InGaAs Quantum-Well MOSFETs for Logic Applications

By

Jianqiang Lin

B. Eng., Electrical Engineering, National University of Singapore, Singapore, 2007
M. Eng., Electrical Engineering, National University of Singapore, Singapore, 2009

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
at the

Massachusetts Institute of Technology

June 2015
© 2015 Massachusetts Institute of Technology
All rights reserved

Signature of Author

Department of Electrical Engineering and Computer Science
May 20, 2015

Certified by

Jesús A. del Alamo
Professor of Electrical Engineering
Thesis Supervisor

Certified by

Dimitri A. Antoniadis
Professor of Electrical Engineering
Thesis Supervisor

Accepted by

Leslie A. Kolodziejski
Professor of Electrical Engineering
Chair, Department Committee on Graduate Students
InGaAs Quantum-Well MOSFETs for Logic Applications

By

Jianqiang Lin

Submitted to the Department of Electrical Engineering and Computer Science on May 20, 2015 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Electrical Engineering

Abstract:

InGaAs is a promising candidate as an n-type channel material for future CMOS due to its superior electron transport properties. Great progress has taken place recently in demonstrating InGaAs MOSFETs for this goal. Among possible InGaAs MOSFET architectures, the recessed-gate design is an attractive option due to its scalability and simplicity. In this thesis, a novel self-aligned recessed-gate fabrication process for scaled InGaAs Quantum-Well MOSFETs (QW-MOSFETs) is developed. The device architectural design emphasizes scalability, performance and manufacturability by making extensive use of dry etching and Si-compatible materials. The fabrication sequence yields precise control of all critical transistor dimensions. This work achieved InGaAs MOSFETs with the shortest gate length ($L_g=20$ nm), and MOSFET arrays with the smallest contact size ($L_c=40$ nm) and smallest pitch size ($L_p=150$ nm), at the time when they were made. Using a wafer bonding technique, InGaAs MOSFETs were also integrated onto a silicon substrate.

The fabricated transistors show the potential of InGaAs to yield devices with well-balanced electron transport, electrostatic integrity and parasitic resistance. A device design optimized for transport exhibits a transconductance of $3.1 \, \text{mS}/\mu\text{m}$, a value that matches the best III-V high-electron-mobility transistors (HEMTs). The precise fabrication technology developed in this work enables a detailed study of the impact of channel thickness scaling on device performance.

The scaled III-V device architecture achieved in this work has also enabled new device physics studies relevant for the application of InGaAs transistors for future logic. A particularly important one is OFF-state leakage. For the first time, this work has unambiguously identified band-to-band tunneling (BTBT) amplified by a parasitic bipolar effect as the cause of excess OFF-state leakage current in these transistors. This finding has important implications for future device design.

Thesis supervisor: Jesús A. del Alamo
Title: Professor of Electrical Engineering

Thesis supervisor: Dimitri A. Antoniadis
Title: Professor of Electrical Engineering
Acknowledgments

First and foremost, I would like to express my sincere gratitude to my thesis advisors, Professor Dimitri Antoniadis and Professor Jesús del Alamo for their continuous support, advice and encouragement throughout the course of my thesis research. I have benefited enormously by their deep insight in the field, and patience and persistence in the pursuit of excellence. I would like to also thank Professor Tomás Palacios and Professor Judy Hoyt for their advice and support in my thesis.

I would also like to express my gratitude toward all in the del Alamo group and Antoniadis group who have been helping me in various aspects ever since I joined the group. The previous effort of del Alamo group and Antoniadis group has formed the foundation of this work. Thanks to Yufei Wu, Evelina Polyzoeva, Niamh Waldron, Jorg Scholvin, John Hennessy, Xiaowei Cai, Jamie Teherani, Ujwal Krishna, Usha Gogineni, Xin Zhao, Tao Yu, Shireen Warnock, Alex Guo, Shaloo Rakheja, Lan Wei, Chia-Yu Chen, Alon Vardy, Xia Ling, Wenjie Lu, Tomohiro Yoshida, Sefa Demirtas, Dae-Hyun Kim, Tae-Woo Kim, Donghyun Jin, Jungwoo Joh, and Luke Guo. I would like to also thank students from Palacios group and Hoyt group who have helped me a lot with fabrication and characterization. Thanks to Xu Zhang, Winston Chern, Han Wang, Bin Lu, Pouya Hashemi, Feng Gao, Allen Hsu, Will Chang, Omair Saadat, Daniel Piedra, Lili Yu, and Mohamed Azize.

My sincere thanks also go to Dr. Sufi Zafar and Dr. Dae-gyu Park for offering me two summer internship opportunities in their groups in IBM Research Center and guiding me to work on the diverse and exciting projects.

Over the past six years, I have enjoyed using the world-class facility for device fabrication and simulation. I owe a debt of gratitude to the staffs of the MTL and SEBL. Without their strong support, moving this project forward would have not been possible. I would like to acknowledge Vicky Diadiuk, Mark Mondol, Bernard Alamariu, Bill Maloney, Bob Bicchieri, Dave Terry, Deborah Hodges-Pabon, Dennis Ward, Donal Jamieson, Eric Lim, Gary Riggott, Kurt Broderick, Kristofer Payer, Mike Hobbs, Paudely Zamora, Paul McGrath, Paul Tierney, Ryan O'Keefe, and Tim Turner.

I've met lots of interesting people at MIT. Many of them have become good friends and made my time in MIT so colorful. I want to thank all of my friends - too many to name everybody.

Most of all, I thank my parents for their never-ending support and encouragement. I hope that my accomplishments have made them proud.

This work was made possible by following funding grants: DTRA, Donner Chair at MIT, E3S STC, Intel Corp., MIT SMART/LEES, and SRC/FCRP MSD. The III-V/Silicon integration has been a collaborative effort with IBM Research, Zurich.
## Contents

### Chapter 1: Introduction

1.1 Motivation of III-V CMOS ................................................................. 17
  1.1.1 Overview of transport-enhanced technology in CMOS devices .............. 17
  1.1.2 Advantages of the III-V channel transistors .............................................. 20
1.2 History and Current Status of III-V FETs ............................................ 22
  1.2.1 Overview of III-V FETs ........................................................................... 22
  1.2.2 HEMT as a logic device ............................................................................. 24
  1.2.3 Grand challenges for III-V MOSFET ......................................................... 26
1.3 Thesis Goals and Organization ............................................................. 27

### Chapter 2: Process Development

2.1 Process Overview ............................................................................... 29
2.2 Contact Module .................................................................................. 35
  2.2.1 Contact consideration ............................................................................... 35
  2.2.2 W barrier for undercut immunity ................................................................. 36
  2.2.3 Metal film resistivity and contact resistivity ............................................... 37
2.3 Pre-III-V Recess Module ................................................................. 40
  2.3.1 RIE damage and its recovery ................................................................. 40
  2.3.2 Gate edge roughness control ................................................................. 43
  2.3.3 Precise lateral control for access region ................................................. 44
2.4 III-V Recess Module ......................................................................... 45
  2.4.1 Drawback of conventional III-V recess .................................................. 45
  2.4.2 III-V dry etch .......................................................................................... 48
  2.4.3 Digital etch ............................................................................................. 51
  2.4.4 Precise channel thickness and access thickness control ......................... 53
2.5 Gate Stack Module ............................................................................ 57
  2.5.1 EOT scaling ............................................................................................ 57
Chapter 2: Introduction ................................................................................. 60

2.5.2 Interface consideration ........................................................................ 60

2.6 Back-end Process Module ....................................................................... 61

2.7 Chapter Summary .................................................................................... 63

Chapter 3: InGaAs MOSFET Performance Analysis ..................................... 65

3.1 Introduction .............................................................................................. 65

3.2 Impact of Access Region Design ............................................................... 65

3.2.1 ON-state characteristics ....................................................................... 66

3.2.2 OFF-state characteristics ..................................................................... 69

3.3 Impact of Channel Thickness Design ....................................................... 70

3.3.1 ON-state characteristics ....................................................................... 72

3.3.2 OFF-state characteristics ..................................................................... 79

3.4 Performance Benchmarking ..................................................................... 83

3.5 Microwave Characteristics ....................................................................... 86

3.6 Chapter Summary ..................................................................................... 89

Chapter 4: Tight-Pitch MOSFET Arrays and Nanoscale Contacts ............... 91

4.1 Introduction .............................................................................................. 91

4.2 Array Design and Fabrication ................................................................. 91

4.3 Result and Discussion ............................................................................. 95

4.4 Chapter Summary ..................................................................................... 101

Chapter 5: III-V Integration on Silicon Substrate ........................................ 102

5.1 Introduction .............................................................................................. 102

5.2 Fabrication Procedures ........................................................................... 103

5.3 Electrical Characteristics ......................................................................... 106

5.4 Device Benchmarking ............................................................................. 109

5.5 Chapter Summary ..................................................................................... 109
Chapter 6: Physics and Mitigation of Excess OFF-state Current..... 111

6.1 $I_{\text{off}}$ in High-Transconductance III-V MOSFETs ................................. 111

6.1.1 Introduction ........................................................................................................ 111
6.1.2 A strategic approach ........................................................................................ 113

6.2 Evidence of Gate-induced Drain Leakage (GIDL) ..................................... 114

6.2.1 Physics of GIDL ................................................................................................... 114
6.2.2 Experiment ............................................................................................................ 117

6.3 Evidence of Bipolar Gain Effect ................................................................. 120

6.3.1 Physics of bipolar gain effect ................................................................................ 120
6.3.2 Experiment ............................................................................................................ 122

6.4 Coupled BTBT and Bipolar Gain Effect .................................................. 126

6.4.1 Simulation setup.................................................................................................... 126
6.4.2 BTBT-induced source-channel barrier lowering ........................................... 127
6.4.3 Bipolar gain amplification and $L_g$ dependency .............................................. 129

6.5 Transistor Design for $I_{\text{off}}$ Reduction .................................................. 131

6.5.1 Delta-doping and channel doping ................................................................. 132
6.5.2 InAs composition in channel and cap ............................................................ 136
6.5.3 Carrier lifetime .................................................................................................. 138
6.5.4 Dimensional design of access region ............................................................. 139

6.6 Chapter Summary .......................................................................................... 140

Chapter 7: Conclusions and Suggestions for Further Research ...... 142

7.1 Thesis Summary .............................................................................................. 142

7.2 Suggested Further Research ....................................................................... 144

7.2.1 Transistor redesign and 3D device architectures ........................................... 144
7.2.2 Modeling of the transport and charge control ............................................. 145

Appendix: Process Integration for InGaAs MOSFET ......................... 147

Bibliography ................................................................................................. 151
List of Figures

Fig. 1-1: (a) Illustration of MOSFET $I_d-V_{gs}$ characteristics. (b) ITRS specification for supply voltage ($V_{dd}$) and nMOS saturated threshold voltage ($V_{t,\text{sat}}$) scaling in future logic devices. $V_{t,\text{sat}}$ is the threshold voltage at $V_{ds}=V_{dd}$ [4]. ................................................................. 18

Fig. 1-2: Simplified schematic of a MOSFET and the potential profile along the channel from source to drain when the device is turned on. It portrays the injection velocity of electron on top of the potential barrier where the virtual source is located. ................................................................. 19

Fig. 1-3: Relative change in electron ballistic velocity vs. relative change in mobility for different strain levels based on experimental data from silicon MOSFETs [17]. ....................................................... 21

Fig. 1-4: The highest room-temperature mobility of electrons in inversion layers and quantum wells for different semiconductors is shown as a function of actual semiconductor lattice constant [15].................................................................................................................. 21

Fig. 1-5: Record transconductance for InGaAs HEMT and MOSFET since their invention ([19], updated until the end of 2014). ........................................................................................................... 22

Fig. 1-6: Selected planar InGaAs FETs (HEMT [20]–[24] and MOSFET [25]–[36]) reported in IEDM for the past 7 years. Key features of the architectures are below each image. “Star” highlights the device technology developed in this work. .................................................................................... 23

Fig. 1-7: Source injection velocity for InAs HEMTs and silicon or strained-silicon FETs as a function of gate length [15].................................................................................................................. 24

Fig. 1-8: Cross-sectional SEM image of a typical HEMT [24]. ................................................................. 26

Fig. 1-9: Cross-sectional schematic to illustrate the key features of a planar logic III-V MOSFET. ................................................................................................................................. 27

Fig. 2-1: Description of the typical heterostructure used in this work. The wafer was grown by IntelliEpi Inc. ......................................................................................................................... 30

Fig. 2-2: Illustration of gate recess steps for fabrication of InGaAs QW-MOSFETs. (a) Deposition of ohmic metal and oxide hardmask. (b) Anisotropic SiO$_2$ and W/Mo etching. (c) W/Mo undercut. (d) III-V cap dry etch. (e) Digital etch. (f) Finished gate. ......................................................... 32

Fig. 2-3: (a) False-color cross sectional TEM of a QW-MOSFET around gate edge. The final channel thickness of this device is 4 nm. The access region, $L_{\text{access}}$, and gate-$n^+$ overlapping region, $L_{\text{ov}}$, are indicated. (b) High-resolution TEM, HRTEM, cross-section of the intrinsic gate region. (c) HRTEM of the contact region. Bracket after layer label indicates thickness in nm. .......................... 32

Fig. 2-4: (a) Cross sectional TEM images of a MOSFETs with 20 nm gate length and 50 nm contact-to-contact spacing. (b) Its output characteristics. ........................................................................................................... 35
Fig. 2-5: Cross sectional TEM images of two MOSFETs: (a) with single layer of Mo contact [26], [41], and (b) composite W/Mo contact [27], [28]..................................................................................................................... 37

Fig. 2-6: Benchmarking of (a) film resistivity $\rho$, [21], [41], [50], [53]–[61]; and (b) contact resistivity $\rho_c$, [28], [50]–[52], [55]–[58], [60]–[66], for different contact technologies proposed for n-type InGaAs MOSFETs (see text). .............................................................................................................. 38

Fig. 2-7: Electrical characteristics of QW-MOSFET with (blue) and without (black) RIE damage anneal: (a) Output characteristics, (b) Transfer and transconductance characteristics, (c) Subthreshold characteristics, (d) Effective field mobility. ..................................................................................... 41

Fig. 2-8: C-V characteristics for InGaAs MOSFETs with (a), and without (b) the RIE damage anneal........................................................................................................................................ 42

Fig. 2-9: Optimization of the gate edge roughness. Cross sectional SEM images for the ZEP/SiO2/Ohmic patterns after SiO2 RIE with gas combination of (a) CF$_4$:H$_2$ =3:1, (b) CF$_4$:H$_2$ =4.5:1, (c) CF$_4$:H$_2$ =6:1. (d) Smooth gate edge line of 20 nm is obtained from the optimized dry etchings. ......................................................................................................................... 43

Fig. 2-10: TEM cross sectional images for: (a) $L_g$ = 50 nm device with $L_{access}$ =15 nm, and (b) $L_g$ = 70 nm device with $L_{access}$ =80 nm.............................................................................................................................................................. 44

Fig. 2-11: Cross sectional TEM image of QW-MOSFET with $L_g$ = 30 nm. The semiconductors around the gate region are recessed by selective wet etching............................................................................................................ 45

Fig. 2-12: Performance of a $L_g$=60 nm QW-MOSFET fabricated by selective cap wet etch. (a) Transfer and transconductance characteristics. (b) Output characteristics. (b) Subthreshold characteristics........................................................................................................................................ 46

Fig. 2-13: III-V cap etch rate as a function of substrate bias voltage. Threshold bias voltage for etching is about 60 V ........................................................................................................................................ 47

Fig. 2-14: AFM surface scan for 1 $\mu$m$^2$ area on: (a) virgin wafer, (b) dry-etched sample under baseline conditions, (c) 40°C, (d) 150°C, (e) increased bias of 234 V, (f) increased ratio of N$_2$:Cl$_2$ flow rate, (g) increased pressure and (h) BCl$_3$-based recipe [72]. ...................................................................................................................... 48

Fig. 2-15: TEM cross sectional images of finished self-aligned InGaAs MOSFETs: (a) etched under baseline conditions with a final 4 nm channel thickness; (b) etched at a higher substrate bias (140 V) and a final 8 nm channel thickness. Increased bias results in observable trenching at the two ends of the channel........................................................................................................................................ 49

Fig. 2-16. Illustration of InP digital etch rate calibration. ........................................................................................................................................ 50

Fig. 2-17: Etch rate per cycle of digital etch as a function of oxidation time under O$_2$ plasma, and fitting with Lukeš’s model [77]. ........................................................................................................................................ 51
Fig. 2-18: Precise etching process of the intrinsic region of the device is performed by RIE (left) and DE (right). The RIE stops a few nm above the channel surface. The final channel thickness is controlled by DE with 1 nm precision. ................................................................. 53

Fig. 2-19: Left: flowchart of digital etch calibration process on test sample. Right: schematic illustration for (a) calibration sample after dry etch, (b) one digital etch cycle, and (c) inspection using optical microscope................................................................. 54

Fig. 2-20: Cross-sectional TEM images of two different runs with different final channel thickness targets when combining dry etch and digital etch: (a) 4 nm and (b) 8 nm. A separate example achieved when combining wet etch and digital etch for a total buried channel thickness of 11 nm. High resolution TEM image (a-HR) shows 7 monolayers remaining in the intrinsic channel for (a), and (c-HR) shows 18 monolayers for (c). ........................................................................ 55

Fig. 2-21: Long-channel subthreshold characteristics of the four generations of gate stack (G1-G4). The subthreshold characteristics are measured at 0.5 and 0.05 V respectively............... 59

Fig. 2-22: Gate leakage density, Jg, of the four gate stacks investigated in this work............. 60

Fig. 2-23: (a) Subthreshold characteristics of a long-channel MOSFET with G2 gate stack. The gate length is 300 µm. (b) Low-voltage subthreshold swing vs. dielectric EOT of several planar long-channel III-V MOSFETs and HEMTs. ........................................................................... 61

Fig. 2-24: Cross sectional schematic for via process: (a) double resist for via opening through selective RIE, (b) pad evaporation, (c) pad lift-off................................................................. 62

Fig. 2-25: Back-end process: (a) Via opened after SiO₂ dry etch using double-layer resist. The etch stops at the W/Mo surface without any surface damage or residual material; (b) Via filling by a self-aligned metal pad. ......................................................................................... 63

Fig. 3-1: Schematic illustration of (a) the key parameters of the access region and (b) the intrinsic channel of the device used in this study. TEM cross-sectional image in the access regions of the gate to the source for (c) short-access device and (d) long-access device. ........................................................................ 66

Fig. 3-2: Transconductance characteristics of Lg=70 nm MOSFETs with long-access region and short-access region at Vds=0.5 V. ..................................................................................... 68

Fig. 3-3: Output characteristics of Lg=70 nm MOSFETs with (a) long-access region and (b) short-access region............................................................................................................ 68

Fig. 3-4: Rsd extraction from Ron vs. Lg measurements for (L) long-access device and (S) short access device................................................................. 69

Fig. 3-5: Subthreshold characteristics for InGaAs MOSFET with Lg=70 nm and (a) long-access region and (b) short-access region..................................................... 70
Fig. 3-6: (a) Correlation between number of digital etch cycles and final channel thickness. Examples of (b) buried channel and (c) surface channel designs. The dashed lines beneath the HfO₂ indicate the location of the oxide/semiconductor interface................................. 71

Fig. 3-7: Electrical characteristics of device with \( t_c = 9 \) nm and \( L_g = 80 \) nm: (a) transfer and transconductance characteristics at \( V_{ds} = 0.5 \) V and (b) output characteristics................................................. 72

Fig. 3-8: Peak intrinsic transconductance at \( V_{ds} = 0.5 \) V \( vs. \ L_g \) for different \( t_c \).................................................. 73

Fig. 3-9: (a) \( R_{on} \) \( vs. \ L_g \) for \( t_c \) from 3 nm to 12 nm, in which \( R_{sd} \) is extracted from the y-intersection. (d) \( R_{sd} \) \( vs. \ t_c \)............................................................................................................... 74

Fig. 3-10: Split-C-V measurement on QW-MOSFET with \( L_g = 5 \) µm for different \( t_c \)................. 75

Fig. 3-11: Mobility \( vs. \) sheet charge density. Continuous lines: effective channel mobility in long-channel devices (\( L_g = 5 \) µm) with different channel thickness. Solid symbols: two sets of etch-Hall mobility from two growth test structures (see text) ..................................................... 76

Fig. 3-12: (a-c) Carrier distribution, \( n \), and conduction-band energy, \( E_c \), in the direction perpendicular to the channel. The semiconductor/oxide interface is placed at \( x=0 \). The energy of the first two quantum states is shown in dashed lines. (d) Average distance of channel electron population, \( d_{av} \), with respect to the oxide-semiconductor interface at \( N_s \) of \( 1 \times 10^{10} \) cm\(^{-2} \) (blue) and \( 3 \times 10^{12} \) cm\(^{-2} \) (red). The schematic of \( d_{AV} \) is shown in the inset in (a). (e) Experimental mobility \( vs. \) \( Z \) at \( N_s \) of \( 3 \times 10^{12} \) cm\(^{-2} \) (see text). ............................................................................................. 78

Fig. 3-13: Subthreshold characteristics of two \( L_g = 80 \) nm MOSFET with (a) same device as that in Fig. 3-7 with \( t_c = 9 \) nm and (b) \( t_c = 4 \) nm................................................................. 80

Fig. 3-14: Threshold voltage roll-off behavior for InGaAs MOSFET for \( t_c \) from 3 nm to 12 nm. .............................................................................................................. 80

Fig. 3-15: (a, b) Experimental \( S_{min} \) and DIBL as a function of gate length for \( t_c = 3 \) nm to 12 nm. (c, d) \( S_{min}-S_{min,long} \) and DIBL-DIBL\(_{long} \) as a function of \( \gamma \) based on the “natural channel length” methodology discussed in the text................................................................. 81

Fig. 3-16: Benchmarking of \( S_{min} \) at \( V_{ds} = 0.5 \) V \( vs. \ L_g \) for III-V FETs published at (a) 2013 year-end and (b) 2014 year-end. Ref: [20], [21], [23], [30], [33], [34], [81], [82], [100]–[110].............. 84

Fig. 3-17: Benchmarking of \( g_{m,\text{max}} \) \( vs. \ S_{min} \) at 0.5 V for III-V FETs with \( L_g \leq 80 \) nm showing record device obtained in the device obtained in this thesis. Ref: [20], [23], [30], [81], [82], [100], [101], [104], [109], [110].................. 85

Fig. 3-18: Benchmarking of \( I_{on} \) \( vs. \ L_g \) at \( V_{dd} = 0.5 \) V and \( I_{off} \leq 100 \) nA/µm for III-V FETs. Ref: [24], [29], [30], [76], [80], [82], [101], [103], [105], [109]–[114]............................... 85

Fig. 3-19: SEM top-view of the microwave device................................................................. 87
Fig. 3-20: (a) Transconductance from DC measurements and RF measurements in the same short-access InGaAs MOSFET. (b) Microwave characteristics of $L_g=70$ nm InGaAs MOSFET. $H_{21}$ and $U_g$ vs. frequency for device biased at peak $g_m$ point........................................ 88

Fig. 3-21: At fixed $V_{gs}$ of 0.4 V and different $V_{ds}$ for short-access InGaAs MOSFET with $L_g=70$ nm (a) $H_{21}$ vs. frequency; (b) $f_T$ vs. $V_{gs}$. At fixed $V_{ds}$ of 0.5 V and different $V_{gs}$ for the same device (c) $H_{21}$ vs. frequency; (d) $f_T$ vs. $V_{gs}$. Inset of (d) shows the correlation between $f_T$ and $g_m$. ........................................................................................................ 88

Fig. 4-1: Schematic illustration of process flow for array fabrication: (a) E-beam lithography, (b) Ohmic metal etching and gate-less array, (c) III-V recess, and (d) MOSFET gate structure. The finished devices with pads are shown for (e) MOSFET gate array and (f) gate-less array. ........... 92

Fig. 4-2: Schematic illustration of MOSFET array with (a) 1 cell and (b) 2 cells. ............... 93

Fig. 4-3: (a) Top view of a MOSFET array. (b) Enlarged top view that shows MOSFET array with 3 cells. ................................................................................................................................................. 93

Fig. 4-4: Cross sectional TEM of MOSFET array with different contact length, gate length and pitch size: (a) $L_c=40$ nm, $L_g=130$ nm and $L_p=200$ nm; (b) $L_c=80$ nm, $L_g=40$ nm and $L_p=150$ nm. ................................................................................................................................. 94

Fig. 4-5: Cross sectional SEM of gate-less array with $L_c=130$ nm and $L_p=200$ nm. Inset shows the enlarged image and layer materials.............................................................................................................. 94

Fig. 4-6: (a) $I_d-V_{gs}$ and $g_m-V_{gs}$ characteristics for a MOSFET arrays with $L_g=100$ nm, $L_c=500$ nm and 3 cells. (b) Extraction of $R_{cell}$ from the slope of total resistance vs. number of cells in the gate-less arrays and MOSFET arrays. .................................................................................................................. 96

Fig. 4-7: (a) $R'_{cell}$ vs. $L_c$ in gate-less arrays where $\rho_{12}$ is extracted. (b) $R_{cell}$ vs. $L_c$ in MOSFET arrays where $\rho_{23}$ is extracted. Solid symbol is from experiment and line is from model.......... 97

Fig. 4-8: Cross-section schematic of one cell for (a) MOSFET array and the circuit model for $R_{cell}$, (b) gate-less array and the circuit model $R'_{cell}$....................................................................................................................................... 99

Fig. 4-9: (a) Schematic of the contact region that contains three parallel conducting layers with the specified sheet resistance. (b) The circuit model for the contact region. The same model is used for the MOSFET array and gate-less array with the appropriate boundary conditions. (see text). ............................................................................................................................................... 99

Fig. 4-10: Modeled contact resistance for the technology in this work based on the extracted $\rho_{12}$ and $\rho_{23}$. This is the estimated resistance between nodes A and E in Fig. 4-9..................... 100

Fig. 5-1: Fabrication procedure for InGaAs MOSFETs on III-V-O-I (see text). ................. 104

Fig. 5-2: (a) $\omega-2\theta$ X-ray diffractogram (XRD) acquired on the finished III-V-O-I substrate and aligned on the InP (004) peak. (b) AFM scan of the surface of the finished III-V-O-I substrate with height scale on the right ......................................................... 104
Fig. 5-3: TEM cross sectional picture for section of MOSFET array with a gate pitch of \( L_p = 150 \) nm on III-V-O-I. The device has \( L_g = 50 \) nm, \( L_{\text{access}} = 30 \) nm and \( L_c = 40 \) nm.

