed levels of scaling and circuit complexity, GaN devices offer excellent high frequency/power performance as well as outstanding optoelectronic properties. In addition, GaN technology is matured enough for commercial applications and it is already widely used in a variety of fields such as solid-state lighting, power electronics, and energy harvesting devices as shown in Figure 1-6. The combination of the excellent properties of GaN with the low cost/high integration capabilities of Si technology would allow unprecedented device performance and flexibility, as well as new circuit functionality.

1.4. Project goal and thesis outline

This thesis has two main goals. The first one is to identify the main problems that limit the high frequency performance of GaN transistors and demonstrate novel solutions to overcome them. To understand the high frequency limitations of GaN transistors and to demonstrate state-of-the-art devices, we have performed a combination of systematic analysis of device characteristics

and development of advanced fabrication technologies. The second goal is to demonstrate the first integration of high performance GaN transistors with silicon (100) electronics on the same wafer through a scalable technology.

The thesis is organized as follows:

In chapter 2, the basic operation principles of GaN HEMTs are outlined. The important concepts and figures of merit of DC and RF characteristics are described. Also, a concise description of the standard fabrication process is provided.

In chapter 3, the f_{max} of GaN transistors fabricated at MIT is optimized thanks to the development of new processing technologies that minimize parasitic components in the device operation. These technologies combined a low-damage gate recess, scaled device geometry, and recessed source/drain ohmic contacts to simultaneously enable minimum short-channel effects (i.e. low output conductance, g_o) and very low parasitic resistances. AlGaN/GaN HEMTs on a SiC substrate with record f_{max} are demonstrated. For example, a 60-nm gate-length HEMT with recessed AlGaN barrier exhibited excellent g_o of 10.4 mS/mm, R_{on} of 1.1 ~ 1.2 Ω·mm, and f_{max} of 300 GHz with a breakdown voltage of ~ 20 V. The obtained f_{max} is the highest reported to date for any nitride transistor. The accuracy of the f_{max} value is verified by small-signal modeling based on carefully extracted S-parameters.

In chapter 4, advanced processing technologies are developed to demonstrate state-of-the-art f_T in GaN transistors. The lower-than-expected frequency performance observed in many AlGaN/GaN HEMTs is attributed to a significant drop of the intrinsic small-signal transconductance (RF g_m) with respect to the intrinsic DC g_m. To reduce this RF g_m-collapse and improve high frequency performance, we have developed a new technology based on a combination of vertical gate-recess, oxygen plasma treatment, and lateral gate-etch technique which resulted in AlGaN/GaN HEMTs with a record f_T of 225 GHz for a gate length of L_g = 55 nm, and 162 GHz for an L_g of 110 nm.

In chapter 5, the first on-wafer integration of Si(100) MOSFETs and GaN HEMT is demonstrated. A new layer transfer technology based on wafer bonding and etch-back process

was developed to enable on-wafer integration of both Ga- and N-face GaN HEMTs with Si(100) electronics. Using this technology, GaN and Si devices separated by less than 5 μm from each other have been fabricated, which is suitable for building future heterogeneous integrated circuits.

In chapter 6, summary and conclusions are presented. Future work to further expand the frequency performance of AlGaN/GaN HEMTs beyond mm-wave frequencies is discussed. Also, research directions to investigate RF g_m-collapse and to improve the integration technology are provided.

1.5. References

[1] O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck, "Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 85, no. 6, pp. 3222-3233, Mar. 1999.

[2] Y. Uemoto, D. Shibata, M. Yanagihara, H. Ishida, h. Matsuo, S. Nagai, N. Batta, M. Li, T. Ueda, T. Tanaka, and D. Ueda, "8300V blocking voltage AlGaN/GaN power HFET with thick poly-AlN passivation," *IEEE IEDM Tech. Digest*, pp. 861-864, 2007.

[3] K. H. Chen, B. S. Kang, H. T. Wang, T. P. Lele, F. Ren, Y. L. Wang, C. Y. Chang, S. J. Pearton, D. M. Dennis, J. W. Johnson, P. Rajagopal, J. C. Roberts, E. L. Piner, and K. J. Linthicum, "c-erbB-2 sensing using AlGaN/GaN high electron mobility transistors for breast cancer detection," *Appl. Phys. Lett.*, vol. 92, no. 19, 192103, May. 2008.

[4] F. Medjdoub, J. F. Carlin, M. Gonschorek, E. Feltin, M. A. Py, D. Ducatteau, C. Gaquiere, N. Grandjean, and E. Kohn, "Can InAlN/GaN be an alternative to high power/high temperature AlGaN/GaN devices?," *IEEE IEDM Tech. Digest*, 2006.

[5] I. Akasaki, "Nitride semiconductors- impact on the future world," *Journal of Crystal Growth*, vol. 237-239, pp. 905-911, Apr. 2002.

[6] H. P. Maruska and J. J. Tietjen, "The preparation and properties of vapor-deposited single-crystalline GaN," *Appl. Phys. Lett.*, vol. 15, no. 10, pp. 327-329, Aug. 1969.

[7] H. Amano, N. Sawaki, I. Akasaki, and Y. Toyoda, "Metalorganic vapor phase epitaxial growth of a high quality GaN film using an AlN buffer layer," *Appl. Phys. Lett.*, vol. 48, no. 5, pp. 353-355, Feb. 1986.

[8] H. Amano, M. Kito, K. Hiramatsu, and I. Akasaki, "P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI)," *Jpn. J. Appl. Phys.*, vol. 28, no. 12, pp. L2112-L2114, Dec. 1989.

[9] S. Nakamura, T. Mukai, M. Senoh, and N. Iwasa, "Thermal annealing effects on p-type Mg-doped GaN films," *Jpn. J. Appl. Phys.*, vol. 31, no. 2B, pp. L139-L142, Feb. 1992.

[10] S. Nakamura, M. Senoh, S.Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, and Y. Sugimoto, "InGaN-based multi-quantum-well-structure laser diodes," *Jpn. J. Appl. Phys.*, vol. 35, no. 1B, pp. L74-L76, Jan. 1996.

[11] M. A. Khan, J. M. Van Hove, J. N. Kuznia, and D. T. Olson, "High electron mobility GaN/$Al_xGa_{1-x}N$ heterostructures grown by low-pressure metalorganic chemical vapor deposition," *Appl. Phys. Lett.*, vol. 58, no. 21, pp. 2408-2410, May. 1991.

[12] M. A. Khan, A. Bhattarai, J. N. Kuznia, and D. T. Olson, "High electron mobility transistor based on a GaN-Al_xG_{1-x}aN heterojunction", *Appl. Phys. Lett.*, vol. 63, pp. 1214-1215, 1993.

[13] Y.-F. Wu, A. Saxler, M. Moore, R. P. Smith, S. Sheppard, P. M. Chavarkar, T. Wisleder, U. K. Mishra, and P. Parikh, "30-W/mm GaN HEMTs by field plate optimization," *IEEE Electron Dev. Lett.*, vol. 25, pp. 117-119, 2004.

[14] Y. F. Wu, M. Moore, A. Abrahamsen, M. Jacob-Mitos, P. Parikh, S. Heikman, and A. Burk, "High-voltage millimiter-wave GaN HEMTs with 13.7 W/mm power density," *IEEE IEDM Tech, Digest*, pp. 405-407, 2007.

[15] M. Micovic, A. Kurdoghilian, P. Hashimoto, M. Hu, M. Antcliffe, P. J. Willadsen, W. S. Wong, R. Bowen, I. Milosavljevic, A. Schmitz, Y. Yoon, M. Wetzel, and D. H. Chow, "GaN HFET for W-band power applications," *IEEE IEDM Tech. Digest*, 2006.

[16] Toshiba website: http://www.toshiba.com/taec/news/press_releases/news_pr_10.jsp

[17] M. Feng, S.-C. Shen, D. C. Caruth, J.-J Huang, "Device technologies for RF front-end circuits in next-generation wireless communications," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 354-375, Feb. 2004.

[18] E. O. Johnson, "Physical limitations on frequency and power parameters of transistors," *RCA Review*, vol. 26, pp. 163-177, Jun. 1965.

[19] M. A. Khan, J. N. Kuznia, D. T. Olson, W. J. Schaff, J. W. Burm, and M. S. Shur, "Microwave performance of a 0.25 μm gate AlGaN/GaN heterostructure field effect transistor," *Appl. Phys. Lett.*, vol. 65, pp. 1121-1123, 1994.

[20] Y.-F. Wu, B. P. Keller, S. Keller, N. X. Nguyen, M. Le, C. Nguyen, T. J. Jenkins, L. T. Kehias, S. P. DenBaars, and U. K. Mishra, "Short channel AlGaN/GaN MODFET's with 50-GHz f_T and 1.7-W/mm output-power at 10 GHz," *IEEE Electron Dev. Lett.*, vol. 18, pp. 438-440, 1997.

[21] M. Micovic, N. X. Nguyen, P. Janke, W.-S. Wong, P. Hashimoto, L.-M. McCray, and C. Nguyen, "GaN/AlGaN high electron mobility transistors with f_T of 110 GHz," *Electronic Lett.*, vol. 36, pp. 358-359, 2000.

[22] V. Kumar, W. Lu, R. Schwindt, A. Kuliev, G. Simin, J. Yang, M. A. Khan, and I. Adesida, "AlGaN/GaN HEMTs on SiC with f_T of over 120 GHz," *IEEE Electron Dev. Lett.*, vol. 23, no. 8, pp. 455-457, Aug. 2002.

[23] T. Palacios, E. Snow, Y. Pei, A. Chakraborty, S. Keller, S. P. DenBaars, and U. K. Mishra, "Ge-spacer technology in AlGaN/GaN HEMTs for mm-waver applications," *IEEE IEDM Tech. Digest*, 2005.

[24] T. Palacios, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "AlGaN/GaN high electron mobility transistors with InGaN back-barriers," *IEEE Electron Dev. Lett.*, vol. 27, no. 1, pp. 13-15, Jan. 2006.

[25] M. Higashiwaki, T. Mimura, and T. Matsui, "30-nm-gate AlGaN/GaN heterostructure field-effect transistors with a current-gain cutoff frequency of 181 GHz," *Jpn. J. Appl. Phys.*, vol. 45, no. 42, pp. L1111-L1113, Oct. 2006.

[26] M. Higashiwaki, T. Mimura, and T. Matsui, "AlGaN/GaN heterostructure field-effect transistors on 4H-SiC substrates with current-gain cutoff frequency of 190 GHz," *Appl. Phys. Express*, vol. 1, 021103, 2008.

[27] G. H. Jessen, R. C. Fitch, J. K. Gillespie, G. Via, A. Crespo, D. Langley, D. J. Denninghoff, M. Trejo, and E. R. Heller, "Short-channel effect limitations on high-frequency operation of AlGaN/GaN HEMTs for T-gate devices," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2589-2597, 2007.

[28] T. Suemitsu, K. Shiojima, T. Makimura, and N. Shigekawa, "Intrinsic transit delay and effective electron velocity of AlGaN/GaN high electron mobility transistors," *Jpn. J. Appl. Phys.*, vol. 44, no. 6, pp. L211-L213, Jan. 2005.

[29] Y.-R. Wu, M. Singh, and J. Singh, "Device scaling physics and channel velocities in AlGaN/GaN HFETs: Velocities and effective gate length," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 588-593, Apr. 2006.

[30] M. C. Foisy, P. J. Tasker, B. Hughes, and L. F. Eastman, "The role of inefficient charge modulation in limiting the current-gain cutoff frequency of the MODFET," *IEEE Trans. Electron Devices*, vol. 35, pp. 871-878, 1988.

[31] R. R. Schaller, "Moore's law: past, present, and future," *IEEE Spectrum*, vol. 34, no. 6, pp. 52-59, Jun. 1997.

[32] S. E. Thompson and S. Parthasarathy, "Moore's law: the future of Si microelectronics," *Materials Today*, vol. 9, no. 6, pp. 20-25, Jun. 2006.

[33] Intel website: http://www.intel.com/technology/mooreslaw/

[34] S. Raman, T.-H. Chang, C. L. Dohrman, and M. J. Rosker, "The DARPA COSMOS program: the convergence of InP and silicon CMOS technologies for high-performance mixed-signal," *International Conference on Indium Phosphide and Related Materials (IPRM)*, Jun. 2010.

Chapter 2. Basics of GaN HEMTs

2.1. DC analysis

The High Electron Mobility Transistor (HEMT) is a field effect transistor incorporating a heterojunction between two layers of dissimilar crystalline semiconductors with different band gaps. The first HEMT was developed by Mimura *et al.* in 1980 [1] and the first commercial HEMT was used as a cryogenic low-noise amplifier for the radio telescope in 1985 which later discovered new interstellar molecules in the Taurus Molecular Cloud about 400 light years away [2]. The most important feature of the HEMT is the two-dimensional electron gas (2DEG), a gas of electrons which is induced at the heterojunction interface. Figure 2-1(a) shows the schematic of the widely studied AlGaAs/GaAs HEMT in which the wider band gap AlGaAs layer is n-doped to provide electrons to the 2DEG at the interface. The excess electrons in the AlGaAs layer diffuses toward the adjacent narrower band gap GaAs layer due to its lower energy level. The diffusion of these electrons creates an electric field in the direction to oppose the electron movement. The electron diffusion and electron drift balance each other, creating a junction at equilibrium similar to a p-n junction. Thus, electrons now reside in undoped narrower band gap GaAs layer forming the 2DEG. The 2DEG is tightly confined in the direction perpendicular to the heterojunction interface, however free to move in the parallel direction. The main advantage of the 2DEG in HEMTs is its higher electron mobility than in other field effect transistors such as metal-oxide-semiconductor field effect transistor (MOSFET) and metal-semiconductor field effect transistor (MESFET). The physical separation of the dopants (e.g. in the AlGaAs layer) from the mobile carrier electrons (e.g. in the GaAs layer) allows the 2DEG in HEMTs to have reduced ionized impurity scattering and high mobility.

In nitride semiconductors, a commonly used material combination for HEMT structures is an AlGaN barrier and a GaN channel layer as shown in Figure 2-1(b). The AlGaN/GaN system benefits not only from its high electron mobility (~ 2000 $cm^2/V \cdot s$) but also from the high 2DEG density (~ 10^{13} cm^{-2}) thanks to the large conduction band discontinuity between AlGaN and GaN. Contrary to AlGaAs/GaAs devices, the high 2DEG density is not induced by doping, but

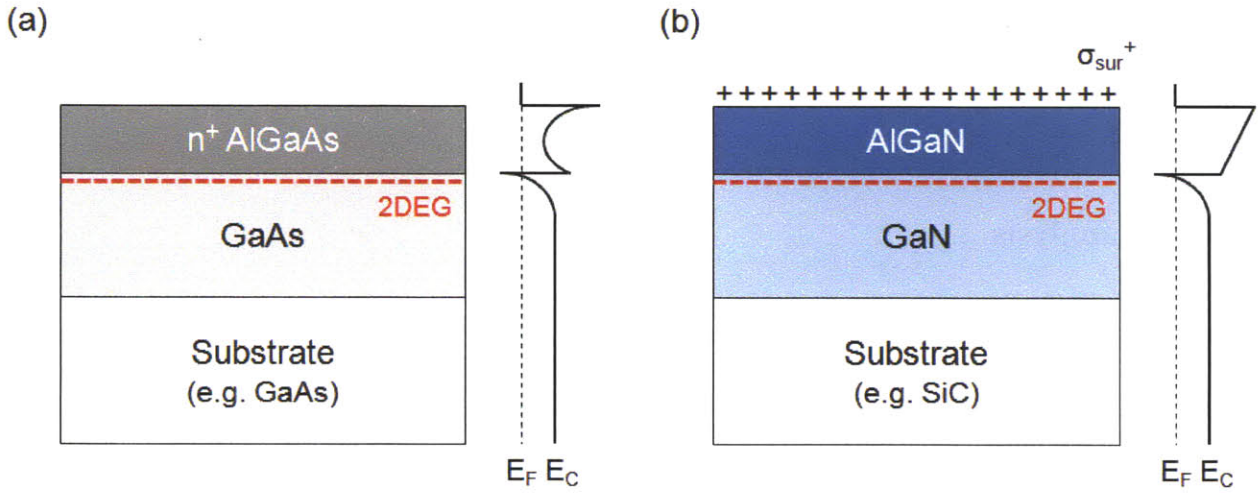


Figure 2-1. Schematic cross-section of (a) AlGaAs/GaAs and (b) AlGaN/GaN heterostructure with associated conduction band diagrams. The 2DEG in AlGaAs/GaAs system is provided from the n-doped AlGaAs layer by the balance of electron drift/diffusion while the 2DEG in AlGaN/GaN system is produced by donor-like AlGaN surface states with the polarization field inside AlGaN layer. Intentional doping is not necessary to form the 2DEG in AlGaN/GaN system.

by donor-like surface states at the AlGaN surface facilitated by spontaneous and piezoelectric polarization electric field inside the AlGaN layer [3]. The 2DEG density in this case is mainly determined by the thickness of AlGaN layer and the amount of aluminum percentage in it.

Figure 2-2 shows a schematic cross-section of a standard AlGaN/GaN HEMT. The current flows from the drain to the source ohmic contacts, and it is controlled by the voltage in the Schottky gate electrode, which varies the 2DEG density. Accurate modeling of carrier transport in AlGaN/GaN HEMTs typically requires solving the Schrödinger and Poisson equations, and sometimes even time-consuming Monte Carlo simulations. However, massive computational modeling/simulation can sometimes obscure the important physical concepts behind the HEMT operation. As an engineer, it is important to have an insight to solve practical challenges timely with an eye toward identifying problems and optimizing performance. In this regard, to develop equations that are compact enough to allow general analysis and problem solving is very useful. Therefore, in this section we will focus on a relatively simple analytical form, yet sufficient to estimate practical operation of GaN HEMTs. For the simple analytical model, the linear charge control approximation, the gradual channel approximation, and two-piece linear velocity-field model are used.

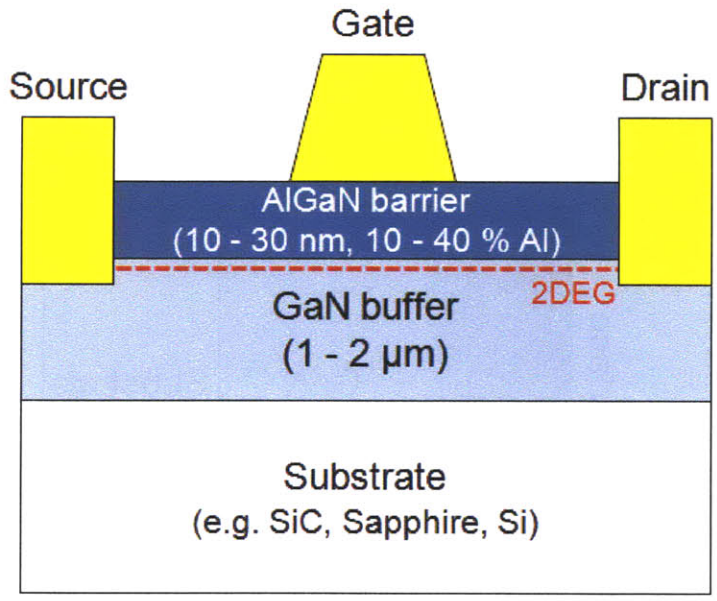


Figure 2-2. Schematic cross-section of a standard AlGaN/GaN HEMT.

2.1.1. Charge control model

The charge distribution in HEMTs is determined by electrostatics and varied by applying a voltage to the gate electrode. A one-dimensional charge control model describes the behavior of the charge in the direction perpendicular to the heterojunction interface as a function of the gate voltage. Figure 2-3 illustrates a schematic diagram of an AlGaN/GaN HEMT structure along with resulting charge distribution and band diagram. From the point of view of charge control, the main difference between AlGaN/GaN HEMTs and conventional HEMTs is the origin of the electrons in the 2DEG. In conventional HEMTs, the channel electrons are provided by the intentionally doped barrier layer, while in AlGaN/GaN HEMTs they come from donor-like surface states. Under linear charge control approximation, the 2DEG charge density is given by

$$qn_s = C(V_{GS} - V_{th}) \tag{2-1}$$

where q is the electronic charge, n_s is the 2DEG carrier density, C is the gate capacitance per unit area, and the V_{th} is the threshold voltage. The 2DEG is assumed to behave as a perfect two-dimensional sheet whose centroid is placed at a distance Δd from the heterointerface ($\Delta d \approx 20$ Å for AlGaN/GaN system [4]). The physical origin of Δd is the spread of the 2DEG wave function at the AlGaN/GaN interface and C is given by

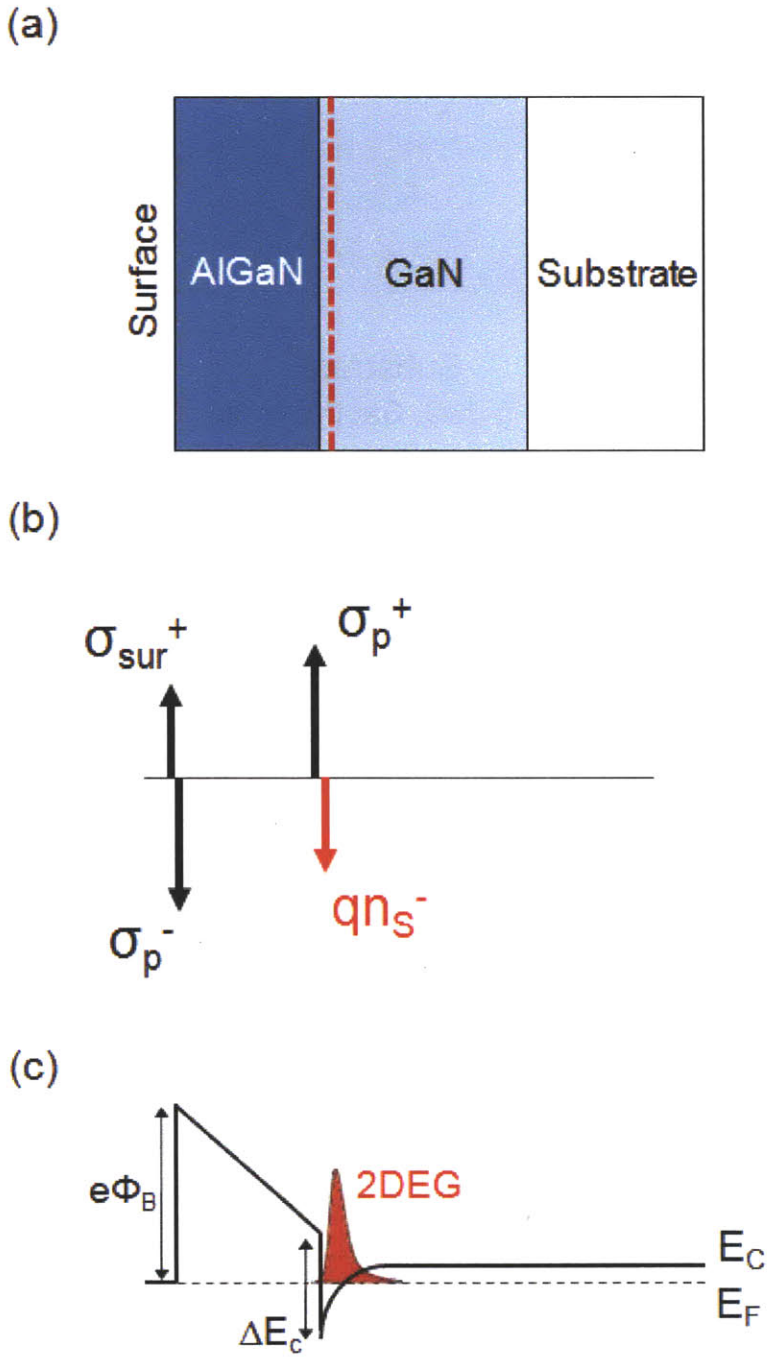


Figure 2-3. (a) Schematic cross-section of a standard AlGaN/GaN heterostructure along with (b) the charge distribution and (c) the conduction band diagram. σ_{sur}^{+} is the positively ionized surface state charge and σ_p is the polarization induced charge.

$$C = \frac{\varepsilon}{d_{AlGaN} + \Delta d} \tag{2-2}$$

where ε is the dielectric constant between the AlGaN surface and the 2DEG, and d_{AlGaN} is the AlGaN barrier thickness. The threshold voltage V_{th} is the gate voltage required to completely deplete the 2DEG carrier density (n_s = 0) at the source end of the gate. V_{th} can be given by following the energy bands from the Fermi level in the metal to the lowest level of the triangular quantum well in the GaN and

$$V_{th} = \emptyset_B - \frac{\sigma_p \cdot d_{AlGaN}}{\varepsilon} - \frac{\Delta E_c}{q} \tag{2-3}$$

where Φ_B is the Schottky barrier height, ΔE_c is the conduction band discontinuity, and σ_p is the net polarization charge at the AlGaN/GaN interface ($\sigma_p = \sigma_{p,AlGaN} - \sigma_{p,GaN}$). Thus, by substituting Eq. (2-2) and Eq. (2-3) into Eq. (2-1), the 2DEG charge density at the source end of the gate as a function of gate voltage V_{GS} can be written as

$$qn_s = \frac{\varepsilon}{d_{AlGaN} + \Delta d}\left(V_{GS} - \emptyset_B + \frac{\sigma_p \cdot d_{AlGaN}}{\varepsilon} + \frac{\Delta E_c}{q}\right) \tag{2-4}$$

2.1.2. Velocity-field relationship

In general, the drift velocity of carriers at high electric fields deviates from the linear relationship observed at low fields. Figure 2-4 shows a plot of average electron drift velocity as a function of applied electric field for several semiconductors [5]. The drift velocity in AlGaN/GaN HEMTs, for example, peaks at approximately 3×10^7 cm/s at an electric field of about 160 kV/cm and then decreases. At even higher electric fields, the drift velocity becomes saturated as in the case of GaAs to a value of about 1×10^7 cm/s (not shown in Figure 2-4). The slope of the drift velocity versus electric field is the differential mobility and the negative differential mobility at high electric field is mainly due to the intervalley electron transfer [6].

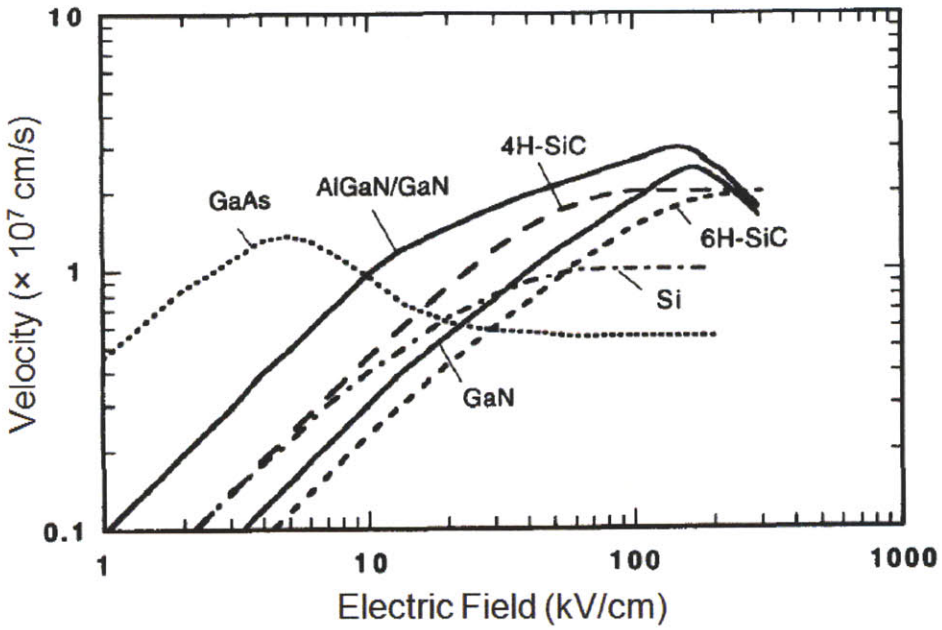


Figure 2-4. Carrier velocity vs. electric field for several semiconductors [5].

Typically, such a velocity-field relationship can be simplified by two-piece linear velocity-field model as shown in Figure 2-5.

$$v = \begin{cases} \mu E(x) & \text{for } E(x) < E_c \\ v_{sat} & \text{for } E(x) > E_c \end{cases} \tag{2-5}$$

where μ is the low field mobility, $E(x)$ is the channel electric field, E_c is the critical electric field for the velocity saturation, and v_{sat} is the carrier saturation velocity.

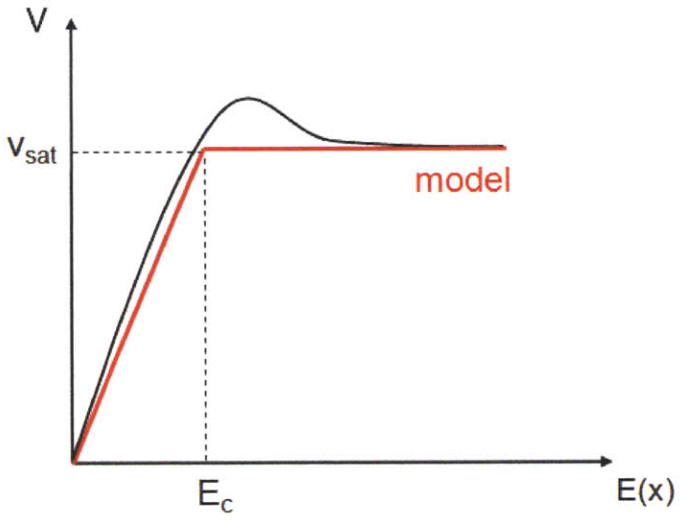


Figure 2-5. The velocity-field relationship in AlGaN/GaN HEMTs is simplified by two-piece linear velocity-field model ($v = \mu E(x)$ for $E(x) < E_C$ and $v = v_{sat}$ for $E(x) > E_C$).

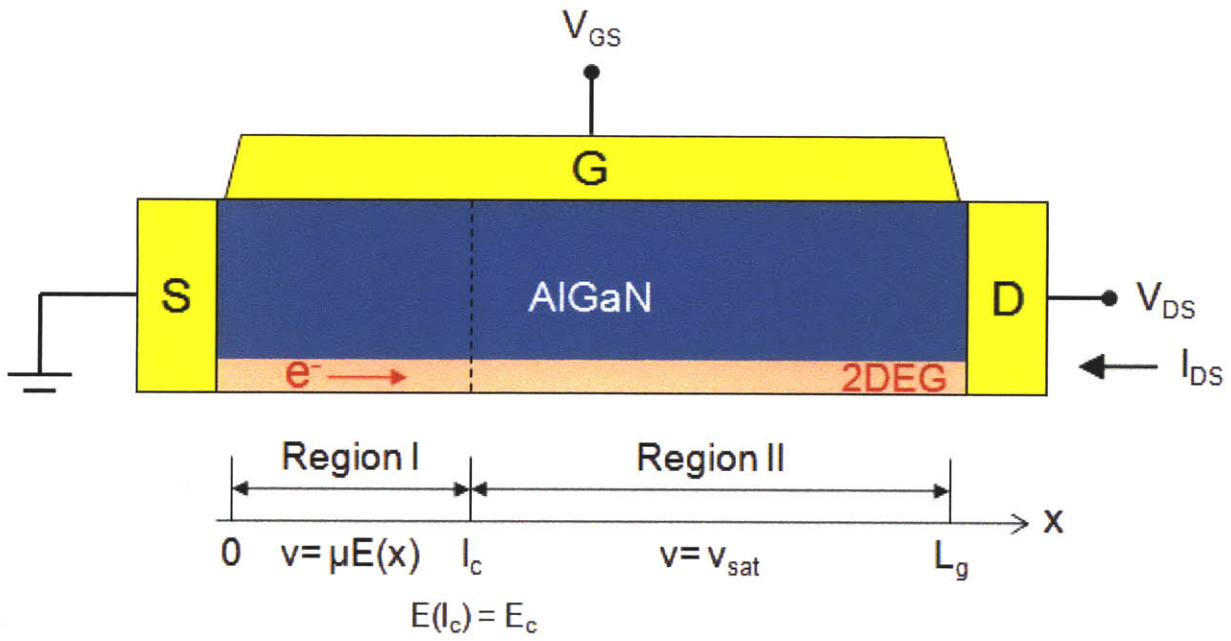


Figure 2-6. Schematic cross-section of the AlGaN/GaN HEMT structure used in this chapter to analyze the operating principle of GaN transistors. Extrinsic source and drain access regions are neglected and the origin of x-axis is set to the source end of the gate.

2.1.3. DC characteristics

The intrinsic AlGaN/GaN HEMT is divided into two regions as shown in Figure 2-6, namely the linear region (Region I) and the velocity saturation region (Region II). During the device operation, a drain bias is applied and an x-dependent potential $V(x)$ (or $E(x) = dV(x)/dx$) exists along the channel. In Region I, the electric field $E(x)$ is smaller than the critical electric field E_c ($E(x) < E_c$) and the electron drift velocity follows $v = \mu E(x)$. The boundary between Region I and Region II is set to the point where the magnitude of the electric field reaches the critical value E_c. In Region II, $E(x)$ is high enough ($E(x) > E_c$) for the electrons to travel with their saturation velocity v_{sat}.

The drain current I_{DS} normalized to the gate width at any position x can be expressed as

$$I_{DS} = qn_s(x)v(x) \tag{2-6}$$

By taking into account the channel potential $V(x)$ between the position x and the source end of the gate, the 2DEG charge density is given by

$$qn_s = C(V_{GS} - V_{th} - V(x)) \tag{2-7}$$

In the Region I ($E(x) < E_c$), substituting Eq. (2-5) and Eq. (2-7) into Eq. (2-6) leads to

$$I_{DS} = C(V_{GS} - V_{th} - V(x)) \cdot \mu \frac{dV(x)}{dx} \tag{2-8}$$

Rearranging the result yields a differential equation with separated variables, and integrating it over the Region I gives,

$$I_{DS} = \frac{\mu C}{x}\left[(V_{GS} - V_{th})V(x) - \frac{V(x)^2}{2}\right] \tag{2-9}$$

When the applied drain bias is small, the Region I extends over the entire channel and by substituting $x = L_g$ and $V(x) = V_{DS}$, the drain current in the linear region of an AlGaN/GaN HEMTs can be written as

$$I_{DS,lin} = \frac{\mu C}{L_g}\left[(V_{GS} - V_{th})V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{2-10}$$

The transconductance is defined as the rate of change in the drain current with respect to V_{GS} and can be obtained by differentiating Eq. (2-10) with respect to V_{GS}.

$$g_{m,lin} = \frac{\mu C}{L_g} V_{DS} \tag{2-11}$$

Now, if the applied drain bias is large enough to induce velocity saturation somewhere in the channel, Eq. (2-10) no longer holds. To derive the drain current in the saturation region, we focus on the boundary condition between Region I and Region II. In Eq. (2-9), solving for $V(x)$ and $E(x)$ gives,

$$V(x) = V_{GS} - V_{th} - \sqrt{(V_{GS} - V_{th})^2 - \frac{2I_{DS}x}{\mu C}} \tag{2-12}$$

$$E(x) = \frac{dV(x)}{dx} = \frac{I_{DS}}{\mu C\sqrt{(V_{GS} - V_{th})^2 - \frac{2I_{DS}x}{\mu C}}} \tag{2-13}$$

At the boundary between Region I and Region II where $x = l_c$, the saturation velocity v_{sat} needs to satisfy $v_{sat} = \mu E(l_c) = \mu E_c$ following the two-piece linear velocity-field model. Thus, from Eq. (2-13) the drain current in the saturation region in AlGaN/GaN HEMTs can be calculated as

$$I_{DS,sat} = v_{sat}C\left[\sqrt{(V_{GS} - V_{th})^2 + (l_cE_c)^2} - l_cE_c\right] \tag{2-14}$$

The saturation current is primarily controlled by the gate voltage V_{GS}. The transconductance in the saturation region is obtained by differentiating Eq. (2-14) with respect to V_{GS}.

$$g_{m,sat} = v_{sat}C\frac{V_{GS} - V_{th}}{\sqrt{(V_{GS} - V_{th})^2 + (l_cE_c)^2}} \tag{2-15}$$

In short gate length devices, l_c becomes small and Eq. (2-14) and Eq. (2-15) reduce to

$$I_{DS,sat} \approx v_{sat}C(V_{GS} - V_{th}) \tag{2-16}$$

$$g_{m,sat} \approx v_{sat}C \tag{2-17}$$

It is noted that the saturation velocity v_{sat} can be replaced by the ballistic velocity v_0 when the device operates in the ballistic region.

It should be noted that only the intrinsic part of AlGaN/GaN HEMT has been considered so far. The effect of the extrinsic components, such as source and drain resistances (R_s and R_d), can be included by redefining V_{GS} and V_{DS} as

$$V_{GS} = V_{GS,ext} - I_{DS}R_s \tag{2-18}$$

$$V_{DS} = V_{DS,ext} - I_{DS}(R_s + R_d) \tag{2-19}$$

where $V_{GS,ext}$ and $V_{DS,ext}$ are externally applied voltages to the gate and drain electrodes, respectively.

2.2. RF analysis

As the operating frequency of the device increases to the microwave range (f = 300 MHz ~ 300 GHz, $\lambda = c / f$ = 1 m ~ 1 mm), standard circuit theory generally cannot be used directly to analyze the device and circuit. At such high frequencies, the wavelength of the signal is comparable to the physical dimension of the device and there exists appreciable phase shift in voltage or current over the physical extent of the device. The device can be analyzed by solving the complete Maxwell's equations, however it is mathematically complicated and often provides more information than what we need for most practical purposes. We are typically interested in terminal quantities such as voltage and current. For this purpose, the two-port network concept and associated small-signal equivalent circuit model are typically used to describe microwave transistors and to analyze frequency performance of the device.

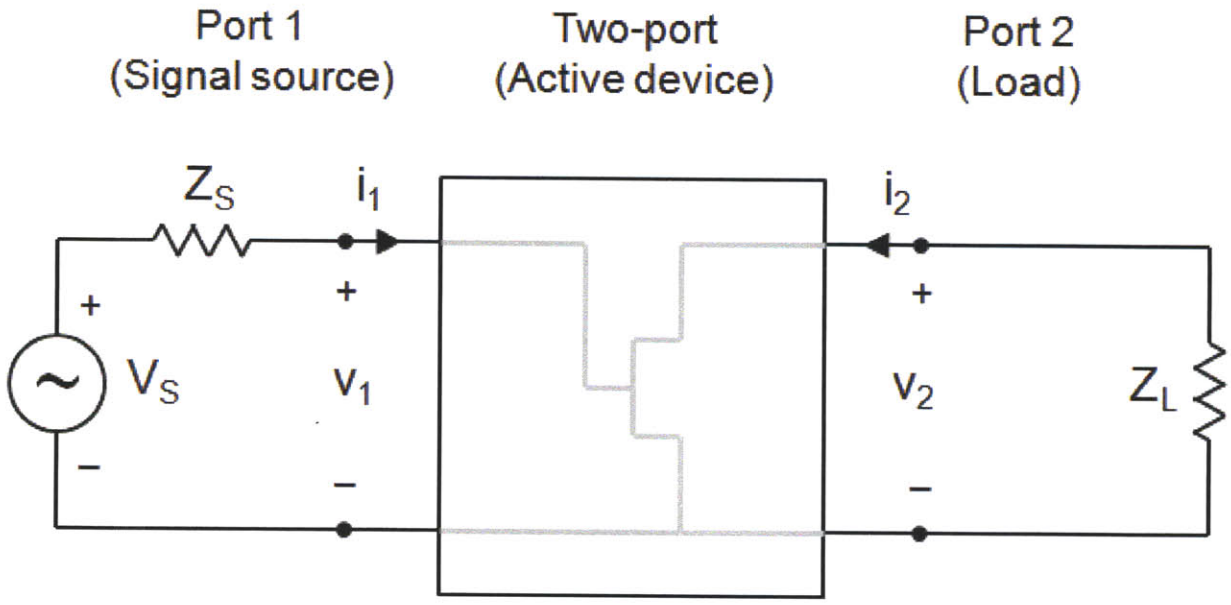


Figure 2-7. A microwave transistor (active device) regarded as a two-port network with signal source and load.

2.2.1. Two-port network

Any active device, including microwave AlGaN/GaN HEMTs, can be treated as a two-port network as shown in Figure 2-7. Port 1 refers to the input connected to a signal source consisting of a voltage source V_S and a source impedance Z_S. Port 2 is the output connected to a load impedance Z_L. The input and output currents of the two-port network are denoted as i_1 and i_2, while the input and output voltages are v_1 and v_2. The two-port network analysis considers the active device as a black box whose input and output voltages and currents can be measured at each port. These terminal quantities can be related to several important figures of merit to characterize any active device. There is a number of parameter sets such as Y-, Z-, H-, or S-parameters which contain essentially the same information, but represent it in a different way. For example, the Y-parameters relate i_1, i_2, v_1, and v_2 under ac short-circuit conditions as

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$i_1 = y_{11}v_1 + y_{12}v_2 \qquad i_2 = y_{21}v_1 + y_{22}v_2$$

$$y_{11} = \left.\frac{i_1}{v_1}\right|_{v_2=0} \qquad y_{12} = \left.\frac{i_1}{v_2}\right|_{v_1=0} \qquad y_{21} = \left.\frac{i_2}{v_1}\right|_{v_2=0} \qquad y_{22} = \left.\frac{i_2}{v_2}\right|_{v_1=0}$$

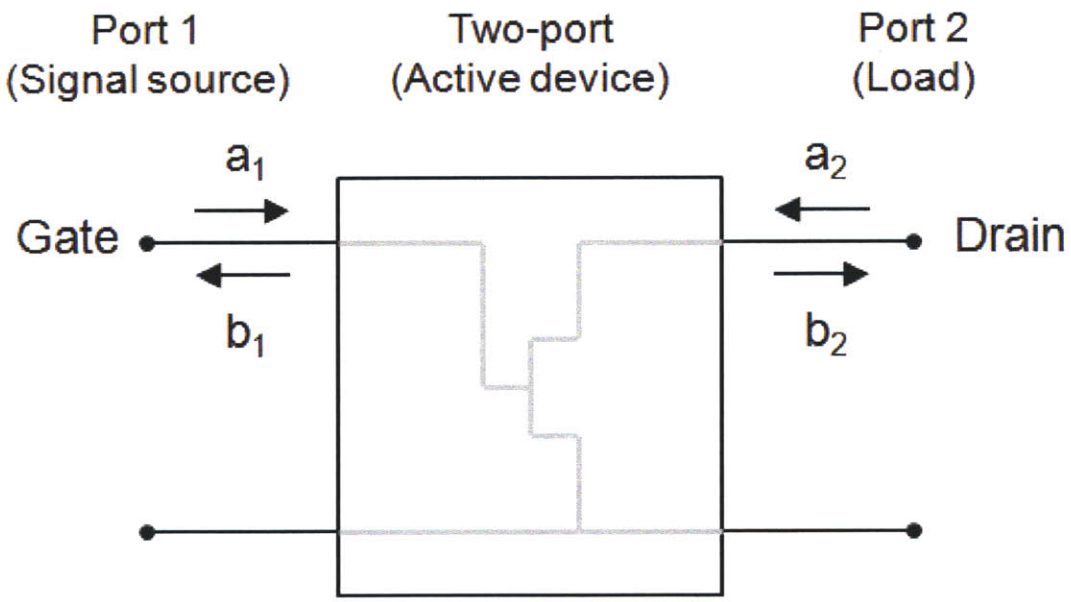


Figure 2-8. Two-port network characterized by the incident (a_1, a_2) and reflected (b_1, b_2) waves (S-parameters).

where y_{11} and y_{22} are the input and output admittances, and y_{12} and y_{21} are reverse and forward transfer admittances, respectively. Since the Y-parameters are more closely related to device physics and conceptually easier to interpret than other parameter sets, they are widely used for microwave devices. However, although Y-parameters are more easily interpretable, it is very difficult to measure external voltages and currents and to realize the required short-circuit terminations in the microwave range. Therefore, another set of parameters called S-parameters is typically used to obtain the device information. Then, the measured S-parameters are converted to Y-parameters to interpret its characteristics. The S-parameters are ratios of the powers of travelling wave signals and can be measured directly with a vector network analyzer. During the S-parameter measurement, each signal can be isolated by the termination with a matched load to eliminate its reflection. The use of the-more-difficult-to-implement short- or open-circuit terminations is, in that way, avoided. The S-parameters of a two-port network are expressed in a similar manner to the Y-parameters.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$b_1 = s_{11}a_1 + s_{12}a_2 \qquad b_2 = s_{21}a_1 + s_{22}a_2$$

$$s_{11} = \left.\frac{b_1}{a_1}\right|_{output\ match} \qquad s_{12} = \left.\frac{b_1}{a_2}\right|_{input\ match} \qquad s_{21} = \left.\frac{b_2}{a_1}\right|_{output\ match} \qquad s_{22} = \left.\frac{b_2}{a_2}\right|_{input\ match}$$

where a and b are the powers of incident and reflected waves as shown in Figure 2-8. s_{11} and s_{22} are the input and output reflection coefficients, and s_{12} and s_{21} are reverse and forward gains, respectively. s_{11} and s_{21} are measured by terminating the output port with a matched load, while s_{12} and s_{22} are extracted by terminating the input port with a matched load. The conversion equations between the different sets of parameters (e.g. S- to Y-parameters) are given in Appendix 1.

2.2.2. Small-signal equivalent circuit

The high frequency operation of the device is often studied in terms of the small-signal equivalent circuit model. The small-signal implies that the magnitude of the signal voltage is on the order k_BT / q (~ 26 mV at T = 300 K). Under such a small-signal condition, nonlinear characteristics of a field-effect transistor such as an AlGaN/GaN HEMT can be linearized and described by a linear small-signal equivalent circuit with lumped elements, as shown in Figure 2-9. Each of the circuit elements has a physical origin in the device operation and their description is listed in Table 2-1. Figure 2-10 illustrates a more complete view of the small-signal equivalent circuit including parasitic pad capacitances and inductances.

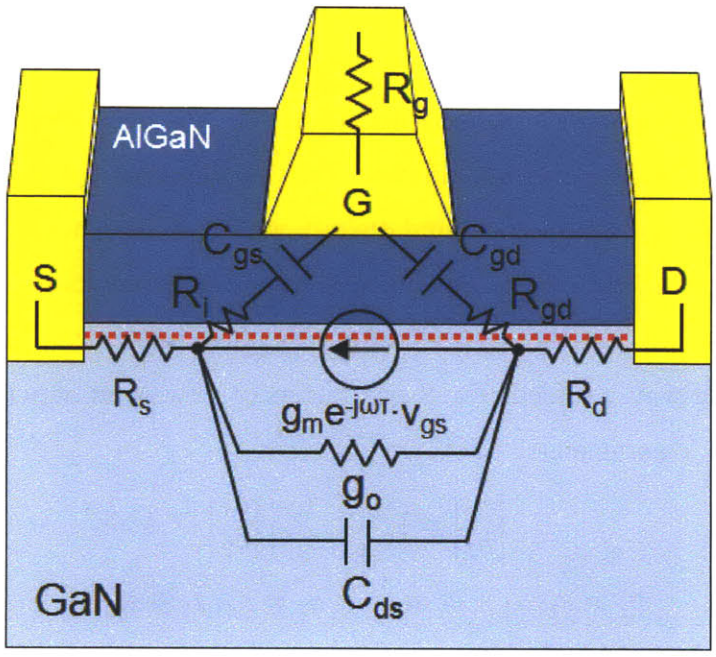


Figure 2-9. Small-signal equivalent circuit elements for a HEMT structure (e.g. AlGaN/GaN HEMT).

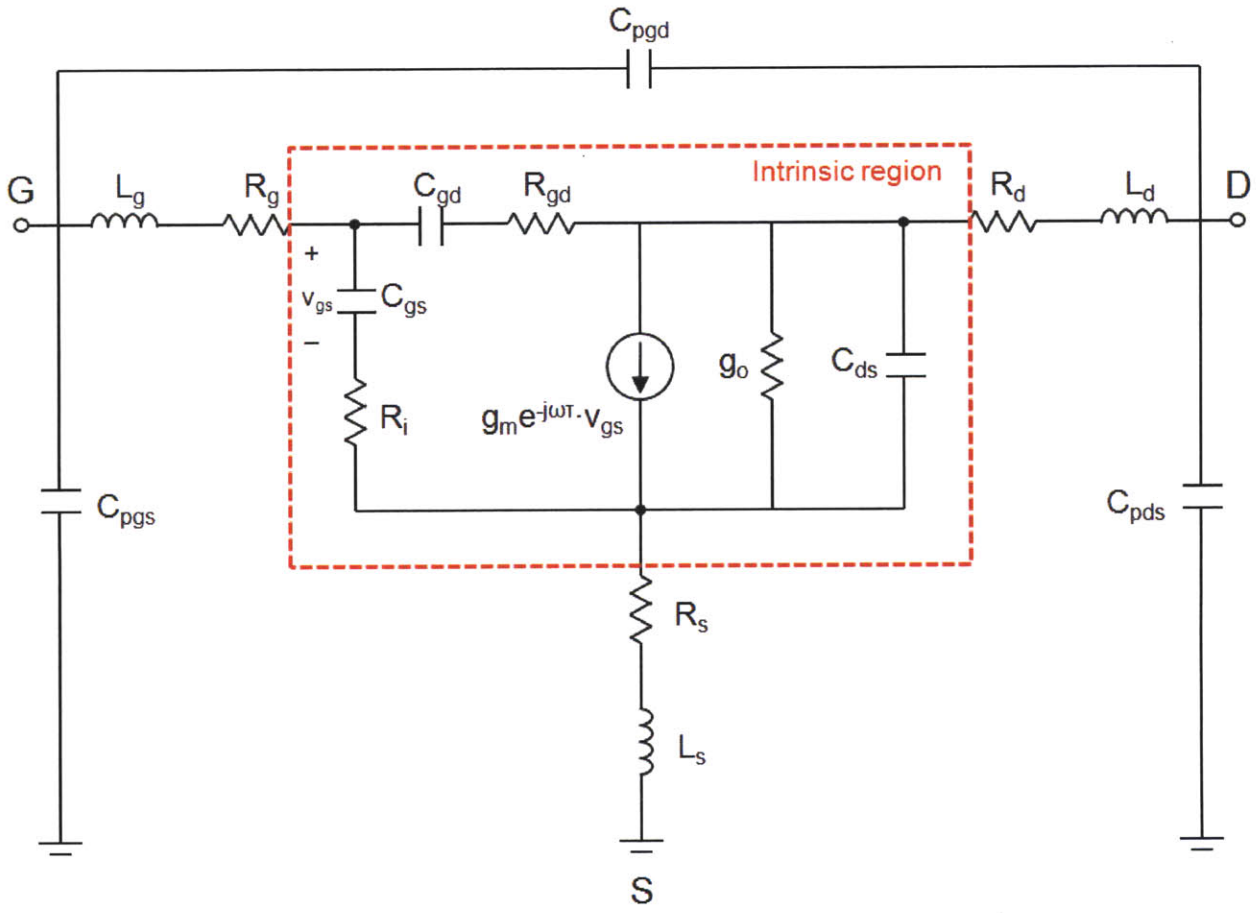


Figure 2-10. More comprehensive view of small-signal equivalent circuit of Figure 2-9. Extrinsic parasitic pad capacitances and inductances are included. Intrinsic region of the device is highlighted by the red dotted line.

		Term	Physical description
Intrinsic	C_{gs}	Gate-source capacitance	Gate charge modulation by changing V_{GS}
	C_{gd}	Gate-drain capacitance	Gate charge modulation by changing V_{DS}
	C_{ds}	Drain-source capacitance	Capacitance between drain and source (e.g. substrate capacitance)
	R_i	Input resistance	Lumped representation of distributed channel resistances
	R_{gd}	Gate-drain resistance	Complement of R_i to reflect symmetrical nature of the device
	g_m	Transconductance	Drain current gain with respect to the change of gate voltage
	τ	Transconductance delay	Time delay between change of gate voltage and drain current
	g_o	Output conductance	Variation of drain current by the change of drain voltage
Extrinsic	R_g	Gate resistance	Resistance of gate metal strip along the gate current flow
	R_d	Drain resistance	Resistance of drain access region and drain ohmic contact
	R_s	Source resistance	Resistance of source access region and source ohmic contact

Table 2-1. Small-signal equivalent circuit components and their physical descriptions.

In a HEMT, the drain current I_{DS} depends on both the gate-source voltage V_{GS} and the gate-drain voltage V_{DS}. The change in I_{DS} caused by small variations of V_{GS} and V_{DS} are defined as transconductance (g_m) and output conductance (g_o), respectively.

$$g_m = \left.\frac{dI_{DS}}{dV_{GS}}\right|_{V_{DS}=const.} \tag{2-20}$$

$$g_o = \left.\frac{dI_{DS}}{dV_{DS}}\right|_{V_{GS}=const.} \tag{2-21}$$

When the variation of V_{GS} is too fast, I_{DS} cannot be changed immediately since it takes time to charge or discharge the associated gate capacitance across the depletion region to produce the change in I_{DS}. Thus, there exists delay time τ between V_{GS} and I_{DS}. At such high frequencies the frequency-dependent transconductance is given by

$$g_m(\omega) = g_m e^{-j\omega\tau} \tag{2-22}$$

where $\omega = 2\pi f$ is the angular frequency, f is the operating frequency, and j is the imaginary unit. On the other hand, the time delay for g_o is negligible because the origin of g_o is not related to

charging and discharging of a capacitance, but is related to the leakage current between source and drain contacts due to high electric field or short-channel effects.

The gate-source capacitance C_{gs} and the gate-drain capacitance C_{gd} describe the change of the charge across the depletion region underneath the gate with varying V_{GS} and V_{DS}, respectively. Due to the charge neutrality condition, the charges induced across the depletion region (Q_{depl}) should be the same as the charges at the gate (Q_G) but with opposite polarity (Q_G = - Q_{depl}).

$$C_{gs} = \left.\frac{dQ_G}{dV_{GS}}\right|_{V_{DS}=const.} \tag{2-23}$$

$$C_{gd} = \left.\frac{dQ_G}{dV_{DS}}\right|_{V_{GS}=const.} \tag{2-24}$$

The drain-source capacitance C_{ds} represents the capacitance between the drain and the source other than its extrinsic pad capacitance (e.g. substrate capacitance) and can be written as

$$C_{ds} = \left.\frac{dQ_D}{dV_{DS}}\right|_{V_{GS}=const.} \tag{2-25}$$

where Q_D is the charge at the drain.

The input (or channel) resistance R_i is a lumped element representation of the distributed channel resistances and it is associated with the time to charge and discharge C_{gs}. The gate-drain resistance R_{gd} is a complement of R_i and reflects the symmetrical nature of the physical device in the equivalent circuit especially when V_{DS} is low (i.e. linear region).

The intrinsic elements C_{gs}, C_{gd}, C_{ds}, R_i, R_{gd}, g_m, τ, and g_o, and extrinsic resistances R_s and R_d are functions of the biasing conditions, whereas other extrinsic elements are independent of the biasing conditions. It is noted that the bias dependency of the extrinsic resistances R_s and R_d, called dynamic access resistances, is one of the distinct properties of AlGaN/GaN HEMTs whose effect needs to be considered in the small-signal equivalent circuit model [7].

Each element in the small-signal equivalent circuit can be correlated with the S-parameters (or their transformed Y-parameters) discussed in the previous section and this relationship will be

discussed in Chapter 4. By doing S-parameter measurements over a wide range of bias points, the entire set of small-signal equivalent circuit elements are obtained, and therefore, the complete device small-signal behavior can be characterized.

2.2.3. RF characteristics

To characterize the RF performance of AlGaN/GaN HEMTs we mainly focus on two of the most important figures of merit at high frequency, the unity current-gain cutoff frequency f_T and the unity power-gain cutoff frequency f_{max}. f_T is the frequency at which the magnitude of short-circuit current gain h_{21} equals unity (or 0 dB) and f_{max} is the frequency at which the unilateral power gain U equals unity (or 0 dB).

