InGaAs MOSFETs for Logic and RF Applications: Reliability, Scalability and Transport Studies

by

Xiaowei Cai

B.S., Optical Engineering, University of Rochester
B.A., Physics, University of Rochester
S.M., Electrical Engineering and Computer Science, MIT

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

June 2019

Signature of author: _____

Department of Electrical Engineering and Computer Science

May 23, 2019

Certified by: _____

Jesús A. del Alamo

Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: _____

Leslie A. Kolodziejski

Professor of Electrical Engineering and Computer Science

Chair, Committee for Graduate Students

InGaAs MOSFETs for Logic and RF Applications: Reliability, Scalability and Transport Studies

by Xiaowei Cai

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

ABSTRACT

InGaAs has emerged as an extraordinary n-channel material due to its superb electron transport properties and low voltage operation. With tremendous advancements over the years, InGaAs MOSFETs have attracted much attention as promising device candidates for both logic and THz applications. However, many challenges remain. This thesis addresses some of the critical issues facing InGaAs MOSFETs and advances the understanding of the limiting factors confronting InGaAs MOSFET technology.

First, it identifies a new instability mechanism in self-aligned InGaAs MOSFETs caused by fluorine migration and passivation of Si dopants in n-InAlAs. This problem is successfully mitigated by eliminating n-InAlAs from the device structure. The new device design achieves improved stability and record device performance.

Second, it evaluates the impact of oxide trapping in InGaAs MOSFETs. A comprehensive PBTI study shows that oxide trapping deteriorates device stability, resulting in threshold voltage shifts and degraded device performance. In addition, oxide trapping also compromises DC device performance. High frequency and fast pulse measurements reveal a rich spectrum of oxide traps with different capture/emission times. Furthermore, oxide trapping also complicates the extraction of fundamental parameters in InGaAs MOSFETs and leads to an underestimation of channel mobility. Thus, a new method has been developed, immune to the impact of oxide traps, to evaluate the intrinsic charge-control relationship of the device, and accurately estimate mobility.

Thirdly, this thesis re-evaluates the impact of channel scaling on device performance and transport in InGaAs planar MOSFETs and FinFETs. In both cases, mobility degradation with channel thickness or fin width scaling is observed to be much less than suggested by conventional CV methods. When the impact of oxide trapping is avoided, mitigated scaling induced degradation is observed and promising intrinsic transistor performance is revealed. Notably, InGaAs FinFETs exhibit $g_{m,max}$ at 1 GHz competitive with current Silicon FinFET technology and high mobility even in very narrow fins (μ_{peak} ~ 570 $cm^2/V \cdot s$ at W_f = 7 nm).

This thesis highlights the importance of mitigating oxide trapping. Further, in light of the results obtained here, the prospects of InGaAs MOSFET technology merit a reassessment.

Thesis supervisor: Jesús A. del Alamo
Professor of Electrical Engineering and Computer Science

Acknowledgments

This thesis would not have been possible without the help, support and love of many. First and foremost, I want to express my sincere gratitude to my advisor Prof. Jesús del Alamo, for taking me as his student, encouraging and challenging me, and providing all the guidance to complete this thesis. His tremendous knowledge and insight sparked my interest in semiconductor device physics when I was a student in his class 6.720; over the years, he has continuously inspired and taught me how to be a researcher through his passion for technology, attention to detail and pursuit of excellence. I am deeply grateful for the kindness, support, patience and dedication that he has shown me. It has been a true honor and privilege to be his student.

I would also like to thank my thesis committee members, Profs. Tayo Akinwande and Tomás Palacios, for taking time to review my work and offering many great suggestions. Outside of my committee, I'd like to thank Prof. Dimitri Antoniadis for much insightful advice on device physics, my academic advisor Prof. Marc Baldo for providing perspective on broader goals in life beyond research, Prof. Qing Hu for teaching me how to do research when I first started graduate school, and Prof. Leslie Kolodziejski for being there for me through many hard times.

During my PhD, I also had the privilege to work with many researchers around the world: Prof. Jesús Grajal and Dr. Jorge Pedros at UPM in Spain; Prof. Jörg Schulze, Linda Hänel and Yasmine Elogail at the University of Stuttgart; Prof. Anthony Muscat and Yissel Contreras Valeriano at the University of Arizona; and Dr. Nerissa Draeger at Lam Research Corp. I have greatly benefited from their generous sharing of knowledge, expertise and facilities. I am particularly thankful to Prof. Grajal, who has taught me everything that I know about RF measurements and theory. I really appreciate his hands-on teaching and humor.

Here at MIT, I am blessed with a wonderful community: MTL. None of my work would be possible without the hard work of all the MTL cleanroom and computer staff. I am particularly thankful to Dennis Ward, Eric Lim, Bob Bicchieri, Kurt Broderick, Mark Mondol and Bill Maloney, for always being willing to lend a helping hand. I also deeply appreciate the MTL administration staff, in particular Debb Hodges-Pabón, Valerie DiNardo, Mara Karapetian, Mary

O’Neil, Elizabeth Kubicki, and Joe Baylon; thank you for making MTL feel less like a lab and more like a big family.

My time at MIT would have been incredibly dry without many wonderful friends. Within the Xtreme Transistors Group, in particular, I would like to thank Jerome Lin for his mentorship during my first year in the group and for teaching me device fabrication and characterization; and Alon Vardi for generously providing the FinFET devices used in this thesis and many advice. I also thank Wenjie Lu, Shireen Warnock, Alex Guo, Yufei Wu, Xin Zhao, Ethan Lee, Taekyong Kim, Aviram Massuda, Youngtack Lee, Moshe Tordjman, Alex Lednev, and Dongsung Choi for many years of friendship, support and help. To my friends in Bldg 39, thank you for the great conversations and long hours spent together in the cleanroom and measurement lab: Winston Chern, Ujwal Radhakrishna, Ahmad Zubair, Cosmi Lin, Sameer Joglekar, Lili Yu, Daniel Piedra, Omair Saadat, Marek Hempel, Charles Mackin, Parker Gould, Mitchell Hsing, Saima Siddiqui, Joseph Finley, Jamie Teherani and Jorg Scholvin. To my other friends at MIT, thank you for all the laughs and tears we’ve shared together over the last 7 years: Samantha Strasser, Reyu Sakakibara, Julian Straub, Joel Jean, Joseph Kim, Vincent Xue, Anjuli Jain Figueroa, Dave Burghoff and Shengxi Huang.

I am grateful to all the brothers and sisters in Christ that God has placed in my life, both at the ABSK and Park Street Church. Thank you for all the prayers and love that you have poured out on Ivan and myself. I especially want to thank Irene unni, Ce ge, Hyejin smn and Pastor Joseph for teaching me the word of God and leading me to Christ.

To my family back in China, thank you for raising me in such love, especially my parents. Without them and their tremendous sacrifice, I would not be able to come so far today. I also want to thank my husband Ivan, for his love and support throughout the years.

Finally, and most importantly, I am forever indebted to God, for His love and faithfulness.

This work was sponsored by DTRA, Lam Research, SRC and MIT MISTI.

Contents

List of Figures

List of Tables

CHAPTER 1. Introduction

Following the invention of the bipolar junction transistor by Shockley in 1948 and the integrated circuit (IC) by Kilby in 1958, semiconductor technology has grown tremendously and expanded into every aspect of society. For the past 50 years, the prediction made by Gordon Moore in 1965 that transistor density in an integrated circuit doubles roughly every two years [1] (shown in Figure 1.1 [2])–known as "Moore's Law"— has been the driving force behind the success of the microelectronics industry. The drive towards ever-smaller transistors has led to exponential improvements in transistor density, switching speed and power efficiency of modern electronics, along with the benefit of reduced cost. Silicon has been the main workhorse in achieving such

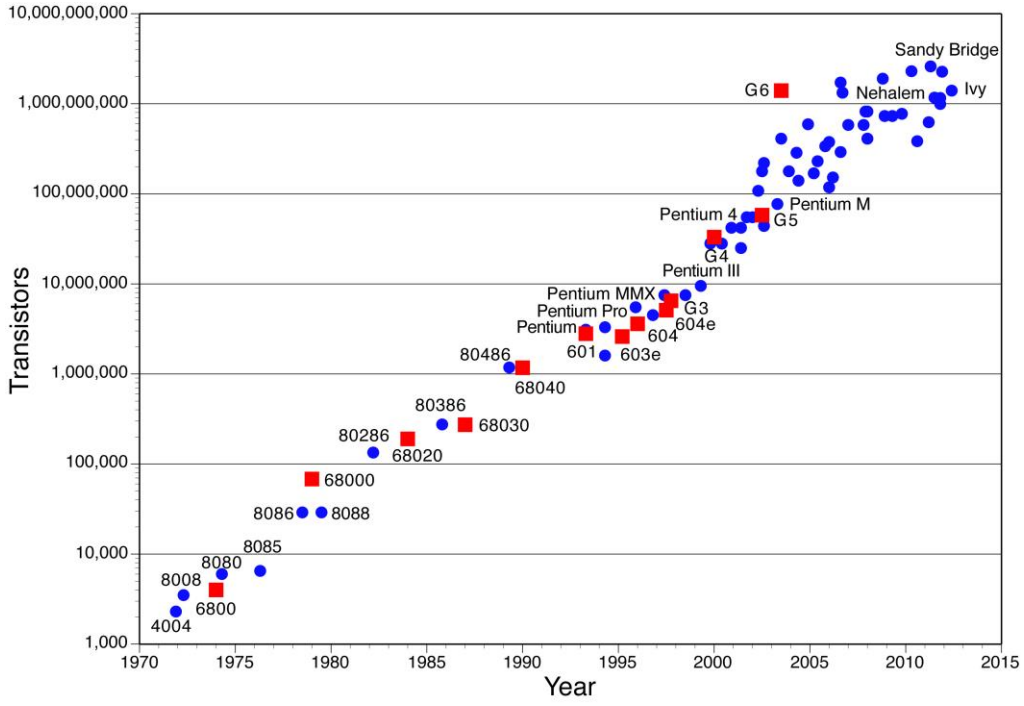

Figure 1.1: Moore's law: the number of transistors on integrated circuit chips over the years (reproduced from [2]).

progress in semiconductor technology. Nonetheless, research efforts are intensifying to find alternative materials for future electronics [3].

1.1 Motivation for III-V Transistors

1.1.1 Transport Advantages

III-V compound semiconductors have attracted much attention throughout the history of electronics [4], [5]. InGaAs, in particular, has shown promise as an n-channel material due to its extraordinary transport properties compared to Si. As a result of the smaller effective mass compared to Si ($In_{0.53}Ga_{0.47}As$ $m_e^* \sim 0.041$ versus Si $m_t^* \sim 0.19$ for electrons [6]), InGaAs exhibits much higher electron mobility ($\mu \propto 1/m^*$) and injection velocity ($v_{\text{inj}} \propto \sqrt{1/m^*}$).

Figure 1.2 shows the room temperature inversion-layer mobility of electrons and holes in various III-V materials, in comparison with Si [4]. Depending on composition and strain, the electron mobility of $In_xGa_{1-x}As$ ranges from 6,000 to 30,000 $cm^2/V{\cdot}s$, far superior to that of Si. $In_xGa_{1-x}As$ can be grown on a lattice-mismatched pseudomorphic InP substrate, and the In composition can be varied to tailor the material and electronic properties. For example, as *x* changes from 0.53 to 1, biaxial compressive strain changes from 0% to 3.1%, and the bandgap changes from 0.81 to 1.42 eV [6]. This opens up the potential for bandgap engineering in various applications.

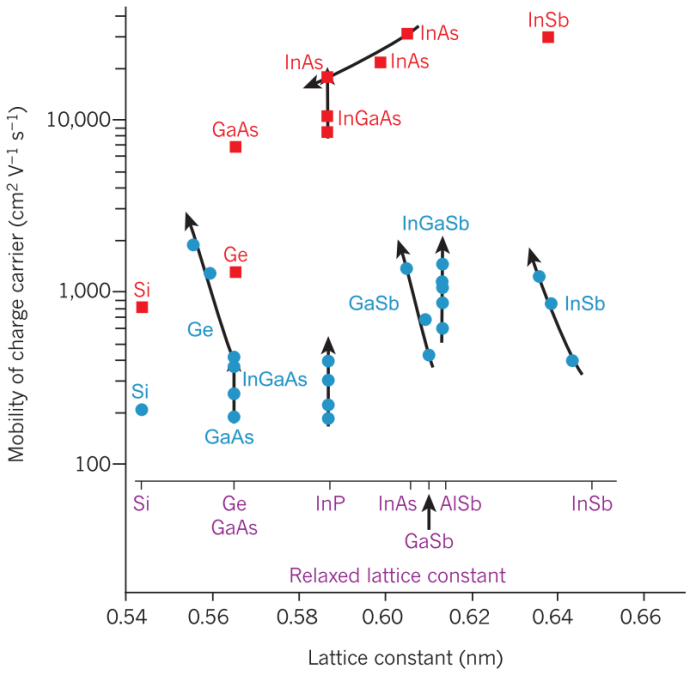

Figure 1.2: Room-temperature inversion layer mobility of electrons (red) and holes (blue) as a function of actual semiconductor lattice constant. The arrows represent increasing compressive biaxial strain (reproduced from [4]).

I_{on} in modern nanoscale FETs is governed by the product of carrier concentration and injection velocity v_{inj}, as the device operates in the regime of nonequilibrium and quasi-ballistic transport [3], [7]. Figure 1.3 shows injection velocity v_{inj} measured in InGaAs HEMTs at V_{DS} = 0.5 V and Si MOSFETs at V_{DS} = 1.1-1.3 V [4]. InGaAs as an n-channel material can deliver at least twice the injection velocity of strained-silicon at less than half of the supply voltage. As In composition increases, v_{inj} further increases. In InAs HEMTs, v_{inj} exceeds 3.5×10^7 cm/s at V_{DS} = 0.5 V. This enables a reduction in drain-source voltage without any compromise in device performance. Thus, InGaAs promises high performance at low power consumption. This advantage applies to both logic and high-speed applications, and suggests the potential of InGaAs to drive continued device performance improvement and scaling.

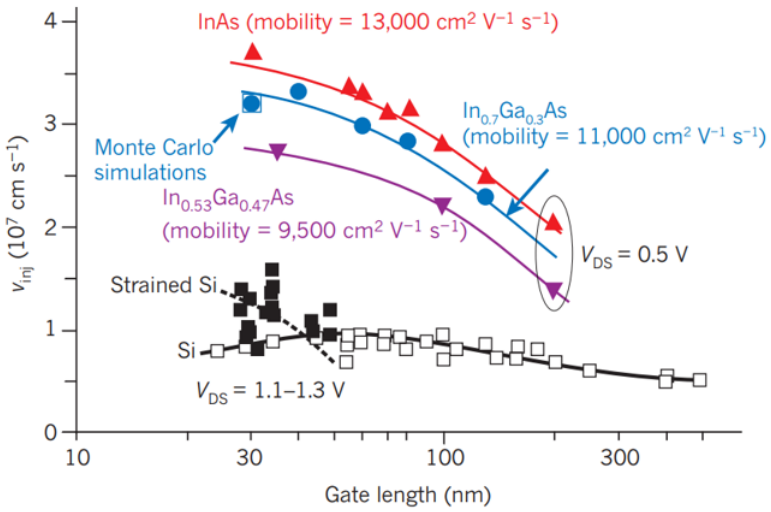

Figure 1.3: Electron injection velocity in InGaAs and InAs HEMTs (at V_{DS} = 0.5 V), and Si MOSFETs (at V_{DS} = 1.1-1.3 V) as a function of gate length (reproduced from [4]).

1.1.2 RF applications

III-V transistors have been widely used for radio-frequency (RF) and millimeter-wave applications such as wireless and satellite communication [8]–[10]. Such technologies include GaAs and InGaAs metal-semiconductor field-effect transistors (MESFETs), hetero-junction bipolar transistors (HBT), and high electron mobility transistors (HEMTs). InGaAs HEMTs in particular are known for their high mobility, high frequency performance, and low noise characteristics [8], [9].

The unique architecture of a HEMT features a heterostructure charge-supplying barrier layer outside the channel to achieve carrier transport with minimal scattering in the channel. However, improvement in III-V HEMTs has been stagnating as gate leakage current limits the scaling of barrier layer thickness, which further limits the reduction in parasitic resistance [8], [11].

The solution to this quandary is MOSFETs [8], [11]. Replacing the heterostructure barrier layer in HEMTs with a wide-bandgap high-k dielectric, the MOSFET structure can both suppress gate leakage current and reduce parasitic resistances. The scaling of oxide thickness in MOSFETs can enable further shrinking of lateral dimensions and improve RF performance.

1.1.3 Logic Applications

In addition to RF applications, there has been tremendous interest and research efforts in exploring InGaAs as an alternative n-channel material for future nanoscale CMOS logic technology [4], [12]. The superior transport and low voltage operation of InGaAs is particularly attractive, as we have entered the era of power constrained scaling. In recent years, Si technology has faced increasing difficulty in continuing to scale down device footprint while maintaining reasonable power dissipation [13]. CMOS power consumption is divided into two components,

$$\begin{aligned} P_{\text{total}} &= P_S + P_D \\ &= I_{\text{off}} V_{DD} + \alpha f C V_{DD}^2 \end{aligned} \tag{1.1}$$

where P_S is the static power consumption, P_D is the dynamic power consumption, I_{off} is the OFF-state leakage current, V_{DD} is the supply voltage, α is the activity factor, f is the clock frequency, and C is the load capacitance. Increasing short channel effects in modern highly scaled transistors lead to increasing I_{off}, and hence increased static power consumption P_S. The scaling down of V_{DD} has saturated at around 0.7 V [14], as further decreases would sacrifice ON-state device performance. This has inhibited the reduction of dynamic power consumption P_D. InGaAs can directly address power constrained scaling by delivering high drive currents at significantly lower voltage operation than modern Si CMOS [4].

Replacing Si with a new channel material such as InGaAs would not be the first major innovation used to continue CMOS scaling. At 90 nm node, SiGe source/drain stressors were introduced to enhance channel mobility [15]. At 45 nm node, high-k/metal gate (HKMG) was adopted to scale EOT without excessive gate leakage [16], [17]. At 22 nm node, planar devices became obsolete, and were replaced by tri-gate transistors [18]. Extraordinary innovations will be needed to further extend Moore's Law beyond 7 nm.

1.2 Progress in InGaAs MOSFETs

The development of the III-V MOSFET was largely delayed for decades, while the III-V HEMT has been a successful electronic technology for nearly 40 years since its first demonstration by Mimura et al. in 1980 [19]. The main roadblock was the poor III-V/oxide interface, which can cause Fermi-level pinning. The lack of a good native oxide in III-V results in a large number of defects (Ga and As oxides, elemental As, As-As dimers, Ga dangling bonds etc. [20], [21]), preventing the modulation of surface potential in the channel. A solution was found in the introduction of atomic layer deposition (ALD) in III-V MOSFET device fabrication in 2003 [22]. High-k gate dielectric deposition by ALD was already a mature manufacturing process in the Si industry at that time [16]. Due to the "self-cleaning effect", defective native oxides on III-V surfaces can be substantially removed in the first few cycles of ALD [20], [23], yielding a Fermi-level unpinned interface.

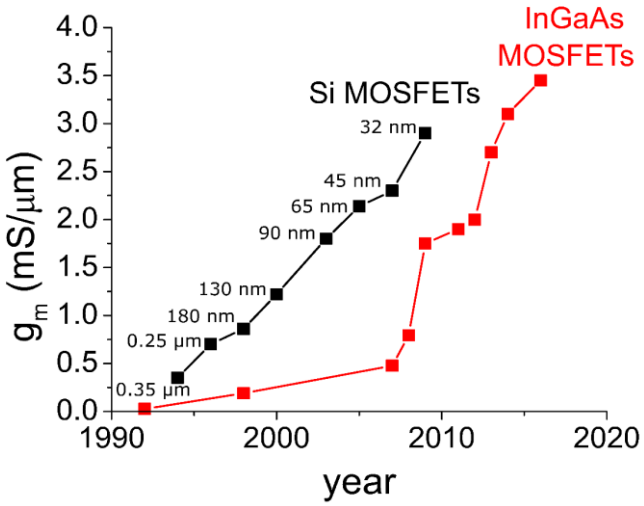

Figure 1.4: Transconductance comparison of inversion-type InGaAs MOSFETs and Si MOSFETs (with InAs composition between 0 and 1) vs. year (reproduced from [12]).

Following the introduction of ALD, InGaAs MOSFET technology advanced rapidly, propelled by the development of many other key enabling technologies. These include refractory ohmic contacts with low contact resistivity [24], water and alcohol based digital etch [25], [26], and precision dry etching [27]. Figure 1.4 charts the transconductance improvement of planar InGaAs MOSFETs in comparison to planar Si MOSFETs over the years [12]. A record transconductance of 3.45 mS/mm was demonstrated in 2016, exceeding the last planar Si technology of Intel at the 32 nm node [12].

To target sub-7 nm CMOS applications, research interest in InGaAs MOSFETs has shifted from planar [28]–[33] to 3D device architectures, including FinFET [34], [35], tri-gate [36], and Gate-All-Around transistor designs [37]–[40]. Recently, the most aggressively scaled InGaAs FinFETs to date were demonstrated with fin widths down to 2.5 nm, incorporating for the first time atomic layer etching (ALE) with *in situ* ALD [35]. The remarkable electrostatic control and improvement in device performance displays the great promise of III-V for CMOS technology. In addition, InAs

MOSFETs targeting RF applications have achieved f_{max} of 410 GHz and f_T = 357 GHz [41]. All this progress demonstrates the promising performance and scalability of InGaAs technology.

1.3 Challenges

Despite the tremendous progress in InGaAs MOSFETs, many challenges remain. These include device reliability, gate-stack quality, mobility degradation with intrinsic channel scaling, OFF-state leakage due to band-to-band tunneling, and V_t variation due to quantization effects [34], [42]–[45], etc. Understanding and mitigating these problems is crucial for the adoption of InGaAs MOSFET technology.

This thesis addresses three of the major challenges: device reliability, oxide trapping, and performance degradation as a result of channel scaling. It is critical to find the root cause of each issue, and differentiate between intrinsic factors arising from fundamental material/transport limitations of InGaAs, and extrinsic factors such as imperfections introduced during device fabrication. Only by excluding these extrinsic factors can we accurately assess the intrinsic potential of InGaAs MOSFET technology.

1.3.1 Device Reliability

While the process technology and performance of InGaAs MOSFETs continues to improve, there is increasing concern regarding the electrical reliability of this device technology, due to bias temperature instability, ion contamination, radiation damage, etc. [43], [46], [47]. Reliable operation of the transistor is critical in actual applications.

In InGaAs MOSFETs, there have been multiple reports of positive threshold voltage (V_t) shifts coupled with transconductance (g_m) degradation following positive gate voltage stress. This is known as Positive-Bias Temperature Instability (PBTI) [48]–[50]. In our self-aligned devices fabricated by precision Reactive Ion Etching (RIE), in contrast, we observe a prominent but fully reversible g_m enhancement under PBTI conditions.

We hypothesize that this phenomenon is caused by fluorine contamination introduced during device fabrication. As F-based RIE and chemical treatments are important in III-V device fabrication, understanding and mitigating the effects of F contamination on performance and reliability are crucial for the adoption of InGaAs MOSFETs for future CMOS and RF applications.

1.3.2 Oxide Trapping

One of the most difficult challenges in III-V MOSFETs is to achieve a high quality gate oxide and III-V/oxide interface. The gate stack, consisting of a metal gate, a high-k dielectric and semiconductor channel, is at the heart of a MOSFET. The lack of a good InGaAs native oxide coupled with processing temperature constraints often leads to a defective gate oxide and oxide/semiconductor interface.

The distribution of preexisting oxide traps in the bulk of the oxide can be depicted as the dark shadow in Figure 1.5. They are believed to originate from the oxygen vacancies in the high-k oxide [43]. When a positive gate voltage is applied, electrons in the channel can tunnel into empty trap states in the oxide. Depending on their energy location inside the oxide bandgap and distance from the interface, different states are associated with different capture/emission time constants. During

device operation, electrons are trapped in these defect states instead of contributing to conduction in the channel, resulting in a reduction in free carrier density and compromised device performance. The interaction of carriers and traps can also degrade device reliability over time. Reports of hysteresis, threshold voltage instability and frequency dispersion in InGaAs MOSFETs are manifestations of this trapping behavior [43], [51], [52]. Furthermore, trapping can also complicate the extraction of fundamental parameters, such as carrier mobility [53]. Considering the extensive and detrimental impact of oxide trapping in InGaAs MOSFETs, understanding it is important.

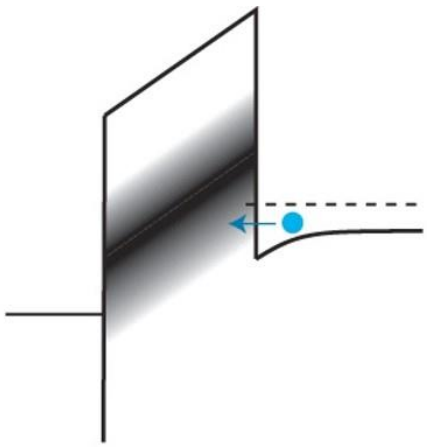

Figure 1.5: Energy band diagram of InGaAs MOSFETs in ON-state.

1.3.3 Performance Degradation with Channel Scaling

For both RF and logic applications, continued channel scaling of MOSFETs to smaller dimensions can be beneficial. It can improve electrostatic control over short-channel-effects (SCE), as well as reduce footprint in 3D device structures. However, many have reported device performance degradation with intrinsic channel scaling in InGaAs MOSFETs [34], [42], [54]. This is often attributed to transport loss, such as degraded mobility. Mobility data are often extracted from CV

measurements and further used to calibrate TCAD models [55]–[58], which subsequently predict that InGaAs MOSFET technology is unscalable [56]–[58]. However, as many researchers observe large frequency dispersion up to GHz frequency which reveals severe charge trapping in the oxide [51], [59], conventional mobility extraction method using CV measurements taken at MHz frequency becomes questionable. Furthermore, large unresolved discrepancies between mobility extracted using CV and Hall measurements are also reported [60], [61]. To correctly assess the scalability of InGaAs technology, we need to separate out the role of oxide trapping and study the intrinsic scaling of transport in InGaAs.

1.4 Thesis Overview

As stated in section 1.3, this thesis addresses the challenges regarding device reliability, oxide trapping and mobility degradation in scaled InGaAs MOSFETs. The remainder of this thesis is organized as follows.

Chapter 2 describes the heterostructures and process flow for the planar InGaAs MOSFETs used in this thesis. A simplified process is developed to fabricate planar InGaAs MOSFETs with a fast throughput for device physics study.

Chapter 3 investigates the anomalous instability caused by fluorine contamination in planar InGaAs MOSFETs. A comprehensive reliability study is carried out using temperature-dependent electrical stress experiments, along with independent material characterization by secondary ion mass spectroscopy (SIMS). A new device structure that is immune to fluorine is introduced and shows improved device performance and reliability.

Chapter 4 studies the impact of oxide trapping in InGaAs MOSFETs. A PBTI study is carried out to characterize device reliability deterioration caused by oxide trapping. Furthermore, high frequency measurements reveal that device DC performance is degraded and channel mobility is severely under-estimated in conventional CV method due to oxide trapping.

Chapter 5 establishes the methodology to restore charge control relationship free from the impact of oxide trapping and accurately estimate mobility. The impact of channel thickness scaling in planar InGaAs MOSFETs and fin width scaling in InGaAs FinFETs are investigated. Preliminary results on device treatment with forming gas annealing are discussed, highlighting the importance of oxide trap mitigation.

Chapter 6 summarizes the key contributions and makes suggestions for future research.

CHAPTER 2. Process Overview for Planar InGaAs MOSFETs

This chapter describes the heterostructures and process flow for the planar InGaAs MOSFETs used in this thesis. The heterostructures are grown by IntelliEpi. Inc. All fabrication steps take place in MIT's Microsystems Technology Laboratory (MTL) and Scanning Electron Beam Lithography (SEBL) facilities.

2.1 Heterostructures

The initial heterostructures are grown by molecular beam epitaxy (MBE) on 3-inch semi-insulating (100) InP substrates. From top to bottom, the heterostructures usually contain a composite n^+ cap of n-$In_{0.7}Ga_{0.3}As$/ n-$In_{0.53}Ga_{0.47}As$/n-InP, an undoped InP barrier, an undoped channel $In_xGa_{1-x}As$, and a 400-nm thick InAlAs buffer layer.

Figure 2.1 shows the three heterostructure designs used in this work. Heterostructure A has a uniform intrinsic channel composition and is used in the channel thickness dependent studies in chapters 4 and 5. Heterostructure B and C are used to compare the impact of fluorine on in InGaAs MOSFETs with and without an n-InAlAs layer, and study device reliability in chapters 3 and 4.

A

Material	Thickness (nm)	Dopant	Level (cm^{-3})
$In_{0.7}Ga_{0.3}As$	5	Si	3E19
$In_{0.53}Ga_{0.47}As$	20	Si	3E19
InP	2	Si	2E19
InP	2		UID
$In_{0.7}Ga_{0.3}As$	10		UID
$In_{0.52}Al_{0.48}As$	30		UID
InP	4		UID
$In_{0.52}Al_{0.48}As$	400		UID
InP substrate			S.I.

n^+ cap
channel
buffer

B

Material	Thickness (nm)	Dopant	Level (cm^{-3})
$In_{0.7}Ga_{0.3}As$	5	Si	3E19
$In_{0.53}Ga_{0.47}As$	5	Si	3E19
InP	9	Si	4E19
$In_{0.52}Al_{0.48}As$	3	Si	4E19
InP	3		UID
$In_{0.7}Ga_{0.3}As$	3		UID
InAs	2		UID
$In_{0.7}Ga_{0.3}As$	5		UID
$In_{0.52}Al_{0.48}As$	5		UID
δ-doping 1E12 cm^{-2}		Si	
$In_{0.52}Al_{0.48}As$	25		UID
InP	6		UID
$In_{0.52}Al_{0.48}As$	400		UID
InP substrate			S.I.

C

Material	Thickness (nm)	Dopant	Level (cm^{-3})
$In_{0.7}Ga_{0.3}As$	5	Si	3E19
$In_{0.53}Ga_{0.47}As$	5	Si	3E19
InP	14	Si	4E19
$In_{0.7}Ga_{0.3}As$	3		UID
InAs	2		UID
$In_{0.7}Ga_{0.3}As$	5		UID
$In_{0.52}Al_{0.48}As$	25		UID
InP	4		UID
$In_{0.52}Al_{0.48}As$	400		UID
InP substrate			S.I.

Figure 2.1: Three different heterostructures used in this work.

2.2 MOSFET Designs

A contact-first, gate-last, self-aligned architecture has been developed first for InGaAs HEMTs [62], [63] and later for InGaAs MOSFETs [28]–[30], [64]. The fabrication starts with sputtering the W/Mo contact, depositing the hardmask SiO_2 and patterning them using F-based Reactive Ion Etching (RIE). To achieve self-aligned gate and contacts, the cap is subsequently recessed by a combination of Cl-based RIE and digital etch. After the intrinsic channel is exposed, gate oxide and metal are deposited to form the gate stack.

Figure 2.2 shows the schematic cross section of a self-aligned planar InGaAs MOSFET. The contact-first approach not only yields low contact resistance, but also keeps the oxide-semiconductor surface from extensive processing and protects the gate stack quality. The quantum

confined thin channel further improves the electrostatic control of the gate over the channel. The self-aligned architectures, along with extensive use of RIE enables very tight contact-gate spacing, and subsequently low source and drain resistance parasitics and device footprint. InGaAs planar MOSFETs performance has greatly improved over the years and devices fabricated in such manner has achieved record performance [31]. In addition, the development of precision RIE dry etch and digital etch has also allowed nanometer-precision control of the critical device dimensions, including gate length (L_g), channel thickness (t_c) and access region length (L_{access}). This has enabled the planar devices to be a great vehicle for device physics study.

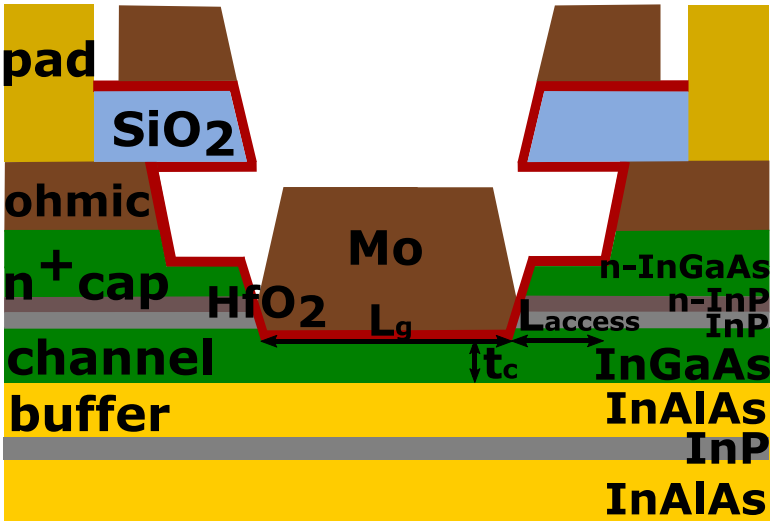

Figure 2.2: Schematic cross section of a self-aligned planar InGaAs MOSFET.

In high performance planar InGaAs MOSFETs such as those used in Chapters 3 and 4, extremely scaled gate oxide is used to boost the ON-performance, but this often results in high gate leakage current in long channel devices. Furthermore, to minimize device parasitics and achieve a highly self-aligned architecture, Electron beam (E-beam) lithography is heavily used and anisotropic RIE dry etch is carefully timed to recess the III-V n^+ cap. These steps result in a complicated and lengthy fabrication process.

For the purpose of device physics study in Chapter 5, a simplified process flow has been developed for fast throughput. Planar InGaAs MOSFETs can be fabricated with a wide range of gate length (from 50 nm - 15 μm) and various channel thicknesses, with a single E-beam lithography step. Many process steps are designed so that the devices are optimal for device physics study, instead of high performance. For example, a thicker gate oxide is used to reduce gate leakage. The recess of the III-V n^+ cap is finished with a selective wet etch to achieve an optimal semiconductor-oxide interface, at the cost of higher access resistance.

2.3 Process Overview

The diagram in Figure 2.3 shows the front-end process steps [65]. Details of each step are described below and can be referred in Appendix A.