Fig. 5-4: Transconductance characteristics for an InGaAs MOSFET with gate length of 70 nm at \( V_{bs} = -2, 0 \) and 3 V. In all figures, \( V_{ds} = 0.5 \) V.

Fig. 5-5: Output characteristics for the same InGaAs MOSFET as in Fig. 5-4 at \( V_{bs} = -2, 0 \) and 3 V.

Fig. 5-6: Subthreshold characteristics for the same InGaAs MOSFET as in Fig. 5-4 at \( V_{bs} = -2, 0 \) and 3 V. In all figures, \( V_{ds} = 0.05 \) and 0.5 V.

Fig. 5-7: (a) Minimum subthreshold swing at \( V_{ds} = 0.5 \) V and (b) DIBL vs. \( V_{bs} \) for the InGaAs MOSFET of Fig. 5-4.

Fig. 5-8: (a) Threshold voltage modulation by \( V_{bs} \) for a \( L_g = 1 \) \( \mu \)m device, and (b) source and drain series resistance under different \( V_{bs} \) from a plot of \( R_{on} \) vs \( L_g \) that interpolates to \( L_g = 0 \).

Fig. 5-9: (a) \( S_{\text{min}} \) vs. \( L_g \) and (b) DIBL vs. \( L_g \) for this work and recently published III-V MOSFETs on insulator. Ref: [104], [123]–[125].

Fig. 6-1: Examples of high-transconductance InGaAs MOSFETs that exhibits excess drain leakage in the OFF-state at high \( V_{ds} \). (a) Regrown-SD InGaAs MOSFET with \( L_g = 55 \) nm [30]. (b) Short-access device in this work with \( L_g = 70 \) nm.

Fig. 6-2: Comparison of subthreshold characteristics in long channel MOSFETs with A: short-access and \( L_g = 500 \) nm and B: long-access and \( L_g = 200 \) nm. The devices were fabricated at the same time on the same sample.

Fig. 6-3: (a) Schematic illustration of BTBT in an InGaAs MOSFET. The device is biased in the OFF-state with \( V_{gs} < V_t \) and \( V_{ds} = V_{dd} \). (b) Corresponding energy band diagram along the channel from source to drain. The star indicates the high field region where BTBT occurs. The arrows indicate the BTBT-generated hole flux \( (F_h) \) and electron flux \( (F_e) \) that constitute the BTBT current. (c) Classic GIDL characteristics at high \( V_{ds} \) [128].

Fig. 6-4: Subthreshold swings of a \( L_g = 500 \) nm short-access MOSFET vs. \( I_d \) at different temperatures. The ideality factor is defined as \( n = S_{\text{min}} / S_{\text{ideal}} \). Almost ideal subthreshold swing below \( T = 150 \) K indicates that interface states are frozen.

Fig. 6-5: Subthreshold characteristics of a \( L_g = 500 \) nm short-access MOSFET at (a) 300 K, (b) 150 K, and (c) 77 K.

Fig. 6-6: Low-temperature dependence of \( I_d \) vs. \( E_g^{3/2} \) in the OFF-state. The straight line that is obtained suggests a tunneling process across the \( \text{In}_{0.7}\text{Ga}_{0.3}\text{As} \) bandgap and supports the BTBT hypothesis.

Fig. 6-7: (a) \( I_d \) vs. \( V_{gs} \) and (b) \( I_d \) vs. \( [V_{dg} + V_t]^{-1} \) for a short-access device with \( L_g = 500 \) nm at 77 K.
Fig. 6-8: Schematic illustration of BTBT and bipolar gain action in an InGaAs QW-MOSFET biased in the OFF-state. (b) Corresponding energy band diagram along the channel from source to drain. The source-channel barrier height \( q\phi_b \) determines the OFF-state drain current.

Fig. 6-9: (a) Experimental subthreshold characteristics of InGaAs MOSFETs for gate lengths between 80 and 500 nm, measured at \( V_{ds}=0.7 \) V and at 200 K. (b) \( I_d^{-1} \) vs. \( L_g \) at fixed \( V_{ds} \) and \( V_{gs} - V_t \) under the same measurement conditions of (a).

Fig. 6-10: (a) Experimental room temperature (RT) subthreshold and gate current characteristics of long-channel InGaAs MOSFETs. (b) Simulated subthreshold characteristics with and without BTBT model suggesting that BTBT plays a role in the excess off-state current.

Fig. 6-11: (a) BTBT contours for electrons (right circle) and holes (left circle) at the drain edge of the channel. The channel consists of In\(_{0.7}\)Ga\(_{0.3}\)As/InAs/In\(_{0.7}\)Ga\(_{0.3}\)As = 1/2/5 nm (Ch. 2). Electron generation takes place in two regions: around the n+/undoped InGaAs junction as a result of the strong band bending and in the InAs core because of its narrow bandgap. (b) Electron and hole current density corresponding to the same situation as in (a). The currents per unit gate width at point \( x_o \) are indicated on the right.

Fig. 6-12: Cross sectional device structure used in simulations.

Fig. 6-13: (a) Simulated subthreshold \( I_d-V_{gs} \) characteristics with and without BTBT for MOSFETs with gate lengths of 40, 100, and 800 nm. (b) Corresponding source-channel barrier height \( q\phi_b \) as a function of \( V_{gs} \). In all cases, \( V_{ds}=0.6 \) V.

Fig. 6-14: Simulated conduction band-edge along the channel from source to drain \( (E_c) \) at various gate voltages for InGaAs MOSFETs with \( L_g=800 \) nm (a, b) and 40 nm (c,d). In all cases, \( V_{ds} \) is 0.6 V. In (a) and (c), the BTBT model is turned off. In (b) and (d), the BTBT model is turned on.

Fig. 6-15: (a) Electron \( (I_e) \) and hole \( (I_{BTBT}) \) channel currents and total drain current \( (I_d,T) \) characteristics vs. \( V_{gs} \) for MOSFETs with gate lengths of 40, 100, and 800 nm at \( V_{ds}=0.6 \) V. (b) Enlarged view for the box region in (a).

Fig. 6-16: \( \beta^{-1} \) vs. \( L_g \) at \( V_{ds}=0.6 \) V for \( V_{gs} \) values between -0.45 and -0.25 V.

Fig. 6-17: Subthreshold \( I_d-V_{gs} \), and \( I_{BTBT}-V_{gs} \) characteristics for \( L_g=40 \) nm for different (a) backside \( \delta \)-doping type and doping level, and (b) channel uniform p-type doping level. For both simulations, \( V_{ds} \) is 0.6 V.

Fig. 6-18: \( \beta \) vs. channel uniform p-doping \( (N_A \) at the bottom x axis) and backside \( \delta \)-doping level \( (N_B \) at the top x axis) at \( V_{gs} = -0.4 \) V and \( V_{ds} = 0.6 \) V. \( L_g \) for this device is 40 nm. Negative \( \delta \)-doping values indicate n-type doping.

Fig. 6-19: Energy band diagram below the gate around the source in the direction perpendicular to the surface.
Fig. 6-20: (a) Simulated energy band diagrams in the perpendicular direction in the channel under the gate for three different types of backside $\delta$-doping (at the same $V_{gs}-V_t$). For clarity, the band diagrams are matched at the oxide-semiconductor interface. (b) $\beta$ vs. $q\Phi_c$ at fixed $V_{gs}-V_t = -0.35$ V and $V_{ds}=0.6$ V in semilog scale indicates the exponential dependency of $\beta$ on $q\Phi_c$. $q\Phi_c$ is extracted in the center of the channel ($z=0$). ................................................................. 135

Fig. 6-21: Energy band diagram under the contact along the perpendicular direction for $x_{cap}<x_{ch}$, ($E_{g, cap}>E_{g, ch}$). ............................................................................................................................... 137

Fig. 6-22: (a) Extracted $I_d$ and $I_{BTBT}$ and (b) $\beta$ vs. $x_{cap}$ at $V_{gs} = -0.35$ V and $V_{ds}=0.6$ V. $L_g$ for this device is 40 nm and $x_{ch}=0.7$.............................................................. 137

Fig. 6-23: (a) $I_d$ and $I_{BTBT}$ and (b) $\beta$ vs. $x_{ch}$ at $V_{gs} = -0.35$ V and $V_{ds}=0.6$ V. $L_g$ for this device is 40 nm and $x_{cap}=0.53$.............................................................. 138

Fig. 6-24: (a) Extracted $I_d$ and $I_{BTBT}$ and (b) $\beta$ vs. $\tau$ at $V_{gs} = -0.45$ V and $V_{ds}=0.6$ V. $L_g$ for this device is 40 nm. .............................................................. 139

Fig. 6-25: $I_d$ extracted at $V_{gs} = -0.35$ V and $V_{ds}=0.6$ V for (a) varying $L_{access}$ and (b) varying $t_{cap}$ in otherwise a baseline device design. .............................................................. 140
Chapter 1: Introduction

1.1 Motivation of III-V CMOS

1.1.1 Overview of transport-enhanced technology in CMOS devices

The scaling of dimensions and voltages the CMOS technology has led to increased device density, faster switching speed and enhanced power efficiency. Areal scaling by a factor of two approximately every two years is generally known as Moore’s Law [1] and has been ruling the microelectronics industry for the past several decades. While Moore’s Law has many performance and economic virtues, it has also resulted in increased power density approaching unsustainable levels. In this regard, power density is recognized as a key constraint for the continuation of Moore’s Law, which would threaten to halt the microelectronics revolution [2], [3]. To address this problem, computation must be performed in a more energy-efficient manner in which the performance gain of integrated circuits is no longer relying on increased power density.

One way to reduce the power density in integrated circuits is by reducing the supply voltage, $V_{dd}$. The span of supply voltage can be divided into two regimes, the subthreshold regime and the gate overdrive regime, which are separated by the threshold voltage, $V_t$. This is illustrated in the typical MOSFET current transfer characteristics in Fig. 1-1 (a). $V_{dd}$ reduction can be realized by reducing the subthreshold component, $V_t$, or the gate overdrive component, $V_{dd}-V_t$. However, because of the non-scalability of the rate of change of current with gate voltage, defined as the “subthreshold swing” measured in decades of current change per volt that is imposed by the limit of thermionic emission, there is little room for $V_t$ reduction for logic MOSFETs. In fact, $V_t$ is expected to remain constant or increase slightly for transistors with
smaller dimensions because short channel effects such as increase of subthreshold swing, $S$, and drain-induced-barrier lower, DIBL, and increased threshold voltage variation will require higher $V_t$ to achieve a given $I_{OFF}$ target. This is illustrated in Fig. 1-1 (b) which shows the prediction of $V_{dd}$ and $V_t$ from the International Technology Roadmap of Semiconductors, ITRS [4]. It indicates that the supply voltage $V_{dd}$ needs to decrease from 0.8 V to 0.6 V from the 21 nm node to the 7.5 nm node in the next 10 years, while the saturation threshold voltage, $V_{t,sat}$, of the n-type MOSFETs has to increase from 0.21 V to 0.24 V. The consequence of this is reduction of the gate overdrive, $V_{dd} - V_t$, which will eventually compromise the drive current in the ON-state. Furthermore, the aggressive pitch size scaling deteriorates the series resistance [5]–[7]. The contribution of the extrinsic parasitic components to the total ON-resistance will keep increasing, and further degrade the drive current. To compensate the drive current loss, transport-enhanced channel transistors have been proposed and are currently used or explored for modern CMOS logic devices.

Fig. 1-1: (a) Illustration of MOSFET $I_d$-$V_{gs}$ characteristics. (b) ITRS specification for supply voltage ($V_{dd}$) and nMOS saturated threshold voltage ($V_{t,sat}$) scaling in future logic devices. $V_{t,sat}$ is the threshold voltage at $V_{ds}=V_{dd}$ [4].

For modern deeply-scaled transistors, the drive current, $I_{on}$, can be described by (1-1):

\[
I_{on} = v_{inj} Q_{inv} \\
\approx v_{inj} C_{inv} (V_{gs} - V_t)
\]

(1-1)
V_{gs-V_t} is the gate overdrive and C_{inv}(V_{gs-V_t}) gives the inversion layer charge at the virtual source. For a given gate capacitance and gate overdrive, the drive current is proportional to the source injection velocity, v_{inj} [6]. The source injection velocity is a parameter that governs carrier transport in field-effect transistors. Its physical meaning is illustrated in Fig. 1-2 which depicts a typical MOSFET biased in strong inversion in saturation. The conduction band potential from source to drain is illustrated. The “virtual source” is located at the top of the barrier, x = x_o [6]–[8].

![Fig. 1-2: Simplified schematic of a MOSFET and the potential profile along the channel from source to drain when the device is turned on. It portraits the injection velocity of electron on top of the potential barrier where the virtual source is located.](image_url)

For a given material system, the maximum possible value for v_{inj} is determined by the unidirectional thermal velocity v_0 toward the channel for carriers at x = x_o. This case is referred to as ballistic transport. In a modern logic transistor whose channel length is close to or smaller than the mean-free-path, the transistor operates near the ballistic transport limit. For conventional n-type MOSFETs, due to the relatively heavy effective mass of electron and low thermal velocity in relaxed silicon, v_{inj} saturates around 1\times10^7 \text{ cm/s} [6]–[8].

An effective way to break this bound is through transport-enhanced channel engineering. The most well-known technology of this type is strained-silicon, which was adopted in manufacturing since early 2000s [9]–[11]. The electron and hole velocity enhancement through
strained-silicon is limited by the available methods to couple increased amount of uniaxial strain into the channel [12]. A lot of researches have been directed to introduce higher strain in the silicon channel. For instance, in the multiple-stressor scheme, Si:C stressors in the source and drain and SiGe stressors beneath the channel are introduced to enhance the uniaxial tensile strain in the n-type silicon channel [13]. Such technology can bring up the source injection velocity in n-type MOSFETs to about $1.6 \times 10^7$ cm/s [12], [14].

Despite the great success of strained-silicon technology in the last decade, the quest for high device density in advanced technology nodes makes strain engineering increasingly difficult. The mechanical strain and performance gain has started to diminish due to aggressive transistor pitch scaling. To continue Moore’s Law of scaling, it is imperative to find effective way to enhance carrier transport in scaled dimensions. In this regard, the use of alternative materials that have superior transport properties for MOSFET channel would be advantageous. Because of the extraordinary electron transport properties of certain III-V compound semiconductors, III-V’s are considered a promising candidate as an n-type channel material for future CMOS logic [15].

1.1.1 Advantages of the III-V channel transistors

Mobility is an important metric to evaluate the transport properties of a material system in the drift-diffusion limit. However, mobility is also closely related to the ballistic velocity in the ballistic transport limit through the effective mass [16]. When the increase in mobility is purely due to the decrease in the effective mass, ballistic velocity exhibits a power law dependence on mobility, $v_0 \propto \mu^{1/2}$. Hence, mobility can be considered as a “proxy” for source injection velocity as shown in Fig. 1-3 [12].

The mobility of many III-V compound semiconductors is far greater than that of silicon. Fig. 1-4 shows the highest room-temperature mobility of electrons in inversion layers and quantum wells as a function of the actual lattice constant for several semiconductors of interest [15]. The highest electron mobility of 32,000 cm$^2$/V·s can be found in InSb and unstrained-InAs. However, it is difficult to grow InSb and unstrained-InAs on pseudomorphic substrates and build
device prototypes due to large lattice mismatch to commonly available substrates such as GaAs and InP. On the other hand, reasonably thick strained- InAs and In\textsubscript{x}Ga\textsubscript{1-x}As can be grown on a lattice-mismatched pseudomorphic InP substrate. The InAs composition $x$ can vary from 0.53 to 1. Biaxial compressive strain of 3.1\% results when InAs is grown on InP. In this case, the mobility of InAs is 18,000 cm\textsuperscript{2}/V·s. The change of InAs mobility due to strain is indicated by the left-pointing arrow in Fig. 1-4. For In\textsubscript{x}Ga\textsubscript{1-x}As on InP, the impact of InAs composition in mobility is indicated by the upward pointing arrow. This work will focus on In\textsubscript{x}Ga\textsubscript{1-x}As channels that are grown on InP substrate.

Fig. 1-3: Relative change in electron ballistic velocity vs. relative change in mobility for different strain levels based on experimental data from silicon MOSFETs [17].

Fig. 1-4: The highest room-temperature mobility of electrons in inversion layers and quantum wells for different semiconductors is shown as a function of actual semiconductor lattice constant [15].
1.2 History and Current Status of III-V FETs

1.2.1 Overview of III-V FETs

A brief overview of the InGaAs field-effective transistor, FET, is given here. Two main architectures of InGaAs FETs are Schottky-gate high-electron-mobility transistors, HEMT, and insulated-gate MOSFET. Both of them were invented in the early 1980s. Fig. 1-5 graphs the record transconductance, $g_m$, as a representative metric to evaluate these two device technologies. Here, InGaAs refers to any composition of the ternary from pure GaAs to pure InAs. HEMTs made persistent progress immediately after their first demonstration in 1980. Since then, they have been delivering the highest transconductance and cut-off frequency in the family of III-V FETs [18], and have also been turned into a very successful transistor technology. On the contrary, the InGaAs MOSFETs were evolving very slowly in the 1980s and 1990s. This was until mid-2000s when a rapid performance improvement started to happen because the prospect of III-V CMOS started to attract tremendous research interest around the world. III-V CMOS research greatly accelerated the device innovation of InGaAs MOSFETs. MOSFETs have demonstrated their potential and exceeded the transconductance performance of HEMT in 2014. This record InGaAs MOSFET device, among a few others, was developed in the course of this thesis research.

![Fig. 1-5: Record transconductance for InGaAs HEMT and MOSFET since their invention ([19], updated until the end of 2014).](image)
Since the mid-2000s, a variety of III-V FET architectures were invented and investigated, including HEMT and MOSFET. They were used for the study of CMOS-relevant process integration and device physics. As a quick illustration of the research progress,

Fig. 1-6 shows the planar InGaAs FETs prototype devices, including HEMT and MOSFET, from the International Electron Device Meeting, IEDM, in the past 8 years. Besides this, many other device engineering innovations, such as the 3D device prototypes, have been reported across the literature.

![Fig. 1-6: Selected planar InGaAs FETs prototypes devices, including HEMT and MOSFET, from the International Electron Device Meeting, IEDM, in the past 7 years. Key features of the architectures are below each image. “Star” highlights the device technology developed in this work.](image-url)
1.2.2 HEMT as a logic device

HEMT is a popular III-V device architecture for several unique electronic applications including millimeter wave and power amplifiers [37]. Owing to its similar operating principle as the MOSFET and the maturity and simplicity of its process, HEMT has also been used as the platform for device physics study for transistors with III-V CMOS channels. Those devices indeed have generated profound new understandings. As shown in Fig. 1-7, HEMTs have shown that the III-V channel can deliver at least 2 times $v_{\text{inj}}$ compared to or strained-silicon at reduced drain voltages [15]. Increased InAs composition in $\text{In}_x\text{Ga}_{1-x}\text{As}$ leads to $v_{\text{inj}}$ enhancement. These results are evidence to support the superior electron transport properties of this materials system and showcase the advantage for future high performance and low power CMOS applications.

![Fig. 1-7: Source injection velocity for InAs HEMTs and silicon or strained-silicon FETs as a function of gate length [15].](image)

The 30-year research on HEMTs has converged to a mature fabrication process and device architecture that exploits the transport advantage of the III-V channel for high frequency applications. Accordingly, the process and structure of HEMTs has remained relatively unchanged over the years, as shown in the HEMT section of
Fig. 1-6. However, as the complete cross-sectional SEM image of a HEMT device shown in Fig. 1-8 [24] highlights, such structure is not suitable for the CMOS applications as discussed below.

The first problem is associated with the gate leakage. In high density circuit, the gate leakage density should be bound below 1 A/cm². HEMTs use the “Schottky gate” technology in which the gate metal is directly placed on top of a wide bandgap semiconductor or barrier; usually InAlAs is used as the barrier. Due to the small band offsets, carriers can easily tunnel across the barrier if it is thin. In this case, the gate leakage can easily exceed 100 A/cm² in the ON-state [20]–[24]. The gate leakage limits how thin the barrier can be scaled. This compromises the short channel effect (SCE) control when the lateral dimensions of the device are scaled down.

Secondly, the removal of the n-dope semiconductor above the channel in the gate region is called “gate recess”. Typical gate recess is done by selective wet etch, which results in an appreciable lateral extent and an “underlap” device architecture as shown in Fig. 1-8. This is in contrast to the modern CMOS “overlap” architecture in which the $n^+$ region is slightly overlapped with the gate. The underlap region limits the device footprint and increases series resistance.