At the current-gain cutoff frequency f_T, the magnitude of current gain equals unity ($|h_{21}| = |i_2 / i_1| = 1$) with the output short-circuited. Figure 2-11 shows the corresponding intrinsic part of the small-signal equivalent circuit. The input current i_1 and the output current i_2 are given by

$$i_1 = \frac{V_{gs}}{\frac{1}{j\omega C_{gs}} + R_i} + \frac{V_{gs}}{\frac{1}{j\omega C_{gd}} + R_{gd}} \tag{2-26}$$

$$i_2 = g_m e^{-j\omega\tau} v_{gs} = g_m e^{-j\omega\tau} \frac{\frac{1}{j\omega C_{gs}}}{\frac{1}{j\omega C_{gs}} + R_i} V_{gs} \tag{2-27}$$

The magnitude of the short-circuit current gain $|h_{21}|$ is calculated as

$$|h_{21}| = \left|\frac{i_2}{i_1}\right| = \left|\frac{(1 + j\omega R_{gd} C_{gd}) g_m e^{-j\omega\tau}}{j\omega(C_{gs} + C_{gd}) - \omega^2 C_{gs} C_{gd}(R_i + R_{gd})}\right| \cong \frac{g_m}{2\pi f(C_{gs} + C_{gd})} \tag{2-28}$$

by considering that the term $\omega R_{gd} C_{gd}$ is typically much less than unity and $\omega(C_{gs} + C_{gd}) >> \omega^2 C_{gs} C_{gd}(R_i + R_{gd})$. Therefore, for $|h_{21}|=1$, the intrinsic current gain cutoff frequency f_T is

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \tag{2-29}$$

In short gate length devices, an alternative expression for f_T, with a more direct physical meaning, can be written by substituting $g_m = v_{sat} C = v_{sat}(C_{gs}+C_{gd}) / L_g$ from Eq. (2-17) (C is the gate capacitance per unit area while $(C_{gs} + C_{gd})$ is the gate capacitance per unit gate width):

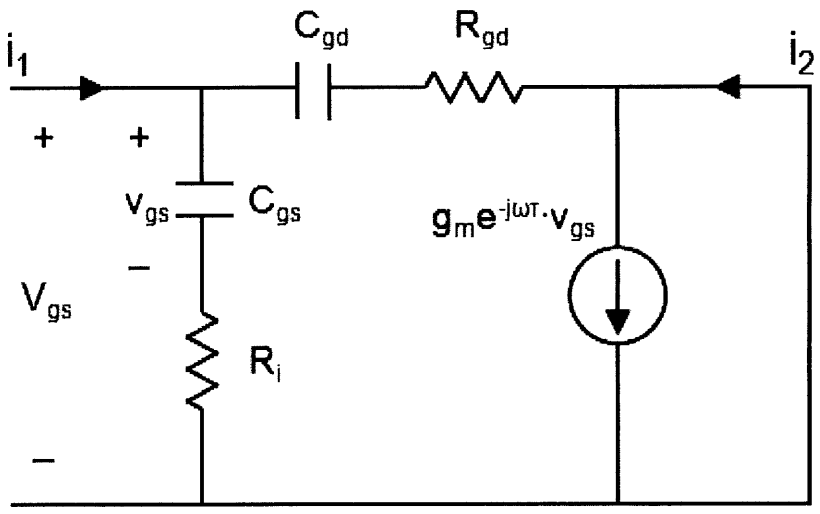


Figure 2-11. Simplified intrinsic part of the small-signal equivalent circuit used to derive the analytical expression of f_T.

$$f_T = \frac{v_{sat}}{2\pi L_g} \tag{2-30}$$

This expression shows that, to first order, the current-gain cutoff frequency can be maximized by reducing the gate length and increasing the electron velocity. However, this expression of f_T ignores all the extrinsic circuit elements such as R_s, R_d, and g_o. These extrinsic elements often limit the available f_T especially when the gate length is very small. Taking into account the extrinsic elements, a more rigorous form of f_T can be written as [8]:

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})(1 + (R_s + R_d)g_o) + g_m C_{gd}(R_s + R_d)} \tag{2-31}$$

For the power-gain cutoff frequency f_{max}, both current gain and voltage gain need to be considered. Figure 2-12 shows a simplified small-signal equivalent circuit commonly used to estimate f_{max}. The short circuit current gain $|h_{21}|$ and the voltage gain $|A_v|$ of this circuit are readily calculated as

$$|h_{21}| = \left|\frac{i_2}{i_1}\right| = \frac{g_m}{2\pi f C_{gs}} = \frac{f_T}{f} \tag{2-32}$$

$$|A_v| = \left|\frac{v_2}{v_1}\right| = \frac{g_m(r_o // R_L)}{\sqrt{1 + \omega^2 {C_{gs}}^2 (R_g + R_i)^2}} \cong \frac{g_m(r_o // R_L)}{\omega C_{gs}(R_g + R_i)} = \frac{f_T}{f}\frac{(r_o // R_L)}{(R_g + R_i)} \tag{2-33}$$

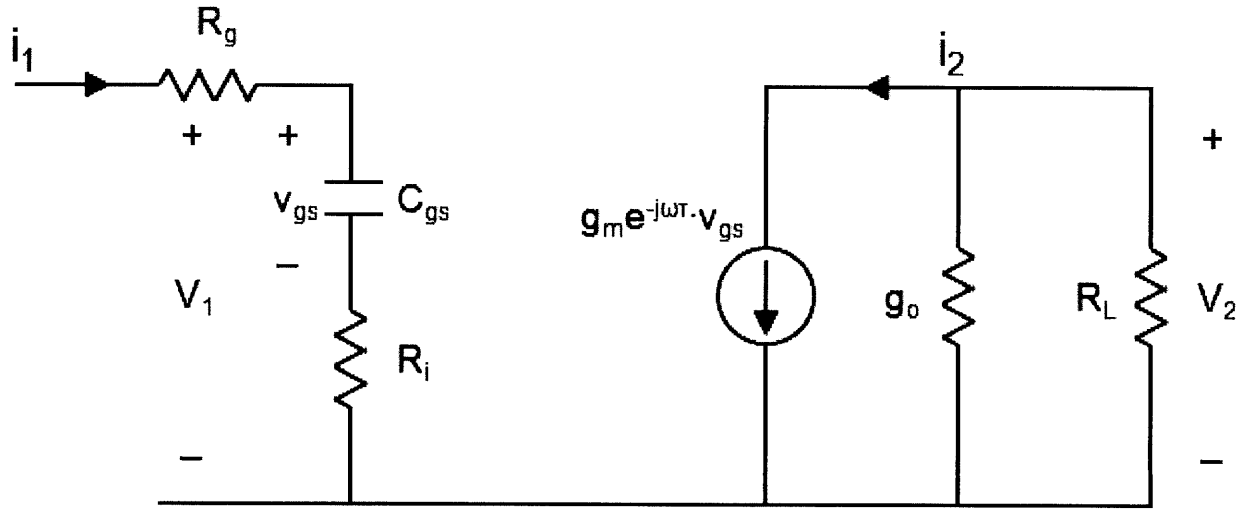


Figure 2-12. Simplified intrinsic part of the small-signal equivalent circuit used to derive the analytical expression of f_{max}. The gate resistance R_g and output resistance R_L are included due to their relevance.

where $r_o = 1/g_o$. We can assume $\omega^2 C_{gs}{}^2 (R_g + R_i)^2 >> 1$ for a device at high frequency. The maximum power gain is obtained at the matched load case, when $r_o = R_L$. In this case, half of the output current flows through the load resistor and the other half through the output resistance of the device. Then the power gain $|G_p|$ at the load becomes

$$|G_p| = \frac{|h_{21}|}{2}|A_v| = \left(\frac{f_T}{f}\right)^2 \frac{1}{4g_o(R_g + R_i)} \tag{2-34}$$

Thus, for $|G_p|$=1, the power-gain cutoff frequency f_{max} is

$$f_{max} = \frac{f_T}{2}\sqrt{\frac{1}{g_o(R_g + R_i)}} \tag{2-35}$$

The power gain cutoff frequency can be increased by improving f_T, reducing g_o (or increasing output resistance $r_o = 1/g_o$), and minimizing R_g and R_i. A more complete form of f_{max} is given in [9]:

$$f_{max} \cong \frac{f_T}{2\sqrt{(R_i + R_s + R_g)g_o + (2\pi f_T)R_g C_{gd}}} \tag{2-36}$$

2.3. Fabrication technology

Nitride semiconductors are widely used in many applications due to their superior electrical properties as well as excellent chemical and thermal stability. Such a high stability allows nitride semiconductors to be very attractive for applications in harsh environments, for instance high temperature and high pressure. However, because of its great stability, the device fabrication of nitride semiconductors is quite challenging. For example, etching nitride-based material is not straightforward. The lack of a reproducible wet etching technique forces the use of dry etching such as chlorine-based plasma reactive ion etching. However, dry etching is commonly prone to create surface damage and degrade electrical properties of nitride semiconductors. Forming good ohmic contacts on nitride semiconductors is also difficult. The ohmic contact resistance is very sensitive to the choice of metals (e.g. Ti, Al, Ni, Au, Mo, etc), their stacking order, and even thickness. Patterning submicron gates by e-beam lithography on nitride semiconductors is another difficulty due to the insulating nature of these materials and the large atomic mass of gallium, which increases electron back-scattering.

In order to obtain a good device performance, it is imperative to have good device isolation, low ohmic contact resistances, and reliable submicron gates. They are the most basic processing steps to ensure reliable device operation and Figure 2-13 summarizes a standard process flow for AlGaN/GaN HEMT fabrication. More advanced technologies to improve high frequency performance of AlGaN/GaN HEMTs such as recessed ohmic, vertical gate-recess, oxygen plasma treatment, lateral gate-etch technique will be introduced in Chapter 3 and Chapter 4.

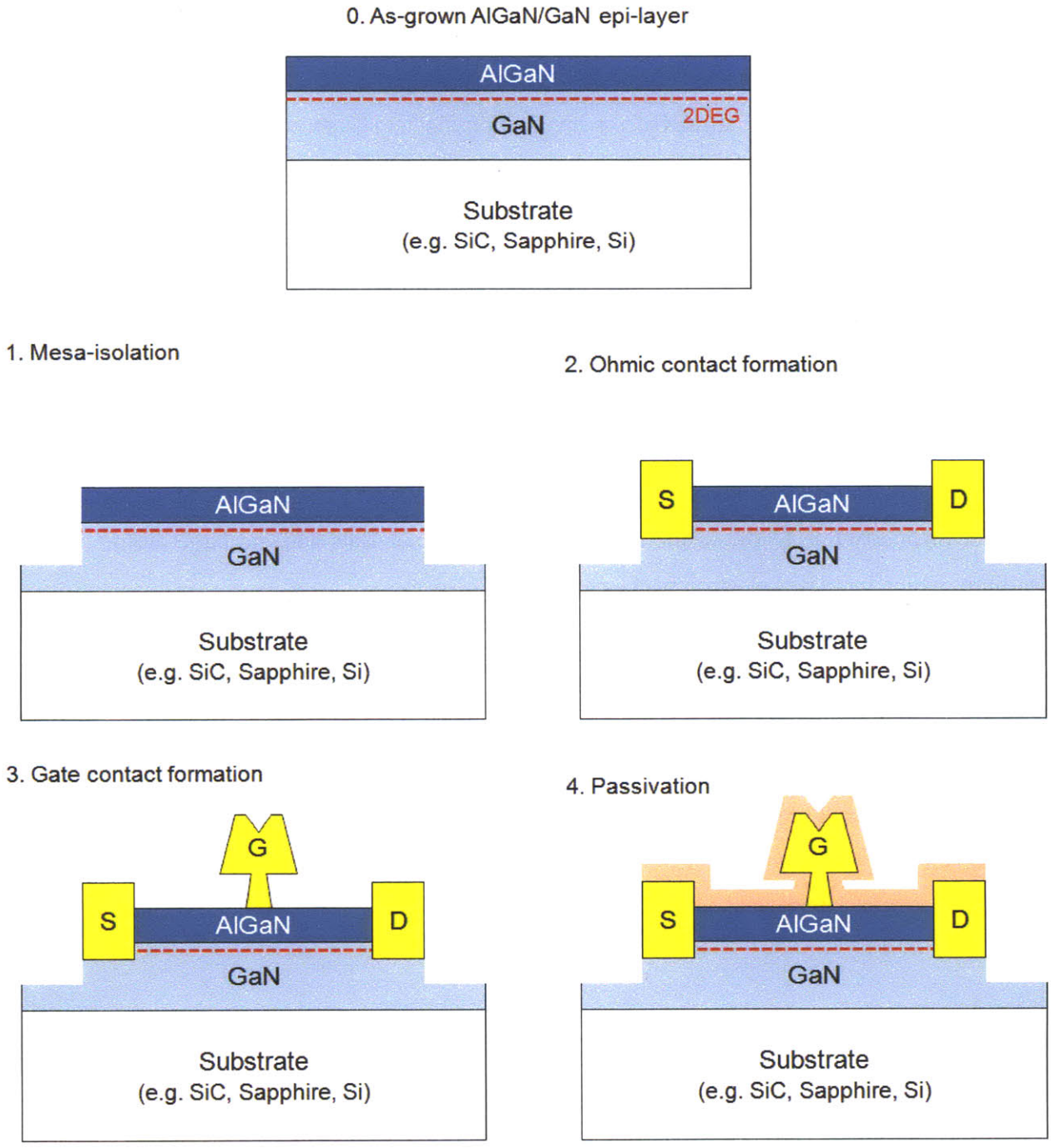


Figure 2-13. Basic process flow of AlGaN/GaN HEMT fabrication.

2.3.1. Isolation

In general, AlGaN and GaN layers are epitaxially grown over the entire wafer and the resulting AlGaN/GaN heterostructure forms a 2DEG everywhere in the wafer. Isolation is the process to electrically isolate one device from the other as shown in Figure 2-14. Poorly isolated transistors show high leakage currents, which causes soft pinch-off and poor gate modulation during device operation. For complete isolation, the AlGaN and GaN epitaxial layers need to be etched down to fully remove the 2DEG. This results in a mesa structure with a typical step height about 100 ~ 150 nm. It should be noted that although device isolation can also be achieved by ion implantation [10], in this work we used mesa-type isolation due to its simplicity.

Owing to the chemical inertness of AlGaN and GaN, no suitable wet etching method for the mesa-isolation exists currently. Thus, we use dry etching based on electron cyclotron resonance reactive ion etching (ECR-RIE) with Cl_2/BCl_3 gas mixture, which is one of the most popular ways to etch GaN-based materials. A photoresist is used as an etch mask to selectively etch AlGaN and GaN at the desired locations. Several processing parameters such as ECR/RF powers, chamber pressure, and temperature need to be optimized to achieve a reliable dry etching and mesa-isolation. Figure 2-15 summarizes the dry etching characteristics as a function of the different processing parameters. Too much ECR power (plasma density) or excessive use of RF power (chuck bias) often cause severe problems in removing the photoresist mask after the etch

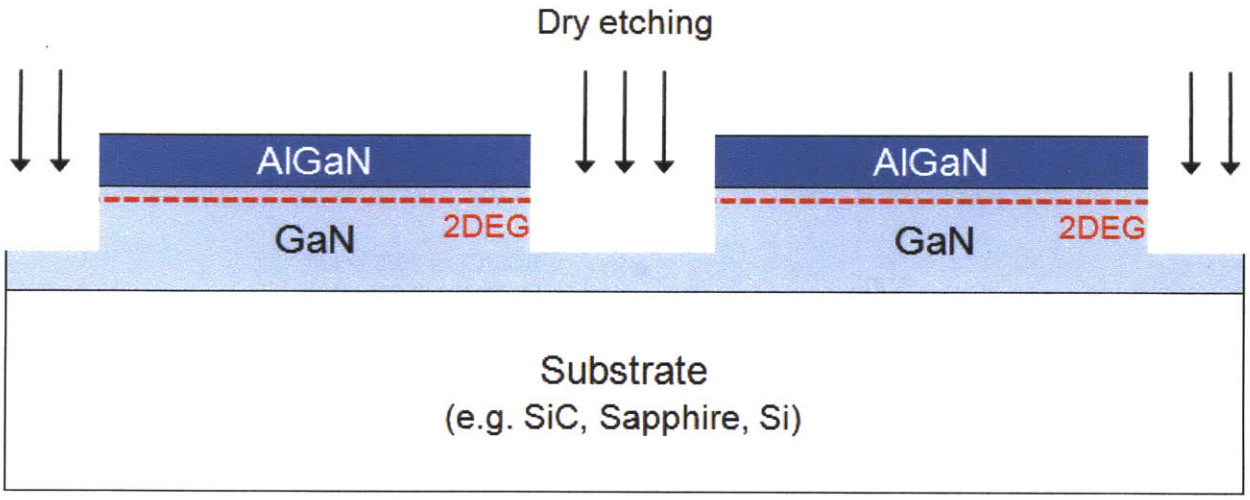


Figure 2-14. Schematic illustration of mesa-isolation. The mesa-isolation is achieved by ECR-RIE dry etching with Cl_2/BCl_3 gas mixture and typical depth for the electrical isolation is 100 ~ 150 nm.

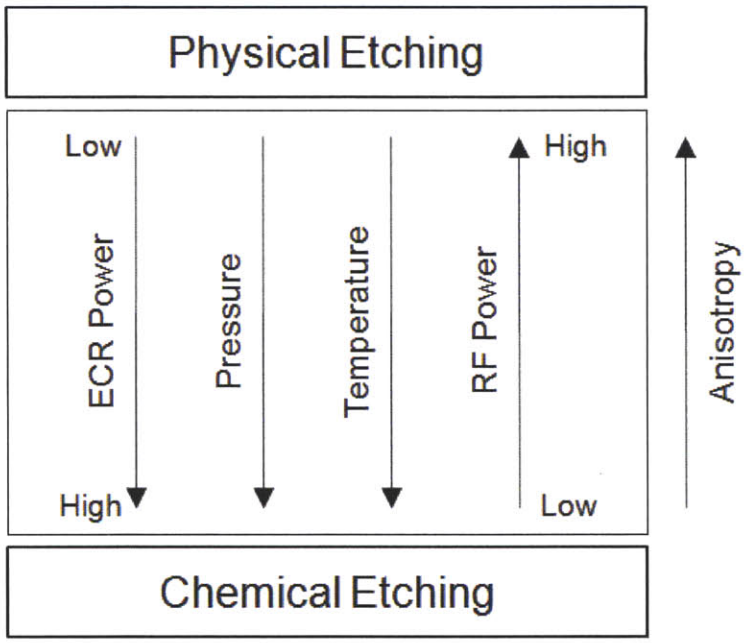


Figure 2-15. Trends of dry-etching characteristics depending on several different processing parameters.

due to a change in its chemical properties. Therefore, a trade-off between physical and chemical etching is required for reliable mesa-isolation process. Figure 2-16 shows the recipes used throughout the thesis work for the mesa-isolation.

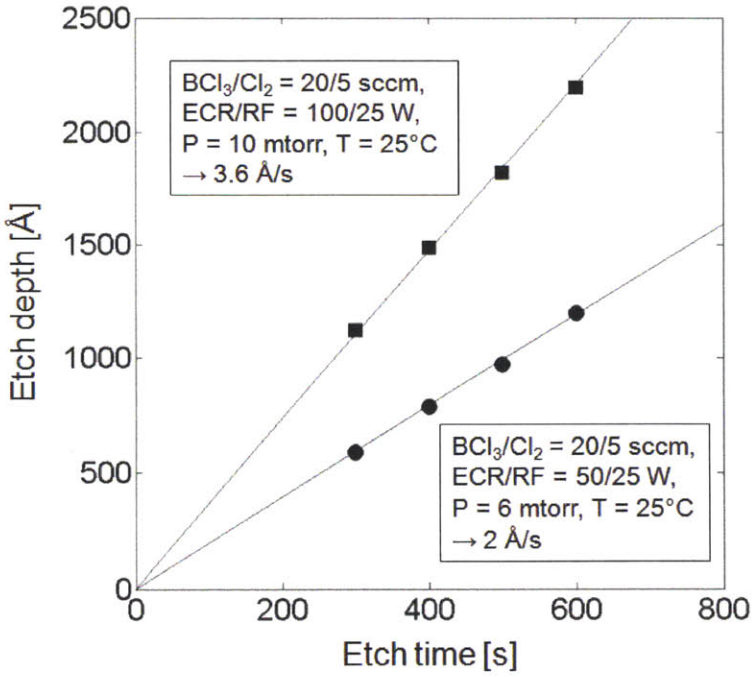


Figure 2-16. Dry-etching recipe developed in this work for reliable mesa-isolation. Process conditions are carefully optimized considering trade-offs in Figure 2-15.

2.3.2. Ohmic contact

The ohmic metal-semiconductor contact resistance is a key parameter that largely affects many device characteristics. Some of them include knee voltage, maximum drain current, extrinsic transconductance, f_T and f_{max}. The ohmic contact resistance must be minimized to improve these characteristics. It is also important to have ohmic contacts with low surface roughness and sharp edge acuity to facilitate the gate lithography process, especially for scaled devices (e.g. devices with source-to-drain distance of less than 2 μm). In addition, high thermal stability is often required to ensure high voltage and high temperature operations.

The ohmic contact, providing a linear I-V characteristic in both forward and reverse biasing directions, is formed by increasing the probability of carrier transfer across the Schottky barrier at the metal-semiconductor interface. It can be achieved by engineering the Schottky barrier, for instance by lowering the height or reducing its width to increase probability of tunneling through the barrier. In AlGaN/GaN heterostructure the Schottky barrier height is often fixed due to the Fermi-level pinning, on the other hand the barrier width is determined by the level of doping inside the AlGaN layer. Thus, the ohmic contact can be obtained by engineering the barrier width, for example, by alloying specific metals at the AlGaN surface as shown in Figure 2-17. Among the many contact metallization schemes reported in the literature, the Ti/Al-based contact with rapid thermal annealing (RTA) is the most widely studied. In this work, we use the Ti/Al/Ni/Au ohmic contacts (metal sequence from the bottom) annealed by RTA. Some advanced technologies such as ohmic recess and Si/Ge insertion will be discussed in Chapter 3.

Ti is known as a key metal for the formation of ohmic contacts in AlGaN/GaN heterostructure [11]. When annealed at high temperatures (800 ~ 900 °C), it forms TiN at the Ti/AlGaN interface, introducing N-vacancies in the semiconductor, which are known to act as n-type doping. The resultant n-type doped AlGaN and the conductive TiN facilitate tunneling mechanism of carriers at the interface. Also, Ti has good adhesion to nitride semiconductors and plays an additional role to getter oxide impurity from the AlGaN [12]. Thus, when it is coupled with high conductivity metals such as Al, alloyed Ti/Al can form a good ohmic contact to AlGaN/GaN heterostructure. However, since both Ti and Al are easily oxidized, particularly

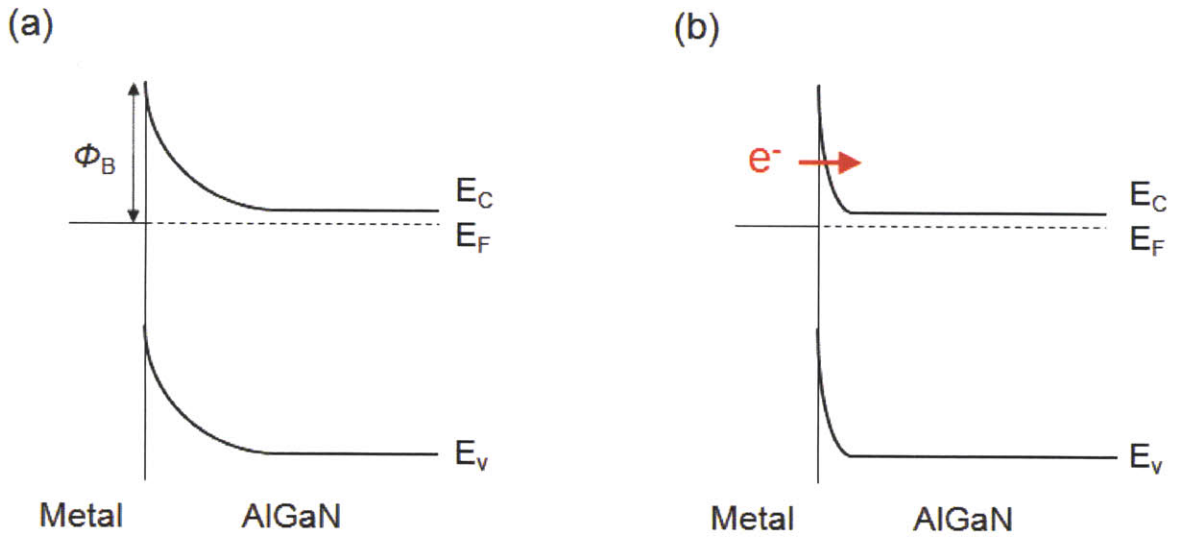


Figure 2-17. (a) As-deposited metals on AlGaN surface typically forms a Schottky contact. (b) Alloying metals such as Ti/Al reduces the barrier width and increases tunneling probability of electrons forming the ohmic contact.

during the RTA, a Ni/Au bilayer cap is often used to prevent oxidation. Owing to its excellent thermal stability, Ni acts as a barrier during the RTA against the out-diffusion and in-diffusion of Ti/Al and Au, respectively. The final Au overcoat helps to prevent oxidation while maintaining a high electrical conductivity.

The optimization of the Ti/Al/Ni/Au ohmic contact technology (contact resistance, surface roughness, edge acuity, etc.) requires a systematic study to find out the best combination of metal thicknesses and RTA conditions. In the last couple of years, many samples have been processed at MIT and the Ti/Al/Ni/Au (200/1000/250/500 Å) annealed at 870 °C for 30 s in N_2 ambient was chosen as the standard metallization scheme for AlGaN/GaN HEMTs. This metallization scheme routinely gives contact resistance of 0.4 ~ 0.5 Ω·mm (Figure 2-18) and an acceptable surface roughness and edge acuity for the following normal gate process.

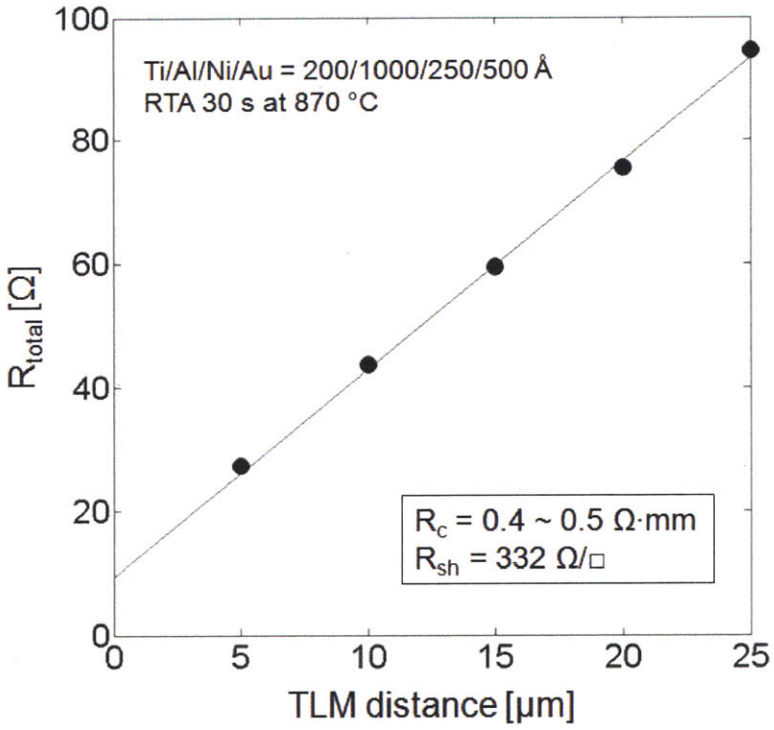


Figure 2-18. Transfer length method (TLM) measurement of typical ohmic contacts on AlGaN/GaN HEMTs. R_c of 0.4 ~ 0.5 Ω·mm is routinely obtained using this scheme for the AlGaN/GaN samples from different vendors.

2.3.3. Gate contact

The gate process is one of the most important steps especially for high frequency device operation. For example, a key figure of merit in RF characteristics is f_T and it heavily depends on the gate properties as shown in Eq. (2-29), namely current gain by the gate modulation (g_m) and the capacitances by the gate contact (C_{gs} and C_{gd}). Also, f_{max} can be greatly improved by reducing the gate resistance as shown in Eq. (2-36). Because of its importance, the gate process is often the most demanding and elaborate step in the entire device process. Typically, the gates are fabricated with a T-shape cross section to reduce the gate length and the gate resistance at the same time. In this section, we introduce the basic deep-submicron T-gate technology used throughout this thesis work.

After ohmic contacts are formed, deep-submicron T-gates (gate length of less than 100 nm) are fabricated between the source and drain ohmic contacts by using a Raith 150 30 keV e-beam

lithography system. We used a tri-layer resist of ZEP/PMGI/ZEP with a two-step e-beam exposure process and the process flow is summarized in Figure 2-19. Figure 2-20 shows a successfully fabricated T-gate with a gate length of 80 nm. While there exists a bi-layer process that allows the fabrication of the T-shape gate with a single e-beam exposure, the tri-layer scheme offers a much more reproducible lift-off process and guarantees smaller gate lengths due to less e-beam broadening and forward scattering. It is also noted that it is possible to use a PMMA/Copolymer/PMMA resist stack, however we chose the ZEP-based tri-layer mainly because of its 2 ~ 3 times better dry-etching stability, which will be discussed in Chapter 3.

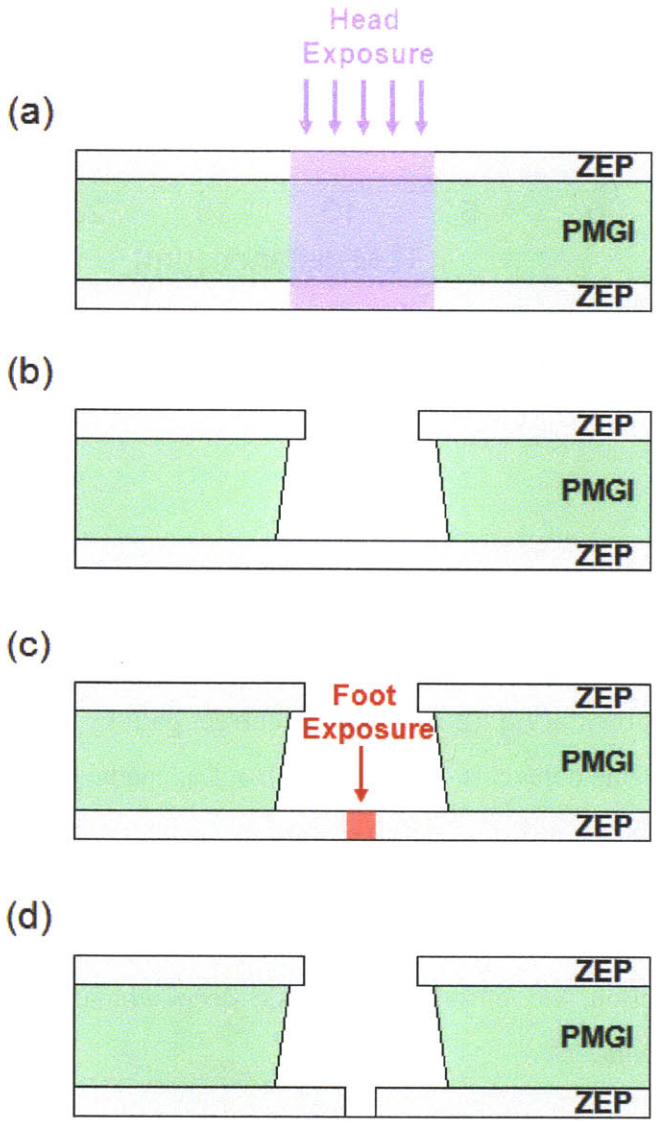


Figure 2-19. Conventional e-beam lithography process for submicron T-shape gate. (a) head exposure, (b) development of top ZEP and middle PMGI in sequence for the head, (c) foot exposure, and (d) development of bottom ZEP for the foot.

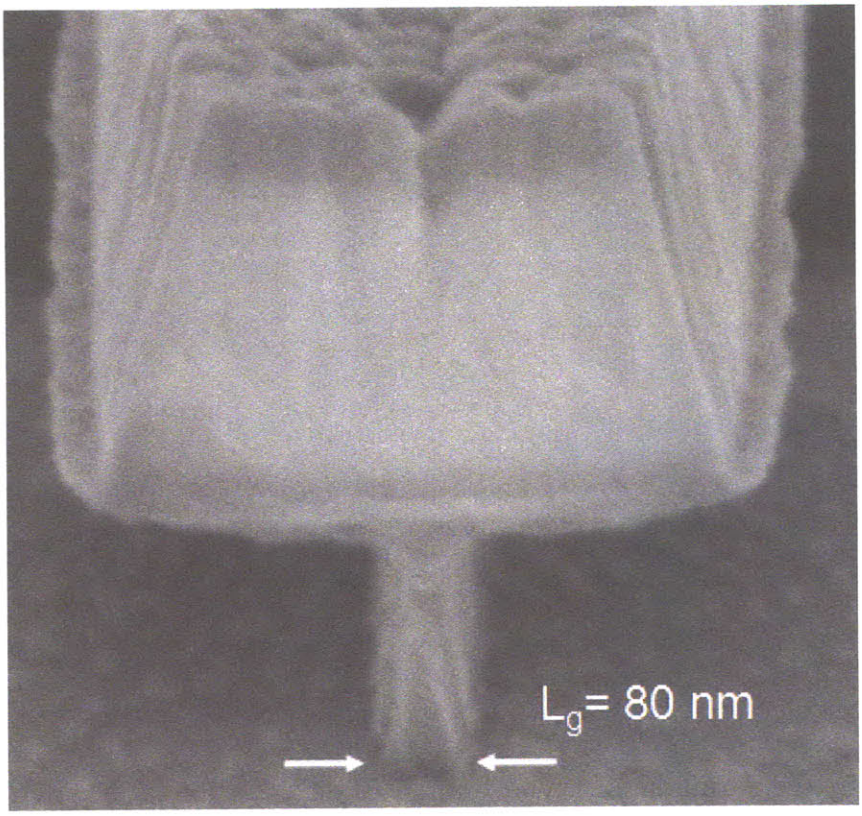


Figure 2-20. Tilted cross-section SEM image of fabricated 80 nm T-gate following the steps in Figure 2-19.

The exposure dose and development conditions were optimized through a dose matrix experiment with time-varying development. After the first exposure (head exposure), a mixture of methyl-isobutyl-ketone (MIBK) and methyl ethyl ketone (MEK) was used for 90 s to develop the top ZEP. After rinsing with isopropanol (IPA), the sample was dipped in CD26 developer for 30 s to develop the middle PMGI layer. The excellent selectivity of this developer avoids the development of the bottom ZEP. After the second exposure (foot exposure), a mixture of MIBK and IPA was used for 90 s at 0 ~ 5 °C (cold development) to develop the bottom ZEP. Once that the T-shape pattern is developed, we apply a simple edge trimming process before depositing the gate metals to improve the process yield and to further reduce the gate length (< 60 nm). The simple edge trimming step developed in this work effectively improves process yield and enables to fabricate T-gates with shorter gate lengths. As shown in Figure 2-21, evaporated gate metal fills the foot part of the T-shape pattern with a triangular shape and a disconnection between the foot and head part of the T-gate can occur when the footprint becomes too narrow (i.e. L_g is too small). To fix this problem, a low power ECR-RIE with Cl_2/BCl_3 gas mixture at high pressure

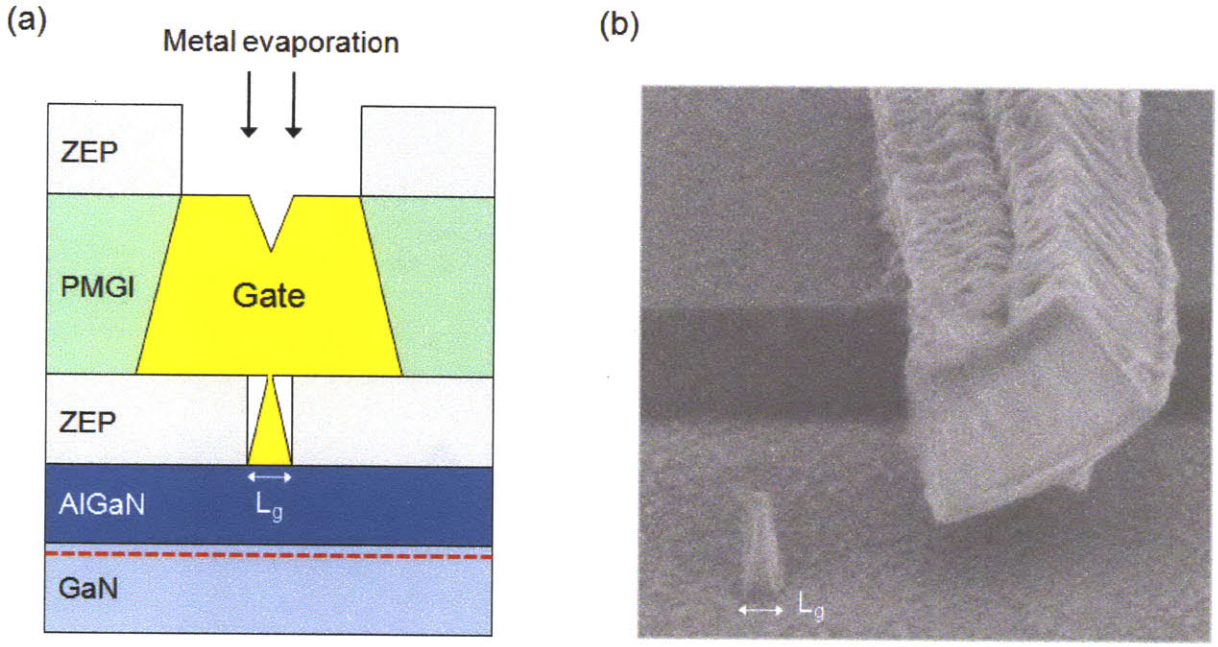


Figure 2-21. (a) Schematic illustration of how evaporated gate metal fills the T-shape pattern. If the foot pattern is too small, evaporated metal has difficulties in fully fill up the foot space. (b) Tilted cross-section SEM after removing e-beam resists for lift-off. The gate falls down and the device yield gets poor.

was used to selectively trim the edge part of the ZEP layer on top of the AlGaN surface. This process changes the sharp corner of the bottom ZEP to a rounded shape as described in Figure 2-22(a) and makes it easier for the gate metals to fill up the foot space.

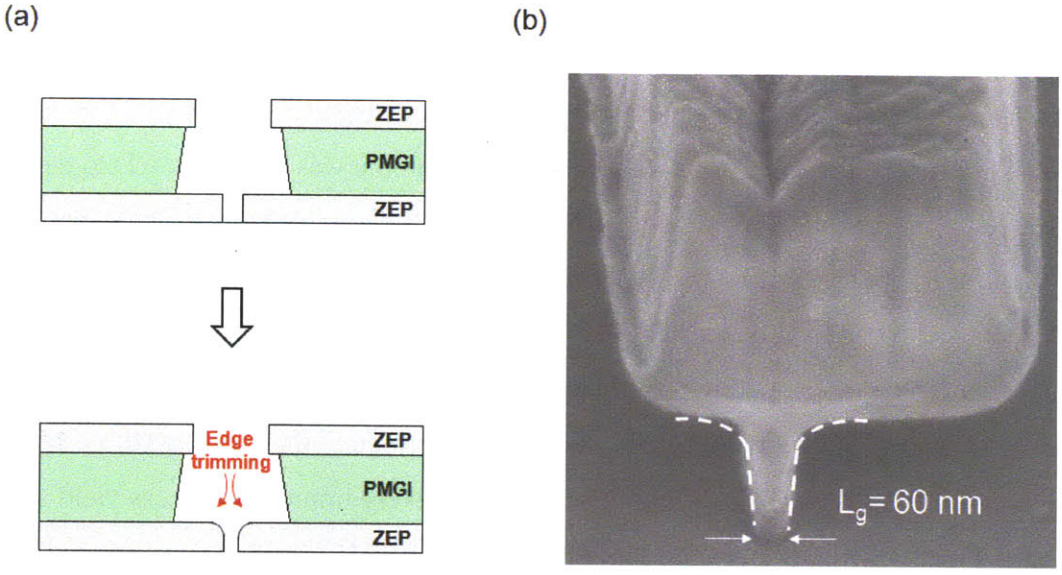


Figure 2-22. (a) Improved gate process by adding simple edge trimming step before depositing the gate metals. (b) Tilted cross-section SEM image of the resultant 60-nm gate length T-gate. This simple technique largely improves process yield and minimum gate length that can be achieved.

Finally, a Ni/Au (200/2300 Å) multilayer is deposited as the Schottky gate contact. Since Ni has a high work function (~ 5.2 eV) and moderate resistivity (7×10^{-8} Ω·m), Ni is expected to set a high Schottky barrier at the Ni/AlGaN interface (unless there is Fermi-level pinning) which reduces gate leakage current during device operation. A thick Au cap layer provides low resistivity (2.4×10^{-8} Ω·m) to the gate electrode. Figure 2-22(b) shows a tilted cross-section SEM image of a successfully fabricated 60 nm T-gate. Using this technology, it is possible to fabricate gate lengths below 60 nm, however 60 nm gate length was the shortest gate length that can be reproducibly achieved.

2.3.4. Passivation

We already discussed the presence of surface states at the AlGaN surface in section 2.1. It is widely known for AlGaN/GaN HEMTs that the surface states between the gate and the drain can trap electrons from the gate metal under high voltage via tunneling and hopping mechanism [13]. Under high frequency operation of the device, these trapped electrons cannot be de-trapped from the states and deplete the channel electrons, resulting in a reduction of the drain current. This phenomenon is called "current collapse" and it is a major concern limiting microwave output power of GaN devices [14].

One of the technological solutions to reduce the current collapse is to passivate the surface states with Si_xN_y or SiO_2 layers in order to decrease the surface state density or make them inaccessible to electrons that tunnel from the gate. However, a Si_xN_y or SiO_2 passivation layer thicker than 100 nm is typically required to fully passivate the device. Such a thick layer inevitably introduces huge amount of gate fringing capacitances that largely degrade the high frequency performance. To reduce these capacitances, our group has recently developed a new passivation scheme based on 25-nm-thick Al_2O_3 deposited by atomic layer deposition (ALD) [15][16]. According to the simulation, the gate capacitance is 20% lower in the new Al_2O_3 passivation than in the conventional Si_3N_4 passivation due to the lower thickness required for the Al_2O_3 layer (25 nm vs. 100 nm) while there is no big difference in the dielectric constant ($\varepsilon_r(Al_2O_3)$ = 10 vs. $\varepsilon_r(Si_3N_4)$ = 7.5). Thin Al_2O_3 passivation layers are therefore promising candidates for high frequency devices where the gate capacitances need to be minimized.

Finally, in order to complete the device fabrication and be able to measure electrical characteristics of the transistor, the gate, source, and drain contacts should be exposed after the passivation. An ECR-RIE dry etching with Cl_2/BCl_3 gas mixture can be used to etch the Al_2O_3 in the contact regions using a normal positive photoresist as a mask.

Although we have developed the new Al_2O_3 passivation technology which is good for device reliability and less detrimental to its high frequency performance than conventional Si_3N_4 passivation, we do not apply it to our devices throughout this work. Since passivation step is generally very sensitive to the process conditions such as surface pre-treatment, film thickness, film stress, temperature etc., the reproducibility of breakdown voltage, gate leakage current, and effectiveness of dispersion elimination are undependable after the passivation [17]. In order to identify the main problems and explore ultimate potential of high frequency performance of AlGaN/GaN HEMTs, reproducibility and repeatability of the devices are imperative. Therefore, unless otherwise specified we exclude passivation step in the process of AlGaN/GaN HEMTs which increases uncertainties in their analysis.

2.4. References

[1] T. Mimura, S. Hiyamizu, T. Fujii, and K. Nanbu, "A new field-effect transistor with selectively doped GaAs/n-$Al_xGa_{1-x}As$ heterojunctions," *Jpn. J. Appl. Phys.*, vol. 19, no. 5, pp. L225-L227, May. 1980.

[2] T. Mimura, "The early history of the high electron mobility transistor (HEMT)," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 3, pp. 780-782, Mar. 2002.

[3] J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck, and U. K. Mishra, "Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors," *Appl. Phys. Lett.*, vol. 77, no. 2, pp. 250-252, Jul. 2000.

[4] U. K. Mishra and J. Singh, "Semiconductor device physics and design," Springer, 2008.

[5] R. J. Trew, M. W. Shin, and V. Gatto, "High power applications for GaN-based devices," *Solid-State Electronics*, vol. 41, no. 10, pp. 1561-1567, Oct. 1997.

[6] B. Gelmont, K. Kim, and M. Shur, "Monte Carlo simulation of electron transport in gallium nitride," *J. Appl. Phys.*, vol. 74, no. 3, pp. 1818-1821, Aug. 1993.

[7] T. Palacios, S. Rajan, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "Influence of the dynamic access resistance in the g_m and f_T linearity of AlGaN/GaN HEMTs," *IEEE Trans. Electron Devices*, vol. 52, no. 10, pp. 2117-2123, Oct. 2005.

[8] P. J. Tasker and B. Hughes, "Importance of source and drain resistance to the maximum fT of millimeter-wave MODFET's," *IEEE Electron Device Lett.*, vol. 10, no. 7, pp. 291-293, Jul. 1989.

[9] F. Schwierz and J. J. Liou, "Modern microwave transistors: theory, design, and performance," Wiley, 2003.

[10] J.-Y. Shiu, J.-C. Huang, V. Desmaris, C.-T. Chang, C-Y. Lu, K. Kumakura, T. Makimoto, H. Zirath, N. Rorsman, and E. Y. Chang, "Oxygen ion implantation isolation planar process for AlGaN/GaN HEMTs," *IEEE Electron Device Lett.*, vol. 28, no. 6, pp. 476-478, June. 2007.

[11] F. M. Mohammed, L. Wang, I. Adesida, and E. Piner, "The role of barrier layer on ohmic performance of Ti/Al-based contact metallizations on AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 100, no. 2, 023708, Jul. 2006.

[12] J. Yun, K. Choi, K. Mathur, V. Kuryatkov, B. Borisov, G. Kipshidze, S. Nikishin, and H. Temkin, "Low-resistance ohmic contacts to digital alloys of n-AlGaN/AlN," *IEEE Electron Device Lett.*, vol. 27, no. 1, pp. 22-24, Jan. 2006.

[13] C. Roff, J. Benedikt, P. J. Tasker, D. J. Wallis, K. P. Hilton, J. O. Maclean, D. G. Hayes, M. J. Uren, and T. Marin, "Analysis of DC-RF dispersion in AlGaN/GaN HFETs using RF waveform engineering," *IEEE Trans. Electron Devices*, vol. 56, no. 1, pp. 13-19, Jan. 2009.

[14] R. Vetury, N. Q. Zhang, S. Keller, and U. K. Mishra, "The impact of surface states on the DC and RF characteristics of AlGaN/GaN HFETs," *IEEE Trans. Electron Devices*, vol. 48, no. 3, pp. 560-566, Mar. 2001.

[15] O. I. Saadat, J. W. Chung, E. L. Piner, and T. Palacios, "Gate-first AlGaN/GaN HEMT technology for high-frequency applications," *IEEE Electron Device Lett.*, vol. 30, no. 12, pp. 1254-1256, Dec. 2009.

[16] J. W. Chung, O. I. Saadat, J. M. Tirado, X. Gao, S. Guo, and T. Palacios, "Gate-Recessed InAlN/GaN HEMTs on SiC Substrate with Al_2O_3 Passivation", *IEEE Electron Device Lett.*, vol. 30, no. 9, pp. 904-906, Sept. 2009.

[17] L. Shen, R. Coffie, D. Buttari, S. Heikman, A. Chakraborty, A. Chini, S. Keller, S. P. DenBaars, and U. K. Mishra, "High-power polarization-engineered GaN/AlGaN/GaN HEMTs without surface passivation," *IEEE Electron Device Lett.*, vol. 25, no. 1, pp. 7-9, Jan. 2004.

Chapter 3. Optimizing Parasitic Elements for Improving f_{max}

3.1. Introduction

An important function of a transistor is to amplify electronic signals such as current and voltage, and provide increased amounts of power to the output. The power gain of the transistor is defined as the ratio of an output power to an input power. Such a power gain is typically frequency dependent and its magnitude rolls off at high frequencies at a slope of -20 dB/decade for any transistor [1]. One of the primary goals in RF power amplifier is to achieve high power gain at high frequencies, namely to increase the maximum power-gain cutoff frequency (f_{max}). f_{max} is the maximum frequency at which the transistor still provides the power gain (≥ 1), and can be expressed as shown in Chapter 2:

$$f_{max} \cong \frac{f_T}{2\sqrt{(R_i + R_s + R_g)g_o + (2\pi f_T)R_g C_{gd}}} \tag{3-1}$$

where f_T is the current-gain cutoff frequency, C_{gd} is the gate-drain capacitance, and g_o is the output conductance, while R_i, R_s, and R_g represent the input, source, and gate resistances, respectively. In spite of the great progress in high frequency performance of GaN HEMTs achieved during the last few years [2], f_{max} still needs to be increased for them to be suitable for millimeter-wave applications (30 ~ 300 GHz).

To maximize f_{max}, each parameter in Eq. (3-1) needs to be carefully optimized. f_T, for example, has been extensively studied in AlGaN/GaN HEMTs and several groups reported largely improved f_T (> 150 GHz) by reducing the gate length down to 30 nm [3]. However, despite its great importance, other parasitic elements have barely been investigated. As shown in Eq. (3-1), f_{max} is not only a function of f_T, but also a strong function of several parasitic elements such as R_i, R_s, R_g, C_{gd}, and g_o. In this chapter, we describe new device design and fabrication technologies to minimize these parasitic elements to achieve state-of-the-art f_{max}.

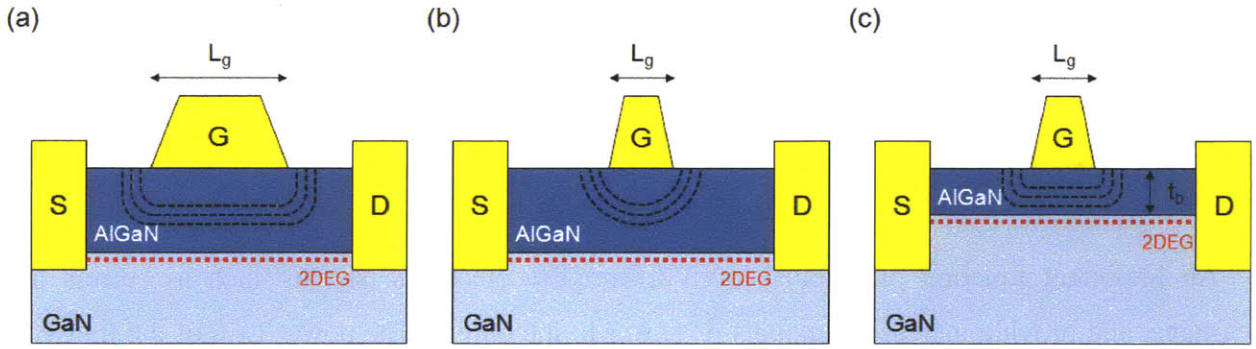


Figure 3-1. Schematic illustratioin of potential distribution in (a) long gate length and (b) short gate length devices. In this diagram, the source and drain are assumed to be short-circuited. At short gate lengths, the potential lines no longer resemble that of a planar capacitor. (c) By scaling both L_g and t_b, the original potential distribution can be re-created.

In HEMTs, the simple theory of their operation as described in Chapter 2 must be modified when the gate length is small. The short-channel effects (i.e. effects that arise at very short gate lengths) are mainly due to a two-dimensional field distribution and high electric fields in the channel region [4]. In Chapter 2, we assumed the gradual channel approximation which indicates the lateral field (E_x) in the channel underneath the gate is much smaller than the vertical field (E_y) ($E_x << E_y$). In other words, the channel field varies very little along the gate length (L_g) over a distance of the barrier thickness (t_b) which requires $t_b << L_g$. However, when L_g is very small, the gradual channel approximation is no longer valid and the field distribution in the channel becomes two-dimensional with both E_x and E_y components. The short-channel effects often complicate device characterization and degrade device performance by causing threshold-voltage shift, poor pinch-off, increased output conductance, and reduced transconductance. These effects can be suppressed by scaling the barrier thickness t_b along with L_g satisfying $t_b << L_g$ and increasing gate aspect ratio L_g / t_b as shown in Figure 3-1.

The short-channel effects also play an important role in the high-frequency characteristics of HEMTs [5]. For example, increased output conductance (g_o) largely degrades f_{max} (Eq. (3-1)). Nevertheless, improvement of short-channel effects has been seldom a path to improve the high frequency performance of GaN HEMTs. In this work, we applied a low damage gate recess technology to effectively suppress the short-channel effects by improving the aspect ratio (L_g / t_b). The resultant L_g = 60 nm HEMT showed excellent output characteristics with a very low g_o of

10.4 mS/mm and good pinch-off behavior. Also, the closer gate-to-channel distance helped reduce R_i, which determines the charging time of the gate-source capacitance (C_{gs}).

The parasitic resistances and capacitances are also very important in that they can significantly degrade the high frequency performance [6][7]. To minimize parasitic source resistance R_s, we combined a short source-to-drain distance and recessed source/drain ohmic contacts [8]. A new Si/Ge-based recessed ohmic contact technology was also introduced to improve surface roughness over the conventional ohmic contacts maintaining low contact resistances. The reduced surface roughness allowed the reduction of the source-to-drain distance down to ~1 μm. The reduced geometry, in combination with an optimized ohmic contact resistance of less than 0.2 Ω·mm resulted in devices with a very low on-resistance of 1.1 ~ 1.2 Ω·mm and a knee voltage of just 2 V. In addition, an optimized T-shaped gate was fabricated to reduce both R_g and C_{gd} by maximizing head size while minimizing the foot print of the T-gate.

Finally, by integrating all these technologies a recessed 60-nm gate length AlGaN/GaN HEMT exhibited a state-of-the-art f_{max} of 300 GHz which shows the great importance of engineering parasitic elements to improve f_{max}.

3.2. Optimizing parasitic elements

In this section, several processing technologies are developed and optimized to minimize the parasitic elements R_i, R_s, R_g, C_{gd}, and g_o in AlGaN/GaN HEMTs. Each element was understood by its physical origin and appropriate technologies were developed to minimize its detrimental contribution to the f_{max} in AlGaN/GaN HEMTs. Finally, process integration was accomplished by considering trade-offs among different technologies developed in this work.

3.2.1. Source resistance (R_s)

Figure 3-2 shows a cross-section schematic diagram of an AlGaN/GaN HEMT with R_s and R_d identified. Each of these resistances consists of two components: channel resistance of the access

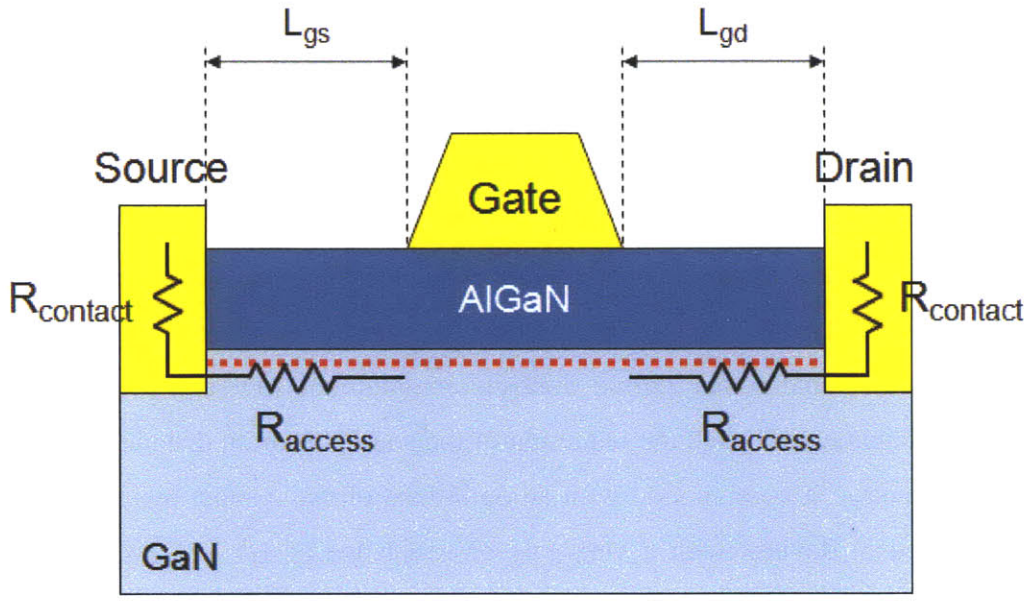


Figure 3-2. A cross-section schematic diagram of an AlGaN/GaN HEMT with source and drain resistances (R_s and R_d). Each of R_s and R_d consists of ohmic contact resistance ($R_{contact}$) and access resistance (R_{access}).

region (of length L_{gs} or L_{gd}) and ohmic contact resistance. For example, to first order, the source resistance R_s can be expressed as

$$R_s = R_{access} + R_{contact} = R_{sh}\left(\frac{L_{gs}}{W_g}\right) + R_c\left(\frac{1}{W_g}\right) \tag{3-2}$$

where R_{sh} is the sheet resistance of the channel (Ω/□), R_c is the ohmic contact resistance (Ω·mm), and W_g is the gate width. R_s and R_d can be directly measured by using the gate current injection method [9]. $R_{contact}$ is obtained by measuring R_c through the four-point transfer length measurement (TLM) [10] and R_{access} is calculated by subtracting $R_{contact}$ from R_s for a given bias condition. It is noted that due to the bias dependent dynamic behavior of the access resistance in AlGaN/GaN HEMTs [11], the R_{access} calculated from R_{sh} through the four-point TLM technique may not be accurate at certain bias conditions. It is also noteworthy that Eq. (3-2) is valid for low voltage drop across R_{access} (i.e. conditions under which R_{access} can be considered linear).