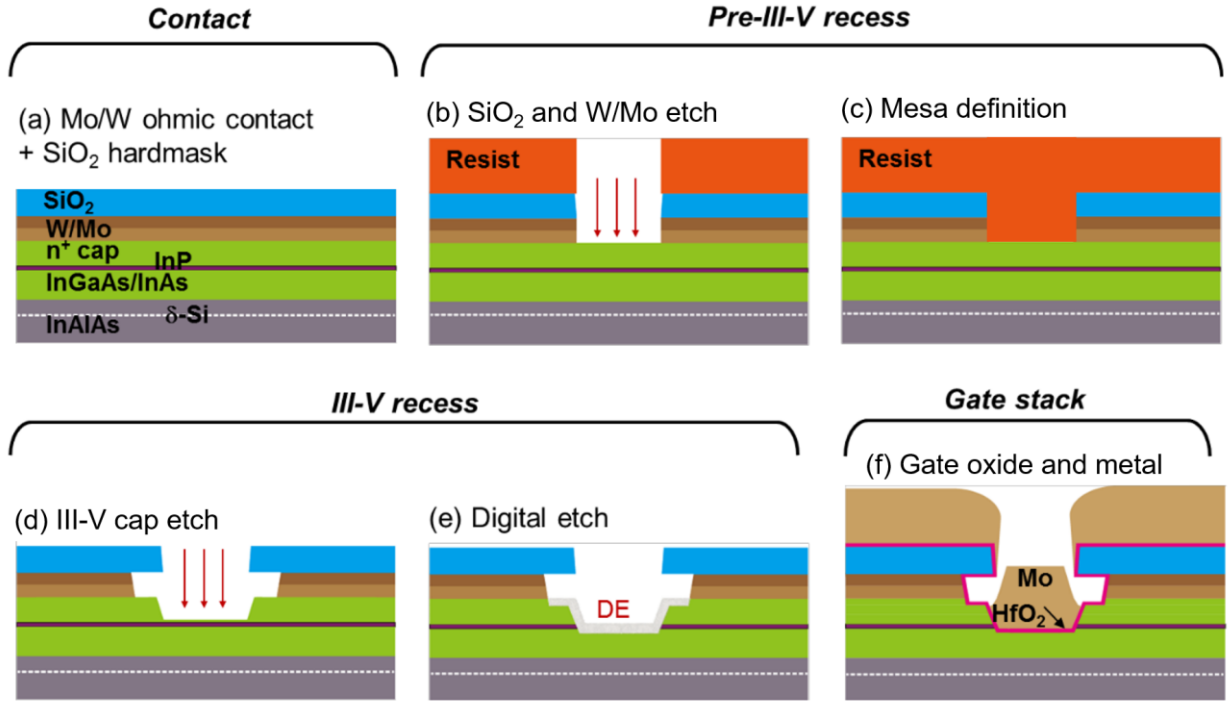

Figure 2.3: Illustration of front-end fabrication for InGaAs planar MOSFETs (a) ohmic metal and SiO_2 hardmask deposition (b) SiO_2 and W/Mo etch (c) mesa definition (d) III-V cap etch (e) digital etch and (f) gate oxide and metal deposition, adapted from [65].

Contact Deposition: After removing the native oxide by diluted HCl, a blanket of W/Mo (10/20 nm) ohmic contact bilayer is deposited by sputtering. Then, a 60 nm-thick SiO_2 hardmask is deposited by chemical vapor deposition (CVD).

Gate Foot Definition: Electron beam (E-beam) lithography is used to pattern the "gate foot" area. Positive E-beam resist ZEP520A or GL2000 can be used. Feature size ranging from 50 nm to 10's of micron can be achieved in a single lithography step after careful calibration of the dose for various feature sizes. A reduced dose is required for larger feature sizes due to the proximity effect. After patterning the resist, CF_4:H_2 and SF_6:O_2 anisotropic RIE dry etches are used to remove the SiO_2 hardmask and the W/Mo contact, respectively, stopping selectively on the III-V surface, shown in Figure 2.4 (a).

Mesa Isolation: Mesa can be patterned using photolithography with positive photoresist SPR 700 and subsequently defined through a series of RIE dry etches (CF_4:H_2 for SiO_2, SF_6:O_2 for W/Mo contact and Cl_2:N_2 for III-V).

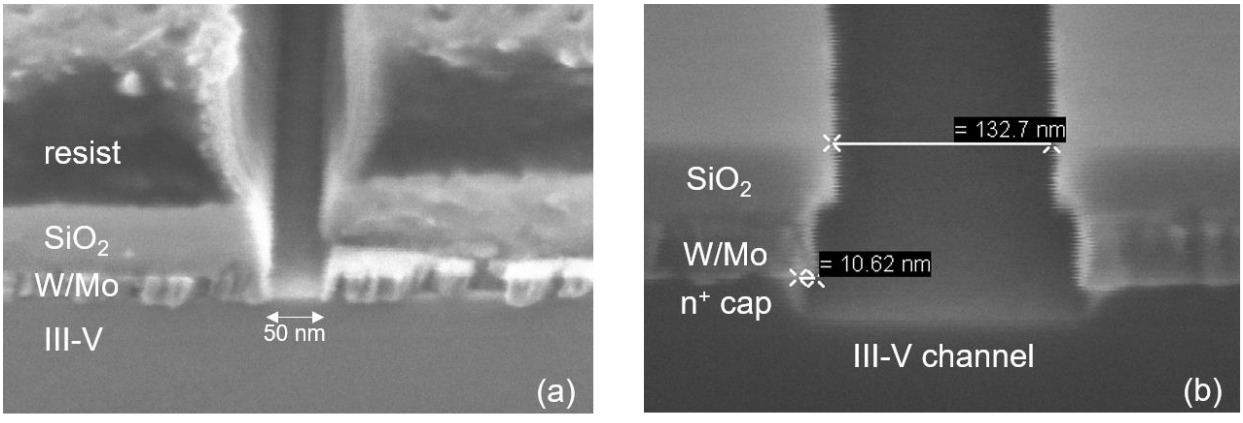

Figure 2.4: SEM images (a) after gate foot definition (b) after III-V n^+ cap removal using a combination of dry and wet etch.

III-V Recess: A combination of Cl_2:N_2 anisotropic RIE dry etch and selective wet etch are used to recess the composite III-V n^+ cap and stop on the InP layer. First, the Cl-based RIE is used to remove most of the n^+ cap. Its etch rate is calibrated such that the etch stops ~5-10 nm above the InP layer. Then, the citric acid:H_2O_2 wet etch is carefully timed to remove any remaining n^+ cap and stop selectively on the InP layer. In such manner, a minimal undercut and a high quality surface are achieved, shown in Figure 2.4 (b).

Digital Etch (DE): DE [25] is used to remove the InP layer and thin down the channel to a desired thickness. Each DE cycle consists of an oxidation step of the III-V surface through low power oxygen plasma and an oxide removal step using H_2SO_4:H_2O. It removes III-V material non-selectively at a rate of ~1 nm/cycle.

Gate Stack Formation: Immediately after the last DE cycle, high-k dielectric HfO_2 is deposited using atomic layer deposition (ALD) at 250 °C as the gate oxide and 35 nm-thick Mo is sputtered as the gate metal. Sputtering is used instead of evaporation to maximize the coverage of any undercut region of the channel due to the wet etch during the III-V recess. The gate hat region is subsequently patterned by photolithography with image reversal photoresist AZ5214E. Pad metal Ti/Au is deposited through lift-off and used as a hardmask to remove any Mo outside the gate hat region through RIE dry etch (SF_6:O_2).

Contact Pad: The contact pad for the source and drain are patterned using photolithography. HfO_2 and SiO_2 are removed through BCl_3:Ar RIE dry etch and BOE wet etch, respectively. Finally, the

contact metal Ti/Au are deposited by lift-off, to make contact with ohmic contact W/Mo. This step is not shown in Figure 2.3.

One improvement to this process could be to include an additional step of inter-level dielectric (ILD) deposition after the front-end processes, to further improve the device stability.

2.4 Summary

We have discussed the heterostructures and MOSFET architectures chosen for the different studies in the subsequent chapters. High-performance device structures are used in the investigation of fluorine-induced instability and the impact of oxide trapping in chapter 3 and 4, respectively. A simplified process flow that has been developed for further device physics studies in chapter 5. It is optimized for fast turn-around time, low gate leakage, and clean semiconductor-oxide interface.

CHAPTER 3. Fluorine-induced Instability

Chapter 2 has described the heterostructures, fabrication process and self-aligned structure of the planar InGaAs MOSFETs investigated in this thesis. In this chapter, we will focus on devices with n-InAlAs in the ledge. We report, for the first time, a prominent but fully reversible enhancement in transconductance after applying positive gate stress in these devices. We attribute this instability to fluorine (F) introduced during the gate recess process.

We first discuss the literature findings of fluorine-induced donor passivation and instability in n-InAlAs. Following that, we present a comprehensive study of device instability under various electrical stress conditions. This includes the positive gate voltage stress, off-state stress and temperature dependent stress. Furthermore, to independently confirm fluorine in our devices and induced donor passivation, secondary ion mass spectroscopy (SIMS) and transmission line method (TLM) structures are used. Finally, we introduce a new device design that is immune to the impact of fluorine.

3.1 Background: Fluorine in III-V Materials

The impact of Fluorine in III-V materials has been well studied in the 1990's. Many observe electrical degradation in InAlAs/InGaAs high electron mobility transistors (HEMT) under thermal stress. This degradation was found to be caused by F thermally diffusing and passivating the donors in the n-doped InAlAs layer.

3.1.1 Thermal Migration and Donor Passivation of Fluorine

Figure 3.1 shows a schematic diagram that describes the donor passivation mechanism [66]. F can thermally diffuse to Si:InAlAs as an interstitial impurity, capture a free electron released by the Si donor and form a F^- ion. The negatively charged F ions are attracted to the positively charged Si donors and form F-Si complexes. Thus, Si donors are effectively passivated and the carrier concentration in the InAlAs layer dramatically decreases with increasing annealing time or temperature.

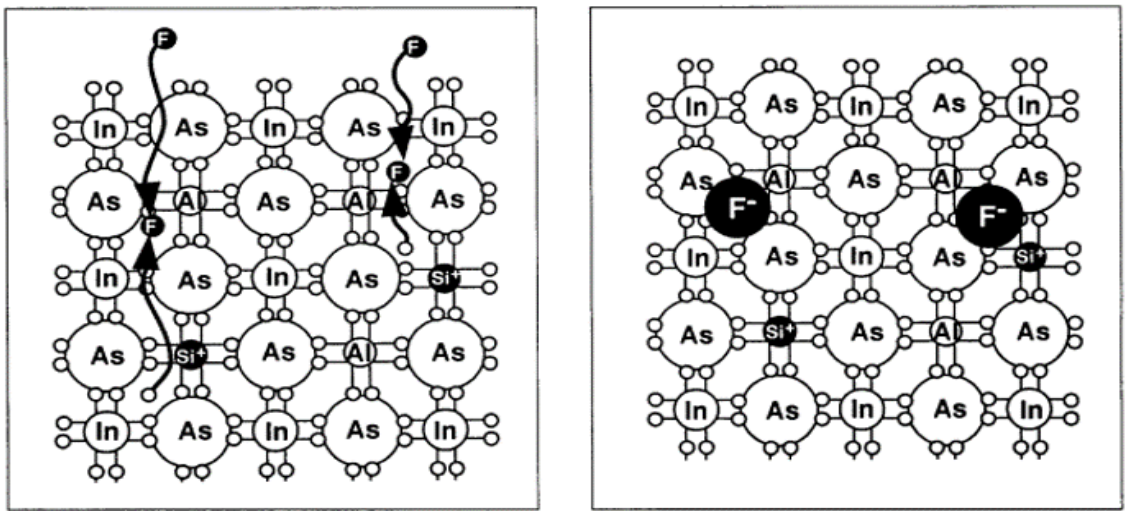

Figure 3.1: schematic diagram of donor passivation mechanism by fluorine in n-InAlAs layers: (left) thermal diffusion of fluorine and (right) negatively charged fluorine attracted to positively charged silicon, adapter from [66].

The effect of Fluorine in different III-V alloys has also been studied [67]. F only affects n-InAlAs, but not other alloys. This is because InAlAs has a strong tendency to ionize and the difference in the ionic radii between the III (In, Al) and V (As) groups creates enough interstitial space in the crystal for the fluorine to pass. In addition, the difference in the ionic radii between the different

group III materials (Al, In) creates dense and sparse regions for the localization of fluorine. Table 3.1 shows the ionic radii for Al, In and As.

	Al	In	As
ionic radius (nm)	0.049	0.081	0.211

Table 3.1: Ionic radii for Al, In and As, adapted from [68].

3.1.2 Fluorine-induced Thermal and Electrical Instability

The F-Si complex has been experimentally observed, through optical absorption measurements [67]. However, it is weakly bounded, and can easily dissociate, resulting in further thermal and electrical instability.

Figure 3.2 shows that as annealing time increases, carrier concentration decreases as a result of donor passivation by fluorine [68]. The effect can be reversed by re-annealing the sample at 400 °C under ultra-high vacuum. As the re-annealing time increases, F diffuses away, reactivating the donors and leading to higher carrier concentration.

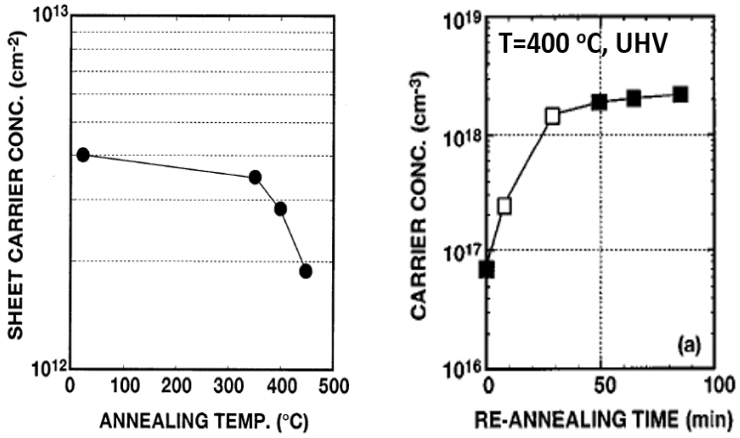

Figure 3.2: sheet career concentration vs annealing temperature (left) and re-annealing time under ultra-high vacuum at 400 °C (right) in n-doped InAlAs layer, adapted from [68].

In addition, under an applied electric field, F bound to Si can also dissociate and migrate [69]. Figure 3.3 shows the results from a bias temperature experiment where a voltage is applied across n-InAlAs structure. As time increases, the voltage between the anode and intermediate terminal increases. This indicates that F migrates preferentially towards the positive terminal under the electric field.

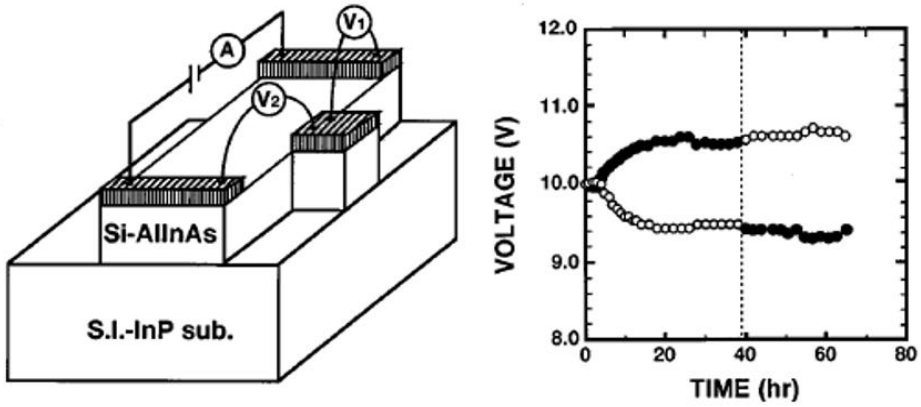

Figure 3.3: A constant voltage of 20 V is supplied across a Si-InAlAs/InP sample (0.6 × 2.7 mm^2) at 250 °C in nitrogen atmosphere. The voltage is measured between the anode and intermediate terminal (closed circles) and between cathode and intermediate terminal (open circles). The bias direction is reversed at the point of 39 h, adapted from [69].

3.2 Fluorine in InGaAs MOSFETs

3.2.1 Self-aligned InGaAs MOSFETs

The InGaAs MOSFETs used in this study were first reported in IEDM 2013 [29]. Cross sectional views of the device schematic and TEM image are depicted in Figure 3.4. The initial heterostructure is described in Figure 2.1(b). Devices are fabricated in a similar manner as described in Chapter 2, with an additional thermal annealing step at 340 °C for 15 min to repair

RIE damage prior to the gate stack deposition. The gate stack consists of 2.5 nm-thick HfO_2 and E-beam evaporated 30 nm-thick Mo. The final device has channel thickness of 7 nm and access length of 85 nm linking the channel and extrinsic portion of the device.

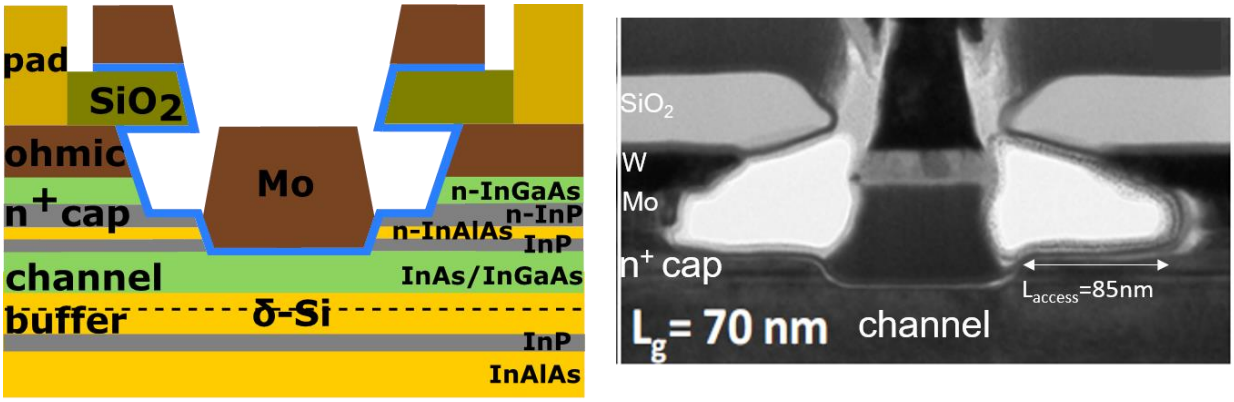

Figure 3.4: (left) Cross sectional schematic of the self-aligned InGaAs MOSFETs used in this study. (right) TEM cross section of a finished device of gate length L_g = 70 nm and access length L_{access} = 85 nm, adapted from [29].

3.2.2 Fluorine Introduced in Device Fabrication

Si doped InAlAs is present in the cap and access regions of the studied InGaAs MOSFETs. F can be introduced into the device during W/Mo contact etch which uses SF_6/O_2 RIE. Later, through

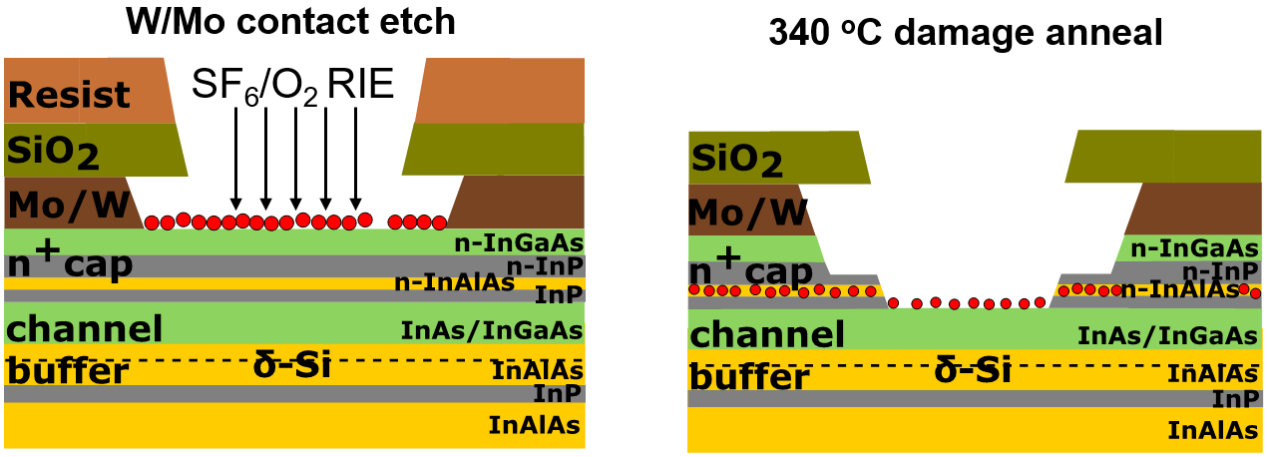

Figure 3.5: Fluorine can be introduced during the SF_6/O_2 reactive ion etching for the W/Mo contact etch (left), and can thermally diffuse into Si-InAlAs in the cap through the 340 °C damage annealing (right).

the 340 °C damage annealing step, F can thermally diffuse into Si-InAlAs in the cap. Thus, we expect F to concentrate under the gate and in the access regions, shown as the red circles in Figure 3.5. As a result, device might be unstable due to F migration.

3.3 Electrical Instability

To study the device instability under prolonged voltage stress, we have developed an automated stress-characterization measurement scheme. A constant bias is applied to the device and periodically interrupted for I_d-V_{gs} characterization. This allows us to study the evolution of device behavior during stress. In principle, the stress bias should be adjusted to maintain the same gate overdrive stress voltage throughout the stress experiment. However, in practice, threshold voltage changes less than 10% of the nominal gate overdrive stress voltage, which is negligible. Thus, the stress bias is kept the same throughout the stress experiment. After reaching the total stress time, the voltage stress is removed, all terminals are grounded and I_d-V_{gs} measurements are taken periodically to study the device recovery. A final 20 min, 70 °C thermal step is used to store the device and to assess if there is any permanent degradation. Key device parameters are also extracted, including linear threshold voltage ($V_{t,lin}$, defined at 1 μA/μm, V_{ds} = 0.05 V), peak transconductance ($g_{m,max}$ at V_{ds} = 0.5 V), minimum subthreshold swing (S_{min} at V_{ds} = 0.05 V), and ON-resistance (R_{on} at V_{gt} > 0.6 V, V_{ds} = 0.05 V).

3.3.1 Forward Gate Voltage Stress

We first perform the forward gate stress experiments, in which a positive gate overdrive voltage is applied ($V_{gt} = V_{gs} - V_t > 0$, V_{ds} = 0 V) followed by device recovery with all terminals grounded.

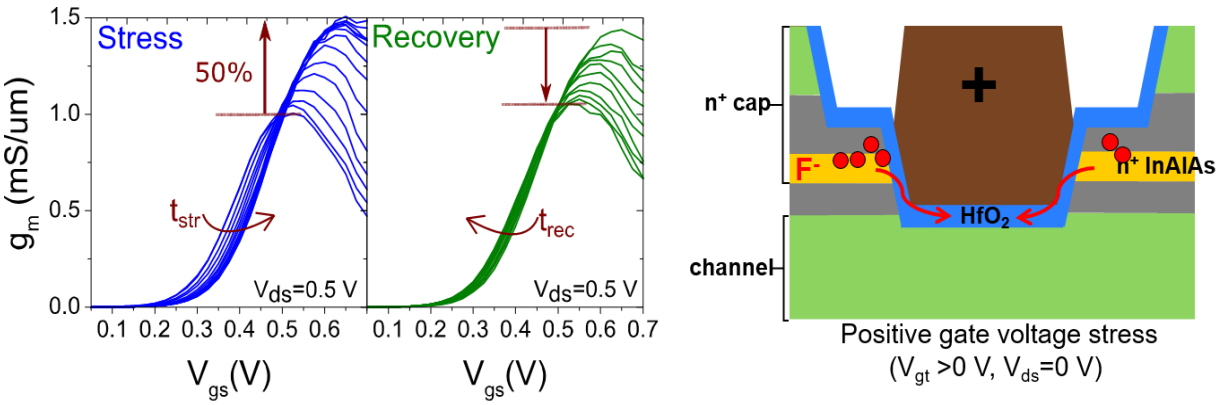

Figure 3.6: Left: evolution of g_m characteristics during forward gate stress (under V_{gt} = 0.8 V, V_{ds} = 0 V for 1.5 h at RT) and recovery (under V_{gs} = 0 V, V_{ds} = 0 V for 2 h at RT). For the time stamp for each line, refer to Figure 3.7. Right: illustration of fluorine ion migration due to forward gate bias with $V_{gt} > 0$.

Figure 3.6 shows the evolution of g_m characteristics of a typical device during 1.5 h stress at V_{gt} = 0.8 V and subsequent 2 h recovery, all at room temperature (RT). During stress, V_t shifts positively, similar to what is observed in PBTI experiments due to oxide trapping [48]–[50]. In contrast, rather than g_m degradation, there is a remarkable increase in $g_{m,max}$ of as much as 50%. We believe that this is due to migration of fluorine ions under electric field. During stress, the positively biased gate terminal attracts negatively charged F ions. F drifts from access regions into the gate oxide, reactivating Si dopants in n-InAlAs ledge. This leads to a higher electron concentration in the access region, decreased R_{on} and increased g_m. During the subsequent recovery where all the terminals are grounded, F diffuses back to n-InAlAs and re-passivates Si donors. This reverts the device back to the virgin state and leads to a near complete recovery of g_m and V_t.

Figure 3.7 shows the impact of different gate overdrive stress voltage ($V_{gt,stress}$) on $g_{m,max}$, $V_{t,lin}$, and R_{on}. For moderate stress time (< 10^3 sec), $g_{m,max}$ increases, $V_{t,lin}$ shifts positively, and R_{on} decreases. S changes negligibly. All changes are enhanced by stress voltage. After prolonged stress time (>

10^3 sec) at high stress voltage ($V_{gt,str}$ = 1 V), S_{min}, $g_{m,max}$ and R_{on} start to degrade due to a separate mechanism. During recovery, $g_{m,max}$, $V_{t,lin}$ S_{min} and R_{on} proceed to return to their initial values. After thermal annealing at 70 °C, all the devices reach complete recovery, except those subjected to harsh stress conditions ($V_{gt,str}$ = 1 V).

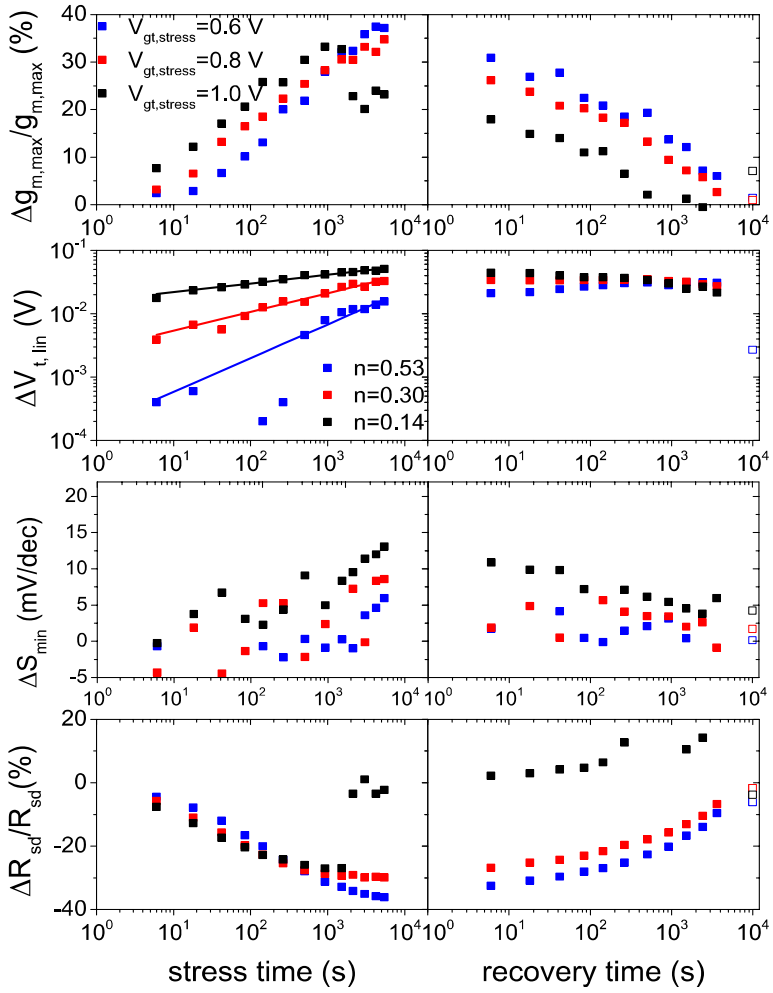

Figure 3.7: Evolution of $g_{m,max}$, $V_{t,lin}$, S_{min} and R_{on} during 2-h positive gate voltage stress at different V_{gt} with V_{ds} = 0 V at RT (left column) and during recovery where all terminals are grounded (right column). The hollow symbols represents data after thermal annealing at 70 °C. The fits to $V_{t,lin}$ represent a saturating power law characteristic of PBTI with the indicated time exponent *n*.

Figure 3.8 shows $\Delta g_{m,max}/g_{m,max,0}$ versus $\Delta V_{t,lin}$, at different stress voltages, observed in this experiment (left), compared with classic PBTI behavior (right [48]). We found the relation between the two parameters to be different under different stress conditions. This is in contrast to classic PBTI behavior where a universal relationship is observed between the two, across different stress voltages V_g. This confirms that PBTI alone cannot explain our observations.

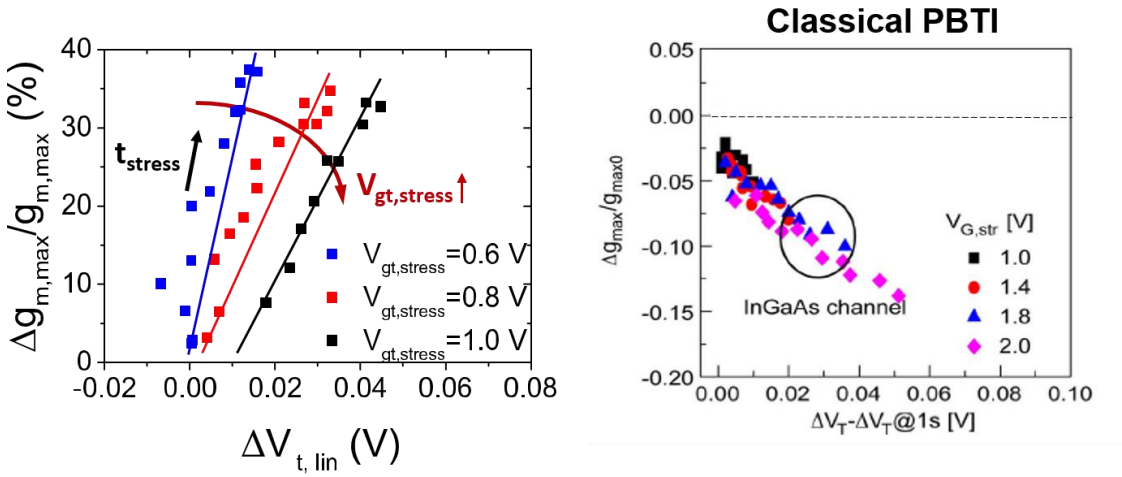

Figure 3.8: Correlation between $\Delta g_{m,max}/g_{m,max0}$ and $\Delta V_{t,lin}$, observed in this experiment (left), and in classical PBTI behavior, adapted from [48] (right).

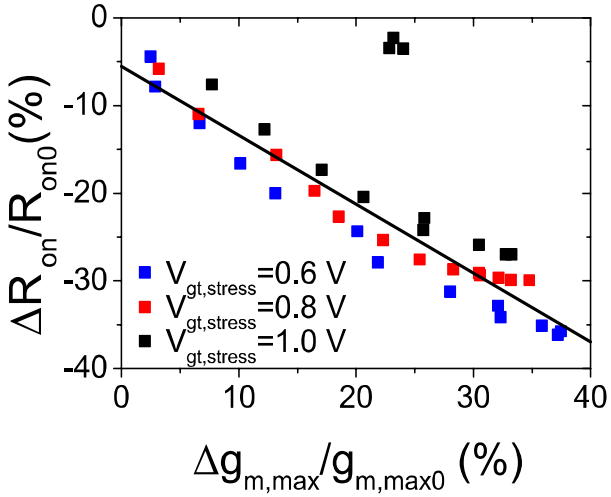

Figure 3.9: Correlation between $\Delta g_{m,max}/g_{m,max0}$ and $\Delta R_{on}/R_{on0}$ at different $V_{gt,stress}$.

In contrast, when we plot $\Delta g_{m,max}/g_{m,max,0}$ versus $\Delta R_{on}/R_{on0}$ at different stress voltage (Figure 3.9), we find that the two are inversely correlated under all but the harshest stress conditions. This suggests a strong connection between g_m instability and the extrinsic portion of the device, consistent with the fluorine hypothesis.

3.3.2 OFF-state Stress

We have also performed OFF-state stress experiments with high V_{ds} and zero V_{gt}. I_d-V_{gs} characteristics were collected in the normal mode and after swapping source and drain. Figure 3.10 shows g_m evolution during stress at V_{gt} = 0 V and V_{ds} = 0.7 V for 2 h at RT. As stress time increases, the lateral E field sweeps fluorine away from the source and gate oxide towards the drain. On the source side, Si dopants are reactivated and the access resistance is reduced, resulting in an increase in g_m in the forward mode. On the drain side, Si dopants are further passivated and the access resistance is increased, resulting in a decrease in g_m in the reversed mode where the source and drain terminals are reversed in the measurement. In both scenarios, V_t shifts negatively. This

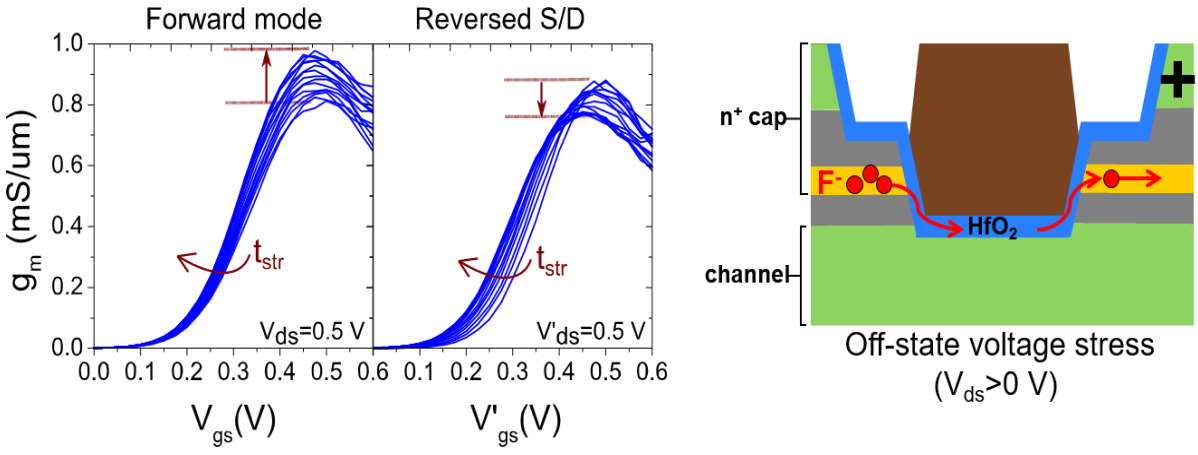

Figure 3.10: Evolution of g_m characteristics during OFF-state stress (under V_{ds} = 0.7 V, V_{gt} = 0 V for 2 h at RT), measured in normal configuration and with source and drain terminals reversed (left). For the time stamp for each line, refer to Figure 3.7. Illustration of fluorine ion migration due to OFF-state bias with V_{ds} > 0 (right).

confirms again a strong connection between g_m instability and the extrinsic portion of the device, inexplicable by established PBTI phenomena.