Thirdly, since footprint is not a major concern for HEMT, typical fabrication of the gate and source/drain usually uses two separate masks. This violates the self-alignment principle required for the fabrication of front-end logic transistors. A long access region must be reserved for the lithography alignment tolerance, directly leading to a large source-to-drain spacing, approximately 2 μm in the case of Fig. 1-8.

Fourthly, HEMTs use Au-containing alloy to form the gate, source and drain electrodes while the pattern definition uses lift-off processes which result in the violation of CMOS-compatibility in both materials and process.
The grand challenges for III-V MOSFETs in logic applications are discussed in detail in this section. Unlike silicon MOSFETs and III-V HEMTs which has been turned into successful technologies, there has only been very limited research effort on III-V MOSFETs before the mid-2000s, and those research efforts were rather unsuccessful. This was mainly due to the lack of a stable oxide on III-V for gate insulator. High-quality gate stack is the heart of a MOSFET. This challenge must be first addressed before the potential of III-V for CMOS application can be realized. Recent advances in atomic layer deposition, ALD, have provided the critical enabling technology to address this challenge [38], [39]. Since the gate insulator is a deposited material, the oxide/III-V interface quality is a serious concern which directly affects current drive and SCE immunity in scaled transistors.

Pitch dimension is a requirement imposed by the high-density logic applications. The gate pitch of modern silicon MOSFET is well below 100 nm, within which there are three functional regions: one gate, one silicide contact, and two spacers. In the III-V terminology, they are equivalent to the gate, the ohmic contact, and the access regions as illustrated in Fig. 1-9. The lengths of those functional regions must be controlled precisely and in a self-aligned manner.

For CMOS-compatibility and manufacturability, neither lift-off nor Au is allowed for the CMOS front-end process. Nanoscale contact technology on III-V transistor has to be developed.
In addition to the process requirements, the III-V contact must also be able to deliver low film resistivity and low contact resistivity as required by future CMOS devices [4]. Due to those constraints, the source and drain contacts for III-V MOSFET remain a great challenge.

The integration of III-V MOSFETs onto a silicon substrate is also an imperative, but very challenging topic. As shown in Fig. 1-4, the large lattice mismatch between InGaAs and silicon prevents direct pseudomorphic growth. An effective approach for co-integration is still to be developed.

Fig. 1-9: Cross-sectional schematic to illustrate the key features of a planar logic III-V MOSFET.

1.3 Thesis Goals and Organization

III-V MOSFETs displaying well-balanced electron transport, electrostatic integrity and parasitic resistance together with potential for high device density and tight pitch is highly desired. The goal of this thesis is to realize planar InGaAs QW-MOSFETs that possess well-balanced electron transport, electrostatic integrity and parasitic resistance, together with the potential for pitch scaling and silicon integration. Following this, the relevant device physics of InGaAs QW-MOSFETs in digital logic applications will be investigated.
Chapter 2 summarizes the process integration. It first gives an overview of the self-aligned gate-last process. The process is then grouped into five process modules: the contact module, the pre-III-V recess module, the III-V recess module, the gate stack module and the back-end process module. The critical concern and key innovations of each process module are described in detail.

Chapter 3 analyzes the impact of various device architectural designs to the MOSFET performance. The trade-off between drive current and SCE control is identified when the designs of the access region and channel thickness are considered. Following this, device benchmarking with other recently published III-V MOSFETs are presented. The RF characteristics of these devices will be analyzed.

Chapter 4 describes the fabrication and modeling of the tight-pitch MOSFET arrays. Contacts with dimensions down to 40 nm are demonstrated in these MOSFET arrays. A special measurement scheme for the fabricated array structure and its model are developed to study the nanoscale contact behavior. This study reveals extremely low contact resistivity between the Mo contact and n+ InGaAs and shows that more work in this direction is necessary.

Chapter 5 describes the integration of InGaAs MOSFETs on insulator on the silicon substrate through a wafer bonding technique. InGaAs MOSFETs are fabricated and their performance is analyzed.

Chapter 6 investigates the origin of excess OFF-state current in InGaAs QW-MOSFETs. The origin is identified to be band-to-band tunneling, BTBT, coupled with parasitic bipolar gain effect as a result of the floating-body channel. Through TCAD simulations, the physics and mitigation schemes of the excess OFF-state current in InGaAs MOSFETs are explored.

Finally, the conclusions of this thesis and suggestions for future research are presented in Chapter 7.
Chapter 2: Process Development

2.1 Process Overview

Following the guidelines in Ch. 1, this chapter describes a self-aligned gate-last process for scaled InGaAs QW-MOSFETs that emphasizes scalability, performance and manufacturability by making extensive use of dry etch and CMOS-compatible materials. The grand challenges being addressed in this integration scheme are:

1. High-level self-alignment that uses only one mask level to define all the critical transistor dimensions
2. III-V/high-κ gate stack that suppresses gate leakage current and delivers steep subthreshold swing
3. Lateral transistor scaling that produces the smallest gate length and smallest device footprint for any type of III-V FET at the time when they were made
4. Vertical transistor scaling that leads to accurate thickness control with sub-1-nm precision
5. Wet-etch free in the entire process
6. Lift-off free and Au-free in the process front-end

In this chapter, a generic process flow will be first described, followed by a detailed discussion of the critical process unit modules. Processes that are different from this generic process flow will be specified as necessary.

The starting heterostructures are grown by molecular beam epitaxy, MBE, on a 3-inch semi-insulating (100) InP substrate by IntelliEpi Inc. A typical design is described in Fig. 2-1. From bottom to top, it comprises of a ~400 nm thick InAlAs buffer with a silicon δ-doped sheet
placed 5 nm beneath the interface with the channel, an undoped In$_{0.7}$Ga$_{0.3}$As/InAs/In$_{0.7}$Ga$_{0.3}$As channel with corresponding layer thickness of 5/2/3 nm, a 3-nm undoped InP barrier, and a composite n$^+$ cap comprised of n$^+$ InAlAs, n$^+$ InP and n$^+$ In$_{0.7}$Ga$_{0.3}$As. Different heterostructures have been used for process development and will be identified in the discussion.

<table>
<thead>
<tr>
<th>Material</th>
<th>Thickness (nm)</th>
<th>Dopant</th>
<th>Level (/cm$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In$<em>{0.7}$Ga$</em>{0.3}$As</td>
<td>5</td>
<td>Si</td>
<td>3x10$^{19}$</td>
</tr>
<tr>
<td>In$<em>{0.25}$Ga$</em>{0.47}$As</td>
<td>6</td>
<td>Si</td>
<td>3x10$^{19}$</td>
</tr>
<tr>
<td>InP</td>
<td>9</td>
<td>Si</td>
<td>2x10$^{19}$</td>
</tr>
<tr>
<td>In$<em>{0.62}$Al$</em>{0.38}$As</td>
<td>3</td>
<td>Si</td>
<td>2x10$^{19}$</td>
</tr>
<tr>
<td>InP</td>
<td>3</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>In$<em>{0.7}$Ga$</em>{0.3}$As</td>
<td>3</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>InAs</td>
<td>2</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>In$<em>{0.7}$Ga$</em>{0.3}$As</td>
<td>5</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>In$<em>{0.65}$Al$</em>{0.35}$As</td>
<td>5</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>Si (\delta)-doping</td>
<td>Si</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In$<em>{0.52}$Al$</em>{0.48}$As</td>
<td>25</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>InP</td>
<td>6</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>In$<em>{0.52}$Al$</em>{0.48}$As</td>
<td>400</td>
<td>UD</td>
<td></td>
</tr>
<tr>
<td>Si InP Substrate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2-1: Description of the typical heterostructure used in this work. The wafer was grown by IntelliEpi Inc..

The fabrication leverages a self-aligned process developed for high-frequency InGaAs HEMTs [21], [23]. In essence, this is a contact-first, gate-last process in which the metal contacts are formed first, and the intrinsic region is created by etching of the contact and cap layers using SiO$_2$ as pattern mask. The key steps are illustrated in Fig. 2-2. And detailed recipes are described in the Appendix.

Immediately after native oxide stripping by diluted HCl, a W/Mo ohmic contact bilayer is deposited by sputtering. Then SiO$_2$ hardmask is deposited by CVD (Fig. 2-2a). Alignment marks are patterned using E-beam lithography, EBL, in an Elionix ELS-F125 system. The built-in algorithm in Elionix is used for subsequent EBL alignment. It can achieve high precision alignment down to a few tens of nm.
The “gate-foot” patterning that defines the intrinsic gate region is a critical lithography step. Calibrated EBL is used to achieve line features down to 20 nm. Positive E-beam resist Zep520A\(^1\) from Zeon Chemicals is used for this step. Critical exposure parameter is the area dose and must be carefully adjusted due to the following reasons: Firstly, the Rutherford backscattering probability increases with substrate nuclear charge. Hence InP results in relative higher proximity effects than silicon. For the same pattern, the dose for an InP sample should be lower than that for silicon. In this work, the nominal area dose used to exposure Zep for features of a few hundred nm to a few µm is 400 µC/cm\(^2\). Secondly, the proximity effect is also affected by the feature size. Area dose needs to be reduced if line features are above 2 µm because of the increased proximity effect, and to be increased if line features fall below 200 nm. If only long channel devices are fabricated (L\(_g\) ≥ 2 µm), photo lithography is used for gate-foot definition.

After gate-foot patterning, the SiO\(_2\) mask and the W/Mo contact stack are etched by CF\(_4\):H\(_2\) and SF\(_6\):O\(_2\) anisotropic RIE, respectively (Fig. 2-2b) [21]. The conditions of all RIE are listed in Appendix. Then the resist is removed and a CF\(_4\):O\(_2\) isotropic RIE is applied to laterally undercut the ohmic metal in a controlled fashion (Fig. 2-2c) [21], [40]. This is achieved by controlling the etch rate and etch time. Since the gate will be made self-aligned to the SiO\(_2\) edge, the length of the device access regions which separate the contact from the gate can be precisely defined in this step. This step is also referred as the “pull-in”. The above F-based RIEs of W and Mo stops well at the III-V surface. But these steps result in extensive damage in the intrinsic portion of the device. Proper annealing, in this case 340 °C in N\(_2\), is required to repair the damage induced by F-based RIE [41]–[43] in the intrinsic part of the device. In the generic process, it is done in the middle of the III-V recess (after dry etch and before digital etch).

Mesa isolation is patterned using photo lithography and dry etchings. The Mesa dry etchings include the F-based SiO\(_2\) and Mo RIE, and the Cl-based III-V RIE. Following this step, the resist has to be completely removed as the following process step will be at high temperature.

\(^1\) Supply of resist Zep520A was discontinued. Resist GL2000 is used as the replacement.
Fig. 2-2: Illustration of gate recess steps for fabrication of InGaAs QW-MOSFETs. (a) Deposition of ohmic metal and oxide hardmask. (b) Anisotropic SiO₂ and W/Mo etching. (c) W/Mo undercut. (d) III-V cap dry etch. (e) Digital etch. (f) Finished gate.

Fig. 2-3: (a) False-color cross sectional TEM of a QW-MOSFET around gate edge. The final channel thickness of this device is 4 nm. The access region, \( L_{\text{access}} \), and gate-\( n^+ \) overlapping region, \( L_{\text{ov}} \), are indicated. (b) High-resolution TEM, HRTEM, cross-section of the intrinsic gate region. (c) HRTEM of the contact region. Bracket after layer label indicates thickness in nm.
To recess the III-V cap, anisotropic Cl-based RIE is used. This process is not selective to the III-V layers presented in this heterostructure. By calibrating the etch rate, the III-V dry etch can be stopped a few nanometers above the channel (Fig. 2-2d), followed by a RIE annealing [41]. Beyond this, digital etch, DE, is used to precisely bring the recess to the desired depth [44]. This determines the thickness of the channel as well as the thickness of the access region, as shown in Fig. 2-2 (e). DE is a self-limiting process and the etch rate is ~1 nm/cycle. Relying on a calibration method, the lower InP surface can be precisely located. Then, the channel thickness can be controlled with 1 nm precision by counting the required additional digital etch steps.

After the last DE cycle, a fresh semiconductor surface is exposed in a final H2SO4 cleaning step and this is immediately followed by ALD to deposit high-κ dielectric HfO2, Al2O3, or both. The high-κ is deposited at 250°C and there is no high temperature step after gate dielectric deposition. Evaporated Mo is used as gate metal. The deposition is shadowed by the SiO2 hardmask and thus it is self-aligned to the gate-foot pattern. Mo gate electrode is a CMOS compatible technology and has been extensively studied for high-κ metal gate, HKMG, integration in silicon MOSFET [45], [46]. After this, the gate-head is defined by EBL. Residual evaporated Mo is removed by dry etch. Some residual Mo remains on top of the SiO2 hardmask as shown in Fig. 2-2 (f), which would introduce an increased extrinsic overlapping capacitance but this is not a problem when studying I-V characteristics. The device is completed by interconnect via and pad formation, namely the back-end process.

In Fig. 2-2 the front-end gate process steps are divided into four modules. They consist of the contact module, the pre-III-V recess module, the III-V recess module and the gate stack module. Note that the back-end process module is not shown in this figure. The key process development of these process modules is discussed in more detail in the following sections.

The process discussed here has several unique advantages. The entire process is wet-etch free. The process steps in the front-end (before interconnect) are Au-free and lift-off free. The thermal budget is very low. Maximum temperature is 340 °C in the RIE-damage annealing, and no high temperature step is applied after the gate dielectric is deposited.
Fig. 2-3 (a) shows a cross-sectional transmission electron microscopy, TEM, image of the transition region between the intrinsic channel on the left and the contact on the right in a finished device. False color has been added to highlight the different layers. The channel thickness, $t_c$, is 4 nm. The access region is defined as the spacing between the edge of the intrinsic channel and the edge of the contact. Its length is denoted as $L_{\text{access}}$. For this device, $L_{\text{access}}$ is 15 nm. Within this small access region, a gate-$n^+$ overlap, $L_{ov}$, is created by introducing a slight angle and sample rotation during gate evaporation. $L_{ov}$ spans a lateral dimension of about 7 nm. The actual $n^+$-gate overlap distance is slightly longer due to the side-wall angle. This design with a short access region that partially overlaps with the gate saves device footprint, reduces parasitic series resistance, and maintains sufficient isolation between the gate and the source-drain (SD) contacts.

Fig. 2-3 (b) and (c) are high-resolution TEM, HRTEM, images for the intrinsic channel and the extrinsic contact region, respectively. Fig. 2-3 (b) shows that a uniform channel and a flat surface are obtained after III-V recess. The channel is about 7 atomic monolayers thick. This agrees with the targeted channel thickness for this device of 4 nm and showcases the excellent accuracy of this process. The 2.5 nm HfO$_2$ gate dielectric used in this case is also highlighted. Between the dielectric and the channel, an interfacial oxide of 0.5 – 1 nm is observed. This is probably formed during the transition time from the last native oxide stripping and the ALD chamber. It has been reported that ALD-HfO$_2$ using TEMAH precursor, as is the case here, results in a self-cleaning effect on the InGaAs surface. This refers to the partial removal of the native oxide which mitigates Fermi-level pinning at the HfO$_2$/InGaAs interface [47]. In this case, residual native oxide remains on the surface of InGaAs. Fig. 2-3 (c) also shows the Mo ohmic contact on InGaAs. A smooth, un-alloyed contact interface free of any interfacial oxide is evident.

This process has delivered functioning InGaAs QW-MOSFETs with 20 nm gate length and 50 nm contact-to-contact spacing as shown in Fig. 2-4 (a). At the time when the device was first reported, it was the smallest among all planar InGaAs MOSFETs. The output characteristics of this device are shown in Fig. 2-4 (b).
Fig. 2-4: (a) Cross sectional TEM images of a MOSFETs with 20 nm gate length and 50 nm contact-to-contact spacing. (b) Its output characteristics.

2.2 Contact Module

2.2.1 Contact consideration

CMOS devices have stringent requirements for their contacts. The first one is self-alignment. The contact must be defined at the same mask level as the gate. Secondly, materials selection and fabrication procedure should be CMOS-compatible. The contact technology must also be able to deliver low film resistivity and low contact resistivity. In 2018, ITRS HP devices define that the maximum allowable parasitic series resistance is 60 $\Omega \mu\text{m}$ (one side) at a contact size of about 10 nm [4]. Due to that challenging goal and other constraints, III-V contacts remain a difficult research topic to date.

Contact in early n-type InGaAs MOSFET is usually inherited from HEMTs [20], which is based on Au-alloy and fabricated by lift-off [38], [48], [49]. Some MOSFET prototypes featuring self-alignment actually do not include a self-aligned ohmic contact. The level of self-alignment in those devices is merely up to the heavily doped source and drain, SD, made by either implantation [48], [49] or selective epitaxial regrowth [30].
Recently, refractory-metal non-alloyed ohmic contacts on n-type InGaAs HEMT and MOSFETs have made attracted increasing research attention due to their low contact resistivity and process simplicity [21], [23], [26]–[28], [50]. NanoTLM [51] and Fin contact structure [52] were fabricated to study the Mo contact behaviors. Ch. 4 of this thesis develops a new method to study the characteristics of Mo contact with InGaAs MOSFET.

The refractory-metal contacts can be etched by a F-based reactive ion etching, RIE, with high selectivity over InGaAs and therefore this process is an ideal candidate for self-aligned transistor architectures. This was first demonstrated in a self-aligned HEMT with W contact. The MOSFET fabrication in this work has adopted and modified the refractory-metal contact technology developed for self-aligned HEMTs in [21].

### 2.2.2 W barrier for undercut immunity

The self-aligned InGaAs MOSFET features a W/Mo bilayer ohmic contact. This bilayer approach was developed to correct the formation of a deep lateral undercut during F-based RIE in the first-generation devices that had only pure Mo contacts [26], [41]. Fig. 2-5 (a) shows a deep gash of a length of ~100 nm emerging from the edge of the contact. This is found to be originated from an interfacial Mo oxide layer of a thickness of about 10 nm that is formed during SiO₂ deposition. The etch rate of this interfacial layer during F-based RIE is much faster than Mo.

A solution to this problem was found by inserting a thin W cap layer above the Mo in the same sputtering step. W oxidizes more slowly than Mo. The dry etching characteristics of W are similar to Mo in F-based RIE. As shown in Fig. 2-5 (b), F-based RIE does not cause any lateral undercut and yields nearly vertical ohmic contact sidewalls. This is critical for tight-pitch contact fabrication for the device structure to be discussed in Ch. 4. For devices with 20 nm gate length, a 50 nm contact-to-contact spacing has been achieved (Fig. 2-5b).
2.2.3 Metal film resistivity and contact resistivity

The film resistivity is defined as the product of sheet resistance of the metal or metal alloy and its thickness, $\rho = R_{sh} \cdot t$. ITRS specifies the requirement for the sheet resistance and thickness of future silicide [4]. In essence, this is equivalent to the resistivity of the ohmic metal film. It is found that the deposition method and condition play important role in the resistivity of the metal films. Table 2-1 lists the sheet resistance of 30 nm metal films, W or Mo, deposited by evaporation and reactive sputtering. It is noted that the evaporated Mo films have sheet resistance, $30 \ \Omega/\square$, which is high for a film thickness of 30 nm. This is equivalent to a film resistivity of $\rho = 9 \times 10^{-5} \ \Omega \cdot \text{cm}$.

Table 2-1: Comparison of sheet resistance for refractory metals films of 30 nm thickness

<table>
<thead>
<tr>
<th>Machine and conditions</th>
<th>$R_{sh}$ (Ω/□)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ebeam Mo</td>
<td>EbeamAu</td>
</tr>
<tr>
<td>Ebeam Mo</td>
<td>EbeamCMOS</td>
</tr>
<tr>
<td>Sputtered Mo</td>
<td>EML sputterer (pre-sputtering 1 min, base pressure= $10^{-5}$ Pa)</td>
</tr>
<tr>
<td>Sputtered Mo</td>
<td>EML sputterer (pre-sputtering 15 min, base pressure= $10^{-7}$ Pa)</td>
</tr>
<tr>
<td>Sputtered W</td>
<td>EML sputterer (pre-sputtering 15 min, base pressure= $10^{-7}$ Pa)</td>
</tr>
<tr>
<td>Sputtered Mo</td>
<td>TRL sputterer</td>
</tr>
<tr>
<td>Sputtered W</td>
<td>TRL sputterer</td>
</tr>
</tbody>
</table>
Fig. 2-6: Benchmarking of (a) film resistivity $\rho$, [21], [41], [50], [53]–[61]; and (b) contact resistivity $\rho_c$, [28], [50]–[52], [55]–[58], [60]–[66], for different contact technologies proposed for n-type InGaAs MOSFETs (see text).

In the early process, the sputtered Mo was deposited with conditions typical to other metal films in the sputtering system: base pressure of $10^{-5}$ Pa after a 3-hour main-chamber pump time, pre-sputtering for 1 min with a closed shutter. The obtained Mo film displayed $R_{sh}$ of 17.7 $\Omega/\square$ for a film thickness of 30 nm. This is equivalent to a film resistivity of $5.3 \times 10^{-5} \Omega \cdot \text{cm}$. This
value is lower than that obtained by evaporation. Further optimization is carried out by bringing down the base pressure to $10^{-7}$ Pa through a 15-hour pump time. An \textit{in-situ} pre-sputtering of 20 min is also applied immediately before the device run. With this additional optimization, the Mo resistivity is reduced to $5 \Omega/\square$ for the same film thickness. Low resistivity is also obtained for W films after similar deposition conditions. After optimization, the equivalent film resistivity of W and Mo is about $1 - 2 \times 10^{-5} \Omega \cdot \text{cm}$. This represents 80% reduction with respect to the evaporation method. The oxygen content of the final Mo film is related to the level of oxygen contamination, and the microstructure of the films [53], [67].

The contact technology studied in this work is benchmarked against other technologies recently published for n-type InGaAs MOSFETs or other similar devices in Fig. 2-6. The comparison is between refractory metals, non-Au based InGaAs metal alloys, and Au-based contacts. Two properties of the contact metal are graphed: the metal film resistivity, $\rho$, and the contact resistivity, $\rho_c$.

The film resistivity $\rho$ is shown in Fig. 2-6 (a). For comparison, the 2018 ITRS requirement is also specified: $\rho=1.6 \times 10^{-5} \Omega \cdot \text{cm}$, equal to that of Nickel Silicide [4]. The film resistivity comparison reflects the fact that, most of the ohmic metals or metal alloys used in InGaAs MOSFETs to date are much beyond the acceptable value for $\rho$. The typical film resistivity of Ni-InGaAs, Co-InGaAs and Pd-InGaAs are between $7 - 15 \times 10^{-5} \Omega \cdot \text{cm}$, about 5 – 10 times the ITRS requirement. In comparison, the refractory metals in this work have achieved film resistivity that satisfies the ITRS requirement. As mentioned, the deposition method plays an important role in determining the resistivity [53], [67]. The high-resistivity Mo films in Fig. 2-6 (a) are deposited by evaporation [41].

The reported results for the contact resistivity of InGaAs alloys and Au-based contacts on n-type InGaAs vary dramatically from $5 \times 10^{-9}$ to $6 \times 10^{-3} \Omega \cdot \text{cm}^2$. They are summarized in Fig. 2-6 (b). The lowest contact resistivity of $4 \times 10^{-9} \Omega \cdot \text{cm}$ is reported for Pd-InGaAs [63], and $5 \times 10^{-9} \Omega \cdot \text{cm}$ for Au-based contact [62]. In this work, contact resistivity is extracted from tight-pitch MOSFET arrays to be discussed in Ch. 4. A contact resistivity of $\rho_c=8 \times 10^{-9} \Omega \cdot \text{cm}^2$ has been
obtained. This is consistent with several other experiments using different test structures for the Mo contact on n\textsuperscript{+} InGaAs [51], [52], [66]. The extremely low contact resistivity is an attractive property of Mo contacts.