In this work, R_{access} was first reduced by scaling source-to-drain distance (simultaneously L_{gs}) to minimize R_s. Although R_{access} is proportional to the source-to-drain distance, this distance cannot be arbitrarily reduced, as it impacts the gate lithography which needs to be aligned

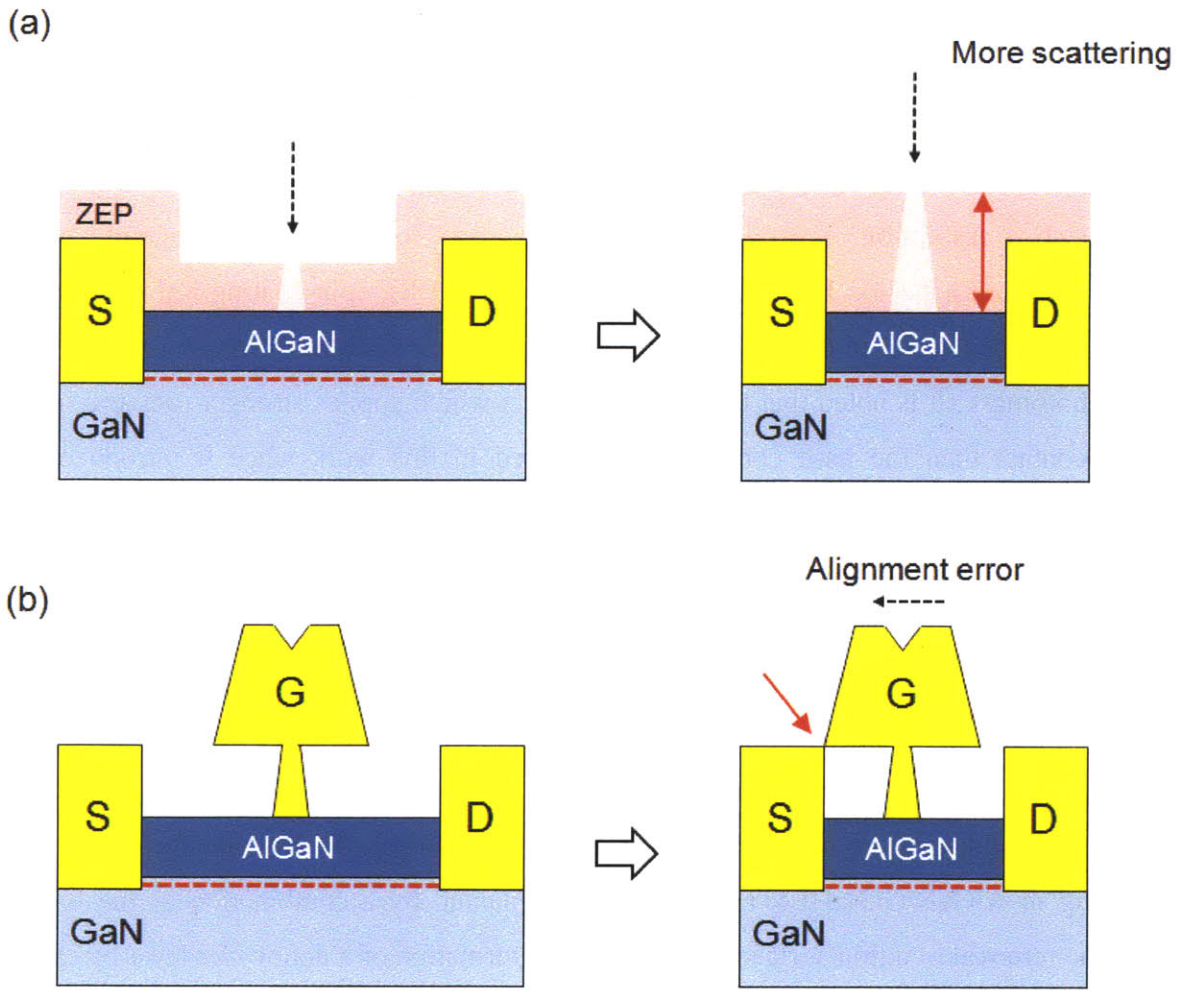


Figure 3-3. Challenges in scaling source-to-drain distance. (a) Due to the thicker resist, e-beam lithography conditions need to be modified and it is harder to pattern small gate lengths. (b) Small alignment error during the gate process can cause short-circuits between the gate and either source and drain contact.

between the source and drain contact. Figure 3-3 illustrates issues in reducing source-to-drain distance in GaN HEMTs. As the source-to-drain contacts get closer to each other, spin-coated e-beam resist becomes thicker than expected and the parameters for e-beam lithography (dose, development time, etc.) need to be adjusted accordingly. However, due to the thick ohmic metal stacks required in GaN HEMTs (e.g. Ti/Al/Ni/Au > 200 nm) the thickness of e-beam resist becomes too thick at one point and this limits the minimum gate length that can be defined by e-beam lithography (Figure 3-3(a)). Moreover, a very tight source-to-drain distance can cause poor device yield. During the gate process, small alignment errors can lead to short-circuits between the gate contact and either the source or drain contact (Figure 3-3(b)).

By considering these issues, a source-to-drain distance of 1 μm was determined to provide a reasonably small R_{access} with a small gate length (< 100 nm) and high device yield. The source and drain contact regions were defined by photo-lithography and the process parameters were carefully optimized to produce a 1 μm distance with acute edge profile in a reproducible way. To achieve such a resolution, several techniques were used for the photo-lithography such as promoting surface adhesion by hexamethyldisilazane (HMDS), spin-coating a thin photoresist, removing edge bead prior to the pattern exposure, and making a hard contact between the sample and the photomask. It is noted that the vacuum contact which applies stronger pressure for more intimate contact than the hard contact was not used in this work since it introduces small misalignments and can damage (or scratch) the mask. Figure 3-4 shows a top-down SEM image of an optimized ohmic contact lithography to achieve 1 μm source-to-drain distance.

Secondly, $R_{contact}$ was reduced by developing a recessed ohmic technology. A mild ion etching prior to the ohmic metallization was used in early n-type GaN devices to improve their ohmic contact resistivity [12]. The application of a similar technique to AlGaN/GaN devices was also demonstrated by several groups later [8][13]. The improvement of ohmic contact resistance by ion etching in AlGaN/GaN HEMTs is typically explained by a combination of the following effects: the removal of a thin surface oxide layer, the formation of a donor-like layer by changing surface chemistry, and the increment of tunneling current by thinning the AlGaN layer. However, excessive etch (or removal) of the AlGaN layer results in the degradation of ohmic contact

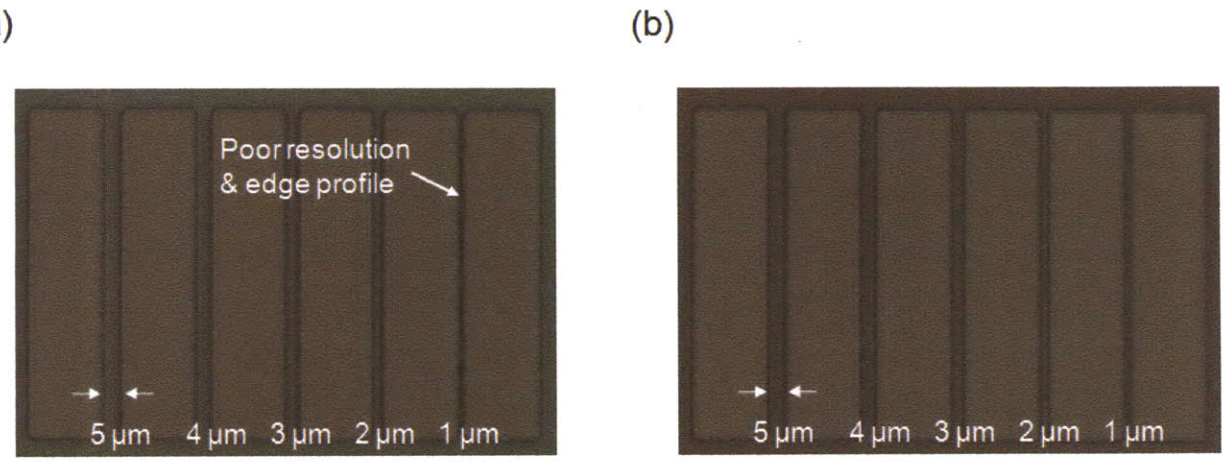


Figure 3-4. (a) Un-optimized photolithography through contact aligner cannot produce 1 μm resolution. (b) By optimizing photolithography (HMDS, thin photoresist, edge bead removal), 1 μm resolution and sharp edge acuity were obtained.

resistances due to the depletion of the 2DEG charges underneath the ohmic contact regions which reduces tunneling electrons. Thus, it is important to find optimum remaining AlGaN barrier thickness after the recess to minimize the ohmic contact resistance.

Another important optimization parameter to obtain the lowest ohmic contact resistance in AlGaN/GaN HEMTs is alloying temperature. In general, it is difficult to form an ohmic contact to wide bandgap materials such as AlGaN and GaN because work functions of most metals are not low enough to form a low Schottky barrier to allow current transport. Thus, thermionic emission hardly takes place and as-deposited metals show Schottky behavior to the AlGaN surface. Ohmic contacts to III-N semiconductors can be made by rapid thermal annealing (RTA) of metals at high temperatures (800 ~ 1000 °C). During the RTA, the metal-AlGaN interface is heavily doped by their surface reaction and this largely increases tunneling probability through the significant band bending at the interface. Conventional ohmic metals consist of Ti/Al/Ni/Au although sometimes Ni can be substituted by other metals such as Ti, Mo, Pt, and Ir. It is believed that when annealed at high temperatures, Ti forms TiN at the Ti/AlGaN interface causing N-vacancies, which dope the AlGaN layer n-type [14]. As a result, the n-type doped AlGaN and the conductive TiN facilitate tunneling mechanism of carriers at the interface. In fact, the formation of TiN and resultant ohmic contact resistance is very sensitive to the RTA temperature and therefore the optimum RTA temperature was investigated to minimize the contact resistance.

In this work, the AlGaN barrier was slowly etched using a low power electron cyclotron resonance reactive ion etching (ECR-RIE) with Cl_2/BCl_3 gas mixture. To minimize the damage induced by ion bombardment, RF bias was kept low (60 ~ 70 V) while ECR power was set to achieve an etch rate of 1 nm/min (~100 W). After the recess, a Ti/Al/Ni/Au (20/100/25/50 nm) metal stack was deposited, followed by RTA for 30 s in N_2 atmosphere. The remaining AlGaN barrier thickness was measured by atomic force microscopy (AFM) in regions of the sample where the ohmic metals were not deposited. RTA temperature and remaining AlGaN barrier thickness were experimentally optimized to have the lowest contact resistance as shown in Figure 3-5. The sheet resistance (R_{sh}) was not affected by the recessed ohmic technology and was kept almost constant (~ 360 Ω/□). It is noted that although RTA at 750 °C for 30 s was also

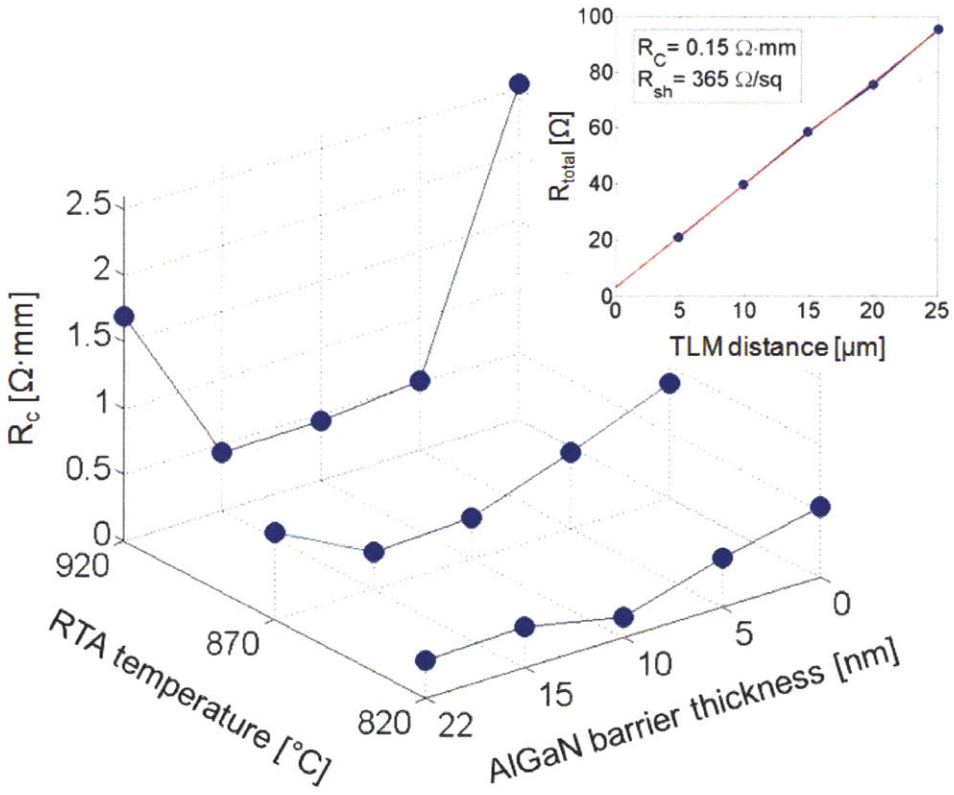


Figure 3-5. TLM-measured contact resistance (R_c) as function of RTA temperature and remaining AlGaN barrier thickness. The optimum R_c of 0.15 Ω·mm was obtained at 10 nm AlGaN barrier after annealing at 820 °C for 30 s in N_2 ambient. The sheet resistance (R_{sh}) was almost constant (~ 360 Ω/□) in all cases.

investigated, this condition was turned out to be not sufficient to form the ohmic contact. In the devices with no remaining AlGaN barrier under the contacts, the ohmic contact was formed laterally between the annealed metals and the 2DEG in the unrecessed AlGaN/GaN access region [15].

The best condition among the range we explored was found at 820 °C with 10 nm AlGaN barrier yielding an R_c of 0.15 Ω·mm, which is one of the lowest contact resistances reported in the literature. As expected, at each RTA temperature the contact resistance was initially decreased when reducing the AlGaN barrier thickness due to the removal of a surface oxide layer and the increase in tunneling probability. On the other hand, an increase in contact resistance was observed for deeper etches of AlGaN barrier and it is associated with the depletion of the 2DEG charges underneath the ohmic contact regions, which cause a reduction in tunneling current.

Despite the very low contact resistance achieved by optimizing the recessed ohmic technology, the ohmic contact showed a rough surface morphology similar to conventional alloy contacts reported in the literature. A smooth ohmic surface is very important to increase processing yield and reproducibility of the gate technology, especially for short source-to-drain distances (1 μm) as discussed in Figure 3-3. A rough surface morphology or the formation of the bulges at the ohmic surface is mainly caused by the reaction of Al with Ni or Au during the RTA [16]. It can be prevented by inserting high melting point refractory metals such as tungsten (W) as a capping layer on top of Al to suppress the reaction of Al [17][18]. However, such a change generally increases contact resistance and it is difficult to achieve both smooth surface roughness and low contact resistance at the same time.

To address this challenge a new recessed ohmic contact metallurgy based on SiGe-based alloyed contacts (Si/Ge/Ti/Al/Ni/Au = 2/2/20/100/25/50 nm) was developed. The enhancement of the contact resistance by inserting a thin layer of Si under the Ti/Al/Ni/Au metallization was already demonstrated by other groups [19][20][21]. The improvement of the contact resistance is believed to be due to the local diffusion induced Si-doping [21] and the accelerated penetration of the metal contacts into the AlGaN barrier which increases tunneling probability [19]. Possible mechanism of the facilitated down diffusion of Ni and Au can be similar to the gettering process in Si electronics [22]. During the RTA, the bottom Si layer creates defect sites which can attract metals such as Ni and Au. Although reference [19][21] does not discuss the change of surface roughness by inserting Si layer, we expected the accelerated down diffusion of Ni and Au to improve the surface morphology by preventing the reaction with Al at the contact surface. Ge was added in this work to enhance the doping of the contact regions even more. Finally, by combining this Si/Ge doping/gettering layers and recessed ohmic technique, we could reduce both ohmic contact resistance (by 55 %, R_c = 0.21 Ω·mm) and surface roughness of the contacts (by 83 %) over conventional non-recessed Ti/Al/Ni/Au ohmic contacts as shown in Figure 3-6. It is noted that Ge alone could not be used for this purpose due to its poor adhesion to the AlGaN surface.

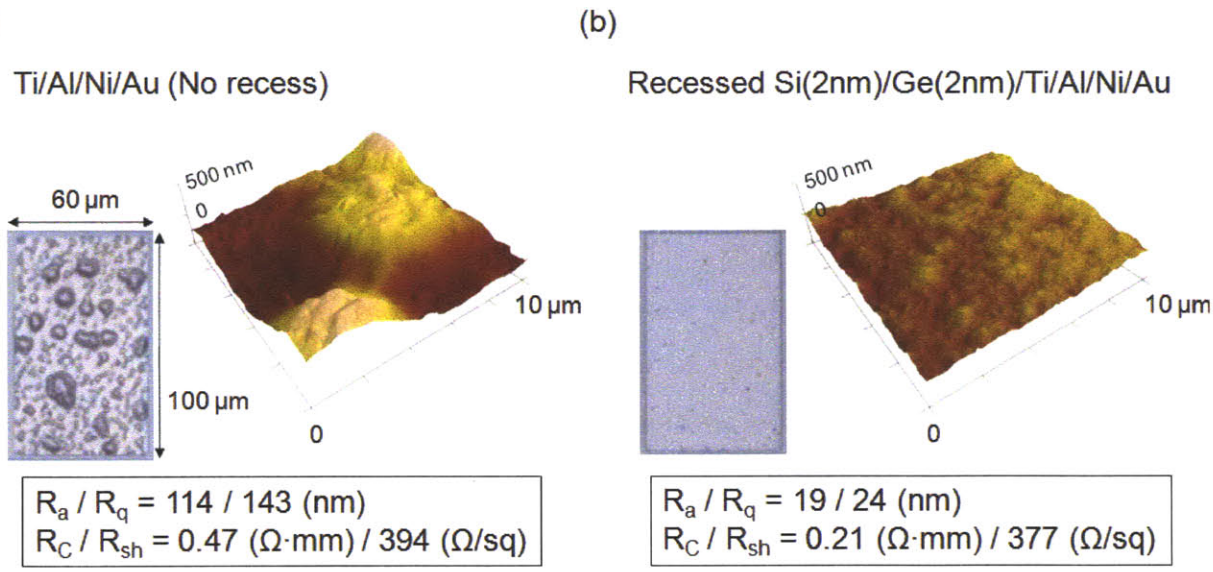


Figure 3-6. Comparison between (a) conventional and (b) recessed Si(2nm)/Ge(2nm)/Ti/Al/Ni/Au ohmic contacts. Both of them were alloyed at 820 °C. 55% smaller contact resistance (R_c) and 83% smoother surface roughness were observed for the new ohmic technology by TLM, AFM, and microscopy images (inset).

As a result, the combination of 1 μm source-to-drain distance and optimized Si/Ge-based recessed ohmic technology produced a very low R_s of 0.4 Ω·mm.

3.2.2. Gate resistance and gate-drain capacitance (R_g, C_{gd})

The resistance of the gate electrode is determined by its dimensions (Figure 3-7) and by the operating frequency. Due to skin effects, at high frequencies the gate current tends to flow through the outermost part of the gate metal at an average depth equal to the skin depth. The skin depth is expressed as [23],

$$\delta = \sqrt{\frac{2}{\omega(\mu_r \mu_o)\sigma}} \tag{3-3}$$

where $\omega = 2\pi f$ is the angular frequency and f is the operating frequency while μ_r , μ_o, and σ are relative permeability, permeability of vacuum ($4\pi \times 10^{-7}$ H/m), and conductivity of the gate metal, respectively. Table 3-1 shows the skin depth of typical gate metals in AlGaN/GaN HEMTs. In our devices, the height and width of the gate electrode are normally less than the skin depth in the interested microwave frequency ranges.

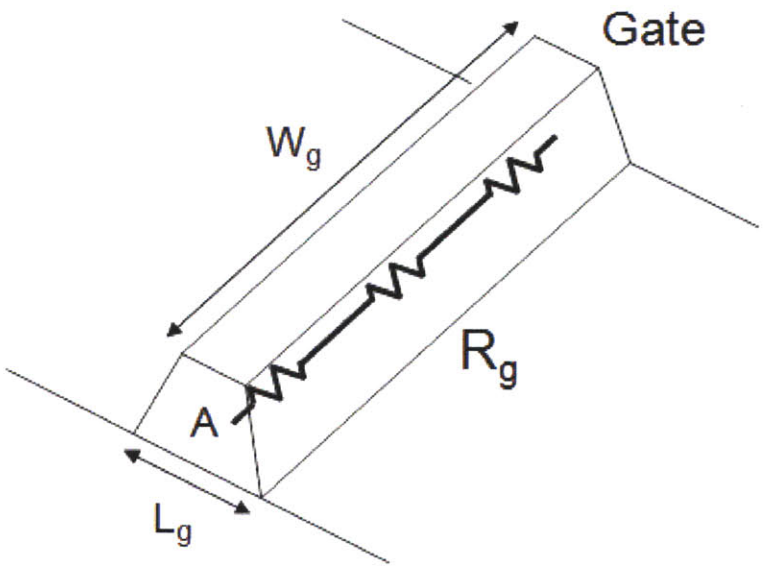


Figure 3-7. Illustration of the gate resistance R_g which is largely dependent on the geometry of the gate strip.

Thus, the entire cross-sectional area of the gate metal contributes to its conductance and the gate resistance at microwave frequencies can be estimated as [23],

$$R_g = \frac{1}{3} \cdot \frac{1}{N_f{}^2} \cdot \left(\rho \frac{W_g}{A} \right) \tag{3-4}$$

where N_f is the number of gate fingers, ρ is the resistivity of the gate metals ($\rho = 1 / \sigma$), W_g is the gate width, and A is the cross-sectional area of the gate strip. The factor of 1/3 accounts for the fact that when the microwave signal is applied to the gate, not all charges have to move from the feed point (gate pad) to the very end of the gate strip. Since these charges are distributed along the gate under small-signal condition, the effective series resistance of the gate strip is less than

Metal	μ_r	σ [$\times 10^7$ S/m]	δ @10 GHz [μm]	δ @100 GHz [μm]
Ni	100 ~ 600	1.43	0.05 ~ 0.13	0.02 ~ 0.04
Cu	1	5.81	0.66	0.21
Au	1	4.10	0.79	0.25
Pt	1	0.95	1.63	0.52
Ti	1	0.24	3.26	1.03

Table 3-1. Relative permeability, conductivity, and skin depth at 10 GHz and 100 GHz of the typical gate metals. It is noted that Ni is a ferromagnetic material and its relative permeability is higher than other gate metals.

that of DC condition. The factor of $1/N_f$ takes into account the effect of splitting gate fingers and connecting them in parallel. The fabricated devices were designed to have two gate fingers (N_f = 2). It is also noted that since a Ni/Au bi-layer gate metal stack was used in this work, two resistors in parallel need to be considered to estimate the total gate resistivity.

For very short gate lengths, rectangular gates lead to unacceptably high values of R_g due to the small A, therefore T-shape gates are much more advantageous to minimize R_g as shown in Figure 3-8. For example, R_g can be reduced more than ten times by fabricating T-shape gate for 60 nm gate length devices. Although T-shape gates introduce additional parasitic capacitances due to the large head part of the T-gate, its effect can be minimized by increasing the gate foot stem height.

The gate-drain capacitance C_{gd} can be minimized by reducing the gate length. Following the analytical expression of C_{gd} in MESFETs [23], C_{gd} in AlGaN/GaN HEMTs in the saturation region can be approximated by

$$C_{gd} = \frac{2\varepsilon W_g}{1 + \frac{2X_{dep}}{L_g}} \tag{3-5}$$

where X_{dep} is the lateral extension of the depletion region at the drain side of the gate. Figures 3-

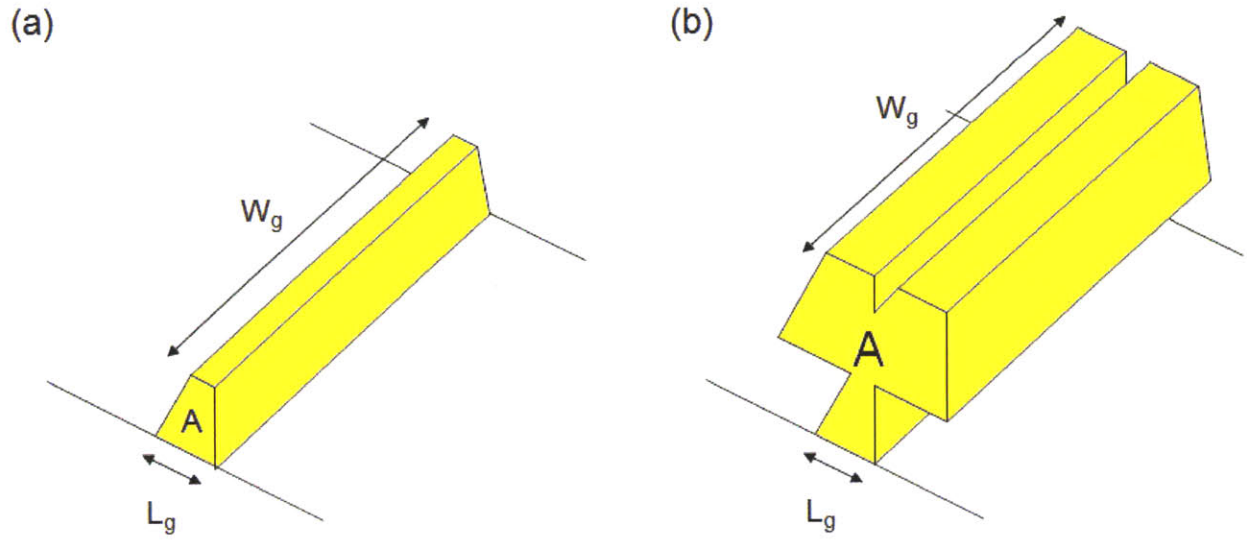


Figure 3-8. Comparison of the gate resistance R_g between (a) rectangular gate and (b) T-shape gate. With the same gate length, T-shape gate has much lower gate resistance due to the larger cross-section area of the gate strip.

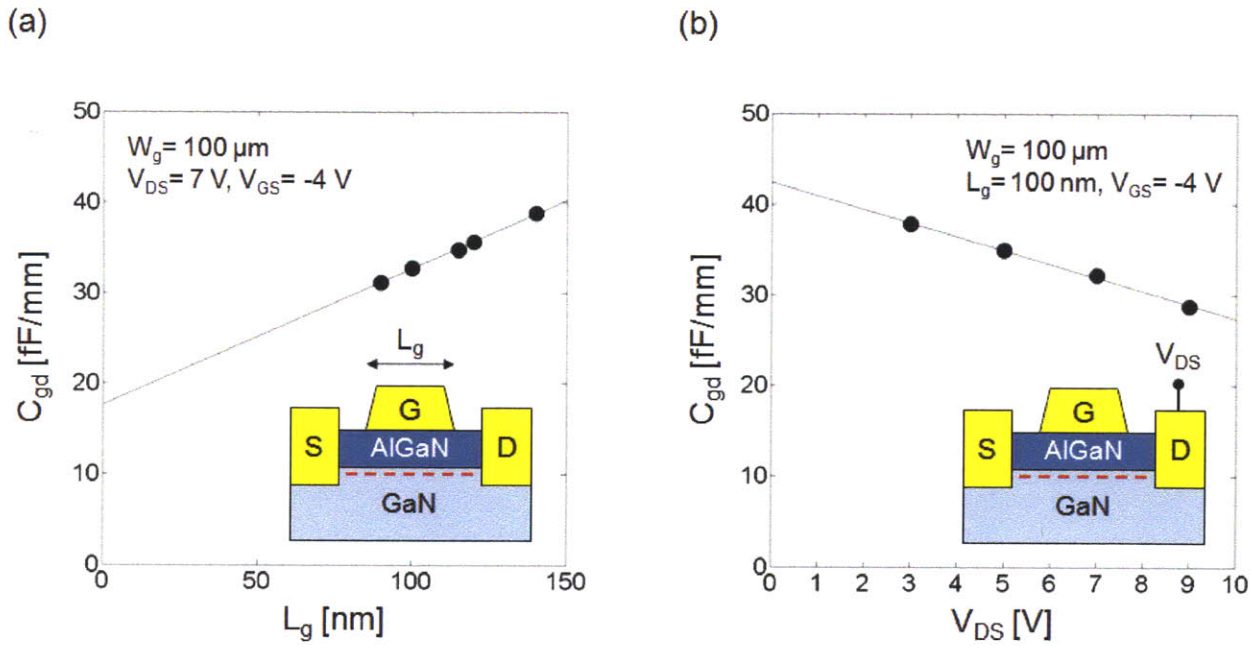


Figure 3-9. Trend of C_{gd} as a function of (a) gate length L_g and (b) drain bias V_{DS} (or X_{dep}). C_{gd} was extracted from the measured S-parameters.

9(a) and 3-9(b) show measured C_{gd} as a function of L_g and V_{DS} (or X_{dep}), respectively to illustrate the trend of C_{gd}. The measured C_{gd} becomes smaller with reducing L_g and increasing V_{DS} (or increasing X_{dep}) as Eq. (3-5) predicts. Therefore, the fabrication of deep-submicron T-shape gate not only lowers R_g, but also reduces C_{gd}.

The conventional technology for the fabrication of T-shape gates was already presented in Chapter 2. It should be highlighted that the edge trimming process which was developed to improve the process yield and to further reduce the gate length is also very effective to fabricate T-gates with large head size while maintaining a short gate length. The increased junction area between head and foot part of the T-gate improves the stability of larger gate heads. By considering process yield and uniformity, the head size was optimized to H_g = 500 ~ 600 nm to reduce R_g while the gate stem height was increased to T_g = 200 nm to minimize the parasitic gate fringing capacitances.

Therefore, a reliable deep-submicron T-gate (L_g = 60 nm, H_g = 500 ~600 nm, T_g = 200 nm) was fabricated to minimize both gate resistance R_g and gate-drain capacitance C_{gd} as shown in Figure 3-10.

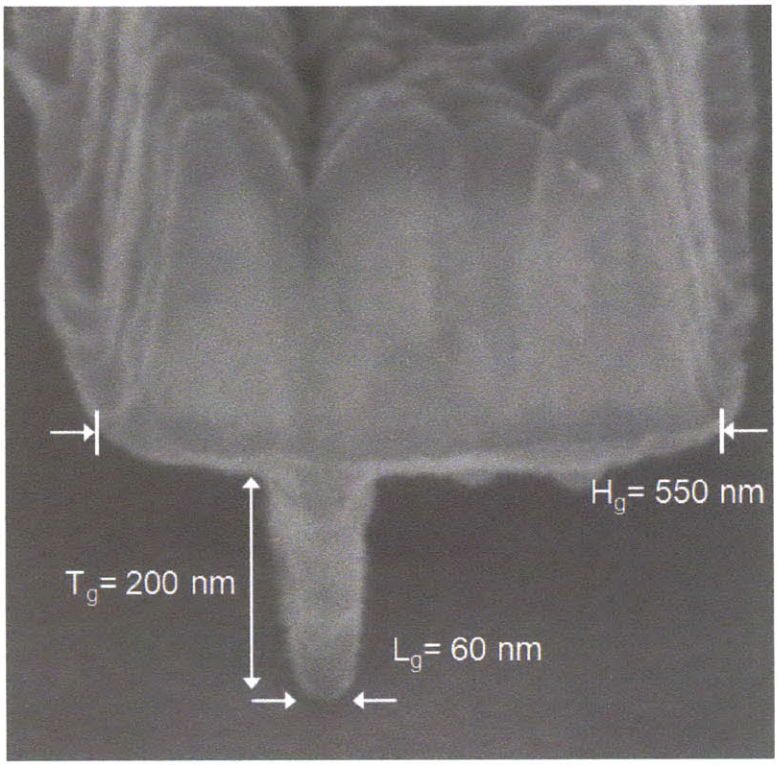


Figure 3-10. A tilted cross-sectional SEM image of 60-nm T-shape gate. The large head size (H_g = 550 nm) reduces gate resistance and the increased stem height (T_g = 200 nm) minimizes the parasitic effects from the increased head size.

3.2.3. Input resistance and output conductance (R_i, g_o)

The input resistance R_i is the lumped element representation of the distributed channel resistances in the small-signal equivalent circuit. At microwave frequencies, the gate region in AlGaN/GaN HEMTs can be modeled by a distributed network as shown in Figure 3-11. The capacitances are distributed along the channel resistances and for simplicity the gate inductance and the leakage resistance were neglected. The channel resistance R_{ch} for the device operation in the saturation region can be estimated as [23],

$$R_{ch} \approx \frac{3v_{sat}L_g}{\mu I_{DS}} \tag{3-6}$$

where v_{sat} is the carrier saturation velocity, μ is the low field mobility, and I_{DS} is the drain current (or channel current). In short gate length devices, I_{DS} can be substituted by Eq. (2-16).

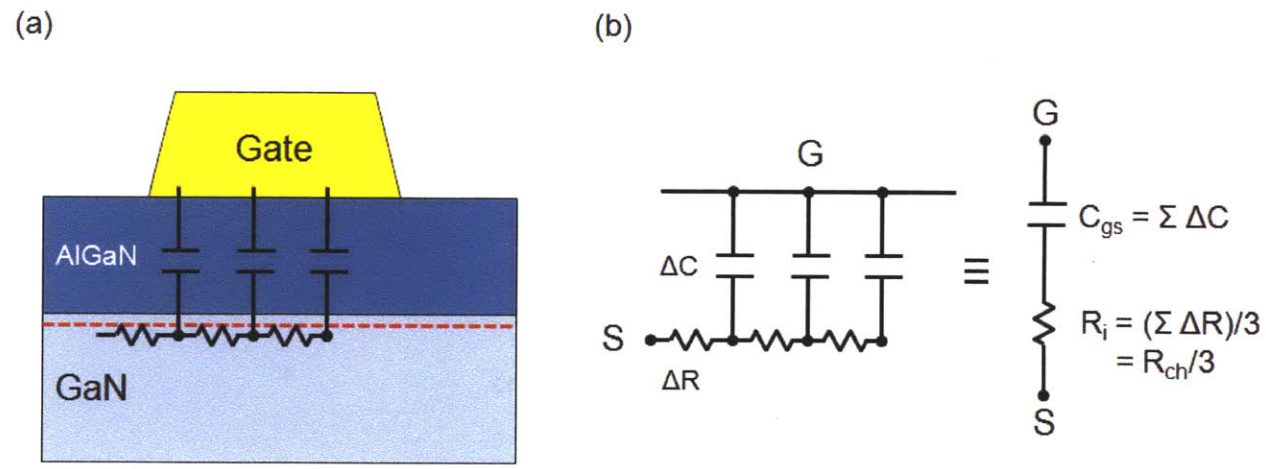


Figure 3-11. (a) Illustration of distributed RC network model underneath the gate at microwave frequencies. (b) Lumped element representation of the distributed RC network to explain the origin of the input resistance R_i.

$$R_{ch} \approx \frac{3v_{sat}L_g}{\mu I_{DS}} = \frac{3L_g}{\mu C(V_{GS}-V_{th})} = \frac{3L_g}{\mu(V_{GS}-V_{th})}\left(\frac{2}{3}\frac{\varepsilon_b}{t_b}\right)^{-1} \tag{3-7}$$

where ε_b is the dielectric constant of the AlGaN barrier and t_b is the AlGaN barrier thickness (or gate-to-channel distance). It is noted that the factor of 2/3 was included to account for the value of C_{gs} in the saturation region (C_{gs} = 2/3(ε_b/t_b)) [24]. It is well known result of distributed network theory that the equivalent small-signal resistance is one-third of the total DC resistance (R_{ch}) [23]. Thus, in short gate length devices, the input resistance R_i can be written as

$$R_i = \frac{1}{3}R_{ch} = \frac{3L_g t_b}{2\mu\varepsilon_b(V_{GS}-V_{th})} \tag{3-8}$$

Therefore, R_i can be minimized by reducing t_b under the gate region and a low damage gate-recess technology can be used for this purpose. It is noted that the effect of V_{th}-shift by reducing t_b is less effective to change R_i. The R_i is often called the gate charging resistance because it acts as a charging resistance for the gate-source capacitance C_{gs}. The lower the gate-to-channel distance t_b, the smaller the RC time constant for charging and discharging C_{gs}, which is equivalent to smaller R_i. Figure 3-12(a) shows experimental data of R_i to show this trend.

The output conductance g_o is defined as the variation of the drain current caused by a variation of V_{DS} at fixed V_{GS}. It is caused by the lack of gate control over the channel electrons or poor

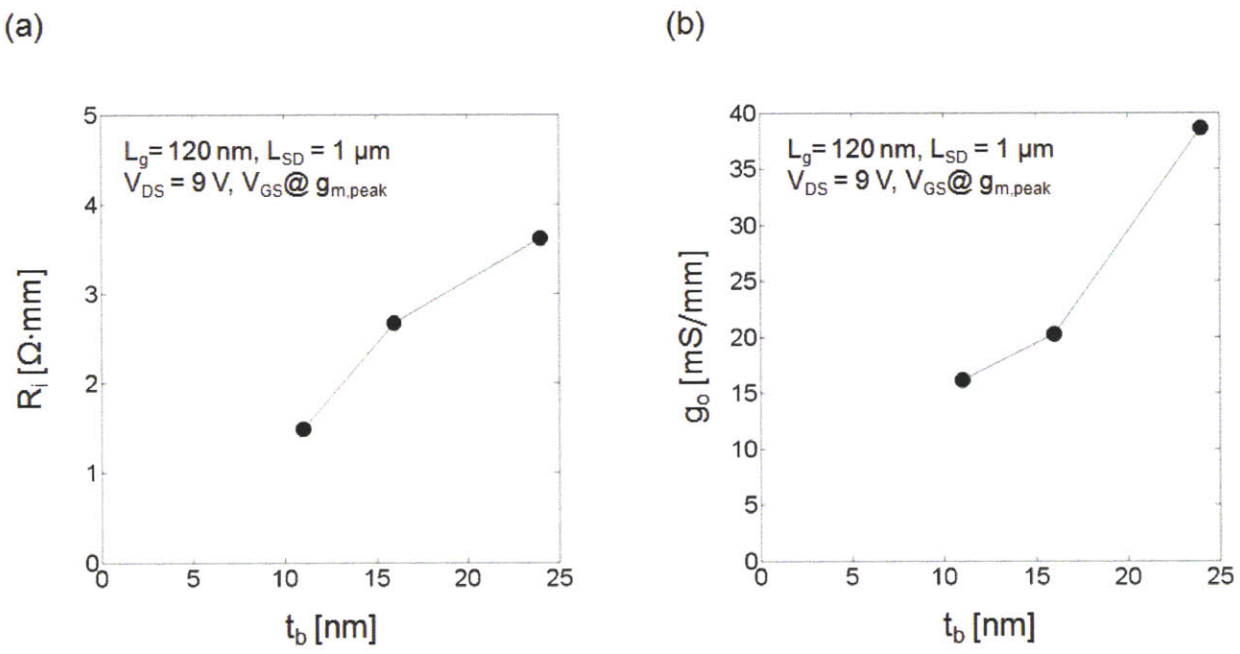


Figure 3-12. Trend of (a) R_i and (b) g_o as a function of AlGaN barrier thickness t_b. As reducing t_b, R_i and g_o become smaller due to the reduced time to charge and discharge C_{gs} and improved aspect ratio L_g / t_b, respectively.

channel confinement at the AlGaN/GaN interface. For example, as the gate length is reduced for a given gate-to-channel distance, the so-called short-channel effects arise due to poor gate aspect ratio. As discussed in Section 3.1., this increased output conductance in the small gate length devices is attributed to the two-dimensional field distribution and high electric fields in the channel region. Another important cause of high output conductance in AlGaN/GaN HEMTs is the buffer leakage. Under high bias conditions, the electric field increases rapidly at the gate-drain region, reaching values as high as 1 ~ 3 MV/cm, especially at the drain side of the gate edge. At such a high field, channel electrons gain very high energy and the channel tends to spread into GaN buffer. In other words, some of the hot electrons start to flow through the GaN buffer instead of the channel at the AlGaN/GaN interface, giving rise to a buffer leakage current. The buffer leakage current increases at high V_{DS} and it may constitute several tens of percent of the total drain current. The expression of g_o can then be approximated by taking into account two leakage currents between the source and the drain, namely channel leakage current $I_{ch,l}$ and buffer leakage current $I_{bf,l}$.

$$g_o = \left.\frac{dI_{DS}}{dV_{DS}}\right|_{V_{GS}} = \left.\frac{dI_{ch,l}}{dV_{DS}}\right|_{V_{GS}} + \left.\frac{dI_{bf,l}}{dV_{DS}}\right|_{V_{GS}} \tag{3-9}$$

Thus, it is important to minimize both channel leakage and buffer leakage current. However, the common techniques to suppress buffer leakage current often require the modification of the device structure, such as the use of a p-type buffer layer [25] or InGaN back barrier [26]. These structures are challenging to grow without degrading the transport properties of the original structure. Another possible solution for the buffer leakage current is to apply field plates to spread and reduce the high electric field in the gate-drain region [27]. However, the use of field plate structures introduces large gate capacitances which in turn degrade high frequency performance of GaN HEMTs. Due to these concerns in the high frequency characteristics, we mainly focus on reducing channel leakage current rather than buffer leakage current to improve g_o.

As already discussed, the channel leakage current originates from the short-channel effects which can be circumvented by increasing the gate aspect ratio (L_g / t_b). For a given gate length, a closer gate-to-channel distance (or lower t_b) enhances effective gate field strength over the channel electrons and improves immunity to short-channel effects. Figure 3-12(b) shows the output conductance g_o as a function of AlGaN barrier thickness (or gate-to-channel distance) at a given gate length and it is shown that the thinner barrier has smaller g_o mainly due to the improved aspect ratio.

Although the use of a thin AlGaN barrier is advantageous for reducing both R_i and g_o of the device, the thin AlGaN thickness lowers the charge density in the channel which in turn degrades the source and drain access resistances (R_s and R_d). Therefore, it is better to reduce the AlGaN barrier thickness locally only underneath the gate preserving low R_s and R_s at the access regions. In this regard, a low-damage gate recess technology was developed to minimize R_i and g_o.

First of all, electron-beam lithography was used to define a T-shaped gate, using a trilayer resist stack made of either PMMA/Copolymer/PMMA or ZEP/PMGI/ZEP. Once the T-shape gate pattern is defined, a low power electron cyclotron resonance reactive ion etching (ECR-RIE) with Cl_2/BCl_3 gas mixture was used for the gate recess. To minimize the damage induced by ion bombardment, RF bias was kept low (60 ~ 70 V) while the ECR power was set to achieve an etch rate of 1 nm/min (~ 100 W) for AlGaN barrier. The etch rate of PMMA and ZEP e-beam

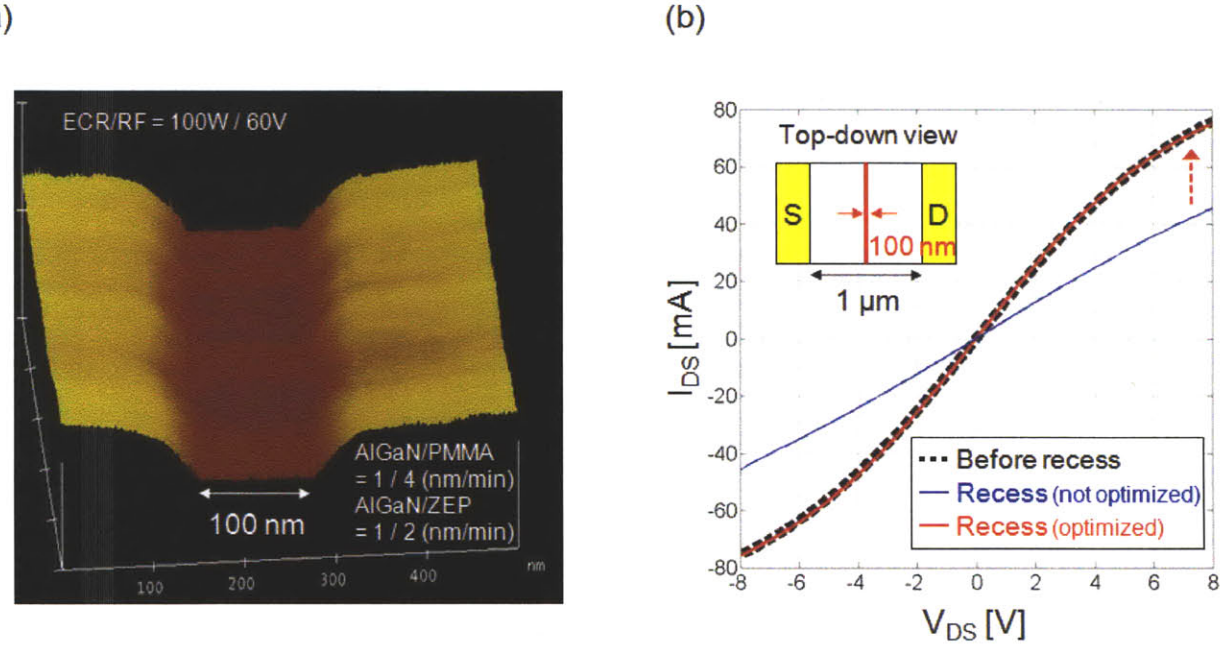


Figure 3-13. (a) AFM image of the AlGaN surface after applying low-damage gate recess technology developed in this work. (b) The measured current between 1 μm separated source and drain contacts after the gate recess of 100 nm length in ungated region. No significant degradation of the current was observed after the optimized low-damage gate recess.

resists was 4 nm/min and 2 nm/min, respectively (the etch selectivity over AlGaN was 1:4 and 1:2, respectively). It is noted that for a very short gate pattern, typically L_g < 150 nm, the etch rate of AlGaN becomes significantly reduced and more time is needed to achieve the target recess depth. This effect is known as aspect ratio dependent etching (ARDE) [22] and the etch rate is different for different aspect ratios, with a lower etch rate for higher aspect ratios (either smaller gate length or deeper recess depth). Hence, it is preferred to use the ZEP as a etch mask rather than PMMA to allow a longer recess time. Figure 3-13 shows AFM image of recessed AlGaN surface by the developed low-damage gate recess technology. We did not observe noticeable degradation in the source-to-drain current after the recess and no post annealing step was followed. This is due to the small recess region (< 100 nm) compared to relatively long source-to-drain distance (1 μm) and confirms the low plasma damage from the optimized gate recess technology. After the gate recess, Ni/Au metal stack was deposited for the gate contact.

Therefore, carefully optimized low damage gate recess technology was used to minimize the input resistance R_i and output conductance g_o.

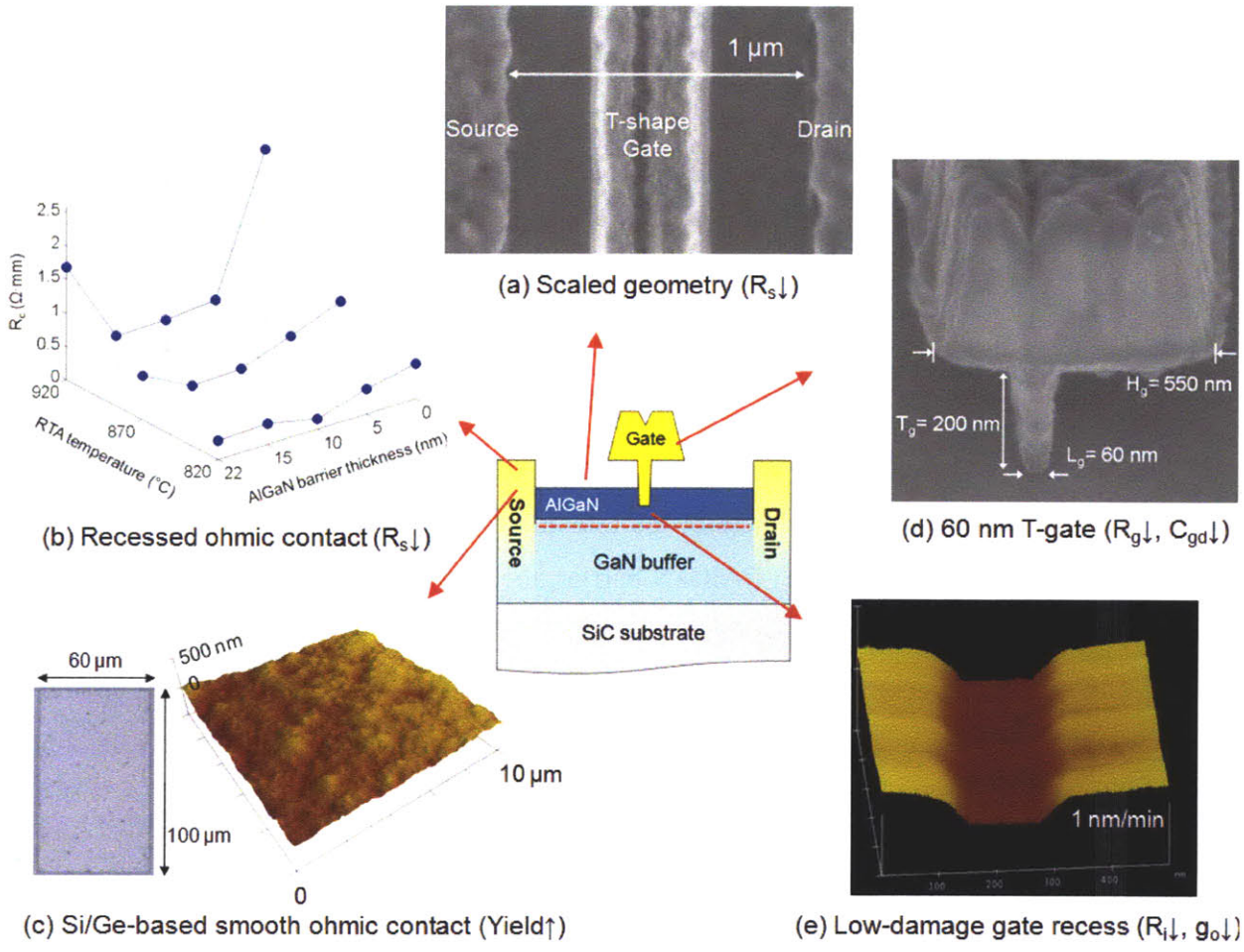


Figure 3-14. Summary of developed technologies to improve f_{max} in AlGaN/GaN HEMTs.

3.2.4. Process integration

Figure 3-14 summarizes the technologies developed and integrated in this work to minimize parasitic elements R_i, R_s, R_g, C_{gd}, and g_o to maximize f_{max} in AlGaN/GaN HEMTs. The AlGaN/GaN transistor structure was grown on a SiC substrate by molecular beam epitaxy (MBE). This structure produced a 2DEG with a total charge density of 8×10^{12} /cm^2 and electron mobility of 2200 cm^2/V·s as measured on unpassivated samples using van der Pauw structures at room temperature. This mobility and charge density translate to a 2DEG sheet resistance of 356 Ω/□. Device fabrication began with mesa isolation using a Cl_2/BCl_3 plasma-based dry etch. Then, source and drain regions with a 1 μm-separation (Figure 3-14(a)) are defined by photolithography and SiGe-based recessed ohmic technique was used to reduce the ohmic

contact resistances ($R_s\downarrow$) and improve ohmic surface roughness (yield$\uparrow$) (Figure 3-14(b) and (c)). Electron-beam lithography was applied to pattern a deep-submicron T-shaped gate with small gate length and large head size ($R_g\downarrow$, $C_{gd}\downarrow$) (Figure 3-14(d)) and the low-damage gate recess technology was used to reduce the gate-to-channel distance ($R_i\downarrow$, $g_o\downarrow$) (Figure 3-14(e)). Finally, the device fabrication is completed by depositing a Ni/Au metal stack for the gate contact.

3.3. Device characteristics

In this section, device characteristics of a recessed AlGaN/GaN HEMT with 60-nm physical gate length are presented. The thickness of the remaining AlGaN barrier layer was estimated to be 11 nm considering the etch rate of AlGaN barrier, the positive shift of threshold voltage, and the increase of transconductance. The fabricated HEMT shows one of the highest f_{max} reported to date in any nitride semiconductors thanks to well optimized parasitic elements. (It should be noted that Hughes Research Laboratories recently has announced higher f_{max} devices in shorter gate length of 40 nm [28].)

3.3.1. DC characteristics

Figure 3-15 shows the DC characteristics of fabricated 60-nm gate-recessed AlGaN/GaN HEMT. The short source-to-drain distance of 1 μm in combination with the optimized recessed ohmic contact allowed a very low on-resistance of 1.1 ~ 1.2 Ω·mm and a knee voltage of only 2 V. Also, the short-channel effects were effectively suppressed by the low-damage gate recess technology as demonstrated by the low output conductance and the good pinch-off characteristics (I_{on} / I_{off} ~ 5×10^3). This improvement is mainly attributed to better gate control over the channel electrons by improved aspect ratio between the gate length and AlGaN barrier thickness. The maximum drain current at V_{GS} = 2 V exceeded 1.2 A/mm and the peak extrinsic transconductance at V_{DS} = 5 V was 410 mS/mm. The two-terminal gate-drain breakdown voltage defined at 1 mA/mm reverse gate current was measured as ~ 20 V. It is noted that the breakdown voltage and the recess depth are mainly limited by the gate leakage current. The high gate leakage current in AlGaN/GaN HEMTs can be reduced by applying a gate dielectric, such as Ga_2O_3 by oxygen plasma treatment which will be discussed in Chapter 4.

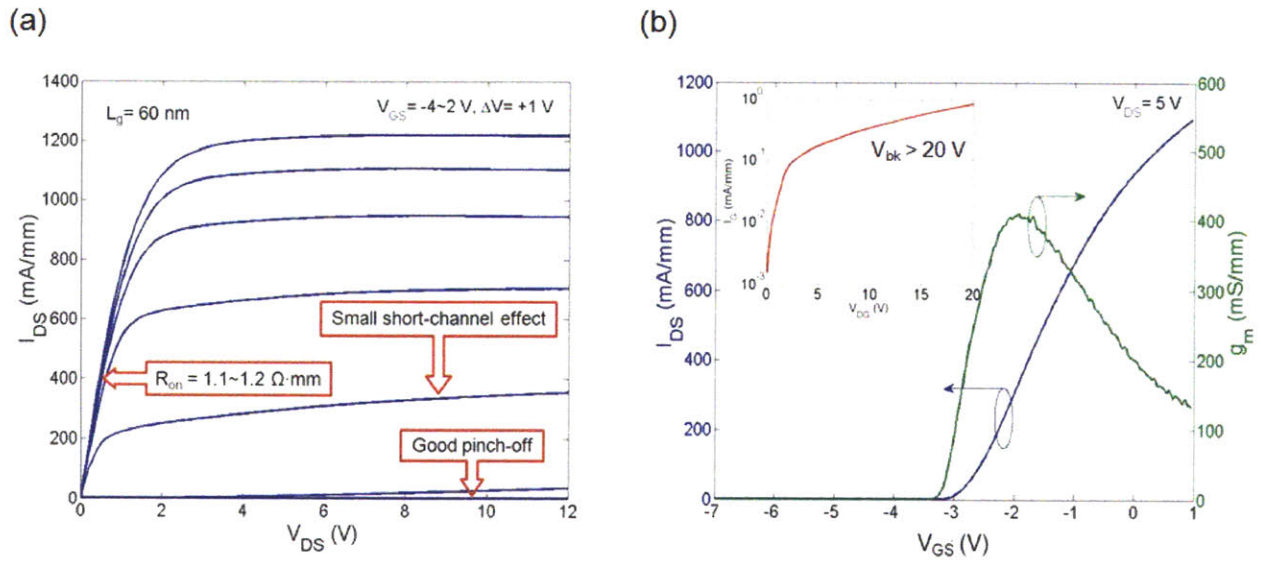


Figure 3-15. (a) DC characteristics of 60-nm gate-recessed AlGaN/GaN HEMT. (a) Output characteristics with a good pinch-off, low R_{on} of 1.1 ~ 1.2 Ω·mm, knee voltage of just 2 V, and small output conductance (small short-channel effect). (b) Transfer characteristics at V_{DS} = 5 V and (inset) the two-terminal gate-drain breakdown voltage of ~ 20 V defined at 1 mA/mm reverse gate current.

3.3.2. RF characteristics

The RF performance of these devices was characterized from 0.45 to 50 GHz using an Agilent Technologies N5230A network analyzer. The system was calibrated with a short-open-load-through (SOLT) calibration standard. The calibration was verified by insuring that both S_{12} and S_{21} of the through standard are less than ±0.01 dB and that both S_{11} and S_{22} are less than -45 dB within the measured frequency range after the calibration [5]. On-wafer open and short patterns were used to subtract the effect of parasitic pad capacitances and inductances from the measured S-parameters [29]. Due to the effect of the pad capacitances, the f_T and f_{max} values calculated from the extrinsic S-parameters were 10 % lower than the intrinsic values. The reproducibility error in our measurements due to the calibration was less than 5 %.