3.3.3 Temperature-dependent Voltage Stress

To study the temperature dependence of F-induced instability, we repeated the positive gate voltage stress (V_{gt} = 0.8 V, V_{ds} = 0 V for 2 h) and subsequent recovery (V_{gs} = 0 V, V_{ds} = 0 V for 1.5 h) under various temperature (-40 °C, -10 °C, 25 °C, and 50 °C).

In Figure 3.11, the left column shows the change in g_m and change in R_{on} as a function of stress time. Similar to what we have observed before, as stress time increases, g_m increases and R_{on} decreases. In addition, both the enhancement of g_m and reduction of R_{on} are accelerated with stress

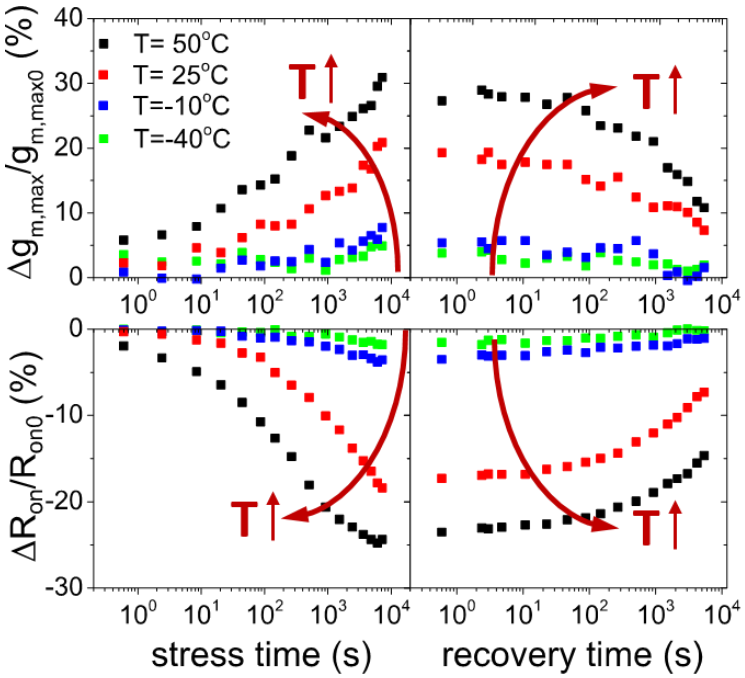

Figure 3.11: Evolution of $g_{m,max}$ and R_{on} during 2-h positive gate voltage stress (V_{gt} = 0.8 V, V_{ds} = 0 V) at different stress temperatures (-40 °C, -10 °C, 25 °C, and 50 °C), and subsequent 1.5-h recovery at the same temperatures.

temperature indicated by the red arrow. The right column shows the recovery of g_m and R_{on}, when the device is resting at the same stress temperature. Likewise, the recovery of g_m and R_{on} are enhanced by increasing temperature. This suggests F migration can be thermally activated.

From the Arrhenius plot of ΔR_{on}, shown in Figure 3.12, activation energy E_A can be extracted to be around 0.23 eV. This is consistent with the estimation of F-Si ionization energy from the literature [70].

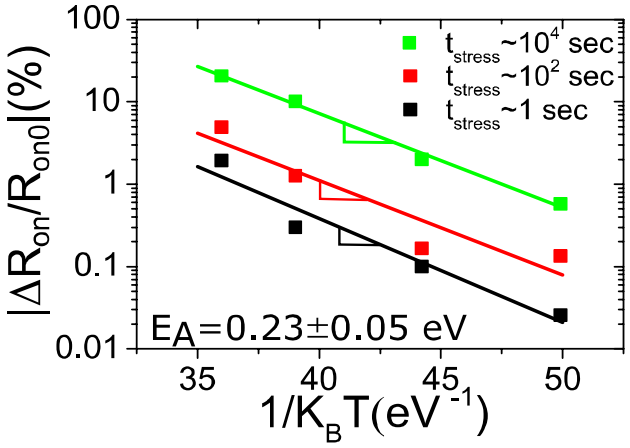

Figure 3.12: Arrhenius plot of $\Delta R_{on}/R_{on0}$ measured during stress. Activation energy E_A is extracted to be 0.23 ± 0.05 eV.

3.4 Independent Verification

To independently confirm that F is introduced during MOSFET device processing and causes donor passivation, we have carried out secondary ion mass spectroscopy (SIMS) and transmission line method (TLM) measurements.

3.4.1 Secondary Ion Mass Spectroscopy (SIMS)

To verify F in our devices using SIMS measurements, we have prepared three samples of a

heterostructure containing a 3 nm-thick n-InAlAs layer. Sample A is never exposed to F. Sample B is exposed to Mo sputtering and F-based RIE. Sample C is further annealed at 350 °C for 1 min. SIMS analysis of these samples is shown in Figure 3.13. Compared to A, only samples B and C which have been exposed to F-RIE show a high surface concentration of F. Sample C, which has been further thermally annealed shows an additional pile-up of F in the n-InAlAs layer that integrates to a concentration of 5.3×10^{14} cm^{-2}. This is much higher than in earlier studies in HEMTs where F is introduced through HF treatments [66], [68], instead of F-RIE, as done here. SIMS also shows that F does not reach the Si δ-doped layer beneath the channel. From these results, we expect a similar F pile up in the n-InAlAs layer of the MOSFET access region (Figure 3.5).

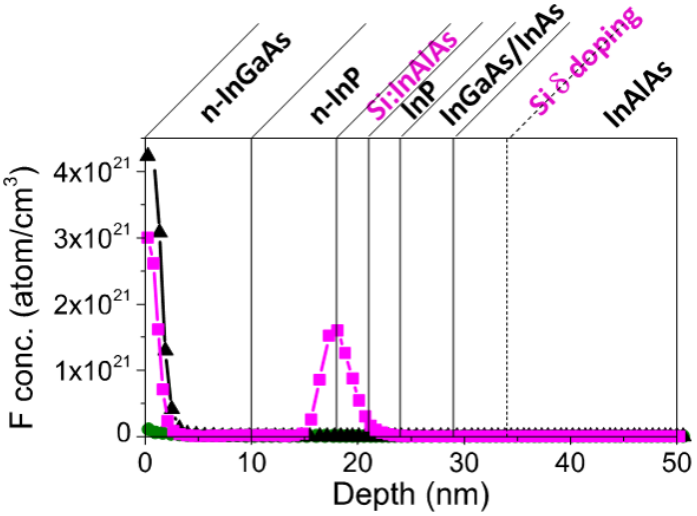

Figure 3.13: Fluorine concentration versus depth from SIMS for samples A (never exposed to F), B (F-based RIE) and C (F-based RIE + 350 °C anneal for 1 min). The layers are labeled on the top. A prominent F peak in the Si-doped InAlAs layer is observed in sample C.

3.4.2 Transmission Line Method (TLM) Measurements

To verify F-induced donor passivation, we have prepared two types of TLMs on a sample with an n-InGaAs/n-InP/n-InAlAs (15/3/3 nm) cap, shown in Figure 3.14(a). The first type uses sputtered

Mo contacts etched by SF_6/O_2 RIE (as in the MOSFET process). The second type uses lift-off Mo contacts, and hence has no exposure to F. Both TLMs are measured before and after annealing at 350 °C for 1 min. Total resistance is plotted as a function of contact spacing in Figure 3.14(b) and the semiconductor sheet resistance (R_{sh}) can be extracted from the slope. Before annealing, the F-RIE sample exhibits a R_{sh} of 176 Ω/□, 30% higher than that of the lift-off sample. After annealing, R_{sh} of the lift-off sample decreased somehow, while R_{sh} of the F RIE sample increased by 3x. This result verifies F induced donor passivation in our devices due to the fabrication process. It is also consistent with literature findings that a thermal step is needed for F to migrate to Si-doped InAlAs and passivate the Si donors [8-10].

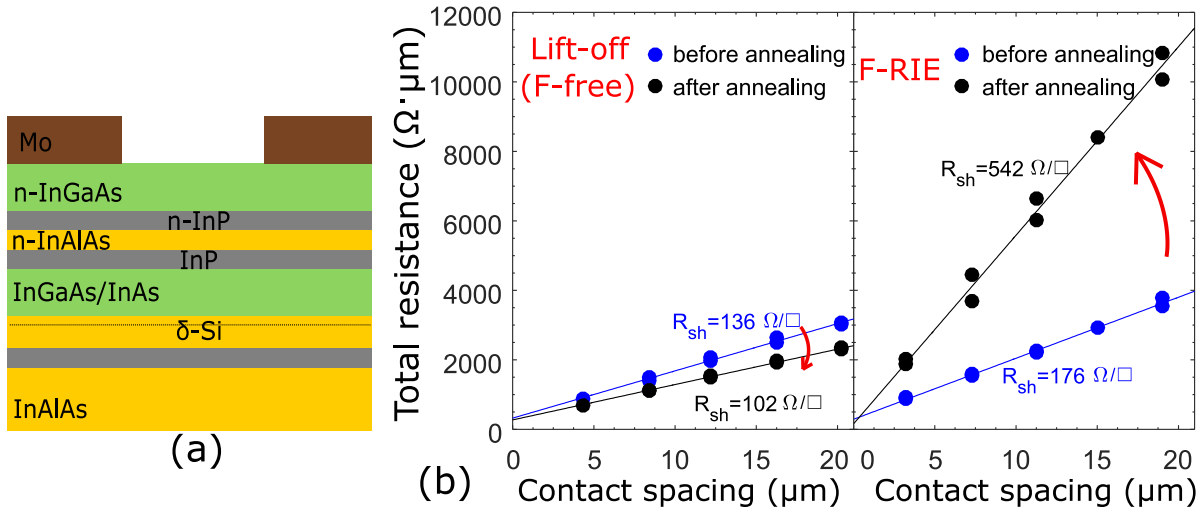

Figure 3.14: (a) Schematic of TLM structure. (b) TLM resistance versus contact distance before and after annealing on (left) samples fabricated by lift-off (no F exposure), and (right) samples that have undergone F-based RIE.

3.5 Mitigation

Now that we have independently verified the presence of fluorine in our devices and its impact on donor passivation and device instability, it is important to identify ways to mitigate this problem.

We can either eliminate F in our device processing or create new device structure that is immune to the impact of F. Considering that F is a key element in semiconductor manufacturing, mitigation of F induced donor passivation necessitates the elimination of n-InAlAs from the device structure. Despite its wide use in InGaAs FETs, this material is not essential.

3.5.1 New Device Architecture

We have designed and fabricated an alternative MOSFET structure that uses n-InP in the access region, rather than the conventional n-InAlAs (Figure 3.15). The original and the new designs share the identical channel design, gate stack and top contact layer. To further boost the performance, a shorter access region of 15 nm is fabricated to reduce the access resistance. Without the n-InAlAs layer, we expect the device will not suffer from F-induced donor passivation and associated instability.

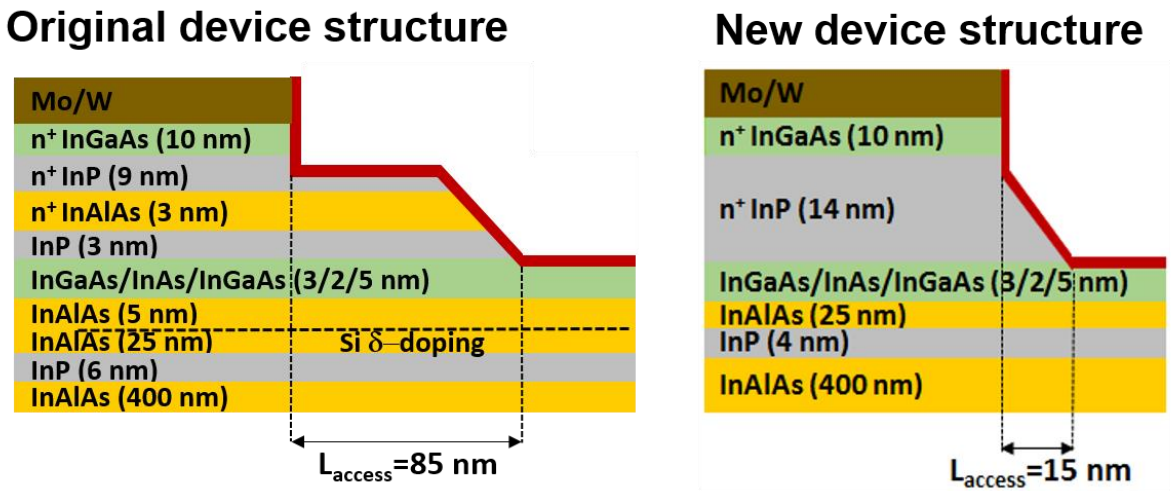

Figure 3.15: Design of the original (left) and new (right) device structure. InAlAs layers in new device structure are undoped.

3.5.2 Improved Device Instability

To compare the device instability of the new device structure with the original device structure, we repeat the same stress experiments on the new sample. We indeed find the new device design to be much more stable.

Under forward gate stress (V_{gt} = 0.8 V and V_{ds} = 0 V for 2 h), shown in Figure 3.16, the new design shows a small positive V_t shift and a $g_{m,max}$ decrease (up to 15%), as opposed to g_m enhancement. This is largely recovered after thermal detrapping at 70 °C. This is classic PBTI behavior attributed to electron trapping in the oxide near the semiconductor-oxide interface [48].

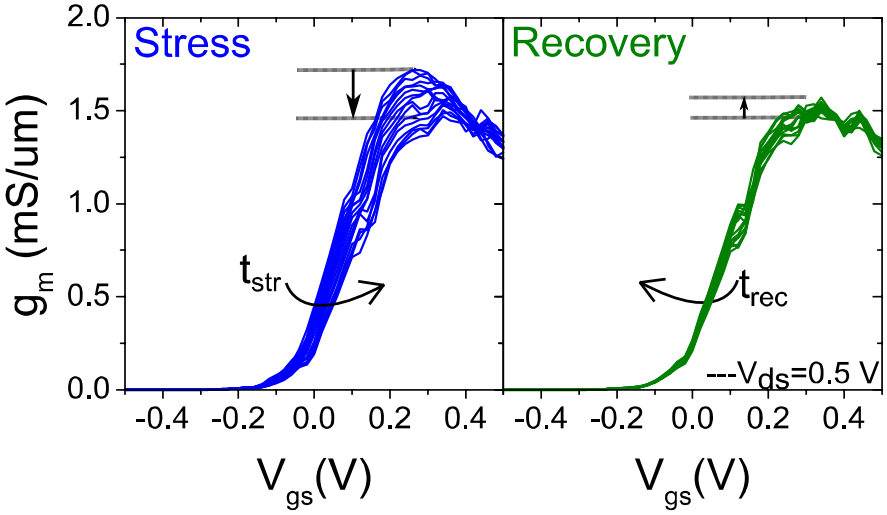

Figure 3.16: Evolution of g_m characteristics (left) during gate stress at V_{gt} = 0.8 V and V_{ds} = 0 V for 2 h and (right) during 1.5-h recovery at RT.

Under OFF-state stress (V_{gt} = 0 V and V_{ds} = 0.7 V for 2 h), shown in Figure 3.17, the new sample shows minimal changes in both the normal measurement configuration and reversed Source/Drain measurement configuration, again confirming the absence of F-induced instability.

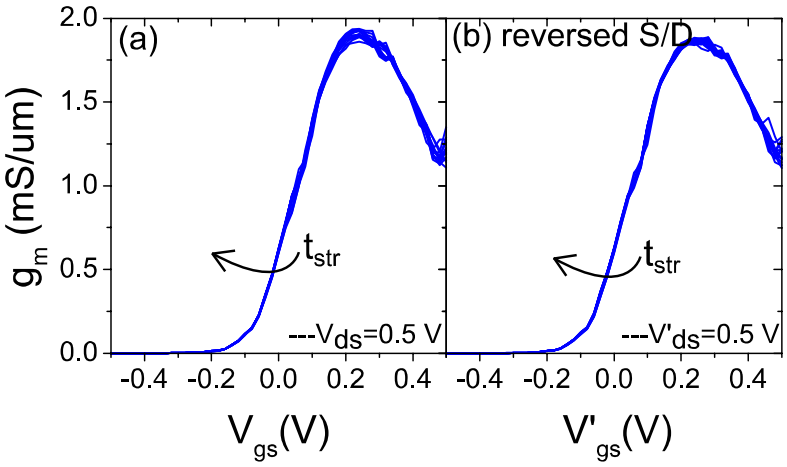

Figure 3.17: Evolution of g_m characteristics during OFF-state stress for 2 h at RT under V_{gt} = 0 V, V_{ds} = 0.7 V, (a) measured in normal configuration, and (b) with source and drain terminals reversed.

Figure 3.18 shows $\Delta V_{t,lin}$ in both samples during forward gate stress (V_{gt} = 0.8 V and V_{ds} = 0 V for 2 h) at different stress temperature. The original sample which suffers from F migration shows a prominent temperature-enhanced V_t shift. In contrast, the new sample is much more stable with temperature, showing only a weak temperature dependence.

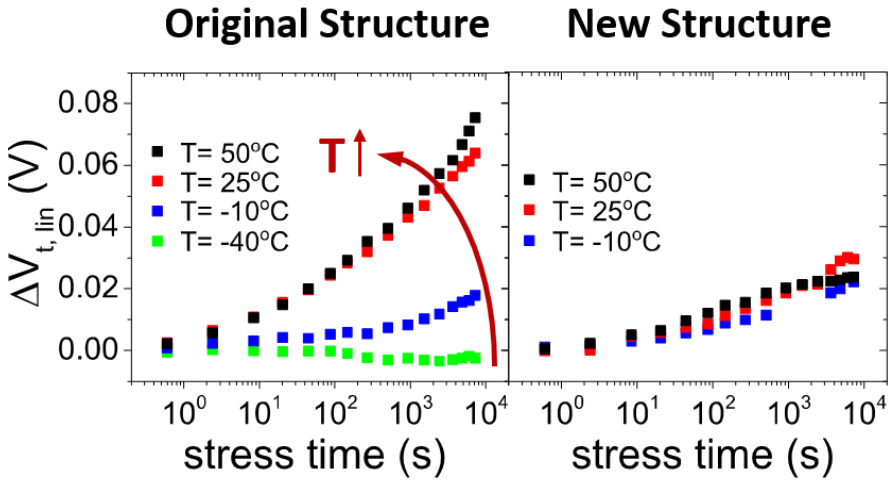

Figure 3.18: Comparison of temperature dependence in $\Delta V_{t,lin}$ evolution during 2-h positive gate voltage stress (V_{gt} = 0.8 V, V_{ds} = 0 V) at different stress temperatures between (left) original sample and (right) new sample.

3.5.3 Record Performance

In addition to the improved electrical stability, the new device design also achieves record performance. The absence of Si:InAlAs in the n^+ ledge grants immunity to F passivation, resulting in a higher electron concentration in the access regions in virgin devices. This has led to a record low ON-resistance R_{on} = 190 Ω–µm and record high transconductance of $g_{m,max}$ = 3.45 mS/µm in a device with 70 nm gate length [71]. Figure 3.19 shows its output, transfer and g_m characteristics.

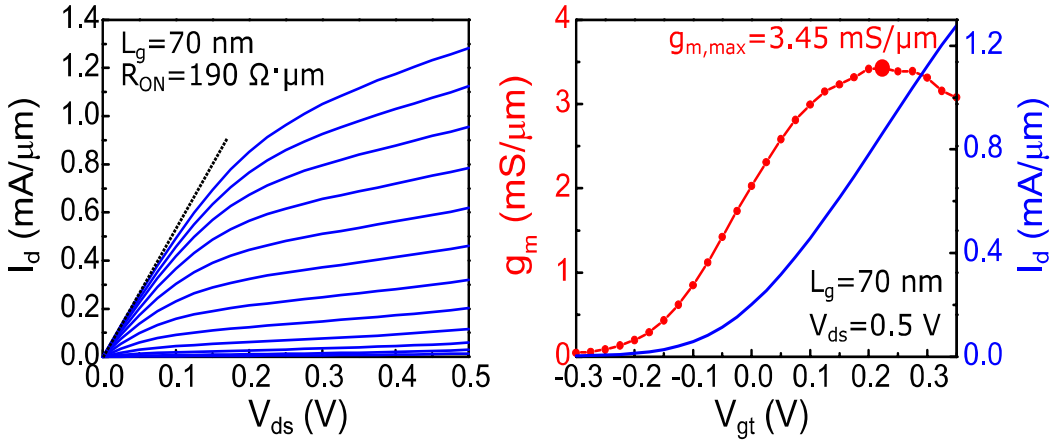

Figure 3.19: L_g = 70 nm InGaAs MOSFET from new sample (left) output I_d-V_{ds} with V_{gt} = -0.2 to 0.35 V in 0.05 V steps and (right) g_m and transfer characteristics at V_{ds} = 0.5 V. $g_{m,max}$ = 3.45 mS/µm, adapted from [71].

3.6 Summary

In this chapter, we have identified a new instability mechanism in self-aligned InGaAs MOSFETs caused by F^- migration and (de)passivation of Si dopants in n-InAlAs. Positive gate voltage stress leads to reversible g_m enhancement, contrary to the well-established PBTI phenomena. We have

successfully mitigated this problem by eliminating n-InAlAs from the device structure. The new device design not only shows improved stability, but also achieves record device performance.

In chapter 4, we will use the new device design that is free from the impact of fluorine to study the impact of oxide trapping in InGaAs MOSFETs.

CHAPTER 4. Impact of Oxide Trapping

4.1 Introduction

Chapter 3 has discussed the anomalous instability due to fluorine contamination in planar InGaAs MOSFETs, peculiar to devices that contain an n-InAlAs layer. A new device structure has been proposed and has successfully mitigated this issue.

With the new device structure, this chapter will investigate the impact of oxide trapping in InGaAs MOSFETs. First, we study its impact on device reliability in InGaAs MOSFETs, namely positive bias temperature instability. Then, we investigate its impact on device performance. A combination of pulsed IV, lock-in and *S*-parameter measurements is used to study the dynamic device performance. Finally, we show that oxide trapping also significantly complicates the extraction of fundamental transport parameters such as mobility.

4.2 Positive Bias Temperature Instability (PBTI) in InGaAs MOSFETs

Reliable device operation is a critical concern for InGaAs MOSFETs, due to the large number of defects in the oxide and at the semiconductor-oxide interface. Furthermore, as EOT scales down in advanced MOSFETs, the electric field across the oxide during device operation continues to increase, accelerating Positive Bias Temperature Instability (PBTI). Here, we study PBTI in InGaAs MOSFETs with HfO_2 gate dielectric of EOT~0.5 nm.

The device structure and initial heterostructure have been shown previously in Figure 2.2 and Figure 2.1(c), respectively. The gate stack consists of 2.5 nm thick HfO_2 deposited by ALD at 250 °C and 35 nm thick E-beam evaporated Mo. The final channel consists of 2 nm InAs on top of 5 nm $In_{0.7}Ga_{0.3}As$. Devices used in this study have 200 nm gate length. Virgin devices exhibit $V_{t,lin}$ ~ -0.1 V, $g_{m,max}$ ~ 1.7 mS/µm, S_{min} ~ 93 mV/dec. We use the same stress-characterization scheme as discussed in Section 3.3.

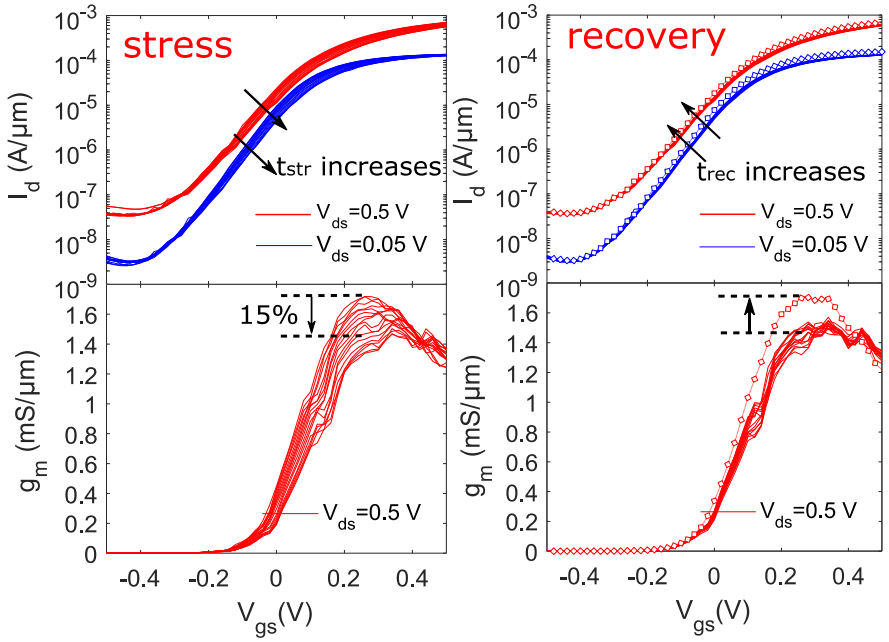

Figure 4.1: Evolution of subthreshold and transconductance characteristics of a typical device (L_g=200 nm) during stress (left) and recovery (right). Stress condition is V_{gt} = 0.7 V, V_{ds} = 0 V at room temperature. For the time stamp for each line, refer to Figure 4.2. Hollow symbols represent device characteristics after thermal step at 70 °C for 20 min.

Figure 4.1 shows the evolution of subthreshold and transconductance characteristics of a typical device during moderate stress (V_{gt} = 0.7 V, V_{ds} = 0 V) and during recovery. During stress, V_t shifts positively, and g_m and S_{min} degrade. During recovery, the device characteristics move in the

direction of restoration. After a thermal detrapping step (hollow symbols), the device reaches full recovery. This is consistent with electron trapping in pre-existing oxide traps.

4.2.1 Gate Voltage Dependence

We investigate the impact of gate stress voltage by repeating the experiment with different gate overdrive stress voltages (V_{gt} ~ 0.6 V, 0.8 V, 1V, 1.1 V). Figure 4.2 shows the evolution of $g_{m,max}$, $V_{t,lin}$, and S_{min} during stress and recovery. V_t shifts and g_m degradation are enhanced by stress voltage and stress time. As the gate overdrive stress voltage increases, the electric field across the

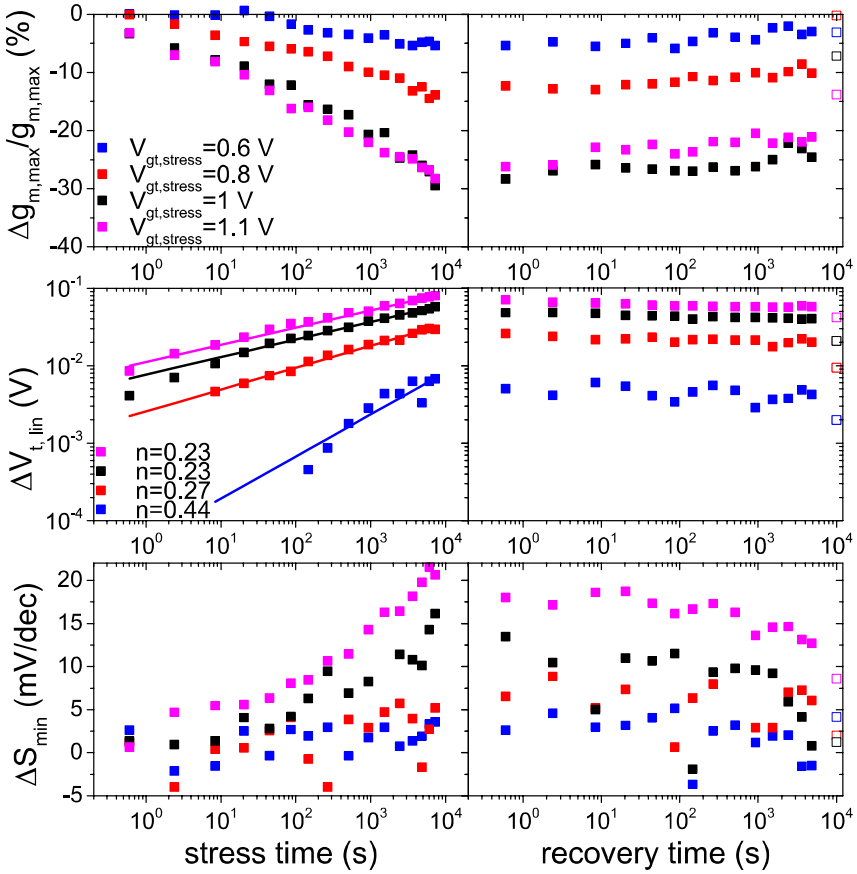

Figure 4.2: Evolution of $g_{m,max}$, $V_{t,lin}$ and S_{min} during 2-h positive gate voltage stress at different V_{gt} with V_{ds} = 0 V at RT (left column) and during 1.5-h recovery where all terminals are grounded (right column). The hollow symbols represents data after thermal annealing at 70 °C. The fits to $V_{t,lin}$ represent a saturating power law characteristic of PBTI with the indicated time exponent *n*.

oxide increases, allowing the carriers to interact with a wider range of defect levels [43]. As stress time increases, defect levels further away from the oxide/semiconductor interface into the oxide are filled. We also observe minimal change in S_{min} for moderate stress voltages ($V_{gt,stress}$ = 0.6 and 0.8 V), but a clear degradation at harsher stress conditions ($V_{gt,stress}$ = 1 and 1.1 V). After the stress is removed, V_t, g_m, and S_{min} all start to recover. For the 1.1 V stress, after the thermal detrapping step, we still observe residual positive $\Delta V_{t,lin}$ and degraded S_{min}. This suggests generation of interface states in the oxide, a form of permanent damage to the device. For other stress conditions, the recovery becomes nearly complete.

The threshold voltage shift ΔV_t has a strong dependence on stress time (t_{str}), stress voltage ($V_{gt,stress}$) and temperature (T). It can be approximated by

$$\Delta V_t \approx A_0 V_{gt,stress}{}^{\gamma} t_{str}{}^{n} \exp\left(-\frac{E_A}{KT}\right) \tag{4.1}$$

where A, n, γ and E_A are the prefactor, time exponent, voltage exponent, and activation energy, respectively [49], [50]. In Figure 4.3, we plot $\Delta V_{t,lin}$ vs. t_{str} and $V_{gt,stress}$ on a log-log scale, and find that $\Delta V_{t,lin}$ in InGaAs MOSFETs does follow the expected power law dependence on stress time

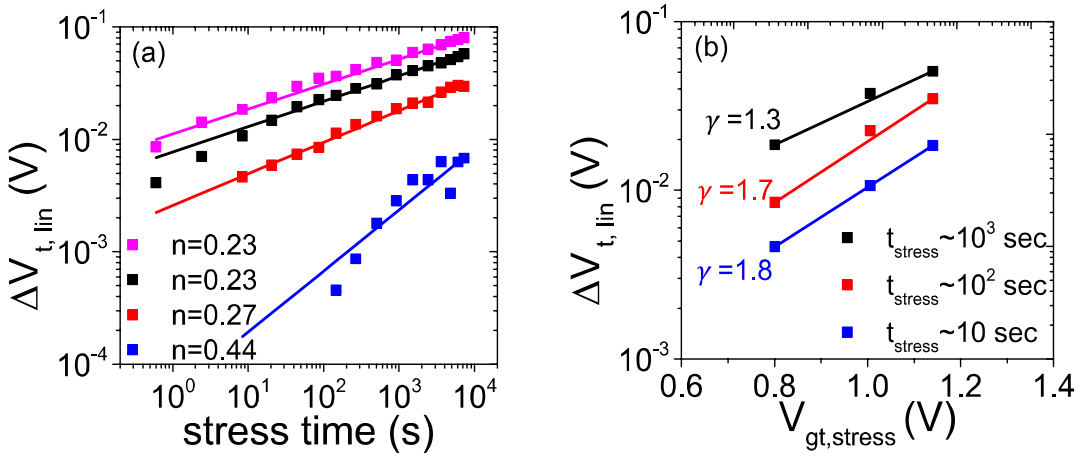

Figure 4.3: (a) $\Delta V_{t,lin}$ vs. stress time, (b) $\Delta V_{t,lin}$ vs. $V_{gt,stress}$. Time exponent n is extracted between 0.23 and 0.44, and voltage exponent γ is extracted between 1.3 and 1.8.

and stress voltage. The extracted time exponent n is between 0.23 and 0.44, and the voltage exponent γ is between 1.3 and 1.8. These values are similar to prior findings in InGaAs MOSFET PBTI studies [49]. Compared to Si ($\gamma \sim 6–7$), InGaAs MOSFETs have a much smaller γ value [50], exhibiting significant ΔV_t at low stress voltage.

We can predict the time-to-failure (TTF) of these devices at different gate stress voltages. Figure 4.4 shows the extrapolated time to 30 mV V_t shift at room temperature as a function of gate overdrive voltages. For 10-year reliable operation, the operating gate overdrive voltages cannot exceeds 0.4 V.