2.3 Pre-III-V Recess Module

2.3.1 RIE damage and its recovery

In this MOSFET fabrication, RIE is extensively used for pattern definition to achieve a high-degree of self-alignment. In this regard, RIE damage is a serious concern. An earlier simplified version of this process was used to study RIE-related damage and its recovery. The simplified heterostructure has 15 nm unstrained In\textsubscript{0.53}Ga\textsubscript{0.47}As channel. The cap is made of one single layer of n\textsuperscript{+} In\textsubscript{0.53}Ga\textsubscript{0.47}As. A relaxed gate stack design is used. The gate insulator consists of 4-nm Al\textsubscript{2}O\textsubscript{3} and 3-nm InP. The III-V cap dry recess is replaced by selective wet etch. This experiment is carried out to compare the device characteristics with and without the damage anneal. The anneal condition is 340 °C for 15 min and it is performed immediately before cap wet etching [41]. This experiment has showed that RIE damage can indeed be severe, but it can be eliminated to some extends by a pre-gate anneal step.

Fig. 2-7 (a) shows transfer characteristics of a 2-\textmu m-gate length self-aligned QW-MOSFET without and with the RIE damage anneal. RIE results in severe damage in the device and leads to performance degradation. The device without damage anneal shows the peak transconductance, \( g_{m,\text{max}} \), of 49 \( \mu \text{S/\mu m} \) at \( V_{ds} = 0.5 \) V. For devices with RIE damage anneal, \( g_{m,\text{max}} \) improved to 205 \( \mu \text{S/\mu m} \) at \( V_{ds} = 0.5 \) V. The RIE damage anneal also shifts the threshold voltage toward the positive direction. Due to the F-based RIE, increased ON-resistance is observed. The output characteristics of the QW-MOSFET are shown in Fig. 2-7 (b). Without damage anneal, \( R_{on} \) of the 2-\textmu m-L\textsubscript{g} transistor is about 10 K\Omega\cdot\mu m and that with damage anneal is dramatically reduced to 2.5 K\Omega\cdot\mu m.
Fig. 2-7 (c) shows significant subthreshold swing, S, improvement with RIE damage anneal. S for the device without RIE damage anneal is 300 mV/dec. It improves to 95 mV/dec after the anneal is applied. Fig. 2-7 (d) shows channel mobility before and after anneal. While the RIE damage does not affect significantly the mobility at high carrier density, presumably because the damage-induced charges are screened, the effect is significant at low density. This is consistent with the significant impact on the ON-resistance (Fig. 2-7b). The peak mobility is 2780 cm$^2$/V.s for the annealed sample, which compares well with the results reported by other authors in similar non-RIE’ed buried-In$_{0.53}$Ga$_{0.47}$As-channel devices.

Fig. 2-7: Electrical characteristics of QW-MOSFET with (blue) and without (black) RIE damage anneal: (a) Output characteristics, (b) Transfer and transconductance characteristics, (c) Subthreshold characteristics, (d) Effective field mobility.
Fig. 2-8 shows the capacitance vs. $V_g$ of two finished MOSFETs (a) with RIE damage anneal and (b) without RIE damage anneal. The gate lengths of these devices are 20 µm. The capacitance is measured between the gate and the shorted source and drain. The incorporation of damage annealing yields a larger capacitance in the ON-state and lower frequency dispersion around threshold. This is an indication of reduction in the interface traps in the midgap. This observation is also consistent with the threshold voltage shift, mobility degradation and subthreshold swing degradation in the devices without RIE-damage annealing.

The incorporation of Fluorine and Oxygen in these types of heterostructures is known to cause reductions in carrier concentration and mobility [43], due to dopant passivation and increased scattering. This work shows that RIE also severely impacts the subthreshold characteristics, suggesting additional interface damage. A relatively benign thermal step before gate oxide deposition could mitigate this damage to some extent.
2.3.2 Gate edge roughness control

In the gate recess, SiO₂ and W/Mo are etched by F-based RIE. It is observed that the etched pattern is significantly influenced by the RIE conditions. Since the SiO₂ hardmask is used for gate and contacts patterning, the line roughness would translate to gate length variation. Severe gate length variation has been observed in HEMT devices fabricated by a similar SiO₂ hardmask process [40], and it limits the smallest gate length that can be made. The gate length variation also introduces complications in device analysis. Severe gate edge roughness is also observed in the fabrication of the early self-aligned InGaAs MOSFETs in this work. The origin of this problem was investigated in the course of this work.

Fig. 2-9: Optimization of the gate edge roughness. Cross sectional SEM images for the ZEP/SiO₂/Ohmic patterns after SiO₂ RIE with gas combination of (a) CF₄:H₂ =3:1, (b) CF₄:H₂ =4.5:1, (c) CF₄:H₂ =6:1. (d) Smooth gate edge line of 20 nm is obtained from the optimized dry etchings.
H₂ is usually added to SiO₂ RIE to enhance the selectivity and anisotropicity. It is found that the H₂ to CF₄ ratio plays important role in the etch profile when the Zep is used as resist. If the H₂ flow rate is too high, it results in significant resist crosslinking and degrades the line features. Fig. 2-9 shows the etched patterns under different ratio of gas flow rate. The CF₄ flow rate is fixed at 25 sccm and that of H₂ varies. The initial resist thickness is the same for all cases: 250 nm Zep. For CF₄:H₂ =3:1, the resist deforms severely during RIE as shown in Fig. 2-9 (a). The deformed resist shadows the subsequent RIE and results in incomplete removal of SiO₂. For narrow-line pattern, for instance 20 nm in Fig. 2-9 (a), the gate region will be completely shadowed.

When the ratio of CF₄:H₂ flow rate increases to 4.5:1, the resist deformation is less (Fig. 2-9b). At a ratio of 6:1, the resist is free from any deformation and pattern transfer achieves maximum fidelity. In this case, the etching selectivity is 2.5:1, higher etch rate for Zep. As for the case in Fig. 2-9 (c) 100-nm-thick resist remains when the oxide is etched through, indicating thinner resist is possible. Thin resist is required in the fabrication of tight-pitch MOSFETs which are discussed in Ch. 4. By understanding the origins of gate edge roughness and optimizing the RIE, extremely smooth and small recessed gate lines can be fabricated (Fig. 2-9d).

### 2.3.3 Precise lateral control for access region

The ohmic metals, W and Mo, are etched in two steps, an anisotropic RIE and an isotropic RIE. When only anisotropic RIE is applied, W/Mo is perfectly self-aligned to the SiO₂. Fig. 2-10 (a) shows an example of a fabricated device. Subsequently, the ohmic metals can be pulled in with respect to the SiO₂ edge with an isotropic RIE. This is achieved by adding O₂ during RIE [22], [40]. This device is shown in Fig. 2-10 (b).

The isotropic RIE results in the following consequences. Firstly, the RIE time controls the extent of the lateral recess which eventually sets the final length of the device access regions (L_access). When only anisotropic RIE is used, the minimum L_access achieved is 15 nm (Fig. 2-10a). Long isotropic pull-in results in increased access region length, as shown in Fig. 2-10 (b),
$L_{\text{access}} = 80$ nm. This forms two air spacers at the two sides of the gate, which maintains electrical isolation between metals in the contact and gate. Secondly, after the ohmic metal is pulled in, the III-V surface in the access region is also exposed, and this surface will be subjected to subsequent digital etch. As a result, in the access region, the III-V cap will be thinned down and the sheet carrier density will be reduced.

![TEM cross sectional images](image)

Fig. 2-10: TEM cross sectional images for: (a) $L_g = 50$ nm device with $L_{\text{access}} = 15$ nm, and (b) $L_g = 70$ nm device with $L_{\text{access}} = 80$ nm.

### 2.4 III-V Recess Module

#### 2.4.1 Drawback of conventional III-V recess

Selective wet etching is commonly used for semiconductor recess in HEMT fabrication. Due to the isotropic nature of wet etching, this usually creates a lateral undercut of about 100 nm at both sides of the gate and results in an “underlap” device architecture as shown in Fig. 1-8 of Ch. 1. The early version of MOSFET fabrication in this work adopted a similar selective wet etch method to remove the cap in the gate region. Nevertheless, the recipe was optimized to achieve minimum wet-etch undercut. The etchant solution is mixed $\text{H}_2\text{O}_2$:citric acid as shown in the recipes in the Appendix. By carefully adjusting the ratio of these two etchants and controlling the immersion time, the lateral undercut is minimized. The smallest undercut achieved is 30 nm as shown in Fig. 2-11. However, the wet-etch undercut region is uncapped (not covered by the $n^+$
cap) and ungated (not covered by the gate). As a result, the series resistance of such device remains a severe concern especially under high current condition. The evidence is shown below.

The transfer and transconductance characteristics of a 60-nm-Lg MOSFET fabricated by the wet-etched-cap process are shown in Fig. 2-12 (a). The gate insulator is 1 nm InP plus 2 nm HfO2. The peak transconductance, $g_{m,max}$, at $V_{ds}=0.5$ V is 1.53 mS/µm. The drain current corresponding to $g_{m,max}$ is 0.2 mA/µm. The transfer curves show negligible hysteresis, indicating high-quality gate insulator with low bulk traps and interface traps. Peak $g_m$ occurs at 0.2 mA/µm. Rapid drop in $g_m$ is observed when current is above 0.2 mA/µm. This experimental observation will be named the “$g_m$ compression” in this thesis. Fig. 2-12 (b) shows the output characteristics of this device. Although the device exhibits sharp saturation and good turn off, relatively high $R_{on}$ of 485 $\Omega$.µm is observed. $R_{on}$ is measured at $V_{gs}-V_t=0.6$ V. In the output characteristics at a current level above 0.2 mA/µm, the increase of current from incremental $V_{gs}$ starts to saturate, which is consistent with the “$g_m$ compression” in Fig. 2-12 (a). The subthreshold and gate current characteristics are shown in Fig. 2-12 (c). A subthreshold swing of 89 mV/dec is obtained at $V_{ds}=0.05$ V, and 102 mV/dec at 0.5 V. Those are low values for the gate length of this device. Drain induced barrier lowering, DIBL, is 156 mV/V.

It is expected that the existence of this uncapped access region causes severe ON-state current degradation. At high current level, the $g_m$ compression could be the result of source starvation due to low carrier density in the access region that results in non-equilibrium “source contact” [68]–[70]. For short channel effect control, the low carrier concentration in the underlap device architecture allows electric field from the gate to penetrate into the access region. Hence the effective channel length is longer than the physical gate length, $L_{eff}>L_g$. This explains the superior short channel effect control of this device.
Fig. 2-11: Cross sectional TEM image of QW-MOSFET with $L_g=30$ nm. The semiconductors around the gate region are recessed by selective wet etching.

Fig. 2-12. Performance of a $L_g=60$ nm QW-MOSFET fabricated by selective cap wet etch. (a) Transfer and transconductance characteristics. (b) Output characteristics. (b) Subthreshold characteristics.
In the perspective of process integration, wet-etch is incompatible with CMOS manufacturing because it suffers from process variability and nonscalable device footprint. Hence, the conventional wet-etched gate recess process must be replaced in the III-V MOSFET for logic applications. To realize this, a new two-step III-V recess that combines III-V dry etch and digital etch is introduced in this work.

2.4.2 III-V dry etch

The development of III-V dry etch is described here. The Cl-based III-V dry-etch originated on the III-V nanowire and III-V fin fabrication [71], [72]. The conditions have been modified to produce a smooth and trench-free recess profile. The III-V dry etch is carried out in a SAMCO RIE-200iP ICP system. The main conditions of the baseline recipe are: gas flow Cl2:N2 of 10:3 sccm, pressure of 0.2 Pa, ICP power of 20 W, substrate bias voltage of 130 V, and chuck temperature of 120 °C. Other conditions are shown in the Appendix. Under these conditions, the etch rate for the composite InGaAs/InAlAs/InP cap is about 0.2 nm/s. The resulting surface roughness, RMS, is 0.29 nm, slightly higher than the virgin wafer of RMS=0.17 nm (Fig. 2-14 a,b). A larger bias voltage results in a higher etch rate (Fig. 2-13), but at the expense of increased surface roughness (Fig. 2-14e). The threshold bias for etching is 60 V (Fig. 2-13). The recipe shows very little temperature dependence for surface roughness (Fig. 2-14 c,d) and etch rate.

Besides bias voltage, increasing the N2:Cl2 gas ratio and the pressure both roughen the surface (Fig. 2-14 f,g) at a similar etch rate. A BCl3-based etch recipe suitable for high-aspect-ratio nanowire fabrication [72] was also investigated and discarded due to significant surface roughening (Fig. 2-14 h).

The use of high bias also introduces significant trenching at the two edges of the recess pattern. Trenching is caused by ion deflection during dry etch and is commonly observed in high-aspect-ratio III-V dry etch [73]. Fig. 2-14 shows cross section SEM images after dry etch using (a) baseline conditions and (b) a high bias condition of 234 V. Significant trenching is evident for the dry etch in the later bias.
Fig. 2-13: III-V cap etch rate as a function of substrate bias voltage. Threshold bias voltage for etching is about 60 V.

Fig. 2-14: AFM surface scan for 1 \( \mu m^2 \) area on: (a) virgin wafer, (b) dry-etched sample under baseline conditions, (c) 40°C, (d) 150°C, (e) increased bias of 234 V, (f) increased ratio of \( N_2:Cl_2 \) flow rate, (g) increased pressure and (h) \( BCl_3 \)-based recipe [72].
Fig. 2-14: Cross-section SEM images of dry etched samples using: (a) baseline III-V etch conditions, and (b) higher bias condition at 234 V. The recess opening is 100 nm.

Fig. 2-15: TEM cross sectional images of finished self-aligned InGaAs MOSFETs: (a) etched under baseline conditions with a final 4 nm channel thickness; (b) etched at a higher substrate bias (140 V) and a final 8 nm channel thickness. Increased bias results in observable trenching at the two ends of the channel.

Fig. 2-15 shows cross sectional TEM images of two finished devices etched under different conditions. The two devices have a similar gate length of about 50 nm. The device recessed at high bias of 140 V in Fig. 2-15 (b) shows slight trenching when compared with the flat channel observed in Fig. 2-15 (a) which is obtained at 120 V. These two devices are reported in [27] and [28] respectively. This indicates that trenching, as consequences of dry etch, will carry on to the finished device and most likely affect the device characteristics.
2.4.3 Digital etch

Digital etch techniques have been proposed to etch silicon and III-Vs [9-12]. In this approach, the two elemental components of etching, oxidation and oxide removal, are applied separately. For III-Vs, a typical oxidation agent is an H₂O₂ solution while the native oxide is stripped by acids [74], [75]. This scheme can result in chemical cross contamination that leads to poor process repeatability. Besides, with an H₂O₂ solution as oxidant, this technique only applies to InGaAs and cannot be used for InP which greatly limits its usefulness. Recently, UV ozone was used for oxidation, followed by HCl wet etch [76]. This addresses the cross contamination issue, but still cannot be used for multilayer structures that include InP. Here, a new digital etch process is demonstrated that allows the precise non-selective etching of InAlAs, InGaAs, and InP layer structures with sub-1 nm precision.

The digital etch approach in this work consists of native III-V oxide formation through exposure to low power oxygen plasma in a commercial asher system. This is followed by oxide removal in diluted H₂SO₄, H₂SO₄:H₂O=1:1, for 30 s at room temperature followed by DI rinse. By repeating these two steps, the semiconductor can be thinned down in a controlled manner.

![Fig. 2-16. Illustration of InP digital etch rate calibration.](image)

Fig. 2-16 illustrates the approach to calibrating the etch rate of the thin InP barrier in the gate recess region of a buried-channel MOSFET. First (Fig. 2-16a), the top InGaAs cap is removed by well-established selective wet etching, leaving a gate recess opening. Following this, multiple cycles of digital etch are performed (Fig. 2-16 b,c). To calibrate the etch rate, one needs
to know when the InP barrier is breached. For this, after a given number of cycles, a small portion of the sample is cleaved and dipped in H$_2$O$_2$:H$_3$PO$_4$:H$_2$O (1:1:25) for 30 s. In this calibration, this solution etches InGaAs over InP with a selectivity of at least 100 to 1. If even a small amount of InP (>0.5 nm) is left (Fig. 2-16c), this etchant only removes the InGaAs cap and the sample appears almost perfectly flat under an optical microscope (Fig. 2-16d). On the other hand, if the InP barrier has been breached, the H$_3$PO$_4$ solution produces readily visible trenches in the underlying InGaAs layer (Fig. 2-16e). If the thickness of the InP layer is known from MBE calibrations or TEM, the etch rate can then be determined.

Fig. 2-17 shows measured InP digital etch rate per cycle as a function of oxygen exposure time. The InP etch rate increases for the first 150 s of oxidation time but it then saturates to a rate of about 0.9 nm/cycle. Since the oxide removal step completely removes the grown oxide, the etch rate is directly proportional to the oxide thickness. Therefore, the saturation behavior in Fig. 2-17 reflects the self-limiting nature of the oxidation process in which once the oxide thickness is thick enough, the diffusion of oxidant species is sharply curtailed. For much longer oxidation time, the sample heats up and the oxidation rate increases somehow. The data between 50 to 500 s follows Lukeš’ rate law for GaAs oxidation:

\[ d=A+B\ln(t+t_o) \]

with fitting parameters \( A \approx 0 \), \( B = 0.19 \text{ nm/dec} \), and \( t_o = -37 \text{ s} \) [77]. The negative \( t_o \) value reflects the incubation time of the O$_2$ plasma oxidation process.

The error bars in Fig. 2-17 reflect two sources of uncertainty. First, the discrete nature of digital etch cannot resolve fractions of an etch cycle. Using a thicker InP layer for calibration can reduce this error. Second, non-uniformities in the etch rate or layer thickness can result in the sample not clearing in its entirety within the same etch cycle. The uncertainty in the etch rate seen in Fig. 2-17 is not a significant problem when it comes to precise etching of a given structure as shown below.
Fig. 2-17: Etch rate per cycle of digital etch as a function of oxidation time under O$_2$ plasma, and fitting with Lukeš’ model [77].

2.4.4 Precise channel thickness and access thickness control

The full gate recess developed in this work consists of two steps. The concept is depicted in Fig. 2-18. For illustration purposes, a simplified III-V layer structure is used with an n$^+$-InGaAs cap, a thin InP etch stop layer and an InGaAs channel. In a first step, the III-V dry etch is used to remove partially the cap. It stops a few nm above the InP layer. This is achieved by time control for a given etch rate. From this point on, digital etch is used to remove material in a self-limiting manner at an etch rate of $\sim$1 nm per cycle. Multiple cycles of digital etch are performed until the desired final recess depth is reached.

Fig. 2-18: Precise etching process of the intrinsic region of the device is performed by RIE (left) and DE (right). The RIE stops a few nm above the channel surface. The final channel thickness is controlled by DE with 1 nm precision.
Through this method 1 nm layers thickness control can be achieved. But it does not automatically result in accurate etch depth control. The reason is because the III-V dry etch and digital etch are lack of selectivity to the III-V materials in this heterostructure. To achieve accurate etch depth control, a novel calibration method is used.

It is again relied on precisely locating the bottom InGaAs/InP interface right above the channel to know when to stop the digital etch. The method is modified from the procedure in determining the etch rate of digital etch as illustrated in Fig. 2-16. Its sequence is shown in the flowchart in Fig. 2-19. A test sample from the same heterostructure is patterned with photoresist
and joined to the real device sample for III-V dry etching (Fig. 2-18a). The photoresist is then removed by acetone. This test sample, as shown in Fig. 2-19, is used to determine the number of digital etch cycles “N” that is needed to reach the lower InP interface from the point left after dry etch, a procedure identical to that of Fig. 2-16. Different from the etch rate test, this information is used differently. When “N” is known, it locates the bottom InGaAs/InP interface in the device sample with a precision equal to that of digital etch, about 1 nm. With this information, one can then estimate the number of cycles that it takes to achieve a given channel thickness, whether buried channel or surface channel, in the actual device sample.

Fig. 2-20: Cross-sectional TEM images of two different runs with different final channel thickness targets when combining dry etch and digital etch: (a) 4 nm and (b) 8 nm. A separate example achieved when combining wet etch and digital etch for a total buried channel thickness of 11 nm. High resolution TEM image (a-HR) shows 7 monolayers remaining in the intrinsic channel for (a), and (c-HR) shows 18 monolayers for (c).

This method has been verified by examining the cross section of finished devices through TEM in different runs with different final channel thickness targets, as shown in Fig. 2-20 (a)
and (b). Both of the devices are built on the same heterostructure that has an as-grown channel thickness of 10 nm as seen in the extrinsic portion of the devices. The intrinsic gate region is recessed by a combination of dry etch and digital etch. The respective TEM images confirm the targeted channel thicknesses, 4 nm and 8 nm, respectively. In the sample with channel thickness of 4 nm, the high-resolution TEM shows that the channel is 7 atomic monolayers thick (Fig. 2-20 a-HR).

The same calibration will work if the first step of III-V dry etch in Fig. 2-19 is replaced by wet etch. An example is shown in Fig. 2-20 (c) and Fig. 2-20 (c-HR). This sample has the same barrier and channel design as the baseline device heterostructure in Fig. 2-1, but a modified cap structure. From bottom to the top, the cap includes $n^+$ InAlAs (3 nm), $n^+$ InP (3 nm) and $n^+$ In$_{0.53}$Ga$_{0.47}$As (25 nm). In the first step (N=0), selective wet etch is used to remove the 25 $n^+$ In$_{0.53}$Ga$_{0.47}$As and stops in the $n^+$ InP. Following this, digital etch is applied to remove the III-V until 1 nm of i-InP is left. This results in 18 monolayers left in the channel (Fig. 2-20 c-HR). The imaging method depicted in Fig. 2-19 has the advantage of being fast (w.r.t. TEM) and accurate (w.r.t step profiling).

As discussed in the pre-III-V recess module, the lateral pull-in of W/Mo also exposes the III-V surface in the access region. Since the digital etch is an isotropic process, the exposed access region subjected to digital etching will also be thinned down. By carefully adjusting the amount of dry etch and digital etch, one can simultaneously determine the channel thickness, $t_c$, and access region thickness, $t_{access}$. This is clear in Fig. 2-10. The long-access device shows a $t_{access}$ only half of the cap thickness. Consequently, the sheet charge density in the access region is significantly lower than that in the cap. The estimated sheet charge density of the long-access device in Fig. 2-10 (b) is to be $2 \times 10^{12}$ cm$^{-2}$. The access region design has strong impact on the device ON- and OFF-state performance, as will be discussed in Ch. 3.
2.5 Gate Stack Module

Gate stack engineering constitutes an important component for MOSFET research. High-k metal gate, HKMG, has been widely used in the recent generation of CMOS devices. In this thesis, HKMG/InGaAs gate stack has been investigated through four generations of process development as discussed below.