First of all, RF performance of the 160-nm gate length device was characterized. Figure 3-16(a) shows forward current gain $|h_{21}|^2$, Mason's unilateral gain U, and maximum stable gain MSG against frequency at the bias point near its peak transconductance (V_{DS} = 16 V and V_{GS} = -1.8 V). A very high f_{max} of 260 GHz was measured by extrapolating measured U with an ideal slope of -

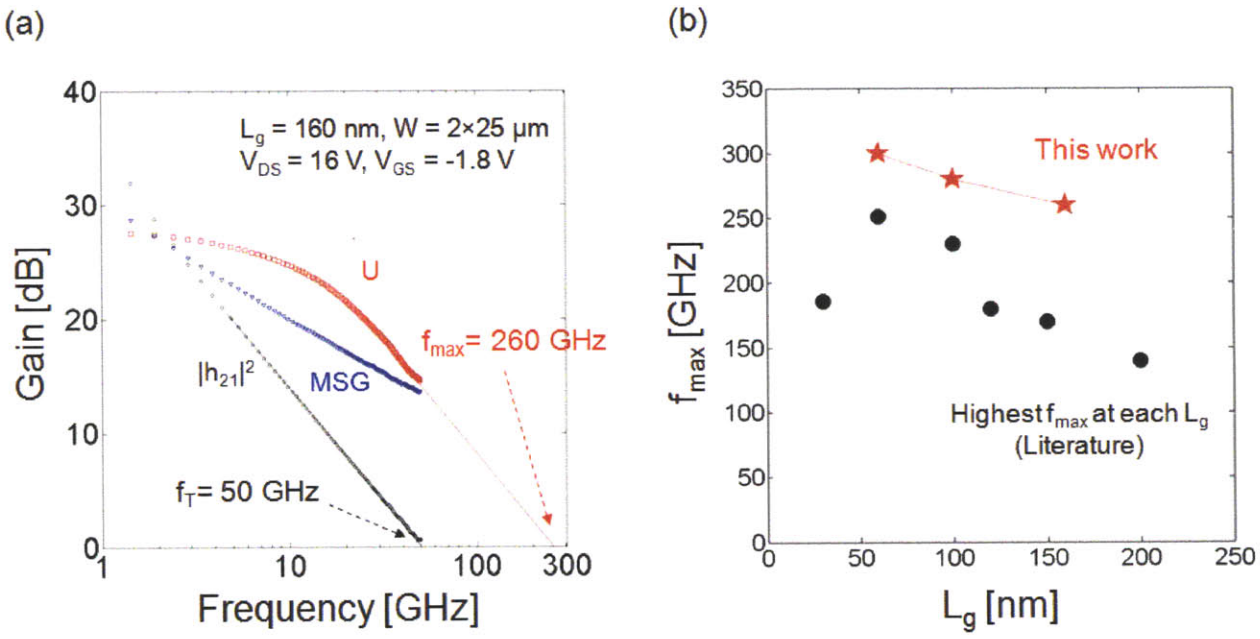


Figure 3-16. (a) RF performance of unpassivated 160-nm gate length HEMT showing f_T = 50 GHz and f_{max} = 260 GHz. (b) The highest f_{max} values at each gate lengh reported in the literature with f_{max} values obtained in this work. f_{max} value of 160-nm device in this work surpasses previous record f_{max} obtained with L_g = 60 nm which shows great importance of minimizing parasitic elements for high f_{max}.

20 dB/dec using a least-square fit. In Figure 3-16(b), the highest f_{max} values at each gate lengh reported in the literature are plotted along with f_{max} values obtained in this work for comparison. The highest f_{max} reported so far in nitride transistors was 251 GHz in 4 nm barrier AlGaN/GaN HEMTs with 60-nm gate length [2]. Despite the relatively long gate length of 160 nm in this work, the obtained f_{max} was higher than the previous record data. This is due to the reduced parasitic elements and shows the great importance of optimizing parasitic elements to maximize f_{max} in AlGaN/GaN HEMTs.

The RF performance of a 60-nm gate length device was also characterized at V_{DS} = 16 V and V_{GS} = -2.2 V as shown in Figure 3-17. The combination of optimized parasitic elements and small gate length resulted in a record f_{max} of 300 GHz. To achieve this high f_{max}, scaled device geometry, recessed source/drain ohmic contacts, and low-damage gate recess technology are harmoniously integrated to simultaneously enable minimum R_i, R_s, R_g, C_{gd}, and g_o. Figure 3-18 shows f_{max} of the same device at different bias points.

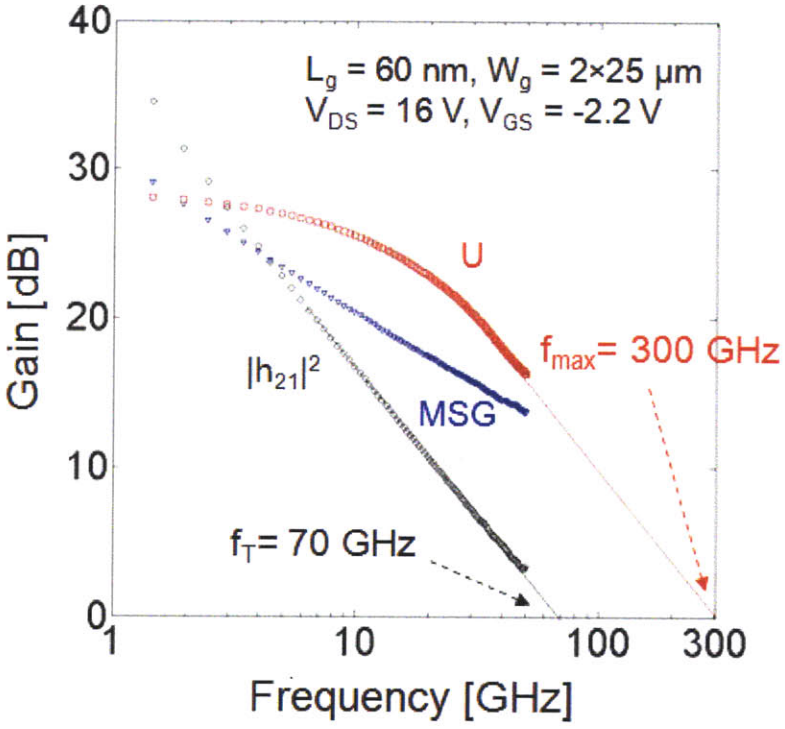


Figure 3-17. RF performance of unpassivated 60-nm gate length AlGaN/GaN HEMT showing f_T = 70 GHz and f_{max} = 300 GHz. The f_T and f_{max} values are extrapolated following a -20 dB/dec ideal decrease with frequency.

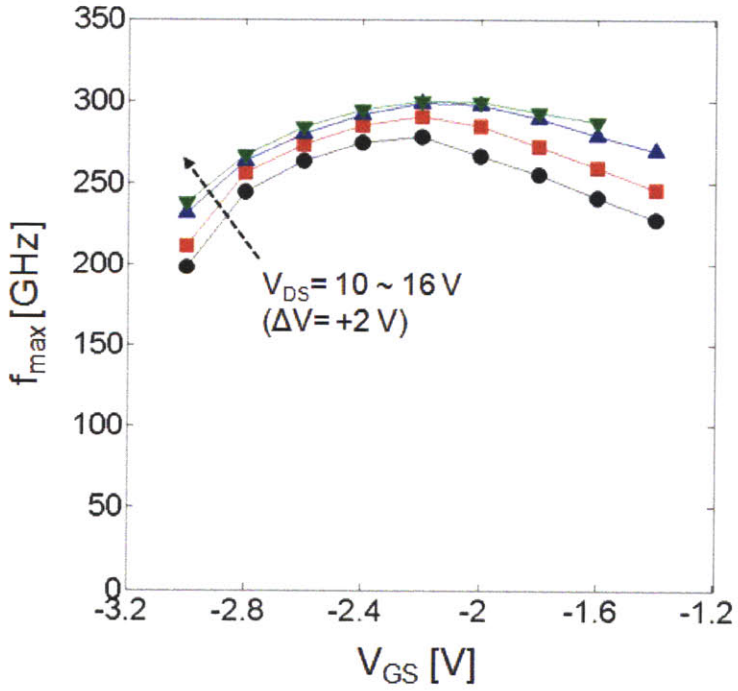
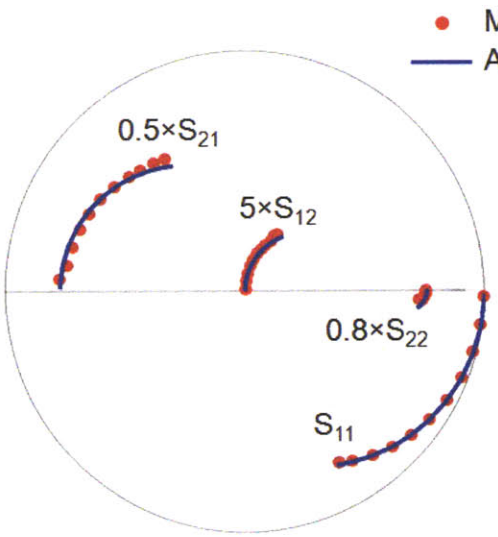


Figure 3-18. Bias dependent f_{max} values of the same device in Figure 3-17.



Intrinsic parameters	Calculated f_T	Measured f_T
g_m = 369.8 mS/mm g_o = 10.4 mS/mm C_{gs} = 740.8 fF/mm C_{gd} = 42.6 fF/mm R_i = 0.45 Ω·mm R_g = 0.23 Ω·mm R_s = 0.42 Ω·mm R_d = 0.57 Ω·mm	73 GHz	70 GHz
	Calculated f_{max}	**Measured f_{max}**
	288 GHz	300 GHz

Figure 3-19. Comparison between simulated and measured S-parameters at V_{DS} = 16 V, V_{GS} = -2.2 V. The Smith chart shows an excellent agreement between the measurement and simulation which verifies the validity of extracted small-signal parameters. The table summarizes the extracted intrinsic parameters and calculated f_T and f_{max}.

The f_T and f_{max} values were verified through S-parameter simulations in the 0.45 ~ 50 GHz range using Advanced Design Software (ADS) as shown in Figure 3-19. The small discrepancy between measurements and simulations in f_{max} (i.e. 300 GHz vs. 288 GHz) could originate from incomplete small-signal model used in this work. The relatively low f_T value is primarily due to a significantly lower-than-expected intrinsic RF transconductance of ~ 370 mS/mm (vs. calculated intrinsic transconductance from Figure 3-15(b) ~ 490 mS/mm). This phenomenon is called RF g_m-collapse and it will be discussed in Chapter 4.

3.4. References

[1] F. Schwierz and J. J. Liou, "Modern microwave transistors: theory, design, and performance," Wiley, 2003.

[2] M. Higashiwaki, T. Mimura, and T. Matsui, "AlGaN/GaN heterostructure field-effect transistors on 4H-SiC substrates with current-gain cutoff frequency of 190 GHz," *Appl. Phys. Express*, vol. 1, 021103, Feb. 2008.

[3] M. Higashiwaki, T. Mimura, and T. Matsui, "30-nm-gate AlGaN/GaN heterostructure field-effect transistors with a current-gain cutoff frequency of 181 GHz," *Jpn. J. Appl. Phys.*, vol. 45, no. 42, pp. L1111-L1113, Oct. 2006.

[4] S. M. Sze, "Physics of semiconductor devices," Wiley-Interscience, 1981.

[5] D.-H. Kim and J. A. del Alamo, "30-nm InAs pseudomorphic HEMTs on an InP substrate with a current-gain cutoff frequency of 628 GHz," *IEEE Electron Device Lett.*, vol. 29, no. 8, pp. 830-833, Aug. 2008.

[6] P. J. Tasker and B. Hughes, "Importance of source and drain resistance to the maximum f_T of millimeter-wave MODFET's," *IEEE Electron Device Lett.*, vol. 10, no. 7, pp. 291-293, July 1989.

[7] Nidhi, T. Palacios, A. Chakraborty, S. Keller, and U. K. Mishra, "Study of impact of access resistance on high-frequency performance of AlGaN/GaN HEMTs by measurements at low temperatures," *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 877-880, Nov. 2006.

[8] D. Buttari, A. Chini, G. Meneghesso, E. Zanoni, B. Moran, S. Heikman, N. Q. Zhang, L. Shen, R. Coffie, S. P. DenBaars, and U. K. Mishra, "Systematic characterization of Cl_2 reactive ion etching for improved ohmics in AlGaN/GaN HEMTs," *IEEE Electron Device Lett.*, vol. 23, no. 2, pp. 76-78, Feb. 2002.

[9] D. R. Greenberg and J. A. del Alamo, "Nonlinear source and drain resistance in recessed-gate heterostructure field-effect transistors," *IEEE Trans. Electron Devices*, vol. 43, no. 8, pp. 1304-1306, Aug. 1996.

[10] D. K. Schroder, "Semiconductor material and device characterization," Wiley, 2006.

[11] T. Palacios, S. Rajan, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "Influence of the dynamic access resistance in the g_m and f_T linearity of AlGaN/GaN HEMTs," *IEEE Trans. Electron Devices*, vol. 52, no. 10, pp. 2117-2123, Oct. 2005.

[12] Z. Fan, S. N. Mohammad, W. Kim, O. Aktas, A. E. Botchkarev, and H. Morkoc, "Very low resistance multilayer ohmic contact to n-GaN," *Appl. Phys. Lett.*, vol. 68, no. 12, pp. 1672-1674, Mar. 1996.

[13] J. W. Chung, W. E. Hoke, E. M. Chumbes, and T. Palacios, "AlGaN/GaN HEMT with 300 GHz f_{max}," *IEEE Electron Device Lett.*, vol. 31, no. 3, pp. 195-197, Mar. 2010.

[14] F. M. Mohammed, L. Wang, I. Adesida, and E. Piner, "The role of barrier layer on ohmic performance of Ti/Al-based contact metallizations on AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 100, no. 2, 023708, Jul. 2006.

[15] L. Wang, D.-H. Kim, and I. Adesida, "Direct contact mechanism of ohmic metallization to AlGaN/GaN heterostructures via ohmic area recess etching," *Appl. Phys. Lett.*, vol. 95, no. 17, 172107, Oct. 2009.

[16] R. Gong, J. Wang, S. Liu, Z. Dong, M. Yu, C. P. Wen, Y. Cai, and B. Zhang, "Analysis of surface roughness in Ti/Al/Ni/Au ohmic contact to AlGaN/GaN high electron mobility transistors," *Appl. Phys. Lett.*, vol. 97, no. 6, 062115, Aug. 2010.

[17] H. C. Lee, J. W. Bae, and G. Y. Yeom, "Thermally stable Ti/Al/W/Au multilayer ohmic contacts on n-type GaN," *J. Kr. Phy. Soc.*, vol. 51, no. 3, pp. 1046-1049, Sept. 2007.

[18] B. Luo, F. Ren, R. C. Fitch, J. K. Gillespie, T. Jenkins, J. Sewell, D. Via, A. Crespo, A. G. Baca, R. D. Briggs, D. Gotthold, R. Birkhahn, B. Peres, and S. J. Pearton, "Improved morphology for ohmic contacts to AlGaN/GaN high electron mobility transistors using WSi_x- or W-based metallization," *Appl. Phys. Lett.*, vol. 82, no. 22, pp. 3910-3912, Jun. 2003.

[19] V. Desmaris, J.-Y. Shiu, C.-Y Lu, N. Rorsman, H. Zirath, E.-Y. Chang, "Transmission electron microscopy assessment of the Si enhancement of Ti/Al/Ni/Au ohmic contacts to undoped AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 100, no. 3, 034904, Aug. 2006.

[20] N. Onojima, N. Hirose, T. Mimura, and T. Matsui, "Effect of Si deposition on AlGaN barrier surfaces in GaN heterostructure field-effect transistors," *Appl. Phys. Express.*, vol. 1, 071101, Jun. 2008.

[21] C. Youn and K. Kang, "Low-resistance ohmic contacts to AlGaN/GaN heterostructure using Si/Ti/Al/Cu/Au multilayer metal scheme," *Jpn. J. Appl. Phys.*, vol. 39, no. 7A, pp. 3955-3956, Jul. 2000.

[22] J. D. Plummer, M. D. Deal, and P. B. Griffin, "Silicon VLSI technology: fundamentals, practice, and modeling," Prentice Hall, 2000.

[23] P. H. Ladbrooke, "MMIC design: GaAs FETs and HEMTs," Artech House, 1989.

[24] D. A. Hodges, H. G. Jackson, and R. A. Saleh, "Analysis and design of digital integrated circuits," McGraw-Hill, 2004.

[25] D. F. Storm, D. S. Katzer, S. C. Binari, E. R. Glaser, B. V. Shanabrook, and J. A. Roussos, "Reduction of buffer layer conduction near plasma-assisted molecular-beam epitaxy grown GaN/AlN interfaces by beryllium doping," *Appl. Phys. Lett.*, vol. 81, no. 20, pp. 3819-3821, Nov. 2002.

[26] T. Palacios, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "AlGaN/GaN high electron mobility transistors with InGaN back-barriers," *IEEE Electron Dev. Lett.*, vol. 27, no. 1, pp. 13-15, Jan. 2006.

[27] H. Xing, Y. Dora, A. Chini, S. Heikman, S. Keller, and U. K. Mishra, "High breakdown voltage AlGaN-GaN HEMTs achieved by multiple field plates," *IEEE Electron Dev. Lett.*, vol. 25, no. 4, pp. 161-163, Apr. 2004.

[28] K. Shinohara, A. Corrion, D. Regan, I. Milosavljevic, D. Brown, S. Burnham, P. J. Willadsen, C. Butler, A. Schmitz, D. Wheeler, A. Fung, and M. Micovic, "220GHz f_T and 400GHz f_{max} in 40-nm GaN DH-HEMTs with re-grown ohmic," *IEEE IEDM Tech. Digest*, pp. 672-675, 2010.

[29] H. Matsuzaki, T. Maruyama, T. Koasugi, H. Takahashi, M. Tokumitsu, and T. Enoki, "Lateral scale down of InGaAs/InAs composite-channel HEMTs with tungsten-based tiered ohmic structure for 2-S/mm g_m and 500-GHz f_T," *IEEE Trans. Electron. Devices*, vol. 54, no. 3, pp. 378-384, Mar. 2007.

Chapter 4. Advanced Gate Technologies for Improving f_T

4.1. Introduction

GaN-based high electron mobility transistors (HEMTs) have emerged as one of the prime candidates for solid-state power amplifiers at frequencies above 30 GHz. Much of the excitement about its prospect comes from the unique combination of high electron velocity (v_{peak} ~ 2.5 × 10^7 cm/s) and high breakdown electric field (~ 3.3 MV/cm) [1][2]. In spite of the recent progress in maximum operating frequency [3][4], there are still important issues that need to be overcome to further extend operating frequencies of GaN HEMTs to mm-wave ranges (30 ~ 300 GHz). One of the biggest challenges in mm-wave GaN HEMTs is that the typical frequency response of these devices, as measured by the current-gain cutoff frequency (f_T), is lower than what the intrinsic material properties predict. In principle, deep-submicron AlGaN/GaN HEMTs (e.g. L_g < 50 nm) are expected to show f_T of more than 200 GHz. However, much lower f_T values have been frequently seen in the literature, for example f_T = 139 GHz for L_g = 45 nm [5], f_T = 145 GHz for L_g = 25 nm [6], etc. In Chapter 3, we reported AlGaN/GaN HEMTs fabricated at MIT with L_g = 60 nm and just f_T = 70 GHz, although they demonstrated a record f_{max} of 300 GHz by minimizing parasitic elements. The goal of this chapter is to understand what is limiting the frequency performance of AlGaN/GaN HEMTs, and then develop new approaches to improve f_T.

Several hypotheses have been proposed to explain the lower-than-expected f_T in GaN HEMTs. Jessen *et al.* [7] and Guerra *et al.* [8] indicated that f_T of AlGaN/GaN HEMTs is mainly limited by the short-channel effects. To avoid undesirable short-channel effects, a high aspect ratio (the ratio between gate length (L_g) and barrier thickness (t_b), namely L_g / t_b) needs to be maintained and they estimated a required aspect ratio of 10 ~ 15 for GaN HEMTs from empirical model and Monte Carlo simulation. This ratio is much higher than the aspect ratio required in InGaAs devices (L_g / t_b = 4 ~ 7.5) to suppress short-channel effects. Thus, they concluded that the reason for the lower-than-expected f_T in GaN HEMTs is the stringent requirement of the aspect ratio, which is difficult to meet in most submicron GaN HEMTs due to the thick barrier layer.

An alternative explanation is based on the concept of modulation efficiency proposed by Foisy *et al.* [9] and Nguyen *et al.* [10]. It describes charge transfer in the situation where a change in the gate voltage does not exclusively result in a change in the 2DEG charge concentration. The modulation efficiency (η) represents how efficiently a small change in the gate voltage (modulation of the gate signal) can produce a change in the drain current. In other words, η indicates how many charges can effectively contribute to the drain current out of the total modulated charges (Q_{tot}). When an ac signal is applied to the gate, the modulated gate charges are imaged at electrically active sites in the device (e.g. channel electrons, interface states, traps in AlGaN barrier, etc). Only the mobile charge in the channel (Q_s) contributes to the drain current, thus the modulation efficiency (η) can be defined as

$$\eta = \frac{\partial Q_s}{\partial Q_{tot}} \tag{4-1}$$

So far we assumed $\eta = 1$ (or $Q_s = Q_{tot}$), however in practical device operations η can be less than 1 (or $Q_s < Q_{total}$) and now we can rewrite g_m and f_T incorporating η as

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = v_{sat}\frac{\partial Q_s}{\partial V_{GS}} = v_{sat}\frac{\partial Q_{tot}}{\partial V_{GS}}\left(\frac{\partial Q_s}{\partial Q_{tot}}\right) = v_{sat}C\eta \tag{4-2}$$

$$f_T = \frac{g_m}{2\pi C_g} = \frac{v_{sat}C\eta}{2\pi(CL_g)} = \frac{v_{sat}}{2\pi L_g}\eta \tag{4-3}$$

Foisy *et al.* and Nguyen *et al.* assumed electrically active sites such as interface states, traps, and defects can respond to the high frequency signals and η can be largely degraded by parasitic charge modulations which do not contribute to the drain current. AlGaN/GaN structures are known to have much higher trap density (or impurity levels) than other material systems [11][12] and these inefficient charge modulations could affect its high frequency characteristics.

This chapter studies in detail these two hypotheses on state-of-the-art GaN transistors. We will show that a large portion of the poor frequency performance of AlGaN/GaN HEMTs can be attributed to the drop in the intrinsic small-signal transconductance (RF g_m) with respect to the intrinsic DC g_m in many of these devices. To mitigate this RF g_m-collapse, the interface between the gate metal and the AlGaN surface needs to be carefully engineered. Harmonious scaling of vertical and lateral dimension of the gate is also important to maximize frequency performance by suppressing short-channel effects. In this regard, we have developed a new gate technology

that solves the RF g_m-collapse problem and the short-channel effects in order to demonstrate state-of-the-art f_T in AlGaN/GaN HEMTs.

4.2. Small-signal parameter extraction

To understand the RF characteristics of AlGaN/GaN HEMTs, it is necessary to accurately extract their small-signal equivalent circuit parameters and verify their frequency response. For example, f_T and f_{max} are determined by several small-signal parameters (Figure 2-9):

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})(1 + (R_s + R_d)g_o) + g_m C_{gd}(R_s + R_d)} \tag{4-4}$$

$$f_{max} \cong \frac{f_T}{2\sqrt{(R_i + R_s + R_g)g_o + (2\pi f_T)R_g C_{gd}}} \tag{4-5}$$

By extracting each small-signal parameter, we can distinguish contributions of each element to f_T and f_{max}. Thus, an accurate extraction of small-signal parameters is very useful for identifying problems in device operation, designing new structures, evaluating process technologies, and optimizing device performance.

Several methods have been proposed to extract small-signal parameters of microwave FETs from measured S-parameters [13]. However, due to non-conventional properties of submicron AlGaN/GaN HEMTs such as high gate leakage current [14] and dynamic access resistances [15], the existing extraction methods are not accurate and need to be adjusted accordingly. In this section, we introduce a new method for the extraction of the intrinsic small-signal parameters of AlGaN/GaN HEMTs (C_{gs}, C_{gd}, C_{ds}, R_i, R_{gd}, g_m, τ, and g_o) using a modified small-signal equivalent circuit, on-wafer open/short de-embedding, and the gate current injection technique [16] to measure the access resistances. This method is simple yet accurate, showing good agreement between the measured and simulated S-parameters over a large frequency range.

4.2.1. Modified small-signal equivalent circuit

When studying the high frequency performance of GaN HEMTs, it is important to choose an appropriate small-signal equivalent circuit model that accounts for all the unique physical phenomena observed in these devices. It is well known, for example, that AlGaN/GaN HEMTs have a large gate leakage current mainly due to the large density of traps and dislocations in this material system [17]. This gate leakage current is two to three orders of magnitude higher than that of AlGaAs/GaAs HEMTs [14]. The large gate leakage current of AlGaN/GaN HEMTs affects their high frequency characteristics and it is necessary to incorporate this effect into the small-signal equivalent circuit. Therefore, two additional resistances R_{lgd} (= $1/g_{lgd}$) and R_{lgs} (= $1/g_{lgs}$) were included into the intrinsic elements. It will be shown that the addition of R_{lgd} and R_{lgs} is effective to accurately model frequency behavior of the device later in this section. Figure 4-1 shows the complete small-signal equivalent circuit used to model AlGaN/GaN HEMTs in this work. This equivalent circuit can be divided into three parts:

(i) The intrinsic region consisting of C_{gs}, C_{gd}, C_{ds}, R_i, R_{gd}, R_{lgd}, R_{lgs}, g_m, τ, and g_o. These intrinsic elements are dependent on the applied bias conditions.

(ii) The active region consisting of the intrinsic region plus R_g, R_d, and R_s. R_g is constant while R_d and R_s are bias dependent due to the dynamic access resistance observed in AlGaN/GaN HEMTs [15]. In general, the f_T and f_{max} measured from the active region are reported in the literature and compared for the benchmark.

(iii) The extrinsic region consisting of L_g, L_d, L_s, C_{pgs}, C_{pgd}, and C_{pds}. The extrinsic elements are independent of the applied bias conditions.

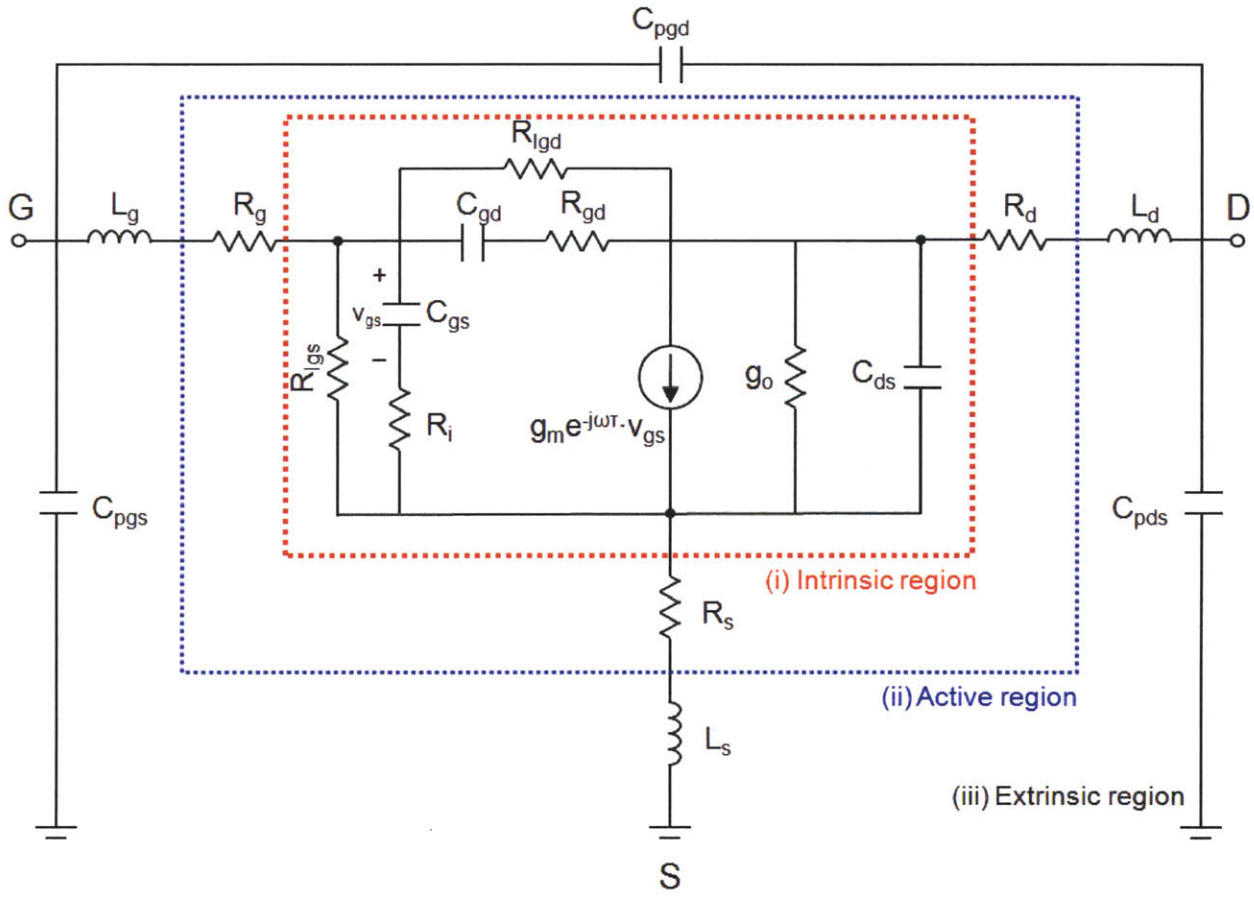


Figure 4-1. Complete small-signal equivalent circuit for an AlGaN/GaN HEMT used in this work. The circuit is divided into three regions: (i) Intrinsic region, (ii) active region, and (iii) extrinsic region. To extract S-parameters of the intrinsic region, parasitic pad components from the extrinsic region and gate, drain, and source resistances (R_g, R_d, and R_s) from the active region need to be de-embedded.

4.2.2. On-wafer de-embedding

To extract the different elements of the intrinsic small-signal equivalent circuit, the S-parameters were first measured with an Agilent Technologies N5230A network analyzer over the frequency range from 0.5 to 40 GHz. The network was calibrated with either a short-open-load-through (SOLT) or a Line-Reflect-Reflect-Match (LRRM) standard and the calibration was verified by ensuring that both S_{12} and S_{21} of the through standard are less than $\pm$0.01 dB and that both S_{11} and S_{22} are below -45 dB within the measured frequency range after the calibration. Since this calibration does not involve any device structure, the S-parameter measurement of the device includes the extrinsic part of the device such as pad capacitances and inductances as well

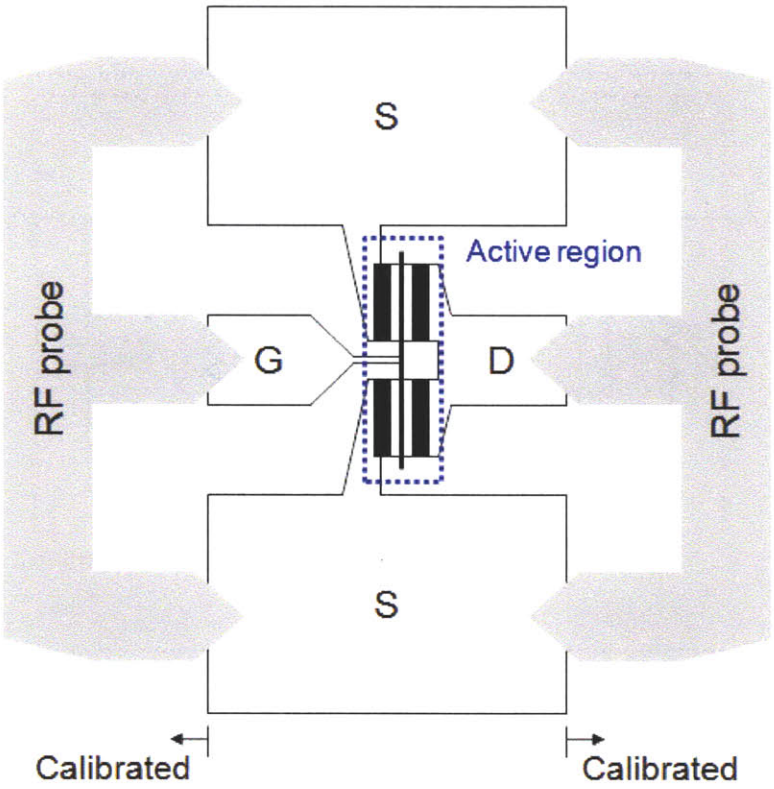


Figure 4-2. Top down layout of a two-finger T-shape AlGaN/GaN HEMT under RF measurement. Measured S-parameter includes the parasitic pad components of the extrinsic region.

as gate, drain, and source resistances as shown in Figure 4-2. Therefore, in order to obtain the S-parameters of the intrinsic device the extrinsic pad components and R_g, R_d, and R_s of the device need to be carefully de-embedded.

On-wafer open and short structures were used to de-embed the effect of parasitic pad capacitances and inductances. The open and short structures and the AlGaN/GaN HEMTs were fabricated at the same time on the same wafer. The open structure was formed by removing the active region of the device leaving only the pad structure, and the short structure is constructed by replacing the active region of the device with a metal layer, short-circuiting in this way the source, drain, and gate pads. It should be noted that it is very important to use identical dimensions for both the open and short structures, and for the measured devices (e.g. pad size, width of active region, source-drain distance for open structure) to perform accurate de-embedding. Otherwise, this error could significantly overestimate or underestimate the frequency characteristics of the device.

The top-down schematic layouts and model representation of the device under test (DUT), open, and short structures are described in Figure 4-3. The DUT was modeled by an impedance matrix of the active device (Z_A), and parasitic impedance (Z_1, Z_2, Z_3) and admittance (Y_1, Y_2, Y_3) of the extrinsic probing pad. It is noted that Z_A does not represent the intrinsic parameter yet since it still includes the impedances of gate, source, and drain resistances as shown in Figure 4-1. By comparing Figure 4-3(a) to Figure 4-3(b) and Figure 4-3(c), the impedance matrix of the active device Z_A is found from the following relationships

$$Z_A = Z[Y_z] - Z_{123} \tag{4-6}$$

where Y_z is the total admittance matrix of Z_A, Z_1, Z_2, and Z_3 and $Z[Y_z]$ is the impedance matrix transformed from its admittance matrix Y_z following the Y to Z conversion rule described in Appendix 1. Z_{123} is the total impedance matrix of Z_1, Z_2, and Z_3. Y_z and Z_{123} can be obtained using the open and short structures as

$$Y_z = Y_{DUT} - Y_{open} \tag{4-7}$$

$$Z_{123} = Z[Y_{short} - Y_{open}] \tag{4-8}$$

By substituting Eq. (4-7) and Eq. (4-8) into Eq. (4-6), the impedance matrix of the active device Z_A excluding pad components (pad capacitances and inductances) is obtained by

$$Z_A = Z\left[Y_{DUT} - Y_{open}\right] - Z[Y_{short} - Y_{open}] \tag{4-9}$$

As mentioned before, Z_A is not the intrinsic parameter due to R_g, R_d, and R_s. The impedance matrix of the intrinsic device Z_{int}, therefore, can be extracted by carefully subtracting these resistances as

$$Z_{int} = \begin{bmatrix} Z_{A,11} - R_g - R_s & Z_{A,12} - R_s \\ Z_{A,21} - R_s & Z_{A,22} - R_d - R_s \end{bmatrix} \tag{4-10}$$

Bias dependent dynamic access resistances R_d and R_s were measured using the gate current injection technique and constant R_g is extracted from the Z-parameters at the zero-biased condition (V_{ds} = 0 V and V_{gs} = 0 V) [18]. Finally, the obtained Z-parameter matrix of the intrinsic device Z_{int} was transformed to Y-parameter Y_{int} which was used to extract all the intrinsic small-signal elements in the next section.

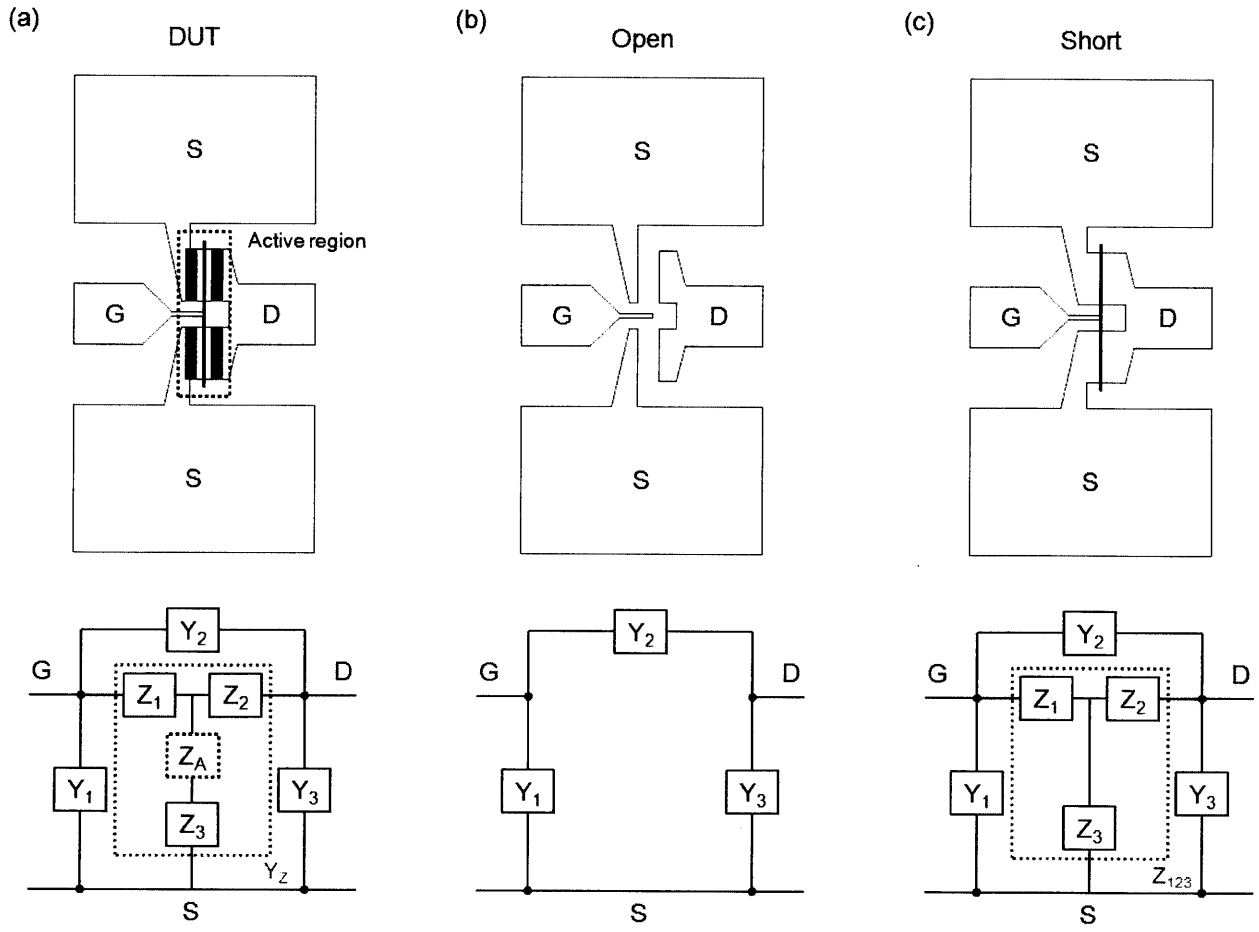


Figure 4-3. Geometry and model representation of (a) DUT, (b) open, and (c) short structures. The impedance matrix of the active region (Z_A) can be obtained by de-embedding extrinsic pad impedances (Z_1, Z_2, Z_3) and admittances (Y_1, Y_2, Y_3) by $Z_A = Z[Y_{DUT} - Y_{open}] - Z[Y_{short} - Y_{open}]$.

4.2.3. Intrinsic parameter extraction

The small-signal equivalent circuit elements of the DUT can be correlated to its Y-parameters through the following expressions [19].

$$Y_{int} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$$

$$D_1 = 1 + \omega^2 {C_{gs}}^2 {R_i}^2 \qquad D_2 = 1 + \omega^2 {C_{gd}}^2 {R_{gd}}^2$$

$$Y_{11} = g_{lgs} + g_{lgd} + \frac{\omega^2 R_i {C_{gs}}^2}{D_1} + \frac{\omega^2 R_{gd} {C_{gd}}^2}{D_2} + j\omega\left(\frac{C_{gs}}{D_1} + \frac{C_{gd}}{D_2}\right)$$

$$Y_{12} = -g_{lgd} - \frac{\omega^2 R_{gd} {C_{gd}}^2}{D_2} - j\omega \frac{C_{gd}}{D_2} \tag{4-11}$$

$$Y_{21} = -g_{lgd} + \frac{g_m e^{-j\omega\tau}}{1 + j\omega R_i C_{gs}} - j \frac{\omega C_{gd}}{1 + j\omega R_{gd} C_{gd}}$$

$$Y_{22} = g_{lgd} + g_o + \frac{\omega^2 R_{gd} {C_{gd}}^2}{D_2} + j\omega \left(C_{ds} + \frac{C_{gd}}{D_2} \right)$$

By separating these complex equations into their real and imaginary parts, ten intrinsic elements can be determined as listed below. These equations are closed form expressions and do not require any excessive computation. Detailed mathematical derivations are given in [19].

$$g_{lgd} = -Re(Y_{12}) \qquad \text{@ low frequencies (e.g. 50 MHz)}$$

$$g_{lgs} = Re(Y_{11}) + Re(Y_{12}) \qquad \text{@ low frequencies (e.g. 50 MHz)}$$

$$C_{gs} = \frac{Im(Y_{11}) + Im(Y_{12})}{\omega} \left[1 + \left(\frac{Re(Y_{11}) + Re(Y_{12}) - g_{lgs}}{Im(Y_{11}) + Im(Y_{12})} \right)^2 \right]$$

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

$$C_{ds} = \frac{Im(Y_{22}) + Im(Y_{12})}{\omega} \tag{4-12}$$

$$R_i = \frac{Re(Y_{11}) + Re(Y_{12}) - g_{lgs}}{\omega C_{gs} \left(Im(Y_{11}) + Im(Y_{12}) \right)}$$

$$R_{gd} = \frac{Re(Y_{12}) + g_{lgd}}{\omega C_{gd} Im(Y_{12})}$$

$$g_o = Re(Y_{22}) + Re(Y_{12})$$

$$g_m = \sqrt{\left(\left(Re(Y_{21}) - Re(Y_{12}) \right)^2 + \left(Im(Y_{21}) - Im(Y_{12}) \right)^2 \right) \left(1 + \omega^2 {C_{gs}}^2 {R_i}^2 \right)}$$

$$\tau = \frac{1}{\omega} \sin^{-1} \left(\frac{Im(Y_{12}) - Im(Y_{21}) - \omega C_{gs} R_i \left(Re(Y_{21}) - Re(Y_{12}) \right)}{g_m} \right)$$

To verify the accuracy of above-described method, extracted intrinsic small-signal equivalent circuit parameters were used to reproduce measured S-parameters of AlGaN/GaN HEMTs over a frequency range from 0.5 to 40 GHz using Advanced Design Software (ADS). One of the

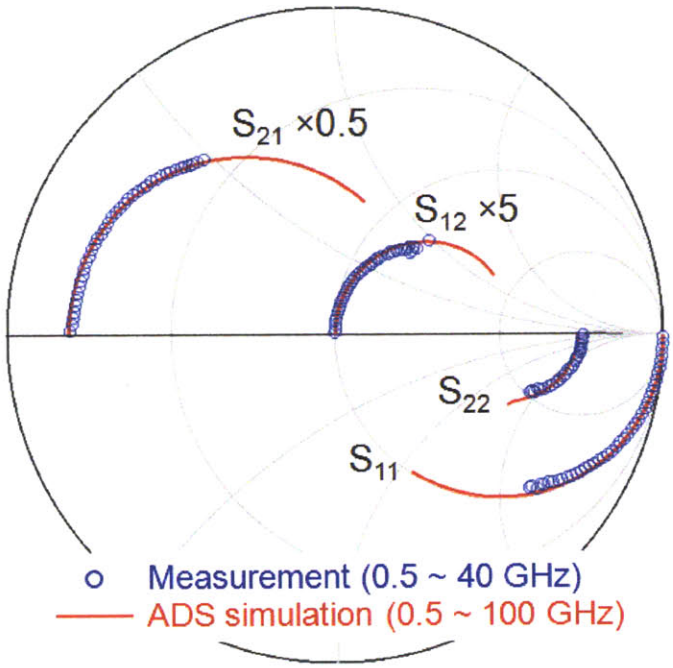


Figure 4-4. Smith chart illustrating an excellent agreement between measured and simulated S-parameters of an AlGaN/GaN HEMT with L_g = 140 nm.

fabricated AlGaN/GaN HEMTs with L_g = 140 nm which showed f_T and f_{max} of 110 GHz and 102 GHz, respectively was used for this verification. Figure 4-4 shows the Smith chart illustrating an excellent agreement between measured and simulated S-parameters. Also, several figures of merit in high frequency performance including forward current gain $|h_{21}|^2$, Mason's unilateral gain U, maximum stable/available gain *MSG/MAG*, stability factor k, f_T and f_{max} were well predicted over a large frequency range as shown in Figure 4-5. It is highlighted that the addition of R_{lgd} and R_{lgs} into the small-signal equivalent circuit plays a significant role to accurately model the frequency characteristics of the device. As shown in Figure 4-6, without R_{lgd} and R_{lgs} it is not possible to precisely model the frequency behavior of the device especially at frequencies below 5 GHz. Therefore, the effect of a large gate leakage current in AlGaN/GaN HEMTs has to be incorporated into the small-signal equivalent circuit model. The modified small-signal equivalent circuit and the extraction method described in this section produce accurate intrinsic small-signal elements which is very important to understand frequency behavior of AlGaN/GaN HEMTs.

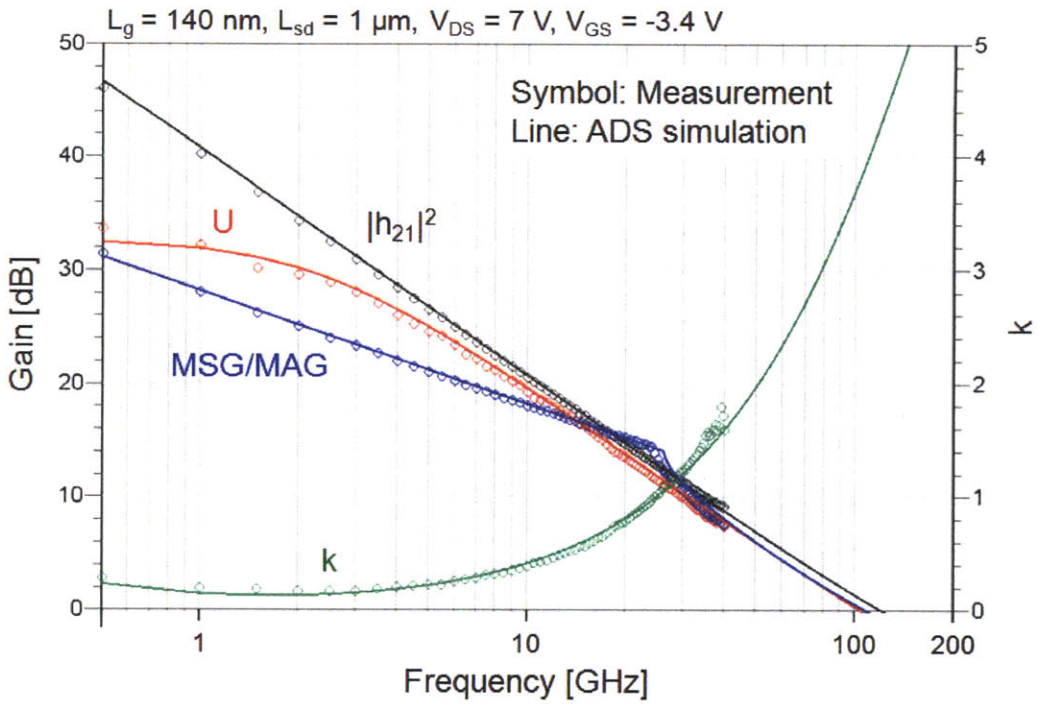


Figure 4-5. Several high frequency characteristics ($|h_{21}|^2$, U, *MSG/MAG*, k, f_T, and f_{max}) showing an excellent agreement between measured and simulated S-parameters of an AlGaN/GaN HEMT with L_g = 140 nm.

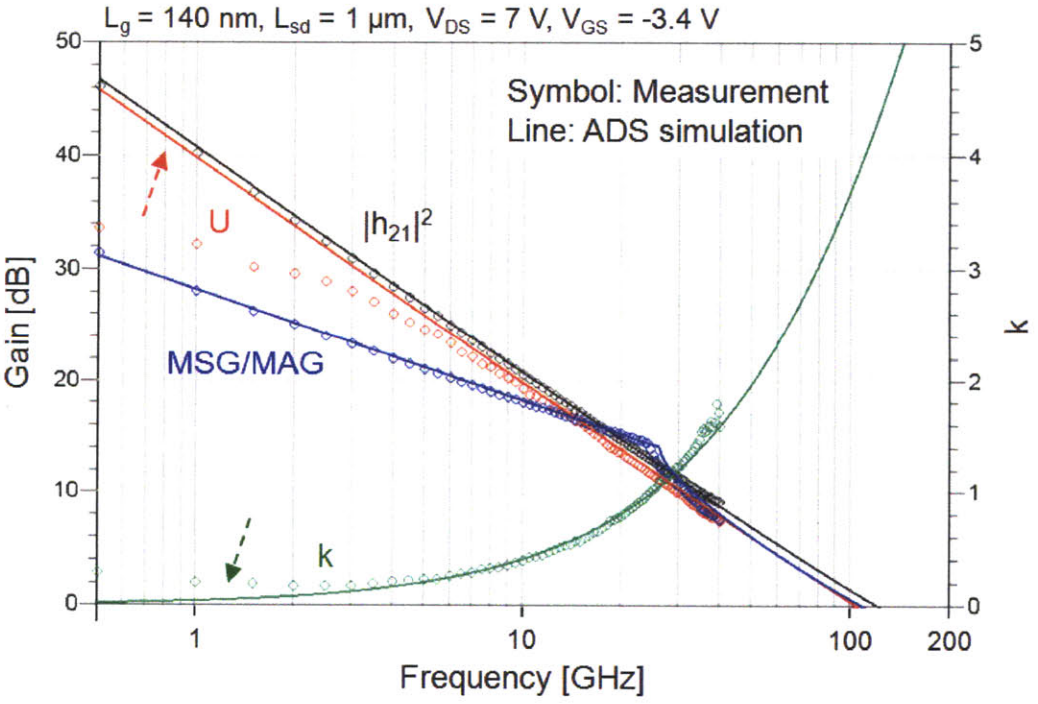


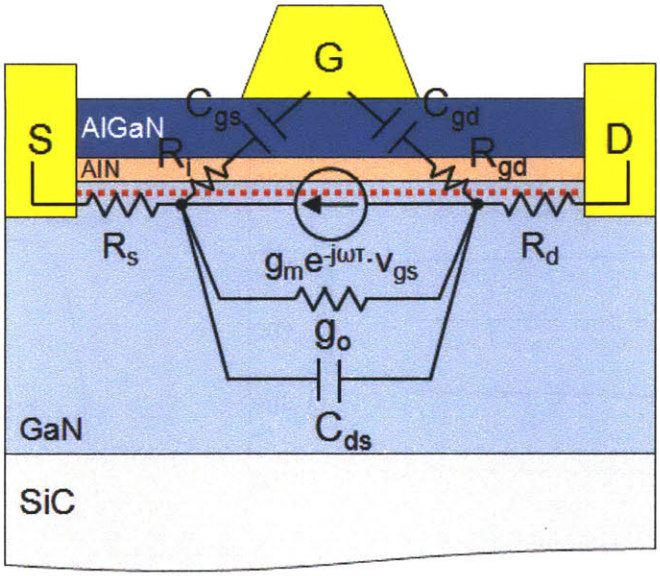
Figure 4-6. Importance of R_{lgd} and R_{lgs} to accurately model measured S-parameters. Without them (leakage current), it is not possible to precisely model the frequency behavior of the AlGaN/GaN HEMTs especially at frequencies below 5 GHz.

4.3. RF g_m-collapse

To understand the origin of the lower-than-expected f_T in AlGaN/GaN HEMTs, we carefully extracted intrinsic small-signal parameters following the steps described in the previous section. Each intrinsic parameter was thoroughly investigated and special attention was paid to the intrinsic transconductance (g_m) and gate capacitances (C_{gs} and C_{gd}) due to their large contribution to the f_T expression as shown in Eq. (2-29). Interestingly, we observed significant degradation of g_m at high frequency while other intrinsic parameters maintained reasonable values. This intriguing phenomenon, known as RF g_m-collapse, will be the main focus of this section. We attribute the lower-than-expected f_T in AlGaN/GaN HEMTs to RF g_m-collapse. To explain the origin of the RF g_m-collapse nitrogen vacancy (V_N)-related defects at the AlGaN surface are proposed as possible origin of RF g_m-collapse. To suppress RF g_m-collapse, we applied several plasma treatment technologies to the AlGaN surface and found that the oxygen plasma treatment is a very effective way to improve the RF g_m-collapse and thus f_T. The proposed oxygen plasma treatment significantly improves the DC and RF characteristics of AlGaN/GaN HEMTs.

4.3.1. Summary of observations

Figure 4-7 shows the fabricated standard AlGaN/GaN HEMTs studied in this section. AlGaN/GaN heterostructures were grown on SiC substrates by metal-organic chemical vapor deposition (MOCVD) at Cree, Inc. In these samples, the AlGaN barrier had a total thickness of 24 nm and a 1 nm AlN interlayer between the GaN and the AlGaN barrier was used to improve the surface morphology and to increase the electron mobility in the channel by reducing the alloy related interface roughness and scattering [20]. A rectangular gate was fabricated instead of a T-gate to simplify the process flow and the devices were not passivated to better estimate expected intrinsic parameters such as g_m, C_{gs} and C_{gd} without considering additional effects by the passivation layer.



$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})(1 + (R_s + R_d)g_o) + g_m C_{gd}(R_s + R_d)}$$

Figure 4-7. A schematic cross-section of a standard AlGaN/GaN HEMT with the small-signal equivalent circuit elements studied in this section.

As described in Chapter 2, a standard AlGaN/GaN HEMT technology does not involve any gate engineering steps such as insertion of gate dielectrics, gate-recess, plasma treatment, etc. In other words, the metal gate is directly deposited on the AlGaN surface forming Schottky metal-semiconductor junction. We confirmed normal DC behaviors of the fabricated devices and Figure 4-8 shows DC characteristics of a typical device with L_g = 200 nm and L_{sd} = 2 μm. Although the DC performance was reasonable for the device geometry, the RF performance was lower than expected as shown in Figure 4-9. Measured f_T was just 50 GHz while the expected value exceeds 80 GHz. Measured f_{max} was also low mainly due to the low f_T. The short-channel effects do not play a significant role in this device owing to the high aspect ratio of 8. This is confirmed by the low drain induced barrier lowering (DIBL) of 25 mV/V extracted from the subthreshold characteristics in this device. Therefore, a different mechanism is limiting the frequency performance of the device. To identify the origin of the lower-than-expected f_T, the intrinsic small-signal parameters were carefully extracted. As shown in Figure 4-10, the accuracy of extracted intrinsic small-signal parameters was confirmed by ADS simulation.