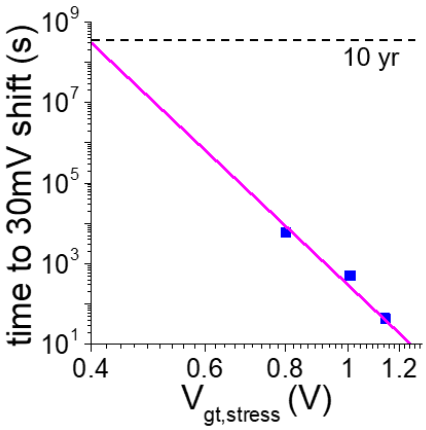

Figure 4.4: Extrapolated time to ΔV_t = 30 mV at various gate overdrive voltages.

We have also found a strong correlation between $\Delta g_{max}/g_{max0}$ and $\Delta V_{t,lin}$ for most of the stress conditions, shown in Figure 4.5 (a). This confirms that majority of charge trapping occurs near the III-V/oxide interface, affecting both ΔV_t and g_m. For harsh stress conditions ($V_{gs,stress}$ =1.1 V, t_{str} > 1000 s), ΔV_{tlin} is higher at a given $\Delta g_{max}/g_{max0}$. This indicates charge trapping also occurs inside the high-k dielectric, contributing additional $\Delta V_{t,lin}$ but having negligible impact on transport. Furthermore, we have found a consistent correlation between $\Delta V_{t,lin}$ after 2-h stress and $\Delta V_{t,lin}$ after

1.5-h recovery at different $V_{gt,stress}$ (Figure 4.5(b)), a trend observed in other PBTI studies [48]. The recovery is independent of the stress condition, but proportional to the amount of trapped charge.

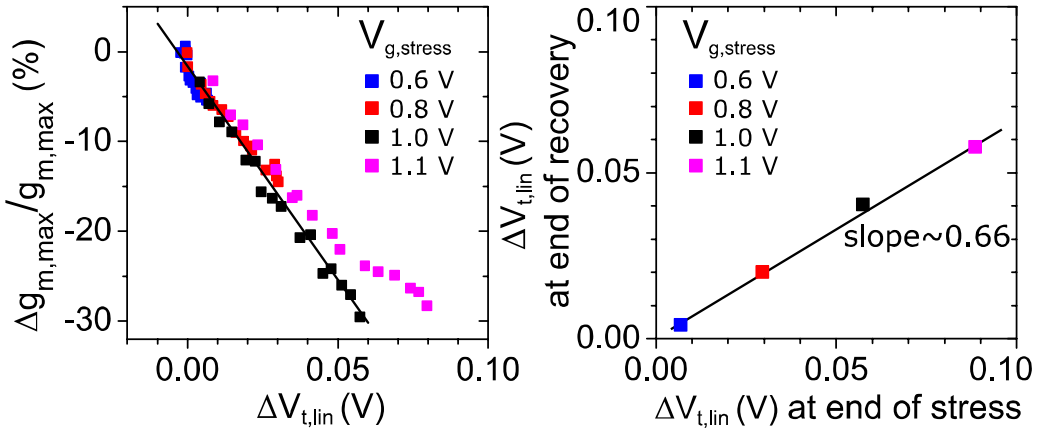

Figure 4.5: (left) strong correlation between $\Delta g_{max}/g_{max0}$ and $\Delta V_{t,lin}$ at different $V_{gt,stress}$ and (right) trend of $\Delta V_{t,lin}$ after 1.5-h recovery vs $\Delta V_{t,lin}$ after 2-h stress at different $V_{gt,stress}$, similar to other PBTI studies.

4.2.2 Temperature Dependence

We also study the role of stress temperature in PBTI by repeating the positive gate voltage stress (V_{gt} = 0.8 V, V_{ds} = 0 V for 2 h) and subsequent recovery (V_{gt} = 0 V, V_{ds} = 0 V for 1.5 h) under various temperatures (-10 °C, 25 °C, and 50 °C).

Figure 4.6 shows the change in g_{max}, $V_{t,lin}$ and S_{min} as a function of stress and recovery time at different temperatures. Device degradation is evidently enhanced by stress temperature. At a high stress temperature of 50 °C, Δg_{max}, ΔS_{min} and $\Delta V_{t,lin}$ are more significant. In comparison, at lower stress temperatures (-10 °C and 25 °C), the degradation is smaller and there is minimal change in S_{min}. During recovery where the device is resting at the same stress temperature, the recovery rate seems to be similar at various temperatures or slightly higher at 50 °C. After thermal detrapping,

the device stressed at 50 °C exhibits permanent damage, as g_{max}, S_{min} and $V_{t,lin}$ only partially recover.

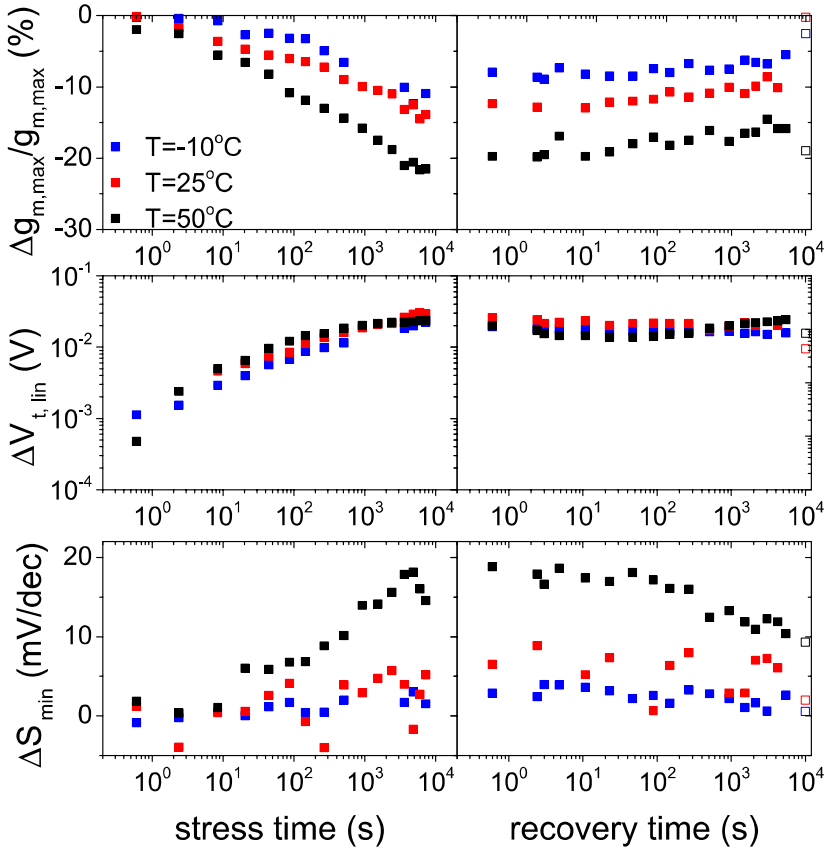

Figure 4.6: Evolution of g_{max}, $V_{t,lin}$ and S_{min} during 2-h positive gate voltage stress at different stress temperatures (T = -10 °C, 25 °C, 50 °C) with $V_{gt,str}$ = 0.8 V, V_{ds} = 0 V, and during subsequent 1.5-h recovery at the same stress temperatures.

Figure 4.7 shows the Arrhenius plot of $\Delta V_{t,lin}$, where $\log(\Delta V_{t,lin})$ is plotted as a function of $1/k_BT$ at different stress times. From the slope, the activation energy E_A is extracted to be 0.062 ± 0.004 eV. This is slightly lower than the other reported values in the literature: 0.08-0.13 eV [49], [50], likely due to the quantum-well nature of our device design that reduces the defect barrier offset [49]. The weak temperature dependence is characteristics of border traps that communicate with the channel through tunneling.

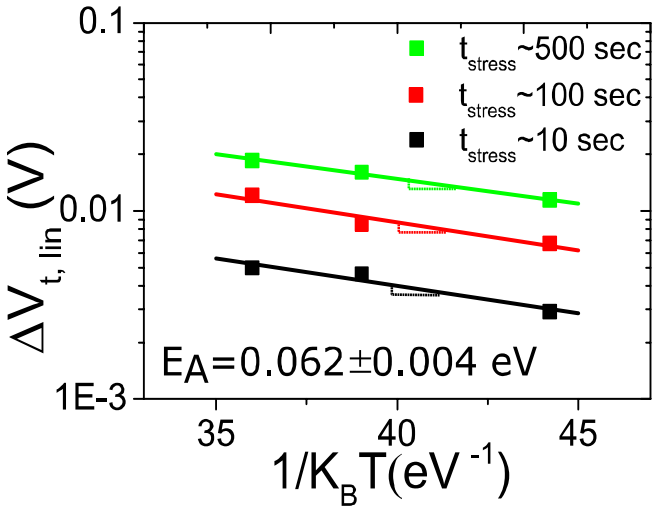

Figure 4.7: Arrhenius plot of $\Delta V_{t,lin}$ measured during stress. Activation energy E_A is extracted to be 0.062 ± 0.004 eV.

4.3 Deteriorated Device Performance

From the study of PBTI, we have found that oxide trapping after prolonged voltage stress (> 10^{-1} s) can induce a threshold voltage shift and degrade device performance. Due to the measurement delay in the sense-stress-sense techniques used to study PBTI, degradation on time scales < 10^{-1} s due to fast trapping cannot be captured. Various reports have shown fast electron trapping in defects close to the semiconductor-oxide interface can also lead to degraded DC device performance and severe frequency dispersion in InGaAs MOSFETs [51], [72]–[74]. To understand the impact of oxide traps on device performance, we study the dynamic performance of InGaAs MOSFETs.

The InGaAs MOSFETs used in this study have an $In_{0.7}Ga_{0.3}As$ intrinsic channel. The initial heterostructure is described in Figure 2.1(a). Devices are fabricated in a similar manner as described in Chapter 2. Different channel thickness (nominal values of t_c = 8, 6, 4, 2 nm) are

obtained via a highly reproducible timed dry etch, and used to study the t_c dependence on the impact of oxide trapping. A high-quality III-V surface with a roughness of ~0.5 nm is achieved prior to gate stack formation, shown in Figure 4.8. The gate stack consists of 2.5 nm-thick HfO_2 deposited by ALD at 250 °C and E-beam evaporated 30 nm-thick Mo.

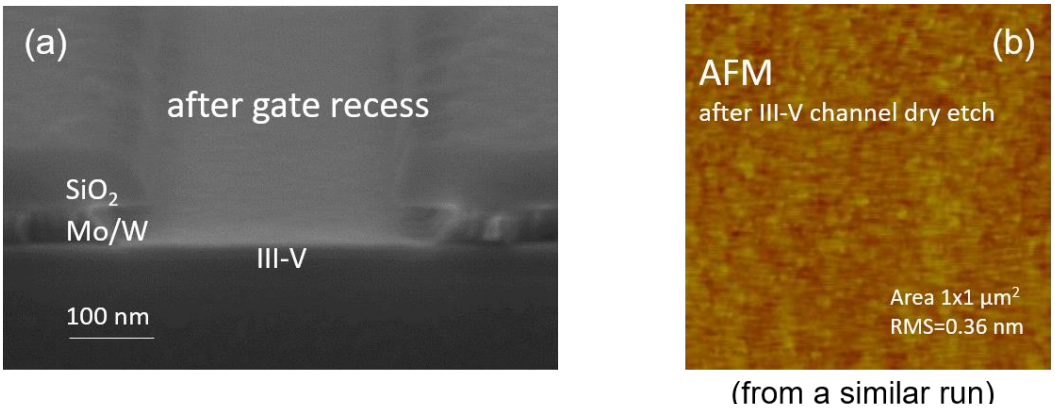

Figure 4.8: (a) SEM cross-section after pre-III-V recess. (b) AFM after III-V dry etch showing a surface roughness of ~0.5 nm.

4.3.1 Measurement Techniques

In this section, we describes the three measurement techniques used to probe the dynamics of oxide traps and study their impact on device performance.

4.3.1.1 Pulsed IV Measurement

The dynamics of oxide trapping can be probed using time-dependent pulsed measurement. Pulsed and DC I_d-V_{gs} measurements are carried out using an Auriga Microwave AU4750 Pulse I/V system. In pulsed IV measurements, a train of short voltage pulses is applied to the gate, as shown in Figure 4.9; we use pulse widths from 100 μs down to 1 μs. The drain voltage bias can be applied either

continuously or in a pulsed manner, without affecting the measured I_d. The limitation of this method is the available pulse widths in the instrument, which cannot be shorter than 200 ns.

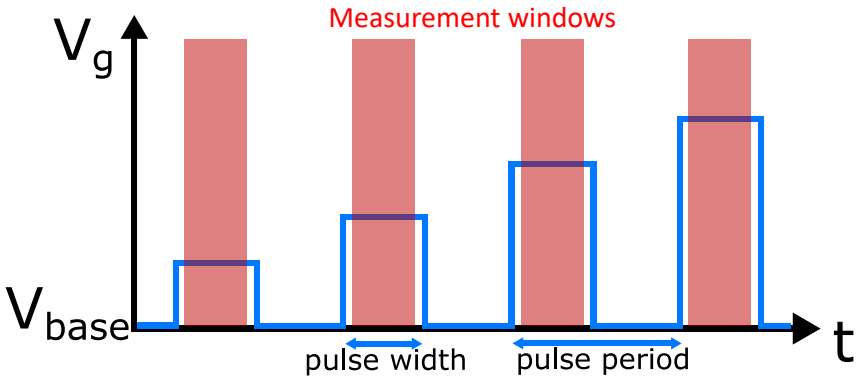

Figure 4.9: Illustration of a pulse-train applied to the gate. Pulse width ranges from 100 μs to 1 μs.

4.3.1.2 AC Transconductance Measurement

The dynamics of oxide trapping can also be probed using frequency-dependent measurements. Frequency dispersion in transconductance (g_m) reflects the distribution of capture/emission time constants in oxide traps. Lock-in and *S*-parameter measurements are used to measure AC g_m over the frequency ranges from 25 KHz to 100 MHz and 100 MHz to 50 GHz, respectively. The devices used here have 50 Ω ground-signal-ground (GSG) pad configuration (shown in Figure 4.10), which resembles a coplanar waveguide and provides high frequency response.

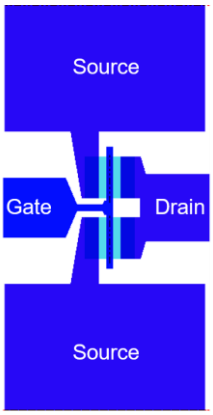

Figure 4.10: Ground-signal-ground (GSG) pad layout.

Lock-in

Figure 4.11 shows the diagram of the lock-in measurement setup covering a frequency range from 25 KHz to 100 MHz. A small AC sinusoidal signal (v_{gs}) superimposed on a DC bias (V_{GS}) is applied to the gate using a function generator (Tektronix AFG 3102). A semiconductor parameter analyzer (Keysight B1500) is used to apply the DC drain voltage (V_{DS}) and measure the DC current (I_D). The AC current (i_d) is amplified and converted to an AC voltage through a transimpedance amplifier (Femto Messtechnik DHPCA-100), and subsequently measured through a lock-in amplifier (Stanford Research SR844).

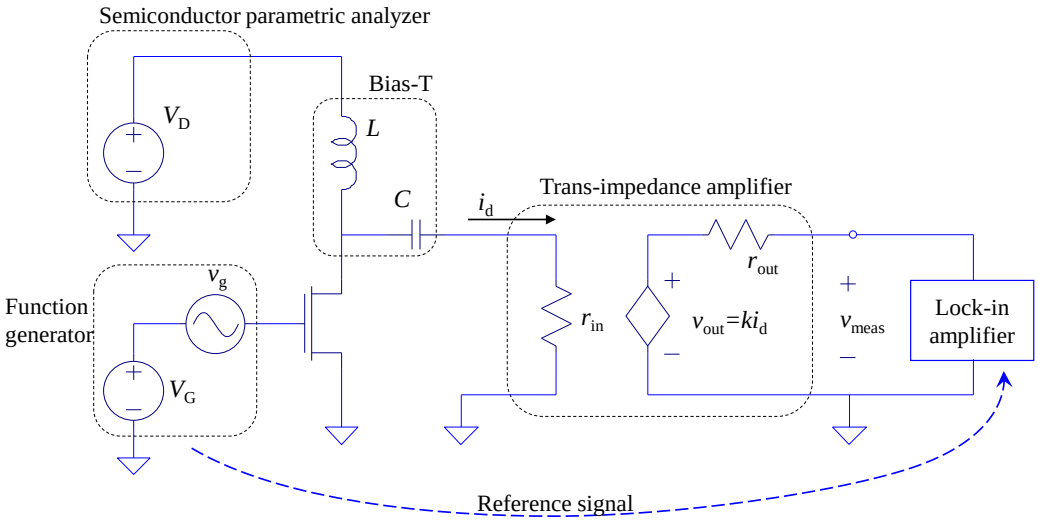

Figure 4.11: Diagram of the lock-in measurement setup

As many instruments have varying frequency response, the transfer function needs to be characterized first. We use the function generator to drive an off-wafer through structure with a known 50 Ω resistance and characterize the transfer function $H(f)$ according to

$$H(f) = \frac{v_{\text{meas}}(f)}{v_{gs} \times \frac{1}{50\Omega} \times k} \tag{4.2}$$

where k is the gain of the transimpedance amplifier (units of V/A). Then, the device under test

(DUT) is measured. The AC g_m can be calculated as the ratio of the AC drain current (i_d) over the AC gate voltage (v_{gs}), according to

$$g_m(f) = \frac{1}{H(f)} \frac{i_d(f)}{v_{gs}} = \frac{1}{H(f)} \frac{v_{\text{meas}}(f)}{k\, v_{gs}} \tag{4.3}$$

Intrinsic transconductance ($g_{m,i}$) can be subsequently calculated considering the source and drain resistance, output conductance and input impedance of the trans-impedance amplifier [75]. For device operation in saturation, the effect of output conductance is negligible, and $g_{m,i}$ can be calculated using

$$g_{m,i} = \frac{g_m}{1 - R_s g_m} \tag{4.4}$$

S-parameter[1]

S-parameter measurements are carried out using an Agilent N5230A network analyzer covering 100 MHz to 50 GHz. A semiconductor parameter analyzer (Keysight B1500) is used to apply the DC biases (V_{DS}, V_{GS}) through bias tees and to measure the DC currents (I_D, I_G). The network analyzer is calibrated using an off-wafer 2-port line-reflect-reflect-match (LRRM) calibration procedure with an impedance standard substrate (Microtech Cascade 101-190 C) to remove the parasitics due to the cables and probe tips. We explain here the high-level procedure used to extract intrinsic circuit elements; mathematical details are deferred to appendix B.

A transistor consists of three nested components:

[1] This work is done in collaboration with Professor Jesús Grajal from Universidad Politécnica de Madrid.

- The core is the *intrinsic device*, which is modeled with bias-dependent components.
- Surrounding it is the *extrinsic device*, which represents the effect of non-ideal contacts, access regions and gate overlap regions with the source/drain.
- Lastly, surrounding both of these are the *pads*, which represent the connection to the measurement system.

The extrinsic device and pads may be modeled with bias and frequency independent components.

The objective is to remove the influence of the extrinsic device and pads from the intrinsic device, and to extract intrinsic device parameters:

- The process begins with S-parameter measurement of the device under test (DUT), DUT biased under "cold" condition (V_{DS} = 0 V), and "open" and "short" structures.
- Next, *S*-parameter of the on-wafer "open" and "short" structures are used to de-embed the effect of the pad parasitics from the DUT (details explained in appendix B). These are pseudo-devices designed to have the same contact pad structure as the DUT. Compared to the DUT, active region is removed in the "open" structure, and replaced with a metal connection shorting together the source, drain, and gate in the "short" structure.
- Next, *S*-parameters of the de-embedded "cold" FET are used to remove the effect of extrinsic device parasitics (details explained in appendix B). When biased under "cold" condition without current flow, the device is dominated by the extrinsic device parasitics.
- Then, the extracted parasitics are then optimized using the Advanced Design System (ADS) software package from Keysight Technologies to minimize any mismatch with the *S*-parameter data.
- Finally, the optimized parasitics representing the extrinsic device are removed from the de-

embedded DUT *S*-parameter, and intrinsic circuit elements are extracted (details explained in appendix B).

4.3.2 Results

4.3.2.1 Pulse Measurements

Figure 4.12 shows the transfer characteristics measured under DC and pulsed conditions in long channel planar InGaAs MOSFET devices (t_c = 8 nm and 2 nm). Pulsed current is always higher than under DC, confirming the compromised DC device performance due to oxide trapping in InGaAs MOSFETs. As the pulse width shortens, drain current continues to increase prominently, as slow oxide traps fail to follow the faster gate signal. Due to oxide trapping, a higher V_{gs} is necessary in DC operation compared to pulsed operation to achieve the same drain current; this results a "stretch-out" of the I_d-V_{gs} characteristics along the gate voltage axis by ΔV_{ox}.

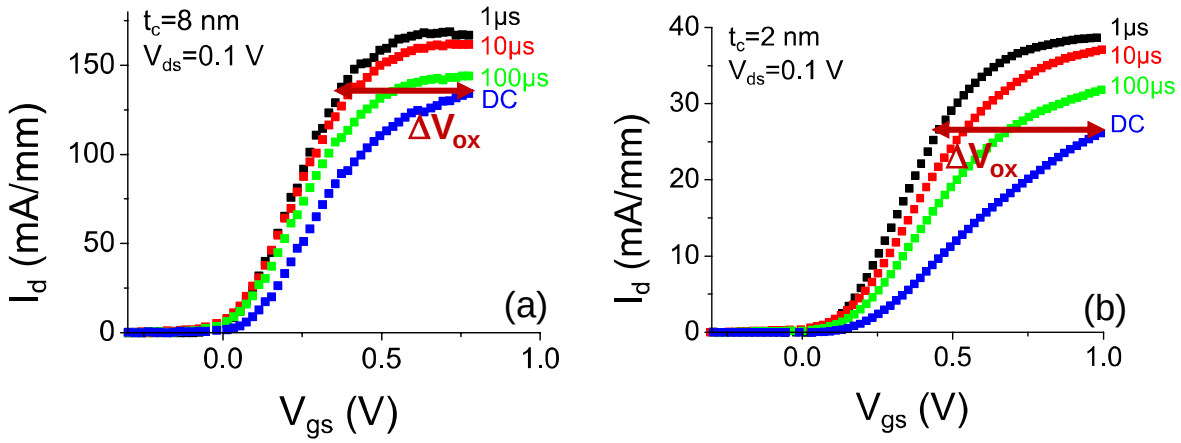

Figure 4.12: Pulsed and DC transfer characteristics of long-channel planar InGaAs MOSFETs with (a) t_c = 8 nm and (b) t_c = 2 nm.

In addition, as the channel thickness shrinks from 8 nm to 2 nm, there is a larger enhancement in I_d under pulsed operation compared to DC and a more significant stretch-out ΔV_{ox}. This indicates more severe oxide trapping and more compromised DC performance in thinner channel devices.

4.3.2.2 Transconductance Dispersion

The intrinsic transconductance ($g_{m,i}$) from DC to 10 GHz is measured and calculated using a combination of parametric, lock-in and *S*-parameter measurements. Figure 4.13 shows the $g_{m,i}$ dispersion for a planar InGaAs MOSFET with t_c = 6 nm and L_g = 200 nm, biased at V_{GS} = 0.4 V and V_{DS} = 0.5 V.

We observe severe frequency dispersion over the entire frequency range, reflecting a rich spectrum of oxide traps with different capture and emission times. Such dispersion in accumulation can be mostly explained by traps inside the high-k dielectric which communicates with carriers in the channel through tunneling. The closer traps lie to the interface, the faster carriers can tunnel in and out. At a fixed DC bias, the small AC signal probes the time response of traps in the vicinity of the

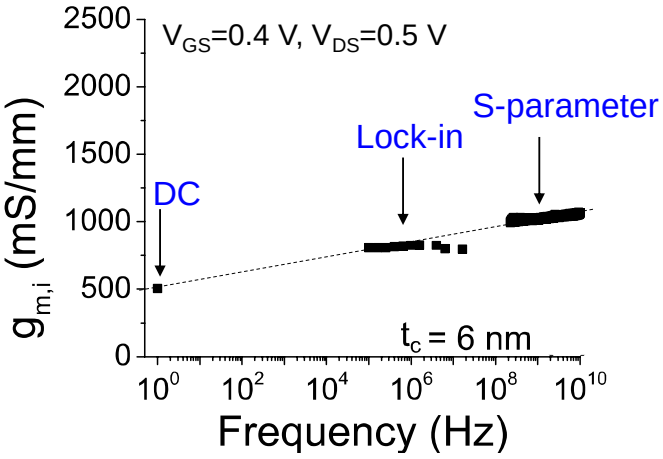

Figure 4.13: Intrinsic transconductance ($g_{m,i}$) as a function of frequency in a planar InGaAs MOSFET with t_c = 6 nm and L_g = 200 nm. Parametric, lock-in and *S*-parameter measurements are used.

Fermi level. As the AC frequency increases, $g_{m,i}$ increases. This is because of the insufficient time for electrons to fill and empty the slow traps, resulting in a higher electron concentration in the channel and higher transconductance compared to DC. Similar behavior is also reported in other InGaAs MOSFETs [59], [72], [76].

In contrast, dispersion-free g_m is observed in Si n-MOSFETs and Si-passivated Ge p-MOSFET, due to the defect-free SiO_2 interfacial layer [72]. As these frequency-dependent measurements probe fast traps close to the semiconductor-oxide interface, the interface quality and defect density near the interface are crucial. Other oxide trapping mechanisms can lead to an opposite phenomenon, where g_m decreases with increasing frequency. For instance, in AlGaN/GaN MOS-HEMT, electron trapping in the oxide can occur with carriers from the gate electrode rather than from the channel, resulting in negative g_m dispersion [77], [78].

Gate Voltage Dependence

We examine the impact of gate voltage on $g_{m,i}$ dispersion. Figure 4.14 shows the $g_{m,i}$ dispersion

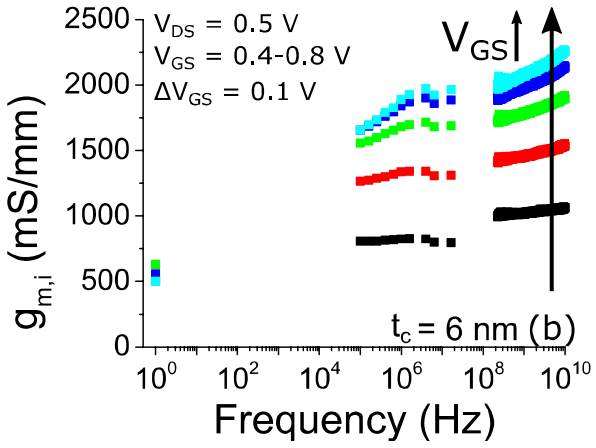

Figure 4.14: Frequency dispersion of $g_{m,i}$ at different DC gate bias (V_{GS} = 0.4-0.8 V and V_{DS} = 0.5 V) in device with t_c = 6 nm.

for a planar InGaAs MOSFET with t_c = 6 nm and L_g = 200 nm, biased at different DC gate bias (V_{GS} = 0.4-0.8 V and V_{DS} = 0.5 V). As the DC gate bias increases, the Fermi level increases inside the conduction band, probing traps at higher energy levels inside the oxide bandgap. The greater dispersion observed at higher V_{GS} indicates more severe trapping at higher energy. This is consistent with the understanding of higher density of border traps at higher energy inside the oxide bandgap [43].

Channel Thickness Dependence

We also find that $g_{m,i}$ dispersion has a strong dependence on channel thickness t_c. Figure 4.15 shows that frequency dispersion of $g_{m,i}$ at a fixed DC gate overdrive bias as a function of channel thickness, normalized to the DC value. As t_c decreases, subband energy increases due to quantization (V_t increases, not shown here). This enables access to border traps at higher energy levels inside the oxide bandgap. Thus, $g_{m,i}$ dispersion increases, having a similar effect as increasing gate bias.

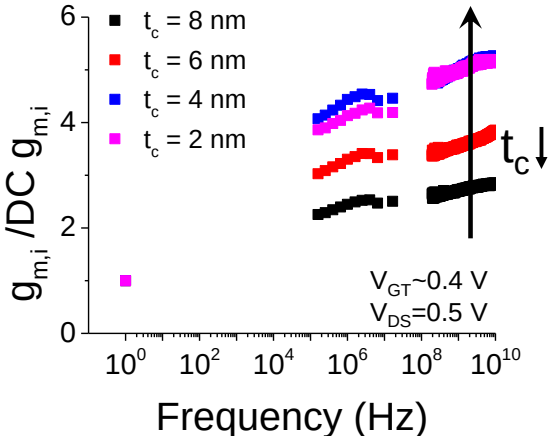

Figure 4.15: Frequency dispersion of $g_{m,i}$/DC $g_{m,i}$ in different t_c at similar gate overdrive (V_{GT} = V_{GS} - V_t ~ 0.4 V).

The strong dependence of frequency-dispersion on t_c is further shown in Figure 4.16, which

compares the $g_{m,i}$ characteristics in saturation regime measured at DC and at 10 GHz. We find that the peak $g_{m,i}$ at 10 GHz is 3 times higher than DC for t_c = 8 nm, rising to as much as 7 times as t_c decreases to 2 nm.

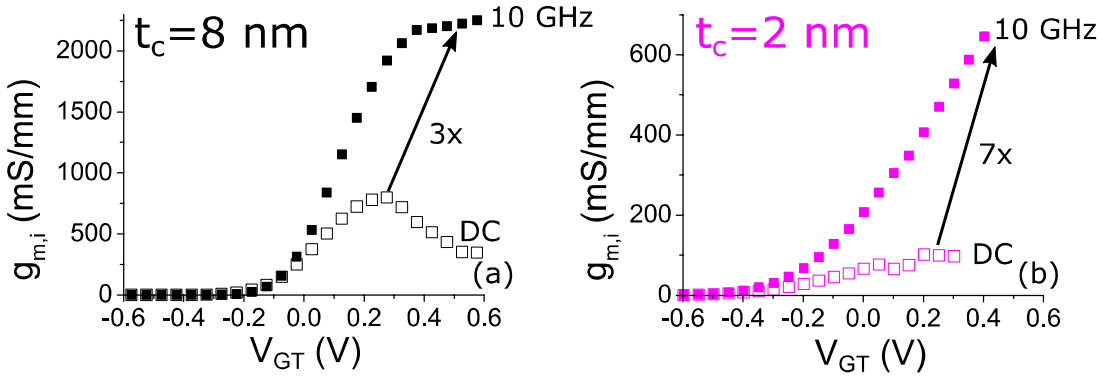

Figure 4.16: Intrinsic transconductance ($g_{m,i}$) versus V_{GT} measured at 10 GHz and at DC, at V_{DS} = 0.5 V . Peak $g_{m,i}$ enhances from DC to 10 GHz by (a) 3 times at t_c = 8 nm and (b) 7 times at t_c = 2 nm.

Implications for Intrinsic Device Performance

So far, we have observed that AC $g_{m,i}$ is consistently higher than DC. The large contrast between DC and high frequency AC behavior suggests that border traps are extremely harmful to device performance, especially in thin channel devices. This is depicted in Figure 4.17, which shows the transconductance characteristics between DC and AC measurement. At DC (Figure 4.17(a)), $g_{m,i}$ degrades severely as channel thickness decreases. In comparison, Figure 4.17(b) shows that at 10 GHz, where most of oxide traps are expected to be unresponsive, the degradation of $g_{m,i}$ is much slower. In fact, significant degradation only starts at 4 nm channel thickness or less. This leads us to the critical conclusion that in current InGaAs FET technology, devices have much higher performance potential that simple DC measurements would suggest. The limit to actual realizable performance may not be due to any intrinsic material limitations, but rather the traps at the interface

and in the oxide. Whereas the former is fundamental, the latter are imperfections in device process that may be addressable through engineering process improvement.

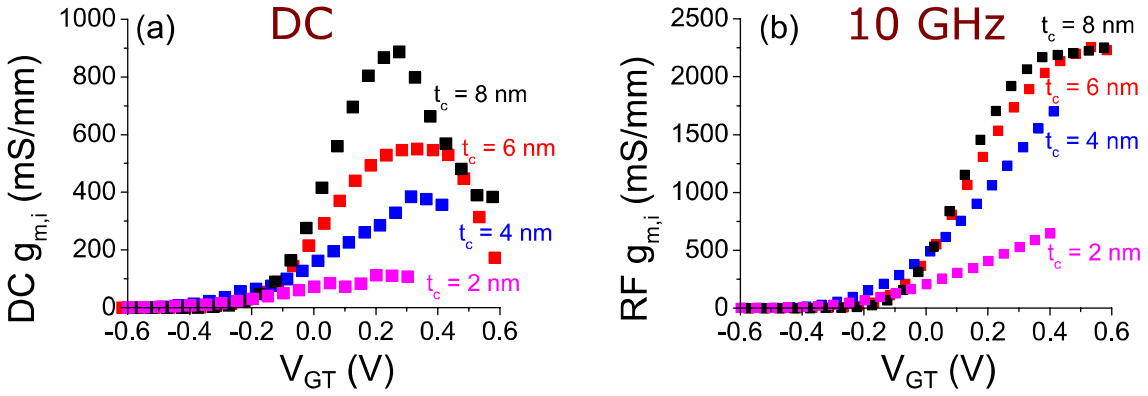

Figure 4.17: Comparison of $g_{m,i}$ characteristics (V_{DS} = 0.5 V) between different t_c at (a) 10 GHz and (b) DC. Degradation of $g_{m,i}$ with t_c is less severe at 10 GHz than DC.

4.4 Complication in Mobility Extraction

Mobility is a key metric to study device transport. Split CV measurement at MHz frequency range is commonly used to estimate the carrier density in the channel and subsequently extract channel mobility.

Figure 4.18 shows typical CV characteristics of a long channel InGaAs MOSFET device measured from 1 kHz to 1 MHz. It is commonly assumed that no significant frequency dispersion occurs beyond MHz frequencies and CV taken at MHz frequency measures only free carriers. However, from the study of g_m dispersion and pulsed measurements, we have found that oxide trapping can occur on a wide range of time scales and respond to AC frequencies up to 10 GHz. This indicates that CV taken at MHz frequency is contaminated by oxide trapping. This trapping effect complicates the estimation of free carriers and mobility.