2.5.1 EOT scaling

Fig. 2-21 and Fig. 2-22 show the long-channel subthreshold $I_d-V_{gs}$ characteristics and gate leakage density $J_g-V_{gs}$ characteristics of the four generations of gate stacks being investigated in this work, namely G1-G4. The cross sectional schematics of the gate stacks are shown in the respective inset. The bracket indicates the dielectric and barrier from which the Equivalent Oxide Thickness, EOT, is calculated. EOT is based on the thickness and dielectric constant of the respective materials. An interfacial oxide of 0.5–1 nm can be formed between the semiconductor and deposited dielectric (Fig. 2-3). This is not included in the EOT calculation here. The relative dielectric constants, $\varepsilon_r$, of the ALD dielectrics are determined from silicon MOS capacitors fabricated in the same ALD system. The relative dielectric constant $\varepsilon_r$ of Al$_2$O$_3$ is 7~8 and that for HfO$_2$ is 20~23. For InP barrier, the value of static dielectric constant is used, $\varepsilon_r=13$. For the “channel” indicated in the insets of Fig. 2-21, devices of G1 and G2 use 15 nm unstrained In$_{0.53}$Ga$_{0.47}$As, and G3 and G4 use the composite channel given by Fig. 2-1. Besides, G1-G3 use wet etch for cap recess and single Mo as ohmic contact. G4 uses the dry etch/digital etch for cap recess and W/Mo composite ohmic contact. Although the channel configuration and extrinsic portion in those devices are different, in the study of gate stack integrity in which interface characteristics and gate leakage suppression are the main concern, those factors become less relevant.

In the first generation (G1), the main objective is to develop a complete gate-last integration scheme. Gate leakage suppression is the main concern. Hence, a relaxed gate stack
with thick Al₂O₃ (4 nm) and thick InP barrier (3 nm) is used. The total EOT for this gate stack is 3.5 nm. For gate length of 2 µm, the subthreshold swing is 95 mV/dec. The gate leakage for this gate stack is extremely low. J₉ is below 5x10⁻⁶ A/cm² for V₉ span from -0.5 V to 0.5 V. This gives much room for further EOT scaling.

In the second generation, G2, gate length and EOT scaling are performed simultaneously. The high-κ dielectrics in G2 comprise of 2 nm HfO₂ and a thin (0.5 nm) Al₂O₃. This maintains the same interface with III-V as G1 but increases the dielectric constant. The total EOT is reduced to 0.9 nm. As a result, although the L₉ is reduced by more than 10X from 2 µm to 140 nm, a reduced subthreshold swing of 85 mV/dec is still obtained. The cost for this is a notable increase in gate leakage as shown in Fig. 2-22. Nevertheless, J₉ in G2 remains an acceptable value below 10⁻² A/cm² for the entire gate voltage range.

Before end of 2012, most InGaAs MOSFET prototypes used Al₂O₃ for the interface with III-V [30], [33], [54] because it was believed that Al₂O₃ can form a better interface with InGaAs. Some research in early 2012 reported that HfO₂ can form an III-V interface with low interface traps as well [78], [79]. In the third generation of gate stack engineering, G3, the Al₂O₃ is removed. The resultant gate stack is shown in Fig. 2-21 (G3). The total EOT is further reduced to 0.6 nm. It is observed that at the same gate length, L₉=140 nm, the subthreshold swing remains unchanged. This indicates HfO₂ can be directly used on III-V and result in similar performance as Al₂O₃. At the same time, the gate leakage only increases slightly. This was the first demonstration of using pure HfO₂ gate dielectric on III-V to achieve scaled device down to L₉=30 nm. This set of devices matched the record of subthreshold swing and ON-current at the time when it was made. Since then, there have been increasing reports of using HfO₂ as gate dielectrics for high-performance scaled InGaAs MOSFETs [80], [81]. This confirms that the HfO₂ direct integration scheme is a promising option.

From G1-G3, the dielectrics lie on an InP barrier resulting in buried channel devices. The new III-V recess in this work enables precise removal of the III-V layers down to 1 nm. Hence the InP and part of the InGaAs channel can be removed in a controlled manner. In the fourth
generation, G4, the InP is removed through additional digital etch. The HfO₂ dielectric is directly on the InGaAs channel, resulting in a surface-channel transistor. Fig. 2-21 reflects that the HfO₂/InGaAs gate stack results in almost identical subthreshold characteristics as that with Al₂O₃/InP and HfO₂/InP at the same gate length of L_g=140 nm. By removing the InP, the total EOT is further reduced to 0.5 nm. Trade-offs between the transport and SCE start to emerge when comparing the buried channel and surface channel designs. This will be discussed in Ch. 3.

Despite of the aggressive EOT scaling over generations, the gate leakage remained low as shown in Fig. 2-22. The maximum gate leakage for the given gate voltage span is below 10⁻²
A/cm\(^2\). The III-V MOSFETs shown here reduce the gate leakage density by at least 4 orders of magnitude with respect to HEMT, in which the gate leakage can be above 10\(^2\) A/cm\(^2\) in the ON-state [23], [40].

![Gate leakage density, J\(_g\), of the four gate stacks investigated in this work.](image)

2.5.2 Interface consideration

Long channel InGaAs MOSFETs with \(L_g\) from 2 \(\mu\)m to 300 \(\mu\)m are fabricated for G2 gate stack. Due to the long gate length, those devices are free from short channel effects and can serve as ideal test structures for the study of III-V/high-\(\kappa\) interface. In the long-channel MOSFETs, \(L_g=300\ \mu\)m, nearly ideal \(S\) of 69 mV/dec at \(V_{ds}=50\ \text{mV}\) is obtained (Fig. 2-23a). This result is among the best ever reported for planar III-V MOSFETs. As shown in Fig. 2-23 (b) that graphs \(S\) vs. dielectric EOT for long-channel III-V FETs at low \(V_{ds}\), this is the second best value among all planar InGaAs MOSFETs. The best value was 66 mV/dec from Ref. [82].

Fig. 2-23 (b) also illustrates the importance of \(D_n\) and dielectric scaling. To the first order, the subthreshold swing of the MOSFET is related to its dielectric and interface trap density by the following equations:

\[
S = 60 \left(1 + \frac{C_{it}}{C_{ox}}\right)
\]

\[
= 60 \left(1 + \frac{t_{EOT}q^2D_{it}}{\varepsilon_0\varepsilon_{SiO2}}\right) \tag{2-1}
\]
$C_{it}$ is the capacitance associated with the interface traps at an energy level around midgap, numerically equal to $q^2D_{it}$. $C_{ox}$ is the oxide capacitance for the deposited high-$\kappa$ dielectric(s) and does not include the InP barrier. This expression has neglect the impact of the quantum effect that will affect the capacitance and charge centroid. But in the subthreshold region, the this impact will be small. In spite of multiple RIE steps, F-based RIE in this case, in the intrinsic gate portion of the devices, the gate stack still delivers a high-quality interface. As indicated in Fig. 2-23 (b), the midgap $D_{it}$ of the gate stack is between mid-$10^{12}$ and $1\times10^{13}$ cm$^{-2}$eV$^{-1}$.

![Fig. 2-23](image)

Fig. 2-23: (a) Subthreshold characteristics of a long-channel MOSFET with G2 gate stack. The gate length is 300 $\mu$m. (b) Low-voltage subthreshold swing vs. dielectric EOT of several planar long-channel III-V MOSFETs and HEMTs.

## 2.6 Back-end Process Module

The final process module for device fabrication is the back-end process. It is done after the gate fabrication. It comprises of interconnect via and pad formation. The back-end process should be carefully designed so that it does not introduce additional series resistance or damage the intrinsic device. In the earlier version of device fabrication, diluted hydrofluoric acid, HF, was used to etch the oxides and open the interconnect via. The oxides to remove include SiO$_2$.
hardmask and the ALD high-κ, HfO2 and Al2O3. High-κ dielectrics is hard to remove by HF [83]. Because of this, residual oxides can remain on the ohmic surface that contributes to the total series resistance. In addition, HF can attack the resist during wet etch. If this happens, the HF can reach the intrinsic device and damage it.

To solve this problem, a wet-etch free via process is adopted. Many researches have been devoted to investigate dry etching method of HfO2 and other high-κ materials [84], [85]. The Cl-based RIE is used here. The process flow is as followed. A double layer resist is first patterned by EBL as shown in Fig. 2-24 (a). The HfO2 is first opened through Cl-based RIE. Then the SiO2 are etched by the previous F-based RIE. By properly adjusting the etching condition (in Appendix), the second RIE for SiO2 has high selectivity (>20) over the refractory metals W and Mo. Sufficient overetch is used to remove the dielectrics completely and it will not cause damage in the ohmic metals. Cross-sectional SEM images after via opening are shown in Fig. 2-25 (a). The high-κ and SiO2 are completely removed. As seen, a clean W surface is observed without any damage or residual material. The via pattern is self-aligned to the top layer of the Zep resist. To realized a self-aligned via filling, PMGI is added beneath the Zep so that it create a retrograde profile as shown in Fig. 2-25 (a). The same lithography step can be used to lift-off the evaporated pad as shown in Fig. 2-24 (b) and (c) and Fig. 2-25 (b). Pad formation is the only lift-off step used in the entire device fabrication.

Fig. 2-24: Cross sectional schematic for via process: (a) double resist for via opening through selective RIE, (b) pad evaporation, (c) pad lift-off.
2.7 Chapter Summary

The conventional III-V fabrication process is not suitable for CMOS as discussed in Ch. 1. To address these problems, the design and fabrication of new type of self-aligned recessed-gate InGaAs MOSFETs are discussed. The approach in this work takes into consideration scalability, performance, and CMOS-compatibility. The process is wet-etch free, Au-free and lift-off free in the front-end. It has high-level self-alignment and very low thermal budget. The key feature and development of five process modules is discussed in details.

In the contact module, W barrier is added between the SiO₂ and Mo to prevent the formation of a long lateral undercut during RIE. Optimization of the deposition technique results in 80% reduction in the sheet resistance of the W and Mo films. The Mo contact resistivity on n⁺ InGaAs is benchmarked against other contact technologies. The contact resistivity value of \( \rho_c = 8 \times 10^{-9} \, \Omega \cdot \text{cm}^2 \) is found to be among the lowest ever demonstrated. The low film resistivity and low contact resistivity highlight the promise of refractory metals contacts.

In the pre-III-V recess module, SiO₂ and Mo are etched by F-based RIE. The RIE process can achieve extremely smooth line edge for small gate line down to 20 nm. The F-based RIE
damage is observed and a method to recover it is demonstrated. Precisely-controlled access region can be fabricated by the isotropic RIE. The shortest access region is 15 nm.

The III-V recess module, the drawback of selective wet etch in conventional III-V gate recess is discussed first. It results in long undercut and the underlap device architecture. The low carrier concentration in the access region is responsible for the low transconductance and \( g_m \) compression. To address this problem, a wet-etch free III-V recess process is developed. This is a two-step process that comprises of III-V dry etch and digital etch. It is shown that III-V dry etch can achieve smooth etched ending surfaces and a trenching-free profile. Then by using a digital etch technique, the semiconductor can be thinned down in the step of 1 nm. The special calibration method developed in this work enables the precise control of the etch depth.

In the gate stack module, four generations of gate stack development are carried out, with EOT scaled from 3.5 nm to 0.5 nm. In the long channel regime, the MOSFET shows extremely low subthreshold swing of 69 mV/dec. Extremely low leakage current are achieved even for the deeply scaled dielectric thickness.

In the back-end process module, the drawback of wet-etch via is first discussed. A two-step dry-etch via opening process is developed. The etch stops on the ohmic metal and results in damage-free surface. Self-aligned via filling is realized by using double-layer resist and a lift-off process.

The device fabrication technology forms the basis for the rest of this thesis. Ch. 3 focuses on the analysis and performance benchmarking of the InGaAs MOSFETs fabricated here. The tight-pitch MOSFET arrays in Ch. 4 and III-V integration with silicon in Ch. 5 are built upon the device technology developed in this chapter. The precise control of the access region leads to the record-high-transconductance InGaAs MOSFETs, from which the excess OFF-state drain leakage is observed and studied for the first time. This is to be discussed in Ch. 6.
Chapter 3: InGaAs MOSFET Performance Analysis

3.1 Introduction

Ch. 2 describes the techniques developed to precisely control the dimension of the access region of InGaAs QW-MOSFETs. MOSFETs with two types of access regions are fabricated and analyzed in this chapter. This study highlights the importance of access region design.

Using a novel CMOS-compatible process scheme that affords precise layer thickness control down to 1 nm, self-aligned InGaAs QW-MOSFETs with channel thickness between 3 nm and 12 nm, and gate length between 40 nm and 5 \( \mu \)m has been fabricated. It is observed that channel thickness has strong impact on device electrical characteristics. It affects the ON-state and OFF-state characteristics in a different direction, which suggests the trade-offs in device design. With the device technology innovations described in Ch. 2, the InGaAs MOSFETs in this work have achieved unprecedented transistor performance in their logic Figure-of-Merits, FOMs. This will be shown in the Section on device benchmarking. Lastly, the microwave characteristics of the InGaAs MOSFETs fabricated here will also be discussed.

3.2 Impact of Access Region Design

The parameters that characterize the access region are its thickness, \( t_{\text{access}} \), and length, \( L_{\text{access}} \), as illustrated in Fig. 3-1 (a). The thickness of the access region determines its sheet carrier density, \( N_s \).
The access region has a large impact on the device characteristics. To study this, devices are fabricated with identical intrinsic region but two different sets of dimensions of the access region (Fig. 3-1 c,d) which will be identified as short- and long-access devices. The channel thickness in the intrinsic region is $t_c=8$ nm. The detailed channel layer structures are shown in Fig. 3-1 (b). The gate stack has 2.5-nm HfO$_2$ directly laid on the InGaAs channel. For the short-access device (Fig. 3-1c) $L_{access}=15$ nm and $t_{access}=20$ nm; for the long access device (Fig. 3-1d), $L_{access}=80$ nm and $t_{access}=10$ nm.

### 3.2.1 ON-state characteristics

The transconductance and output characteristics of the two devices with the same gate length, $L_g=70$ nm, are shown in Fig. 3-2 and Fig. 3-3. At $V_{ds}=0.5$ V, the long-access device
delivers a peak transconductance $g_{m,\text{max}}$ of 1.9 mS/µm. At higher gate overdrive, $g_m$ drops rapidly after the peak. The current that corresponds to the peak $g_m$ is 0.3 mA/µm. $R_{on}$ of 340 Ω.µm is obtained from the output characteristics. The “$g_m$ compression” in the long-access device is similar to what has been observed in wet-etch-cap MOSFETs in Sec. 2.4.1. As discussed in Sec. 2.4.1, this can be the indication of source starvation.

For the short-access device, its $g_{m,\text{max}}$ is 2.7 mS/µm at $V_{ds}=0.5$ V as shown in Fig. 3-2. This was the record transconductance for all III-V MOSFETs when this device was made in 2013. At $V_{ds}$ of 0.65 V, $g_{m,\text{max}}$ reached 2.9 mS/µm, while further increase of $V_{ds}$ above 0.65 V did not increase $g_{m,\text{max}}$ any further. In contrast to the long-access device, the $g_m$ of short-access device stays at high value even at current level of 0.8 mA/µm. This device only shows slight “$g_m$ compression” at current level above 0.8 mA/µm. This can also be observed in the output characteristics of Fig. 3-3 (b). The short-access devices exhibit extremely low $R_{on}$ of 206 Ω.µm.

At low current level, the access region behaves as a resistor. The extrinsic source and drain resistance, $R_{sd}$, can be obtained by plotting $R_{on}$ vs $L_g$ and interpolate $R_{on}$ to $L_g=0$. $R_{on}$ is measured at high gate overdrive, $V_{gs}-V_t=0.7$ V, and low drain bias, $V_{ds}=0.05$ V. The extracted $R_{sd}$ of the long-access device is 302 Ω.µm, and it reduces to 206 Ω.µm for the short-access device. In comparison with the cap-wet-etched device in which $R_{sd}=450$ Ω.µm, these values suggest that the new gate recess method proposed in this work can effectively reduce the source and drain series resistance. $R_{sd}=206$ Ω.µm represents one of the lowest values ever demonstrated on III-V FETs of any kind. Due to device variation, $R_{on}$ in some short-channel “hero devices” can be slightly smaller than $R_{sd}$. While the higher $R_{on}$ can be explained by the higher $R_{sd}$ of the long-access device, the higher $g_m$ compression by itself cannot. It is however possible that both stem from reduction of active carrier dose in the access region. These results indicate the importance of the access region design for the purpose of high drive current and high transconductance. Device engineering to maintain high carrier concentration in the access regions and prevents source starvation is the key to fully realize high ON-state performance for InGaAs MOSFETs.
Fig. 3-2: Transconductance characteristics of $L_g=70$ nm MOSFETs with long-access region and short-access region at $V_{ds}=0.5$ V.

Fig. 3-3: Output characteristics of $L_g=70$ nm MOSFETs with (a) long-access region and (b) short-access region.
3.2.2 OFF-state characteristics

The subthreshold characteristics of the long- and sort-access devices are shown in Fig. 3-5. The long-access device with gate length of 70 nm features excellent short-channel effects with a minimum subthreshold swing $S=90 \text{ mV/dec}$ at $V_{ds}=50 \text{ mV}$ and $94 \text{ mV/dec}$ at $0.5 \text{ V}$. This is one of the lowest $S$ demonstrated to date at this gate length among all III-V MOSFETs. The drain-induced barrier lowering, DIBL, is $130 \text{ mV/V}$. The short-access device leads to increased short-channel effects for comparable gate length. The minimum subthreshold swing of this device is $S=108 \text{ mV/dec}$ at $V_{ds}=50 \text{ mV}$ and $138 \text{ mV/dec}$ at $0.5 \text{ V}$. DIBL is $249 \text{ mV/V}$. In addition to the increased SCE, it is also observed that the device cannot be completely turned off to $100 \text{ nA/µm}$ at $V_{ds}=0.5 \text{ V}$. This excess OFF-state drain leakage is attributed to the band-to-band tunneling and a parasitic bipolar gain effect. These phenomena are very sensitive to the access region design, and the impact is even more significant for deeply-scaled devices. A detailed study will be presented in Ch. 6. This work has revealed the trade-off between high drive current and SCE control that emerges from the access regions of an InGaAs MOSFET.
3.3 Impact of Channel Thickness Design

In the quest for balanced ON-state and OFF-state performance, there has been increasing concern about the superior intrinsic transport properties of III-V semiconductors being degraded by aggressive structural design directed to mitigate short-channel effects, SCE. Split-C-V measurements on long-channel devices have shown that ultra-thin channels suffer from significant mobility degradation [86]. This can be the result of increased scattering and a heavier carrier effective mass from enhanced quantization and band non-parabolicity. These mechanisms should also affect short-channel transistors with gate lengths comparable with the mean-free-path, or close to the ballistic limit. However, to date, there have been few experimental studies on high-performance InGaAs MOSFETs of different channel thickness with gate lengths that span from the drift-diffusion regime to the ballistic transport regime, and address both carrier transport and SCE control. The device technology demonstrated in Ch. 2 enables the tight control of the channel thickness through a precision recess method. At the same time, the gate length of
the devices can be scaled down to 20 nm. This device technology then represents an ideal platform to study the tradeoffs involved in channel thickness design as the gate length is scaled.

For this analysis, devices with channel thickness between 3 nm and 12 nm, and gate length between 40 nm and 5 µm are fabricated. Accurate channel thickness is the heart in this study. To realize this, during the digital etch, a procedure is used to pinpoint the precise location of the top surface of the InP etch-stop layer (Ch. 2). From that point on, the number of digital etch cycles needed to achieve a target channel thickness can be calculated. Here $t_c$ denotes the total channel thickness left after III-V recess, that is, the total semiconductor thickness above the InAlAs buffer in the intrinsic region of the device. The correlation between the number of digital etch cycles and the final channel thickness, and examples of buried channel and surface channel designs are shown in Fig. 3-6 (a) –(c). In this heterostructure, for $t_c$ between 11 – 13 nm, the high-κ dielectric interface is located on the InP barrier, resulting in a buried channel device (Fig. 3-6b). For $t_c$ of 10 nm or below, the high-κ dielectric lies directly on the quantum-well channel, resulting in a surface channel device (Fig. 3-6c). For $t_c$ below 5 nm, the high-mobility InAs core in the channel is entirely removed. The estimated $t_c$ involves 1 nm uncertainty due to the discrete nature of the digital etching technique.

![Fig. 3-6](image-url)

Fig. 3-6: (a) Correlation between number of digital etch cycles and final channel thickness. Examples of (b) buried channel and (c) surface channel designs. The dashed lines beneath the HfO$_2$ indicate the location of the oxide/semiconductor interface.
3.3.1 ON-state characteristics

Electrical characteristics of a short-channel transistor with the highest ON-state performance are shown in Fig. 3-7. This device has $t_c=9$ nm, which is surface-channel, and $L_g=80$ nm. In Fig. 3-7 (a), the transconductance characteristics at $V_{ds}=0.5$ V show a peak value, $g_{m,\text{max}}$ of $3.1$ mS/µm. This is the record value among all III-V FETs including HEMT and MOSFETs to date. The output characteristics in Fig. 3-7 (b) show an excellent ON-resistance of $190\ \Omega\cdot\mu\text{m}$.

![Fig. 3-7](image)

The intrinsic peak transconductance ($g_{m,i,\text{max}}$) is used as the metric to study the scaling behavior of ON-state performance. The intrinsic transconductance ($g_{m,i}$) is estimated without considering the effect of the output conductance, because in the worst case (highest output conductance) of the shortest gate length device of $L_g=40$ nm, this correction results in a relative small error (8%) in the estimation of $g_{m,i,\text{max}}$ [87]. Fig. 3-8 graphs $g_{m,i,\text{max}}$ vs. $L_g$ for families of devices with different $t_c$. It is found that $g_{m,i,\text{max}}$ follows a classic gate length scaling behavior: it increases as $L_g$ decreases and it eventually saturates. For $t_c>7$ nm, $g_{m,i,\text{max}}$ decreases as the gate length further decreases. This is likely to be the result of significant degradation of the electrostatic gate control to the channel in scaled gate length.
The measured source resistance $R_s$ is used for extracting $g_{mi}$. This is obtained from a measurement of the ON-resistance, $R_{on}$ which is performed at a high gate overdrive of 0.8 V. Fig. 3-9 (a) shows $R_{on}$ as a function of $L_g$ for $t_c$ from 12 nm to 3 nm. For all $t_c$ values, a linear dependence of $R_{on}$ with $L_g$ is obtained. The intercept of these lines with the y-axis gives the value of $R_{sd}$, the sum of the source and drain resistance. Fig. 3-9 (b) shows $R_{sd}$ as a function of $t_c$. For $t_c > 7$ nm the devices exhibit $R_{sd}$ of $\approx 250 \, \Omega \cdot \mu m$ more or less independent of $t_c$. This excellent result is due to the n$^+$-gate overlap design and good contact resistance (Ch. 2). For thin channels, $R_{sd}$ increases rapidly as $t_c$ reaches 4 nm and below. This is most likely due to the increased spreading resistance associated with the link region at the gate edge of the channel (Fig. 2-3a). Due to the symmetry of the device, it is assumed that $R_s$ is half of $R_{sd}$.

![Fig. 3-8: Peak intrinsic transconductance at $V_{ds}=0.5$ V vs. $L_g$ for different $t_c$.](image)

Fig. 3-8 reveals that carrier transport in the intrinsic device is greatly influenced by $t_c$. The highest $g_{mi,max}$ is obtained for $t_c=9$ nm which is the thickest surface-channel device that have been fabricated. For channels thicker than this, the devices are of buried-channel type and $g_{mi,max}$ decreases due to a loss in gate-channel capacitance (Fig. 3-10). For channels thinner than this, $g_{mi,max}$ decreases rapidly. This reveals worsening transport characteristics. It is interesting to note that $g_{mi,max}$ for the shortest devices is nearly identical for $t_c=7$ and 12 nm, indicating that the
advantage in transport properties balances the disadvantage in capacitance, while at the longer lengths the transport advantage outweighs the capacitance disadvantage and the thicker devices exhibit higher $g_{\text{mi,max}}$.

Fig. 3-9: (a) $R_{\text{on}}$ vs. $L_g$ for $t_c$ from 3 nm to 12 nm, in which $R_{\text{sd}}$ is extracted from the y-intersection. (d) $R_{\text{sd}}$ vs. $t_c$.