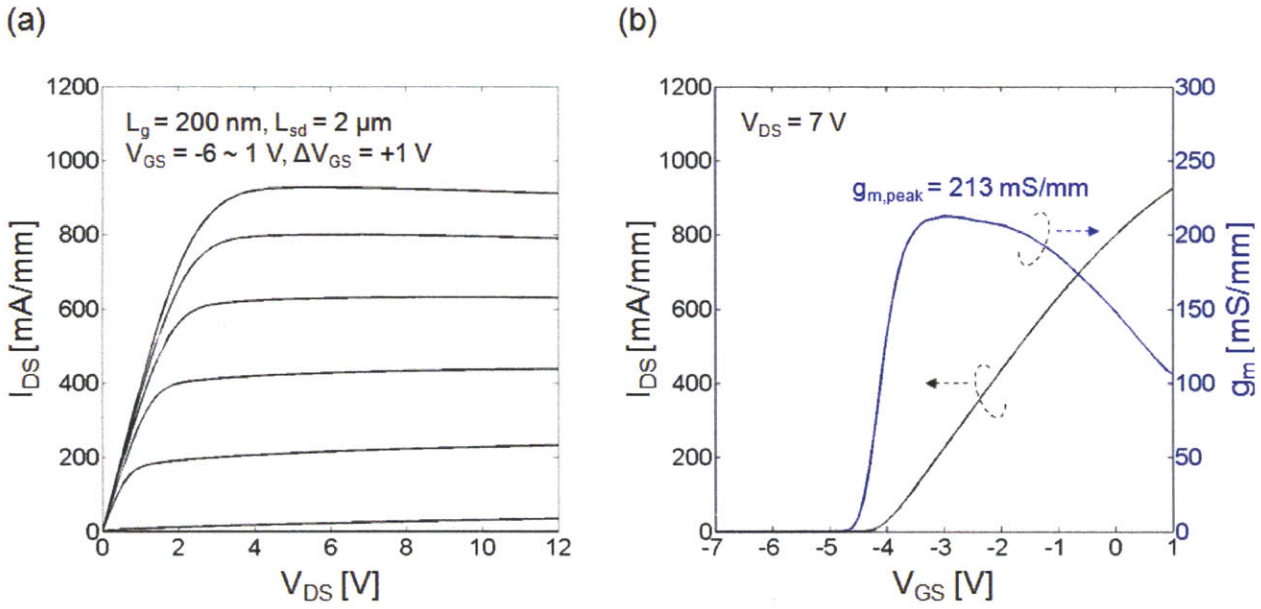


Figure 4-8. DC characteristics of the fabricated AlGaN/GaN HEMT with L_g = 200 nm and L_{sd} = 2 μm. (a) I-V and (b) transfer characteristics.

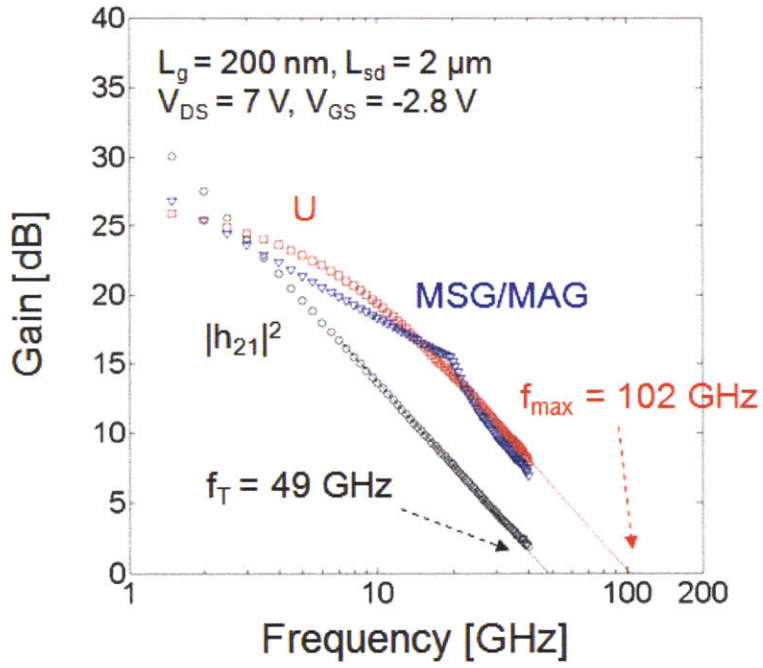


Figure 4-9. RF characteristics of the fabricated AlGaN/GaN HEMT with L_g = 200 nm and L_{sd} = 2 μm. Lower-than-expected f_T of 49 GHz was measured.

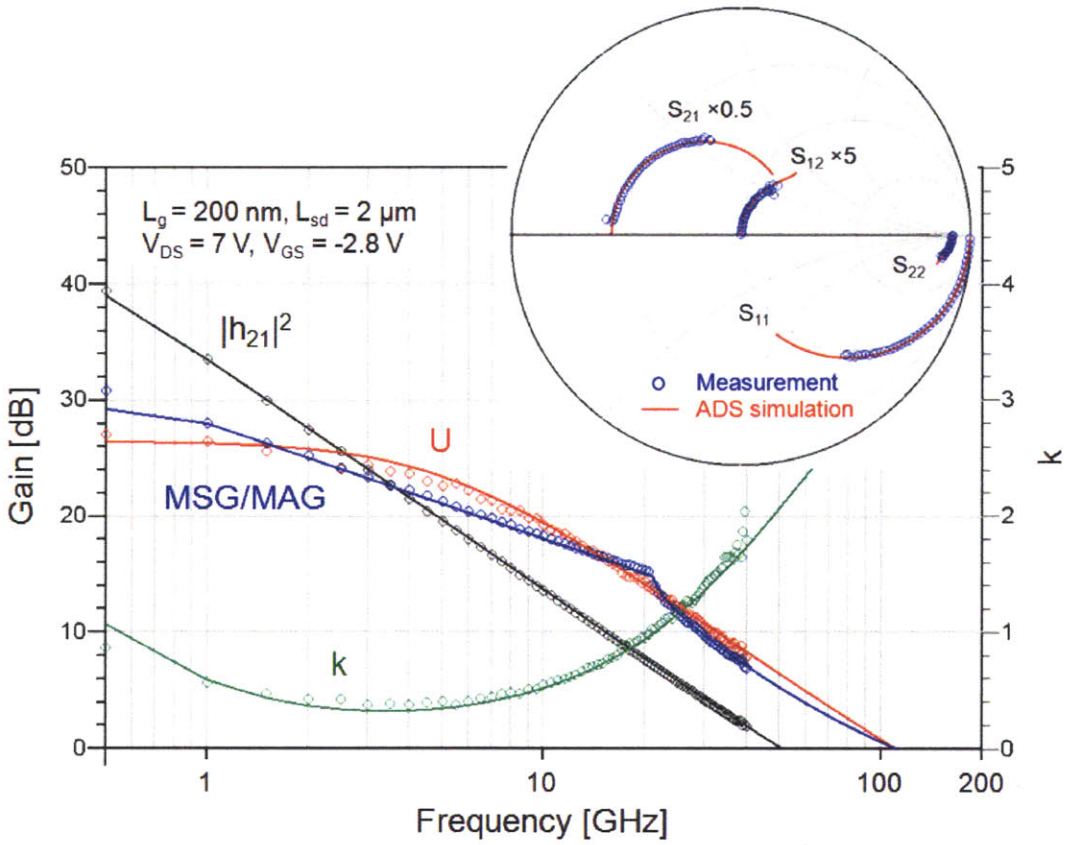


Figure 4-10. Intrinsic small-signal parameters were carefully extracted and the accuracy was verified by ADS simulation. A good agreement between the measured and simulated S-parameters was confirmed.

Figure 4-11 compares extracted intrinsic C_{gs}, C_{gd}, g_m, and g_o to their expected values. It is noted that the intrinsic parameters, in the ideal case, should be independent of the measured frequency. The expected intrinsic C_{gs} and C_{gd} normalized to the gate width were estimated from the expressions in Chapter 3,

$$C_{gs} = \frac{2}{3}\left(\varepsilon_b \frac{L_g}{t_b}\right) \tag{4-13}$$

$$C_{gd} = \frac{2\varepsilon_b}{1 + \frac{2X_{dep}}{L_g}} \tag{4-14}$$

where ε_b is the dielectric constant of the barrier, L_g is the gate length, and t_b is the thickness of the barrier. The factor of 2/3 in the expression of C_{gs} accounts for the charge partitioning in the device when this is operating in the saturation region [21]. X_{dep} represents the extension of the depletion region in the drain access region. X_{dep} is assumed to be 0.5 μm at given bias condition (V_{DG} ~ 11 V) [22]. Although X_{dep} is not rigorously verified, its variation does not significantly

affect to f_T (or lower-than-expected f_T) due to relatively small contribution of C_{gd} to f_T. Strictly speaking, the extracted intrinsic C_{gs} and C_{gd} include external fringing capacitances which are not considered in these two equations. Thus, as will be shown in Section 4.5.2., these fringing capacitances were extracted ($C_{gs,ex}$ and $C_{gd,ex}$) and added to the Eq. (4-13) and Eq. (4-14). The expected intrinsic transconductance at high frequency (RF g_m or $g_{m,i}$) was calculated from the measured extrinsic DC g_m by taking into account R_s, R_d, and g_o [23].

$$g_{m,i} = \frac{g_m(1 + (R_s + R_d)g_o)}{1 - g_m R_s} \tag{4-15}$$

As shown in Figure 4-11, we found that extracted intrinsic $g_{m,i}$ was significantly lower than its expected value, while extracted intrinsic C_{gs} and C_{gd} were well within the expected ranges. It is noted that g_o was not analytically calculated, however it was also well within the reasonable range considering the slope of the IV characteristics in Figure 4-8(a) at the bias point. Thus, only the intrinsic $g_{m,i}$ largely deviates from its expected value. Since f_T is proportional to $g_{m,i}$, f_T becomes largely degraded by this collapse of $g_{m,i}$ at high frequency, the RF g_m-collapse. Therefore, we attribute the lower-than-expect f_T in our devices to RF g_m-collapse. The RF g_m-collapse was repeatedly observed in several wafers from four different vendors although the amount of collapse varied. Thus, we believe that the RF g_m-collapse plays an important role in

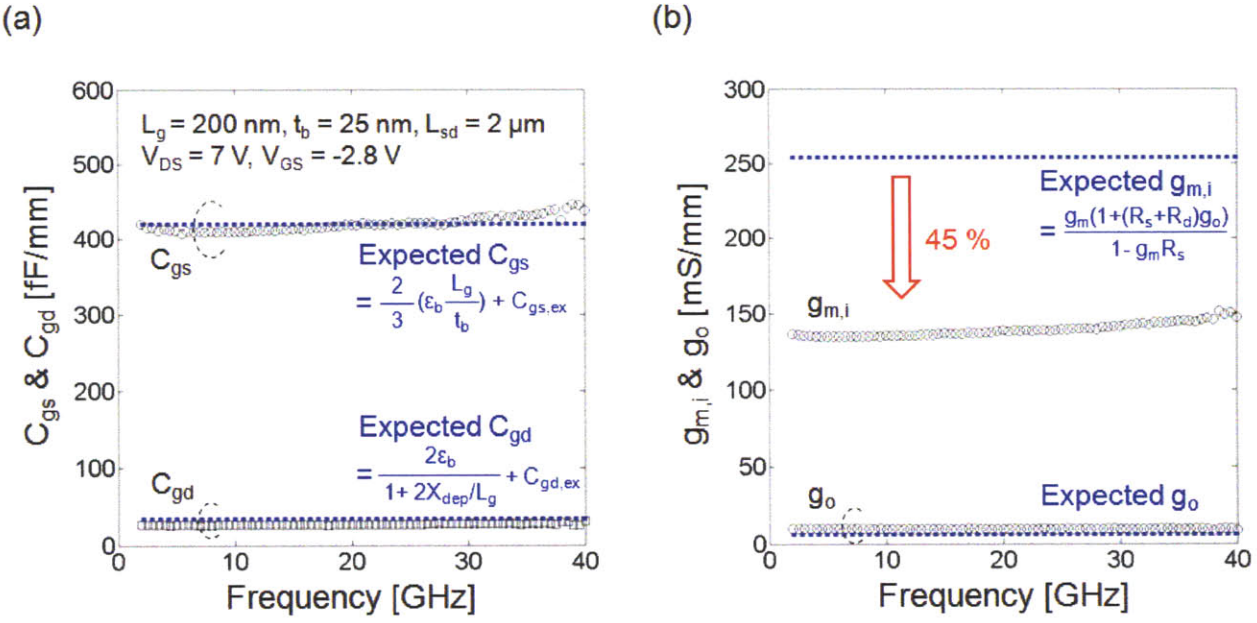


Figure 4-11. Comparison between extracted small signal parameters and their expected values. (a) C_{gs} and C_{gd}, (b) $g_{m,i}$ and g_o.

explaining the large variation of f_T's reported in the literature at a given gate length.

4.3.2. Effect of plasma treatment

The RF g_m-collapse was found to be strongly affected by surface plasma treatments performed in the gate region immediately before gate metal deposition. As shown in Figure 4-12, several plasma treatments were applied to the gate region. Typical plasma treatments used in GaN device processing such as Cl_2/BCl_3, CF_4, and O_2 plasma treatments were tested. The processing conditions of each plasma treatment were carefully optimized to prevent any negative effect in the material properties underneath the gate (Table 4-1). For example, the power of each plasma treatment was kept very low to avoid degradation in the channel transport often caused by plasma induced damage or ion bombardment. After the plasma treatment, Ni/Au gate metals were deposited and the RF g_m-collapse was investigated in each sample.

Figure 4-13 compares expected and extracted intrinsic g_m of each plasma treated sample as a function of measured frequency. As expected, a significant drop of RF g_m was observed in the device without treatment. Although the Cl_2/BCl_3 plasma treatment increased DC g_m by slightly recessing the gate region, it could not recover RF g_m to its expected level. On the other hand, devices with CF_4 and O_2 plasma treatment largely improved RF g_m compared to the one without treatment and successfully eliminated RF g_m-collapse.

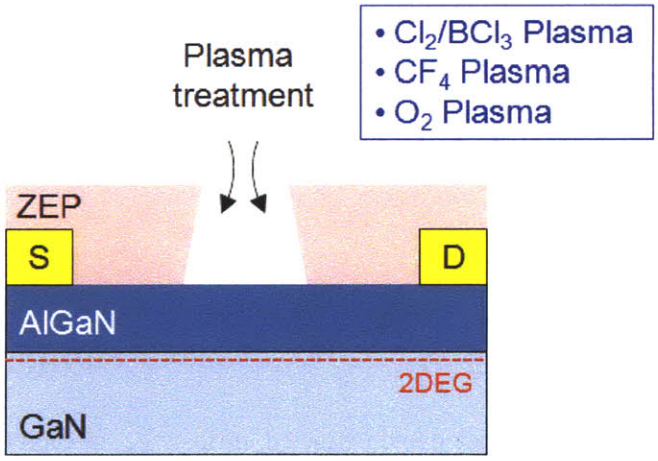


Figure 4-12. Several plasma treatments were applied before gate metallization to investigate their effects to RF g_m-collapse.

*Temperature = 25 °C

Plasma Treatment	System	Condition
Cl_2/BCl_3 Plasma	PlasmaQuest ECR-RIE	Cl_2/BCl_3 = 5/15 sccm, Pressure = 4 mtorr, ECR /RF = 100 W/60 V
CF_4 Plasma	PlasmaQuest ECR-RIE	CF_4 = 25 sccm, Pressure = 4 mtorr, ECR /RF = 100 W/60 V
O_2 Plasma	Branson IPC plasma asher	Pressure < 0.4 torr Power = 800 W

Table 4-1. Optimized conditions to study effect of different plasma treatment to RF g_m-collapse.

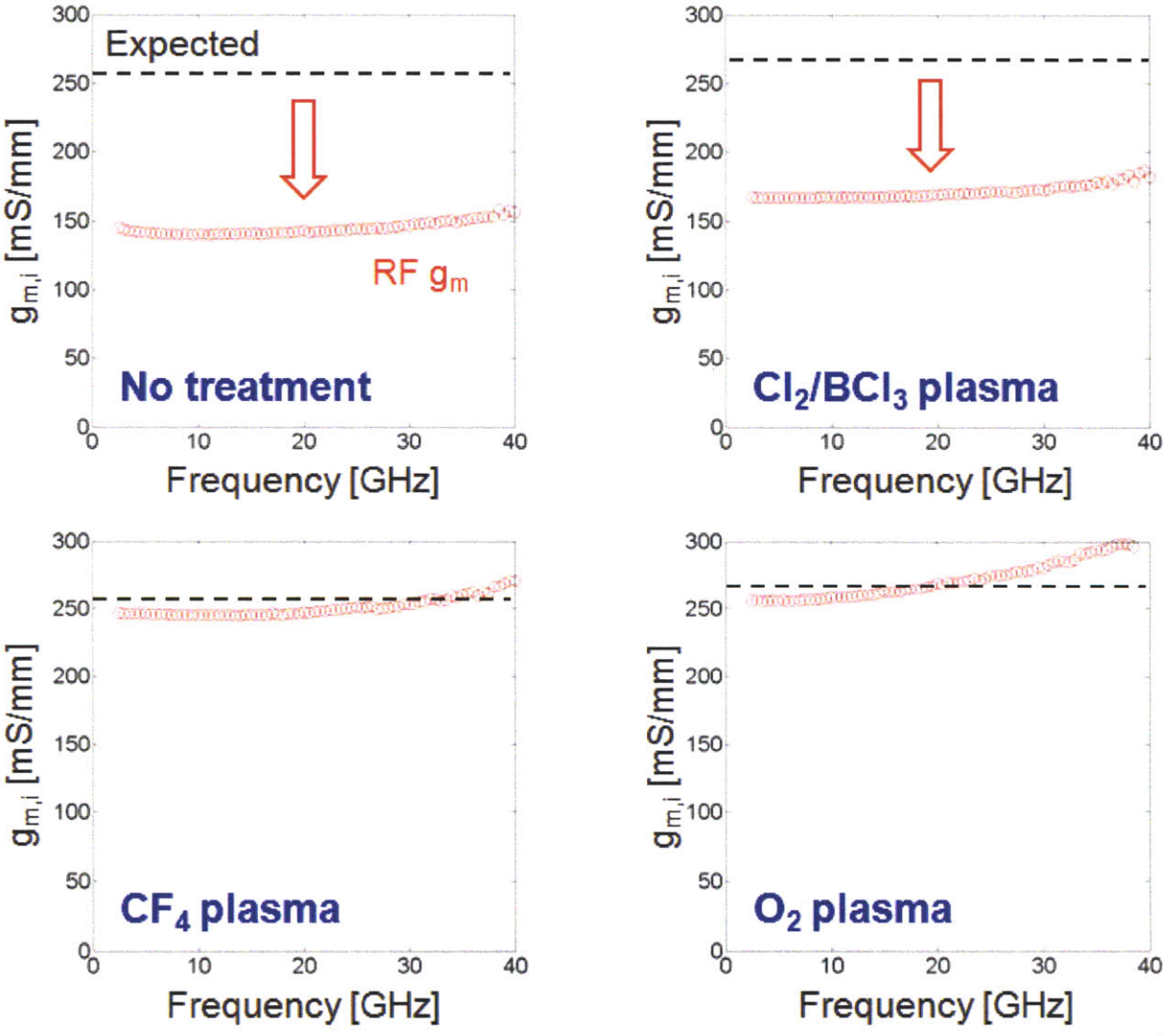


Figure 4-13. Comparison between expected and extracted intrinsic $g_{m,i}$ of each plasma treated sample as a function of measured frequency. Both CF_4 plasma and O_2 plasma treatments are effective to suppress RF g_m-collapse.

Instrument:	PHI Quantum 2000
X-ray source:	Monochromated Alk$_\alpha$ 1486.6eV
Acceptance Angle:	±23°
Take-off angle:	45°
Analysis area:	1400μm x 300μm
Charge Correction:	C1s 284.8 eV

Table 4-2. Measurement conditions used during the XPS analysis of plasma-treated samples.

X-ray Photoelectron Spectroscopy (XPS) measurements were conducted on the AlGaN surface after each plasma treatment to study the impact of each treatment on the AlGaN barrier. The measurement conditions are given in Table 4-2 and XPS data is quantified using relative sensitivity factors and a model that assumes a homogeneous layer. Photoelectrons are generated within the X-ray penetration depth (typically many microns), but only the photoelectrons within the top three photoelectron escape depths are detected. Escape depths are on the order of 1.5 ~ 3.5 nm, which leads to an analysis depth of 5 ~ 10 nm. Typically, 95% of the signal originates from within this depth.

Table 4-3 summarizes the atomic concentrations of the elements detected in the top 5 ~ 10 nm of AlGaN barrier by the XPS measurements. The values given are normalized to 100 % and the detection limits are approximately 0.05 to 1.0 atomic %. Major factors affecting the detection limits are the element itself (heavier elements generally have lower detection limits), interferences (can include photoelectron peaks and Auger electron peaks from other elements), and background (mainly caused by signal from electrons that have lost energy to the matrix). Gallium has two distinct signals found on opposite ends of the energy range and these are the Ga3d signal with a low binding energy (i.e. photoelectrons have a high kinetic energy) and the Ga2p3 with a high binding energy (i.e. photoelectrons have a low kinetic energy). The kinetic energy differences result in the Ga2p3 signal originating from shallower depths than the Ga3d signal. If the sample is homogeneous, quantifying with either signal will result in identical concentrations. Conversely, a heterogeneous sample may result in very different Ga concentrations depending on the signal used. For this reason, samples were quantified using both Ga3d and Ga2p3 signals as shown in Table 4-3(a) and Table 4-3(b), respectively.

(a) Ga3d signal

	RF g_m-collapse	Al	Ga	N	O	F	C	Cl	Si	B
As-grown	Yes	9.0	30.1	28.0	20.2	1.6	10.7	0.2	0.3	-
Cl_2/BCl_3 Plasma	Yes	9.5	31.4	36.1	8.2	0.8	10.0	2.2	-	1.7
CF_4 Plasma	No	8.6	27.1	22.6	14.4	14.4	12.3	0.6	-	-
O_2 Plasma	No	8.8	29.1	22.4	26.8	2.0	10.4	0.2	0.4	-

(b) Ga2p3 signal

	RF g_m-collapse	Al	Ga	N	O	F	C	Cl	Si	B
As-grown	Yes	9.3	27.8	28.9	20.8	1.7	11.0	0.2	0.3	-
Cl_2/BCl_3 Plasma	Yes	10.1	27.3	38.3	8.7	0.9	10.6	2.4	-	1.8
CF_4 Plasma	No	9.2	22.1	24.2	15.4	15.4	13.2	0.6	-	-
O_2 Plasma	No	9.1	26.6	23.2	27.7	2.1	10.8	0.2	0.4	-

Table 4-3. Atomic concentrations of the elements at the AlGaN surface after each plasma treatment detected by the XPS using (a) Ga3d signal and (b) Ga2P3 signal. XPS detects all elements with an atomic number of 3 (Li) and above. XPS does not detect H or He because the diameter of these orbitals is so small, reducing the catch probability to almost zero. A dash line '-' indicates the element is not detected.

As already shown in Figure 4-13, only CF_4 and O_2 plasma treatments suppress RF g_m-collapse. The main change observed in the samples after plasma treatment was a reduced percentage of N atoms, which was accompanied by an increase in the percentage of F and O. Since the electronegativity (tendency of an atom to attract electrons) of F and O is higher than N (N < O < F), the introduced F and O could attract the valence electrons of Ga more strongly than N, forming GaF_3 and Ga_2O_3. In fact, the formation of GaF_3 and Ga_2O_3 at the AlGaN surface was confirmed by XPS measurements after CF_4 and O_2 plasma treatment, respectively. While F and O exchange their sites with N, it is also possible that nitrogen vacancies (V_N) can be filled by F or O, therefore reducing V_N-related defects as well as dangling bonds near the AlGaN surface. The existence of nitrogen vacancies (V_N) in AlGaN or GaN has been previously reported by several researchers and the V_N-related defects are known as donor-like states [24][25] (positively

charged when empty and neutral when filled) causing strong Fermi level pinning at the AlGaN surface, reduction of the barrier height and increased leakage currents at the AlGaN Schottky interface, as well as higher current collapse in AlGaN/GaN HEMTs [26]. The Cl_2/BCl_3 plasma treatment acted in a way completely different from CF_4 and O_2. In the case of Cl_2/BCl_3, XPS measurements revealed an increase in the percentage of N atoms, while reducing the amount of O. This indicates that there was no chemical reaction or consumption of N or V_N. Therefore, based on the XPS analysis we *speculate* RF g_m-collapse could be caused by V_N-related defects.

4.3.3. Origin of RF g_m-collapse

Frequency dependent transconductance behavior was previously observed and studied in several material systems such as GaAs, SiC, and GaN [27][28][29]. This phenomenon is called transconductance dispersion (or g_m dispersion) and two types of g_m dispersion have been identified in the past, positive and negative. The positive (negative) g_m dispersion means that tranconductance at the high frequencies is higher (lower) than that at the low frequencies. The transition frequency is typically in the range of $10 \sim 10^5$ Hz. In GaAs FETs, the positive g_m dispersion is believed to be due to the *DX* center under the gate and the negative g_m dispersion is thought to be due to the surface states existing on the ungated source/drain access regions [30].

The positive g_m dispersion is explained as following. At low frequencies, electrons in the DX center respond to the applied alternating signal and can be fully modulated. During the positive (negative) cycle of the signal, energy level of the *DX* center (E_{DX}) becomes lower (higher) than the Fermi level in the channel (E_F) and thus electrons are captured by (emitted from) *DX* centers. This results in the reduction of the 2DEG charge modulation in the channel and lowers transconductance. At high frequencies, on the other hand, electrons in the *DX* centers are unable to follow the applied alternating signal. Thus, the alternating signal applied to the gate fully modulates the 2DEG charges resulting in the increase of transconductance.

The negative g_m dispersion is explained as following. At low frequencies, electrons captured at the surface states in the source/drain access regions can follow the applied alternating signal. Thus, depletion layers under the access regions are simultaneously modulated with the region

underneath the gate. At high frequencies, on the other hand, the electrons are not able to follow the signal, and the depletion region underneath the gate is only modulated, resulting in the decrease of the transconductance. The negative g_m dispersion is dependent on the surface leakage current from the gate and can be suppressed by the surface passivation with Si_3N_4 [31].

What we observe in our AlGaN/GaN HEMTs, RF g_m-collapse, is similar to the negative g_m dispersion reported in the literature. However, although more analysis needs to be carried out, we believe RF g_m-collapse is a clearly different phenomenon because of following reasons. First of all, the transition frequency of RF g_m-collapse (~10^7 Hz) is much higher than that of transconductance dispersion (~10^3 Hz). Also, there was no clear correlation between gate leakage and RF g_m-collapse opposite to what has been seen in the g_m dispersion [32]. Moreover, source/drain access regions are not likely responsible for the RF g_m-collapse since a surface treatment applied only underneath the gate (± 20 nm) successfully eliminated RF g_m-collapse. More importantly, RF g_m-collapse was also observed in fully passivated samples without evident current collapse, although better statistics from more samples are required to fully prove it.

To understand the origin of RF g_m-collapse, we first looked at the expression of intrinsic g_m given by measured Y-parameters (transformed from measured S-parameters) as

$$g_m = \sqrt{\left(\left(Re(Y_{21}) - Re(Y_{12})\right)^2 + \left(Im(Y_{21}) - Im(Y_{12})\right)^2\right)\left(1 + \omega^2 {C_{gs}}^2 {R_i}^2\right)} \qquad (4\text{-}16)$$

We focused on the device with L_g = 200 nm, W_g = 100 μm, and L_{sd} = 2 μm, This device showed an RF g_m-collapse similar to the one shown in Figure 4-11. In this device, the extracted C_{gs} and R_i were 41 fF and 23 Ω which results in $(1 + \omega^2 {C_{gs}}^2 R_i^2) \approx 1$ throughout the measured frequency range. Also, Re(Y_{21}) and Im(Y_{21}) dominate the equation since they are in the order of 10^{-2} ~ 10^{-3} while Re(Y_{12}), Im(Y_{12}) are in the order of 10^{-4} ~ 10^{-6}. Now, the expression for g_m can be simplified by

$$g_m \cong \sqrt{(Re(Y_{21})^2 + Im(Y_{21})^2)} \qquad (4\text{-}17)$$

Thus, g_m is mainly determined by the forward transfer admittance (or forward gain) Y_{21}.

From Appendix 1, Y_{21} is related to the measured S-parameter as

$$Y_{21} = \frac{1}{Z_0} \frac{-2S_{21}}{(1+S_{11})(1+S_{22}) - S_{12}S_{21}} \tag{4-18}$$

where Z_0 is the characteristic impedance, typically Z_0 = 50 Ω. By carefully analyzing the measured S-parameters, we can identify the reason for low Y_{21} or low intrinsic g_m. Figure 4-14 shows measured S-parameters as a function of frequency from 10 MHz to 40 GHz. The measurement was divided into two frequency regions, high frequency range 1 ~ 40 GHz and lower frequency range 0.01 ~ 1 GHz. Interestingly, the magnitude of the forward gain S_{21} dropped significantly with increasing frequency up to 1 GHz while other S-parameters showed expected frequency dependencies. Figure 4-15 highlights this drop of $|S_{21}|$ by comparing it to its expected frequency behavior. $|S_{21}|$ value was initially as high as its expected value, however when increasing frequency over 10 MHz, $|S_{21}|$ rapidly decreases largely deviating from its expected value. Since Y_{21} is directly proportional to S_{21} (Eq. 4-18), this drop of S_{21} causes low Y_{21} and, ultimately, low intrinsic g_m. It is noted that the change in S_{21} does not affect Y_{11}, Y_{12}, and Y_{22} significantly due to its small contribution to their S-to-Y transformation (Appendix 1). Thus, the drop of S_{21} with frequency hardly changes other intrinsic parameters such as g_o, C_{gs} and C_{gd}, in good agreement with the observed frequency behaviors of C_{gs}, C_{gd}, g_m and g_o in Figure 4-11. Therefore, RF g_m-collapse in AlGaN/GaN HEMTs is mainly due to the degradation of the forward gain S_{21} at high frequency operation. In other words, the incident power is inefficiently lost in the device producing less forward gain at high frequency.

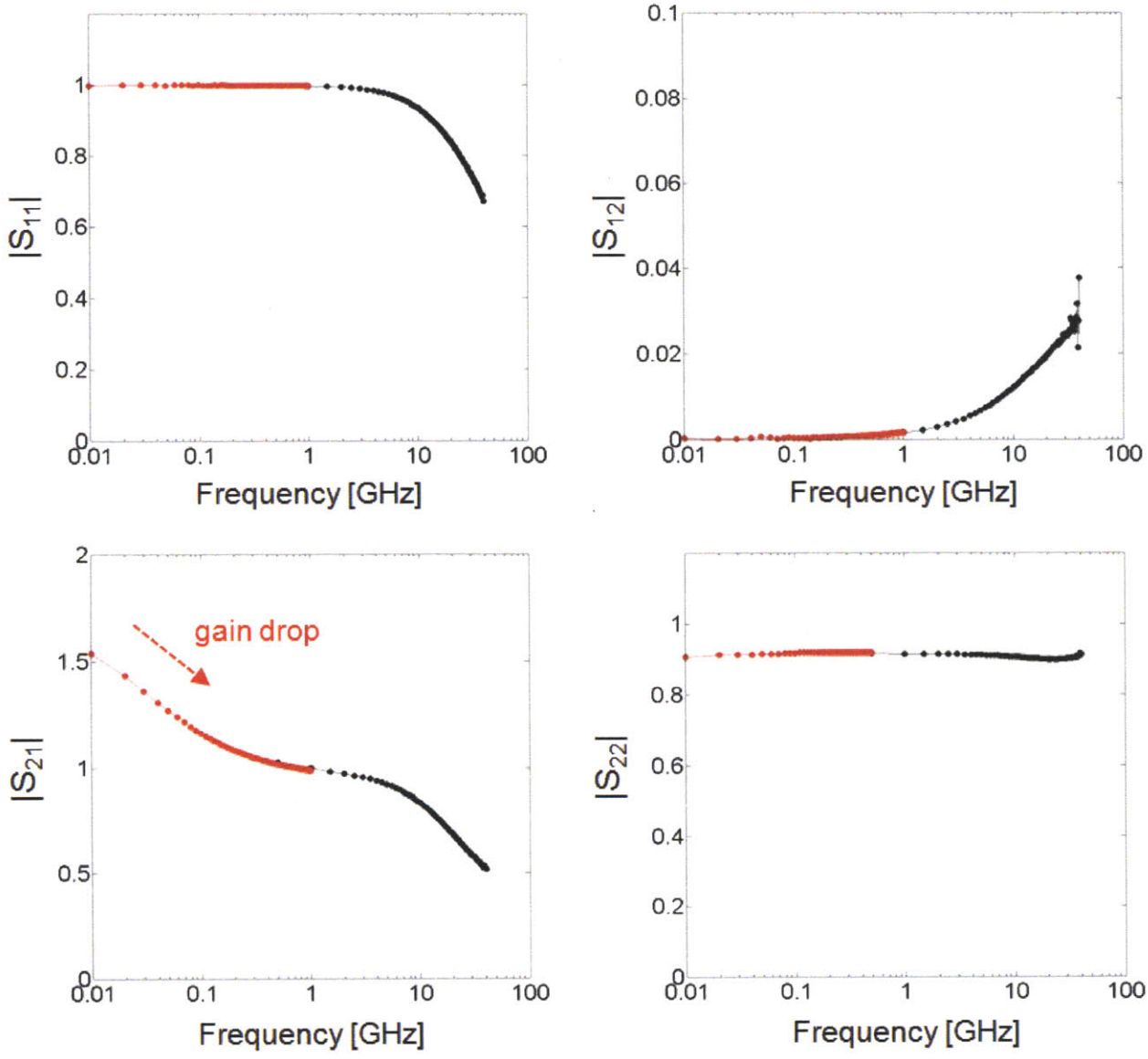


Figure 4-14. Magnitude of measured S-parameters as a function of frequency. The S-parameter measurement was carried out in two frequency regions, high frequency range 1 ~ 40 GHz and lower frequency range 0.01 ~ 1 GHz. Piecewise continuous curves indicates RF g_m-collapse is not a function of time, but a function of frequency. The drop of S_{21} causes low Y_{21} and, ultimately, low intrinsic g_m.

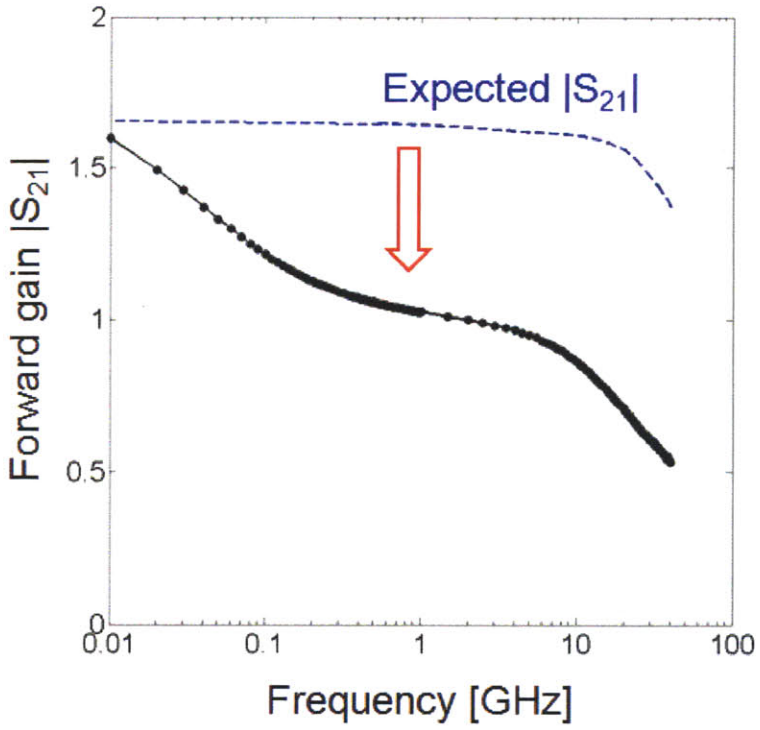


Figure 4-15. A large deviation of forward gain $|S_{21}|$ with respect to its expected frequency dependency.

Measured S-parameters were modeled by ADS simulation. Figure 4-16 shows the modified small-signal equivalent circuit designed to model RF g_m-collapse in AlGaN/GaN HEMTs. Additional voltage controlled current source (VCCS) with negative transconductance g_m' and transconductance delay τ' of 0.01 ~ 1 GHz range was inserted in series with the original VCCS to simulate RF g_m-coallpse. The net g_m of the device drops as increasing the frequency due to the negative g_m' and the transition frequency of RF g_m-collapse is set by τ'. This additional VCCS successfully models measured S-parameters by only degrading $|S_{21}|$ (or net g_m), but barely changing other S-parameters (or C_{gs}, C_{gd}, g_o, etc) as shown in Figure 4-17. Thus, RF g_m-collapse seems to be mainly caused by decreasing intrinsic g_m as increasing the frequency and not associated with other small-signal elements such as C_{gs}, C_{gd}, and g_o. It is also highlighted that modifying other small-signal equivalent circuit elements such as extrinsic R_s and R_d cannot model RF g_m-collapse since it changes not only S_{21} but also other S-parameters. Therefore, access regions of the devices are not responsible for RF-g_m-collapse and this is another evidence that RF g_m-collapse is different from the negative g_m dispersion reported in the literature. Although RF g_m-collapse was investigated in this section, further theoretical and experimental studies are necessary to clarify the mechanism of RF g_m-collapse.

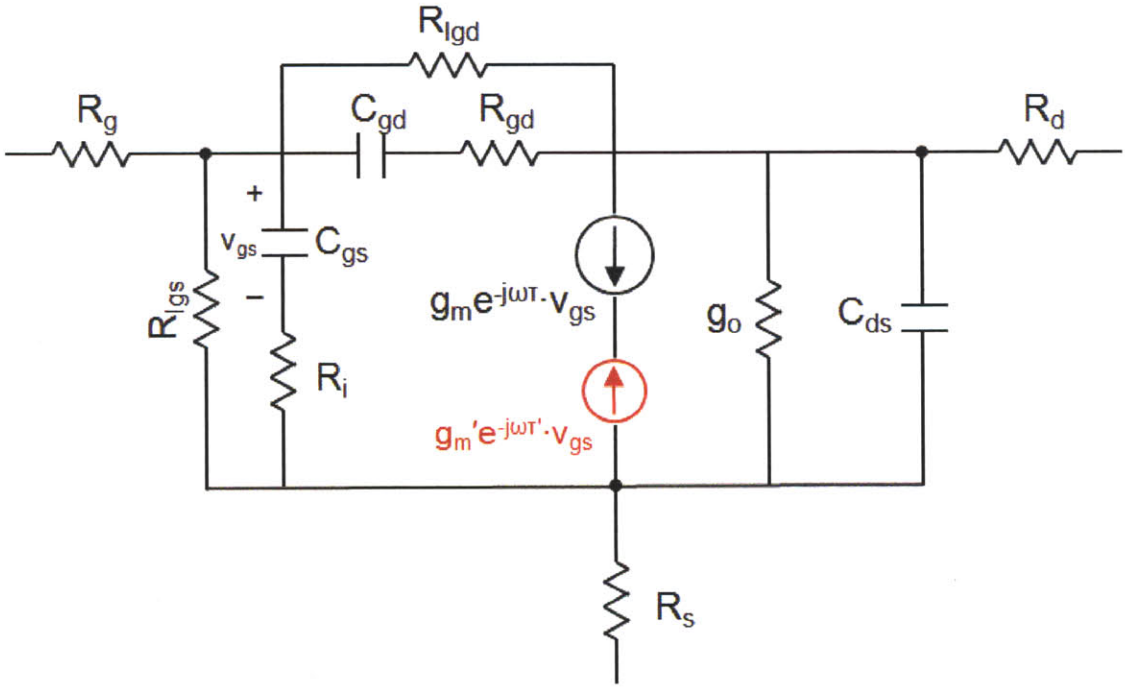


Figure 4-16. Modified small-signal equivalent circuit of the active region designed to model RF g_m-collapse in AlGaN/GaN HEMTs. Voltage controlled current source (VCCS) with negative transconductance g_m' and transconductance delay τ' of 0.01 ~ 1 GHz range was added to the circuit.

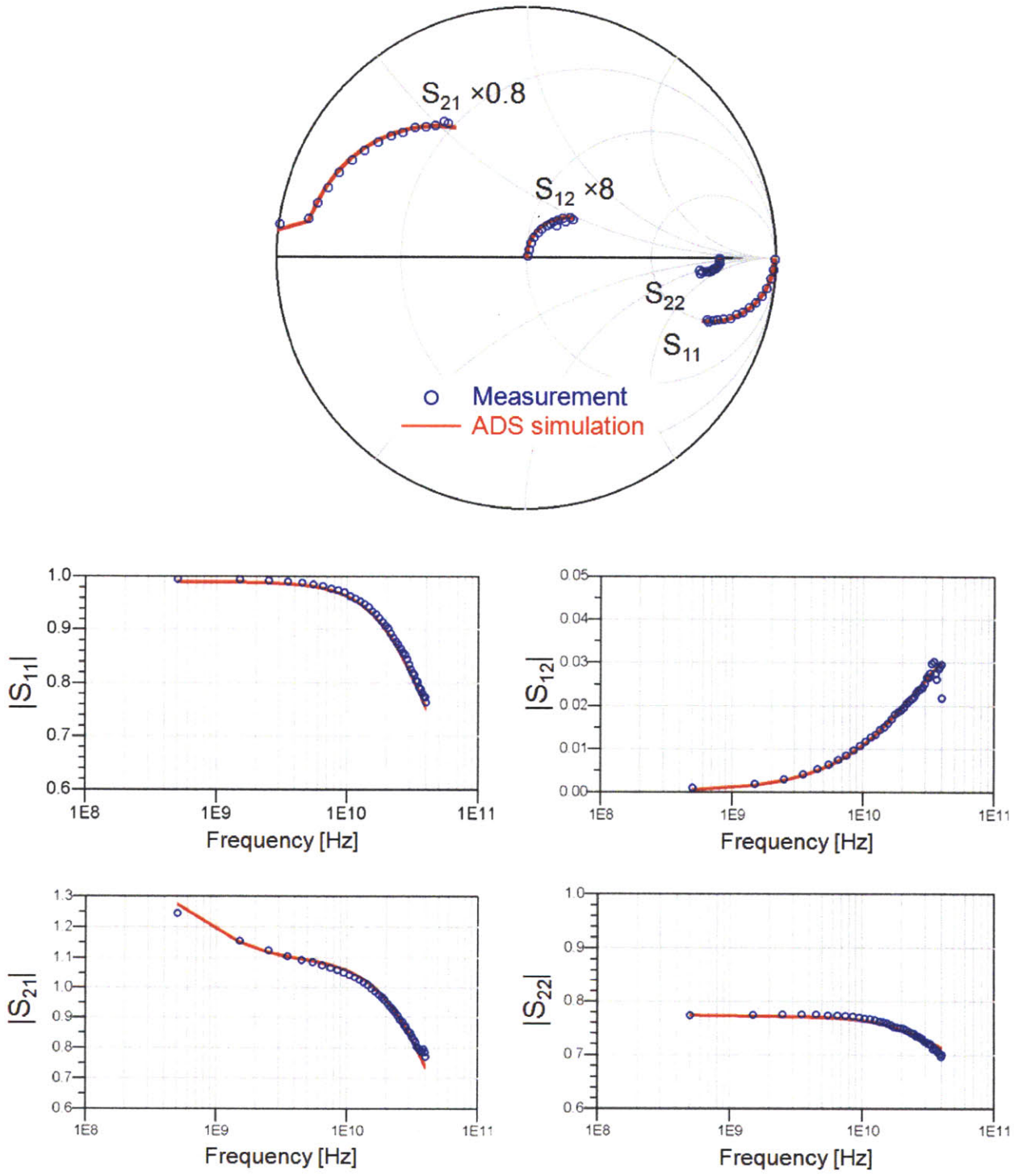


Figure 4-17. Comparison between measured and ADS modeled S-parameters of modified small-signal equivalent circuit in Figure 4-16. There is a good agreement between measurement and ADS simulation and it confirms RF g_m-collapse is not associated with other small-signal elements such as C_{gs}, C_{gd}, and g_o.

4.3.4. Oxygen plasma treatment

Although both CF_4 plasma and O_2 plasma treatments are effective in terms of preventing RF g_m-collapse, O_2 plasma treatment was used to improve f_T in AlGaN/GaN HEMTs since it is more controllable and reproducible than CF_4 plasma treatment. The CF_4 plasma treatment was previously studied by other researchers mainly to realize enhancement mode operation in AlGaN/GaN HEMTs [33][34]. The incorporation of F ions in the AlGaN barrier by the CF_4 plasma treatment is known to be responsible for the positive shift of threshold voltage [34]. However, CF_4 plasma treatment often results in mobility degradation in the 2DEG channel and reduction in the on-current, which can in many cases be partially recovered by a high-temperature annealing. Also, the amount of threshold-voltage shift is sensitive to the treatment condition and hard to control precisely. O_2 plasma treatment, on the other hand, is a very attractive option for preventing RF g_m-collapse and at the same time it improves the mobility, prevents threshold-voltage variations, and reduces gate leakage current without degrading on-current as will be shown in this section.

O_2 plasma treatment was applied to the exposed AlGaN surface right before the gate metallization of AlGaN/GaN HEMTs. The O_2 plasma treatment promotes the formation of Ga_2O_3 over Al_2O_3 at the AlGaN surface and it was confirmed by XPS measurement (Table 4-4). By taking cross-sectional high-resolution transmission electron microscopy (HR-TEM) of the fabricated device, a thin oxide layer (could be combination of Ga_2O_3 and Al_2O_3) was also observed between the gate metal and the AlGaN surface as shown in Figure 4-18. The process for the formation of this oxide layer was experimentally confirmed to be self-terminating, as predicted by other reports [35]. The oxide thickness saturated at about 5 nm after 20 minutes of O_2 plasma treatment. The oxide thickness introduces a negative threshold-voltage shift in the transistor characteristics consistent with the change in the gate-to-channel distance.

Oxygen chemical state in % of total O

	Ga_2O_3, $Ga_xO_yN_z$	Al_2O_3, $Al_xO_yN_z$, AlOOH, Organic O
As-grown	29	71
O_2 plasma	58	42

Table 4-4. Percentage of the total atomic concentration of the oxygen detected at AlGaN surface.

The formation of the gate oxide effectively reduces the gate leakage current by at least one order of magnitude without degrading the on-current (Figure 4-19) [36]. In addition, the oxide layer enhances the effective mobility (or electron velocity) possibly due to the reduction of a parallel conduction by improving the gate control over the channel and/or because of the change in electric field distribution by the gate dielectric [37]. The improved transport properties by the O_2 plasma treatment are summarized in Figure 4-20. Figure 4-20(a) shows slightly improved drift mobility over a wide range of charge density measured from FAT-FET and Figure 4-20(b) shows more than 15% increased peak transconductance in the HEMT with L_g = 200 nm.

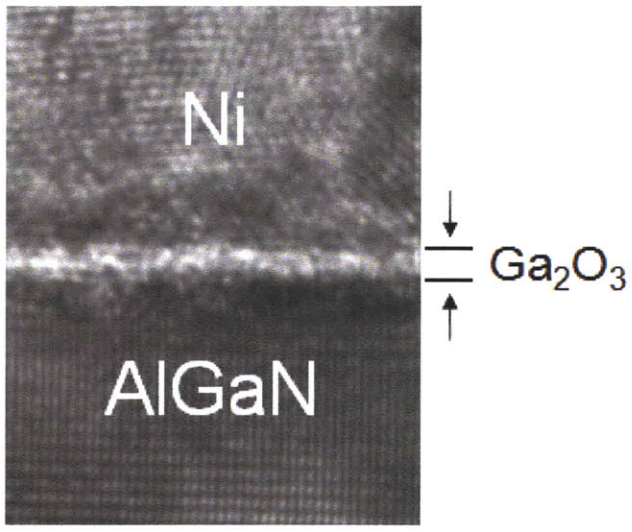


Figure 4-18. A cross-section HR-TEM image of HEMTs at the interface between gate metal (Ni) and the AlGaN surface. 1 ~ 5 nm of Ga_2O_3 layer depending on the treatment time was confirmed in conjunction with XPS measurement shown in Table 4-4.

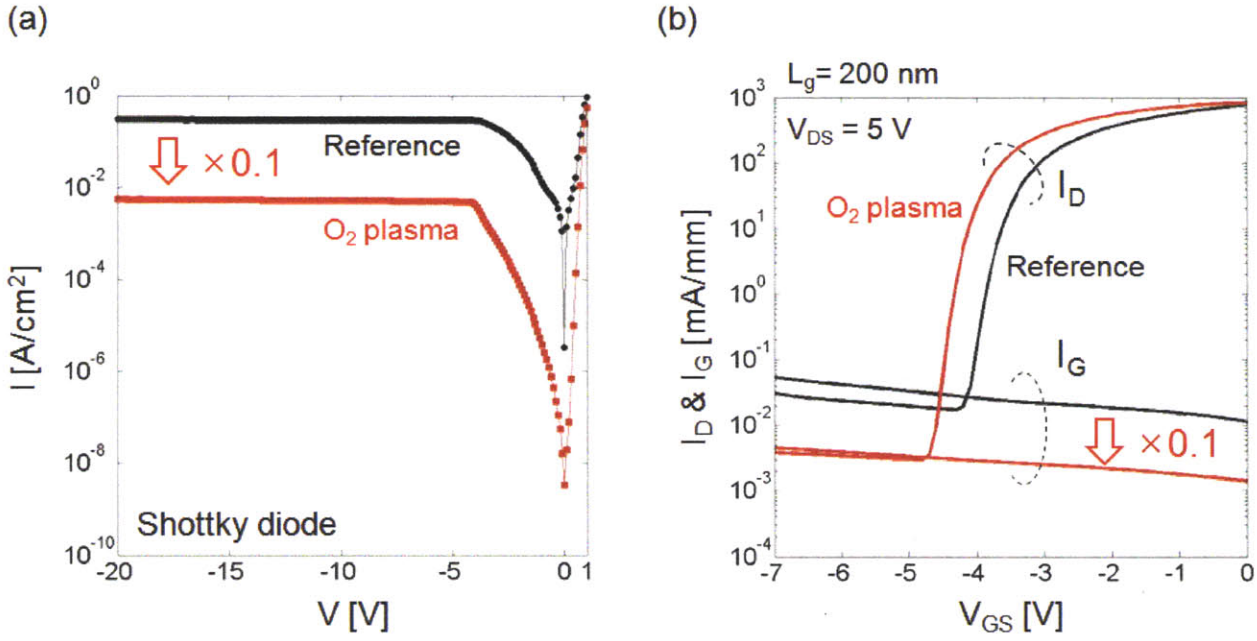


Figure 4-19. Reduced gate leakage current by O_2 plasma treatment in (a) Schottky diode and (b) HEMT with L_g = 200 nm.

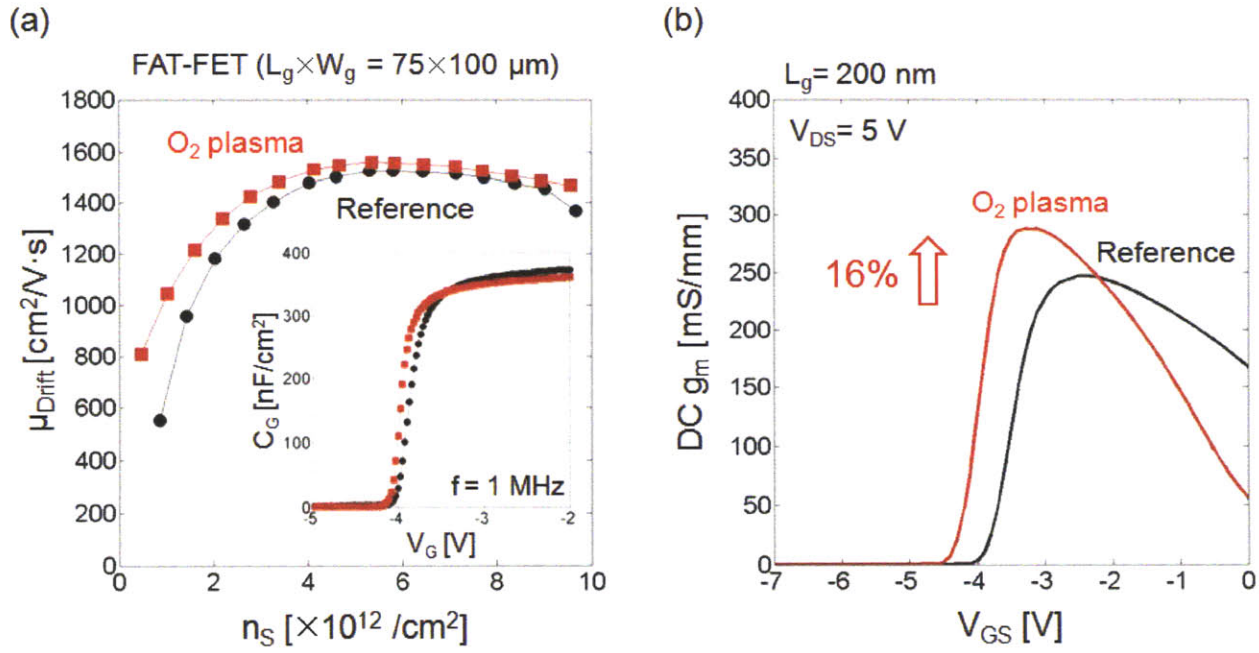


Figure 4-20. Improved transport properties by O_2 plasma treatment measured from (a) FAT-FET and (b) HEMT with L_g = 200 nm.

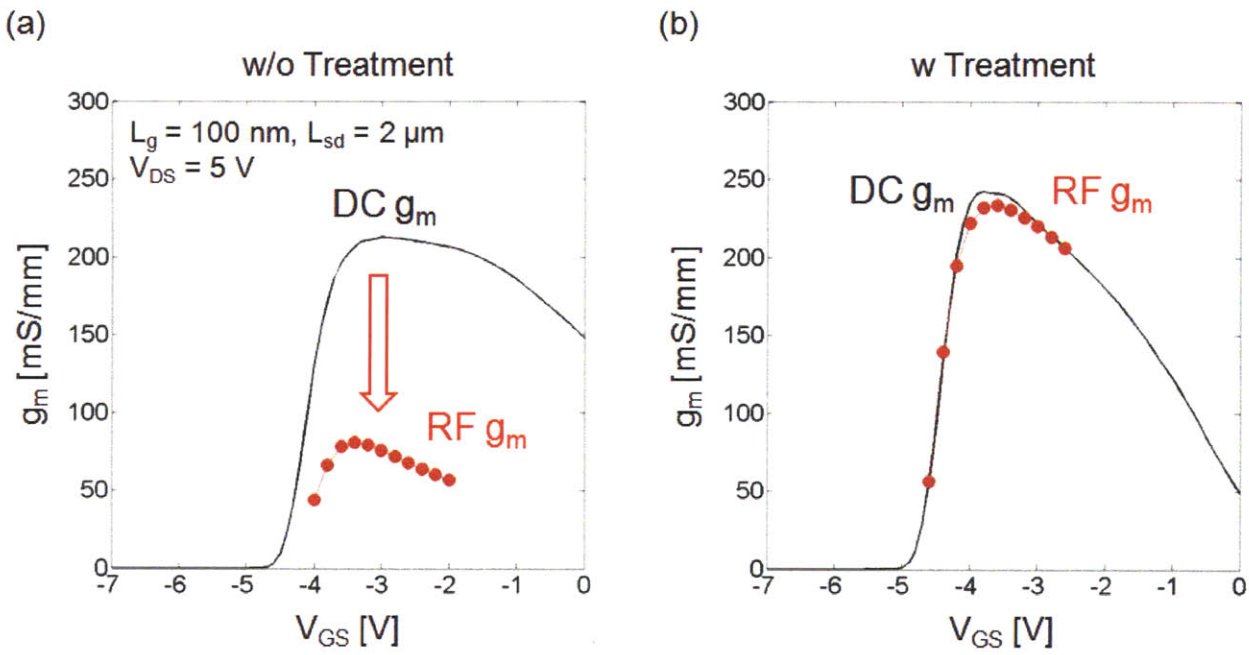


Figure 4-21. Comparison between DC and RF extrinsic transconductance of a device with L_g = 100 nm (a) without and (b) with O_2 plasma treatment over different bias conditions. O_2 plasma treatment recovers RF g_m to the DC g_m level and eliminates RF g_m -collapse.

The most important effect of O_2 plasma treatment for improving high frequency performance of AlGaN/GaN HEMTs is suppressing RF g_m-collapse. Figure 4-21 compares DC and RF extrinsic transconductance of a device with L_g = 100 nm with and without O_2 plasma treatment. As expected, without any treatment there is a significant drop in RF g_m compared to DC g_m (Figure 4-21(a)). This problem can be eliminated by applying O_2 plasma treatment to the gate region and as a result RF g_m almost matches DC g_m over the entire range of measured V_{GS} (Figure 4-21(b)).

Finally, as a result of all the advantages of using O_2 plasma treatment described so far, f_T and its scaling behavior largely improves. Figure 4-22(a) compares f_T of L_g = 170 nm devices with and without O_2 plasma treatment. Due to the suppression of RF g_m-collapse after the O_2 plasma treatment, the f_T of the O_2 plasma-treated sample is almost two times higher than the device without the treatment. Figure 4-22(b) shows scaling behavior of f_T for the devices with and without O_2 plasma treatment. Without the treatment, the f_T of the devices was barely affected by scaling L_g. This is due to the degradation of g_m by both RF g_m-collapse and short-channel effects.