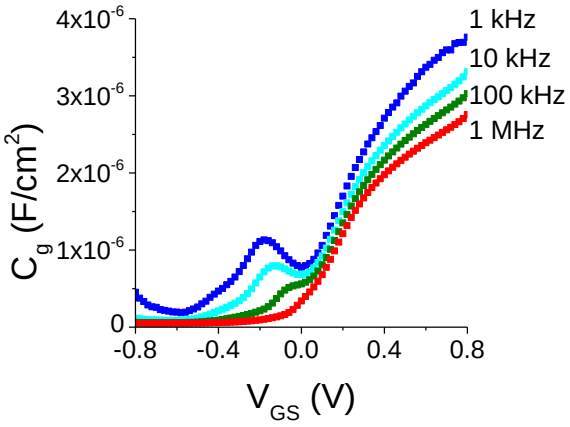

Figure 4.18: Capacitance voltage (CV) measurements of a long channel device at 1 kHz, 10 kHz, 100 kHz and 1 MHz.

In this section, we will first discuss the conventional method for mobility estimation using IV-CV measurement and reveal how oxide trapping underestimates mobility. Then, we will discuss other existing methods for mobility estimation based on gated-Hall and pulsed IV measurements, along with their advantages and disadvantages.

4.4.1 Conventional IV-CV

The conventional method for mobility estimation uses IV-CV measurements with CV taken at MHz frequency when the transistor is biased in the linear regime. The sheet carrier density (N_s) is calculated by integrating the measured capacitance (C_g) over gate voltage bias (V_{GS}),

$$Q_n = qN_s = \int C_g \, dV_{GS} \tag{4.5}$$

This results in an over-estimation of N_s due to two reasons. First, at MHz frequency where CV is commonly taken, prominent oxide trapping is still observed (Figure 4.13). This suggests that

standard CV measures not only conducting carriers, but also a great deal of trapped carriers. Secondly, the V_{GS} axis is severely stretched out due to oxide trapping (Figure 4.12). Integrating over V_{GS} not only integrates mobile carriers but also trapped carriers. As N_s is overestimated, it results in an underestimation of mobility, according to

$$\mu = \frac{I_D}{\frac{W}{L} V_{DS,i} Q_n} \tag{4.6}$$

where $V_{DS,i} = V_{DS} - R_{sd} \times I_D$. We see that oxide trapping not only complicates mobility extraction but also can lead to underestimation of mobility using the conventional method. Alternative methods are needed to accurately estimate mobility.

4.4.2 Gated-Hall

To evaluate only free carriers, Hall measurement can be used [79], [80]. We carry out Hall measurements on gated-Hall structures fabricated on the same wafer as the MOSFETs. Figure 4.19 shows the gated-Hall structure. When we apply the source, drain and gate voltages, it acts like a normal transistor and electrons flow from source to drain. When a vertical magnetic field is applied, the Lorentz force deflects the electrons to one side of the Hall bar, resulting in a voltage difference across the width of the Hall bar. By measuring this Hall voltage (V_{Hall}) and drain current (I_D), carrier density N_s and mobility μ_{Hall} can be calculated using

$$Q_n = qN_s = r_{\text{Hall}} \frac{I_D B}{V_{\text{Hall}}} \tag{4.7}$$

$$\mu_{\text{Hall}} = \frac{I_D}{\frac{W}{L} V_{DS,i} Q_n} \tag{4.8}$$

where the Hall scattering factor r_{Hall} is assumed to be unity.

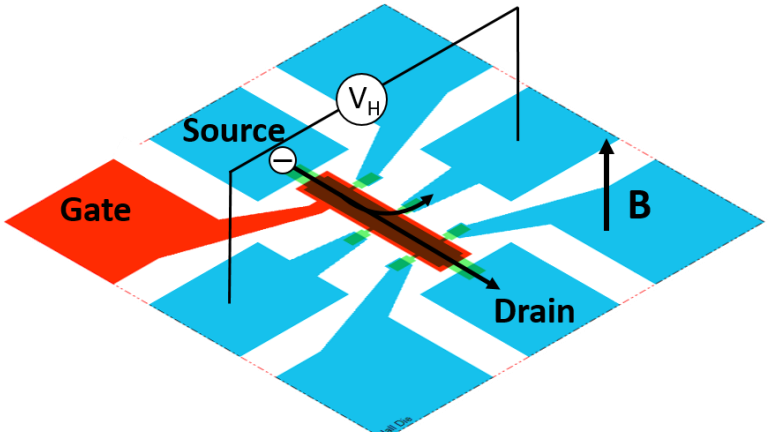

Figure 4.19: Schematic of the gated-Hall measurement.

The advantage of gated-Hall measurement is its sensitivity to mobile carriers alone, as only mobile carriers respond to the magnetic field. Figure 4.20 shows a comparison of mobility estimated from gated-measurement and using conventional IV-CV at 4 MHz for InGaAs MOSFET with 2 nm channel thickness. Indeed, we find that the conventional IV-CV overestimates N_s and underestimates mobility compared to gated-Hall. Similar drastic inconsistency between CV and Hall measurements have also been reported by other researchers [61], [80]. Gated-Hall measurement reveals the true electrostatic and transport in the channel, with the uncertainty of Hall scattering factor. N_s rises initially and starts to saturate later, indicating severe oxide trapping at high gate overdrive voltage (V_{GT}). The gate becomes unable to modulate the surface potential in the channel but continue to trap electrons in the oxide. On the other hand, at t_c = 2 nm, the peak Hall mobility is as high as 2500 $cm^2/V{\cdot}s$, 6-times greater than the mobility estimated from IV-CV.

This again suggests that the low DC $g_{m,i}$ at t_c = 2 nm (shown in Figure 4.17) is likely due to the insufficient supply of mobile charge rather than poor transport. This is similar to the findings in GaAs/AlGaAs MODFETs that inefficient charge modulation limits the current-gain cutoff frequency (f_T) [81].

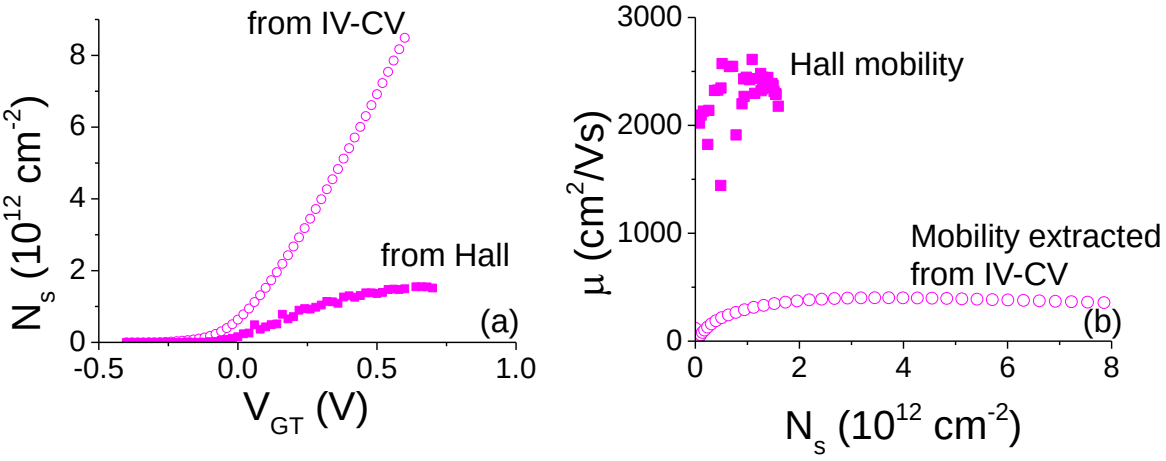

Figure 4.20: Comparison of (a) N_s vs. V_{GT} (b) mobility vs. N_s, extracted from Hall and conventional CV (4 MHz) measurements for InGaAs MOSFET with t_c = 2 nm.

However, Hall measurement has its own disadvantages. The first is the uncertainty of the Hall scattering factor. Many theoretical calculations and experimental studies have shown that the Hall scattering factor in III-V material is close to unity at room temperature [82]–[84]. Thus, a common practice is to assume r_{Hall} = 1 and recognize that the derived "Hall mobility" might be somehow different from drift mobility. Calculating or measuring r_{Hall} rigorously is quite difficult, as r_{Hall} depends on scattering mechanisms and subsequently carrier density, temperature, magnetic field, etc [82], [83]. The quantum well nature of the InGaAs channel complicates this further.

The second disadvantage arises from the difficulty in applying the Hall technique to devices with vertical architecture. As the magnetic field needs to be perpendicular to the plane of conduction, it is only easily applicable to planar FETs.

4.4.3 Pulsed Measurement

A combination of parametric C_g-V_{GS} and pulsed I_d-V_{gs} measurements have recently been proposed to mitigate the effect of oxide trapping and extract mobility in thin-channel InGaAs MOSFETs [53]. As C_g-V_{GS} and DC *IV* are measured at the same sweep rate, the two measurements share the same V_{GS} axes, stretched out by oxide trapping. The difference between pulsed and DC I_d-V_{gs} characteristics can be used to determine the stretch-out ΔV_{ox}, and reconstruct a C_g –V_{GS}^* curve without stretch-out ($V_{GS}^* = V_{GS} - \Delta V_{ox}$), as shown in Figure 4.21 (a) and (b). Figure 4.21 (b) and (c) show the CV characteristics and extracted mobility with and without the stretch-out correction for an InGaAs MOSFET with t_c = 2 nm. After correction, the extracted mobility is 2-times higher, but still significantly lower than that from Hall measurement. This indicates that the correction is incomplete, and shows this technique is limited by minimum pulse width; in the frequency domain, microsecond pulses are only able to observe dispersion up to MHz, whereas we observe severe dispersion even at GHz frequencies.

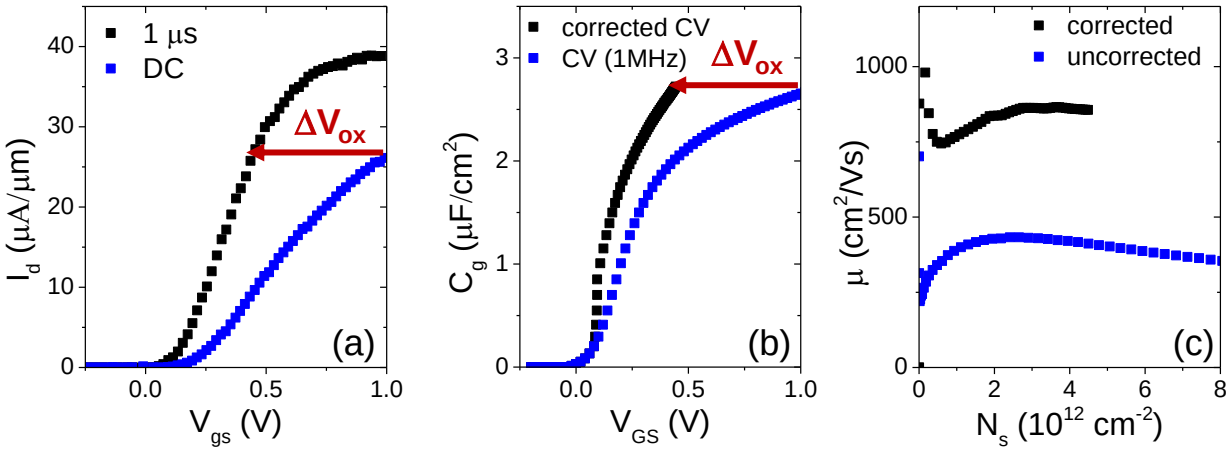

Figure 4.21: (a) Transfer characteristics measured at V_{ds} = 0.1 V under pulsed (pulse width = 1 µs) and DC condition. (b) Measured and corrected CV characteristics after accounting for DC stretch-out ΔV_{ox}. (c) Extracted mobility versus carrier concentration with and without stretch-out correction.

The conclusion from this work is that current characterization techniques other than Hall are unable to accurately extract important transport parameters in MOSFETs in the presence of severe oxide trapping. In the next chapter, we present a new technique that addresses this and can also be used in 3D device structures, such as FinFETs.

4.5 Summary

We have found that oxide trapping deteriorates device stability. Prolonged voltage stress leads to electron trapping in the pre-existing defects in the oxide, resulting in threshold voltage shift and degraded device performance. Under harsh stress conditions, permanent damage can occur through the generation of interface states in the oxide.

In addition, we have also observed large frequency dispersion in transconductance and compromised DC device performance caused by severe oxide trapping in thin-channel InGaAs MOSFETs. When the impact of the border traps is avoided in RF and pulsed measurements, we have found much less degradation with scaling and promising device performance even at 2 nm channel thickness.

Lastly, we have found that oxide trapping also complicates the understanding of charge control in the channel. It results in an over-estimation of charge and under-estimation of mobility when using the conventional IV-CV method.

In Chapter 5, we will establish methods to overcome the impact of oxide trapping and accurately extract carrier concentration and mobility. We will subsequently study the impact of channel scaling on intrinsic device performance and transport.

CHAPTER 5. Impact of Channel Scaling

5.1 Introduction

Chapter 4 has revealed that InGaAs MOSFETs suffer severely from oxide trapping. It degrades device reliability and performance, and obscures true mobility. Considering that oxide quality could possibly be ameliorated through advancements in process technology, the potential and scalability of InGaAs technology should be reassessed.

In Chapter 5, we study the impact of channel scaling on transport and device performance in InGaAs MOSFETs. We first start with planar InGaAs MOSFETs and establish the methodology to recover the correct charge-control relationship and to estimate mobility free from the influence of border traps. We use this methodology to evaluate the impact of channel thickness scaling on device performance and transport in planar InGaAs MOSFETs. Although planar MOSFETs will not be the device architecture for future sub-7 nm logic technology node, their relatively simple structure and well-understood device operation enable them to serve as an excellent platform for device physics study. This will also help us understand the impact of scaling in other device architectures. Then, we apply the same methodology to InGaAs FinFETs and study the impact of fin width scaling on device performance and transport. Finally, we estimate oxide trap density and discuss preliminary results on oxide trap mitigation using forming gas anneal.

5.2 New Method for Mobility Extraction

As discussed in Chapter 4, existing methods for mobility extraction have various shortcomings. Here, we present a new method to extract mobile carrier density and mobility using concurrent *S*-parameter and DC IV measurements in a transistor biased in the linear regime (RF-I_D method).

The calculation of mobility requires an accurate estimation of mobile carrier concentration. Sheet carrier concentration (N_s) calculation can be generally estimated in the following way

$$Q_n = qN_s = \int C_{g,i}\, dV_{GS,i} \tag{5.1}$$

where $C_{g,i} = C_{gs,i} + C_{gd,i}$ is the gate capacitance in the linear regime, and $V_{GS,i}$ is the intrinsic gate-source voltage. However, as shown in Chapter 4, from conventional CV measurement of InGaAs MOSFETs, $C_{g,i}$ is contaminated by trap capacitance and V_{GS} is stretched out due to oxide trapping. This results in an incorrect estimation of N_s.

To avoid measuring oxide traps, trap-free $C_{g,i}$ can be extracted at GHz frequencies from *S*-parameter characterization. To avoid stretch-out in $V_{GS,i}$, the integration must be done over I_D instead of $V_{GS,i}$. The differential for drain current dI_D is given by

$$\begin{aligned} dI_D &= \left.\frac{\partial I_D}{\partial V_{GS,i}}\right|_{V_{DS,i}} dV_{GS,i} + \left.\frac{\partial I_D}{\partial V_{DS,i}}\right|_{V_{GS,i}} dV_{DS,i} \\ &= g_{m,i} dV_{GS,i} + g_{d,i} dV_{DS,i} \end{aligned} \tag{5.2}$$

For a transistor with finite source-drain resistance (R_{sd}), the intrinsic voltage $V_{DS,i}$ is related to the applied voltage V_{DS} by

$$V_{DS,i} = V_{DS} - I_D R_{sd} \tag{5.3}$$

This implies the differential

$$dV_{DS,i} = dV_{DS} - dI_D R_{sd} = -dI_D R_{sd} \tag{5.4}$$

where $dV_{DS} = 0$ because experiments are done with constant drain voltage V_{DS}. Despite that $dV_{DS} = 0$, the non-zero source-drain resistance leads to changes in the intrinsic voltage $V_{DS,i}$ as I_D changes. Inserting equation 5.4 into equation 5.2 yields

$$dI_D = g_{m,i} dV_{GS,i} - dI_D R_{sd} g_{d,i} \tag{5.5}$$

and

$$dV_{GS,i} = \frac{1 + g_{d,i} R_{sd}}{g_{m,i}} dI_D \tag{5.6}$$

Therefore, inserting equation 5.6 into 5.1 yields

$$Q_n = qN_s = \int C_{gi} \frac{1 + g_{d,i} R_{sd}}{g_{m,i}} dI_D \tag{5.7}$$

As $g_{m,i}$, $g_{d,i}$ and $C_{g,i}$ are extracted at GHz frequency, where most traps are unresponsive, this approach for N_s extraction is free from the impact of oxide trapping. The mobility μ can be subsequently calculated using

$$\mu = \frac{I_D}{\frac{W}{L_g} V_{DS,i} Q_n} \tag{5.8}$$

where L_g is the gate length and W is the active region width.

To validate our new method, we have compared results of extracted N_s and μ against Hall measurement using gated-Hall structures on the same InGaAs planar-MOSFET wafer. Figure 5.1 shows good agreement between the new method and Hall measurement for both N_s vs. gate overdrive voltage V_{GT} ($V_{GT} = V_{GS} - V_t$) and μ vs. N_s. One important advantage of this method compared to Hall measurement is that it can be applied to any device architecture.

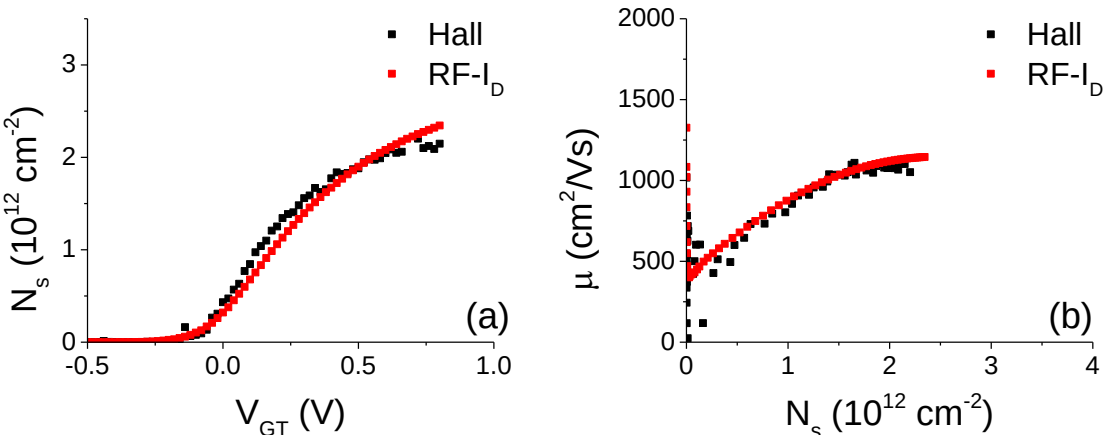

Figure 5.1: (a) carrier concentration and (b) mobility extracted using RF-I_D method (red) and Hall measurement (black) for planar InGaAs MOSFET with t_c = 4 nm.

Another application of this method is that it can be used to estimate the intrinsic gate voltage $V_{GS,i}^*$ in the absence of stretch-out. For this, we can integrate eqn. 5.6 to get

$$V_{GS,i}^* = \int \frac{1 + g_{d,i}R_{sd}}{g_{m,i}} dI_D \tag{5.9}$$

The result for the device in Figure 5.1 is shown in Figure 5.2. To confirm, we have carried out 1D self-consistent Poisson-Schrodinger (P-S) simulations for a planar InGaAs MOSFET with t_c = 4 nm using NextNano. The device structure is described in section 5.3.1. The electronic band structure is modeled using the 8-band $k \cdot p$ method. HfO_2 is parameterized with a dielectric constant of 16, bandgap of 5.6 eV and a conduction band offset with InGaAs of 1.8 eV [85], [86]. We have found that the experimentally extracted N_s vs. V_{GT} does not agree with P-S simulation results, unless V_{GT} is corrected to $V_{GT}^* = V_{GS,i}^* - V_t$, shown in Figure 5.2. The good agreement with theoretical calculation provides additional validation for the new method.

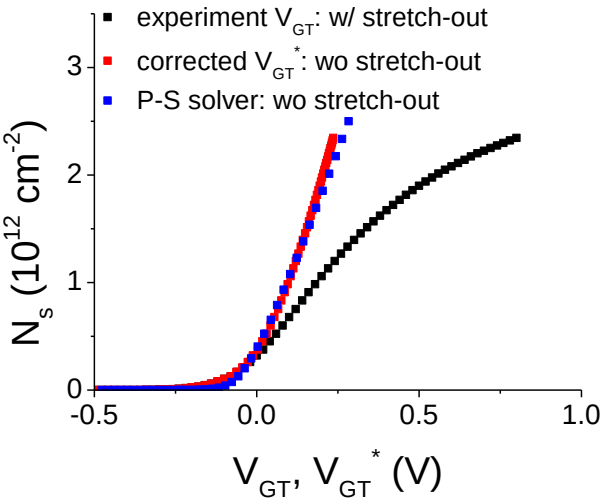

Figure 5.2: N_s as a function of V_{GT}, V_{GT}^* for the experiment in Fig. 5.1. Black: versus measured V_{GT}. Red: versus corrected $V_{GT}^* = V_{GS,i}^* - V_t$. Blue: from Poisson-Schrodinger simulation. Good agreement with P-S simulation is obtained after correcting V_{GT} to V_{GT}^*.

5.3 Channel Thickness Scaling in Planar InGaAs MOSFETs

Channel thickness in a planar MOSFET has a large impact on both transport and electrostatic control, and hence device characteristics. In this section, we study the impact of channel thickness scaling in planar InGaAs MOSFETs with channel thickness between 4 nm and 11 nm. The evolution of DC device characteristics is examined as a function of gate length and channel thickness. The significance of this work lies in the application of RF measurement and the RF-I_D method discussed in section 5.2 to separate the role of oxide trapping, and reveal the intrinsic channel thickness scaling behavior in planar InGaAs MOSFETs.

5.3.1 Device Structure

The InGaAs MOSFETs used in this study have an intrinsic channel $In_{0.7}Ga_{0.3}As$ with various channel thicknesses (t_c = 11, 9, 7, 5, 4 nm). The initial heterostructure is described in Figure 2.1

(a) and device fabrication is described in Chapter 2. To achieve a high-quality semiconductor-oxide interface, a combination of dry and wet etch is used for the III-V n^+ cap recess to stop selectively on the InP layer. Digital etch is then used to remove III-V material at a rate of ~1 nm/cycle and to achieve any targeted t_c. For t_c = 11 nm, the high-k dielectric sits on top of a 1 nm thick InP barrier and 10 nm thick $In_{0.7}Ga_{0.3}As$ channel, resulting in buried channel devices. For t_c = 9, 7, 5, 4 nm, the high-k dielectric sits directly on the $In_{0.7}Ga_{0.3}As$ channel, resulting in surface channel devices. All devices share the same gate stack, consisting of 4 nm-thick HfO_2 deposited by ALD at 250 °C and sputtered 30 nm-thick Mo.

5.3.2 DC Scaling Behavior

We first perform DC characterization. Figure 5.3 shows the output and subthreshold characteristics of a typical device with L_g = 100 nm and t_c = 11 nm. We further extract key figures of merit to study the scaling behavior as a function of channel length L_g and channel thickness t_c.

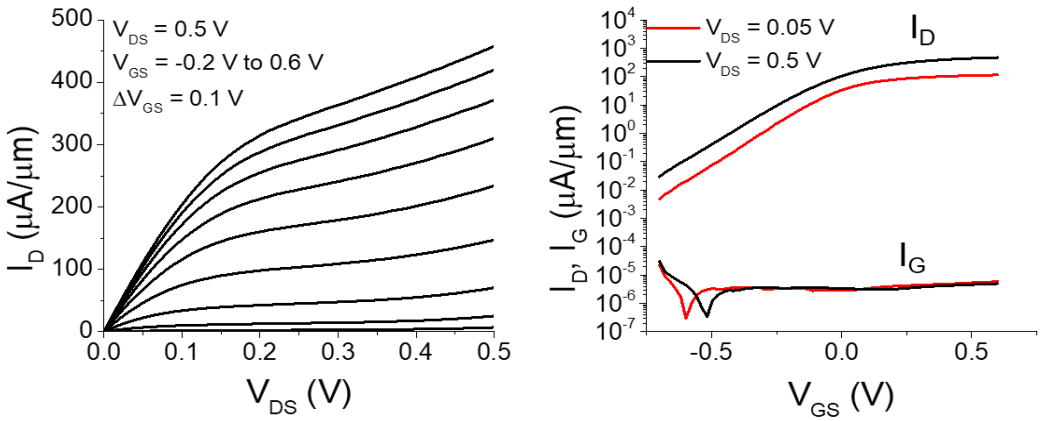

Figure 5.3: Output and subthreshold characteristics of a typical device with L_g = 100 nm and t_c = 11 nm.

DC Transconductance

Figure 5.4 shows peak DC transconductance ($g_{m,max}$) vs. L_g for different t_c. As L_g decreases, ON-state performance enhances leading to higher $g_{m,max}$. This is a classic gate length scaling behavior. However, as t_c scales down from 9 nm to 4 nm in surface channel devices, $g_{m,max}$ degrades, likely due to the enhanced scattering in thinner channels. This is more evident in long-channel devices than short-channel devices, as they suffer more from scattering in the channel. As gate length shortens and approaches mean-free-path, transport in the channel becomes ballistic or quasi-ballistic with minimal scattering [24], [87]. For buried channel (t_c = 11 nm), despite the thicker channel, $g_{m,max}$ is inferior than that in 9 nm surface channel devices. This is likely due to the reduced capacitance and subsequently reduced number of mobile carriers in the channel.

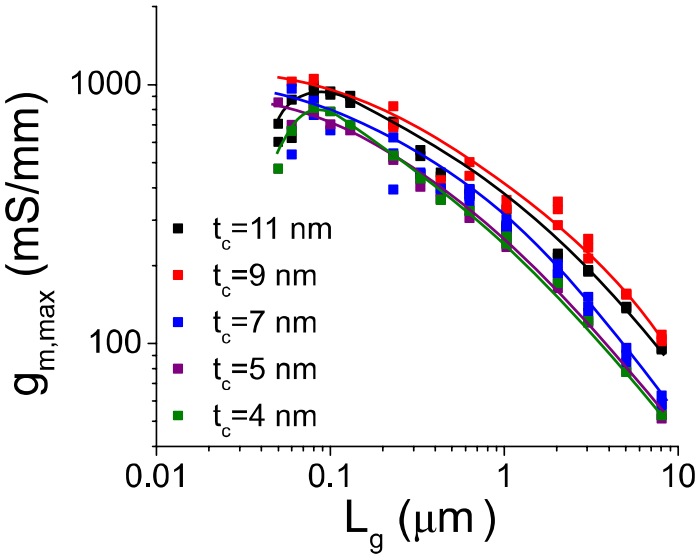

Figure 5.4: Peak transconductance $g_{m,max}$ (V_{DS} = 0.5 V) vs. L_g for different t_c.

ON-resistance

ON-resistance R_{on} is extracted when the device is biased with high gate overdrive voltage in the linear regime (V_{GT} > 0.6 V and V_{DS} = 0.05 V). Figure 5.5(a) shows R_{on} versus L_g for different t_c. As channel length increases, channel resistance increases proportionally, leading to higher R_{on}. Source-drain resistance R_{sd} can be subsequently extracted from extrapolating R_{on} vs. L_g to zero L_g, using R_{on} data from devices with L_g =200 nm to 8 μm. Figure 5.5(b) shows R_{sd} for different t_c. As t_c thins down, the distance between the n^+ cap and intrinsic channel increases, leading to higher access resistance and higher R_{sd}.

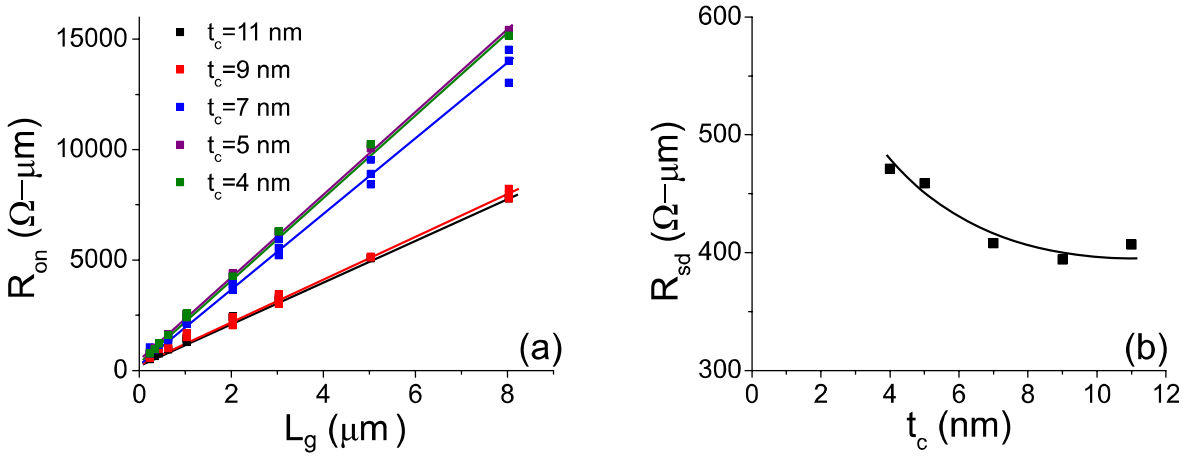

Figure 5.5: (a) R_{on} vs. L_g for different t_c. (b) R_{sd} vs. t_c.

Threshold Voltage and Subthreshold Swing

OFF-state scaling behavior is shown in Figure 5.6. The scaling of linear threshold voltage $V_{t,lin}$ vs. L_g for different t_c is shown in Figure 5.6 (a) and (b). $V_{t,lin}$ is extracted by linearly extrapolating I_D to the V_{GS} axis at the maximum g_m (V_{DS} = 0.05 V). As L_g decreases, V_t decreases, a manifestation of short-channel effects (SCE). Interestingly, V_t in surface channel devices does not monotonically increase as t_c decreases, as expected from quantization. This might be due to spatial variations

across the wafer or non-uniform impact of oxide trapping and interface states in different channel thicknesses. Figure 5.6(b) plots V_t roll-off relative to long channel devices. Evidently, thinner channel devices have a much tighter distribution of V_t. This shows the improved electrostatics in thinner t_c and the benefits of channel thickness scaling in reducing SCE.

Figure 5.6 (c) and (d) depict the distribution of minimum subthreshold swing at V_{DS} = 0.05 V as a function of L_g in different t_c. As expected, S_{lin} worsens as L_g decreases. The dependence of S_{lin} on L_g is weaker with thinner t_c, again confirming better electrostatic control in thinner channels.

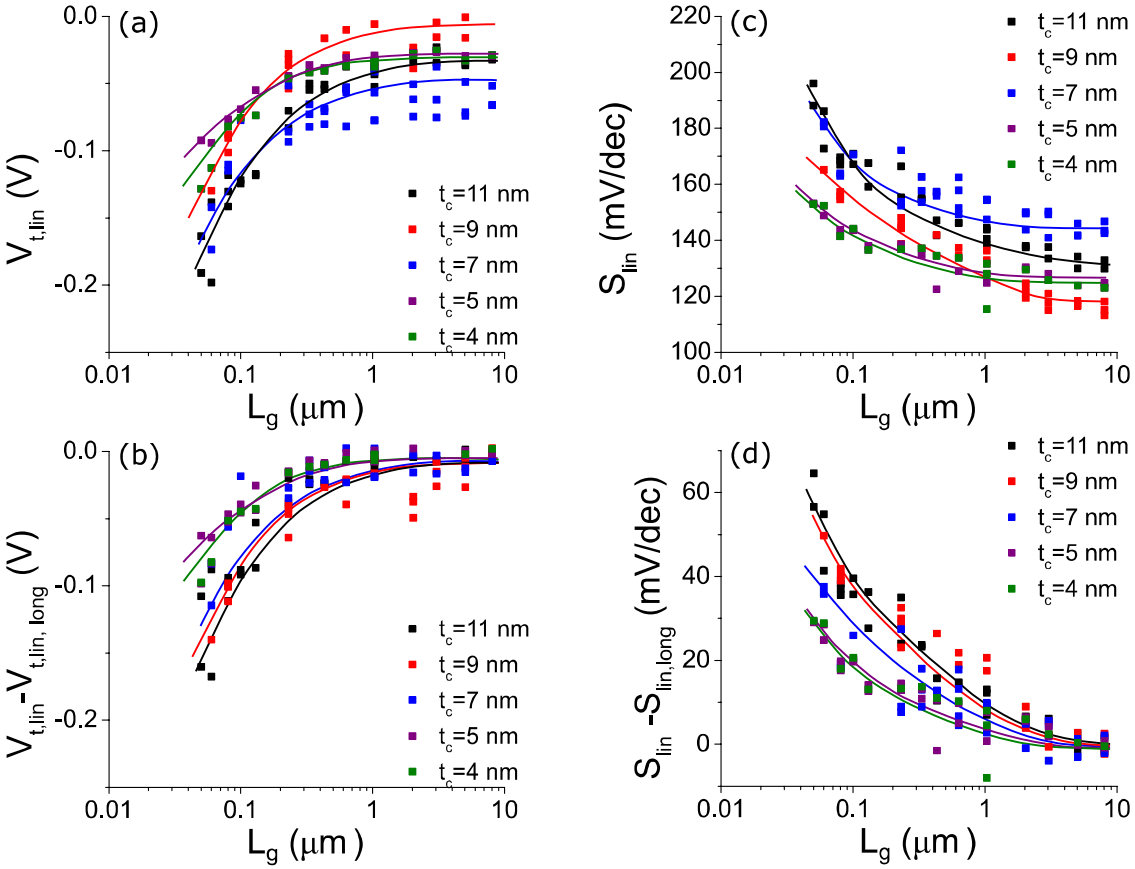

Figure 5.6: The distribution of (a) $V_{t,lin}$ (b) $V_{t,lin}$ - $V_{t,lin,long}$ (c) S_{lin} (d) S_{lin}- $S_{lin,long}$ as a function of L_g for different t_c.

5.3.3 RF Analysis

As discussed in Chapter 4, oxide traps can mask intrinsic device performance. In order to further understand the scaling behavior with channel thickness, we have carried out additional RF measurements at GHz frequencies. At these frequencies, most traps are unresponsive. This enables us to study the intrinsic scaling in device performance and transport.