To further understand charge control and transport in these devices, split-C-V measurements on long-channel MOSFETs, $L_g=5 \, \mu$m, at 1 MHz (Fig. 3-10) have been carried out. While $C_g$ vs. $V_{gs}$ does not saturate, the capacitance range values representing the ON-state increase monotonically as the channel is thinned down. The ON-state capacitance increases monotonically as the channel is thinned down. Two factors contribute to this. First, as $t_c$ is reduced, the centroid of charge moves closer to the gate. Second, for thin enough $t_c$, the InAs core is eventually removed and the electron effective mass increases. This also yields an improved saturation behavior of the C-V characteristics for $t_c=3$ nm.

Independent evidence of the key role of transport in the performance of these devices is obtained from measurements of channel mobility. The effective field mobility, $\mu_{\text{eff}}$, is extracted by the split-C-V method in long-channel devices, $L_g=5$ um, and the results are shown in Fig. 3-11 as a function of sheet electron concentration, $N_s$. The $I_d$-$V_{gs}$ characteristics used in the
mobility extraction were corrected for series resistance. This correction is relatively small: in the worst case (highest $N_s$ at thick $t_c$), it is 14%. The impact of carrier trapping is not taken into account. It has been reported that a high concentration of border traps can exist in the high-κ oxide close to the III-V interface and cause significant trapping especially at high gate voltage [88], [89]. This could lead to error in the mobility extraction at high carrier concentration.

Fig. 3-10: Split-C-V measurement on QW-MOSFET with $L_g=5\ \mu$m for different $t_c$.

Fig. 3-11 also includes measurements of Hall mobility vs. sheet electron concentration from two epi calibration structures that were grown immediately before the actual device heterostructures. These test samples have the same channel and barrier design as the device heterostructure (Fig. 2-1) except a modified cap that consists of $n^+$ InAlAs (3 nm), $n^+$ InP (3 nm) and $n^+$ In$_{0.53}$Ga$_{0.47}$As (25 nm) with silicon doping between $2\times10^{19}$ and $3\times10^{19}$ cm$^{-3}$. Hall mobility measurements are performed as the $n^+$ InGaAs portion of the cap is thinned down through selective wet etching. This results in reduction of $N_s$ and increase in mobility as the relative
fraction of charge in the high-mobility channel increases. Starting with a heterostructure with a high Hall mobility has been found to be essential to obtaining high effective channel mobility of III-V MOSFET [33]. The results presented here confirm this observation.

Fig. 3-11 reveals a nearly progressive deterioration of channel mobility as the channel is thinned down. The highest peak mobility is obtained in the buried channel device with 1 nm InP barrier where $\mu_{\text{eff}}=8800 \text{ cm}^2/\text{V.s}$ at $N_s$ of $2.6 \times 10^{12} \text{ cm}^{-2}$. This is about 80% of the etch-Hall mobility at the same carrier concentration. With an InP barrier of 2 nm, the mobility is slightly lower. In surface channel devices, the mobility is lower than in buried channel designs and it monotonically drops as the channel becomes thinner. The mobility falls rapidly when $t_c$ is reduced below 5 nm.

Fig. 3-11: Mobility vs. sheet charge density. Continuous lines: effective channel mobility in long-channel devices ($L_g=5 \mu\text{m}$) with different channel thickness. Solid symbols: two sets of etch-Hall mobility from two growth test structures (see text).

In order to understand these results, 1D Poisson-Schrödinger (P-S) simulations (Nextnano$^3$ [90]) in the intrinsic region of the device are performed. Fig. 3-12 (a)-(c) show the carrier distribution, $n$, and the conduction-band edge energy, $E_c$, in the direction perpendicular to the channel, $x$, at a sheet electron concentration of $N_s=3\times10^{12} \text{ cm}^{-2}$ for $t_c$ of 11, 7 and 3 nm. For a thick channel, the charge in the channel is broadly spread in space and away from the oxide-
semiconductor interface. In a very thin channel, the electrons are tightly confined against the interface.

To evaluate the role of the charge distribution on transport, the distance of the charge centroid in the channel from the oxide-semiconductor interface, $d_{AV}$, is calculated. This is graphically indicated in Fig. 3-12 (a) and is computed using the following expression:

$$d_{AV} = \frac{1}{N_s} \int_0^{t_c} n(x) \cdot x \, dx$$  \hspace{1cm} (3-1)

In this equation, $n(x)$ is the local carrier density at a depth $x$ from the oxide-semiconductor interface where $x=0$.

Fig. 3-12 (d) graphs $d_{AV}$ as a function of $t_c$ at $N_s$ of $1 \times 10^{10}$ cm$^{-2}$ and $3 \times 10^{12}$ cm$^{-2}$ which are the scenarios close to threshold and ON-state. It is clear that as the channel thickness increases, the charge centroid continuously moves further away from the metal-oxide interface. The trend is more pronounced for buried channels. From these calculations it can be concluded that in thin channel designs, the electrons suffer more from surface roughness scattering as well as remote columbic scattering both of which would degrade the mobility.

Electrons in the InAs core possess the lowest effective mass and hence the higher mobility. The larger the electron fraction in the InAs subchannel, the higher the overall mobility should be. To quantitatively evaluate this, the weighted average of the inverse of the $\Gamma$-valley effective mass, $m_{F}$, of the channel electron, $Z$, in unit of $m_0^{-1}$, is extracted. This is obtained from P-S simulations using:

$$Z = \sum_i \frac{1}{m_i} \frac{1}{N_s} \int_{t_i}^{t_c} n(x) \, dx$$  \hspace{1cm} (3-2)

In this equation, the integrals evaluate the sheet electron concentration in the respective channel layer $i$, which can be InP, InAs or In$_{0.7}$Ga$_{0.3}$As. $m_i$ and $t_i$ are effective mass and thickness of respective channel layer. Effective mass of those materials follow the room-temperature values in [91].
Fig. 3-12: (a-c) Carrier distribution, n, and conduction-band energy, $E_c$, in the direction perpendicular to the channel. The semiconductor/oxide interface is placed at $x=0$. The energy of the first two quantum states is shown in dashed lines. (d) Average distance of channel electron population, $d_{AV}$, with respect to the oxide-semiconductor interface at $N_s$ of $1 \times 10^{10}$ cm$^{-2}$ (blue) and $3 \times 10^{12}$ cm$^{-2}$ (red). The schematic of $d_{AV}$ is shown in the inset in (a). (e) Experimental mobility vs. $Z$ at $N_s$ of $3 \times 10^{12}$ cm$^{-2}$ (see text).

Fig. 3-12 (e) graphs experimental channel mobility vs. $Z$, both at $N_s$ of $3 \times 10^{12}$ cm$^{-2}$. The figure reveals a clear correlation: increased $Z$ results in a higher mobility. However, there is a step increase in mobility for buried-channel designs. This is due to the alleviation of interface scattering experienced by electrons. Among surface channel devices, the mobility monotonically
increases with $Z$ which itself increases with $t_c$. This reflects the reduced effective mass and weaker alloyed scattering associated with a higher relative InAs composition in the channel. For buried channel devices, increased $t_c$ brings about increased penetration of the electron wave function into the InP barrier which results in a reduction of $Z$. This is the likely reason for the drop in mobility from the 11 nm to the 12 nm design.

The channel thickness dependence of the mobility contributes to explaining the strong impact that channel thickness has on transconductance (Fig. 3-8). A complete model for the transconductance requires the development of a charge-control model and transport model that take into account the effective mass under strain in each of the channel layers and includes conduction band non-parabolicity. This detailed study is best done on devices with a simpler channel design than those here.

### 3.3.2 OFF-state characteristics

Despite the advantages of using a thick channel from a transport point of view, the drawback is evident in poor SCE control. As shown in Fig. 3-13 (a), minimum subthreshold swing, $S_{\text{min}}$, of 159 mV/dec at $V_{ds}=0.5$ V and DIBL=310 mV/V are observed in the device that demonstrates record transconductance ($L_g=80$ nm and $t_c=9$ nm). In contrast, a device with of the same gate length but a thinner channel ($L_g=80$ nm and $t_c=4$ nm) exhibits superior subthreshold behavior as seen in Fig. 3-13 (b) where $S=111$ mV/dec at $V_{ds}=0.5$ V and DIBL=126 mV/V.

Threshold voltage roll-off control is a key issue for deeply scaled logic transistors. It is a consequence of short channel effects. Fig. 3-14 graphs $V_{t,\text{sat}}$ as a function of gate length where $V_{t,\text{sat}}$ is defined at $I_d=1\times10^{-6}$ A/µm and $V_{ds}=0.5$ V. It is clear that thicker $t_c$ results in more significant $V_t$ roll-off. At $L_g=100$ nm, $\Delta V_{t,\text{sat}}$ is -0.68 V for $t_c=12$ nm with respect to the long channel value.

Fig. 3-15 (a) and (b) depict $S_{\text{min}}$ at $V_{ds}=0.5$ V and DIBL vs. $L_g$ for devices with varying $t_c$. DIBL is obtained from $V_t$ shift between $V_{ds}=0.05$ V and 0.5 V at a constant current of $I_d=1\times10^{-6}$
A/µm. However, not all devices can be turned off to this current level; hence some data points at short gate lengths are not available. \( S_{\text{min}} \) and DIBL in Fig. 3-15 (a) and (b) follow classic scaling behavior; they are both independent of gate length for long \( L_g \) but degrade rapidly as \( L_g \) scales down beyond a certain value. The onset of \( S_{\text{min}} \) and DIBL degradation occurs at longer gate lengths for devices with larger \( t_c \).

Fig. 3-13: Subthreshold characteristics of two \( L_g = 80 \) nm MOSFET with (a) same device as that in Fig. 3-7 with \( t_c = 9 \) nm and (b) \( t_c = 4 \) nm.

Fig. 3-14: Threshold voltage roll-off behavior for InGaAs MOSFET for \( t_c \) from 3 nm to 12 nm.
These results can be analyzed using classic SCE electrostatics originally developed for fully-depleted SOI MOSFETs by Yan et al. [92]. Electrostatically, the QW-MOSFETs discussed here are similar to FDSOI transistors, where the InAlAs buffer is equivalent to the buried oxide. Based on Yan’s model, the “natural channel length” $\lambda$ is defined that depends on the geometry and the dielectric constant of the various materials involved:

$$\lambda = \sqrt{\frac{t_{\text{ins}}}{t_{\text{c}}} \frac{\varepsilon_{\text{c}}}{\varepsilon_{\text{ins}}}}$$  \hspace{1cm} (3-3)

In this equation, $\varepsilon_{\text{c}}$ and $\varepsilon_{\text{ins}}$ are the relative dielectric constants of the channel and the effective insulator respectively; $t_{\text{c}}$ and $t_{\text{ins}}$ are the thickness of the channel and the effective insulator respectively. Here it departs from Yan’s model: in this case, $\varepsilon_{\text{ins}}$ and $t_{\text{ins}}$ take into
account the HfO$_2$ gate insulator, the interfacial oxide and $d_{av}$. For the interfacial oxide, it is assumed that it has a thickness of 1 nm (Fig. 2-3b) and a relative dielectric constant of 16. The composition and properties of III-V native oxides are not well known. For example, relative dielectric constants between 9.9 and 23.3 have been reported for Ga$_2$O$_3$ [93], [94]. Here, it assumed that the interfacial oxide has a dielectric constant given by the average of these two values. The average distance of channel electrons away from the oxide- semiconductor interface, $d_{av}$, arises from the quantum nature of the channel and the presence of the InP barrier in the buried channel designs. Here $d_{AV}$ at $N_s=1\times10^{10}$ cm$^{-2}$ is used as graphed in Fig. 3-12 (d).

The scaling parameter $\gamma$ that governs SCE is the ratio of the effective channel length and the natural channel length:

$$\gamma = \frac{L_{eff}}{\lambda} \quad (3-4)$$

$L_{eff}$ is related to gate length, but, as argued below is the two are not generally the same.

Fig. 3-15 (c) and (d) show the change of $S_{min}$ and DIBL vs. $\gamma$. The change of $S_{min}$ and DIBL is with respect to the long channel values. To study short channel effects, the long-channel behavior which is caused by other reasons such as floating body effects is removed. Then the data all fall nearly on top of each other suggesting a universal electrostatic scaling trend.

MOSFET designs with better SCE immunity start suffering from SCE degradation at smaller $\gamma$ values. In Si FDSOI MOSFETs, it is reported that well-designed device architectures maintain acceptable SCE down to $\gamma$ of 5 to 10 [92], [95]. Reports on InGaAs HEMTs indicate SCE immunity down to $\gamma=5$. In the fabricated devices, it is found that SCE FOMs start to degrade at $L_g/\lambda \sim 15$. This apparently early SCE degradation of the fabricated InGaAs MOSFETs might arise from two factors. First, here and in the analysis of III-V HEMTs [96], it is assumed that $L_{eff}=L_g$; but HEMTs have an underlapped SD, and therefore the physical gate length $L_g$ is an underestimation of the effective channel length $L_{eff}$, and hence $\gamma$ (Eq. 3-4), while the MOSFETs in this work, as with modern VLSI transistors, are designed with a slight n$^+$-gate overlap and therefore $L_g$ is close to or a slight overestimation of the effective channel length.
Second, the introduction of an n-type \( \delta \)-doped layer in the buffer below the channel is likely to degrade the subthreshold characteristics. This is equivalent to the introduction of positive fixed charge below the channel. In floating-body transistor designs, it has been observed that positive fixed charge beneath the channel results in degraded subthreshold characteristics in short III-V MOSFETs \([27]\), \([97]\), \([98]\) and FDSOI transistors \([99]\). This suggests that appropriate transistor redesign might mitigate SCE in future InGaAs QW-MOSFETs.

This study reveals that planar InGaAs QW-MOSFETs, as fabricated here, even with the thinnest channels, \( L_g \sim 50 \) nm are at the limit of scaling. Similar to silicon MOSFETs, the use of 3D device architectures can increase \( \lambda \) and allow the device to scale to even smaller \( L_g \). In any case, this will require the use of a very thin semiconductor body. This suggests that the trade-off between OFF- and ON-state performance will remain an important engineering challenge \([92]\), \([97]\).

### 3.4 Performance Benchmarking

Recently, tremendous progress has been made in the research of III-V MOSFETs and it is important to assess the potential of this technology class by benchmarking it with proper metrics. This section develops such a benchmarking methodology based on the key figures of merit (FOM) for logic applications. The minimum subthreshold swing, \( S_{\text{min}} \), is regarded as the representative FOM for SCE control, and maximum transconductance, \( g_{m,\text{max}} \), for ON-state performance. Both are measured at drain bias of 0.5 V. The ON-current, defined at \( I_{\text{off}}=100 \) nA/\( \mu \)m and \( V_{dd}=0.5 \) V, is used to gauge the balance between SCE control and ON-state performance of the III-V FETs.

Fig. 3-16 shows the benchmarking of \( S_{\text{min}} \), measured at \( V_{ds} \) of 0.5 V, for III-V FETs reported in recent literature, including HEMT, planar MOSFET and 3D MOSFET (FinFET, trigate and nanowire). The QW-MOSFETs fabricated in this work are graphed using solid triangles. Three types of device, “MIT MOS WE”, “MIT MOS LA” and “MIT MOS SA”, are included. The labels are denoted in the following manner: “MIT MOS WE” for wet-etch-cap
MOSFETs discussed in Sec. 2.4, “MIT MOS LA” for long-access MOSFETs and “MIT MOS SA” for short-access MOSFETs discussed in Sec. 3.1. At 2013 year end, the long-access MOSFETs fabricated in this work represent the lowest subthreshold swing of all III-V planar MOSFETs, indicating superior SCE control. Only a Trigate architecture displayed better SCE scalability [100]. At 2014 year end, new regrown-SD planar MOSFETs demonstrated improved SCE [81], [101].

Fig. 3-17 shows the benchmarking of $g_{m,\text{max}}$ vs. $S_{\text{min}}$. Both device parameters are measured at $V_{\text{ds}}=0.5$ V and only device with $L_g<80$ nm are selected. The long-access MOSFETs have one of the smallest subthreshold swings in this scaled dimension. For the transport FOM, the short-access MOSFETs have demonstrated the highest transconductance. Two versions of short-access device are shown. “MIT MOS SA” is for short-access MOSFETs discussed in Sec. 3.1. These were fabricated in 2013. “MIT MOS SA V2” is for short-access MOSFETs discussed in Sec. 3.2 with channel thickness optimization. Both of these two devices represented the highest transconductance values at the time when they were fabricated. However, the subthreshold swing for these two devices is less satisfactory. The subthreshold characteristic is degraded because of the enhanced band-to-band tunneling and floating body bipolar gain effects in scaled short-access transistors, which will be explained in Ch. 6 in details.

![Fig. 3-16: Benchmarking of $S_{\text{min}}$ at $V_{\text{ds}}=0.5$ V vs. $L_g$ for III-V FETs published at (a) 2013 year-end and (b) 2014 year-end. Ref: [20], [21], [23], [30], [33], [34], [81], [82], [100]–[110].]
Fig. 3-17: Benchmarking of $g_{m,max}$ vs. $S_{min}$ at 0.5 V for III-V FETs with $L_g \leq 80$ nm showing record device obtained in the device obtained in this thesis. Ref: [20], [23], [30], [81], [82], [100], [101], [104], [109], [110].

Fig. 3-18: Benchmarking of $I_{on}$ vs $L_g$ at $V_{dd}=0.5$ V and $I_{off}=100$ nA/µm for III-V FETs. Ref: [24], [29], [30], [76], [80], [82], [101], [103], [105], [109]–[114].
Fig. 3-18 shows the benchmarking of the ON-current, $I_{\text{on}}$, of III-V MOSFETs. $I_{\text{on}}$ is defined at $I_{\text{off}}=100$ nA/µm and $V_{dd}=0.5$ V. This is a stringent criterion. If a device cannot be turned off to 100 nA/µm ($I_{\text{off}}>100$ nA/µm), or ON-current does not reach 0.2 mA/µm ($I_{\text{on}}<0.2$ mA/µm), they will not be placed in this graph. This FOM tests the device for a balance of transport, charge control, scalability, low interface traps and low extrinsic series resistance. Fig. 3-18 summarizes the $I_{\text{on}}$ vs. $L_g$ from reported InGaAs MOSFETs. The lines show the record values of each year. It is clear that the performance of InGaAs MOSFETs is consistently improving over years. The wet-etch-cap MOSFET and long-access MOSFETs has represented the record $I_{\text{on}}$ for any type of III-V MOSFET in year 2012 and 2013. Only very recently, a new type of III-V MOSFETs reported in [81], [101] has reached a higher value.

3.5 Microwave Characteristics

Microwave/RF performance of the short-access MOSFETs was characterized using an HP 8510C network analyzer. It covers a frequency range from 0.5 to 40 GHz. A device with a T-shape microwave layout is shown in Fig. 3-19. The gate width is 100 µm, 50 µm x2. The transconductance from the RF measurement approximately agrees with that obtained from DC measurement except that there is a slight increase and stretching of $g_m$ for RF measurement with respect to the DC measurement. The as-measured extrinsic peak $g_m$ for this device is 2.3 mS/µm from DC measurement, and 2.5 mS/µm from RF measurement (Fig. 3-20a). This can be attributed to the positive gate stress applied to the device. To sample across the required frequency range in RF measurement, the time for which the device were exposed to any gate voltage is much longer than that during DC measurement. This equivalent to positive gate stress which has been observed to increase $g_m$ and shift $V_t$ positively. Another possible reason for the higher $g_m$ under RF measurement is that the effect of $D_{it}$ is diminished at high frequency.
The measured current gain, $H_{21}$, and Mason’s unilateral gain, $U_g$, are shown in Fig. 3-20b. The device is biased at $V_{ds}=0.5$ V and $V_{gs}=0.4$ V. On-chip open and short de-embedding is used to subtract both pad capacitance and inductance components from $H_{21}$ and $U_g$. The open de-embedding only reaches the pad region. The parasitic gate capacitance associated with the overlapping portion of the gate is not de-embedded. The as-measured value of the microwave FOMs are: $f_T=169$ GHz and $f_{max}=110$ GHz. After de-embedding, $f_T=236$ GHz and $f_{max}=123$ GHz. The de-embedded characteristics show a slope of -20 dB/dec decay. As mentioned in Ch. 2, the current MOSFET structures focus on I-V device characteristics, and have not been optimized for overlapping capacitance. Owing to the record high transconductance, there should still be large room for improvement on the microwave FOMs.

Bias dependence of the current gain is investigated. $H_{21}$ is measured at different $V_{gs}$ and $V_{ds}$ as shown in Fig. 3-21 (a). At fixed $V_{gs}$ of 0.4 V, $f_T$ continues increasing up to $V_{ds}$ of 0.5 V. In another measurement on the same device, $V_{gs}$ is treated as variable (Fig. 3-21b). It is found that when $V_{gs}$ is above $V_t$, $f_T$ increase approximately linearly until it saturates at $V_{gs}=0.4$ V. $f_T$ roughly follows the $g_m$ characteristics as shown in the inset of Fig. 3-21 (b).
Fig. 3-20: (a) Transconductance from DC measurements and RF measurements in the same short-access InGaAs MOSFET. (b) Microwave characteristics of $L_g=70$ nm InGaAs MOSFET. $H_{21}$ and $U_g$ vs. frequency for device biased at peak $g_m$ point.

Fig. 3-21: At fixed $V_{gs}$ of 0.4 V and different $V_{ds}$ for short-access InGaAs MOSFET with $L_g=70$ nm (a) $H_{21}$ vs. frequency; (b) $f_T$ vs. $V_{gs}$. At fixed $V_{ds}$ of 0.5 V and different $V_{gs}$ for the same device (c) $H_{21}$ vs. frequency; (d) $f_T$ vs. $V_{gs}$. Inset of (d) shows the correlation between $f_T$ and $g_m$. 
3.6 Chapter Summary

In this chapter, a detailed vertical and lateral scaling study of InGaAs QW-MOSFETs has been carried out. It is found that the access region has strong impact on the device performance. When the carrier density in the access region is low, the device can suffer from source starvation and “$g_m$ compression”. By increasing carrier density using a short-access design, record peak transconductance is achieved. However, this is at the cost of degraded OFF-state characteristics.

Using the novel gate recess method that affords precise channel thickness control down to 1 nm, InGaAs MOSFETs with channel thickness between 3 nm and 12 nm, and gate length between 40 nm and 5 $\mu$m were fabricated. It is found that the intrinsic transconductance follows a classic gate length scaling behavior. Surface channel designs with a relatively thick channel provide the best transconductance characteristics. Buried-channel designs feature lower transconductance presumably due to a loss of gate capacitance. Thin channels suffer from severe transconductance degradation. From a study of channel mobility it was identified that the InAs composition that electrons experience in the channel is a key factor affecting transport characteristics. In contrast with their excellent ON-state performance, the OFF-state characteristics of thick channel devices are significantly compromised. It was found that the subthreshold swing and DIBL follow classic scaling behavior similar to that of Si FDSOI MOSFETs. The current InGaAs QW-MOSFET is at the limit of scaling around $L_g=50$ nm. This indicates the need for transistor redesign and advanced 3D device architectures to deliver further progress.