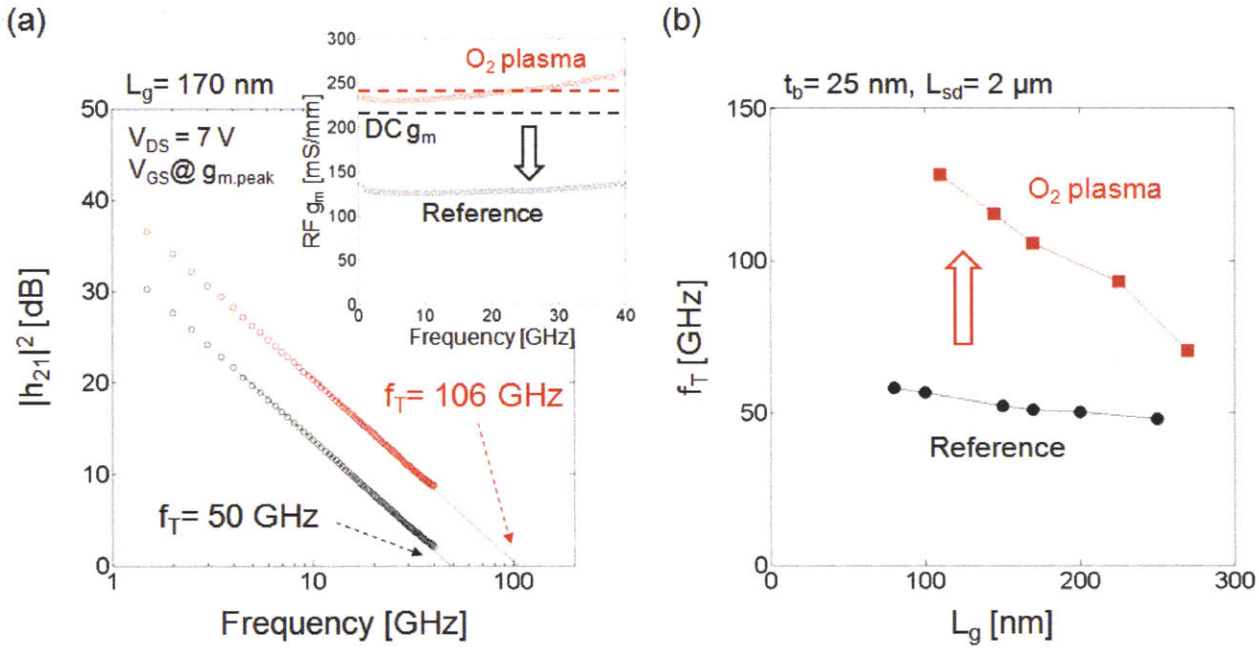


Figure 4-22. O_2 plasma treatment improves (a) f_T a factor of two and (b) scaling behavior of f_T by suppressing RF g_m-collapse.

By applying O_2 plasma treatment, the scaling behavior of f_T is largely improved mainly due to the suppression of RF g_m-collapse.

4.4. Advanced gate technology

To maximize f_T in AlGaN/GaN HEMTs, RF g_m-collapse, short-channel effects, and the gate lengths should be simultaneously minimized. To achieve this goal, RF g_m-collapse needs to be eliminated by applying O_2 plasma treatment. In addition, vertical scaling of AlGaN barrier (t_b) has to be accompanied by lateral scaling of L_g to maintain high aspect ratio (L_g / t_b) and avoid strong short-channel effects. Finally, the already fabricated gate electrode needs to be shortened even more by laterally etching the gate metal. In this section, we presént an advanced gate technology that combines vertical gate-recess, O_2 plasma treatment, and lateral gate-etch to achieve state-of-the-art f_T in AlGaN/GaN HEMTs.

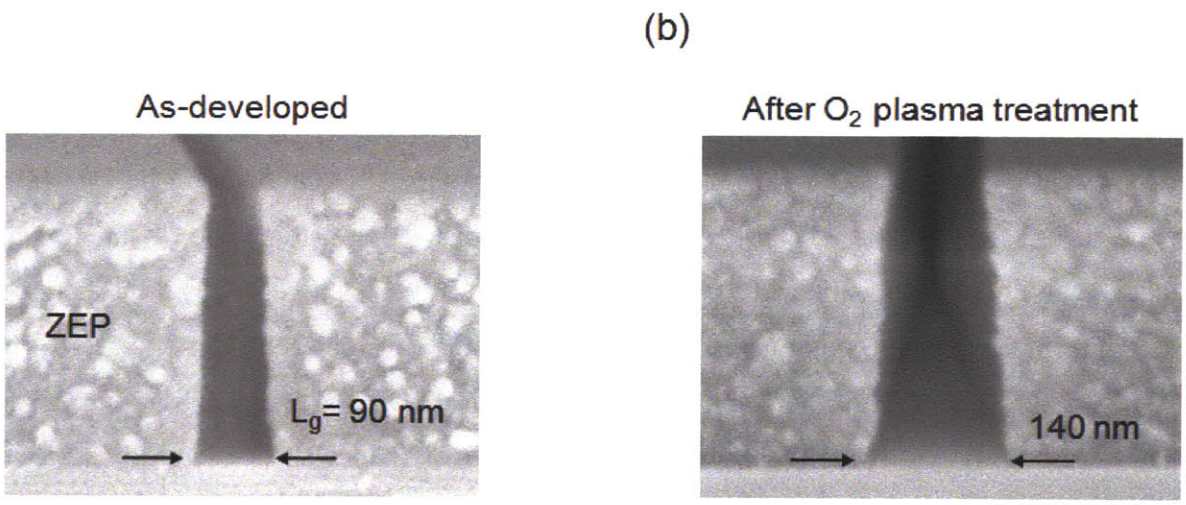


Figure 4-23. O_2 plasma treatment etches organic materials (e.g. photoresist, e-beam resist, etc) and deteriorates as-defined gate profile. Gate length was increased from (a) 90 nm to (b) 140 nm by applying O_2 plasma treatment for 5 minutes.

4.4.1. Al_2O_3-sidewall technology

Although the O_2 plasma treatment is very attractive to improve high frequency performance of AlGaN/GaN HEMTs, there is a process issue that needs to be solved for its practical application. In the standard fabrication process of AlGaN/GaN HEMTs, the oxygen plasma treatment is applied after developing the gate pattern and right before depositing the gate metals. However, applying the O_2 plasma treatment at this step etches away organic materials such as e-beam resist and photoresist. This causes a widening of the gate structure originally defined by the e-beam lithography. Figure 4-23 shows the cross-sectional SEM image of the gate region before the gate metallization and how the O_2 plasma treatment adversely affects the patterned e-beam resist. The as-defined gate length was 90 nm, however the patterned gate length was increased more than 50 % to 140 nm after the O_2 plasma treatment. Also, adding more process steps such as vertical gate recess makes the resist even more vulnerable to the damage (Figure 4-24(a)). Such a wide gate length is not acceptable for high frequency devices.

To solve the problem of e-beam resist removal by the O_2 plasma treatment, an Al_2O_3-sidewall technology has been developed to protect the patterned gate profile from the processing steps, especially O_2 plasma treatment. Atomic layer deposition (ALD) is a self-limiting sequential

surface chemistry that deposits precisely one atomic layer of material per deposition cycle achieving very conformal thin-films even in high aspect ratio (> 40) and complex structures [38]. Figure 4-24(b) shows cross-section SEM image of the patterned gate profile after Al_2O_3 passivation by ALD at 80 °C. It is important to keep the ALD temperature below 100 °C to prevent the reflow of the e-beam resist, which could distort the gate profile. It should be noted that the conformal deposition of 10 nm of Al_2O_3 passivation helps to reduce the defined gate length from 90 nm to 70 nm by creating 10-nm-thick sidewalls on each side. Figure 4-24(b) also shows the cross-section of the gate pattern with 10 nm ALD Al_2O_3 passivation after applying a gate-recess, followed by the O_2 plasma treatment. The Al_2O_3 sidewalls successfully protected the originally defined gate length of 70 nm even after these processing steps. Therefore, ALD Al_2O_3-sidewall technology allows the use of an O_2 plasma treatment without widening the gate length. The next section will discuss the advanced gate technology, based on the ALD Al_2O_3-sidewall technology and O_2 plasma treatment, that we have used in this work to maximize f_T in AlGaN/GaN HEMTs.

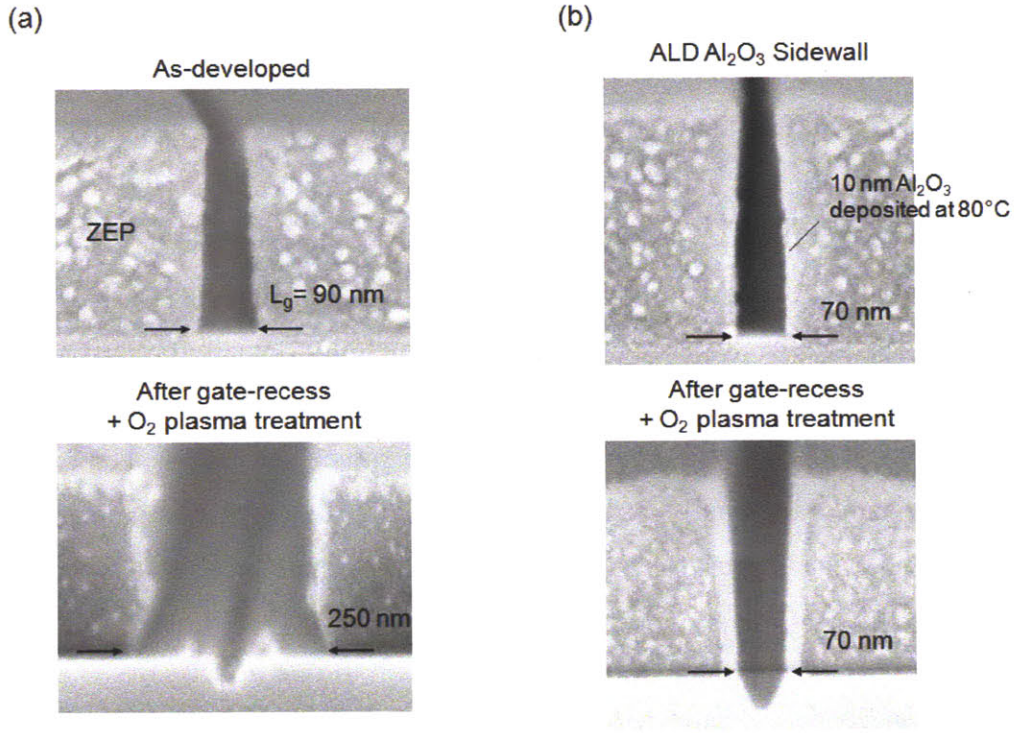


Figure 4-24. (a) Issues in using O_2 plasma treatment and (b) how Al_2O_3-sidewall technology fix the problem. It should be noted that although (b) shows modified angle of the resist profile from its original reverse angled undercut profile, there was no problem with the subsequent lift-off process. We believe the profile is still kept reverse angled or almost vertical even after the gate-recess and O_2 plasma treatment, however it is distorted by the electron beam radiation during the SEM analysis.

4.4.2. New gate process to maximize f_T

Figure 4-25 summarizes the gate technology used in this work. Sub-micron gates were patterned by 30-keV electron-beam lithography on a single ZEP resist layer. The patterned ZEP layer was coated by a 10 nm of Al_2O_3 dielectric layer deposited by atomic layer deposition. A deposition temperature of 80 °C was used to prevent the thermal reflow of the ZEP. Then a low-power gate recess was applied using Cl_2/BCl_3-based electron cyclotron resonance reactive ion etching (ECR-RIE). The gate recess was followed by an O_2 plasma treatment of the exposed AlGaN surface. It should be highlighted that the remaining Al_2O_3 sidewalls play an important role to protect the ZEP layer from the O_2 plasma, preserving the original gate length defined by the e-beam lithography. After the oxygen plasma treatment, a Ni/Au (10/70 nm) metal stack was deposited for the gate contact and, finally, a selective Ni etching was performed to further reduce the physical gate length.

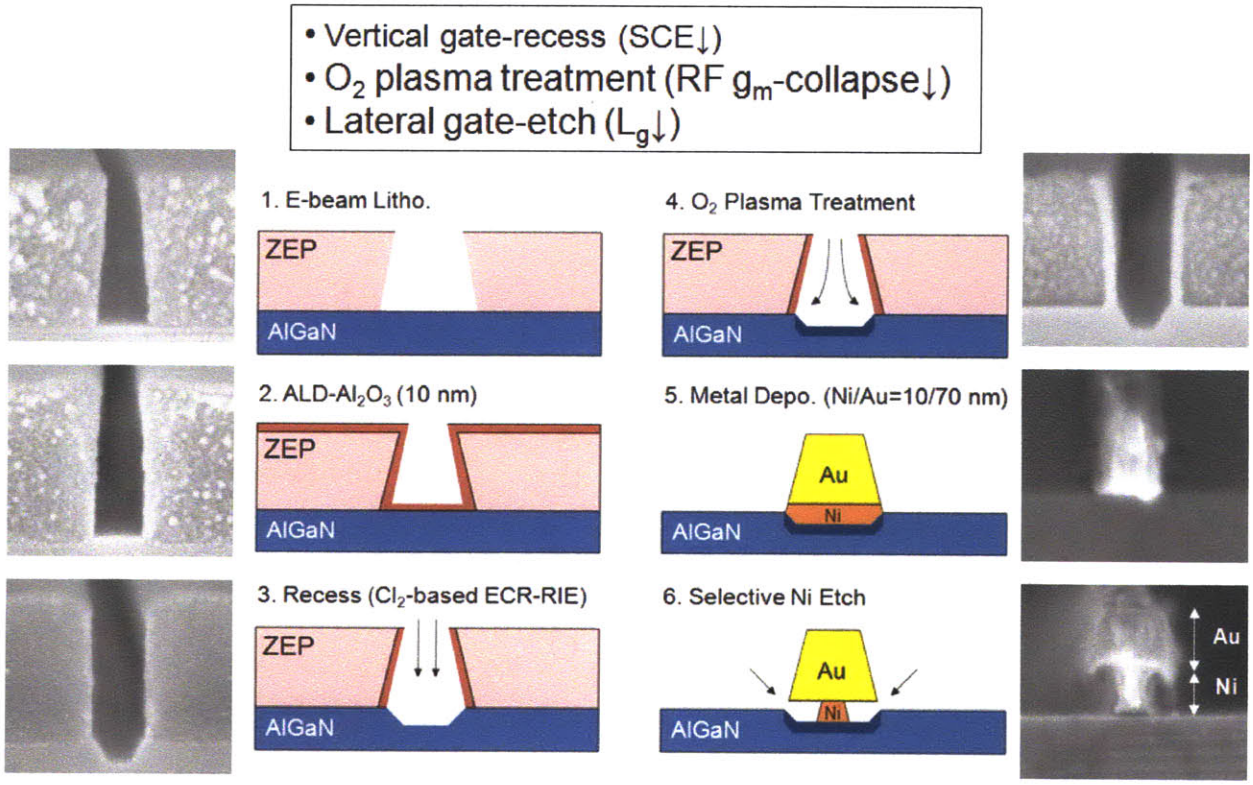


Figure 4-25. New gate technology developed in this work and relevant cross-section SEM images in each step. 10 nm of ALD-Al_2O_3 sidewalls protects the gate length defined by e-beam lithography from the oxygen plasma treatment. After Ni/Au gate metal deposition, the bottom Ni layer was selectively etched to further reduce the physical gate length.

The selective Ni-etch in this work was accomplished by multiple cycles of coating and stripping ZEP. Detailed process is spin coating a single layer of ZEP, baking it on a hot plate at 180 °C, and finally removing ZEP by N-Methyl-2-Pyrrolidone (NMP, $CH_3N(CH_2)_3CO$). The NMP is a chemically stable and powerful solvent used in the electronics industry for many years as a photoresist (or e-beam resist) stripper. This Ni-etch process is a very convenient way to reduce the gate length because it does not degrade any device characteristics. Also, the lateral etch of Ni can be easily controlled by the number of process cycles used. However, the mechanism of the Ni-etch in the above process is not fully understood. There are two possibilities: One is that the Ni surface reacts with ZEP during the baking step and the reacted portion of Ni is stripped by NMP; another possibility is that the Ni surface is oxidized during the baking step and the NiO_x is then removed by NMP. Figure 4-26 shows the improvement of f_T by applying the selective Ni-etch to an AlGaN/GaN HEMT with gate length of 250 nm. After one cycle of the Ni-etch process, f_T of the device increased by 24 % from 74 GHz to 97 GHz mainly thanks to 20 % reduced C_{gs} (or L_g). Due to the increased f_T, f_{max} of the device was also improved by 16% from 108 GHz to 128 GHz (not shown in Figure 4-26).

Figure 4-27 shows top-down SEM image and cross-section TEM image of successfully

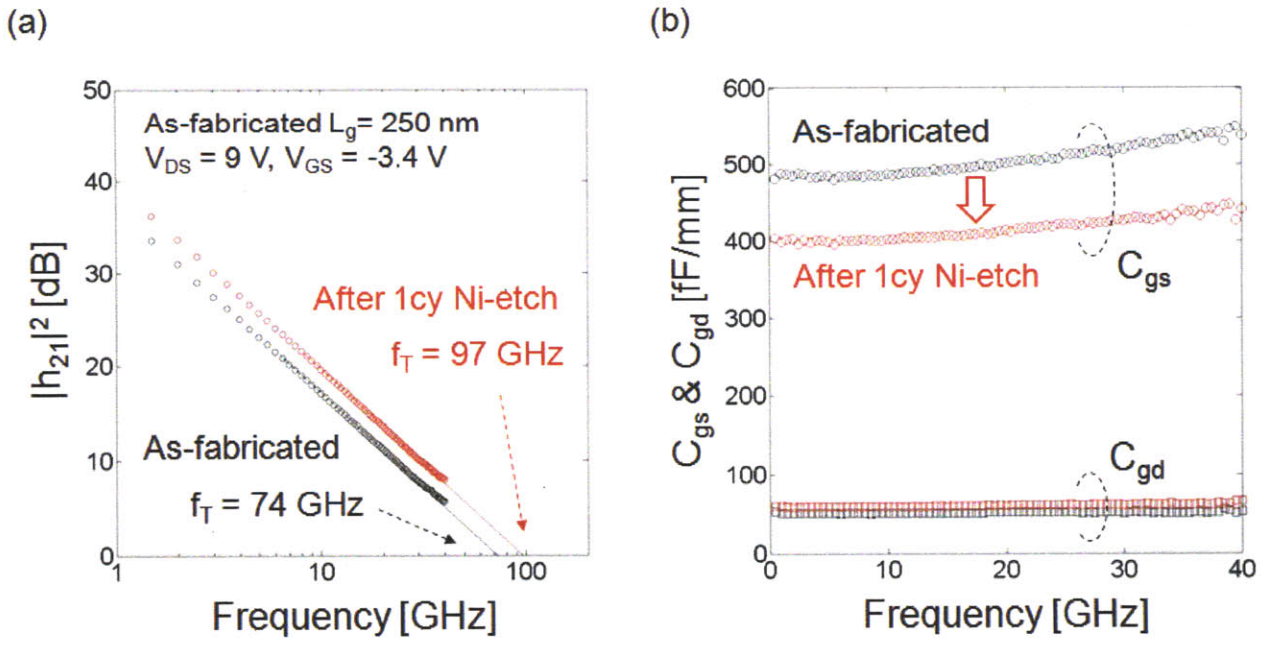


Figure 4-26. (a) Selective lateral Ni-etch of the gate improves f_T. (b) The reduction of the gate length by the lateral Ni-etch was confirmed by the reduced intrinsic C_{gs} extracted from the measured S-parameters.

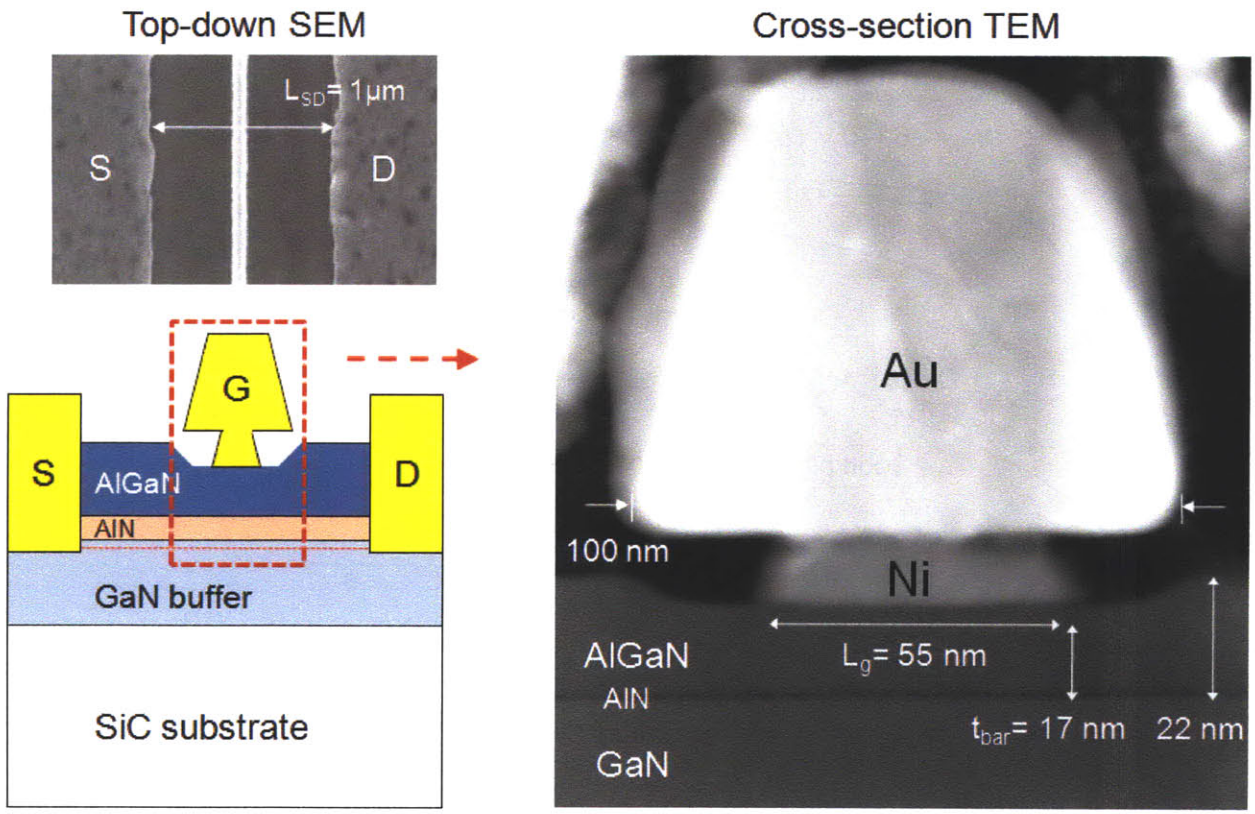


Figure 4-27. Schematic (left) and cross-section STEM image (right) of recessed AlGaN/GaN HEMT. The physical gate length (L_g) after lateral Ni-etch is 55 nm. The barrier thickness (t_b) after vertical gate-recess is 17 nm (1 nm AlN spacer is included.).

fabricated AlGaN/GaN HEMTs following the new gate process in this work. A gate length of 55 nm and the remaining AlGaN barrier thickness of 17 nm were precisely measured from the TEM image. It should be highlighted that the selective Ni-etch reduced the effective gate length by more than 40%.

4.5. Device characteristics

4.5.1. DC characteristics

Figure 4-28 shows the DC characteristics of one of the fabricated AlGaN/GaN HEMTs with L_g of 55 nm, gate-to-channel distance (t_b) of 17 nm, and source-to-drain distance (L_{sd}) of 1 μm. As shown in Figure 4-28(a), the device exhibited a low on resistance of 1 Ω·mm and a low knee voltage of 2 V due to the tight source-to-drain distance (L_{sd} = 1 μm) and the optimized ohmic

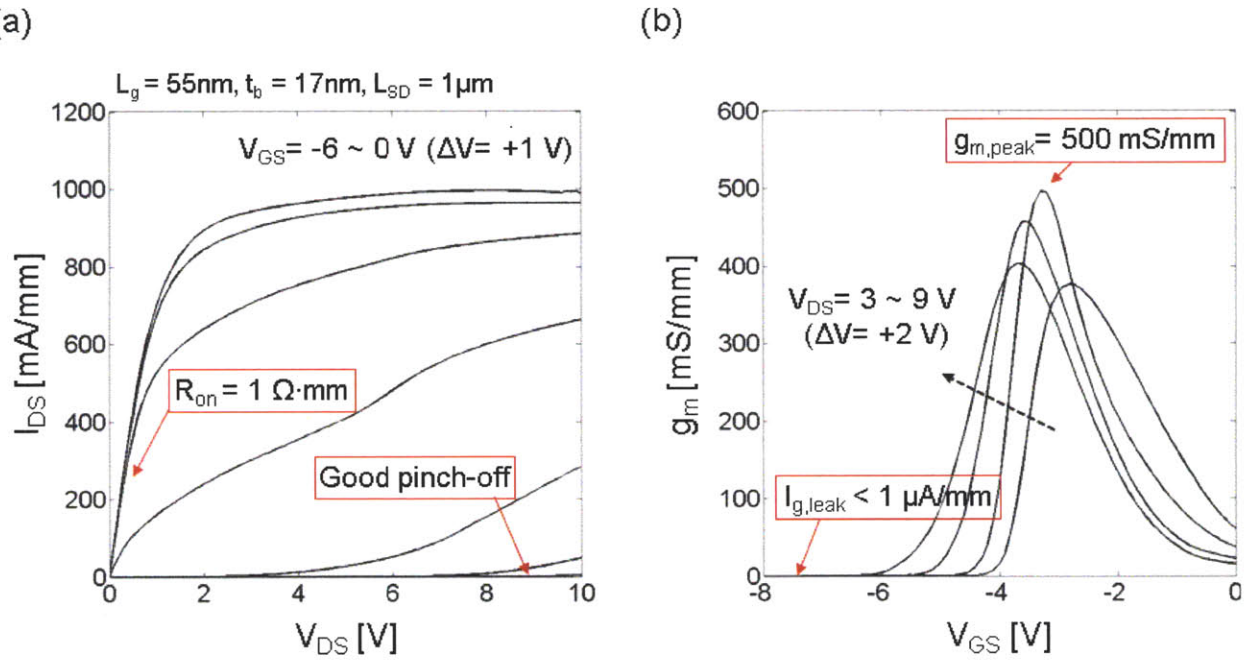


Figure 4-28. DC characteristics of a recessed and oxygen plasma-treated AlGaN/GaN HEMT (L_g = 55 nm, t_b = 17 nm, L_{sd} = 1 μm). (a) I-V and (b) transfer characteristics.

contact resistance (R_c < 0.2 Ω·mm). An excellent pinch-off and gate modulation of the channel current were also confirmed. Figure 4-28(b) shows a high peak transconductance of 500 mS/mm owing to the vertical gate-recess while keeping the gate leakage current below 1 μA/mm thanks to the gate dielectric formed by the O_2 plasma treatment. On the other hand, the appreciable output conductance and drain induced barrier lowering (DIBL) observed in this device indicate that the gate-to-channel distance needs to be further scaled to reduce the short-channel effects.

4.5.2. RF characteristics

The RF performance was characterized from 0.5 to 40 GHz using an Agilent Technologies N5230A network analyzer. The network analyzer was calibrated with a short-open-load-through (SOLT) standard and the calibration was verified by ensuring that both S_{12} and S_{21} of the through standard are less than ±0.01 dB and that both S_{11} and S_{22} are below -45 dB within the measured frequency range after the calibration. On-wafer open and short structures were used to de-embed the effect of parasitic pad capacitances and inductances as described in this chapter. A record f_T of 225 GHz was obtained in devices with a source-to-drain distance of 1 μm by extrapolating

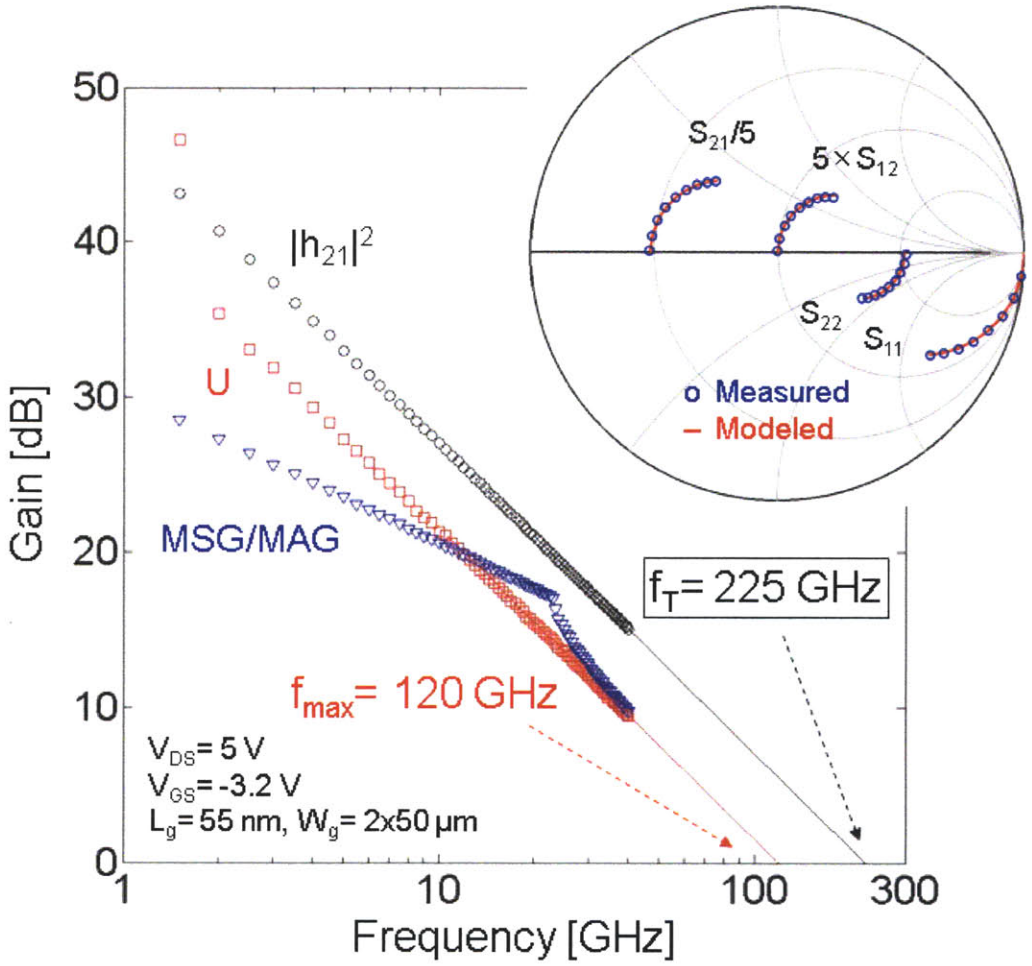


Figure 4-29. RF characteristics ($|h_{21}|^2$, Mason's unilateral gain U, and *MSG/MAG*) against frequency for an AlGaN/GaN HEMT with L_g = 55 nm. The inset shows comparison of measured and modeled S-parameters. The relatively low f_{max} values in our devices are mainly due to the large gate resistance (R_g).

$|h_{21}|^2$ with a slope of -20 dB/dec using a least-square fit as shown in Figure 4-29. For devices with a source-to-drain distance of 2 µm, the f_T was 205 GHz.

The accuracy of the f_T value was confirmed by carefully examining the validity of the extracted small-signal equivalent circuit parameters with respect to their expected values (g_m, C_{gs}, C_{gd}, etc). For comparison, the highest f_T reported so far in nitride transistors was 190 GHz in 6-nm barrier AlGaN/GaN HEMTs with 60-nm gate length [3]. The f_T of 225 GHz reported in this work was achieved in spite of much higher AlGaN barrier thickness (t_b = 17 nm) and stronger short-channel effects. This implies a great potential of the advanced gate technology developed in this work to further improve f_T. Figure 4-30 shows f_T and f_{max} at different bias points and the maximum f_T and f_{max} values were found as 225 GHz and 125 GHz, respectively. The bias

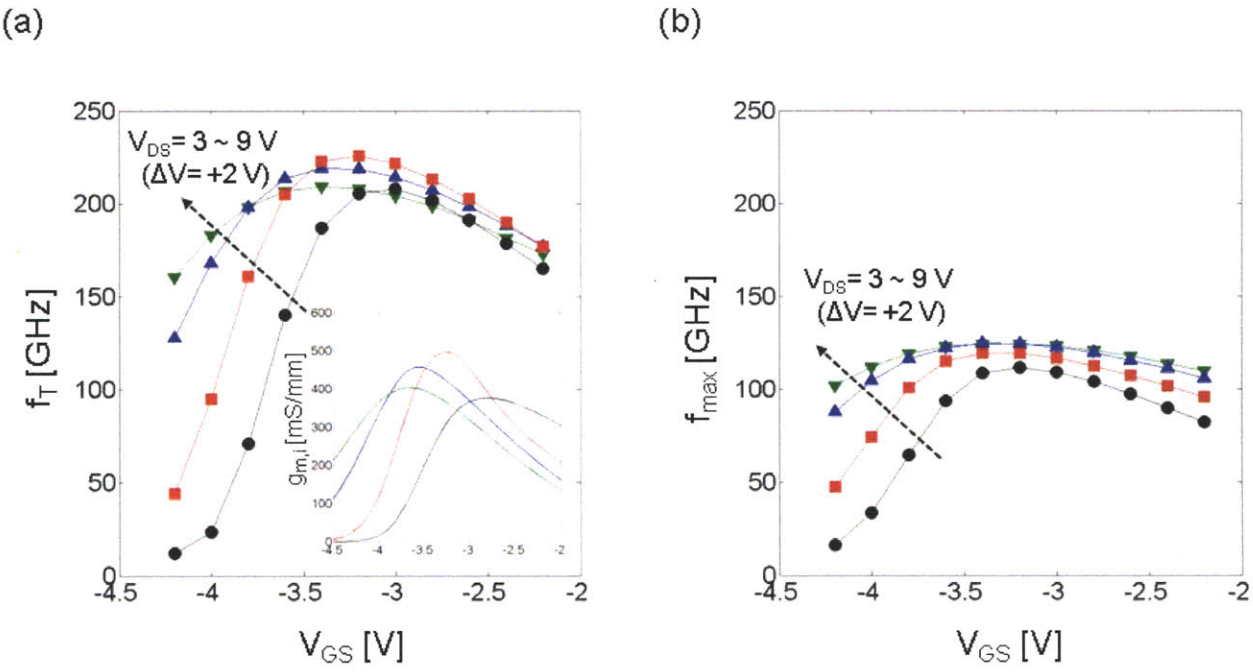


Figure 4-30. Bias dependent (a) f_T and (b) f_{max} values of the same device in Figure 4-29. The bias dependency of f_T was similar to that of intrinsic g_m.

dependency of f_T was similar to that of intrinsic g_m. The relatively low f_{max} values in our devices were mainly attributed to the high gate resistance due to the small cross-section area of the fabricated gate metal, which causes more than ten times higher gate resistance than conventional T-shape gate.

The small-signal equivalent circuit parameters of the fabricated devices were carefully extracted as described in previous sections and the accuracy of the extracted values was verified by comparing the S-parameters of the extracted equivalent circuit with the measured S-parameters over the measured frequencies. The Advanced Design System (ADS) software was used in these simulations.

We first studied the effect of lateral (L_g) and vertical (t_b) scaling to the device characteristics as shown in Figure 4-31. Without any vertical gate recess (t_b = 22 nm), the peak transconductance is reduced with L_g scaling due to the strong short-channel effects [39]. This g_m roll-off also significantly degrades f_T and its scaling behavior. The short-channel effects can be minimized by increasing the aspect ratio (L_g / t_b). As the AlGaN thickness is reduced by the gate recess, the aspect ratio increases and both g_m roll-off and scaling behavior of f_T gradually improves. For

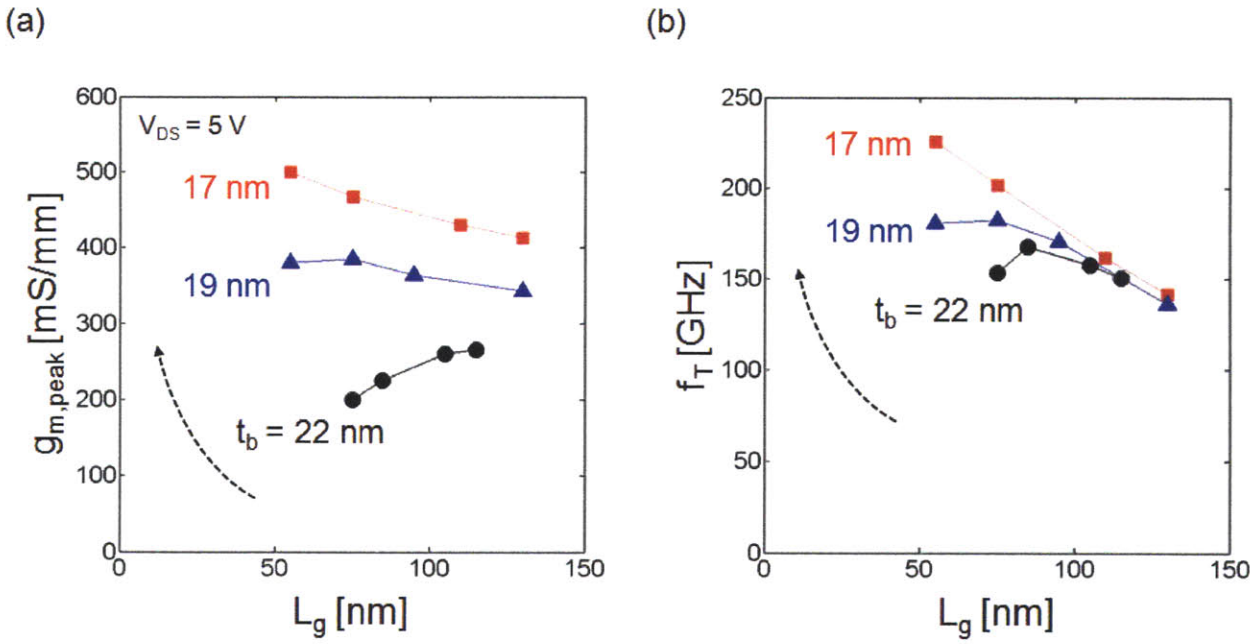


Figure 4-31. (a) Measured DC $g_{m,peak}$ and (b) f_T as a function of L_g for different t_b. The $g_{m,peak}$ and f_T drop as reducing L_g is mainly due to short-channel effects. This drop can be mitigated by reducing t_b, which increases the aspect ratio of the gate (L_g / t_b) and reduces short-channel effects.

example, when L_g = 75 nm, the f_T can be increased by more than 30 % when reducing the AlGaN barrier thickness from 22 nm (f_T = 153 GHz) to 17 nm (f_T = 201.5 GHz). Therefore, it is very important to achieve harmonious scaling of both L_g and t_b simultaneously to maximize f_T in AlGaN/GaN HEMTs.

A detailed delay analysis was performed based on the extracted small-signal parameters to identify how to improve f_T even more. The f_T of a HEMT is inversely proportional to the total delay (τ_{total}), and τ_{total} can be divided into three different components: intrinsic delay (τ_{int}), extrinsic delay (τ_{ext}), and parasitic delay (τ_{par}) [40]:

$$\tau_{total} = \frac{1}{2\pi f_T} = \tau_{int} + \tau_{ext} + \tau_{par} \tag{4-19}$$

$$f_T = \frac{1}{2\pi} \frac{g_m}{C_{gs} + C_{gd} + g_m C_{gd}(R_S + R_D)\left[1 + (1 + \frac{C_{gs}}{C_{gd}})\frac{g_o}{g_m}\right]} \tag{4-20}$$

where τ_{int} is the time taken by the electrons to cross the intrinsic channel region (L_g) right underneath the gate, τ_{ext} is associated with the external fringing gate capacitances and interpreted as the additional transit time due to the extended channel region (ΔL_g), and τ_{par} is the RC time

delay needed to charge and discharge the remaining parasitic portion of the channel region. Each delay can be calculated from extracted small-signal parameters (Figure 4-32) as

$$\tau_{int} = \frac{C_{gs,i} + C_{gd,i}}{g_m} \tag{4-21}$$

$$\tau_{ext} = \frac{C_{gs,ex} + C_{gd,ex}}{g_m} \tag{4-22}$$

$$\tau_{par} = C_{gd}(R_S + R_D)\left[1 + (1 + \frac{C_{gs}}{C_{gd}})\frac{g_o}{g_m}\right] \tag{4-23}$$

where the internal gate capacitances ($C_{gs,i}$, $C_{gd,i}$) without the external gate fringing capacitances ($C_{gs,ex}$, $C_{gd,ex}$) were obtained from the scaling behavior of C_{gs} and C_{gd} as shown in Figure 4-33. Figure 4-34 shows the delay analysis of the devices with 17 nm AlGaN barrier thickness as a function of the gate length. It is found that even for a device with an L_g = 55 nm which demonstrated a record f_T of 225 GHz, 60 % of the total delay was dominated by the intrinsic delay. Therefore, more aggressive scaling of both L_g and t_b, combined with the O_2 plasma treatment will further increase f_T in AlGaN/GaN HEMTs. It is expected that an f_T of 300 GHz is achievable by applying current technology and further scaling L_g without additional efforts to improve the parasitic delays (e.g. R_s and R_d). Much higher performance with f_T in excess of 400 GHz can be obtained by simultaneously reducing device dimensions and parasitic components as shown in Figure 4-35.

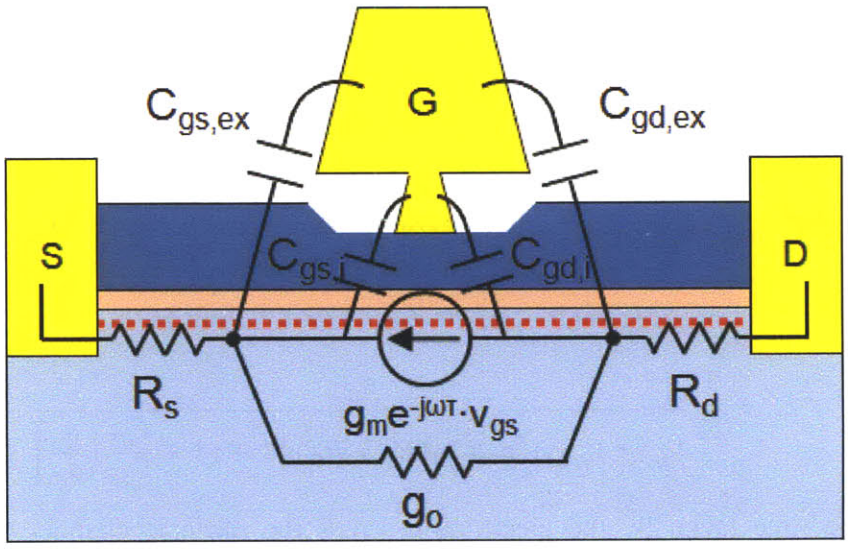


Figure 4-32. A schematic cross-section of AlGaN/GaN HEMTs with intrinsic small-signal parameters for delay analysis.

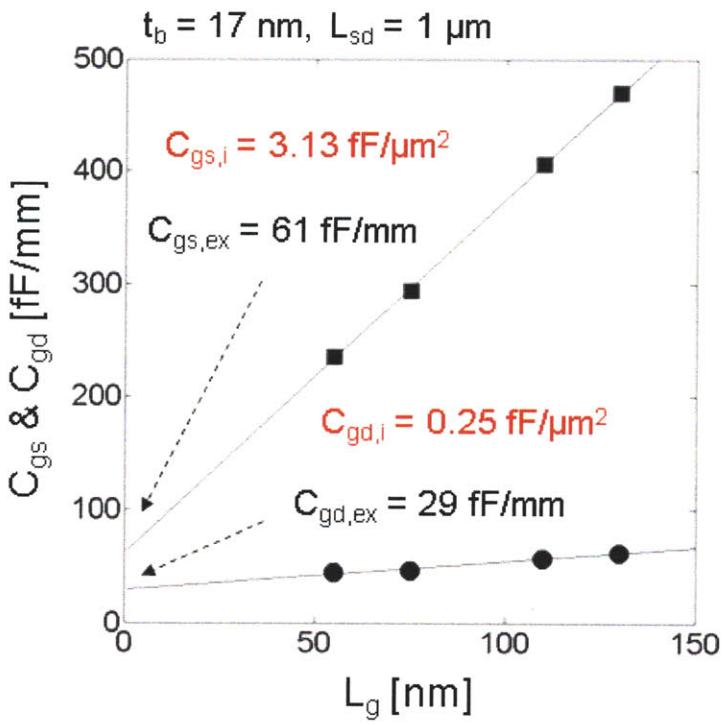


Figure 4-33. External gate fringing capacitances ($C_{gs,ex}$, $C_{gd,ex}$) and internal gate capacitances ($C_{gs,i}$, $C_{gd,i}$) were obtained from the scaling behavior of C_{gs} and C_{gd}.

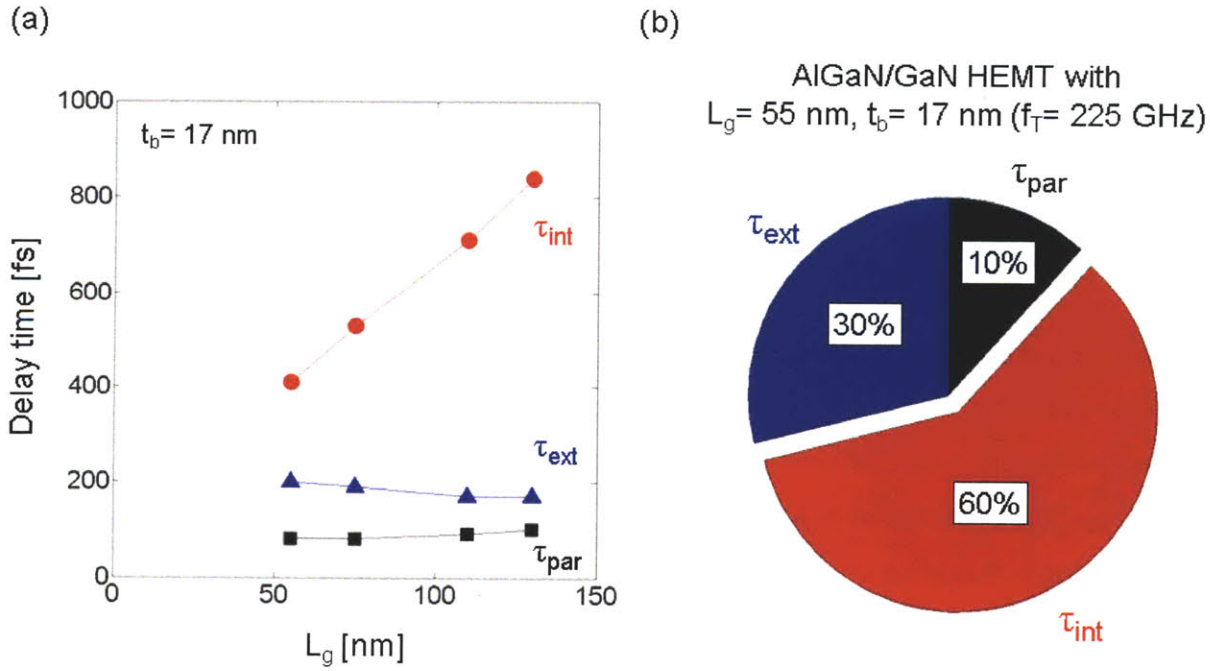


Figure 4-34. (a) Delay analysis of the devices with 17 nm AlGaN barrier thickness as a function of the gate length. (b) It is found that even for a device with an L_g = 55 nm which demonstrated a record f_T of 225 GHz, 60 % of the total delay was dominated by the intrinsic delay τ_{int}.

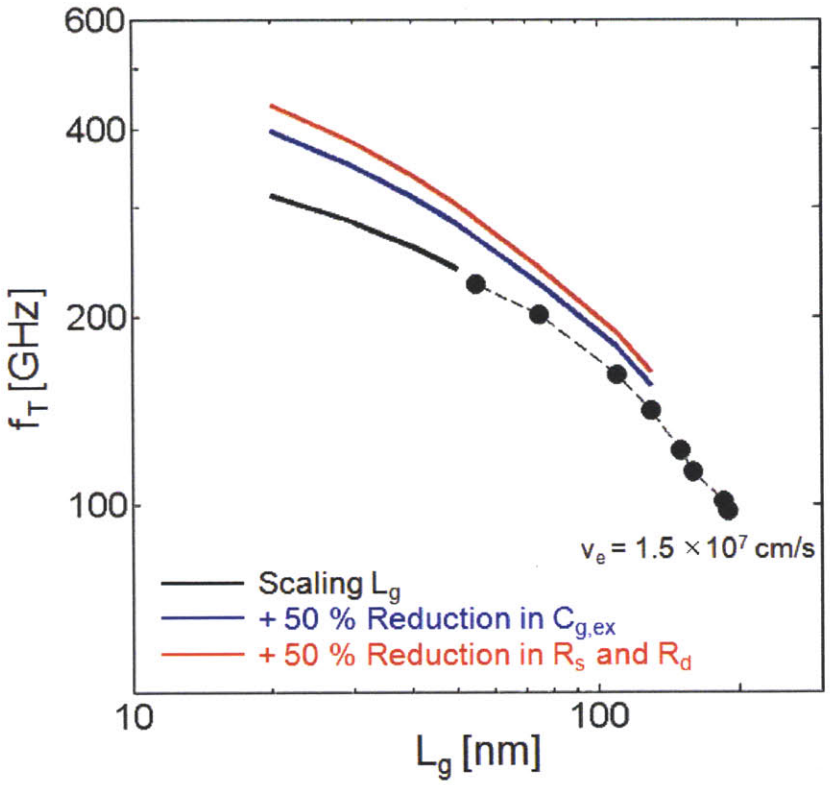


Figure 4-35. Projection of f_T by further improving device parameters such as L_g, $C_{g,ex}$, R_s, and R_d. $f_T > 400$ GHz is likely achievable in GaN HEMTs

The electron velocity can be calculated from the definition of the intrinsic delay Eq. (4-21). In our devices the extracted electron velocity was $1.4 \sim 1.5 \times 10^7$ cm/s.

$$\tau_{int} = \frac{C_{gs,i} + C_{gd,i}}{g_m} = \frac{L_g}{v_e} \tag{4-24}$$

Finally, Figure 4-36 shows the evolution of f_T for the past 15 years since the first high frequency performance of GaN HEMTs was reported in 1994. The f_T of GaN HEMTs has greatly evolved in a short period of time owing to matured material quality and improved fabrication technology. The evolution of f_T in GaN HEMTs marches on continuously by applying new ideas and technologies based on solid understanding of this material system.

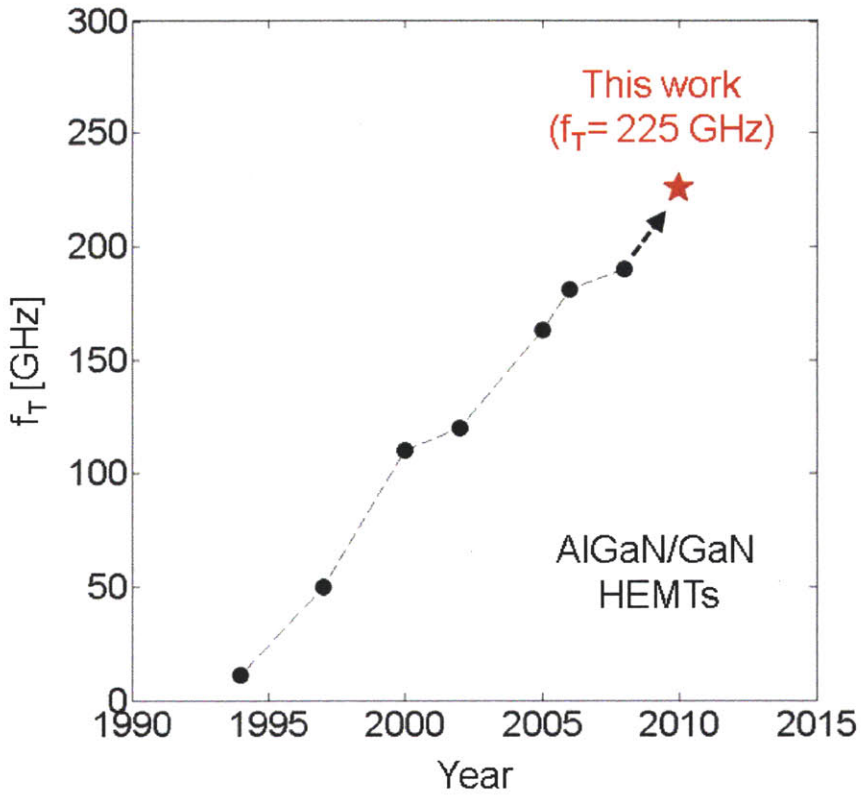


Figure 4-36. Evolution of f_T in AlGaN/GaN HEMTs for the past 15 years. A state-of-the-art frequency performance is reported in this work.

4.6. References

[1] T. Palacios, E. Snow, Y. Pei, A. Chakraborty, S. Keller, S. P. DenBaars, and U. K. Mishra, "Ge-spacer technology in AlGaN/GaN HEMTs for mm-waver applications," *IEEE IEDM Tech. Digest*, 2005.

[2] M. Micovic, A. Kurdoghilian, P. Hashimoto, M. Hu, M. Antcliffe, P. J. Willadsen, W. S. Wong, R. Bowen, I. Milosavljevic, A. Schmitz, Y. Yoon, M. Wetzel, and D. H. Chow, "GaN HFET for W-band power applications," *IEEE IEDM Tech. Digest*, 2006.

[3] M. Higashiwaki, T. Mimura, and T. Matsui, "AlGaN/GaN heterostructure field-effect transistors on 4H-SiC substrates with current-gain cutoff frequency of 190 GHz," *Appl. Phys. Express*, vol. 1, 021103, 2008.

[4] J. W. Chung, W. E. Hoke, E. M. Chumbes, and T. Palacios, "AlGaN/GaN HEMT with 300 GHz f_{max}," *IEEE Electron Device Lett.*, vol. 31, no. 3, pp. 195-197, Mar. 2010.

[5] A. Endoh, Y. Yamashita, N. Hirose, K. Hikosaka, T. Matsui, S. Hiyamizu, and T. Mimura, "High performance AlGaN/GaN metal-insulator-semiconductor high electron mobility transistors fabricated using SiN/SiO_2/SiN triple-layer insulators," *Jpn. J. Appl. Phys.*, vol. 45, no. 4B, pp. 3364-3367, Apr. 2006.

[6] Y. Yamashita, A. Endoh, N. Hirose, K. Hikosaka, T. Matsui, S. Hiyamizu, and T. Mimura, "Effect of bottom SiN thickness for AlGaN/GaN metal-insulator semiconductor high electron mobility transistors using SiN/SiO_2/SiN triple-layer insulators," *Jpn. J. Appl. Phys.*, vol. 45, no. 26, pp. L666-L668, Jun. 2006.

[7] G. H. Jessen, R. C. Fitch, J. K. Gillespie, G. Via, A. Crespo, D. Langley, D. J. Denninghoff, M. Trejo, and E. R. Heller, "Short-channel effect limitations on high-frequency operation of AlGaN/GaN HEMTs for T-gate devices," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2589-2597, 2007.

[8] D. Guerra, R. Akis, F. A. Marino, D. K. Ferry, S. M. Goodnick, and M. Saraniti, "Aspect ratio impact on RF and DC performance of state-of-the-art short-channel GaN and InGaAs HEMTs," *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1217-1219, Nov. 2010.

[9] M. C. Foisy, P. J. Tasker, B. Hughes, and L. F. Eastman, "The role of inefficient charge modulation in limiting the current-gain cutoff frequency of the MODFET," *IEEE Trans. Electron Devices*, vol. 35, pp. 871-878, 1988.

[10] L. D. Nguyen, L. E. Larson, and U. K. Mishra, "Ultra-high-speed modulation-doped field-effect transistors: A tutorial review," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 494-518, Apr. 1992.

[11] E. J. Miller, X. Z. Dang, H. H. Wieder, P. M. Asbeck, E. T. Yu, G. J. Sullivan, and J. M. Redwing, "Trap characterization by gate-drain conductance and capacitance dispersion studies of an AlGaN/GaN heterostructure field-effect transistor," *J. Appl. Phys.*, vol. 87, no. 11, pp. 8070-8073, Jun. 2000.

[12] S. C. Binari, P. B. Klein, and T. E. Kazior, "Trapping effects in GaN and SiC microwave FETs," *Proceedings of the IEEE*, vol. 90, no. 6, pp. 1048-1058, Jun. 2002.

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

[14] S. Mizuno, Y. Ohno, S. Kishimoto, K. Maezawa, and T. Mizutani, "Large gate leakage current in AlGaN/GaN high electron mobility transistors," *Jpn. J. Appl. Phys.*, vol. 41, no. 8, pp. 5125-5126, Aug. 2002.

[15] T. Palacios, S. Rajan, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "Influence of the dynamic access resistance in the g_m and f_T linearity of AlGaN/GaN HEMTs," *IEEE Trans. Electron Devices*, vol. 52, no. 10, pp. 2117-2123, Oct. 2005.