Charge Control Relationship

To study charge control in InGaAs MOSFETs, capacitance-voltage (CV) characteristics of long-channel devices are measured. Figure 5.7 shows CV characteristics for a device with t_c = 11 nm and L_g = 2 μm. From 1 kHz to 4 MHz, conventional split CV measurement is carried out. For 100 MHz and 1 GHz, C_{gi} is extracted from S-parameter measurements and effect of series resistance is removed, as described in Section 4.3.1.2. Consistent with the observation of large frequency dispersion in g_m, large frequency dispersion is observed in CV characteristics in the ON regime. This confirms that severe oxide trapping still occurs at 4 MHz, especially at high V_{GS}, and

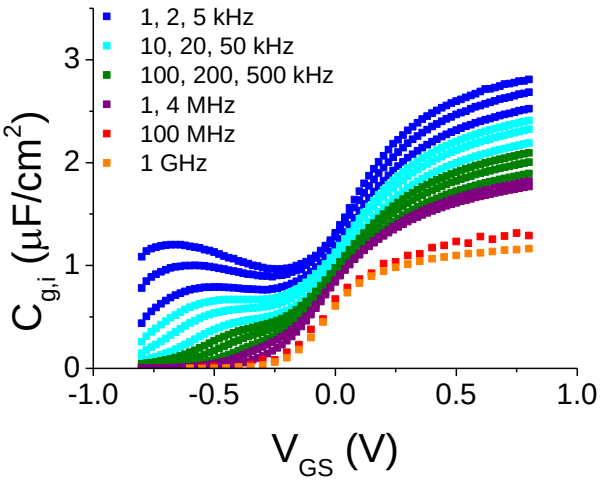

Figure 5.7: Capacitance-voltage characteristics of a planar InGaAs MOSFETs with t_c = 11 nm and L_g = 2 μm from 1 kHz to 1 GHz.

contaminates the results obtained from conventional split-CV measurement. On the other hand, little dispersion is observed beyond 1 GHz.

Figure 5.8 compares CV characteristics for different t_c at 4 MHz and at 1 GHz. To account for different V_t in devices with different t_c, C_g at 4 MHz and C_{gi} at 1 GHz are plotted as a function of gate overdrive voltage V_{GT}. V_t is extracted from DC IV characteristics taken concurrently with CV and *S*-parameter measurements. At high V_{GT}, ON-state capacitance does not saturate, but continues to increase with V_{GT}. On-state capacitance also increases monotonically with decreasing t_c. Both effects are due to the increasing semiconductor capacitance C_s as the charge centroid moves closes to the semiconductor-oxide interface due to increasing V_{GT} or decreasing t_c. For all channel thicknesses, capacitance further decreases as AC frequency increases from 4 MHz to 1 GHz.

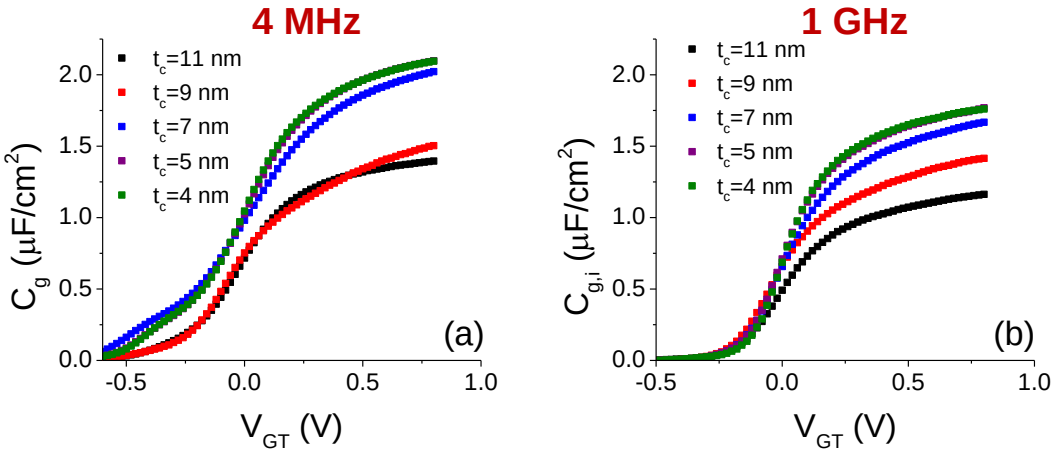

Figure 5.8: Capacitance-voltage characteristics in long channel planar InGaAs MOSFETs of different t_c, measured at (a) 4 MHz and (b) 1 GHz.

Transconductance: DC vs. RF

Transconductance reflects both transport and electrostatic control in the channel. For long channel devices operating in the linear regime, carrier transport is governed by drift-diffusion [24], [88]

and transconductance is approximately the product of mobility and capacitance, as in

$$g_{m,i} = \frac{\partial I_D}{\partial V_{GS,i}}\bigg|_{V_{DS,i}} = \frac{\partial}{\partial V_{GS,i}}\left(\frac{W}{L_g} V_{DS,i}\, Q_n \mu\right) \approx \frac{W}{L_g} V_{DS,i} C_{g,i} \mu \tag{5.10}$$

Figure 5.9 compares intrinsic transconductance ($g_{m,i}$) characteristics (V_{DS} = 0.05 V) vs. V_{GT} in long channel devices (L_g = 2 μm) for different t_c at DC and at 1 GHz. $g_{m,i}$ at 1 GHz is much higher than at DC, and the shape of the $g_{m,i}$ curve is quite different. The rapid degradation of DC $g_{m,i}$ at high gate voltage bias is often attributed to mobility degradation. In contrast, at 1 GHz where the gate voltage is modulated quickly enough to avoid oxide trapping, such degradation is much relieved. In fact, the $g_{m,i}$ characteristic for t_c = 11 nm has not yet reached its peak in the measurement voltage range. The severe degradation at DC can be instead attributed to the inability to modulate the surface potential due to severe oxide trapping, and the consequent degradation in capacitance, rather than mobility. In addition, the percentage degradation in g_{mi} vs. t_c is less severe at 1 GHz compared to DC; for example, from t_c = 9 nm to t_c = 4 nm, maximum g_{mi} (V_{DS} = 0.05 V) decreases by 64% at DC but only by 33% at 1 GHz. Surprisingly, buried channel t_c=11 nm shows the most

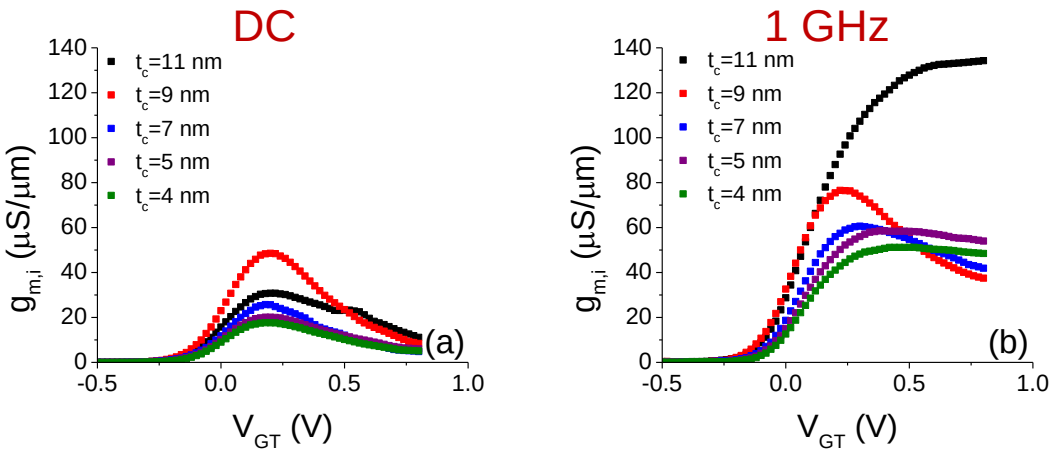

Figure 5.9: transconductance characteristics in long channel planar InGaAs MOSFETs of different t_c, measured with V_{DS} = 0.05 V at (a) DC and (b) 1 GHz.

significant enhancement in $g_{m,i}$ from DC to 1 GHz. The larger dispersion may indicate that interface between HfO_2-InP is worse than HfO_2-InGaAs.

We now examine the maximum extrinsic transconductance ($g_{m,max}$ at V_{DS} = 0.5 V) in short channel devices. As the device is biased in strong inversion in saturation, transport is governed by the injection velocity at the top of the potential barrier between the source and channel [3], [88]. g_m can be approximated as the product of capacitance and injection velocity

$$g_m = \frac{\partial I_D}{\partial V_{GS}}\bigg|_{V_{DS}} = \frac{\partial}{\partial V_{GS}}\left(W\, Q_n v_{inj}\right) \approx W\, C_g v_{inj} \tag{5.11}$$

This is one of the key metrics used to assess the speed and performance of a logic technology. Figure 5.10 shows the t_c scaling of $g_{m,max}$ at DC and at 1 GHz in short-channel devices. Similar to the observation in long-channel devices, $g_{m,max}$ at 1 GHz extracted from *S*-parameters is higher than at DC, revealing the high intrinsic performance that is not realized in DC. Moreover, $g_{m,max}$ at V_{DS} = 0.5 V displays a non-monotonic scaling behavior with t_c, in contrast to at V_{DS} = 0.05 V (Figure 5.9). As t_c decreases, $g_{m,max}$ initially increases, which may be due to better SCE control in

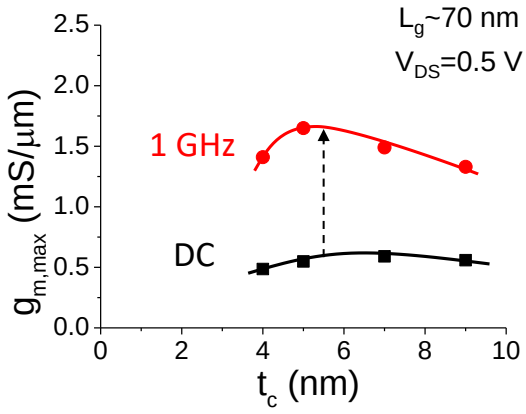

Figure 5.10: DC and 1 GHz $g_{m,max}$ (V_{DS} = 0.5 V) dependence on channel thickness in planar MOSFETs.

thinner channels. As t_c further decreases below 5 nm, $g_{m,max}$ degrades, mostly likely due to the enhanced scattering in extremely thin channels. This also indicates a different channel thickness scaling of v_{inj} compared to μ, which requires further investigation.

5.3.4 Mobility

The advantage of InGaAs channels over Si relies on its superior transport properties, best manifested through its high mobility. Many have reported a rapid mobility degradation with channel thickness scaling in InGaAs MOSFETs with various device architectures [34], [42], [89]. However, the impact of oxide trapping is often not considered. Thus, here we apply the RF-I_D method in long channel devices, to study mobility scaling with channel thickness in planar InGaAs MOSFETs.

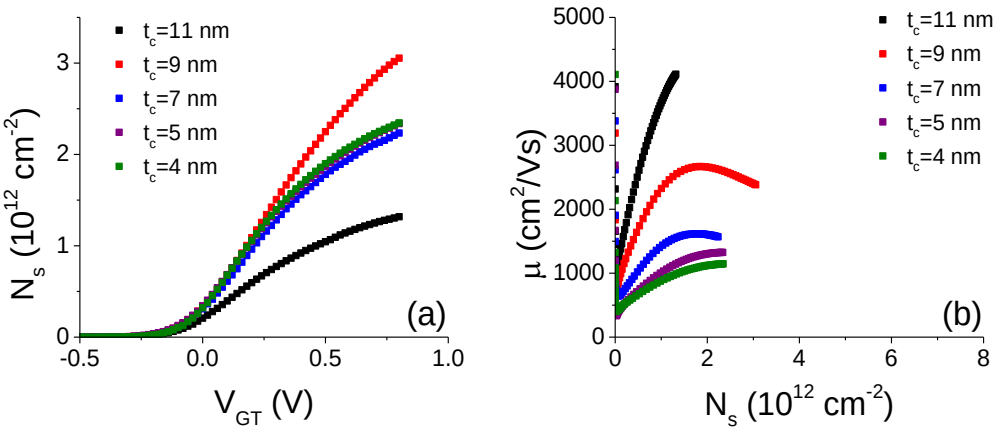

Figure 5.11: (a) N_s vs. V_{gt} (b) μ vs. N_s in planar MOSFETs of different t_c.

Figure 5.11 (a) shows that at the same gate overdrive voltage V_{GT}, carrier concentration N_s shows a non-monotonic dependence with t_c. The buried channel t_c = 11 nm device exhibits the lowest N_s, despite the highest mobility. Among the surface channel devices, N_s is highest for t_c = 9 nm and

drops significantly as t_c decreases to 7 nm, which we ascribe to enhanced oxide trapping in thinner channels. However, as t_c drops further from 7 to 4 nm, a slight increase in N_s is observed, which might be due to increased gate capacitance.

Figure 5.11 (b) shows that mobility μ has a non-monotonic dependence on N_s. At low N_s, μ increases as N_s increases, governed by Coulomb scattering; at high N_s, μ becomes limited by surface roughness scattering, degrading as N_s increases. As t_c decreases, μ monotonically decreases. Several factors aggravated by thinner channels may contribute to this, including enhanced surface roughness scattering, Coulomb scattering from charged oxide traps, and heavier effective mass due to nonparabolicity.

We have also carried out gated-Hall and conventional IV-CV (CV taken at 4 MHz) measurements for comparison. Figure 5.12 shows N_s extracted by three different methods. RF-I_D and Hall show good agreement for all t_c. This indicates again that the RF-I_D method successfully avoids the impact of oxide trapping in all t_c, like Hall measurement does. In contrast, the conventional CV method at 4 MHz severely overestimates N_s by a factor of 3-5. In addition, the shape of the N_s characteristics is drastically different from the other two methods. At high gate voltage, N_s begins to saturate in both the RF-I_D and Hall methods, indicating loss of gate control over the channel due to severe oxide trapping. Conventional CV measurement fails to capture this and N_s continues to increase linearly, as excessive trapped carriers can still be measured at 4 MHz and N_s is incorrectly estimated (equation 4.5).

The accuracy in N_s extraction determines the outcomes of mobility estimation. Figure 5.13 subsequently compares μ extracted from the three different methods. RF-I_D and Hall again show

good agreement for all t_c. As the conventional CV method overestimates N_s, it subsequently underestimates μ. As the channel thins down, the impact of oxide trapping is more severe, aggravating the incorrect estimation of N_s and μ, resulting in an incorrect suggestion of severe mobility degradation.

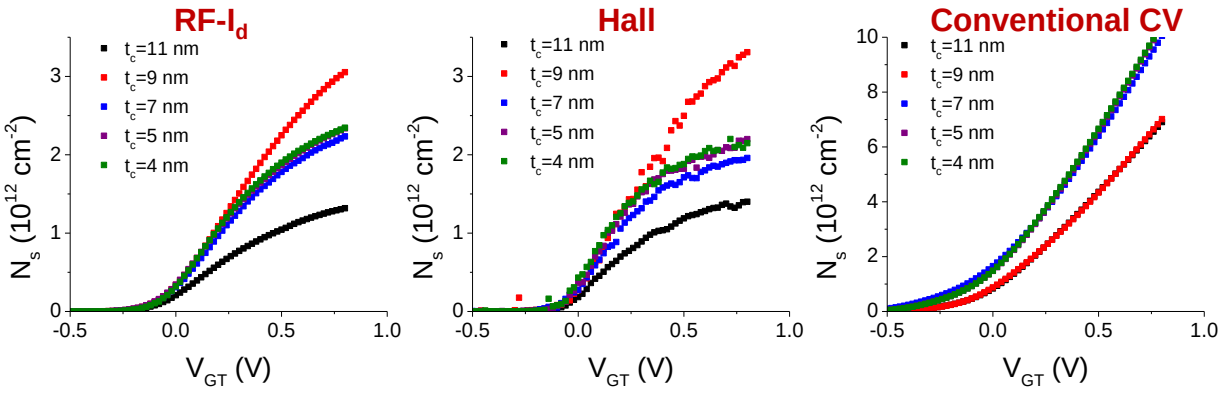

Figure 5.12: Carrier concentration N_s vs. V_{GT} in planar MOSFETs of different t_c using: from left to right, RF-I_D method, gated-Hall measurement, classic IV-CV with CV measured at 4 MHz. Note that the graph of conventional CV has a different vertical scales.

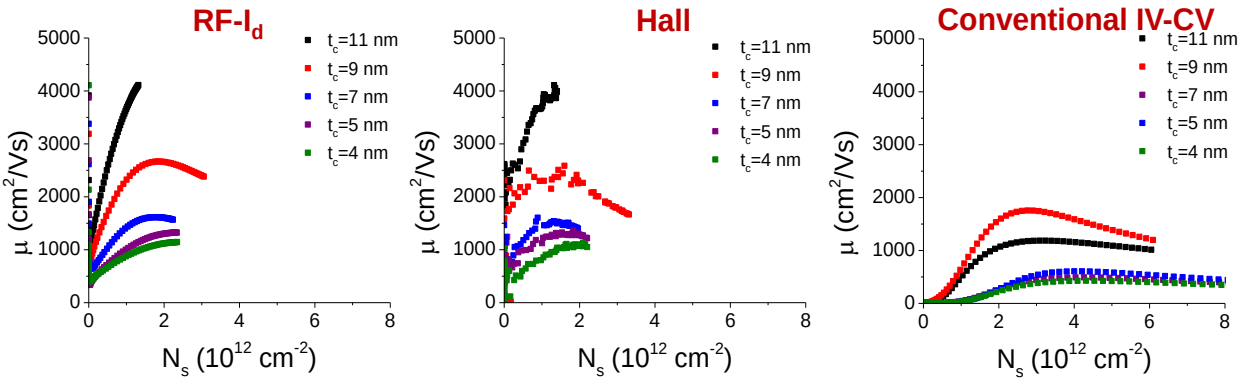

Figure 5.13. Mobility vs. N_s in planar MOSFETs of different t_c using: from left to right, RF-I_D method, gated-Hall measurement, conventional IV-CV with CV measured at 4 MHz.

Unfortunately, conventional CV is still the most common method for mobility estimation in the device community today. Furthermore, many predictive TCAD simulations are calibrated using

mobility data analyzed by conventional CV [55]–[58]. This has resulted in many researchers dismissing InGaAs MOSFET performance as unscalable for future nanoscale CMOS technology [56]–[58]. Our new results indicate this is not the case. Figure 5.14 shows the t_c dependence of peak mobility, extracted using the conventional IV-CV and new RF-I_D methods. We find that with the new method to accurately estimate mobility, the inherent potential of InGaAs is significantly higher than what has been reported so far. μ remains as high as 1000 $cm^2/V{\cdot}s$ even at t_c = 4 nm.

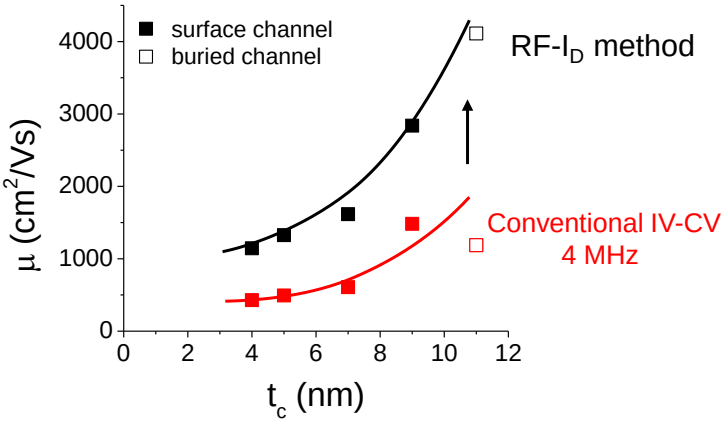

Figure 5.14: Peak mobility versus t_c in planar MOSFETs, comparing RF-I_D (black squares) and the conventional CV method (red square). Solid symbols represent surface channel devices and hollow symbols represent buried channel devices.

5.3.5 Temperature-dependent Study[2]

To study the temperature dependence of oxide trapping behavior and transport in InGaAs MOSFETs, we have carried out simultaneous I_D-V_{GS} and *S*-parameter measurements at different

[2] This work is done in collaboration with Dr. Jorge Pedros and Professor Jesús Grajal from Universidad Politécnica de Madrid (UPM).

temperatures (20 K, 77 K, 150 K, 225 K, RT) under vacuum in a closed-cycle helium-cooled probe station at UPM. Figure 5.15(a) shows the temperature-dependent subthreshold characteristics in long-channel InGaAs MOSFET device (t_c = 9 nm, L_g = 5 μm).

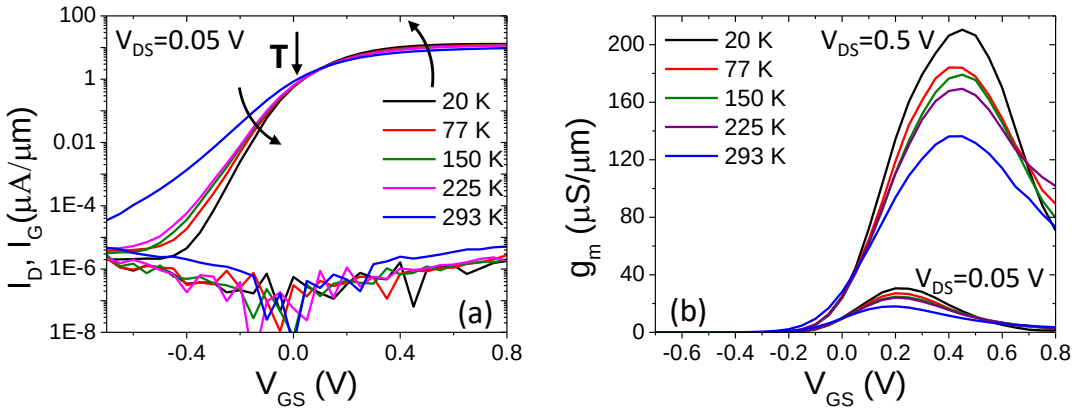

Figure 5.15: Temperature-dependent (a) subthreshold characteristics at V_{DS} = 0.05 V and (b) transconductance characteristics at V_{DS} = 0.05 V and V_{DS} = 0.5 V in long-channel InGaAs MOSFET device (t_c = 9 nm, L_g = 5 μm).

As temperature decreases, OFF-current decreases, along with an improved subthreshold swing. The subthreshold swing follows the equation

$$S = \frac{k_B T}{q}(\ln 10)\left(1 + \frac{C_{it}}{C_{ox}}\right) \tag{5.12}$$

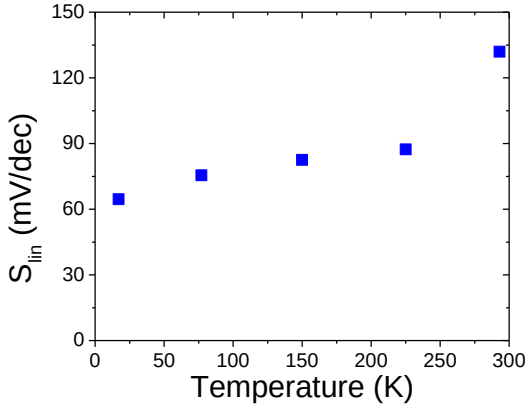

Figure 5.16: Minimum subthreshold swing (V_{DS} = 0.05 V) vs. Temperature for the devices in figure 5.15.

Figure 5.16 shows that a rapid decrease is observed from 293 K to 225 K that exceeds $\frac{k_B \Delta T}{q}(\ln 10)$, likely due to the freezing of the interface traps. However, below 225 K, the subthreshold swing becomes only weakly temperature dependent. This indicates the possible presence of an additional leakage between the source and drain, for example through trap-assisted tunneling.

In the ON state, we observe an enhancement in the current and the DC g_m as the temperature decreases (Figure 5.15). To further investigate this behavior, we extracted $g_{m,i}$ and $C_{g,i}$ at 0.2 GHz (Figure 5.17). Similar to DC, peak $g_{m,i}$ increases as temperature lowers. Capacitance rises more sharply around threshold and ON-state capacitance drops slightly.

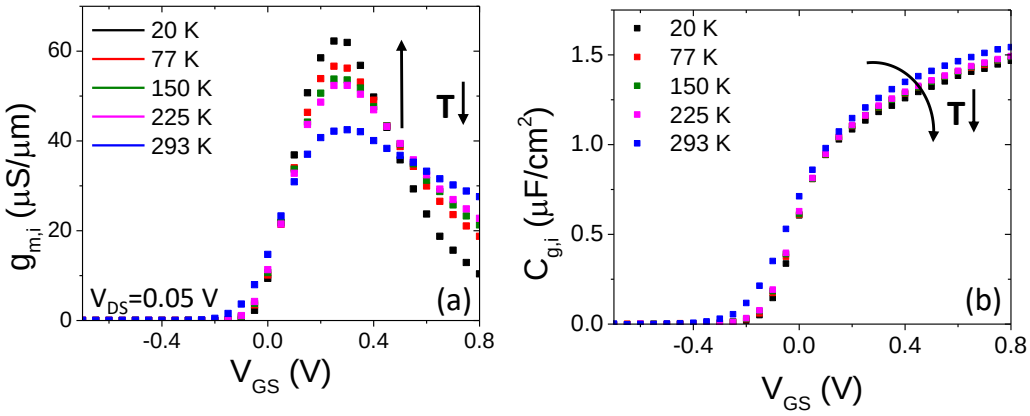

Figure 5.17: Extracted (a) $g_{m,i}$ vs. V_{GS} and (b) $C_{g,i}$ vs. V_{GS} at different temperatures at 0.2 GHz.

To understand the origin of this improvement in peak g_m, we used the RF-I_D method to extract the sheet carrier concentration N_s and mobility at different temperatures. Figure 5.18 shows that N_s only increases slightly with decreasing temperature. Thus, lowering the temperature does not fully alleviate oxide trapping in the ON-state. This is reasonable given that the Fermi level lies in the conduction band in the ON state. The temperature insensitivity is furthermore consistent with the

understanding that electron trapping in the oxide is mediated by tunneling, which is only weakly temperature dependent.

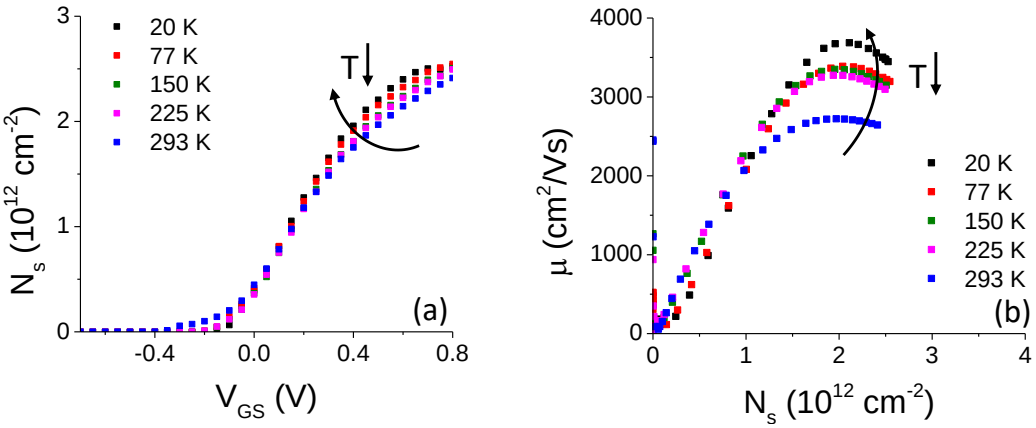

Figure 5.18: Extracted (a) N_s vs. V_{GS} and (b) μ vs. N_s at different temperatures using the RF-I_D method.

We further extract mobility at different temperatures. Figure 5.18(b) shows that lowering temperature significantly improves mobility. This is likely due to the reduction of phonon scattering in the channel, and suggests that the improvement in g_m at lower temperature is due to enhanced transport rather than charge control.

5.4 Fin Width Scaling in InGaAs FinFETs

FinFET is the current state-of-the-art device architecture for logic technology [14]. In the quest to demonstrate III-V technology for logic applications, numerous highly scaled InGaAs FinFETs have been made with dimensions approaching those of the state-of-the-art Si FinFETs [34], [35], [74], [90]. However, unlike planar InGaAs MOSFETs, performance of most InGaAs FinFETs is still lagging behind Si FinFETs. Rapid degradation of transconductance is also observed as fin width scales down [34], [35], raising concern over the scalability of InGaAs FinFETs technology down to 5 nm fin width.

Compared to planar MOSFETs, InGaAs FinFETs fabricated in a top-down approach, suffer from extensive RIE etching on the sidewalls. The damaged sidewalls with excessive roughness and defects may cause extensive oxide trapping during device operation. With the understanding and methodology developed in planar InGaAs MOSFETs, in this section, we examine the impact of oxide trapping in InGaAs FinFETs and study the fin width scaling of intrinsic device performance and mobility.

5.4.1 Device Structure

The InGaAs FinFETs used in this study were first reported at IEDM 2017 [34], featuring fin width (W_f) from 25 nm down to 7 nm and gate length (L_g) down to 60 nm. Figure 5.19 (a) shows the starting heterostructure that includes a heavily-doped cap of 30 nm thick Si:InGaAs, an etch stopper of 4 nm InP, an intrinsic channel of 50 nm thick $In_{0.53}Ga_{0.47}As$ layer and an InAlAs buffer beneath the channel. Device fabrication followed a contact first, gate last self-aligned process similar to the planar InGaAs MOSFET fabrication. A combination of dry etch and digital etch

[25], [27] was used to define high aspect ratio fins to achieve smooth and vertical sidewalls. The HSQ that serves as the dry etch mask remains on top of the fins making these double-gate FinFETs. The gate oxide consists of 1 monolayer of Al_2O_3 and 3 nm HfO_2 (EOT ~ 0.8 nm), deposited by ALD.

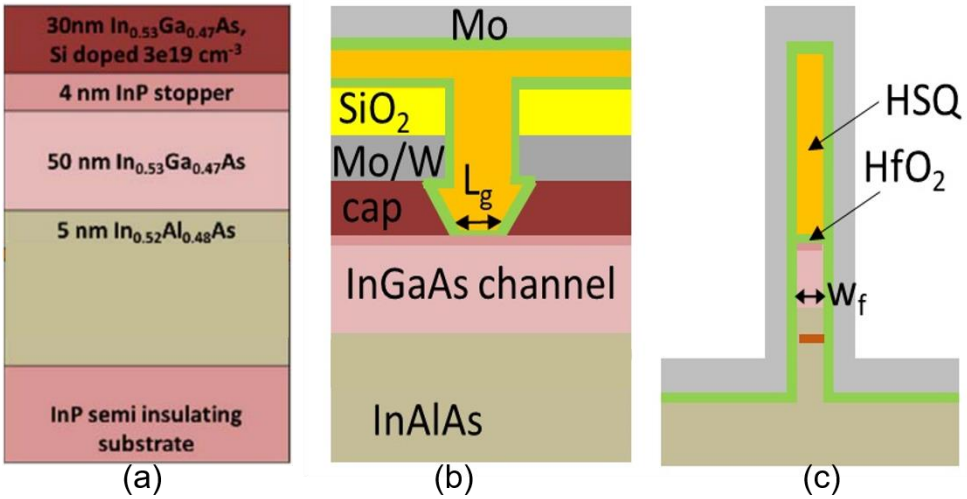

Figure 5.19: Self-aligned InGaAs FinFETs used in this study (a) starting heterostructure; schematic cross section of the (b) along the channel and (c) across the fin direction. The intrinsic channel is $In_{0.53}Ga_{0.47}As$. The gate oxide consists of 1 monolayer of Al_2O_3 and 3 nm HfO_2 (EOT ~ 0.8 nm), deposited by ALD.

5.4.2 Results and Discussion

These are some of the most aggressively scaled and best-performed InGaAs FinFETs ever reported. The scaling behavior of DC device characteristics ($g_{m,}$ S_{sat}, V_t, R_{on}, R_{sd}) has been reported [34] and will not be discussed in details here. In this thesis, we carry out additional pulsed-IV and RF characterization, to reveal the impact of oxide trapping and intrinsic device performance. Intrinsic device parameters ($C_{g,i}$, $g_{m,i}$ etc.) are extracted at 1 GHz using equivalent circuit analysis.

We will also extract carrier density and mobility using the RF-I_D method developed in section 5.2, and examine the impact of fin width scaling.

Pulsed-IV

Transfer characteristics are measured in InGaAs FinFETs with various fin widths, when the gate voltage is applied under DC and pulsed conditions, shown in Figure 5.20. Pulsed current is consistently higher than DC, and as the pulse width shortens, the drain current continues to increase prominently. This is similar to the observation in InGaAs planar devices, confirming strong oxide trapping in InGaAs FinFETs. In addition, as W_f narrows below 13 nm, the enhancement in I_d from DC to pulsed conditions further increases, indicating that narrow fin devices suffer more from oxide trapping.

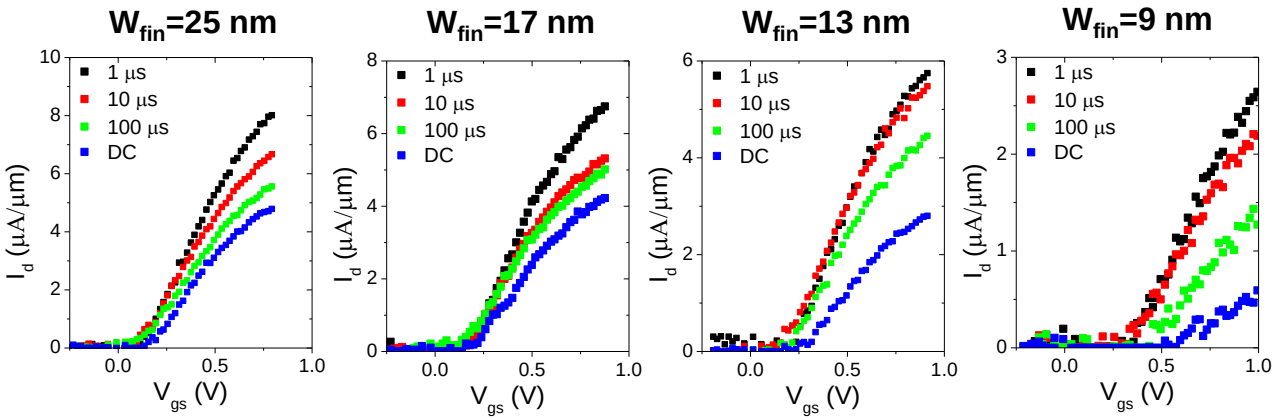

Figure 5.20: Pulsed and DC transfer characteristics (V_{ds} = 0.05 V) for InGaAs FinFETs with various fin width (L_g~5 μm).