The InGaAs MOSFETs in this work have achieved unprecedented transistor performance in terms of logic applications. The long-access devices matched the subthreshold swing and ON-current records for 2013, and the short-access devices achieved new transconductance records for 2013 and 2014. The highest transconductance of 3.1 mS/$\mu$m achieved in this work has exceeded that of HEMTs, for the first time in history.
The high-frequency characteristics of the QW-MOSFETs are explored. The high-transconductance InGaAs QW-MOSFET shows outstanding microwave characteristics with $f_t=236$ GHz.
Chapter 4: Tight-Pitch MOSFET Arrays and Nanoscale Contacts

4.1 Introduction

Modification of the process technology described in Ch. 2 yields a novel InGaAs MOSFET array structure with unprecedented pitch size and contact size for any kind of InGaAs FETs to date. Two different arrays are demonstrated: a gated MOSFET array and a gate-less array. From their analysis the contact resistance of nanoscale contacts in InGaAs MOSFETs can be estimated. This chapter shows the characterization and modeling of these new types of device structures. The study confirms the great promise of the Mo contact scheme, though it suggests that more work is needed to further reduce the total access resistance to deliver the required performance.

4.2 Array Design and Fabrication

Two types of arrays is used to analyze all resistive components of nanoscale self-aligned W/Mo ohmic contacts to InGaAs QW-MOSFETs. These two arrays and their fabrication procedure are sketched in Fig. 4-1. The device structure and fabrication procedure for these arrays is similar to that of self-aligned InGaAs MOSFETs described in Ch. 2. Only a broad outline of the process with the unique details to the fabrication of these test structures is given here.
The heterostructure described in Fig. 2-1 of Ch. 2 is used for this fabrication. The top of the cap layers is In$_{0.7}$Ga$_{0.3}$As doped with silicon to $3 \times 10^{19}$ cm$^{-3}$. In the gate patterning process step, resist Zep-520A of thickness of about 120 nm is used for E-beam lithography. It must ensure that the resist is thick enough to withstand all the subsequent RIE steps, but not too thick to compromise the minimum attainable line spacing (Fig. 4-1a). This is because the high-aspect-ratio resist pattern can be easily deformed in the subsequent process.

After E-beam lithography, SiO$_2$, W and Mo are then etched sequentially. After eliminating the E-beam resist, part of the sample is clipped off for use as “gate-less array” (Fig. 4-1b). The rest of the sample continues with the gate fabrication. The III-V recess process described in Ch. 2 follows (Fig. 4-1c). Then 2.5-nm ALD HfO$_2$ is deposited as gate insulator, followed by Mo gate electrode patterning. The finished gate structure is shown in Fig. 4-1d. Lastly, the via opening and pad deposition are carried out using the self-aligned wet-etch-free process discussed in Sec. 2.6. Vias are defined only on the long contacts at two ends of the arrays and is done for both types of arrays. The schematic of the finished MOSFET array is shown in Fig. 4-3 (e), while for the gate-less array, it is shown in Fig. 4-1 (f).

![Fig. 4-1](image)

**Fig. 4-1**: Schematic illustration of process flow for array fabrication: (a) E-beam lithography, (b) Ohmic metal etching and gate-less array, (c) III-V recess, and (d) MOSFET gate structure. The finished devices with pads are shown for (e) MOSFET gate array and (f) gate-less array.

Arrays with different gate length, contact length and number of cells can be fabricated in this process. The structure referred to as a cell is illustrated in Fig. 4-1. As shown in Fig. 4-1, one
cell contains one contact, one gate length and two access regions. The respective lengths of those regions are given by \( L_c \), \( L_g \), and \( L_{\text{access}} \). The space occupied by one full cell is its pitch size, \( L_p = L_c + L_g + 2 \cdot L_{\text{access}} \). A zero-cell device is then a normal MOSFET with long contacts. This work has fabricated array structures with 1 to 4 cells. In this experiment MOSFET array and gate-less array with contact lengths \( L_c \) vary from 40 nm to 800 nm are fabricated. For all devices, the length of the access region is \( L_{\text{access}} = 15 \) nm. The gate length of the intrinsic gate region \( L_g \) is kept at constant, \( L_g = 100 \) nm, for all devices.

![Schematic illustration of MOSFET array with (a) 1 cell and (b) 2 cells.](image)

Fig. 4-2: Schematic illustration of MOSFET array with (a) 1 cell and (b) 2 cells.

![Top view of a MOSFET array.](image)

Fig. 4-3: (a) Top view of a MOSFET array. (b) Enlarged top view that shows MOSFET array with 3 cells.
Fig. 4-4: Cross sectional TEM of MOSFET array with different contact length, gate length and pitch size: (a) $L_c=40$ nm, $L_g=130$ nm and $L_p=200$ nm; (b) $L_c=80$ nm, $L_g=40$ nm and $L_p=150$ nm.

Fig. 4-5: Cross sectional SEM of gate-less array with $L_c=130$ nm and $L_p=200$ nm. Inset shows the enlarged image and layer materials.
The SEM top views in Fig. 4-3 show the layout of a typical tight-pitch MOSFET gate array with 3 cells. It has a typical MOSFET layout structure except that in the gate region four gates are separated by three inner contacts. All gates are connected to the same gate pad which is biased at $V_g$. All inner contacts are floating. Two outer contacts are connected to source and drain pads which are biased at $V_s$ and $V_d$ respectively. Fig. 4-4 (a) shows a cross sectional TEM of a 1-cell MOSFET array with $L_p=200$ nm. The contact length and gate length are $L_c=40$ nm and $L_g=130$ nm. Fig. 4-4 (b) shows a 2-cell MOSFET array. It has a different combination of contact length and gate length, $L_c=80$ nm and $L_g=40$ nm respectively, that leads to a pitch size of $L_p=150$ nm. These are the smallest contact length and pitch size of any InGaAs FETs to date. In addition, Fig. 4-5 shows a cross sectional SEM of a gate-less array with $L_c=130$ nm and $L_p=200$ nm.

This device structure in this work has some unique advantages. Firstly, it can be easily incorporated into the same chip dedicated to typical MOSFET fabrication and does not add significant complexity to the existing process flow. The test structure directly reflects the level of self-alignment and pitch size that a front-end process can deliver. Secondly, it deals with a complex multi-layer contact configuration, as will be shown in a contact structure that comprises three layers and two contact interfaces in the following section. Lastly, the multiple-cell array structure is a simple method to achieve accurate measurement, which will be discussed later.

### 4.3 Result and Discussion

This section presents the characterization and analysis of the gate-less array and MOSFET array. Fig. 4-6 (a) shows $I_d-V_{gs}$ and $g_m-V_{gs}$ characteristic for a MOSFET array with $L_g=100$ nm, $L_c=500$ nm and 3 cells biased in the linear regime with $V_{ds}=50$ mV. For a MOSFET array, if the gate is biased at a high gate overdrive, the $I_d-V_{gs}$ characteristics plateau, and the transconductance $g_m$ approaches zero. The total resistance is extracted as $R_{total}=V_{ds}/I_d$. Since the
intrinsic channels have identical gate length, the resistance associated with each one of these intrinsic channel is the same.

As shown in Fig. 4-6 (b), the resistance added by the insertion of one cell, $R_{\text{cell}}$, can be extracted from the slope of $R_{\text{total}}$ vs. number of cells, $N_{\text{cell}}$. In this manner, the parasitic resistance associated with the peripherals such as the large contact and probe pad can be eliminated. To measure current through a nanoscale structure, the resistance associated with the peripherals usually plays a significant role. Accurate current-voltage measurement sometimes requires advanced test methods, such as the Kelvin measurement [115] and specially designed TLM devices [51]. Such experiments may demand dedicated structures with air bridge that is difficult to make and model. To enhance the accuracy of the extraction, in this work, serially-connected replicas of identical “cells” is designed. The underlying principle is that these structures extract the incremental resistance component associated with the change of number of cells in the array. So the peripheral components, such as the long contact and the probe pad, do not play a role.

![Fig. 4-6: (a) Id-Vgs and gm-Vgs characteristics for a MOSFET arrays with Lg=100 nm, Lc=500 nm and 3 cells. (b) Extraction of Rcell from the slope of total resistance vs. number of cells in the gate-less arrays and MOSFET arrays.](image)

The extracted $R_{\text{cell}}$ for varying contact lengths is shown in Fig. 1-1 (a) and (b) for the gate-less arrays and MOSFET arrays, respectively. In the gate-less array, $R'_{\text{cell}}$ is about 50 $\Omega \cdot \mu$m.
in the long contact range and this reduces to about 30 Ω·μm when Lc=40 nm. A rapid reduction occurs for contact lengths below 200 nm. In the MOSFET array, a similar trend is observed, but the magnitude of the change in R_{cell} is larger.

Equivalent circuit models are developed to analyze the above R_{cell} results. They are shown in Fig. 4-8 (a) and (b) for one cell in the gated array and gate-less array, respectively. In the MOSFET array under the bias condition selected here, the intrinsic channel region behaves as a resistor. In addition to this, the resistance associated in the access region of the MOSFET is also constant under a low current condition which is the case here. The sum of these resistances is denoted as R_o in Fig. 4-8 (a). The resistance seen between A and B in Fig. 4-8 (a), R_{AB}, depends on the contact length, L_c, of the cell. A and B are the projections of the contact edges to the channel. Variations of L_c reflect the contact scaling behavior. This is similar in the gate-less array (Fig. 4-8b). The access resistance component in the region where the ohmic contact is etched constitutes a constant offset R_o’. The resistance in the contact region is now between C and D, R_{CD}, where C and D are the projections of the contact edges to the cap. In Fig. 1-1, L_c=0 extrapolation corresponds to R_o and R_o’. When they are substracted from R_{cell} and R_{cell}’ respectively, the remaining portions are the resistances R_{AB} and R_{CD}.

![Fig. 4-7: (a) R’_{cell} vs. L_c in gate-less arrays where ρ_{12} is extracted. (b) R_{cell} vs. L_c in MOSFET arrays where ρ_{23} is extracted. Solid symbol is from experiment and line is from model.](image)
Of interest is the contact region, whose cross sectional schematic is shown in Fig. 4-9 (a). It contains three stacking layers: ohmic metal characterized with sheet resistance $R_{sh,m}$, n$^+$ cap with $R_{sh,n}$, and channel with $R_{sh,ch}$. The length of the contact is $L_c$. This multilayer system can be modeled by a resistor network composed of three coupled conducting layers as shown in Fig. 4-9 (b). The contact length $L_c$ is discretized into segments of length $dx$. The model uses $dx$ of 10 nm. This value of $dx$ is selected so that it is small enough to model the smallest contact ($L_c=40$ nm). Then the values of lateral resistors in Fig. 4-9 (b) are $R_{sh}·dx$ and vertical resistors are $ρ_{xy}/dx$, all in unit of $Ω·μm$. $ρ_{xy}$ characterizes the interfacial resistivity between two conducting layers. $ρ_{12}$ is the contact resistivity between the Mo to the n$^+$ cap, and $ρ_{23}$ is that between the n$^+$ cap and the channel. This resistive netlist is generated by Matlab and solved by Hspice.

In this model, the sheet resistance components, $R_{sh}$, of the individual conducting layers are inputs that are obtained from independent measurement. The contact resistivities, $ρ_{xy}$, are the extracted parameters after optimization. The sheet resistance of the channel, $R_{sh,ch}$, is obtained from transistor measurements. For instance, the slope of $R_{on}$ vs $L_g$ in Fig. 3-9 (a) of Ch. 3 reflects $R_{sh,ch}$, where $R_{on}$ is measured with the gate biased at gate overdrive that results in the transistor $g_m$ approaching zero. The sheet resistance of the n$^+$ cap resistance, $R_{sh,n}$, and the metal, $R_{sh,m}$, are obtained from TLM measurements on separate test structures. $R_{sh,n}$ is confirmed with Hall measurement. Their values are specified in Fig. 4-9 (b).

The model of the contact is essentially the same as for the MOSFET array and the gateless array, but the boundary conditions are different. In the MOSFET array, the current enters from terminal A and exits from terminal B as seen in Fig. 4-8 (a). In the gate-less array, the n$^+$ cap is not removed. Here $R_{sh,n}$ of the n$^+$ cap is significantly lower than that of the channel $R_{sh,ch}$, so the current is mostly carried by the cap layer, from terminal C to terminal D (Fig. 4-8b). In the Hspice program, the same netlist for the contact region is generated, but the boundary conditions are set differently for these two types of array. When the two boundary nodes, A to E,
are selected, the rest of the boundary nodes are left open, same as calculating the Thévenin equivalent resistance between the selected nodes.

Fig. 4-8: Cross-section schematic of one cell for (a) MOSFET array and the circuit model for $R_{cell}$, (b) gate-less array and the circuit model $R'_{cell}$.

Fig. 4-9: (a) Schematic of the contact region that contains three parallel conducting layers with the specified sheet resistance. (b) The circuit model for the contact region. The same model is used for the MOSFET array and gate-less array with the appropriate boundary conditions. (see text).
The experimental data in Fig. 1-1 are analyzed using these models. Good agreement between the predictions of the models and the $R_{\text{cell}}$ and $R_{\text{cell}'}$ measurements of Fig. 1-1 can be obtained as indicated by the continuous lines in those graphs. The contact resistivity between Mo and n$^+$ cap that emerges is $\rho_{12} = (8 \pm 2) \times 10^{-9} \Omega \cdot \text{cm}^2$. This is consistent with the value obtained from nanoTLM structure [51]. Comparison of this value against the latest technology development in InGaAs contact is shown in the Contact Module of this thesis (Ch. 2 Sec. 2.2). This work and a few other recent experiments have suggested that Mo is capable of delivering lower contact resistivity compared to other contact metals, such as Au and Ni-InGaAs alloys. This highlights the promise of non-alloy Mo contact for n$^+$-InGaAs. Note that the Landauer limit of any metal contact to n$^+$ InGaAs is about $1 \times 10^{-9} \Omega \cdot \text{cm}^2$ [116]. There is still room for reduction of contact resistivity between Mo and n$^+$ cap.

The contact resistivity between n$^+$ cap and channel is also extracted: $\rho_{23} = (2 \pm 0.8) \times 10^{-8} \Omega \cdot \text{cm}^2$. This value is rather high and indicates that further optimization is required in designing a heterostructure to reduce this portion of the contact resistance.

Fig. 4-10: Modeled contact resistance for the technology in this work based on the extracted $\rho_{12}$ and $\rho_{23}$. This is the estimated resistance between nodes A and E in Fig. 4-9.
Using the extracted values of $\rho_{12}$ and $\rho_{23}$, the MOSFET contact resistance for a given contact length can be estimated by simply extracting the Thévenin equivalent resistance between A and E in Fig. 4-9. The result is shown in Fig. 4-10. $R_c$ shoots up as $L_c$ becomes smaller than the transfer length at about 110 nm. At $L_c$ of 10 nm, the contact resistance is $260 \ \Omega \cdot \mu m$. This does not include the resistance associated with the access regions. As the ITRS goal is $60 \ \Omega \cdot \mu m$ at 10 nm [4], it is imperative to investigate further engineering method to deliver the required performance.

**4.4 Chapter Summary**

For the first time, working front-end tight-pitch InGaAs MOSFET arrays by a self-aligned process have been fabricated and MOSFET arrays achieve 40-nm contact size and 150-nm pitch size has been demonstrated. Arrays with different numbers of cells and contact lengths allow the elimination of parasitic components and the resistance of elemental contacts in the nanoscale can be accurately extracted. An Hspice model is developed to analyze the tight-pitch MOSFET array and gate-less array that takes into account two contact interfaces. Contact resistivities of $8 \times 10^{-9} \ \Omega \cdot \text{cm}^2$ between Mo and $n^+$ cap and $2 \times 10^{-9} \ \Omega \cdot \text{cm}^2$ between $n^+$ cap and channel have been extracted. Model infers a contact resistance from the Mo contact to the channel of $260 \ \Omega \cdot \mu m$ for $L_c=10$ nm. This suggests that more work is required in the InGaAs MOSFET contacts to deliver the required performance.
Chapter 5: III-V Integration on Silicon Substrate

5.1 Introduction

In addition to a CMOS-compatible process for InGaAs MOSFET fabrication, it is of critical importance to develop an effective strategy to integrate those devices onto silicon. Several approaches have been proposed to realize the integration of III-V devices on a silicon substrate. III-V HEMT and MOSFET have also been demonstrated on silicon substrates through direct MBE growth [109], [117], [118]. In this case, high-quality III-V layers are realized by using thick metamorphic buffer. The Aspect-Ratio-Trapping technique was used to grow III-V Fin structures on 200 mm and 300 mm silicon wafer and FinFETs were then fabricated on such structures [119], [120]. Another promising integration scheme is through wafer bonding. Through this technique, III-V-On-Insulator (III-V-O-I) substrate can be fabricated. Various III-V MOSFET structures have been demonstrated using such substrate as platform such as InGaAs MOSFETs with regrown SD [50] and with Schottky SD [121].

In this chapter, a new technique that combines wafer bonding and self-aligned recessed-gate technology is demonstrated. This work is done in collaboration with IBM Research, Zurich, where the substrate fabrication takes place. The III-V-O-I wafer is fabricated through a wafer bonding technique developed in IBM [32]. Then device fabrication is performed in the Microsystems Technology Laboratories at MIT. Ultra-thin-body (UTB) devices built on a SOI-like platform can incorporate a buried gate (back-gate) is formed by the Si substrate which enables modulation of the transistor threshold by application of voltage to it. This dynamic threshold voltage control can be the key for energy-efficient high performance computing and
low power applications. In silicon, this has been demonstrated some time ago as the so-called Silicon-On-Insulator with Active Substrate (SOIAS) technology [122].

This chapter describes the fabrication sequence of thin-body InGaAs QW-MOSFETs on insulator with a p-type Silicon substrate. Self-aligned InGaAs MOSFETs with a tight contact pitch of 150 nm is demonstrated and characterized. It is observed that the back-gate bias with respect to the source (V_{bs}) has a strong influence on various device performances. The threshold voltage (V_t), parasitic series resistance (R_{sd}), drain-induced barrier lowering (DIBL), and subthreshold swing (S), all have strong dependency on V_{bs}.

5.2 Fabrication Procedures

The integration of InGaAs QW-MOSFETs into a III-V-O-I substrate consists of two main steps: substrate fabrication and device fabrication. Substrate fabrication was done at IBM and it follows the method described in [32]. The III-V heterostructure is first grown by MBE on a semi-insulating InP donor wafer. The heterostructure is essentially grown upside down from how it appears in Fig. 5-1 (a). In order of growth it comprises a highly doped n^+ capping layer, an InP etch stopper, a strained In_{0.7}Ga_{0.3}As channel (without InAs core), a thin InAlAs buffer layer, and a 1-nm InGaAs protection layer. This top 1-nm InGaAs protection layer will later serve as interface to the buried oxide, BOX. Little influence is expected on device operation from the presence of this thin layer of InGaAs. A silicon δ-doped sheet with N_s=1×10^{12} cm^{-2} is located 5 nm away from the channel in the buffer layer. In the final device configuration, the thin InAlAs layer will serve as a buffer between the channel and BOX.

Substrate fabrication continues with hydrogen implanted into the heterostructure. The hydrogen peak is located in the InP substrate below the device heterostructure. This is shown by the dash-dot line in Fig. 5-1 (b). After this, a 30 nm-thick Al_2O_3 BOX layer is deposited by ALD on the III-V donor wafer. The wafer is flipped over and bonded to a p-type Si wafer with p-type doping concentration of 1×10^{17} cm^{-3}. Fig. 5-1 (b) illustrates the cross section schematic of this
step. A thermal process is then performed to split the donor wafer from the device layers as shown in Fig. 5-1 (c). A final InP selective etch back is carried out, leaving the device heterostructure with the n⁺ InGaAs contact exposed at the top (Fig. 5-1d). If during MBE certain buffer layers are added between the device layer and the InP donor wafer, the donor wafer can be recycled [9].

Fig. 5-2 summarizes the structural data, including XRD and AFM, acquired at the end of UTB III-V-O-I substrate fabrication. High structural quality is preserved and the smooth surface observed in the AFM measurement confirms the excellent surface quality after bonding.

Fig. 5-1: Fabrication procedure for InGaAs MOSFETs on III-V-O-I (see text).

Fig. 5-2: (a) ω-2θ X-ray diffractogram (XRD) acquired on the finished III-V-O-I substrate and aligned on the InP (004) peak. (b) AFM scan of the surface of the finished III-V-O-I substrate with height scale on the right.
Device fabrication follows the self-aligned tight-pitch process described in Ch. 2. A cross sectional schematic of the InGaAs MOSFETs on III-V-OISAS studied here is shown in Fig. 5-1 (e). A TEM cross-section of a fabricated device is shown in Fig. 5-3. This device has a gate length of 50 nm. The access region length is 30 nm. The gate dielectric is 3.5 nm HfO₂ placed on top of an 8-nm thick In₀.₇Ga₀.₃As channel. The Al₂O₃ BOX is 30-nm thick. The TEM image shows a device in a tight-pitch gate array. The pitch size is 150 nm. One gate pitch comprises the gate region (50 nm), two access regions (30 nm x 2) and contact region (40 nm). Device measurements are performed on devices with large contacts.
5.3 Electrical Characteristics

Fig. 5-4: Transconductance characteristics for an InGaAs MOSFET with gate length of 70 nm at $V_{bs} = -2, 0$ and $3$ V. In all figures, $V_{ds}=0.5$ V.

Fig. 5-5: Output characteristics for the same InGaAs MOSFET as in Fig. 5-4 at $V_{bs} = -2, 0$ and $3$ V.

Fig. 5-6: Subthreshold characteristics for the same InGaAs MOSFET as in Fig. 5-4 at $V_{bs} = -2, 0$ and $3$ V. In all figures, $V_{ds}=0.05$ and $0.5$ V.
Fig. 5-7: (a) Minimum subthreshold swing at $V_{ds}= 0.5$ V and (b) DIBL vs. $V_{bs}$ for the InGaAs MOSFET of Fig. 5-4.

Fig. 5-8: (a) Threshold voltage modulation by $V_{bs}$ for a $L_g=1$ µm device, and (b) source and drain series resistance under different $V_{bs}$ from a plot of $R_{on}$ vs $L_g$ that interpolates to $L_g=0$.

The transconductance $g_m-V_{gs}$ and output $I_d-V_{ds}$ characteristics of a transistor with $L_g=70$ nm are shown in Fig. 5-4 and Fig. 5-5. $V_{bs}$ is biased at -2, 0 and 3 V. The back-gate bias with respect to the source is denoted as $V_{bs}$. At $V_{bs}=3$ V, the device delivers the highest ON-state performance. The highest transconductance is 400 mS/µm. Drive current under the same gate overdrive condition is the highest at $V_{bs}= 3$ V. Output characteristics at the given $V_{bs}$ conditions
show a noticeable increase in output conductance when $V_{ds} > 0.4$ V. This phenomenon is more severe for negative $V_{bs}$.

The subthreshold $I_d-V_{gs}$ characteristics for $V_{ds}=0.05$ and 0.5 V are shown in Fig. 5-6. $V_{bs}$ modulates $V_t$ of the device: a positive $V_t$ shift results from the application of a negative $V_{bs}$ and *vice versa*. The subthreshold swing improves as $V_{bs}$ becomes more negative. The combination of $V_t$ shift and S sharpening at negative $V_{bs}$ yields an extremely low OFF-state drain leakage. At $V_{bs} = -2$ V and $V_{ds}=0.5$ V, an OFF-state current of 0.26 nA/µm is measured at $V_{gs}=-0.4$ V.

At $V_{gsd}=-0.4$ V, the front gate leakage of this device is low: $I_g=0.1$ pA/µm. The leakage current through the BOX is extremely low, a value below the measurement limit. Hence, at negative $V_{bs}$, the back-gate results in a positive shift in $V_t$ and a sharpening in the subthreshold swing at no cost in static leakage. This offers the opportunity to efficiently reduce static power consumption in VLSI circuits, making them attractive candidates for low power applications.