[16] D. R. Greenberg and J. A. del Alamo, "Nonlinear source and drain resistance in recessed-gate heterostructure field-effect transistors," *IEEE Trans. Electron Devices*, vol. 43, no. 8, pp. 1304-1306, Aug. 1996.

[17] S. Karmalkar, D. M. Sathaiya, and M. S. Shur, "Mechanism of the reverse gate leakage in AlGaN/GaN high electron mobility transistors," *Appl. Phys. Lett.*, vol. 82, no. 22, pp. 3976-3978, Jun. 2003.

[18] J. Lu, Y. Wang, L. Ma, and Z. Yu, "A new small-signal modeling and extraction method in AlGaN/GaN HEMTs," *Solid-State Electronics*, vol. 52, pp. 115-120, Sept. 2008.

[19] M. Berroth and R. Bosch, "High-frequency equivalent circuit of GaAs FET's for large-signal application," *IEEE Trans. Microwave Theory Tech.*, vol. 39, no. 2, pp. 224-229, Feb. 1991.

[20] L. Shen, S. Heikman, B. Moran, R. Coffie, N.-Q. Zhang, D. Buttari, I. P. Smorchkova, S. Keller, S. P. DenBaars, and U. K. Mishra, "AlGaN/AlN/GaN high-power microwave HEMT," *IEEE Electron Device Lett.*, vol. 22, no. 10, pp. 457-459, Oct. 2001.

[21] D. A. Hodges, H. G. Jackson, and R. A. Saleh, "Analysis and design of digital integrated circuits," McGraw-Hill, 2004.

[22] R. Vetury, Y.-F. Wu, P. T. Fini, G. Parish, S. Keller, S. P. DenBaars, and U. K. Mishra, "Direct measurement of gate depletion in high breakdown (405V) AlGaN/GaN heterostructure field-effect transistors," *IEEE IEDM Tech. Digest*, pp. 55-58, 1998.

[23] S. Y. Chou and D. A. Antoniadis, "Relationship between measured and intrinsic transconductances of FET's," *IEEE Trans. Electron Devices*, vol. ED-34, no. 2, pp. 448-450, Feb. 1987.

[24] T. Hashizume and R. Nakasaki, "Discrete surface state related to nitrogen-vacancy defect on plasma-treated GaN surfaces," *Appl. Phys. Lett.*, vol. 80, no. 24, pp. 4564-4566, Jun. 2002.

[25] J. Neugebauer and C. G. Van de Walle, "Atomic geometry and electronic structure of native defects in GaN," Phys. Rev. B, vol. 50, no. 11, pp. 8067-8070, Sept. 1994.

[26] T. Hashizume and H. Hasegawa, "Effects of nitrogen deficiency on electronic properties of AlGaN surfaces subjected to thermal and plasma processes," *Appl. Surface Science*, vol. 234, pp. 387-394, Jul. 2004.

[27] K. J. Choi and J.-L. Lee, "Interpretation of transconductance dispersion in GaAs MESFET using deep level transient spectroscopy," *IEEE Trans. Electron Devices*, vol. 48, no. 2, pp. 190-195, Feb. 2001.

[28] S. Mitra, M. V. Rao, and K. A. Jones, "Transconductance frequency dispersion meaaurements on fully implanted 4H-SiC MESFETs," *Solid-State Electronics*, vol. 48, no. 1, pp. 143-147, Jan. 2004.

[29] S. S. Islam, F. M. Anwar, and R. T. Webster, "A physics-based frequency dispersion model of GaN MESFETs," *IEEE Trans. Electron Devices*, vol. 51, no. 6, pp. 846-853, Jun. 2004.

[30] K. J. Choi, J.-L. Lee, and H. M. Yoo, "Effects of deep levels on tronsconductance dispersion in AlGaAs/InGaAs pseudomorphic high electron mobility transistor," *Appl. Phys. Lett.*, vol. 75, no. 11, pp. 1580-1582, Sept. 1999.

[31] P. H. Ladbrooke and S. R. Blight, "Low-field low-frequency dispersion of transconductance in GaAs MESFET's with implications for other rate-dependent anomalies," *IEEE Trans. Electron Devices*, vol. 35, no. 3, pp. 257-267, Mar. 1988.

[32] W. Kruppa and J. B. Boos, "Low-frequency transconductance dispersion in InAlAs/InGaAs/InP HEMT's with single- and double-recessed gate structures," *IEEE Trans. Electron Devices*, vol. 44, no. 5, pp. 687-692, May. 1997.

[33] R. Chu, C. S. Suh, M. H. Wong, N. Fichtenbaum, D. Brown, L. McCarthy, S. Keller, F. Wu, J. S. Speck, and U. K. Mishra, "Impact of CF_4 plasma treatment on GaN," *IEEE Electron Device Lett.*, vol. 28, no. 9, pp. 781-783, Sept. 2007.

[34] Y. Cai, Y. Zhou, K. M. Lau, and K. J. Chen, "Control of threshold voltage of AlGaN/GaN HEMTs by fluoride-based plasma treatment: From depletion mode to enhancement mode," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2207-2215, Sep. 2006.

[35] S. Pal, R. Mahapatra, S. K. Ray, B. R. Chakraborty, S. M. Shivaprasad, S. K. Lahiri, D. N. Bose, "Microwave plasma oxidation of gallium nitride," *Thin Solid Films*, vol. 425, no. 1-2, pp. 20-23, Feb. 2003.

[36] J. W. Chung, J. C. Roberts, E. L. Piner, and T. Palacios, "Effect of gate leakage in the subthreshold characteristics of AlGaN/GaN HEMTs," *IEEE Electron Device Lett.*, vol. 29, no. 11, pp. 1196-1198, Nov. 2008.

[37] P. Kordos, D. Gregusova, R. Stoklas, K. Cico, and J. Novak, "Improved transport properties Al_2O_3/AlGaN/GaN metal-oxide-semiconductor heterostructure field-effect transistor," *Appl. Phys. Lett.*, vol. 90, no. 12, 123513, Mar. 2007.

[38] R. G. Gordon, D. Hausmann, E. Kim, and J. Shepard, "A kinetic model for step coverage by atomic layer deposition in narrow holes or trenches," *Chem. Vap. Deposition*, vol. 9, no. 2, pp. 73-78, Mar. 2003.

[39] O. Breitschadel, L. Kley, H. Grabeldinger, J. T. Hsieh, B. Kuhn, F. Scholz, and H. Schweizer, "Short-channel effects in AlGaN/GaN HEMTs," *Materials Science and Engineering B*, vol. 82, no. 1-3, pp. 238-240, May 2001.

[40] J. A. del Alamo and D.-H. Kim, "Is f_T over 1 THz possible?," *Topical Workshop on Heterostructure Microelectronics*, Aug. 2009.

Chapter 5. On-wafer Integration of GaN and Si(100) Electronics

5.1. Introduction

The integration of III-V compound semiconductors and silicon (100) CMOS technologies has been a long pursued goal. A robust heterogeneous integration technology would make the outstanding analog and mixed-signal performance of compound semiconductor electronics available to VLSI chips where they are difficult to be implemented by Si technology.

GaN-based devices are one of the best candidates for integration with Si. While Si electronics has shown unsurpassed levels of scaling and circuit complexity, GaN devices offer excellent high frequency/power performance as well as outstanding optoelectronic properties [1][2]. The ability to combine these two material systems in the same chip and in very close proximity would allow unprecedented flexibility for advanced applications.

To successfully integrate GaN with Si on the same wafer, several requirements need to be satisfied. First of all, it is desirable to have the integration with (100)-oriented Si. Due to its reduced surface state density, the (100) orientation (with no miscut) is the preferred one for the Si wafers used in microprocessors and VLSI chips. Secondly, the entire process flow needs to be fully compatible with standard fabrication process for the Si devices. For example, the process has to meet the high thermal budget of Si processing which is normally 800 ~ 1000 °C for dopant activation and the entire process needs to be scalable to large wafer diameters, such as 6, 8, and 12 inches. Thirdly, the heterogeneous integration should not result in any degradation or cross contamination in the material properties of either the Si or the GaN devices. Finally, the integration should be achieved in a seamless way, enabling the very high proximity between devices and ensuring high interconnection density.

In this chapter, we demonstrate the first on-wafer integration of Si(100) MOSFETs and GaN HEMTs satisfying all the requirements described above. Many different applications are

envisioned for the proposed technology and some of them are discussed at the end of the chapter, namely on-wafer GaN/Si power regulators and enhancement mode power transistors.

5.2. Challenges and new approach

Previously, several authors have reported heterogeneous integration of Si and GaAs devices (i.e. field effect transistors, light emitting diodes) by the low-temperature selective epitaxial growth of GaAs on a miscut Si(100) substrate [3][4][5]. With similar technology, several groups have reported the growth of GaN structures on miscut Si(100) or Si(110) substrates by molecular beam epitaxy (MBE) [6][7] and metalorganic vapor phase epitaxy (MOVPE) [8][9]. However, this approach is challenging because of the difficulty of growing high quality wurtzite GaN on these Si substrates due to the large lattice mismatch between GaN and Si [9]. Also, typically at least a few microns of GaN thickness is required to grow high quality GaN on Si substrate and when it comes to integrate GaN and Si devices this large step height causes lithography problems to fabricate these devices in very close proximity. Moreover, the use of miscut substrates increases the density of surface states in the Si material, degrading the performance of Si electronics designed therein.

A different approach to achieve the heterogeneous integration of GaN devices with a Si wafer involves transferring an already grown GaN epilayer onto a Si(100) substrate through the removal of the original substrate and subsequent bonding to the Si(100) wafer. Laser lift-off and Au/In/Au bonding layer was used in [10]. Wafer bonding of GaN with Si substrates using $PdIn_3$ and AuGe as an interlayer has also been reported [11][12], however none of these hybrid wafers satisfies the thermal budget of Si processing (~ 1000 °C) since the melting points of $PdIn_3$ and AuGe are ~ 660 °C and ~ 360 °C, respectively. Recently, our group demonstrated the robust wafer bonding of GaN and Si(100) wafers through the use of a SiO_2 interlayer [13]. The thermal stability of this bonding was successfully tested up to 1000 °C, a sufficient thermal budget for Si and GaN processing.

In this chapter, we demonstrate the first integration of Si(100) MOSFETs and GaN HEMTs on the same wafer in very close proximity. The key enabling technology is the fabrication of a

Si(100)-GaN-Si(100) hybrid substrate through a wafer bonding and etch-back process. On this substrate, standard Si MOSFETs were first fabricated by following a conventional Si process technology. Then, the top Si layer was locally removed to expose the AlGaN surface, and standard GaN HEMTs were processed in those regions. It should be highlighted that in this technology, the Si devices were fabricated on Si(100) wafers without any miscut and standard fabrication processes were carried out for both Si and GaN devices without modifying any step. Due to the very high thermal stability of GaN [14], this process did not adversely affect the electrical properties of the embedded GaN layer.

5.3. Layer transfer technology

Wafer bonding offers unprecedented opportunities to enable the heterogeneous integration of GaN devices with other semiconductors. This technology allows us to place the high-quality GaN film on any position within the wafer stack without compromising the film growth conditions. In other words, the optimization of the GaN growth can be decoupled from the integration process of GaN and Si devices. For example, high quality AlGaN/GaN device structures can be grown on their optimum substrate, then placed together with Si(100) substrates using the wafer bonding technique to form a Si(100)-GaN-Si(100) hybrid substrate. This hybrid substrate is the platform for the on-wafer integration of GaN and Si(100) devices, thus it is very critical to have a reliable hybrid substrate to achieve a robust integration.

In this work, we have developed a new wafer bonding technology to fabricate a reliable Si(100)-GaN-Si(100) hybrid substrate. This technology is based on using a hydrogen silsesquioxane (HSQ) intermediate layer between GaN and Si(100) wafers

5.3.1. HSQ interlayer bonding

Wafer-to-wafer bonding has been widely studied in the context of Si microstructure fabrication. The three types of wafer bonding most commonly employed in Si microstructure fabrication are direct bonding, anodic bonding, and intermediate-layer bonding [15]. The *direct wafer bonding* relies on forces that naturally attract each other even without any external forces or fields or

intermediate layers when the wafer surfaces are atomically smooth and flat. Standard wafer bonding of Si wafers in air is attributed to relatively weak intermolecular attractive forces, such as van der Waals forces or forces associated with hydrogen bridge bonds [16]. Typically, a surface roughness below 10 Å and a wafer bow of less than 5 μm are required to ensure a good direct bonding between 4 inch wafers. It is also common practice for directly bonded wafers to go through thermal cycling to enhance the bonding strength. The *anodic wafer bonding* is accomplished by applying an electric field or a voltage of 200 ~ 1000 V between two wafers in direct contact to each other at elevated temperatures of 300 ~ 400 °C. The anodic bonding is often used to bond sodium-containing glass wafers since this method fosters sodium ionic conduction. For example, when bonding a glass wafer to a Si wafer under a high electric field and elevated temperatures, the mobile sodium ions (Na^+) in the glass migrate away from the bonded interface region, leaving behind less mobile oxygen ions at the sodium depletion region. Oxygen ions then diffuse into the silicon surface and form SiO_2 with its permanent covalent bond in the bonded region. The third bonding technology, *the intermediate-layer wafer bonding*, utilizes an interlayer between two wafers to promote the bonding. A wide range of interlayer materials can be used for this purpose including dielectrics, polymers, and metals. This method is typically more tolerant to the surface condition of the initial wafers than the two other bonding technologies.

Among the different types of wafer bonding, the intermediate-layer bonding is the most suitable for GaN and Si wafer bonding. The anodic bonding is not applicable due to the lack of ionic conduction in the GaN wafer. The direct bonding of GaN to a Si wafer is also difficult because of the relatively poor surface roughness of GaN and the large wafer bow typical of GaN wafers. Figure 5-1 shows the surface roughness of an as-grown GaN wafer measured by atomic force microscopy (AFM). Although the surface roughness of GaN is acceptable over small areas (rms surface roughness R_q = 1.6 nm for 2 × 2 μm^2), the surface roughness increases significantly when large areas are considered (R_q = 3.8 nm for 20 × 20 μm^2). When even larger areas are scanned by a white light interferometer measurement, the surface roughness of GaN becomes much worse (R_q = 11 nm for ~ 1 × 1 mm) as shown in Figure 5-2. Over this area, the average peak to valley height (R_z) was more than 50 nm and many irregular bumps were also observed. Another challenge for direct bonding is the wafer bow of GaN wafers. Figure 5-3 compares the

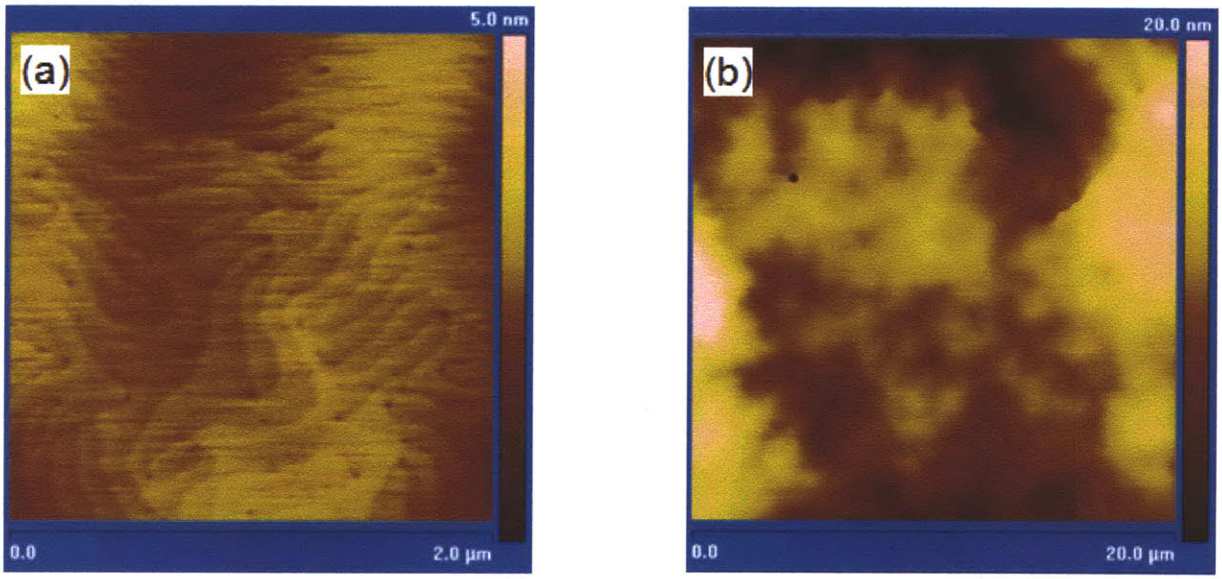


Figure 5-1. AFM image of as-grown GaN surface in (a) small area (2 × 2 μm) and (b) large area (20 × 20 μm). A surface rms surface roughness (R_q) was measured as 1.6 nm and 3.8 nm, respectively [17].

wafer bow of a 4 inch Si(100) wafer and a GaN epitaxial layer grown on a 4 inch Si(111) wafer, measured by laser profilometry. While the Si(100) wafer has less than 10 μm of wafer bow, the GaN wafer has more than 50 μm mainly due to the high strain induced by the lattice mismatch between GaN and Si(111) substrate and the large difference between the thermal expansion coefficients of GaN and Si. To overcome these challenges introduced by the poor surface roughness and the large wafer bow of GaN, we have developed a new intermediate-layer wafer bonding using a hydrogen silsesquioxane (HSQ) as an interlayer.

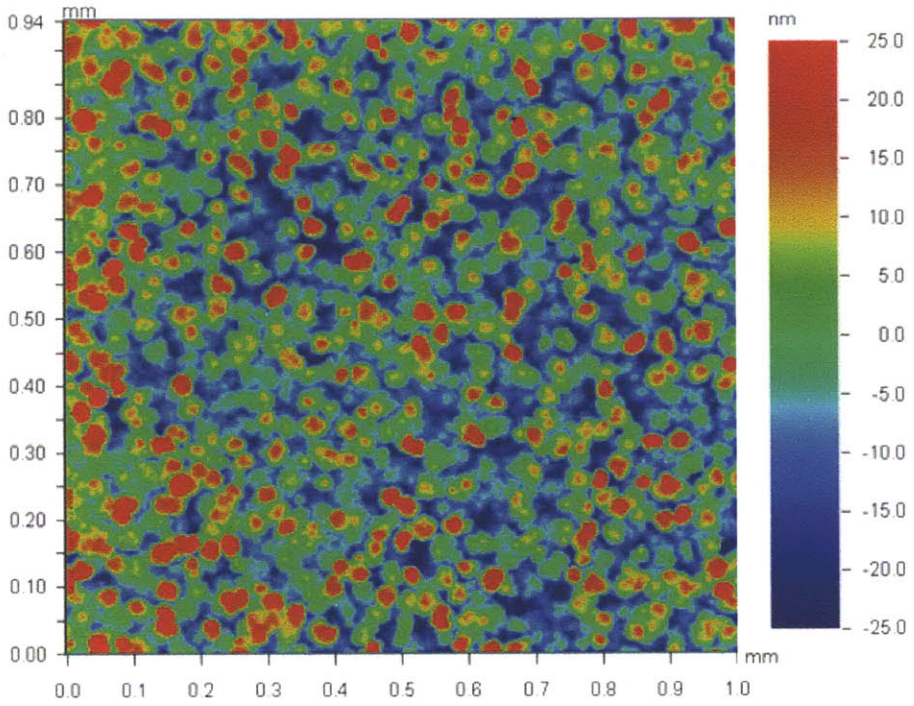


Figure 5-2. Large area (1 × 1 mm^2) interferometer measurement of the as-grown GaN surface. The rms surface roughness was measured as high as 11 nm [17].

Figure 5-4 illustrates the process of HSQ interlayer bonding. HSQ is a flowable oxide with excellent thermal stability, which stands the high thermal budget (800 ~ 1000 °C) required during the processing of GaN and Si devices. The HSQ film is simply spin-coated on the Si(100) carrier wafer to a thickness of less than 1000 Å and baked sequentially on hot plates at 150 °C and 200 °C for 1 minute each. Spin-coated HSQ has a very good surface uniformity (surface roughness <

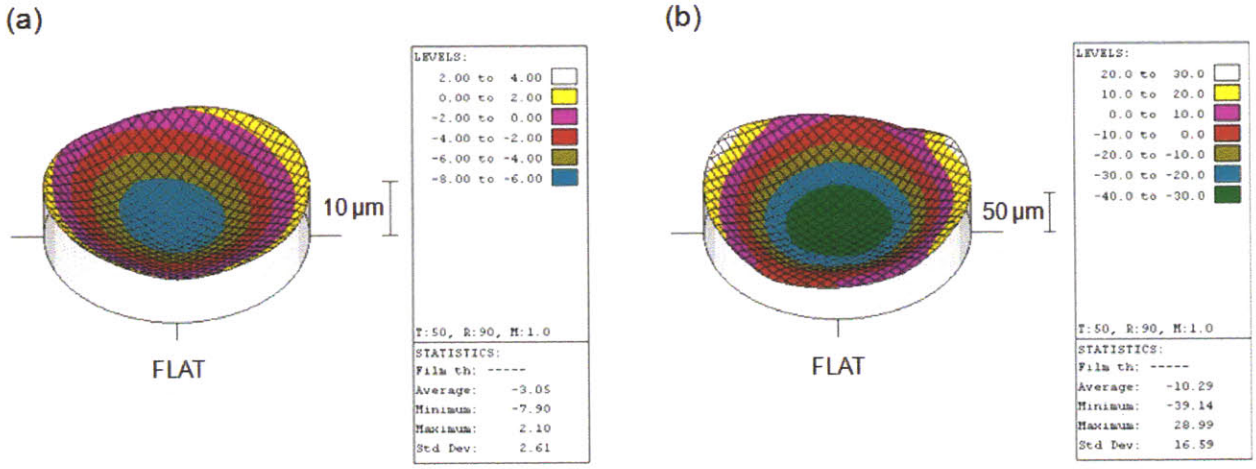


Figure 5-3. Wafer bow of (a) a 4 inch Si(100) wafer and (b) an as-grown 4 inch GaN on Si(111) substrate measured by laser profilometry (KLA-Tencor-Prometrix UV-1280).

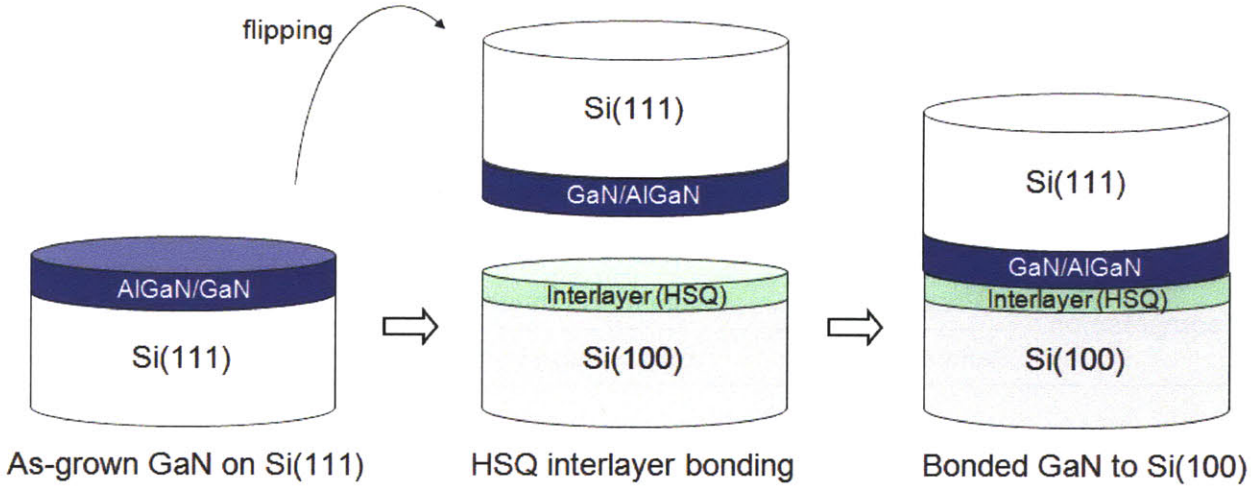


Figure 5-4. Main processing steps of the HSQ interlayer bonding developed in this work. Spin-coated HSQ is utilized for the adhesive interlayer. Thermal compression hardens the HSQ interlayer forming a robust bonding between GaN and Si(100) carrier wafer. It should be noted that other substrates such as SiC, sapphire, and glass can be also used as carrier wafers.

40 Å) which is necessary to have excellent contact and avoid any void formation at the bonding interface. Then, the HSQ-coated Si(100) substrate is put in close contact to the GaN wafer in vacuum and both are thermally compressed against each other at 400 °C for one hour. The elevated temperature hardens the HSQ layer and the uniformly applied compression forms a stable bond between the GaN wafer and the Si(100) carrier wafer as shown in Figure 5-5.

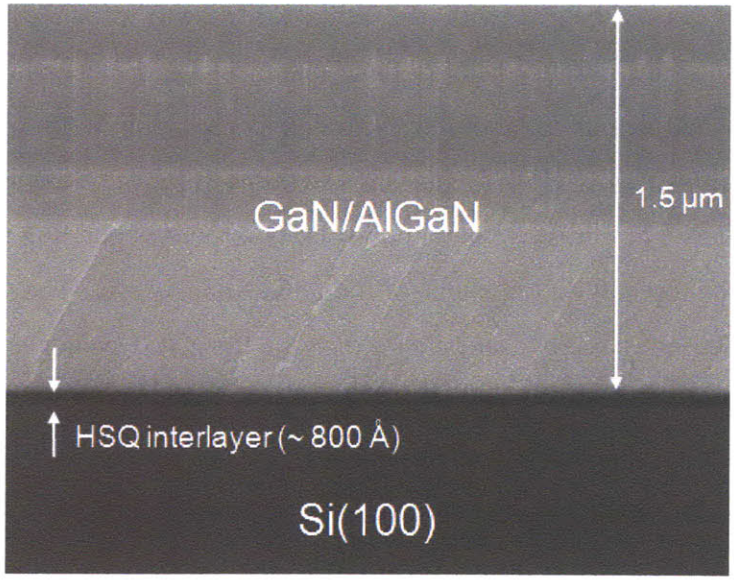


Figure 5-5. Cross-section scanning electron microscope (SEM) image of the interface between GaN and the Si(100) carrier wafer after the HSQ interlayer bonding. The HSQ interlayer enables a uniform and smooth interface.

Step	Parameter	Requirements
HSQ spin-coating	Particle	Avoid any particles from air, wafer, pipet, etc
	HSQ thickness	There exists a minimum critical thickness, typically > 500 Å
	HSQ uniformity	Dynamic dispense and planarization baking
Thermal compression	Void (air gap)	Make bonding at vacuum or low pressure condition
	Pressure	Adjust (too much pressure breaks the wafers)
	Temperature	High enough to harden HSQ, typically 400 °C

Table 5-1. Process considerations and related control parameters at each step of HSQ interlayer bonding.

To prevent any failure in the HSQ interlayer bonding, several processing parameters need to be carefully optimized. Some of them include the thickness of HSQ, bonding temperature, pressure, density of surface particles, air gaps at the interface and wafer bow. The optimum parameters and process considerations at each step of the HSQ interlayer bonding are summarized in Table 5-1.

5.3.2. Si(100)-GaN-Si(100) hybrid substrate

Figure 5-6 summarizes the main steps of the fabrication of Si(100)-GaN-Si(100) hybrid substrates. The fabrication begins with the epitaxial growth of an AlGaN/GaN transistor structure on a Si(111) substrate by metal-organic chemical vapor deposition (MOCVD). These structures were provided by our collaborator, Nitronex Corporation. In these samples, the AlGaN barrier had a total thickness of 175 Å and an Al composition of 26 %. Our technology then removes the original Si(111) substrate and applies HSQ interlayer bonding twice to have Si(100) substrates on both the top and bottom sides of the AlGaN/GaN layer. The top Si(100) layer was obtained from the active Si layer of an silicon-on-insulator (SOI) wafer and has a thickness of 200 nm. The doping of this layer sets the nMOS or pMOS characteristics of the future Si devices and in this work an n-type doping concentration of 10^{15} cm^{-3} was used. The Si substrate (both (111) and (100)-orientation) and buried oxide (BOX) removal in this process can be achieved by several different methods such as laser lift-off, smart cut, selective wet-etch, and selective dry-etch. In this work, we used a deep reactive ion etch (DRIE) with an SF_6-based plasma to selectively remove the original Si(111) substrate over GaN, XeF_2 to selectively remove Si(100)

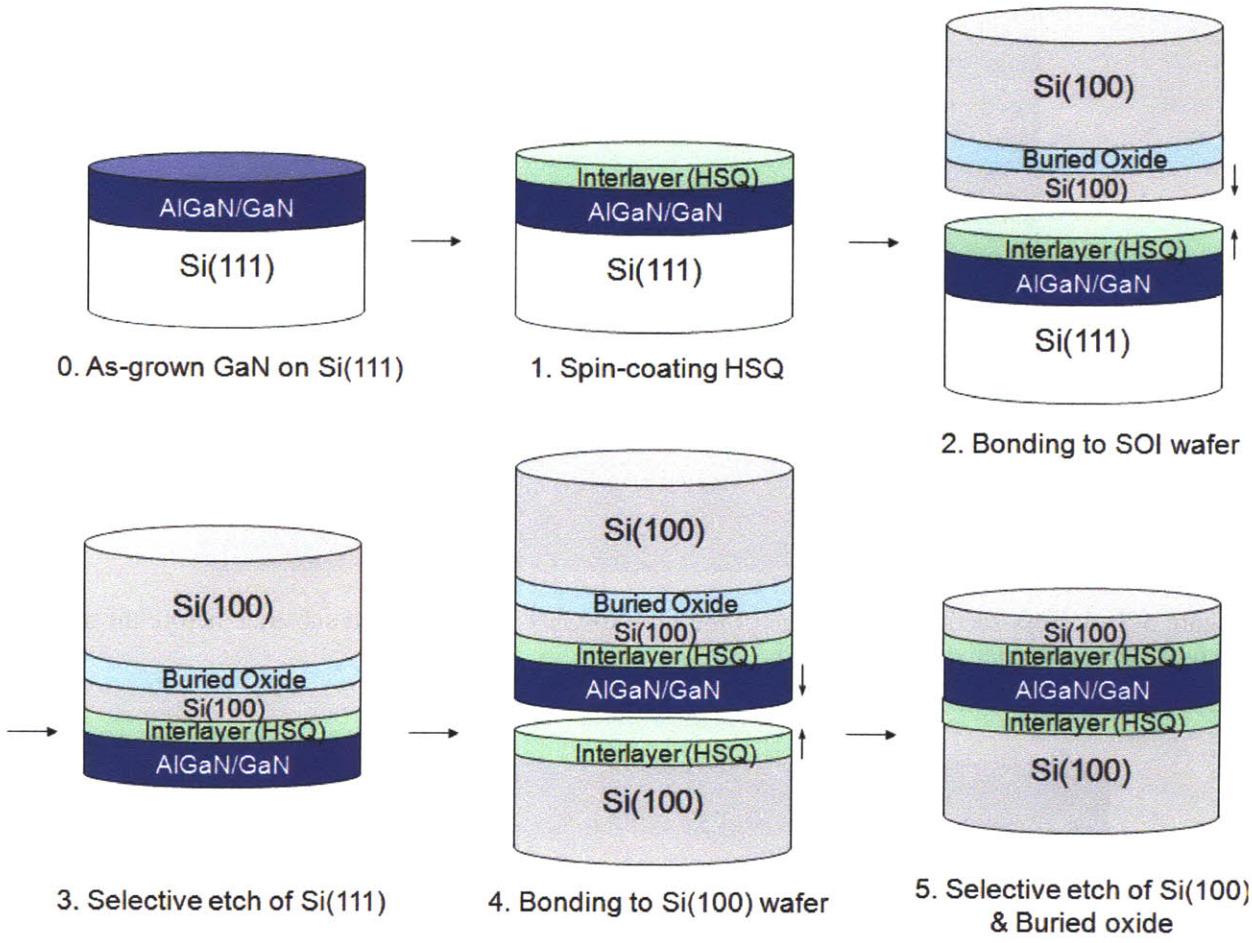


Figure 5-6. Schematic illustration of the main processing steps in the fabrication of Si(100)-GaN-Si(100) hybrid substrates through the layer transfer technology. The thin top Si(100) layer is obtained from the active Si layer of an Silicon-on-Insulator (SOI) wafer. The doping of this layer sets the nMOS or pMOS character of the fabricated devices. The fabrication process of the Si-GaN-Si hybrid substrates can be simplified by leaving the Si(111) substrate and skipping steps 3 and 4.

over SiO_2, and buffered oxide etch (BOE) to selectively etch SiO2 over Si(100). Figures 5-7 and 5-8 show the cross-section SEM of HSQ-bonded wafers before the etch-back process and the completed Si(100)-GaN-Si(100) hybrid substrate, respectively. It is noted that the fabrication process of the hybrid substrate can be simplified by leaving the Si(111) substrate and skipping steps 3 and 4 in Figure 5-6. Hybrid wafers with one inch in diameter have been processed in this work for the initial demonstration although it can be scaled to larger wafer sizes [17].

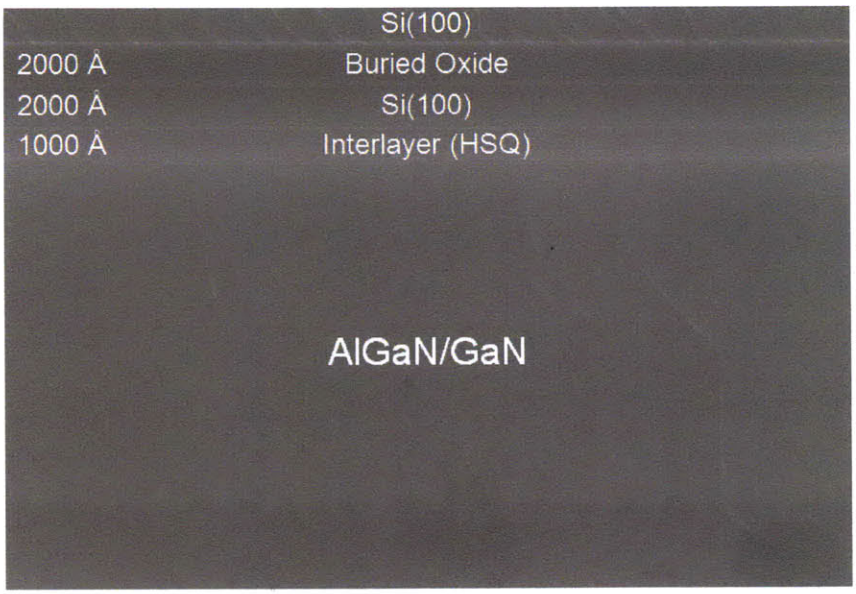


Figure 5-7. A cross-section SEM image of the Si(100)-GaN-Si(100) hybrid substrate before the selective etch-back process (step 4 in Figure 5-6).

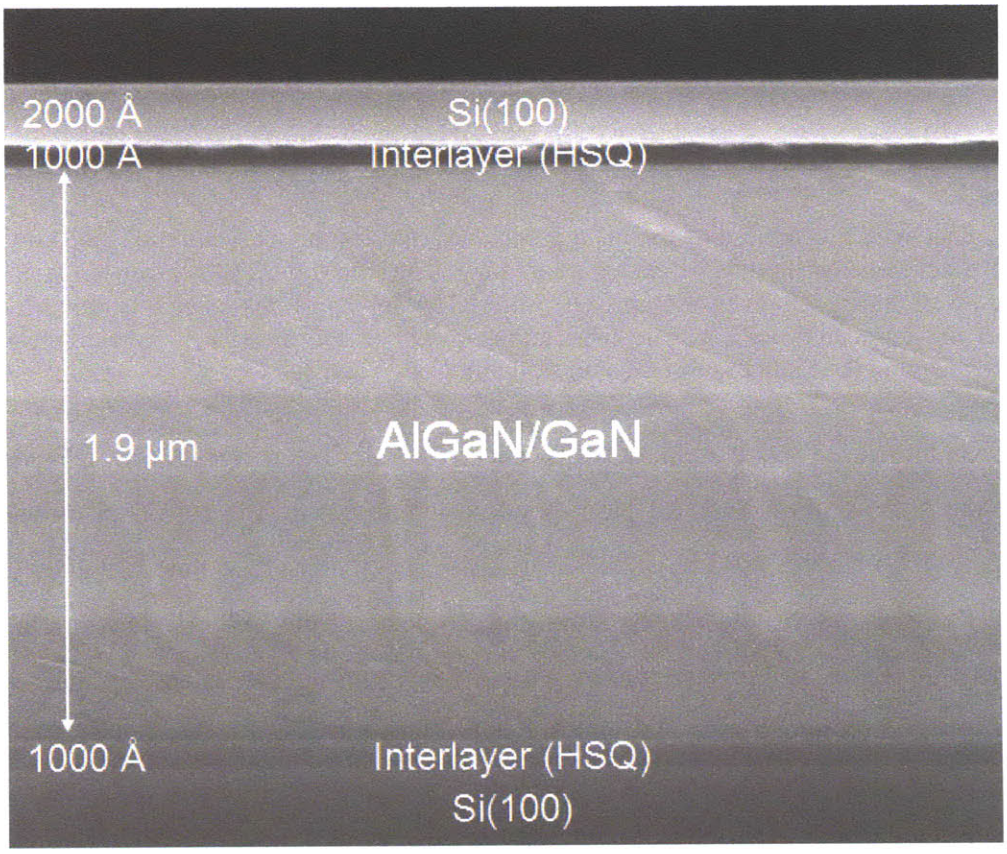


Figure 5-8. A cross-section SEM image of the successfully processed Si(100)-GaN-Si(100) hybrid substrate (step 5 in Figure 5-6).

This basic technology can also be used for the fabrication of novel N-face GaN HEMTs (i.e. reversed polarization field against normal Ga-face GaN HEMTs) and their integration with Si(100) electronics. This topic is discussed in [13].

5.4. On-wafer integration of GaN and Si(100) devices

Once the Si(100)-GaN-Si(100) hybrid substrate has been fabricated (Figure 5-8), device processing starts with the fabrication of Si MOSFETs. Figure 5-9 summarizes the entire process flow for Si MOSFET and GaN HEMT. In this work, we fabricated p-type Si MOSFETs. Device isolation was achieved by field oxide and a 10 nm gate oxide was formed by plasma enhanced chemical vapor deposition (PECVD). Undoped polycrystalline Si was subsequently deposited and anisotropically etched in a Cl_2 plasma to form the gate contact. The source, drain, and gate implantation was achieved at the same time with BF_2 at a dose of 4×10^{15} cm^{-2} and an implantation energy of 10 keV.

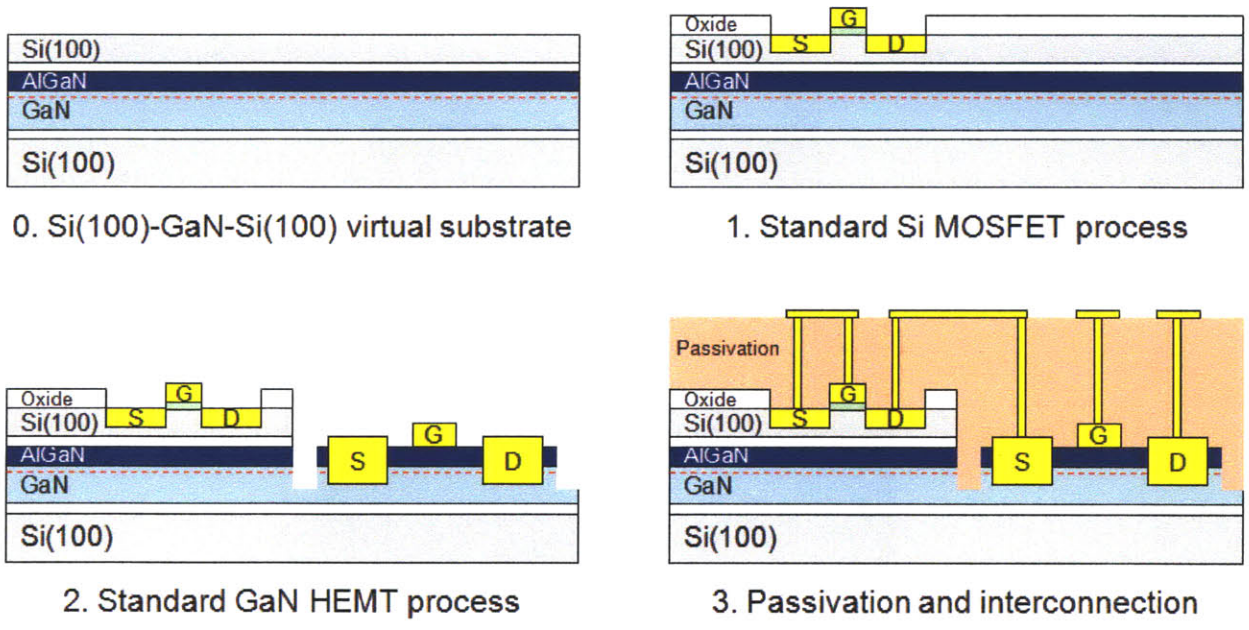


Figure 5-9. Process flow for the fabrication of standard Si MOSFETs and GaN HEMTs on a Si(100)-GaN-Si(100) hybrid substrate. The process is completed by the passivation and interconnection to build hybrid circuits.

After the standard Si MOSFET process, the AlGaN/GaN layer embedded in the hybrid substrate was exposed by etching the top Si(100) and HSQ layers using SF_6 plasma followed by buffered oxide etchant (BOE) in those regions where GaN devices are to be located. The etch selectivity between Si/SiO_2 and AlGaN is excellent and a smooth AlGaN surface was obtained after the etch. Once the AlGaN/GaN layer is exposed, the fabrication of GaN HEMTs is identical to a standard GaN HEMT process. A Ti/Al/Ni/Au multilayer was first deposited for the ohmic contacts. Ohmic metal alloying in the GaN HEMTs and dopant activation in the Si p-MOSFETs were simultaneously accomplished by rapid thermal annealing at 870 °C for 30 s in N_2 atmosphere. A Cl_2/BCl_3 plasma was used for the mesa isolation of the HEMT devices, and then a 2~3 μm-long gate was formed by photolithography and Ni/Au/Ni metallization. Figure 5-10 shows a top-down scanning electron micrograph (SEM) image of the integrated Si p-MOSFETs and GaN HEMT devices after this step. Figure 5-11 shows another SEM image of fabricated devices on the same wafer to build hybrid circuits. The separation between these two devices is just 4 μm. Finally, 500 nm of SiO_2 passivation layer was deposited by PECVD, contact vias were opened in the dielectric, and the contact pads and interconnections were metalized with Ti/Al. It is noted that although Au-bearing metallizations were used for the ohmic and Schottky contacts in this initial demonstration, our group has recently developed a Ti/Al/W-based ohmic metallization which is much more suitable for integration with Si devices.

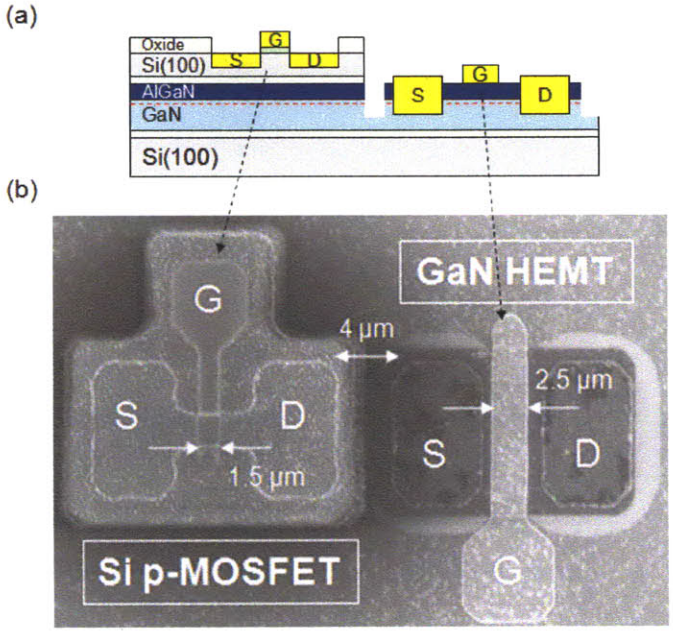


Figure 5-10. (a) A cross-section schematic of fabricated Si p-MOSFETs and GaN HEMTs. (b) A plan-view SEM image of the fabricated transistors.

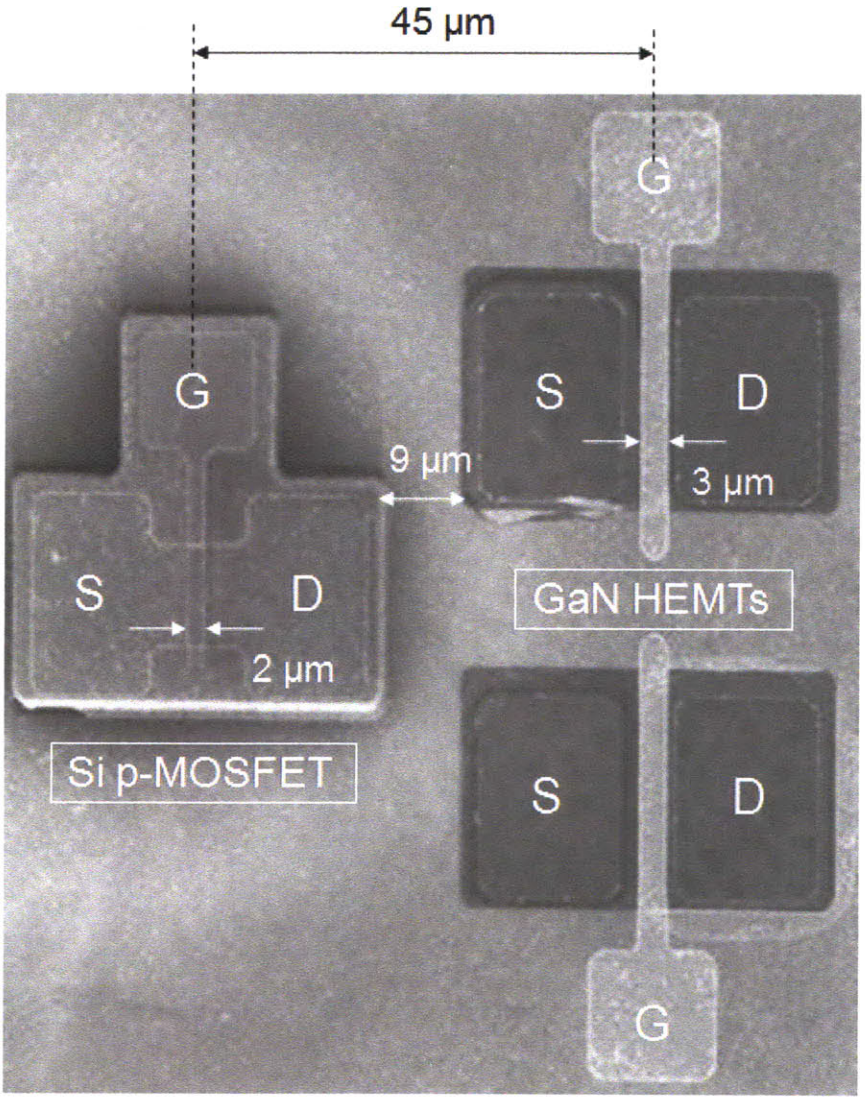


Figure 5-11. A plan-view SEM image of the fabricated transistors on the same wafer to build hybrid circuits.

The drain current versus drain voltage characteristics, as well as the transfer characteristics, of a typical Si p-MOSFET and GaN HEMT fabricated using the technology described above are shown in Figure 5-12. Both devices have good modulation of the drain current by the gate contact as well as low off-currents. However the long gate lengths of these devices limited their maximum output current. The use of a shorter gate length and higher doping levels in the Si active layer are expected to improve the performance of the devices.

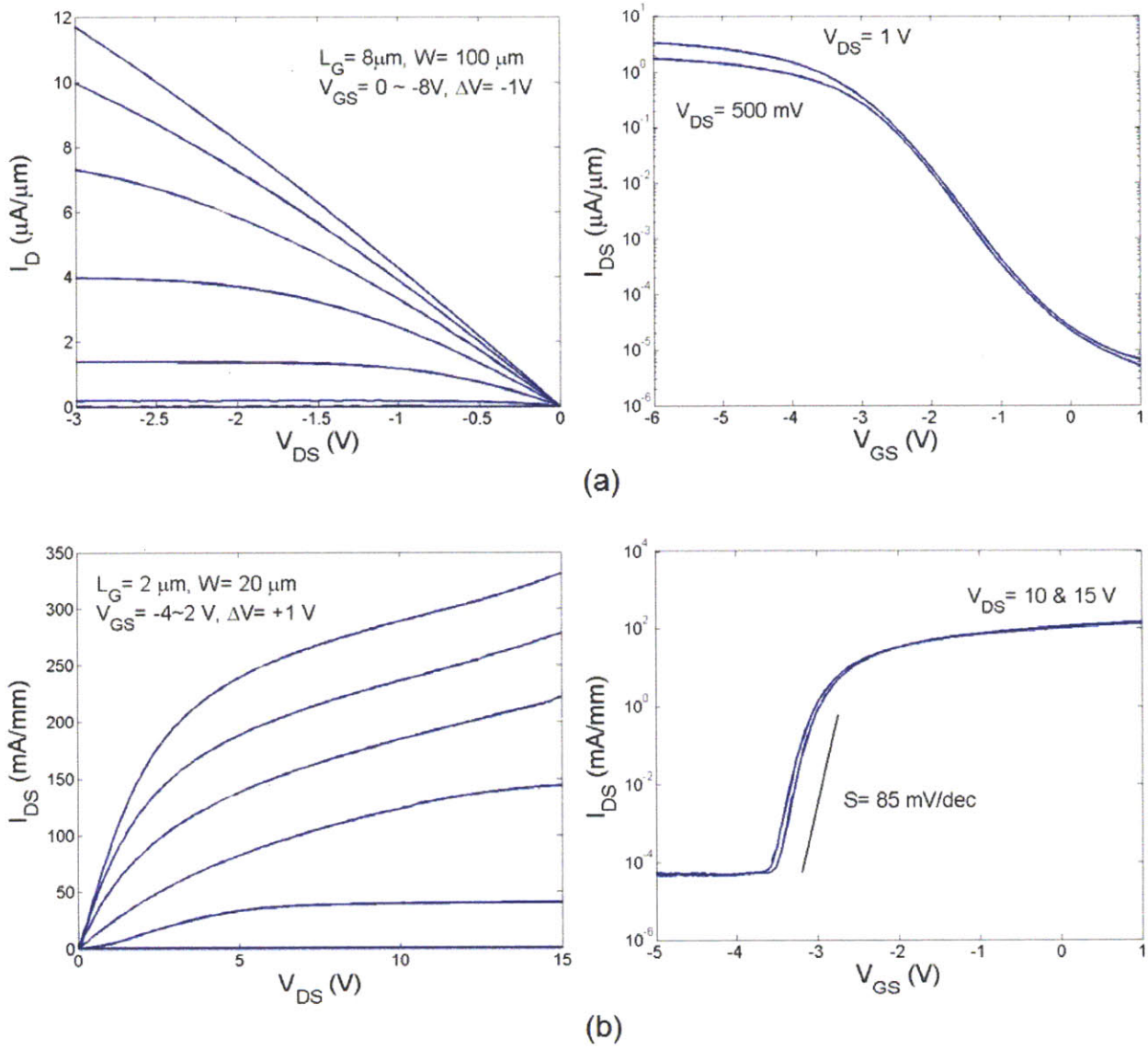


Figure 5-12. (a) DC current-voltage characteristics (left) and transfer curve (right) of a long channel Si p-MOSFET (L_g = 8 μm). (b) DC current-voltage characteristics (left) and transfer curve (right) of a GaN HEMT (L_g = 2 μm).

The effect of the entire fabrication process to the transport properties of the AlGaN/GaN epilayer was evaluated by transfer length measurement (TLM) and compared to conventional devices as shown in Figure 5-13. The contact resistance (R_c) did not change (R_c = 0.4 Ω·mm), however interestingly the sheet resistance (R_{sh}) was reduced by 24%, from 476 Ω/sq to 364 Ω/sq. This improvement in R_{sh} is due to the layer transfer process and this effect was also reported in [13]. The change in strain in the channel is a potential cause for the improvement [18] although additional experiments are needed to fully confirm its effect.

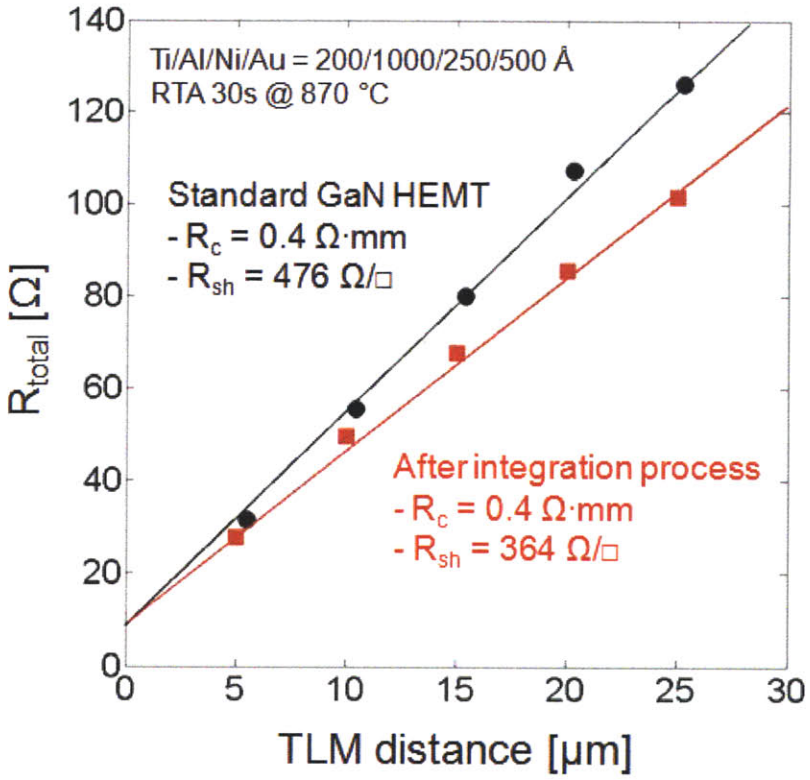


Figure 5-13. The entire integration process improved the sheet resistance (R_{sh}) of AlGaN/GaN HEMTs by 24%, from 476 Ω/sq to 364 Ω/sq while the contact resistance (R_c) was not affected (R_c = 0.4 Ω·mm). Possible change in the channel strain can be related to the improvement.

5.5. New applications

Numerous applications can be envisioned that take advantage of the integration technology demonstrated in this chapter. Some examples of circuits that could benefit from the integration of Nitride and Si electronics on the same wafer include high power digital-to-analog converters, high speed differential amplifiers, normally-off power transistors, and highly-compact power regulator circuits. Among these applications, advanced power distribution scheme in microprocessors and new enhancement-mode power transistors are explained in more detail below.

5.5.1. On-wafer GaN/Si power regulator

Power regulation is an important challenge in modern microprocessors due to the trade-off in power dissipation, operating voltage, and input bias current. To keep a constant power dissipation in microprocessors, the operating voltage has to be decreased as the switching frequency increases, which increases the input current to levels well above 100 A per chip as shown in Figure 5-14 [19]. This high input current not only increases conductive power losses but also uses a large number of the I/O pins available in traditional microprocessor packaging. One of the most promising solutions to this problem is to distribute the input power in the microprocessor at high voltages (and low currents) and then downconvert it to the required low voltages, locally, in highly integrated on-wafer power regulator circuits. The fabrication of an all-Si solution is very challenging due to the low breakdown voltage and frequency performance of Si power electronics. The hybrid integration of GaN and Si devices, on the other hand, would enable on-wafer voltage regulators with unprecedented performance and integration levels. Figure 5-15 shows the proposed architecture for advanced power distribution in the microprocessor and Figure 5-16(a) describes the circuit diagram of one of these new hybrid regulators currently being developed in our group, a tapped-inductor DC-DC converter topology where GaN switches are used in the high voltage / low current part of the circuit and Si power MOSFETs are used in the low voltage / high current regions of the power circuit. The use of a tapped-inductor topology also extends the available duty cycle to allow high voltage conversion ratios and helps the soft-switching of both Si and GaN transistors. Using this topology, our group has designed a hybrid tapped-inductor DC-DC converter with 12:1 conversion ratio and 10 W output power [20]. A very high switching frequency of 300 MHz was chosen to reduce the size of the passive components and to allow for on-chip integration of the entire converter. The circuit operates in the soft-switching mode with duty cycle of 50%. The on resistance and output capacitance in the SPICE model were obtained from measurements and data-sheets. In Figure 5-16(b), the simulation shows that the use of GaN switches in the high voltage part of the circuit instead of Si devices reduces the circuit losses about five-fold. This result demonstrates the great potential of a hybrid GaN/Si power electronic circuit to enable local power conversion in high performance Si electronics.