Capacitance-Voltage Characteristics

Gate capacitance C_g is extracted at 1 GHz from S-parameter measurements, to remove the effect of parasitic resistance. Fin capacitance C_{fin} is extracted from the slope of C_g per fin vs. L_g using

devices with L_g = 120 nm to 2 μm, to remove the parasitic capacitance. Figure 5.21 compares CV characteristics for different W_f at 4 MHz and at 1 GHz. For all fin widths, as AC frequency increases from 4 MHz to 1 GHz, C_{fin} decreases by as much as 30%. This a similar observation as in planar devices and again confirms oxide trapping at MHz frequencies in InGaAs FinFETs. In the ON-state, as W_f decreases, C_{fin} decreases, an opposite trend compared to planar devices. This is an effect of electron population transitioning from surface inversion to volume inversion as W_f narrows. As the double gate starts to share control over one channel, the gate capacitance drops from $2(\frac{1}{C_{ox}}+\frac{1}{C_s})^{-1}$ to $2(\frac{1}{C_{ox}}+\frac{2}{C_s})^{-1}$ [91].

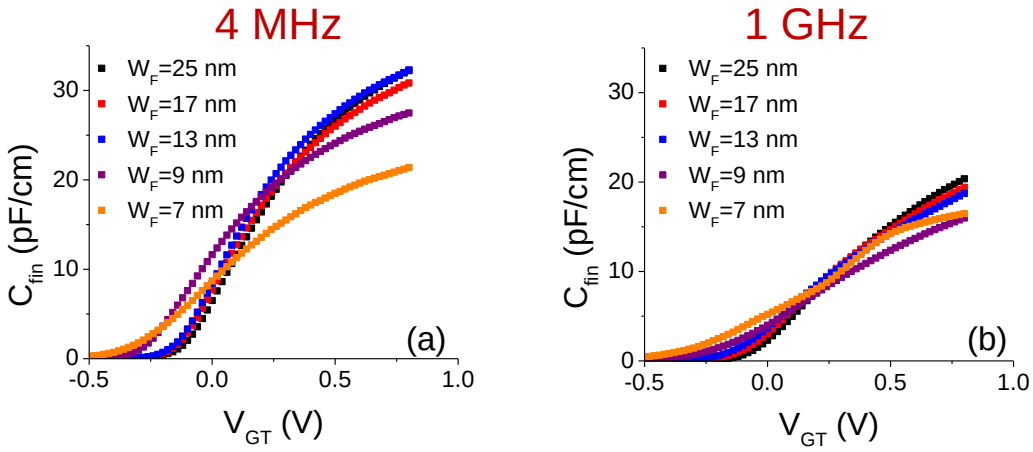

Figure 5.21: Capacitance-voltage characteristics in InGaAs FinFETs of different W_f, measured at (a) 4 MHz and (b) 1 GHz.

Transconductance characteristics

Figure 5.22 shows the intrinsic transconductance $g_{m,i}$ characteristics in the linear regime (V_{DS} = 0.05 V) measured at DC and 1 GHz in long channel devices with different W_f. Transconductance is normalized by the conducting gate periphery. At DC, we observe a rapid degradation in transconductance as fin width narrows or at high gate voltage bias (shown in Figure 5.22(a)). In contrast, when we modulate the AC signal at GHz frequency, the transconductance is enhanced by

3-6 times. This confirms that DC device performance is severely suppressed by oxide trapping in InGaAs FinFETs, similar to the observation in planar InGaAs MOSFETs. At 1 GHz, $g_{m,i}$ characteristics in devices with W_f = 7 and 9 nm have not reached their peaks given the measurement voltage range. This indicates no evident mobility degradation observed yet. In addition, at 1 GHz, the degradation of $g_{m,i}$ with W_f is much slower compared to at DC.

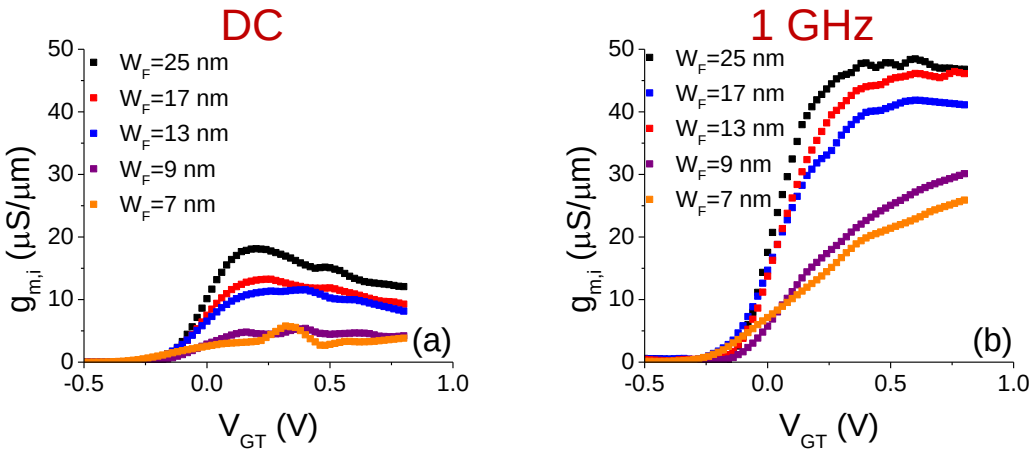

Figure 5.22: Transconductance characteristics in long channel InGaAs FinFET devices with different W_f, measured with V_{DS} = 0.05 V at (a) DC and (b) 1 GHz.

Figure 5.23 compares the scaling of maximum extrinsic transconductance ($g_{m,max}$ at V_{DS} = 0.5 V) with fin width in short-channel InGaAs FinFETs (L_g~60 nm) at DC and 1 GHz. The enhancement of $g_{m,max}$ from DC to 1 GHz is most striking for narrow fins, where 1 GHz $g_{m,max}$ can be 5-times higher than DC and becomes competitive with existing Silicon FinFETs [14] even at a reduced operating voltage and longer L_g. On the other hand, at 1 GHz, we also observe severe degradation as W_f falls below 13 nm, which is an earlier onset compared to planar devices (t_c = 5 nm). The reason for this requires further studies.

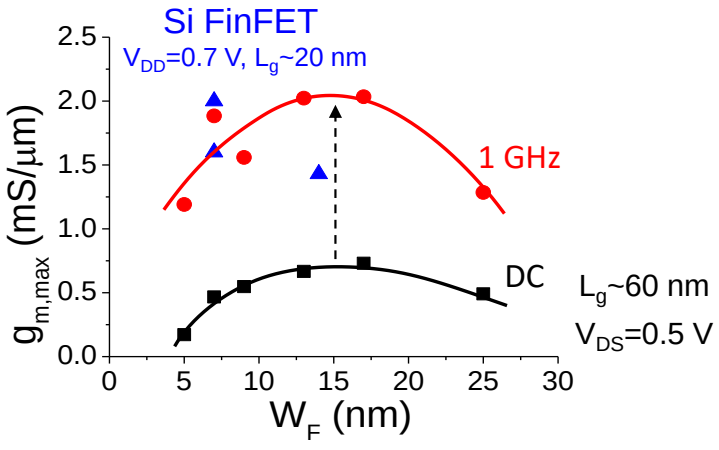

Figure 5.23: Fin width dependence of $g_{m,max}$ (V_{DS} = 0.5 V) in InGaAs FinFETs with L_g = 60 nm at DC (black) and 1 GHz (red). Blue: $g_{m,max}$ (V_{DD} = 0.7 V) in silicon FinFET with L_g = 20 nm.

Carrier Density

Having verified the RF-I_D technique in planar InGaAs MOSFETs, now we can confidently apply it to InGaAs FinFETs and extract carrier density N_s. Here, N_s is normalized by the channel height (H_c) and gate length (L_g) per fin. To facilitate our understanding, 2D P-S simulations in the fin cross-section [34] are shown in Figure 5.24 with W_f = 9 and 25 nm. For wide fins, electrons populate on both sides of the fins forming two channels. Thus, we expect N_s to be twice as in

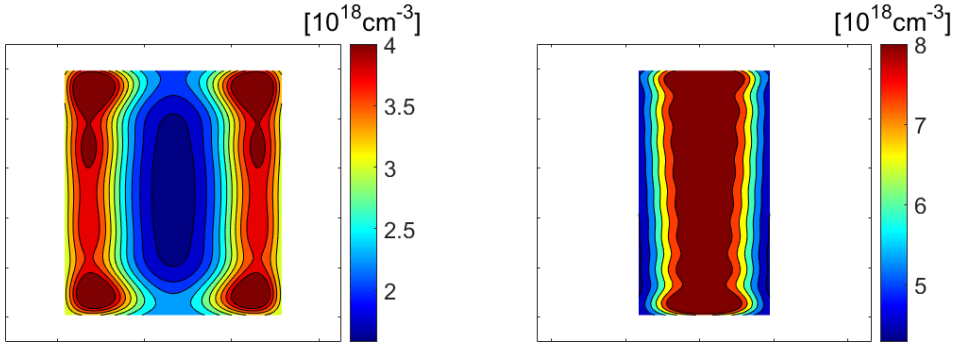

Figure 5.24: Charge distribution in W_f = 25 nm (left) and W_f = 9 nm (right) fins during the ON state ($N_s = 6 \times 10^{12}$ cm^{-2}) from P-S simulation, adopted from [34].

equivalent planar devices. As the fin narrows, electrons start conducting through the body of the fin, forming one channel in the center of the body as a result of volume inversion.

Figure 5.25 shows that N_s extracted from RF-I_D method decreases as W_f narrows, consistent with the qualitative understanding provided by P-S simulations. In contrast, N_s estimated by conventional CV method overlap on top of each other for various W_f, and the scaling behavior with W_f is masked by oxide trapping. In addition, the over-estimation of N_s by conventional CV continue to worsen as frequency decreases from 1 GHz to 4 MHz.

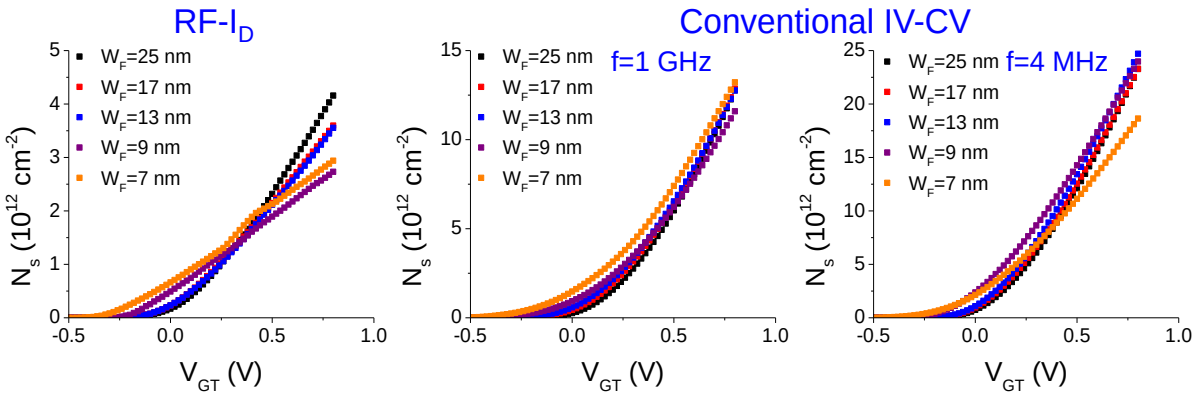

Figure 5.25: N_s vs. V_{GT} in InGaAs FinFETs of different W_f using: from left to right, RF-I_D method, classic IV-CV with CV measured at 1 GHz and 4 MHz. Note the vertical scales are different.

Mobility

We further extract mobility using RF-I_D methods for various fin widths and compare the results with conventional IV-CV methods in Figure 5.26. Similar to what we have observed in planar devices, mobility is grossly underestimated using conventional IV-CV method. This is still true even when we increase CV frequency from 4 MHz to1 GHz. As W_f narrows, the impact of oxide trapping is even more severe, resulting in an incorrect suggestion of severe mobility degradation.

In contrast, with the new method to accurately estimate mobility, mobility in extremely-scaled InGaAs fins is still very promising. At W_f = 7 nm, mobility as high as 570 $cm^2/V \cdot s$ is obtained.

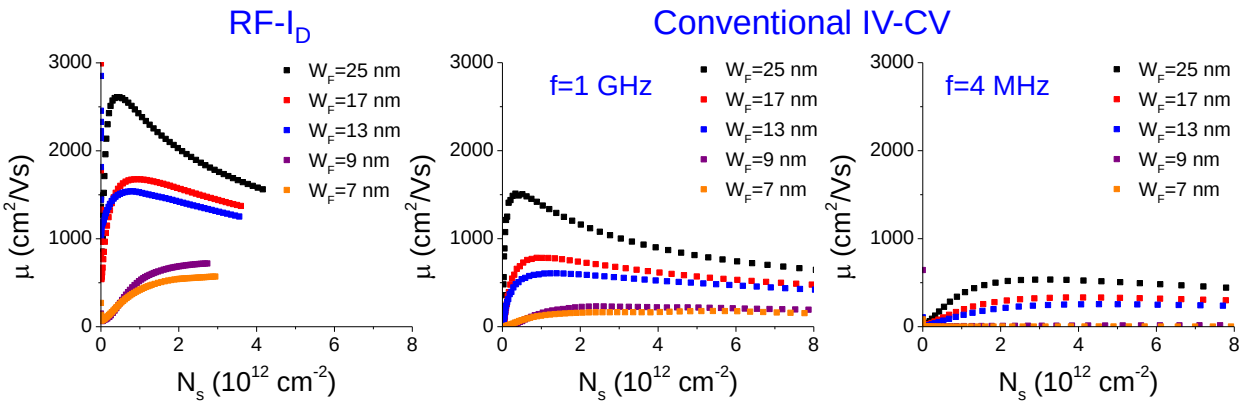

Figure 5.26: Mobility vs. N_s in InGaAs FinFETs of different W_f using: from left to right, RF-I_D method, classic IV-CV with CV measured at 1 GHz and 4 MHz.

Figure 5.27 shows the W_f dependence of estimated mobility at $N_s \sim 1.7\times10^{12}$ cm^{-2} using the conventional CV and RF-I_D methods in InGaAs FinFETs. With the new method, we find that InGaAs FinFETs exhibit very promising inherent potential and scalability, significantly better than what is suggested using conventional IV-CV.

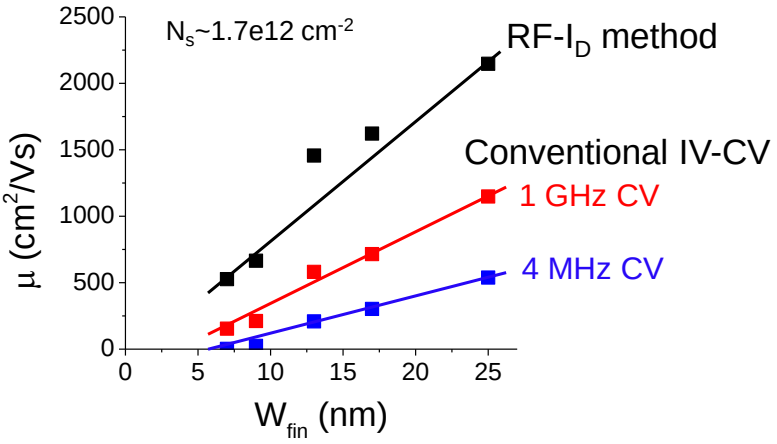

Figure 5.27: Mobility (at N_s ~1.7 × 10^{12} cm^{-2}) versus W_f in InGaAs FinFETs, comparing RF-I_D (black) and conventional IV-CV method with CV taken at 1 GHz (red) and 4 MHz (blue).

5.5 Oxide Traps

5.5.1 Estimation of Trap Density

One additional important benefit of the RF-I_D technique is that it allows us to quantify the amount of oxide trap density N_{ox}. From the difference in extracted carrier concentration between CV (4 MHz) and RF-I_D methods, oxide trap density N_{ox} can be estimated. Figure 5.28 here shows N_{ox} versus mobile carrier density N_s in InGaAs FinFETs with various W_f. We find that N_{ox} is ~ 3-5 times higher than N_s. In addition, N_{ox} increases as W_f scales down. This can be attributed to a combination of severe oxide trapping and the loss of channel carriers due to volume inversion in narrow-fin devices.

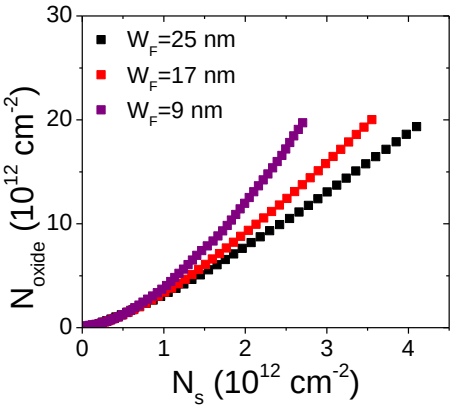

Figure 5.28: Oxide trap density N_{ox} vs. free carrier density N_s in the channel for different fin widths. N_{ox} is estimated from the difference in extracted carrier concentration between CV (4 MHz) and RF-I_D methods.

5.5.2 Mitigation

Many approaches in the literature have shown effectiveness in improving interface/oxide quality in InGaAs, such as sulfur-based surface passivation [92], H_2/N_2 plasma treatment [93]–[95], annealing in H-containing environment [96], [97], interfacial layer between the semiconductor

channel and high-k dielectric [43], [98], [99]. Here, we use forming gas annealing (FGA). We perform sequential FGA with increasing temperature on the same InGaAs FinFET device with W_f = 9 nm and L_g = 5 μm. Its transfer characteristics (V_{ds} = 0.05 V) under pulse and DC conditions are measured before and after each anneal, shown in Figure 5.29. As we anneal the device and increase the FGA temperature from 250 °C to 300 °C, the drain current consistently increases and the device shows less dispersion. This indicates less severe oxide trapping, mitigated by FGA. Whether the reduction in oxide traps have also resulted in less Coulomb scattering and further contributed to the drain current enhancement, requires further mobility analysis. Nonetheless, the significant device performance enhancement through FGA anneal suggests tremendous potential for improvement of InGaAs MOSFET technology.

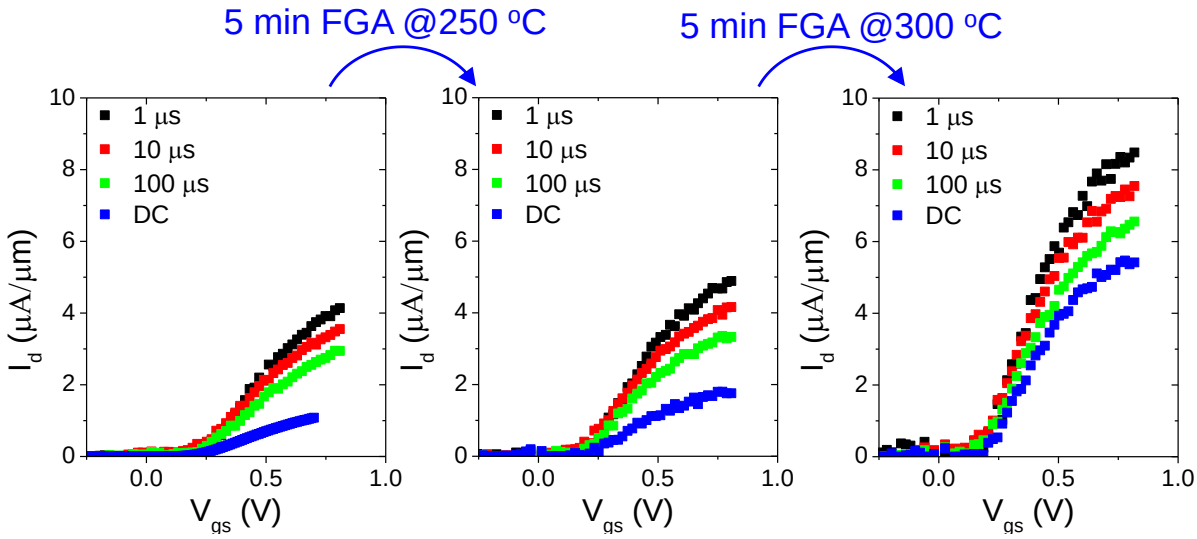

Figure 5.29: Transfer characteristics (V_{ds} = 0.05 V) of InGaAs FinFET device (W_f = 9 nm, L_g = 5 μm) under pulse (1 μs, 10 μs, 100 μs) and DC conditions, before and after 5 min forming gas annealing (FGA) at 250 °C and at 300 °C.

5.6 Summary

In this chapter, we have first established a methodology (RF-I_D) to accurately estimate mobility and charge control relationship that is immune to the impact of oxide traps. This new method is verified to be in excellent agreement with Hall mobility measurements and Poisson-Schrodinger simulations in planar InGaAs MOSFETs.

We use RF measurement techniques to isolate the intrinsic characteristics in InGaAs MOSFETs free from the influence of oxide trapping. Together with RF-I_D method, we study the impact of intrinsic channel scaling on device performance and transport in InGaAs planar MOSFETs and FinFETs. In both cases, we find much less mobility degradation than indicated by conventional CV methods. Rather than mobility, DC device performance is inhibited by the limited N_s due to severe oxide trapping. When the impact of oxide trapping is avoided, we find promising potential in InGaAs MOSFETs. In particularly, InGaAs FinFETs exhibits $g_{m,max}$ at 1 GHz competitive with current Silicon FinFET technology and high mobility even in narrow fins (μ_{peak} ~ 570 $cm^2/V \cdot s$ at W_f = 7 nm).

Preliminary experiments using FGA indicate that oxide trapping could be reduced by process improvements. This suggests the importance of mitigating oxide trapping and reassessing the future of InGaAs technology.

CHAPTER 6. Conclusions and Suggestions for Future Work

6.1 Thesis Summary

This thesis addresses some of the critical challenges facing InGaAs MOSFETs and advances the understanding of the limiting factors of InGaAs MOSFET technology.

It reports, for the first time, a prominent but fully reversible transconductance enhancement after applying positive gate stress in InGaAs MOSFETs. This is in contrast to the well-established PBTI phenomena. A new instability mechanism is identified as electric-field-induced migration of fluorine ions that are introduced during the RIE gate recess process. F is known to passivate Si donors in InAlAs. In our device structure, an n-InAlAs ledge facilitates the link from the contacts to the intrinsic device. To independently confirm the presence of F in our device and induced donor passivation, secondary ion mass spectroscopy (SIMS) and transmission line model (TLM) structures are used. Comprehensive voltage stress experiments have been carried out to thoroughly understand the F-induced device instability, including forward-gate stress, off-state stress, and temperature-dependent stress. The understanding derived has lead us to redesign our InGaAs MOSFETs by eliminating n-InAlAs layers and instead use n-InP ledges. The new device design not only exhibits greatly improved electrical stability, but also record performance.

With the new device structure immune to fluorine, the impact of oxide trapping in InGaAs MOSFETs is evaluated. While the use of ALD has greatly improved the high-k/III-V interface

quality due to the self-cleaning effect, severe electron trapping in oxide traps is still found in InGaAs MOSFETs, particularly in thin-channel devices. A comprehensive PBTI study shows that oxide trapping deteriorates device stability. Prolonged voltage stress leads to electron trapping in the pre-exiting defects in the oxide, resulting in threshold voltage shifts and degraded device performance. Under harsh stress conditions, permanent damage can occur through the generation of interface states in the oxide. Compared to silicon, III-V/high-k gate stacks show excessive charge trapping at lower operating voltages. In addition to deteriorating device stability, oxide trapping also causes frequency dispersion and time-dependent variability, and jeopardizes DC device performance. High frequency and pulse measurements reveal a rich spectrum of oxide traps with different capture/emission times. When the impact of the border traps is avoided, InGaAs MOSFETs show promising performance. Furthermore, oxide trapping complicates the extraction of fundamental parameters in in InGaAs MOSFETs and can lead to severe underestimation of channel mobility. Understanding that oxide trapping significantly obscures the intrinsic potential of InGaAs MOSFETs presses the need to reassess the potential of InGaAs technology.

DC device performance degradation of InGaAs FinFETs is observed as they scale to sub-10 nm fin width. This is often attributed to degradation in intrinsic transport parameters. Understanding the complication of oxide trapping has motivated this thesis to develop a new mobility extraction method, immune to the impact of oxide traps. It enables accurate estimation of the correct charge-control relationship, and hence mobility. The impact of intrinsic channel scaling on device performance and transport in InGaAs planar MOSFETs and FinFETs is re-evaluated. In both cases, much less mobility degradation is observed than indicated by conventional CV methods. Rather than severe transport degradation, this thesis suggests that performance degradation of InGaAs

FinFETs is largely an extrinsic phenomenon caused by severe oxide trapping. Notably, InGaAs FinFETs exhibits $g_{m,max}$ at 1 GHz competitive with current Silicon FinFET technology and high mobility even in narrow fins (μ_{peak} ~ 570 cm^2/V·s at W_f = 7 nm). Preliminary experiments using FGA indicate that oxide trapping could be further reduced by process improvements.

In conclusion, this thesis finds that due to oxide trapping, the potential of scaled InGaAs MOSFETs has been misjudged and grossly underestimated. Performance degradation has been mistakenly attributed to transport loss. This highlights the importance of mitigating oxide trapping, and that the future of InGaAs technology should be reassessed.

6.2 Suggestions for Future Work

6.2.1 TCAD Modeling

TCAD modeling, calibrated against experimental data, plays an important role in predicting the performance of highly scaled devices. Unfortunately, our findings in this thesis suggest that existing TCAD studies of InGaAs have been calibrated against incorrect transport data. A valuable contribution would be to recalibrate existing TCAD models using the corrected charge-control and mobility data of this thesis to reassess the scalability of InGaAs MOSFETs.

In addition, this thesis still finds that performance and transport degrades with scaling in planar devices and in FinFETs. A recalibrated TCAD model would be useful to quantify the contributions of different scattering mechanisms and quantization effects to this degradation as dimensions

shrink. Furthermore, such a TCAD model could also facilitate a comparison between thin-channel planar devices and thin-body FinFETs in order to study the impact of excessive fin etching on FinFET performance.

6.2.2 Improved gate oxide technology

Knowing the detrimental impact of oxide traps on InGaAs MOSFET performance and reliability, process technologies need to be developed to improve the quality of the III-V/high-k semiconductor interface and the high-k dielectric material. The preliminary results of this thesis generated using FGA suggest that this is possible. Another successful effort is the recent development of interfacial layers (IL) and use of IL/$LaSiO_x$/HfO_2 gate stacks to significantly reduce defect density [98], [99]. Another innovation is the incorporation of *in-situ* ALE and ALD processes in InGaAs FinFETs by Lu et al. [35], which has significantly improved device performance. The contribution of reduced oxide trap density using this technique yet needs to be quantified, as the *in-situ* ALE-ALD process avoids air exposure before the gate dielectric deposition.

6.2.3 Semi-analytical modeling for border trap extraction

One extension of this work is to develop a semi-analytical model of border traps in terms of their energy and spatial distribution. This model should be able to explain the experimental observations, including the frequency dispersion, temperature dependence, and voltage stretch-out. Such a model would enable quantifiable comparisons of oxide quality between different gate-stack technologies.

Appendix A: Process Flow of InGaAs MOSFETs

Module	Step	Specification	Recipe	Tool
Contact	Clean	HCl:H_2O (1:3)	30 sec	Acidhood
	Metal dep	Mo 20 nm W 10 nm	XC MO gun 2 test (~0.8 Å/s) linjq w gun3 run (~0.9 Å/s)	TRLAJA
	Oxide dep	SiO_2 60 nm	HFSIO_WL (~85 nm/min)	STSCVD
Gate recess	EBL	ZEP: Anisol	Bake 120°C 3 min Spin 0.7/3.5 krpm, 3/60 sec Bake 180°C 3min	Hotplate Coater Hotplate
		Exposure	DAC600/6e4/DT0.6/1nA (dose 800 μC/cm^2)	Elionix
		Develop	Xylene 1 min and IPA 1 min	Photohood
	Etching	SiO2 etch	CF4 AV (~45 nm/min)	Oxford-100
		Mo/W etch	WJ-SF6/O2-50W (Mo/W:~15 nm/min) (SiO_2:~7.5 nm/min)	Oxford-100
		SEM		SEM
	Resist strip	NMP	Overnight or 2 hr heated water bath	Photohood
		Asher	700 W (~1 μm/40min)	TRLAsher
Mesa	Positive photo	HMDS	program 1	HMDS
		SPR 700	Spin 0.7/3.3krpm, 6/60 sec Bake 80°C 3 min	Coater Hotplate
		Exposure	laser 375, dose 170 mJ/cm^2	MLA
		Develop	MF-CD-26 1 min15 sec	Photohood
	Etching	SiO_2 etch	CF4 AV (~45 nm/min)	Oxford-100
		Mo/W etch	WJ-SF6/O2-50W (Mo/W:~15 nm/min) (SiO_2:~7.5 nm/min)	Oxford-100
		III-V etch (120 °C)	InGaAsJQ/RCP67 (~12 nm/min)	SAMCO
	Resist strip	NMP	Overnight or 2 hr heated water bath	Photohood
		Asher	700 W (~1 μm/40min)	TRLAsher

III-V recess	Etching	III-V dry etch (120 °C)	InGaAsJQ/RCP67 (~12 nm/min)	SAMCO
		III-V wet etch	citric acid: H_2O_2 = 20:1 (~20 nm/min)	Acidhood
		SEM		SEM
	Digital etch	Oxidation	O_2 plasma, 850 W 2:30 min	TRLAsher
		Oxide removal	H2SO4:H2O=1:1 25 sec	Acidhood
Gate stack	ALD	HfO_2	250°C (~1 Å/cycle)	ICL-ALD
	Metal dep	Mo 35 nm	XC MO gun 2 test (~0.8 Å/s)	TRLAJA
	Image reversal photo	AZ5214	Bake 120°C 2 min Spin 0.75/3.75 krpm, 6/30 sec Bake 80°C 2 min	Hotplate Coater Hotplate
		Exposure	laser 405, dose 22 mJ/cm^2 Bake 114°C 1 min flood 40 sec	MLA Hotplate OAI-Flood
		Develop	AZ422 1:30 min	Photohood
	Metal dep	Ti/Au (18/200 nm)	Ti: 1 Å/s, Au: 2 Å/s	EbeamFP
	Liftoff	Acetone	> 3 hr	Photohood
	Etching	Mo etch	WJ-SF6/O2-50W (~15 nm/min)	Oxford-100
Pad	Image reversal photo	AZ5214	Bake 120°C 2 min Spin 0.75/3.75 krpm, 6/30 sec Bake 80°C 2 min	Hotplate Coater Hotplate
		Exposure	laser 405, dose 22 mJ/cm^2 Bake 114°C 1 min flood 40 sec	MLA Hotplate OAI-Flood
		Develop	AZ422 1:30 min	Photohood
	Etching	HfO_2 etch (RT)	InGaAsJQ2/RCP59 1:20 min for 4 nm	SAMCO
		SiO_2 etch	BOE 30 sec (~300 nm/min)	Oxford-100
	Metal dep	Ti/Au (10/180 nm)	Ti: 1 Å/s, Au: 2 Å/s	EbeamFP
	Liftoff	Acetone	> 3 hr	Photohood

Recipes in the process

SiO_2 Etch Tool: Oxford-100 Recipe: CF4-AV Gas flow: CF_4 = 50 sccm Pressure: 10 mTorr Power: 250 W Rate: 45 nm/min (calibration needed)	**Mo/W Etch** Tool: Oxford-100 Recipe: RIE-SF6-WJ-50W Gas flow: SF_6/O_2 = 45/5 sccm Pressure: 20 mTorr Power: 50 W Rate: 15 nm/min (calibration needed)
High-k Dielectric (Al_2O_3/HfO_2) Etch Tool: SAMCO Recipe: 59 Gas flow: BCl_3/Ar = 9/2 sccm Substrate temperature: 40°C ICP/Bias power: 75/75 W Pressure: 0.3 Pa	**III-V Dry Etch** Tool: SAMCO Recipe: 67 Gas flow: Cl_2/N_2 = 10/3 sccm Substrate temperature: 120°C ICP/Bias power: 20/50 W Pressure: 0.2 Pa Rate: 12 nm/min (calibration needed)
Mo Sputtering Tool: AJA-TRL Recipe: XC_Mo_gun2_test Loadlock pressure < $3\cdot10^{-6}$ Torr Main chamber pressure < $3\cdot10^{-6}$ Torr Plasma strike power: 150 W Deposition power: 100 W Rate: 0.8 Å/s (calibration needed)	**W Sputtering** Tool: AJA-TRL Recipe: linjq_W_gun3_Run Loadlock pressure < $3\cdot10^{-6}$ Torr Main chamber pressure < $3\cdot10^{-6}$ Torr Plasma strike power: 150 W Deposition power: 120 W Rate: 0.9 Å/s (calibration needed)
SiO_2 CVD Tool: STS-CVD Recipe: HFSIO_WL Stabilization: 4 min 30 s Pressure: 900 mTorr Gas flow: $N_2O/N_2/SiH_4$= 1420/392/10 sccm Power: 50 W Plate temperature: 300°C Showerhead temperature: 250°C Rate: 85 nm/min (calibration needed)	

Appendix B: RF Parasitic Extraction Procedure

Continuing the discussion in section 4.3.1.2, this appendix explains in detail the procedure to de-embed the pad parasitics, extract extrinsic device parasitics and calculate intrinsic device elements.

B.1 Pad Parasitics De-embedding

We begin by referring to Figure B.1 [100], which shows the two-port equivalent models for the DUT, and open/short de-embedding structures. To remove the effect of the pad capacitances and inductances, the measured S-parameters of the devices-under-test (DUTs) are de-embedded using the on-wafer open and short structures.

Measurement of these three structures yields 3 sets of two-port S-matrices, S_{DUT}, S_{open}, and S_{short}, which may be converted into equivalent two-port admittance matrices Y_{DUT}, Y_{open}, and Y_{short}. The pad parasitics can then be de-embedded from the device using

$$Z_A = Z\left[Y_{\text{DUT}} - Y_{\text{open}}\right] - Z\left[Y_{\text{short}} - Y_{\text{open}}\right] \tag{B.1}$$

where Z_A is the de-embedded device impedance matrix, and $Z[Y]$ represents the equivalent impedance matrix of the admittance matrix Y.

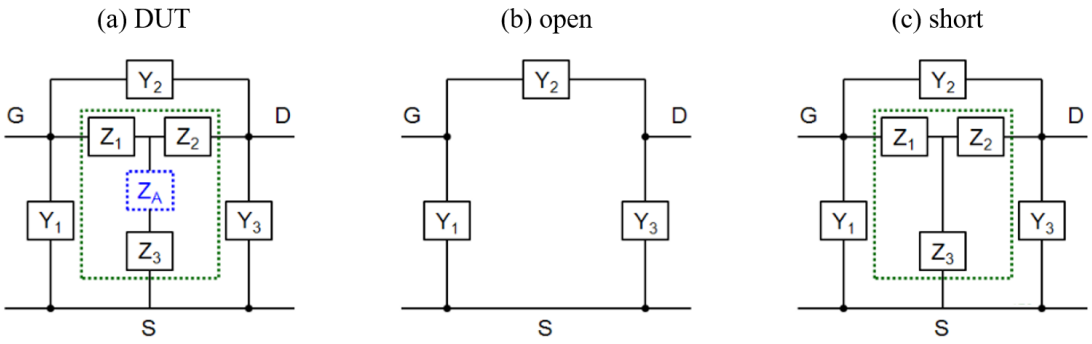

Figure B.1: Impedance model for (a) DUT, (b) open structure and (c) short structure. The active region of the device is represented by Z_A, adapted from [100].

The de-embedding is not perfect for all DUTs, as the pad geometries are slightly different among the short structure and DUTs with different gate length, specifically the size of the gate pad and the distance between the source and drain pads. Nonetheless, after de-embedding, most of the pad parasitics have been removed.

B.2 Extrinsic Device Parasitics Extraction

After de-embedding the effect of pad parasitics, the active region of the device can be represented by a circuit model shown in Figure B.2. Here, we describe the procedures to extract and remove the extrinsic device parasitics (capacitances, resistances, and inductances) [101]–[103]. The procedure uses de-embedded cold FET *S*-parameter data. When biased under "cold" condition (V_{DS} = 0 V) without current flow, the device is dominated by the extrinsic device parasitics.

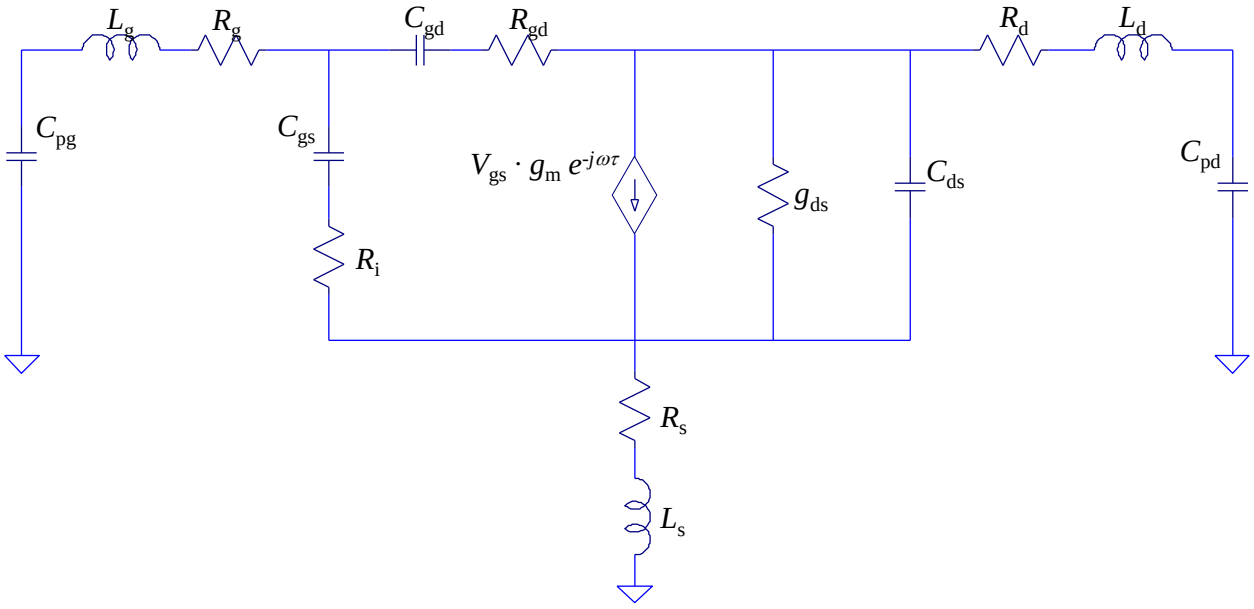

Figure B.2: Small-signal equivalent circuit model used for DUT de-embedding of the extrinsic device parasitics.

Extrinsic device capacitances

A cold conditions below threshold (V_{DS} = 0 V, V_{GS} << V_t), the imaginary part of the *Y*-parameters at low frequencies are dominated by parasitic capacitances. They can be approximated as

$$\mathrm{Im}\{Y_{11}\} = \omega(C_{pg} + 2C_b) \tag{B.2}$$

$$\mathrm{Im}\{Y_{12}\} = \mathrm{Im}\{Y_{21}\} = -\omega C_b \tag{B.3}$$

$$\mathrm{Im}\{Y_{22}\} = \omega(C_{pg} + C_b) \tag{B.4}$$

C_b represents the overlap capacitance between the gate and source/drain that exists even in the OFF-state. Parasitic capacitances can be therefore determined from the slope of $\mathrm{Im}\{Y_{ij}\}$ vs. angular frequency ω, with frequencies below 2 GHz. The impact of parasitic capacitances can then be subtracted out using

$$Y' = \begin{bmatrix} Y_{11} - j\omega C_{pg} & Y_{12} \\ Y_{21} & Y_{22} - j\omega C_{pd} \end{bmatrix} \tag{B.5}$$

Extrinsic device resistances

At cold conditions below threshold (V_{DS} = 0 V, V_{GS} << V_t), the real part of the *Z*-parameters at low frequencies are dominated by parasitic resistances and can be approximated as

$$\mathrm{Re}\{Z_{11}\} = R_g + R_s \tag{B.6}$$

$$\mathrm{Re}\{Z_{12}\} = \mathrm{Re}\{Z_{21}\} = R_s \tag{B.7}$$

$$\mathrm{Re}\{Z_{22}\} = R_d + R_s \tag{B.8}$$

Parasitic resistances can be determined from the slope of $\mathrm{Re}\{Z_{ij}\} \times \omega^2$ vs. ω^2, with frequency range of 3-10 GHz. The impact of parasitic resistances can be subtracted out according to

$$Z' = \begin{bmatrix} Z_{11} - R_g - R_s & Z_{12} - R_s \\ Z_{21} - R_s & Z_{22} - R_d - R_s \end{bmatrix} \quad \text{(B. 9)}$$

Extrinsic device inductances

At cold conditions above threshold (V_{DS} = 0 V, V_{GS} >> V_t), the imaginary part of the *Z*-parameters at high frequencies are dominated by parasitic inductances and can be approximated as

$$\omega \text{Im}\{Z_{11}\} = \omega^2\left(L_g + L_s\right) - \frac{1}{C_g} \quad \text{(B. 10)}$$

$$\text{Im}\{Z_{12}\} = \text{Im}\{Z_{21}\} = \omega L_s \quad \text{(B. 11)}$$

$$\text{Im}\{Z_{22}\} = \omega(L_d + L_s) \quad \text{(B. 12)}$$

Parasitic inductances can be determined from the slope of $\text{Im}\{Z_{ij}\} \times \omega$ vs. ω^2, with frequencies above 30 GHz. The impact of parasitic inductances can be subtracted out using

$$Z' = \begin{bmatrix} Z_{11} - j\omega\left(L_g + L_s\right) & Z_{12} - j\omega L_s \\ Z_{21} - j\omega L_s & Z_{22} - j\omega(L_d + L_s) \end{bmatrix} \quad \text{(B. 13)}$$

Advanced Design System (ADS) software can be used to further optimize the extraction of the parasitic components by minimizing the error between the measured and simulated *S*-parameters. Afterwards, the parasitic components can be removed using equations B.5, B.9 and B.13,

B.3 Intrinsic device parameter extraction

At this point, the *Z*-parameters represent only the intrinsic portion of the device and can be converted to *Y*-parameters. The admittance matrix elements Y_{ij} can be expressed using the intrinsic circuit elements as

$$\mathrm{Y}_{11} = \frac{i_1}{v_1}|_{v_2=0} = \frac{R_i C_{gs}^2 \omega^2}{D_1} + \frac{R_{gd} C_{gd}^2 \omega^2}{D_2} + j\omega\left(\frac{C_{gs}}{D_1} + \frac{C_{gd}}{D_2}\right) \tag{B.14}$$

$$\mathrm{Y}_{12} = \frac{i_1}{v_2}|_{v_1=0} = -\frac{R_{gd} C_{gd}^2 \omega^2}{D_2} - j\omega\frac{C_{gd}}{D_2} \tag{B.15}$$

$$\mathrm{Y}_{21} = \frac{i_2}{v_1}|_{v_2=0} = \frac{g_m e^{-j\omega\tau}}{1 + j\omega R_i C_{gs}} - j\frac{\omega C_{gd}}{1 + j\omega R_{gd} C_{gd}} \tag{B.16}$$

$$\mathrm{Y}_{22} = \frac{i_2}{v_2}|_{v_1=0} = \frac{R_{gd} C_{gd}^2 \omega^2}{D_2} + j\omega\left(C_{ds} + \frac{C_{gd}}{D_2}\right) + g_0 \tag{B.17}$$

where the denominator terms are given by

$$D_1 = 1 + R_i^2 C_{gs}^2 \omega^2 \tag{B.18}$$

$$D_2 = 1 + R_{gd}^2 C_{gd}^2 \omega^2 \tag{B.19}$$

Considering the real and imaginary components of *Y* separately, equations B.14 through B.19 represent a system of 8 equations with 8 unknowns, the unknowns being the intrinsic device parameters. After some algebra, we can express the intrinsic device parameters from the admittance data of the intrinsic device according to

$$C_{gs} = -\frac{1}{\omega \mathrm{Im}\left(\frac{1}{Y_{11} + Y_{12}}\right)} \tag{B.20}$$

$$C_{gd} = \frac{1}{\omega \mathrm{Im}\left(\frac{1}{Y_{12}}\right)} \tag{B.21}$$

$$R_i = \mathrm{Re}\left(\frac{1}{Y_{11} + Y_{12}}\right) \tag{B.22}$$

$$R_{gd} = -\mathrm{Re}\left(\frac{1}{Y_{12}}\right) \tag{B.23}$$

$$\tau = -\frac{1}{\omega}\angle\left\{(Y_{21} - Y_{12})\left[1 + \frac{j\mathrm{Re}(Y_{11} + Y_{12})}{\mathrm{Im}(Y_{11} + Y_{12})}\right]\right\} \tag{B.24}$$

$$g_m = \left| \frac{(Y_{11} + Y_{12})(Y_{21} - Y_{12})}{\mathrm{Im}(Y_{11} + Y_{12}) \exp(-j\omega\tau)} \right| \quad \text{(B. 25)}$$

$$g_0 = \mathrm{Re}(Y_{22} + Y_{12}) \quad \text{(B. 26)}$$

$$C_{ds} = \frac{\mathrm{Im}(Y_{22} + Y_{12})}{\omega} \quad \text{(B. 27)}$$

Bibliography

[1] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronics*, vol. 38, no. 8, p. 114, 1965.

[2] “Moore’s Law.” [Online]. Available: https://archive.education.mrsec.wisc.edu/SlideShow/images/computer/Moores_Law.png. [Accessed: 18-May-2019].

[3] D. A. Antoniadis, I. Aberg, C. Ni Chleirigh, O. M. Nayfeh, A. Khakifirooz, and J. L. Hoyt, “Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations,” *IBM J. Res. Dev.*, vol. 50, no. 4.5, pp. 363–376, 2006.

[4] J. A. del Alamo, “Nanometre-scale electronics with III–V compound semiconductors,” *Nature*, vol. 479, no. 7373, pp. 317–323, 2011.

[5] H. Riel, L. E. Wernersson, M. Hong, and J. A. del Alamo, “III-V compound semiconductor transistors - From planar to nanowire structures,” *MRS Bull.*, vol. 39, no. 8, pp. 668–677, 2014.

[6] “Semiconductors.” [Online]. Available: http://www.ioffe.ru/SVA/NSM/Semicond/. [Accessed: 03-May-2019].

[7] A. Khakifirooz and D. A. Antoniadis, “Transistor Performance Scaling: The Role of Virtual Source Velocity and Its Mobility Dependence,” in *IEDM Tech. Dig.*, 2006, pp. 1–4.

[8] J. A. del Alamo, “Nanometer-scale InGaAs field-effect transistors for THz and CMOS technologies,” *Eur. Solid-State Device Res. Conf.*, pp. 16–21, 2013.

[9] J. A. del Alamo, “The High-Electron Mobility Transistor at 30: impressive accomplishments and exciting prospects,” *CS MANTECH*, pp. 17–22, 2011.

[10] X. Mei *et al.*, “First Demonstration of Amplification at 1 THz Using 25-nm InP High Electron Mobility Transistor Process,” *IEEE Electron Device Lett.*, vol. 36, no. 4, pp. 327–329, 2015.

[11] D. H. Kim *et al.*, “High-speed E-mode InAs QW MOSFETs with Al2O3 insulator for future RF applications,” *IEEE Electron Device Lett.*, vol. 34, no. 2, pp. 196–198, 2013.

[12] J. A. del Alamo, X. Cai, J. Lin, W. Lu, A. Vardi, and X. Zhao, “CMOS beyond Si: Nanometer-Scale III-V MOSFETs,” *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meet.*, pp. 25–29, 2017.

[13] D. J. Frank, “Power-constrained CMOS scaling limits,” *IBM J. Res. Dev.*, vol. 46, no. 2.3, pp. 235–244, Mar. 2002.

[14] C. Auth *et al.*, “A 10nm High Performance and Low-Power CMOS Technology Featuring 3rd Generation FinFET Transistors, Self-Aligned Quad Patterning, Contact over Active Gate and Cobalt Local Interconnects,” *IEDM Tech. Dig.*, pp. 673–676, 2017.

[15] T. Ghani *et al.*, “A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors,” *IEDM Tech. Dig.*, pp. 11.6.1-11.6.3, 2003.

[16] C. Auth *et al.*, “45nm High-k + Metal Gate Strain-Enhanced Transistors,” *Proc. VLSI Tech.*, pp. 216–217, 2008.

[17] K. Mistry *et al.*, “A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging,” *IEDM Tech. Dig.*, pp. 247–250, 2007.

[18] C. Auth *et al.*, “A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors,” *Proc. VLSI Tech.*, pp. 131–132, 2012.

[19] T. Mimura, S. Hiyamizu, T. Fujii, and K. Nanbu, “A New Field-Effect Transistor with Selectively Doped GaAs / n-Al x Ga1- x As Heterojunctions,” *Jpn. J. Appl. Phys.*, vol. 19, no. 5, pp. L225–L227, 1980.

[20] C. L. Hinkle *et al.*, “GaAs interfacial self-cleaning by atomic layer deposition,” *Appl. Phys. Lett.*, vol. 92, p. 071901, 2008.

[21] L. Lin and J. Robertson, “Defect states at III-V semiconductor oxide interfaces,” *Appl. Phys. Lett.*, vol. 98, p. 082903, 2011.

[22] P. D. Ye *et al.*, “GaAs MOSFET with oxide gate dielectric grown by atomic layer deposition,” *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 209–211, 2003.

[23] M. M. Frank *et al.*, “HfO2 and Al2O3 gate dielectrics on GaAs grown by atomic layer deposition,” *Appl. Phys. Lett.*, vol. 86, p. 152904, 2005.

[24] J. Lin, Y. Wu, J. A. del Alamo, and D. A. Antoniadis, “Analysis of Resistance and Mobility in InGaAs Quantum-Well MOSFETs from Ballistic to Diffusive Regimes,” *IEEE Trans. Electron Devices*, vol. 63, no. 4, pp. 1464–1470, 2016.

[25] J. Lin, X. Zhao, D. A. Antoniadis, and J. A. del Alamo, “A novel digital etch technique for

deeply scaled III-V MOSFETs," *IEEE Electron Device Lett.*, vol. 35, no. 4, pp. 440–442, 2014.

[26] W. Lu, X. Zhao, D. Choi, S. El Kazzi, and J. A. del Alamo, "Alcohol-Based Digital Etch for III-V Vertical Nanowires with Sub-10 nm Diameter," *IEEE Electron Device Lett.*, vol. 38, no. 5, pp. 548–551, 2017.

[27] X. Zhao and J. A. del Alamo, "Nanometer-Scale Vertical-Sidewall Reactive Ion Etching of InGaAs for 3-D III-V MOSFETs," *IEEE Electron Device Lett.*, vol. 35, no. 5, pp. 521–523, 2014.

[28] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Sub-30 nm InAs Quantum-Well MOSFETs with self-aligned metal contacts and Sub-1 nm EOT HfO2 insulator," *IEDM Tech. Dig.*, pp. 1–4, 2012.

[29] J. Lin, X. Zhao, T. Yu, D. A. Antoniadis, and J. A. del Alamo, "A new self-aligned quantum-well MOSFET architecture fabricated by a scalable tight-pitch process," *IEDM Tech. Dig.*, pp. 421–424, 2013.

[30] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Novel intrinsic and extrinsic engineering for high-performance high-density self-aligned InGaAs MOSFETs: Precise channel thickness control and sub-40-nm metal contacts," *IEDM Tech. Dig.*, pp. 25.1.1-25.1.4, 2014.

[31] J. Lin, X. Cai, Y. Wu, D. A. Antoniadis, and J. A. del Alamo, "Record Maximum Transconductance of 3.45 mS/um for III-V FETs," *IEEE Electron Device Lett.*, vol. 37, no. 4, pp. 381–384, Apr. 2016.

[32] S. W. Chang *et al.*, "InAs N-MOSFETs with record performance of Ion = 600 μa/μm at Ioff = 100 nA/μm (Vd = 0.5 V)," *IEDM Tech. Dig.*, pp. 417–420, 2013.

[33] S. Lee *et al.*, "Highly Scalable Raised Source/Drain InAs Quantum Well MOSFETs Exhibiting Ion=482 uA/um at Ioff=100 nA/um and VDD=0.5 V," *IEEE Electron Device Lett.*, vol. 35, no. 6, pp. 621–623, 2014.

[34] A. Vardi *et al.*, "Self-Aligned InGaAs FinFETs with 5-nm Fin-Width and 5-nm Gate-Contact Separation," pp. 429–432, 2017.

[35] W. Lu *et al.*, "First Transistor Demonstration of Thermal Atomic Layer Etching: InGaAs FinFETs with sub-5 nm Fin-width Featuring in situ ALE-ALD," *IEDM Tech. Dig.*, pp. 39.1.1-39.1.4, 2018.

[36] M. Radosavljevic *et al.*, "Electrostatics improvement in 3-D tri-gate over ultra-thin body

planar InGaAs quantum well field effect transistors with high-K gate dielectric and scaled gate-to-drain/gate-to-source separation," *IEDM Tech. Dig.*, pp. 765–768, 2011.

[37] X. Zhao, J. Lin, C. Heidelberger, E. A. Fitzgerald, and J. A. del Alamo, "Vertical nanowire InGaAs MOSFETs fabricated by a top-down approach," *IEDM Tech. Dig.*, pp. 695–698, 2013.

[38] N. Waldron *et al.*, "Gate-all-around InGaAs nanowire FETS with peak transconductance of 2200μS/μm at 50nm Lg using a replacement Fin RMG flow," *IEDM Tech. Dig.*, pp. 3111–3114, 2016.

[39] J. J. Gu *et al.*, "20-80nm Channel length InGaAs gate-all-around nanowire MOSFETs with EOT=1.2nm and lowest SS=63mV/dec," *IEDM Tech. Dig.*, pp. 27.6.1-27.6.4, 2012.

[40] X. Zhou *et al.*, "Scalability of InGaAs gate-All-Around FET integrated on 300mm Si platform: Demonstration of channel width down to 7nm and Lg down to 36nm," *Proc. VLSI Tech.*, pp. 2–3, 2016.

[41] J. Wu, Y. Fang, B. Markman, H. Y. Tseng, and M. J. W. Rodwell, "Lg=30 nm InAs Channel MOSFETs Exhibiting fmax=410 GHz and ft=357 GHz," *IEEE Electron Device Lett.*, vol. 39, no. 4, pp. 472–475, 2018.

[42] A. Alian *et al.*, "Impact of the channel thickness on the performance of ultrathin InGaAs channel MOSFET devices," *IEDM Tech. Dig.*, pp. 437–440, 2013.

[43] J. Franco *et al.*, "Characterization of Oxide Defects in InGaAs MOS Gate Stacks for High-Mobility n- Channel MOSFETs (Invited)," *IEDM Tech. Dig.*, pp. 175–178, 2017.

[44] S. Takagi *et al.*, "MOS interface and channel engineering for high-mobility Ge/III-V CMOS," *IEDM Tech. Dig.*, pp. 505–508, 2012.

[45] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Off-State Leakage Induced by Band-to-Band Tunneling and Floating-Body Bipolar Effect in InGaAs Quantum-Well MOSFETs," *IEEE Electron Device Lett.*, vol. 35, no. 12, pp. 1203–1205, 2014.

[46] K. Ni *et al.*, "Single-event transient response of InGaAs MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 6, pp. 3550–3556, 2014.

[47] K. Ni *et al.*, "Gate Bias and Geometry Dependence of Total-Ionizing-Dose Effects in InGaAs Quantum-Well MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 64, no. 1, pp. 239–244, 2017.

[48] S. Deora *et al.*, "Positive bias instability and recovery in InGaAs channel nMOSFETs,"

IEEE Trans. Device Mater. Reliab., vol. 13, no. 4, pp. 507–514, 2013.

[49] N. Agrawal, A. V Thathachary, S. Mahapatra, and S. Datta, "Impact of Varying Indium (x) Concentration and Quantum Confinement on PBTI Reliability in InxGa1−xAs FinFET," *IEEE Electron Device Lett.*, vol. 36, no. 2, pp. 120–122, 2015.

[50] J. Franco *et al.*, "Suitability of high-k gate oxides for III-V devices: A PBTI study in In0.53Ga0.47As devices with Al2O3," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 1–6, 2014.

[51] S. Johansson, J. Mo, and E. Lind, "Characterization of border traps in III-V MOSFETs using an RF transconductance method," *Eur. Solid-State Device Res. Conf.*, pp. 53–56, 2013.

[52] G. Roll, E. Lind, M. Egard, S. Johansson, L. Ohlsson, and L. E. Wernersson, "RF and DC analysis of stressed InGaAs MOSFETs," *IEEE Electron Device Lett.*, vol. 35, no. 2, pp. 181–183, 2014.

[53] E. Cartier *et al.*, "Electron Mobility in Thin In0.53Ga0.47As Channel," *Eur. Solid-State Device Res. Conf.*, pp. 292–295, 2017.

[54] S. H. Kim *et al.*, "High performance sub-20-nm-channel-length extremely-thin body InAs-on-insulator tri-gate MOSFETs with high short channel effect immunity and Vth tunability," *IEDM Tech. Dig.*, pp. 429–432, 2013.

[55] G. Betti Beneventi *et al.*, "A TCAD Low-Field Electron Mobility Model for Thin-Body InGaAs on InP MOSFETs Calibrated on Experimental Characteristics," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3645–3652, 2015.

[56] F. M. Bufler, G. Eneman, N. Collaert, and A. Mocuta, "Monte Carlo Benchmark of In0.53Ga0.47As- and Silicon- FinFETs," *IEDM Tech. Dig.*, pp. 318–321, 2017.

[57] M. Rau *et al.*, "Performance Projection of III-V Ultra-Thin-Body , FinFET , and Nanowire MOSFETs for two Next- Generation Technology Nodes," *IEDM Tech. Dig.*, pp. 758–761, 2016.

[58] M. Poljak, V. Jovanović, D. Grgec, and T. Suligoj, "Assessment of electron mobility in ultrathin-body InGaAs-on-insulator MOSFETs using physics-based modeling," *IEEE Trans. Electron Devices*, vol. 59, no. 6, pp. 1636–1643, 2012.

[59] S. Johansson, M. Berg, K. M. Persson, and E. Lind, "A High-Frequency Transconductance Method for Characterization of High-κ Border Traps in III-V MOSFETs," *IEEE Trans. Electron Devices*, vol. 60, no. 2, pp. 53–56, 2013.

[60] N. Taoka *et al.*, "Impact of Fermi level pinning inside conduction band on electron mobility

in InGaAs MOSFETs and Mobility Enhancement by Pinning Modulation," *IEDM Tech. Dig.*, pp. 610–613, 2013.

[61] T. Chidambaram *et al.*, "Quantification of interface trap density above threshold voltage by gated hall method in InGaAs buried quantum well MOSFET," *Device Res. Conf.*, pp. 127–128, 2013.

[62] N. Waldron, D. Kim, and J. A. del Alamo, "90 nm Self-aligned Enhancement-mode InGaAs HEMT for Logic Applications," *IEDM Tech. Dig.*, pp. 633–636, 2007.

[63] D. Kim, B. Brar, and J. A. del Alamo, "fT = 688 GHz and fmax = 800 GHz in Lg = 40 nm In0.7Ga0.3As MHEMTs with gmmax > 2.7 mS/um," *IEDM Tech. Dig.*, pp. 319–322, 2011.

[64] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "A CMOS-Compatible Fabrication Process for Scaled Self-Aligned InGaAs MOSFETs," *IEEE Trans. Semicond. Manuf.*, pp. 239–242, 2015.

[65] J. Lin, "InGaAs Quantum-Well MOSFETs for Logic Applications," *Thesis, Massachusetts Inst. Technol.*, 2015.

[66] Y. Yamamoto *et al.*, "Donor passivation in n-AlInAs layers by fluorine," *J. Electron. Mater.*, vol. 25, no. 4, pp. 685–690, 1996.

[67] S. T. N. Hayafuji, Y. Yamamoto, N. Fujiii, T. Sonoda, "Incorporation of Fluorine into the N-type (AlxGayIn1-x-y)As system," *Mat. Res. Soc. Symp. Proc.*, vol. 417, pp. 415–420, 1996.

[68] N. Hayafuji, Y. Yamamoto, N. Yoshida, T. Sonoda, S. Takamiya, and S. Mitsui, "Thermal stability of AlInAs/GaInAs/InP heterostructures," *Appl. Phys. Lett.*, vol. 66, p. 863, 1995.

[69] N. Hayafuji, Y. Yamamoto, T. Ishida, and K. Sato, "Degradation mechanism of the AlInAs/GaInAs high electron mobility transistor due to fluorine incorporation," *Appl. Phys. Lett.*, vol. 69, p. 4075, 1996.

[70] A. Taguchi, T. Ohno, and T. Sasaki, "Fluorine atoms in AlAs, GaAs, and InAs: Stable state, diffusion, and carrier passivation," *Phys. Rev. B*, vol. 62, no. 3, pp. 1821–1827, 2000.

[71] J. Lin, X. Cai, Y. Wu, D. A. Antoniadis, and J. A. del Alamo, "Record Maximum Transconductance of 3.45 mS/μm for III–V FETs," *IEEE Electron Device Lett.*, vol. 37, no. 4, pp. 381–384, 2016.

[72] D. Lin *et al.*, "Beyond interface: The impact of oxide border traps on InGaAs and Ge n-MOSFETs," *IEDM Tech. Dig.*, pp. 645–648, 2012.

[73] X. Sun *et al.*, "AC Transconductance Dispersion (ACGD): A Method to Profile Oxide Traps in MOSFETs Without Body Contact," *IEEE Electron Device Lett.*, vol. 33, no. 3, pp. 438–440, 2012.

[74] X. Sun *et al.*, "High Performance and Low Leakage Current InGaAs-on-Silicon FinFETs with 20 nm Gate Length," *Proc. VLSI Tech.*, pp. 40–41, 2017.

[75] S. Y. Chou and D. A. Antoniadis, "Relationship between Measured and Intrinsic Transconductance in FETs," *IEEE Trans. Electron Devices*, vol. ED-34, no. 2, pp. 448–450, 1987.

[76] X. Sun *et al.*, "AC Transconductance: A Novel Method to Characterize Oxide Traps in Advanced FETs without a Body Contact," *IEDM Tech. Dig.*, pp. 462–465, 2012.

[77] X. Sun, O. I. Saadat, K. S. Chang-Liao, T. Palacios, S. Cui, and T. P. Ma, "Study of gate oxide traps in HfO2/ AlGaN / GaN transistors by use of ac transconductance method," *Appl. Phys. Lett.*, vol. 102, p. 103504, 2013.

[78] J. W. Chung, T. W. Kim, and T. Palacios, "Advanced gate technologies for state-of-the-art fT in AlGaN/GaN HEMTs," *IEDM Tech. Dig.*, pp. 676–679, 2010.

[79] T. Chidambaram, D. Veksler, S. Madisetti, M. Yakimov, V. Tokranov, and S. Oktyabrsky, "InGaAs Inversion Layer Mobility and Interface Trap Density From Gated Hall Measurements," *IEEE Electron Device Lett.*, vol. 37, no. 12, pp. 1547–1550, 2016.

[80] N. Taoka *et al.*, "Impact of Fermi level pinning inside conduction band on electron mobility in InGaAs metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 103, p. 143509, 2013.

[81] 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, no. 7, pp. 871–878, 1988.

[82] N. Apsley, "The Hall scattering factor in InP," *Semicond. Sci. Technol.*, vol. 2, pp. 44–49, 1987.

[83] G. E. Stillman, C. M. Wolfe, and J. O. Dimmock, "Hall coefficient factor for polar mode scattering in n-type GaAs," *J. Phys. Chem. Solids*, vol. 31, no. 6, pp. 1199–1204, 1970.

[84] D. A. Anderson and N. Apsley, "The Hall effect in III-V semiconductor assessment," *Semicond. Sci. Technol.*, vol. 1, no. 3, pp. 187–202, 1986.

[85] G. He, X. Chen, and Z. Sun, "Interface engineering and chemistry of Hf-based high-k

dielectrics on III-V substrates," *Surf. Sci. Rep.*, vol. 68, no. 1, pp. 68–107, 2013.

[86] Y. C. Chang *et al.*, "Atomic-layer-deposited HfO2 on In0.53Ga0.47As: Passivation and energy-band parameters," *Appl. Phys. Lett.*, vol. 92, p. 072901, 2008.

[87] D. A. Antoniadis, "On Apparent Electron Mobility in Si nMOSFETs from Diffusive to Ballistic Regime," *IEEE Trans. Electron Devices*, vol. 63, no. 7, pp. 2650–2656, 2016.

[88] M. Lundstrom, "Elementary scattering theory of the Si MOSFET," *IEEE Electron Device Lett.*, vol. 18, no. 7, pp. 361–363, 1997.

[89] S. Kim *et al.*, "High-performance inas-on-insulator n-MOSFETs with Ni-InGaAs S/D realized by contact resistance reduction technology," *IEEE Trans. Electron Devices*, vol. 60, no. 10, pp. 3342–3350, 2013.

[90] A. Vardi, J. Lin, W. Lu, X. Zhao, and J. A. del Alamo, "High aspect ratio InGaAs FinFETs with sub-20 nm fin width," *Proc. VLSI Tech.*, pp. 7–8, 2016.

[91] A. Majumdar, "Semiconductor Capacitance Penalty per Gate in Single- and Double-Gate FETs," *IEEE Electron Device Lett.*, vol. 35, no. 6, pp. 609–611, 2014.

[92] É. O'Connor *et al.*, "A systematic study of (NH4) 2 S passivation (22%, 10%, 5%, or 1%) on the interface properties of the Al2 O3 / In0.53 Ga0.47 As/InP system for n -type and p -type In0.53 Ga0.47 As epitaxial layers," *J. Appl. Phys.*, vol. 109, p. 024101, 2011.

[93] A. D. Carter, W. J. Mitchell, B. J. Thibeault, J. J. M. Law, and M. J. W. Rodwell, "Al2O3 growth on (100) In0.53Ga0.47as initiated by cyclic trimethylaluminum and hydrogen plasma exposures," *Appl. Phys. Express*, vol. 4, p. 091102, 2011.

[94] V. Chobpattana, J. Son, J. J. M. Law, R. Engel-Herbert, C. Y. Huang, and S. Stemmer, "Nitrogen-passivated dielectric/InGaAs interfaces with sub-nm equivalent oxide thickness and low interface trap densities," *Appl. Phys. Lett.*, vol. 102, p. 022907, 2013.

[95] Z. Liu *et al.*, "Effect of H on interface properties of Al2O3/In0.53Ga0.47As," *Appl. Phys. Lett.*, vol. 99, p. 222104, 2014.

[96] M. Si *et al.*, "Effects of forming gas anneal on ultrathin InGaAs nanowire metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 102, p. 093505, 2013.

[97] A. O. H. Gate-stack *et al.*, "Impact of H 2 High-Pressure Annealing Onto InGaAs Quantum-Well Metal – Oxide – Semiconductor Field-Effect Transistors With," *IEEE Electron Device Lett.*, vol. 36, no. 7, pp. 672–674, 2015.

[98] A. Vais *et al.*, "Record mobility (μeff ~3100 cm2/V-s) and reliability performance

(Vov~0.5V for 10yr operation) of In0.53Ga0.47As MOS devices using improved surface preparation and a novel interfacial layer," *Proc. VLSI Tech.*, pp. 1–2, 2016.

[99] S. Sioncke *et al.*, "First demonstration of ~3500 cm2/V-s electron mobility and sufficient BTI reliability (max Vov up to 0.6V) In0.53Ga0.47As nFET using an IL/LaSiOx/HfO2 gate stack," *Proc. VLSI Tech.*, pp. T38–T39, 2017.

[100] J. W. Chung, "Millimeter-wave GaN High Electron Mobility Transistors and Their Integration with Silicon Electronics," 2011.

[101] G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, "A New Method for Determining the FET Small-Signal Equivalent Circuit," *IEEE Trans. Microw. Theory Tech.*, vol. 36, no. 7, pp. 1151–1159, 1988.

[102] G. Crupi, D. M. M. P. Schreurs, A. Raffo, A. Caddemi, and G. Vannini, "A new millimeter-wave small-signal modeling approach for pHEMTs accounting for the output conductance time delay," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 4, pp. 741–746, 2008.

[103] G. Crupi, D. Schreurs, B. Parvais, A. Caddemi, A. Mercha, and S. Decoutere, "Scalable and multibias high frequency modeling of multi-fin FETs," *Solid. State. Electron.*, vol. 50, pp. 1780–1786, 2006.