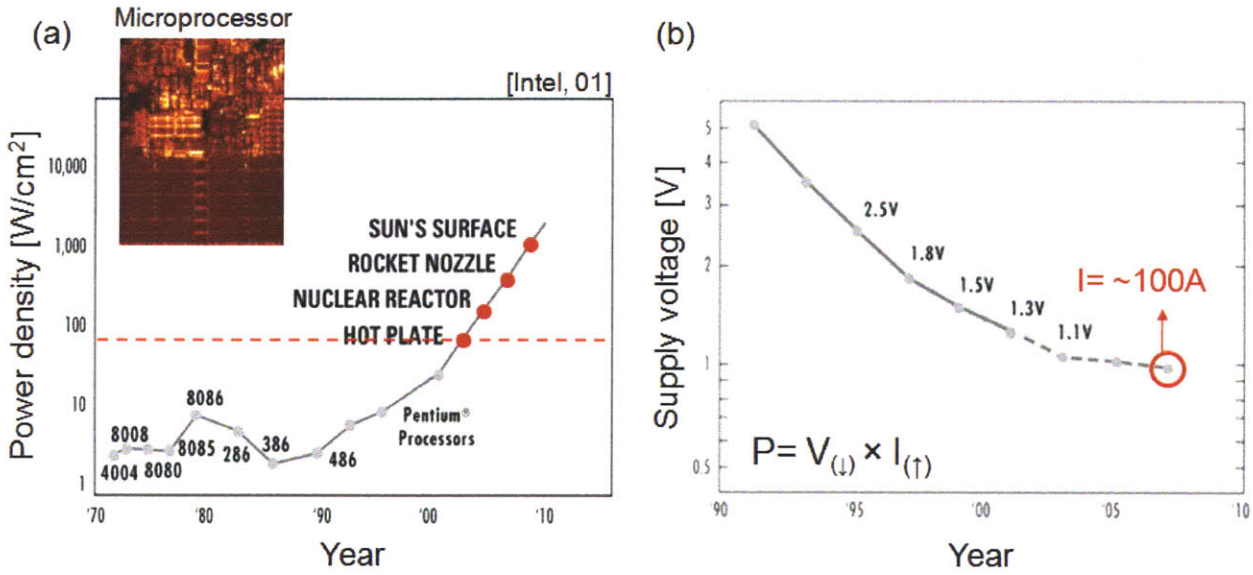


Figure 5-14. (a) Evolution of the power density dissipation in typical microprocessors. The power density has been increasing as the number of transistors increases. (b) To reduce the power dissipation in microprocessors, the supply voltage needs to be reduced. However, this significantly increases driving current which causes more conductive power losses and inefficient consumption of I/O pins.

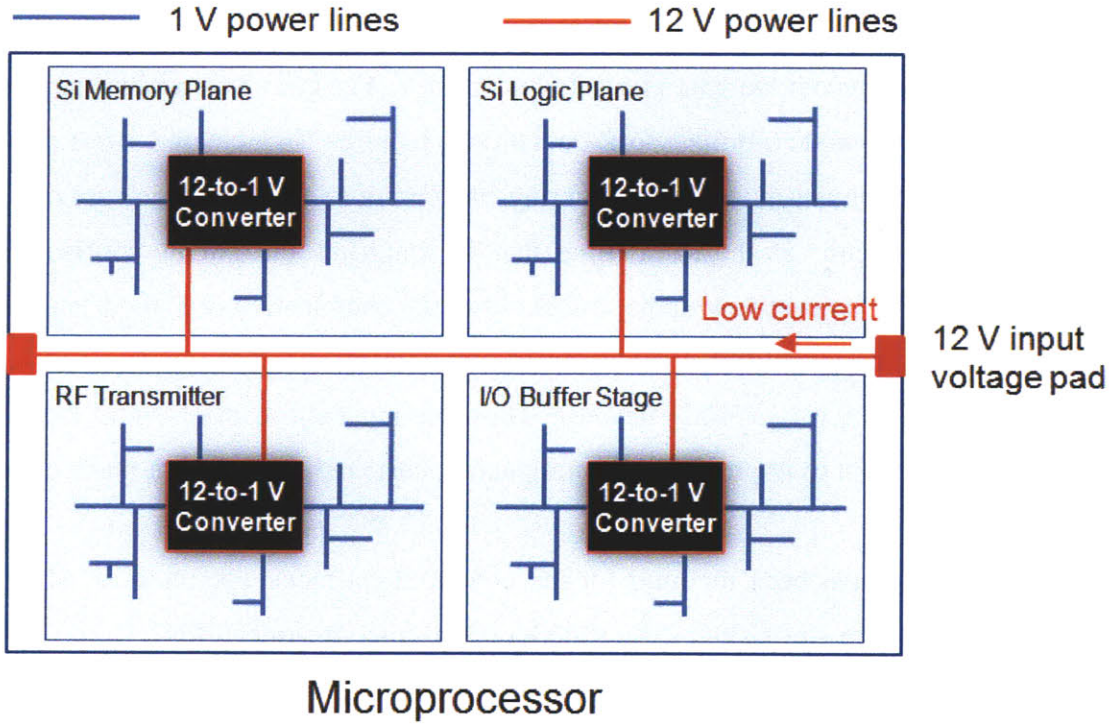
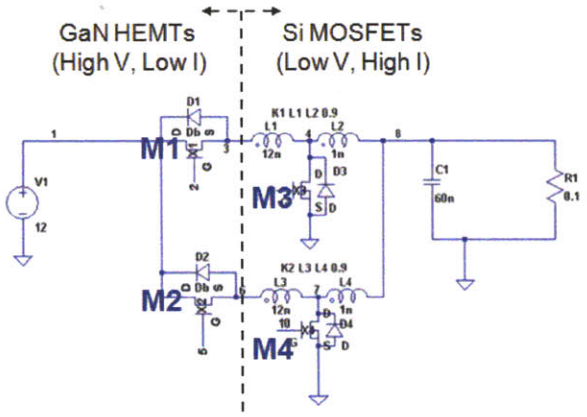


Figure 5-15. Proposed architecture for advanced power distribution in microprocessors. The power is delivered through 12 V power lines and locally downconverted to 1 V to power Si memory logic circuits. The 12-to-1 V conversion is realized by GaN/Si hybrid power electronic circuits.

(a)



(b)

	All Si circuit	GaN/Si hybrid circuit
Efficiency (@300 MHz)	9%	55 %
Losses in inductor	23%	73%
Losses in M1, M2	77%	11%
Losses in M3, M4	0.3%	16%
Maximum V/I in M1, M2	31 V / 7 A	56 V / 3 A
Maximum V/I in M3, M4	6.3 V / 14 A	6.3 V / 18 A

Figure 5-16. (a) Circuit schematic of the new GaN/Si hybrid power converter. M1 and M2 are GaN-based power transistors while M3 and M4 are Si MOSFETs. (b) Simulation of the current and voltage waveforms in the 12:1 V hybrid voltage regulator shows six-fold better efficiency at 300 MHz switching frequency [20].

5.5.2. Enhancement-mode power transistor

Most AlGaN/GaN HEMTs operate in depletion mode (normally-on) due to the 2DEG channel naturally induced underneath the gate at a gate bias of 0 V. For power electronics applications, however, enhancement-mode (normally-off) operation is required for increased safety and circuit simplicity. The enhancement-mode operation brings the circuit to a safe point in case of a failure in the gate driver circuit, at the same time that it simplifies the circuit configuration by eliminating the need of a negative voltage source. Also, the combination of E-mode and D-mode devices on the same chip enables the realization of complementary logic circuits which has been impeded by the lack of p-type GaN devices. Thus, having enhancement-mode AlGaN/GaN HEMTs is very important to improve safety, complexity, cost, and flexibility of many circuits.

Several techniques have been reported for the enhancement-mode operation of AlGaN/GaN HEMTs. Some of them are using a fluoride-based plasma treatment [21], a recessed gate structure [22], a thin AlGaN barrier [23], a p-type gate [24], and a non-polar a-plane channel [25]. Each technology has pros and cons, however in general they sacrifice some of the original performance of GaN HEMTs by modifying the GaN underneath the gate to achieve

enhancement-mode operation. In this section, we describe a new method to achieve enhancement-mode operation of GaN HEMTs by building a hybrid GaN-Si device.

Figure 5-17 shows the schematic of the proposed enhancement-mode power transistor which is a hybrid structure consisting of a low breakdown voltage, normally-off silicon n-MOSFET and a high breakdown voltage, normally-on AlGaN/GaN HEMT. The drain region of the low voltage Si n-MOSFET is connected in series with the source contact of the high voltage AlGaN/GaN HEMT. It is important to note that the gate contact of the AlGaN/GaN HEMT is connected to ground (or to the source contact of the Si n-MOSFET). This concept of cascade arrangement is called Baliga-Pair configuration [26] and it was initially proposed in SiC power switch devices. However, to the best of our knowledge, an integrated version of this configuration has not been demonstrated on the same wafer yet and its operation has been limited to discrete Si n-MOSFET and SiC MESFET components. The integration technology discussed in this chapter can be directly applied to realize Baliga-Pair transistors for enhancement-mode operation of GaN power transistor. Its operating principle is explained in more detail below.

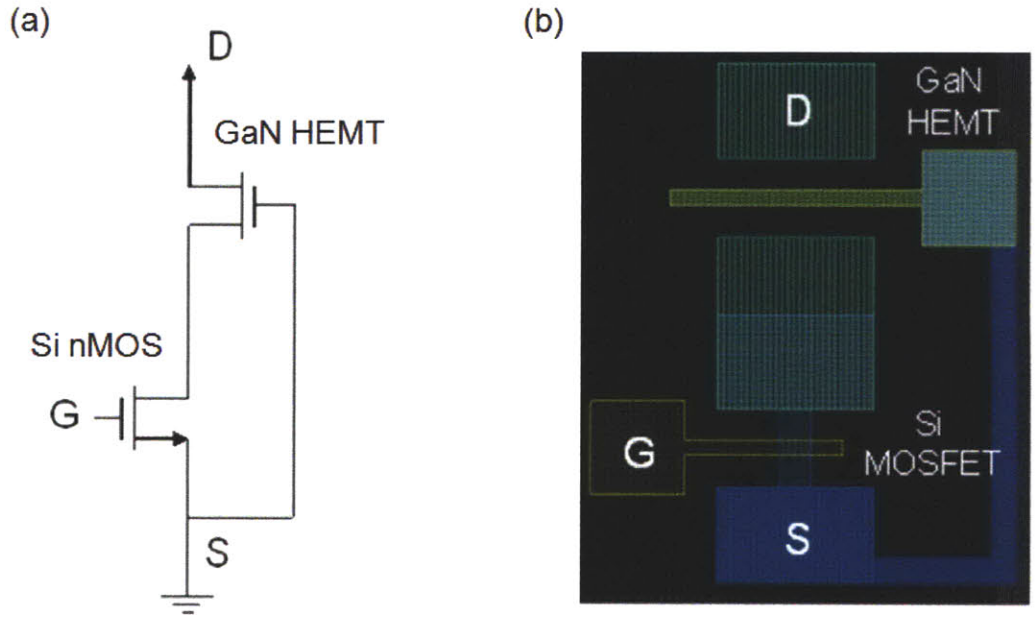


Figure 5-17. (a) Schematic and (b) layout of the proposed enhancement-mode power transistor scheme consisting of silicon MOSFET and GaN HEMT.

When the gate is grounded (V_{GS} = 0 V) and an increasing positive bias is applied to the drain (V_{DS} > 0 V), the hybrid device is in the OFF-state and the applied V_{DS} is initially supported by the Si n-MOSFET until the 2DEG channel of GaN HEMT becomes depleted. Once that the drain voltage of the Si n-MOSFET (or equivalently the source voltage of GaN HEMT) exceeds the voltage required to deplete the 2DEG channel by reverse biasing the gate-source junction of the GaN HEMT, any further increase in V_{DS} becomes sustained by the extension of the depletion region in GaN HEMT. The potential at the drain region of the Si n-MOSFET remains relatively constant and the breakdown characteristics is mainly determined by GaN HEMT. When a positive voltage is applied to the gate (V_{GS} > 0 V) and an increasing positive bias is applied to the drain (V_{DS} > 0 V), the device is the ON-state and the current can flow through the entire channel of the Si n-MOSFET and GaN HEMT. Therefore, a normally-off Si n-MOSFET can be used to control a high voltage normally-on GaN HEMT. This enables supporting large voltages within the GaN HEMT while allowing enhancement-mode operation by the Si n-MOSFET. This configuration does not require any additional process but interconnecting two devices for enhancement-mode operation, hence it does not degrade the intrinsic transport properties of the original devices.

5.6. References

[1] U. K. Mishra, L. Shen, T. E. Kazior, and Y.-F. Wu, "GaN-based RF power devices and amplifiers," *Proceedings of the IEEE*, vol. 96, no. 2, pp. 287-305, Feb. 2008.

[2] S. Nakamura, G. Fasol, and S. J. Pearton, "The blue laser diode: The complete story," Springer, 2000.

[3] R. Fischer, T. Henderson, J. Klem, W. Kopp, C. K. Peng, and H. Morkoc, "Monolithic integration of GaAs/AlGaAs modulation-doped field-effect transistors and N-metal-oxide-semiconductor silicon circuits," *Appl. Phys. Lett.*, vol. 47, pp. 983-985, Nov. 1985.

[4] R. N. Ghosh, B. Griffing, and J. M. Ballantyne, "Monolithic integration of GaAs light-emitting diodes and Si metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 48, pp. 370-371, Feb. 1986.

[5] H. K. Choi, G. W. Turner, and B-Y. Tsaur, "Monolithic integration of Si MOSFET's and GaAs MESFET's," *IEEE Electron Device Lett.*, vol. 7, no. 4, pp. 241-243, Apr. 1986.

[6] S. Joblot, F. Semond, Y. Cordier, P. Lorenzini, and J. Massies, "High-electron-mobility AlGaN/GaN heterostructures grown on Si(001) by molecular-beam epitaxy," ," *Appl. Phys. Lett.*, vol. 87, 133505, Sept. 2005.

[7] Y. Cordier, J. -C. Moreno, N. Baron, E. Frayssinet, S. Chenot, B. Damilano, and F. Semond, "Demonstration of AlGaN/GaN high-electron-mobility transistors grown by molecular beam epitaxy on Si(110)," *IEEE Electron Device Lett.*, vol. 29, no. 11, pp. 1187-1189, Nov. 2008.

[8] F. Schulze, O. Kisel, A. Dadgar, A. Krtschil, J. Blasing, M. Kunze, I. Daumiller, T. Hempel, A. Diez, R. Clos, J. Christen, and A. Krost, "Crystallographic and electric properties of MOVPE-grown AlGaN/GaN-based FETs on Si(001) substrates," *J. Crystal Growth*, vol. 299, pp. 399-403, Feb. 2007.

[9] A. Dadgar, F. Schulze, M. Wienecke, A. Gadanecz, J. Blasing, P. Veit, T. Hempel, A. Diez, J. Christen, and A. Krost, "Epitaxy of GaN on silicon- impact of symmetry and surface reconstruction," *New J. Phys.*, vol. 9, Oct. 2007.

[10] H. Ji, J. Das, M. Germain, and M. Kuball, "Laser lift-off transfer of AlGaN/GaN HEMTs from sapphire onto Si: A thermal perspective," *Solid-State Electronics*, vol. 53, pp. 526-529, Apr. 2009.

[11] W. S. Wong, Y. Cho, E. R. Weber, T. Sands, K. M. Yu, J. Kruger, A. B. Wengrow, and N. W. Cheung, "Structural and optical quality of GaN/metal/Si heterostructures fabricated by excimer laser lift-off," *Appl. Phys. Lett.*, vol. 75, no. 13, pp. 1887-1889, Sep. 1999.

[12] M. Funato, S. Fujita, and S. Fujita, "Integration of GaN with Si using a AuGe-mediated wafer bonding technique," *Appl. Phys. Lett.*, vol. 77, no. 24, pp. 3959-3961, Dec. 2000.

[13] J. W. Chung, E. L. Piner, and T. Palacios, "N-face GaN/AlGaN HEMTs fabricated through layer transfer technology," *IEEE Electron Device Lett.*, vol. 30, no. 2, pp. 113-116, Feb. 2009.

[14] M. Kuball, F. Demangeot, J. Frandon, M. A. Renucci, J. Massies, N. Grandjean, R. L. Aulombard, and O. Briot, "Thermal stability of GaN investigated by Raman scattering," *Appl. Phys. Lett.*, vol. 73, no. 7, pp. 960-962, Aug. 1998.

[15] M. A. Schmidt, "Wafer-to-wafer bonding for microstructure formation," *Proceedings of the IEEE*, vol. 86, pp. 1575-1585, Aug. 1998.

[16] J. Esch, "Wafer direct bonding: from advance substrate engineering to future applications in micro/nanoelectronics," *Proceedings of the IEEE*, vol. 94, no. 12, pp. 2058-2059, Dec. 2006.

[17] K. Ryu, J. W. Chung, B. Lu, and T. Palacios, "Wafer Bonding Technology in Nitride Semiconductors for Applications in Energy and Communications," *218th Electrochemical Society Meeting*, Las Vegas, Oct. 10-15, 2010.

[18] M. Azize and T. Palacios, "Effect of substrate-induced strain in the transport properties of AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 108, no. 2, 023707, Jul. 2010.

[19] V. De and S. Borkar, "Technology and design challenges for low power and high performance," *Proc. International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 163-168, 1999.

[20] B. Lu, D. Perrault, and T. Palacios, "GaN-based power distribution plane in Si electronics," *Proc. Interconnect Focus Center Review*, pp. 182, Oct. 2008.

[21] Y. Cai, Y. Zhou, K. M. Lau, and K. J. Chen, "Control of threshold voltage of AlGaN/GaN HEMTs by fluoride-based plasma treatment: From depletion mode to enhancement mode," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2207-2215, Sep. 2006.

[22] W. Saito, Y. Takada, M. Kuraguchi, K. Tsuda, and I. Omura, "Recessed-gate structure approach toward normally off high-voltage AlGaN/GaN HEMT for power electronics applications," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 356-362, Feb. 2006.

[23] Y. Ohmaki, M. Tanimoto, S. Akamatsu, and T. Mukai, "Enhancement-mode AlGaN/AlN/GaN high electron mobility transistor with low on-state resistance and high breakdown voltage," *Jpn. J. Appl. Phys.*, vol. 45, no. 44, pp. L1168-L1170, Nov. 2006.

[24] N. Tsuyukuchi, K. Nagamatsu, Y. Hirose, M. Iwaya, S. Kamiyama, H. Amano, and I. Akasaki, "Low-leakage-current enhancement-mode AlGaN/GaN heterostructure field-effect transistor using p-type gate contact," *Jpn. J. Appl. Phys.*, vol. 45, no. 11, pp. L319-L321, Mar. 2006.

[25] M. Kuroda, T. Ueda, and T. Tanaka, "Normally-off AlGaN/GaN MIS-HFETs using non-polar a-plane," in *Proc. Ext. Abstr. Solid State Devices Mater.*, pp. 148-149, 2007.

[26] B. J. Baliga, "Fundamentals of power semiconductor devices," Springer, 2008.

Chapter 6. Conclusions and Future Work

6.1. Summary

In this thesis, we have investigated two important topics in GaN HEMTs. First, we have studied high frequency characteristics of AlGaN/GaN HEMTs. In particular, we focused on f_T and f_{max}, two of the most important figures of merit in high frequency performance and investigated them analytically and experimentally. Based on improved physical understanding and advanced process technologies, we have demonstrated state-of-the-art f_T of 225 GHz and f_{max} of 300 GHz in AlGaN/GaN HEMTs. RF g_m-collapse and short-channel effects were effectively suppressed by a new gate technology to improve f_T, while the parasitic components of the device were thoroughly optimized to maximize f_{max}. Second, we have developed a new technology to realize on-wafer integration of GaN and (100)-oriented Si electronics. The novel layer transfer technology allows both Ga- and N-face GaN devices to be integrated with Si(100) devices on the same wafer through a fully Si-compatible process. This integration enables the development of numerous hybrid circuits that take advantage of the high-frequency and power capability of GaN and the unsurpassed circuit scalability and complexity of Si electronics. Some examples of these circuits include on-wafer GaN/Si power regulators, high linearity power amplifiers, and enhancement mode power transistors.

6.1.1. Reduction of parasitic elements for improving f_{max}

One of the key challenges to produce a high-gain RF power amplifier using GaN HEMTs is to increase their maximum power-gain cutoff frequency (f_{max}). Despite the fact that f_{max} is largely affected by several parasitic elements in the device such as R_i, R_s, R_g, C_{gd}, and g_o, systematic study of these parasitic elements has been seldom a path to increase f_{max}. Conventional wisdom for improving f_{max} was to increase f_T and most of the research and development efforts were focused on f_T rather than on f_{max}. Contrary to the conventional approach, in this work, we tried to maximize f_{max} of AlGaN/GaN HEMTs by minimizing the detrimental effects from the parasitic

elements. The combination of 1 μm source-to-drain distance and optimized Si/Ge-based recessed ohmic technology produced a very low source resistance R_s of 0.4 Ω·mm. A carefully optimized low-damage gate recess technology was used to minimize the input resistance R_i and output conductance g_o. Also, a reliable deep-submicron T-gate (L_g = 60 nm, H_g = 500 ~ 600 nm, T_g = 200 nm) was fabricated to minimize both gate resistance R_g and gate-drain capacitance C_{gd}. Finally, all these technologies were harmoniously integrated to simultaneously enable minimum R_i, R_s, R_g, C_{gd}, and g_o in the high frequency operation of the device.

As a result, we have obtained a state-of-the-art f_{max} of 300 GHz with an L_g = 60 nm gate-recessed AlGaN/GaN HEMT. Owing to the low-damage gate recess technology, scaled device geometry, and recessed source/drain ohmic contacts, the short-channel effects (i.e. high output conductance g_o) as well as parasitic resistances were effectively suppressed. A device with a longer gate length (L_g = 160 nm) also showed an excellent f_{max} of 260 GHz (f_T = 50 GHz) surpassing previous record f_{max} of 251 GHz (f_T = 190 GHz) achieved by a device with a much shorter gate length (L_g = 60 nm). This highlights the great importance of optimizing parasitic elements to maximize f_{max}. The accuracy of the reported f_{max} values was verified by small-signal modeling based on carefully extracted S-parameters. The fabrication technology and the record f_{max} described in this work demonstrate the unsurpassed potential of GaN transistors for mm- and sub-mm wave power amplifiers.

6.1.2. Advanced gate technologies for improving f_T

Another important challenge in AlGaN/GaN HEMTs is to understand the origin of the lower-than-expected current-gain cutoff frequency (f_T) frequently observed in this material system throughout the past years. Understanding f_T in GaN HEMTs is important not only to increase f_{max} by improving f_T, but also to benchmark GaN HEMTs with transistors in other material systems, since f_T is closely related to the intrinsic material properties. Although several hypotheses have been proposed to explain the lower-than-expected f_T in GaN HEMTs, it is still not clear what is primarily limiting high frequency performance of GaN HEMTs and more understanding is needed to identify solutions to this problem.

In order to find out the origin of the lower-than-expected f_T in GaN HEMTs, intrinsic small-signal parameters (C_{gs}, C_{gd}, C_{ds}, R_i, R_{gd}, g_m, τ, and g_o) were carefully extracted and evaluated with respect to their expected frequency behaviors. An improved extraction method for intrinsic small-signal parameters was developed by taking into account the effect of the large gate leakage current in AlGaN/GaN HEMTs, on-wafer open/short de-embedding to completely de-embed extrinsic pad components, and gate current injection technique to estimate bias dependent source/drain resistances. This method is simple yet accurate, showing an excellent agreement between the measured and simulated S-parameters (from the extracted intrinsic parameters) over a large frequency range.

After carefully investigating the extracted intrinsic small-signal parameters, we identified a severe degradation of g_m at high frequencies, while the other intrinsic parameters were maintained at reasonable values. This phenomenon, called RF g_m-collapse, is believed to be the cause of the lower-than-expected f_T in AlGaN/GaN HEMTs. From an experimental point of view, the RF g_m-collapse is observed as a decreasing S_{21} with increasing the input signal frequency in the 10 ~ 100 MHz range. From a physical point of view, on the other hand, we postulate that RF g_m-collapse is caused by nitrogen vacancies (V_N) or other defects with shallow energy levels at the AlGaN surface. A possible mechanism could be the lower modulation efficiency at high frequencies due to the *defect*-related donor-like states (e.g. V_N) underneath the gate, however further theoretical and experimental studies are necessary to clarify the mechanism of RF g_m-collapse.

Since RF g_m-collapse is believed to be related to the poor surface condition between gate contact and AlGaN surface, we studied the effect of several plasma treatments applied to the gate region to improve the surface chemistry. Both O_2 plasma and CF_4 plasma treatments were effective in suppressing the RF g_m-collapse. XPS measurements of the AlGaN surface immediately after these plasma treatments revealed that O and F species from the plasma exchange their sites with N. During this process, it is also possible that V_N can be occupied by O or F, therefore reducing V_N-related defects at the AlGaN surface. Although both O_2 plasma and CF_4 plasma treatments are effective in terms of preventing RF g_m-collapse, O_2 plasma treatment was chosen in this work to improve f_T in AlGaN/GaN HEMTs since it is more controllable and

reproducible over CF_4 plasma treatment. This treatment effectively reduces gate leakage current, increases peak transconductance, and eliminates RF g_m-collapse.

Based on the developed O_2 plasma treatment, an advanced gate technology was designed to maximize f_T in AlGaN/GaN HEMTs. The new gate technology features vertical gate-recess to minimize short-channel effects, O_2 plasma treatment to suppress RF g_m-collapse, and selective lateral gate-etch to further shrink the gate length. The resulting AlGaN/GaN HEMT showed a state-of-the-art f_T of 225 GHz for an L_g = 55 nm and t_b = 17 nm. A device with a longer gate length L_g = 110 nm also showed an excellent f_T of 162 GHz which produces a very high $f_T \times L_g$ product of 17.8 GHz·μm, one of the highest reported values at these gate lengths. The accuracy of the obtained f_T values was verified by small-signal modeling based on carefully extracted S-parameters.

Delay analysis was carried out to identify additional room for the improvement of f_T and it is found that even for an L_g = 55 nm device, which demonstrated a record f_T of 225 GHz, 60 % of the total delay was associated to intrinsic delay. Thus, the use of O_2 plasma treatment in combination with more aggressive scaling of both L_g and t_b will further increase f_T in AlGaN/GaN HEMTs. From these results, transistors with f_T in excess of 300 GHz are possible even without reducing the parasitic delays. Much higher performance with f_T in excess of 400 GHz can, of course, be obtained by simultaneously reducing device dimensions and parasitic components.

6.1.3. On-wafer integration of GaN and Si(100) electronics

Although GaN devices offer unsurpassed potential for high frequency and high power electronics, they cannot compete with Si technology in terms of flexibility and circuit complexity. On the other hand, Si microelectronics, traditionally driven by scaling and Moore's law, is approaching a saturation in the available performance due to power dissipation and device dimensions. We believe that the heterogeneous integration of GaN and Si electronics would significantly help to develop a new generation of electronic systems where the best semiconductor is seamlessly used in its optimum application.

GaN-based devices are one of the best candidates for integration with Si electronics. In addition to their excellent performance, the high thermal stability of GaN makes it possible to satisfy the high thermal budget of a standard Si processing. In this work, we have developed a new technology to integrate GaN and (100)-oriented Si (no miscut) electronics on the same wafer. This technology is based on a novel layer transfer process and it allows the heterogeneous integration of GaN and Si devices without having to modify the standard processing technologies of each of these material systems.

The first on-wafer integration of Ga-face AlGaN/GaN HEMTs and Si(100) MOSFETs was demonstrated on a Si(100)-GaN-Si(100) hybrid substrate. The high thermal stability of GaN allowed the fabrication of Si MOSFETs on this substrate without degrading the performance of the GaN epilayers. After the Si devices were fabricated, the nitride epilayer was exposed and the nitride transistors are processed. Using this technology, GaN and Si devices separated by less than 5 μm from each other has been fabricated, which is suitable for building future heterogeneous integrated circuits.

The device-level integration of nitride semiconductors (transistors, LEDs, lasers, etc) and Si-based transistors on the same wafer enables revolutionary advances in circuits and systems. Some of the applications envisioned for this integration include high power digital-to-analog converters, high speed differential amplifiers, normally-off power transistors, and highly-compact power regulator circuits. Among these applications, advanced power distribution scheme in microprocessors and new enhancement-mode power transistors were described in more detail.

6.2. Future work

Although we have extensively studied the high frequency characteristics of AlGaN/GaN HEMTs and developed several process technologies to demonstrate state-of-the-art f_T and f_{max} in these devices, there is still plenty of room for further improvement of their frequency

performance. In addition, the RF g_m-collapse needs to be more clearly understood. Finally, the on-wafer integration technology demonstrated in this work needs to be improved to allow larger wafer diameters and higher yields. The following sections describe our proposed work on these topics.

6.2.1. Improvement of f_T and f_{max}

As discussed in Chapter 4, delay analysis is a very powerful tool to identify which small-signal parameter needs to be engineered to improve f_T in AlGaN/GaN HEMTs. To increase f_T, the total delay (τ_{total}) needs to be minimized, which requires a careful study of the different components of this delay. τ_{total} can be divided into three different components: intrinsic delay (τ_{int}), extrinsic delay (τ_{ext}), and parasitic delay (τ_{par}).

$$f_T = \frac{1}{2\pi\tau_{total}} \tag{4-25}$$

$$\tau_{total} = \tau_{int} + \tau_{ext} + \tau_{par} \tag{4-26}$$

Each delay can be analytically expressed as a function of the small-signal parameters in the active region of the device, C_{gs}, C_{gd}, g_m, g_o, R_s, and R_d.

$$\tau_{int} = \frac{C_{gs,i} + C_{gd,i}}{g_m} = \frac{L_g}{v_e} \tag{4-27}$$

$$\tau_{ext} = \frac{C_{gs,ex} + C_{gd,ex}}{g_m} \tag{4-28}$$

$$\tau_{par} = C_{gd}(R_s + R_d)\left[1 + (1 + \frac{C_{gs}}{C_{gd}})\frac{g_o}{g_m}\right] \tag{4-29}$$

where $C_{gs,i}$ and $C_{gd,i}$ are the internal gate capacitances excluding external gate fringing capacitances $C_{gs,ex}$ and $C_{gd,ex}$ from the total intrinsic gate capacitance C_{gs} and C_{gd}, respectively ($C_{gs} = C_{gs,i} + C_{gs,ex}$, $C_{gd} = C_{gd,i} + C_{gd,ex}$).

The following paragraphs provide a guideline to improve f_T. Most of the proposed techniques have already been tried throughout the thesis, however it is suggested to apply them more aggressively to better quality materials to further improve f_T in AlGaN/GaN HEMTs.

First of all, the RF g_m-collapse needs to be eliminated and RF g_m should be maintained at high value since g_m is in the denominator of the expressions for all three delay components. RF g_m-collapse can be suppressed by applying proper plasma treatment to the AlGaN surface such as O_2 plasma or CF_4 plasma treatment as shown in this work. RF g_m-collapse also could be removed by improving the quality of material in the growth level.

Secondly, intrinsic delay τ_{int} has to be reduced by scaling L_g without degrading g_m. L_g can be reduced by optimizing the e-beam gate lithography and applying more aggressively the lateral Ni-etch developed in this work. To reduce the short-channel effects, it is important to increase the gate-to-channel aspect ratio for which low-damage vertical gate recess technique developed in this work is very useful. The use of a thin barrier (t_b) structure with reasonable channel transport properties as a starting material could also help to minimize the short-channel effects. A thin barrier InAlN/GaN HEMTs can be a promising solution since a high sheet charge density which produces current density over 2 A/mm can be achieved in this device even with a barrier thickness below 10 nm [1]. For example, Sun *et al.* reported a 55-nm gate length InAlN/GaN HEMT with an excellent f_T and f_{max} of 205 GHz and 191 GHz, respectively [2]. The minimum barrier thickness is limited by the gate leakage current, however the gate leakage can also be reduced by the plasma treatment applied to suppress RF g_m-collapse. Our group recently demonstrated a new record f_T of 245 GHz in a 50-nm InAlN/GaN HEMT using O_2 plasma treatment presented in this work [3]. Another way to reduce τ_{int} is to increase v_e, by engineering the mobility μ and the electric field E. μ can be increased by introducing appropriate gate dielectric such as Ga_2O_3 as shown in this work or by engineering the epitaxial structure, for example by inserting a thin AlN layer between AlGaN and GaN to reduce the scattering mechanisms. There exists an optimum E which provides the peak v_e and engineering E in the channel can be done by changing the device structure such as the shape of the gate contact. It is generally difficult to optimize E in the channel, however it could have a tremendous impact in the device performance since there is a still large discrepancy between the measured and expected peak v_e (1.5×10^7 cm/s vs. 3.0×10^7 cm/s).

Thirdly, the extrinsic delay τ_{ext} needs to be reduced by minimizing external gate fringing capacitances. The gate fringing capacitances are mainly caused by the head part of the T-shape

gate and sensitive to its distance from the channel. Therefore, the stem height of the T-gate needs to be increased as much as possible to make this effect negligible.

Finally, parasitic delay τ_{par} should be minimized by reducing R_s, R_d, and g_o. To lower R_s and R_d, in this work we developed a Si/Ge-based recessed ohmic technology, however there are other techniques which could be more effective such as n^+ GaN cap layer [4], selectively regrown ohmic contacts [5], Si-implantation to the source and drain regions [6], etc. It is important to choose appropriate technology to minimize R_s and R_d without adversely affecting other material properties. As discussed in Chapter 3, g_o is closely related to the leakage current between source and drain contacts. Thus, g_o can be suppressed by either engineering the device structure incorporating p-type buffer layer [7] or InGaN back barrier [8] to reduce buffer leakage current or improving short-channel effects applying low-damage gate recess technique to minimize channel leakage current.

As discussed in Chapter 3, the analytical formula of f_{max} can be thought as consisting of two parts: f_T and parasitic components (R_i, R_s, R_g, C_{gd}, and g_o).

$$f_{max} \cong \frac{f_T}{2\sqrt{(R_i + R_s + R_g)g_o + (2\pi f_T)R_g C_{gd}}} \tag{4-30}$$

In this work, we demonstrated the great impact of the parasitic components on the f_{max} in AlGaN/GaN HEMTs. A tight source-to-drain distance ($R_s\downarrow$), a carefully optimized low-damage gate recess technology ($R_i\downarrow$, $g_o\downarrow$), and a reliable T-shaped gate with small gate length and large head size ($C_{gd}\downarrow$, $R_g\downarrow$) have been developed to reduce them. These technologies are compatible with the requirements to improve f_T described in the section 6.2.1. Therefore, by combining all the efforts to improve f_T and technologies developed to optimize parasitic components in this work, we can anticipate to obtain very high f_T and f_{max} at the same time.

6.2.2. Investigation of RF g_m-collapse

In Chapter 4, the lower-than-expected frequency performance observed in many AlGaN/GaN HEMTs was attributed to a significant drop of the intrinsic small-signal transconductance (RF g_m)

at high frequency with respect to the intrinsic DC g_m. This phenomenon was called RF g_m-collapse and we showed that an O_2 plasma treatment applied to the gate region can largely improve f_T of AlGaN/GaN HEMTs by effectively suppressed RF g_m-collapse. Although we described our current understanding on the possible mechanism of RF g_m-collapse and the role of the O_2 plasma treatment in Chapter 4, the cause of the RF g_m-collapse and how O_2 plasma treatment solves it are still not clear. Below we suggest additional experiments which will be useful to clarify RF g_m-collapse.

It is first important to identify whether RF g_m-collapse is a new type of phenomenon which was not previously studied or it is similar to the transconductance dispersion observed in GaAs MESFETs [9]. Although it needs to be more rigorously verified, we believe this is a new phenomenon since the transition frequency of RF g_m-collapse (~ 10^7 Hz) is much higher than that of transconductance dispersion (~ 10^3 Hz). Also, source/drain access regions are not likely responsible for the RF g_m-collapse since RF g_m-collapse was also observed in fully passivated samples without evident current collapse. However, few passivated samples have been studied in this work and better statistics from more samples are required. It is also suggested to investigate the amount of RF g_m-collapse on self-aligned AlGaN/GaN HEMTs (no access regions) or as a function of length of access regions (L_{gd}, L_{gs}). If the access regions are related to transconductance dispersion, self-aligned device should not show RF g_m-collapse.

To identify the origin of RF-g_m collapse, it is also useful to extract the activation energy of the potential trap states. The activation energy will allow us to identify the origin and specific location of the traps (e.g. V_N, V_{Ga}, AlGaN-related, etc). The activation energy can be estimated from the temperature dependent transition frequency of g_m [10]. Unfortunately, in this work we could not obtain a precise activation energy due to limitations in our network analyzer, which has a minimum input frequency of 10 MHz. We did, however, confirm a positive shift in the transition frequency as the temperature was increased, which indicates a contribution from the traps. Therefore, it is encouraged to try extracting the activation energy from the g_m vs. f curve by changing the temperature and using a network analyzer that can go below 10 MHz.

Finally, the development of a physics-based analytical modeling of the RF g_m-collapse would complete the verification of the potential mechanisms. In Chapter 4, we modeled the frequency behavior of the measured S-parameters by introducing in the small-signal equivalent circuit an additional voltage controlled current source (VCCS) with negative g_m' to account for RF g_m-collapse. This model predicts the measured S-parameters and frequency characteristics of AlGaN/GaN HEMTs very well, however it lacks physical insight.

6.2.3. On-wafer integration of GaN and Si(100)

Although the first on-wafer integration of GaN and Si(100) devices was demonstrated in this work, several improvements are required for this technology to become a viable solution for successful integration of GaN and Si(100) electronics.

At present, the new technology has been used to fabricate Si(100)-GaN-Si(100) hybrid substrate that is about one square inch in size. Conventional Si manufacturing processes typically use larger wafers, 8 or 12 inches in diameter, so the research now has to be focused on scaling up the process. Particularly, it is important to have a reliable and reproducible wafer bonding technology even at such a scaled process. When larger wafers are used for the layer transfer technology (wafer bonding and etch-back process), there are more processing parameters to consider such as wafer bow, surface particles, air gap formation, applied pressure, etc. It is suggested that wafers with a less than 10 μm of wafer bow are bonded using high purity HSQ interlayer in a vacuum condition with enough pressure to make a perfect contact. Recently, Kevin Ryu and Hyung-Seok Lee in our group at MIT have succeeded in developing a 4 inch wafer process.

Once a reliable platform for the heterogeneous integration is established (e.g. Si(100)-GaN-Si(100) hybrid substrate), the fabrication technology of GaN and Si devices needs to be optimized to make sure there is no degradation in device performance. In principle, the integration process does not adversely affect the transport properties of each material and no modification of the process is required. In the proof-of-concept demonstration in Chapter 5, only unoptimized long gate length devices were characterized and it is expected that the use of a

shorter gate length and higher doping levels in the Si active layer will improve the device characteristics of both GaN HEMTs and Si MOSFETs.

6.3. References

[1] H. Wang, J. W. Chung, X. Gao, S. Guo, and T. Palacios, "Al_2O_3 passivated InAlN/GaN HEMTs on SiC substrate with record current density and transconductance," *Physica Status Solidi (c)*, vol. 7, no. 10, pp. 2440-2444, Oct. 2010.

[2] H. Sun, A. R. Alt, H. Benedickter, E. Feltin, J.-F. Carlin, M. Gonschorek, N. R. Grandjean, and C. R. Bolognesi, "205-GHz (Al,In)N/GaN HEMTs," *IEEE Electron Device Lett.*, vol. 31, no. 9, pp. 957-959, Sept. 2010.

[3] D. S. Lee, H. Wang, X. Gao, S. Guo, P. Ray, and T. Palacios, "245 GHz InAlN/GaN HEMTs with oxygen plasma treatment," accepted to *IEEE Electron Device Lett.*, 2011.

[4] Y. Pei, L. Shen, T. Palacios, N. A. Fichtenbaum, L. S. McCarthy, S. Keller, S. P. DenBaars, and U. K. Mishra, "Study of the n^+ GaN cap in AlGaN/GaN high electron mobility transistors with reduced source-drain resistance," *Jpn. J. Appl. Phys.*, vol. 46, no. 35, pp. L842-L844, Sept. 2007.

[5] C.-H. Chen, S. Keller, G. Parish, R. Vetury, P. Kozodoy, E. L. Hu, S. P. Denbaars, and U. K. Mishra, "High-transconductance self-aligned AlGaN/GaN modulation-doped field-effect transistors with regrown ohmic contacts," *Appl. Phys. Lett.*, vol. 73, no. 21, pp. 3147-3149, Nov. 1998.

[6] Y. Pei, F. Recht, N. Fichtenbaum, S. Keller, S. P. DenBaars, and U. K. Mishra, "Deep submicron AlGaN/GaN HEMTs with ion implanted source/drain regions and non-alloyed ohmic contacts." *IEEE Electron. Lett.*, vol. 43, no. 25, pp. 1466-1467, Dec. 2007.

[7] D. F. Storm, D. S. Katzer, S. C. Binari, E. R. Glaser, B. V. Shanabrook, and J. A. Roussos, "Reduction of buffer layer conduction near plasma-assisted molecular-beam epitaxy grown GaN/AlN interfaces by beryllium doping," *Appl. Phys. Lett.*, vol. 81, no. 20, pp. 3819-3821, Nov. 2002.

[8] T. Palacios, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "AlGaN/GaN high electron mobility transistors with InGaN back-barriers," *IEEE Electron Dev. Lett.*, vol. 27, no. 1, pp. 13-15, Jan. 2006.

[9] P. H. Ladbrooke and S. R. Blight, "Low-field low-frequency dispersion of transconductance in GaAs MESFET's with implications for other rate-dependent anomalies," *IEEE Trans. Electron. Devices*, vol. 35, no. 3, pp. 257-267, Mar. 1988.

[10] G. Meneghesso, A. Paccagnella, Y. Haddab, C. Canali, and E. Zanoni, "Evidence of interface trap creation by hot-electrons in AlGaAs/GaAs high electron mobility transistors," *Appl. Phys. Lett.*, vol. 69, no. 10, pp. 1411-1413, Sept. 1996.

Appendix 1. Scattering Parameter Conversion Rules

The relationships between S-parameter and Z-, Y-, and H-parameters are listed below. The Z′-, Y′-, and H′-parameters are normalized values with respect to a characteristic impedance Z_0 from the actual Z-, Y-, and H-parameters. Most microwave systems have a characteristic impedance of $Z_0 = 50\ \Omega$, and therefore the conversion rules must take Z_0 into account. By normalizing the scattering parameters, impedance and admittance can be plotted on the same Smith Chart which makes it easy to interpret the system.

- Normalization

$$Z' = \begin{bmatrix} z_{11}' & z_{12}' \\ z_{21}' & z_{22}' \end{bmatrix} = \frac{1}{Z_0}\begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} = \frac{Z}{Z_0}$$

$$Y' = \begin{bmatrix} y_{11}' & y_{12}' \\ y_{21}' & y_{22}' \end{bmatrix} = Z_0\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = Z_0 Y$$

$$H' = \begin{bmatrix} h_{11}' & h_{12}' \\ h_{21}' & h_{22}' \end{bmatrix} = \begin{bmatrix} \frac{h_{11}}{Z_0} & h_{12} \\ h_{21} & Z_0 h_{22} \end{bmatrix}$$

- $S \leftrightarrow Z$

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} \frac{(z_{11}'-1)(z_{22}'+1)-z_{12}'z_{21}'}{(z_{11}'+1)(z_{22}'+1)-z_{12}'z_{21}'} & \frac{2z_{12}'}{(z_{11}'+1)(z_{22}'+1)-z_{12}'z_{21}'} \\ \frac{2z_{21}'}{(z_{11}'+1)(z_{22}'+1)-z_{12}'z_{21}'} & \frac{(z_{11}'+1)(z_{22}'-1)-z_{12}'z_{21}'}{(z_{11}'+1)(z_{22}'+1)-z_{12}'z_{21}'} \end{bmatrix}$$

$$Z' = \begin{bmatrix} z_{11}' & z_{12}' \\ z_{21}' & z_{22}' \end{bmatrix} = \begin{bmatrix} \frac{(1+s_{11})(1-s_{22})+s_{12}s_{21}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} & \frac{2s_{12}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} \\ \frac{2s_{21}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} & \frac{(1-s_{11})(1+s_{22})+s_{12}s_{21}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} \end{bmatrix}$$

- $S \leftrightarrow Y$

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} \frac{(1-y_{11}')(1+y_{22}')+y_{12}'y_{21}'}{(1+y_{11}')(1+y_{22}')-y_{12}'y_{21}'} & \frac{-2y_{12}'}{(1+y_{11}')(1+y_{22}')-y_{12}'y_{21}'} \\ \frac{-2y_{21}'}{(1+y_{11}')(1+y_{22}')-y_{12}'y_{21}'} & \frac{(1+y_{11}')(1-y_{22}')+y_{12}'y_{21}'}{(1+y_{11}')(1+y_{22}')-y_{12}'y_{21}'} \end{bmatrix}$$

$$Y' = \begin{bmatrix} y_{11}' & y_{12}' \\ y_{21}' & y_{22}' \end{bmatrix} = \begin{bmatrix} \dfrac{(1-s_{11})(1+s_{22})+s_{12}s_{21}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}} & \dfrac{-2s_{12}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}} \\ \dfrac{-2s_{21}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}} & \dfrac{(1+s_{11})(1-s_{22})+s_{12}s_{21}}{(1+s_{11})(1+s_{22})-s_{12}s_{21}} \end{bmatrix}$$

- $S \leftrightarrow H$

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} \dfrac{(h_{11}'-1)(h_{22}'+1)-h_{12}'h_{21}'}{(h_{11}'+1)(h_{22}'+1)-h_{12}'h_{21}'} & \dfrac{2h_{12}'}{(h_{11}'+1)(h_{22}'+1)-h_{12}'h_{21}'} \\ \dfrac{-2h_{21}'}{(h_{11}'+1)(h_{22}'+1)-h_{12}'h_{21}'} & \dfrac{(1+h_{11}')(1-h_{22}')+h_{12}'h_{21}'}{(h_{11}'+1)(h_{22}'+1)-h_{12}'h_{21}'} \end{bmatrix}$$

$$H' = \begin{bmatrix} h_{11}' & h_{12}' \\ h_{21}' & h_{22}' \end{bmatrix} = \begin{bmatrix} \dfrac{(1+s_{11})(1+s_{22})-s_{12}s_{21}}{(1-s_{11})(1+s_{22})+s_{12}s_{21}} & \dfrac{2s_{12}}{(1-s_{11})(1+s_{22})+s_{12}s_{21}} \\ \dfrac{-2s_{21}}{(1-s_{11})(1+s_{22})+s_{12}s_{21}} & \dfrac{(1-s_{11})(1-s_{22})-s_{12}s_{21}}{(1-s_{11})(1+s_{22})+s_{12}s_{21}} \end{bmatrix}$$

Appendix 2. Process Flow for High Frequency GaN HEMTs

Mesa isolation

- Solvent cleaning (Acetone - Methanol - IPA)
- Photolithography
 (OCG825 Spin 5/30s@750/3000 rpm, Exp. 5s@Low-Vac contact, Dev. 2m@OCG934)
- Mesa etching (ECR-RIE 600s@BCl_3/Cl_2 = 20/5 sccm, ECR/RF = 50/25 W)
- PR removal (Acetone)

Ohmic contact

- Solvent cleaning (Acetone - Methanol - IPA)
- Photolithography
 (AZ5214 Spin 5/30s@750/4000rpm, Exp. 5s@Hard contact, 2m@115°C, Exp. 80s, Dev. 2m@422MIF)
- Metal evaporation (Si/Ge/Ti/Al/Ni/Au = 20/20/200/1000/250/500 Å)
- Lift off (Acetone)
- RTA (30s@820 °C)

Optical gate contact / Pad

- Solvent cleaning (Acetone - Methanol - IPA)
- Photolithography
 (AZ5214 Spin 5/30s@750/4000rpm, Exp. 5s@Hard contact, 2m@115°C, Exp. 80s, Dev. 2m@422MIF)
- Metal evaporation (Ni/Au = 200/2300 Å)
- Lift off (Acetone)

E-beam gate contact (T-gate in Chapter 3)

- Solvent cleaning (Acetone - Methanol - IPA)
- E-beam lithography

(ZEP/PMGI/ZEP Spin 60s@3000rpm each → 60s@180°C, Head Exp. 30 keV/20μm aperture/Dose 125, Dev. 90s@MIBK:MEK=1:1 → 30s@CD26, Foot Exp. 30 keV/30μm aperture/Dose 400, Dev. 90s@MIBK:IPA=1:3 at 0 ~ 5°C)

- Gate recess (ECR-RIE, BCl_3/Cl_2 = 15/5 sccm, ECR/RF = 100/75 V)
- Metal evaporation (Ni/Au = 200/2300 Å)
- Lift off (NMP)

E-beam gate contact (Δ-gate in Chapter 4)

- Solvent cleaning (Acetone - Methanol - IPA)
- E-beam lithography

(ZEP Spin 60s@3000rpm → 60s@180°C, Exp. 30 keV/30μm aperture/Dose 400, Dev. 90s@MIBK:IPA=1:3 at 0 ~ 5°C)

- Al_2O_3 passivation (ALD Al_2O_3 = 100 Å@80°C)
- Gate recess (ECR-RIE, BCl_3/Cl_2 = 15/5 sccm, ECR/RF = 100/75 V)
- O_2 plasma treatment (Asher 10m@800W)
- Metal evaporation (Ni/Au = 100/700 Å)
- Lift off (NMP)

Passivation

- Solvent cleaning (Acetone - Methanol - IPA)
- Passivation (ALD Al_2O_3 = 250 Å)
- Photolithography

(OCG825 Spin 5/30s@750/3000 rpm, Exp. 5s@Low-Vac contact, Dev. 2m@OCG934)

- Pad open (ECR-RIE 400s@BCl_3/Cl_2 = 15/5 sccm, ECR/RF = 100/75 V)
- PR removal (Acetone)

Appendix 3. Process flow for Si MOSFETs and GaN HEMTs

[Si MOSFET]

A. Active Region			
1.	RCA	Acid-hood	Piranha (5min) + 1:50 HF:H_2O (15sec) + 1:1:6 HCl:H_2O_2:H_2O (5min)
2.	Field oxide depo.	Tube (or sts-CVD)	3000 Å
3.	Coat PR on front side	Coater	OCG825 1μm
4.	Bake	Hotplate	5m@110°C
5.	Remove oxide from backside	Acid-hood	BOE dip
6.	Remove PR by Piranha	Acid-hood	Piranha (10min)
7.	Coat PR	Coater	OCG825 1μm
8.	Bake	Hotplate	5m@110°C
9.	Pattern Active	Ksaligner 2	Hard contact, 30s
10.	Develop PR	Photo-wet-r	2m 30s in OCG934 1:1
11.	Etch Field oxide-step1 (leave 500 Å)	Plasmaquest (or sts1)	Cl_2 or SF_6 plasma
12.	Remove PR by Piranha	Acid-hood	Piranha (10min)
13.	Etch Field oxide-step2 (remove 500 Å oxide)	Acid-hood	1:50 HF:H_2O (depends on etch rate)
B. Gate / G_{ox} / Implantation			
1.	RCA	Acid-hood	Piranha (5min) + 1:50 HF:H_2O (15sec) + 1:1:6 HCl:H_2O_2:H_2O (5min)
2.	Gate Oxide depo.	Tube (or sts-CVD)	100 Å
3.	Poly (undoped) depo.	Tube (or eBeamAu)	1500 Å
4.	Coat PR on front side	Coater	OCG825 1μm
5.	Bake	Hotplate	5m@110°C
6.	Remove Poly from backside	XeF_2	Selective over underneath Gate Oxide
7.	Remove Gate Oxide from backside	Acid-hood	BOE dip
8.	Remove PR by Piranha	Acid-hood	Piranha (10min)
9.	Coat PR	Coater	OCG825 1μm
10.	Bake	Hotplate	5m@110°C
11.	Pattern Gate	Ksaligner 2	Hard contact, 30s
12.	Develop PR	Photo-wet-r	2m 30s in OCG934 1:1
13.	Etch Poly-step1	Plasmaquest (or sts1)	Cl_2 or SF_6 plasma

14.	Etch Poly-step 2	XeF_2	Selective over underneath Gate Oxide
15.	Remove PR by Piranha	Acid-hood	Piranha (10min)
16.	Implantation (Gate, Drain, & Source)	Innovion	

[GaN HEMT]

A. GaN Field Exposure			
1.	Solvent Cleaning	Photo-wet-r	Acetone - Methanol - Iso
2.	Coat PR	Coater	OCG825 1μm
3.	Bake	Hotplate	5min@110°C
4.	Pattern GaN Field	Ksaligner 2	Hard contact, 30s
5.	Develop PR	Photo-wet-r	2m 30s in OCG934 1:1
6.	Etch Oxide/Si/HSQ	sts1	SF_6 plasma (GaN is a etch stop layer)
7.	Etch HSQ residue	Acid-hood	BOE dip
8.	Remove PR	Photo-wet-r	Acetone (or NMP) with ultrasonic
B. Source/Drain			
1.	Solvent Cleaning	Photo-wet-r	Acetone - Methanol - Iso
2.	Coat PR	Coater	AZ5214 (image reversal)
3.	Bake	Hotplate	5min@110°C
4.	Pattern S/D	Ksaligner 2	Hard contact, 30s
5.	2^{nd} Bake	Hotplate	2min@110°C
6.	Flood exposure	Ksaligner 2	Hard contact, 180s
7.	Develop PR	Photo-wet-r	2min in 422MIF
8.	Descum	Asher	5min@1000W
9.	Oxide strip	Acid-hood	1:3 $HCl:H_2O$ (1min)
10.	S/D metal depo.	eBeamFP	Ti/Al/Ni/Au = 200/1000/250/500 Å
11.	Lift-off	Photo-wet-r	Acetone (overnight)
12.	Solvent Cleaning	Photo-wet-r	Acetone - Methanol - Iso
13.	RTA	RTA35	30sec@870°C (S/D ohmic contact & Activation)
C. Mesa Isolation			
1.	Solvent Cleaning	Photo-wet-r	Acetone - Methanol - Iso
2.	Coat PR	Coater	OCG825 1μm
3.	Bake	Hotplate	5min@110°C
4.	Pattern Mesa	Ksaligner 2	Hard contact, 30s
5.	Develop PR	Photo-wet-r	2m 30s in OCG934 1:1
6.	Etch GaN (1800 Å)	Plasmaquest	Cl_2/BCl_3 plasma
7.	Remove PR	Photo-wet-r	Acetone (or NMP) with ultrasonic
D. Gate			
1.	Solvent Cleaning	Photo-wet-r	Acetone - Methanol - Iso
2.	Coat PR	Coater	AZ5214 (image reversal)
3.	Bake	Hotplate	5min@110°C

4.	Pattern Gate	Ksaligner 2	Hard contact, 30s
5.	2nd Bake	Hotplate	2min@110°C
6.	Flood exposure	Ksaligner 2	Hard contact, 180s
7.	Develop PR	Photo-wet-r	2min in 422MIF
8.	Descum	Asher	5min@1000W
9.	Gate metal depo.	eBeamFP	Ni/Au/Ni = 300/2000/500 Å
10.	Lift-off	Photo-wet-r	Acetone (overnight)

[Pad]

A. Via			
1.	Solvent Cleaning	Photo-wet-r	Acetone - Methanol - Iso
2.	Descum	Asher	5min@1000W
3.	Oxide strip	Acid-hood	1:3 HCl:H_2O (1min)
4.	Passivation (ILD)	sts-CVD	SiO_2 or SiN 5000 Å
5.	Coat PR	Coater	AZ5214 (image reversal)
6.	Bake	Hotplate	5min@110°C
7.	Pattern Via	Ksaligner 2	Hard contact, 30s
8.	Develop PR	Photo-wet-r	2m 30s in OCG934 1:1
9.	Etch Via- step1	Plasmaquest (or sts1)	Cl_2 or SF_6 plasma
10.	Etch Via- step2	Acid-hood	BOE
11.	Remove PR	Photo-wet-r	Acetone (or NMP) with ultrasonic
B. Pad			
1.	Premetal cleaning	Acid-hood	Piranha (10min) + 1:50 HF:H_2O (15sec)
2.	Pad metal depo.	eBeamFP	Ti/Al = 1000Å / 1μm
3.	Coat PR	Coater	AZ5214 (image reversal)
4.	Bake	Hotplate	5min@110°C
5.	Pattern Pad	Ksaligner 2	Hard contact, 30s
6.	Develop PR	Photo-wet-r	2m 30s in OCG934 1:1
7.	Etch Pad metal	Acid-hood	Ti/Al etchant
8.	Remove PR	Photo-wet-r	Acetone (or NMP) with ultrasonic
9.	Sinter	Tube	30m@400°C