Fig. 5-7 (a) shows the result of a systematic study of the impact of $V_{bs}$ on the minimum subthreshold swing ($S_{min}$) of the 70 nm device of Fig. 5-4. The drain bias is 0.5 V. At $V_{bs} = -2$ V, $S_{min}$ of 128 mV/dec is obtained. The subthreshold swing increases with $V_{bs}$. DIBL is measured at a constant current $1 \times 10^{-6}$ A/µm and between $V_{ds}$ of 0.05 and 0.5 V. As shown in Fig. 5-7 (b), DIBL exhibits a different trend with respect to subthreshold swing. It decreases from a value of 120 mV/V at negative $V_{bs}$ to about 70 mV/V at positive $V_{bs}$.

There could be multiple factor for the impact of $V_{bs}$ on S and DIBL. The movement of inversion charge centroid and the growth of space charge region in the body can couple with short-channel effects. It is believed at this point that S is more strongly correlated to the inversion charge centroid in the quantum well channel, while DIBL is more strongly correlated to the appearance and growth of a space charge region in the p-type silicon substrate. More work is needed to fully explore this difference in behavior.

Fig. 5-8 (a) shows saturated $V_t$ (at $V_{ds}=0.5$ V) of a long-channel MOSFET ($L_g=1$ µm) against $V_{bs}$. For a $V_{bs}$ that spans between -3 V and 4 V, $V_t$ is modulated by 220 mV. The resistance $R_{sd}$, as extracted from the gate length interpolation method at various $V_{bs}$, is shown in
Fig. 5-8 (b). $R_{sd}$ increases by 2x when $V_{bs}$ changes from 4 V to -4 V. This explains the loss of performance noted above as $V_{gs}$ is made more negative.

### 5.4 Device Benchmarking

Fig. 5-9 (a) and (b) show $S$ and DIBL as a function of $L_g$ benchmarked against recently published MOSFETs on III-V-O-I. $S$ and DIBL are extracted at $V_{bs}=0$ V. For all papers, the values for $S$ are extracted at $V_{ds}=0.5$ V. Device in this work matches the lowest $S$ of 83 mV/dec in the long channel regime. The obtained values of DIBL match the lowest reported values in the short-channel regime ($L_g<100$ nm). In the long channel regime, DIBL is slightly higher than Ref. [123]. The UTB III-V-O-I devices show outstanding short-channel effect control.

![Fig. 5-9: (a) $S_{min}$ vs. $L_g$ and (b) DIBL vs. $L_g$ for this work and recently published III-V MOSFETs on insulator. Ref: [104], [123]–[125].](image-url)

### 5.5 Chapter Summary

Thin-body InGaAs MOSFETs are fabricated on a III-V-On-Insulator with a Silicon active substrate. The p-type active substrate acts as a back gate bias that can modulate the threshold voltage and other electrical characteristics of the device. The impact of $V_{bs}$ on the device characteristics is explored. This includes the behavior of subthreshold swing, DIBL,
threshold voltage, and series resistance. The subthreshold swing and OFF-state leakage current improve as $V_{bs}$ becomes more negative, while the series resistance and ON-state performance improve as $V_{bs}$ becomes more positive. Understanding this should provide device physics insights that can be utilized for more effective device design as well as providing ideas for use at circuit level.
Chapter 6: Physics and Mitigation of Excess OFF-state Current

6.1 \( I_{\text{off}} \) in High-Transconductance III-V MOSFETs

6.1.1 Introduction

In recently demonstrated scaled InGaAs QW-MOSFETs with tight SD spacing around the gate, excessive drain to source leakage current has been observed that prevents the effective turning-off of the transistor [27], [28], [30], [80]. Excess OFF-state current, \( I_{\text{off}} \), is defined as the drain leakage current observed at high \( V_{\text{ds}} \) and \( V_{\text{gs}} \) significantly below \( V_t \). In the rest of this chapter, this bias condition will be referred to as the “OFF-state”. Fig. 6-1 gives two examples. Fig. 6-1 (a) is reported in [30]. This device has an In\(_{0.53}\)Ga\(_{0.47}\)As quantum-well channel and its gate length is 55 nm. It delivers a transconductance of 1.9 mS/\( \mu \)m at \( V_{\text{ds}} \) of 0.5 V, the record value in year 2011. Another example in Fig. 6-1 (b) is the short-access MOSFET discussed in Section 3.1. It delivers a transconductance of 2.7 mS/\( \mu \)m at \( V_{\text{ds}} \) of 0.5 V, the record value in year 2013. As highlighted in Fig. 6-1, \( I_{\text{off}} \), in these two high-transconductance scaled InGaAs MOSFETs cannot be turned off by reducing gate voltage.

The excess OFF-state current observed in Fig. 6-1 is a serious concern that must be addressed before the logic potential of this material system can be realized. As a reference, the ITRS requires \( I_{\text{off}} \) of 10 pA/\( \mu \)m for low-standby-power applications [4]. In this regards, to understand the underlining physics that contribute to the excess \( I_{\text{off}} \) is of great importance. There can be a number of coupled factors that influence \( I_{\text{off}} \). A systematic study of this is yet to be
performed. The MOSFETs fabricated in this work provide an excellent platform to study the physics and mitigation strategies of $I_{\text{off}}$ in InGaAs channel transistors.

![Graphs showing high-transconductance InGaAs MOSFETs with excess drain leakage in the OFF-state at high $V_{ds}$.](image_url)

Fig. 6-1: Examples of high-transconductance InGaAs MOSFETs that exhibit excess drain leakage in the OFF-state at high $V_{ds}$. (a) Regrown-SD InGaAs MOSFET with $L_g=55$ nm [30]. (b) Short-access device in this work with $L_g=70$ nm.

This chapter is organized as follows. First, the impact of short channel effects, buffer leakage, gate leakage and Fermi level pinning are systematically excluded as the origin of the excess OFF-state current. Then, the physics and the experimental evidence behind band-to-band tunneling (BTBT) and bipolar amplification in QW-MOSFETs are presented separately. Following this, a comprehensive model is developed to understand the coupling of these two mechanisms. Simulation results of the OFF-state current that illuminate BTBT-induced source-channel barrier lowering, the bipolar gain amplification and its $L_g$ dependence are then presented. And finally, device design approaches to mitigate the excess OFF-state current problem are explored. In particular, the role is studied of: (1) doping of the channel region; (2) InAs composition of channel and cap; (3) carrier lifetime; and (4) geometrical design of the access regions.
6.1.2 A strategic approach

In Fig. 3-5 of Sec. 3.2.2, the subthreshold characteristics of the long-access and short-access devices are compared. Even though those devices have identical intrinsic region and gate length, the short-access device shows higher excess $I_{\text{off}}$ and cannot be turned off further at negative $V_{gs}$. This indicates that the access region has strong impact on excess $I_{\text{off}}$.

First, devices with long gate lengths were studied so that the role of short-channel effects can be identified. Fig. 6-2 shows the comparison of subthreshold characteristics for MOSFET A with $L_g=500$ nm and short-access $L_{\text{access}}=15$ nm, and MOSFET B with $L_g=200$ nm and long-access $L_{\text{access}}=80$ nm. MOSFET A is on the same chips as the short-access device (Fig. 3-5b) and MOSFET B is on the same chips as the long-access device (Fig. 3-5a). These two chips are fabricated at the same time and on the same heterostructure. The intrinsic channel structures of these two sets of device are identical as described in Section 3.1. The threshold voltages of these two devices are shifted and matched. Useful information can be extracted from this comparison.

These two devices practically have the same subthreshold characteristics except at the “OFF-state” when $V_{gs}$ is significantly lower than $V_t$. Excess $I_{\text{off}}$ appears only for MOSFET A in the form of a relatively flat current level of $\sim 2\times 10^{-8}$ A/µm (line). MOSFET B shows continuous turning-off to a current level below $2\times 10^{-10}$ A/µm (solid dots). Since they have identical intrinsic region (except for the gate length), it is reasonable to argue that the flat $I_{\text{off}}$ in MOSFET A is not due to high interface states or Fermi level pinning at an energy close to the valence band edge.

Since MOSFET B has a shorter gate length, if there is any significant short channel effects, it would have degraded device B more. But this is not the case in Fig. 6-2. Hence, it can be concluded that “short channel effects” are not the cause of the difference in these two devices in the OFF-state. In this discussion, “short channel effects” refer to the classic two-dimensional potential distribution effect such as charge sharing. After the mechanisms at play in long-channel devices are identified, I will come back and discuss how those mechanisms would couple with the channel length and result in a new “short channel effect”.

113
Since these two devices are fabricated on the same heterostructure as the same time, Fig. 6-2 also proves that $I_{\text{off}}$ in the short-access device does not result from buffer leakage. Buffer leakage refers to a leakage path beneath the channel in the buffer layers that is far away from the gate and cannot be turned off by the gate. Buffer leakage was commonly seen in epilayers contaminated by dopants. In fact, the buffer leakage for the layer structure below the channel is much lower than the drain current. It is approximately in the order of $10^{-11}$ A/µm.

Fig. 6-2 also indicates that the gate leakage current is at least two orders of magnitude lower than the drain current level. Its impact on the drain current can be neglected and virtually all the drain current observed here flows directly from drain to source.

![Fig. 6-2: Comparison of subthreshold characteristics in long channel MOSFETs with A: short-access and $L_g=500$ nm and B: long-access and $L_g=200$ nm. The devices were fabricated at the same time on the same sample.](image)

### 6.2 Evidence of Gate-induced Drain Leakage (GIDL)

#### 6.2.1 Physics of GIDL

Transistors in low bandgap materials, such as Ge, usually suffer serious influence from Gate-induced drain leakage (GIDL) [126]. Although the bandgap of InGaAs can be smaller than Ge for high InAs compositions, there had not been any reports of GIDL in InGaAs transistors
before 2013. This can be due to the following reasons. $I_{\text{off}}$ of InGaAs HEMTs, the mainstream InGaAs transistor in the past, is usually set by the leakage of the Schottky gate. Besides, the relaxed requirement on device footprint allows relatively large side recess and relatively low electric fields in those devices [20], [22], [23], [127]. Only recently, much progress has been made in InGaAs-based III-V MOSFETs for logic applications. A great deal of emphasis on those devices has been given to gate length scaling, reduced source and drain series resistances ($R_{\text{sd}}$) and self-aligned designs transistor [27], [28], [30], [80]. In addition, suppression of Fermi-level pinning at the III-V/oxide interface and the use of high-permittivity gate insulator allows the device to be turned off sharply [26], [82]. All these engineering advances have yielded great progress in III-V MOSFET performance. However, all these enhancements have also contributed to an increase in the electric field at the drain edge of the channel. This has allowed GIDL-like characteristics to emerge. For the first time, this work provides supporting evident for the appearance of GIDL in high-performance InGaAs MOSFETs.

![Fig. 6-3: (a) Schematic illustration of BTBT in an InGaAs MOSFET. The device is biased in the OFF-state with $V_{\text{gs}}<V_{\text{t}}$ and $V_{\text{ds}}=V_{\text{dd}}$. (b) Corresponding energy band diagram along the channel from source to drain. The star indicates the high field region where BTBT occurs. The arrows indicate the BTBT-generated hole flux ($F_h$) and electron flux ($F_e$) that constitute the BTBT current. (c) Classic GIDL characteristics at high $V_{\text{ds}}$ [128].](image-url)
GIDL takes place in the high electric field region at the drain-end of the channel and is a manifestation of a BTBT process. The classic GIDL concept is illustrated in Fig. 6-3 (a), which shows a schematic cross section of a simplified InGaAs QW-MOSFET biased in the OFF-state. The corresponding band diagram is shown in Fig. 6-3 (b). The high E-field region around the drain-end of the channel where BTBT takes place in the OFF state is denoted with a star. As sketched by the arrows in Fig. 6-3 (a) and (b), valence electrons from the channel tunnel across this region and are collected by the drain. Overlaid are arrows that indicate electron and hole flow which constitutes the BTBT current. Fig. 6-3 (c) illustrates the classic GIDL behavior in the subthreshold characteristics of a MOSFET. The drain current goes up as Vgs goes down, because the E-field at the gate edge near the drain is increase. In Fig. 6-3 (b), the band bending at the location denoted by a star becomes steeper, and hence BTBT becomes more prominent.

Analytical models have been developed for the GIDL in silicon MOSFETs. Usually such a model comprises two steps. In the first step, it captures the magnitude of E-field in the high field region at the drain. Under some simplified conditions, the E-field (E) is proportional to the gate-to-drain voltage [129]: $E \propto |V_{ds} - V_{ds,\text{sat}}|$. This expression of E has taken threshold voltage into account, where $V_{ds,\text{sat}} = V_{gs} - V_t$. Then $E \propto |V_{dg} + V_t|$. This expression is valid when the n-type channel does not have significant hole accumulation near the drain edge. Otherwise, the accumulated holes would screen the potential from the gate and prevent the E-field from penetrating into the semiconductor. In the second step, BTBT current is calculated using the Keldysh and Kane’s BTBT rate equation [130]–[132] that has a form of:

$$I_{\text{BTBT}} = A \times \exp\left(- B \frac{E_g^3}{|E|}\right)$$

$I_{\text{BTBT}}$ denotes the magnitude of the BTBT current ($I_{\text{BTBT}} = |I_{e,\text{BTBT}}| = |I_{h,\text{BTBT}}|$), where $I_{e,\text{BTBT}}$ is the BTBT electron current (blue arrow in Fig. 6-3 a and b), and $I_{h,\text{BTBT}}$ is the BTBT hole current (red arrow in Fig. 6-3 a and b). Due to charge conservation, the magnitude of $I_{e,\text{BTBT}}$ and $I_{h,\text{BTBT}}$ must be equal. $E_g$ is the bandgap of the channel. $|E|$ is magnitude of the E-field. A and B are two parameters that characterize BTBT for a given material.
6.2.2 Experiment

To further rule out the impact of interface states on the OFF-state device characteristics, devices were characterized at cryogenic temperatures where the interface states would be frozen out. Evidence can be seen in Fig. 6-4 that plots $S$ vs. $I_d$ at $V_{ds}=50$ mV. The theoretical limit of the subthreshold swing, $S_{\text{ideal}}$, is imposed by thermal limit, $kT/q$. The ideal subthreshold swing is 60 mV/dec at room temperature 300 K, 30 mV/dec at 150 K and 15 mV/dec at 77 K. The ideality factor $n$ is defined as the ratio between the minimum measured subthreshold swing, $S_{\text{min}}$, to the ideal subthreshold swing, $n=S_{\text{min}}/S_{\text{ideal}}$. In a thin-body QW-MOSFET with a long channel, the measured subthreshold swing becomes higher than the ideal value because of the presence of interface trap density ($D_{it}$) in the gap of the MOS interface. This is observed at 300 K, and $n=1.26$. At both 77 K and 150 K, subthreshold swings approach ideal values ($n = 1.06$) across many orders of magnitude of $I_d$. This indicates that at temperatures below 150 K, interface states become essentially frozen. The subthreshold characteristics for $V_{ds}$ between 0.3 and 0.6 V are shown in Fig. 6-5. A characteristic GIDL signature in $I_{off}$ is evident at 150 K and 77 K.

Fig. 6-4: Subthreshold swings of a $L_g=500$ nm short-access MOSFET vs. $I_d$ at different temperatures. The ideality factor is defined as $n=S_{\text{min}}/S_{\text{ideal}}$. Almost ideal subthreshold swing below $T=150$ K indicates that interface states are frozen.

117
Fig. 6-5: Subthreshold characteristics of a $L_g=500$ nm short-access MOSFET at (a) 300 K, (b) 150 K, and (c) 77 K.

Fig. 6-6: Low-temperature dependence of $I_d$ vs. $E_g^{3/2}$ in the OFF-state. The straight line that is obtained suggests a tunneling process across the In$_{0.7}$Ga$_{0.3}$As bandgap and supports the BTBT hypothesis.

The GIDL current depends on temperature (bandgap) and bias (E-field) as discussed in Section 6.2.1. The measured data is examined against the expected GIDL dependence. Fig. 6-6 graphs the temperature dependence with the drain current at a given $V_{ds}$ and $V_{gs}$ being plotted against $E_g^{3/2}$. $E_g$ is calculated based on the bandgap model for In$_x$Ga$_{1-x}$As under different
temperatures, where $x=0.7$ [91]. The straight line is consistent with the temperature dependence of the bandgap and supports the BTBT hypothesis.

To examine the bias (E-field) dependence, Fig. 6-7 (a) graphs the subthreshold $I_d$ vs. $V_{gs}$ characteristics at 77 K of the $L_g=500$ nm short-access device. The threshold voltage is $V_t=0$ V. Fig. 6-7 (b) graphs the same subthreshold current characteristics against $[V_{dg}+V_t]^{-1}$ in a semilog scale. The straight line indicates the regime that follows the model prediction. It is observed that within a limited range of operation at low $V_{dg}$, the current follows the expected relationship. When $V_{dg}$ is low, the BTBT current is negligible compared to the drift-diffusion current. So $I_d$ is above the reference line on the right side of the plot. On the other hand, $I_d$ falls below the reference line on the left because holes start to accumulate in the channel. In this case, the relationship $E \propto (V_{dg}+V_t)$ starts to break down. In addition, this simple model assumes that the coupled bipolar gain is not changed for all $V_{gs}$. But in fact the bipolar gain is also changed with gate bias. This will become clear later.

For the first time, a detailed experimental analysis is shown that supports the hypothesis that BTBT is responsible for the excess OFF-state leakage at high drain bias in InGaAs MOSFETs. Since BTBT happens only in the high field region, to suppress GIDL, a solution is to reduce the E-field. The use of long access with reduced carrier concentration can effectively
reduce the field magnitude and $I_{\text{off}}$, as is evident in Fig. 6-2. However, the trade-off of using a long access region is the ON-state performance degradation and increased device footprint. GIDL mitigation in InGaAs has attracted a lot of research attention recently. Other proposed solutions for BTBT suppression include undoped vertical spacers [80] and asymmetric SD design [133].

6.3 Evidence of Bipolar Gain Effect

6.3.1 Physics of bipolar gain effect

In the last section, evidence was shown that excessive $I_{\text{off}}$ is triggered by BTBT in InGaAs MOSFETs with tight source and drain to gate distance. Based on the physics of BTBT, solutions have been proposed to address the GIDL problem. In spite of this, so far no consideration has been given to the quantum-well nature of the channel and the potential for a coupled floating-body effect. A number of anomalies have been traced to this combination of mechanisms such as the kink effect for analog device applications [134], jitter [135] and premature breakdown [136], among others [134]. In InGaAs High-Electron Mobility Transistors (HEMTs), floating body effects associated with hole generation through impact ionization have also been widely reported [137]–[139]. These have been found to result in the kink effect, excess gate leakage, excess and frequency-dependent output conductance, and premature breakdown and burnout [140].

In floating-body Silicon-on-Insulator (SOI) MOSFETs, it is well-known that the presence of a lateral bipolar junction transistor (BJT) with substantial current gain yields a dramatic enhancement of BTBT or impact-ionization generated current [141], [142]. The current gain of the lateral BJT depends strongly on channel length. In this section, it is shown that a phenomenon of this nature also takes place in quantum-well InGaAs MOSFETs and that this is
responsible for the large OFF-state current experimentally observed. This effect is called the bipolar gain effect, and it strongly increases the BTBT current as illustrated in Fig. 6-8.

As discussed in the previous section, valence electrons from the channel tunnel across the high field region and are collected by the drain. Unlike in the bulk MOSFET where BTBT hole current can flow toward the body terminal, the holes that are left behind accumulate in the floating-body channel. This leads to reduction of the source-channel energy barrier which under steady-state conditions allows the BTBT-generated holes to be injected into the source and recombine in the heavily-doped cap or at the source-contact interface.

Fig. 6-8: Schematic illustration of BTBT and bipolar gain action in an InGaAs QW-MOSFET biased in the OFF-state. (b) Corresponding energy band diagram along the channel from source to drain. The source-channel barrier height ($q\phi_b$) determines the OFF-state drain current.

An important consequence of the reduction of the source-channel energy barrier is electron injection from the source into the channel. These electrons diffuse through the channel and contribute an additional drain current component. This is indicated in Fig. 6-8 by the thick blue arrow. In essence, this is a bipolar gain effect similar to what has been observed in floating-body SOI MOSFETs [129], [143]. In a nanoscale device, the gain of the parasitic bipolar
transistor can be quite large and, as a result, the total drain current can be many times higher than the BTBT electron-hole generation current.

6.3.2 Experiment

The coupled bipolar gain effect can be experimentally verified by investigating the gate length dependency of $I_{\text{off}}$. A pure BTBT nature of the OFF-state current would imply absence of gate length dependence. That is not what is experimentally observed. Fig. 6-9 (a) shows experimental $I_{\text{d}}$-$V_{\text{gs}}$ characteristics of transistors with $L_g$ from 80 nm to 500 nm at 200 K and at $V_{ds}=0.7$ V. For clarity, the voltage axis is shifted by the $V_t$ of the respective transistor. Low temperature data is used to freeze-out the interface states and reduce their influence on the subthreshold behavior of the transistor. $I_{\text{off}}$ is observed to increase as $L_g$ is reduced. Fig. 6-9 (b) plots the inverse of $I_d$ vs. $L_g$ for $V_{ds}=0.7$ V and three $V_{gs}$-$V_t$ values. The strong inverse dependence of OFF-state $I_d$ with $L_g$ cannot be explained by a pure BTBT-induced current.

A 2D device model is developed to understand the $L_g$ dependence of the OFF-state current. The 2D device simulator Sentaurus Device by Synopsys [144] is used in these simulations. The coupled Poisson, electron and hole continuity equations are solved self-consistently with a non-local-path BTBT model. Quantum effects are not included in these simulations due to convergence difficulties. It is expected some minor corrections when quantum effects are accounted for as a result of the presence of a quantum “dark region” that pushes the charge centroid away from the front and back interfaces of the channel. The device structure closely follows the short-access MOSFETs, except that in the extrinsic source and drain region, the 3-nm InP etch-stop is substituted by In$_{0.53}$Ga$_{0.47}$As for improved numerical convergence. Recombination is assumed to take place only at the S/D metal contacts which are characterized by an infinite surface recombination velocity. The contacts are long enough for their length not to be relevant. Recombination in the body of the channel and the cap is neglected because the carrier diffusion lengths in these two regions are estimated to be much longer than the gate length and the cap thickness, respectively. This estimation is based on experimental carrier
lifetimes of In$_{0.53}$Ga$_{0.47}$As/InP structures [145] and reasonable mobilities [146], [147], and yields a diffusion length that is over 10 µm in the channel and ~0.2 µm in the n$^+$ cap. Nevertheless the effect of carrier recombination in the body of the semiconductor is explored in a following section. The value of $E_g$ for the In$_x$Ga$_{1-x}$As channel is selected following Vegard’s law based on the bulk values of InAs (0.35 eV) and GaAs (1.42 eV). The parameters that characterize the BTBT process were adjusted to provide order of magnitude agreement with the experimental results. For the purpose of obtaining high-level physical understanding, this was deemed adequate.

Fig. 6-10 (a) shows the experimental subthreshold $I_d$-$V_{gs}$ characteristics of a $L_g$=500 nm short-access MOSFET at room temperature. The simulation result for this device is shown in Fig. 6-10 (b). The rising tail of $I_d$ for Fig. 6-10 (b) with decreasing $V_{gs}$ is observed only when the BTBT model is included; otherwise $I_d$ continues to decrease with $V_{gs}$ following the classic subthreshold behavior. In the simulations that include BTBT, the minimum current level in the OFF-state is close to the experimentally observed values. This strongly suggests that BTBT plays a role in the excess OFF-state current. Both subthreshold swing and the slope of drain leakage are steeper in the simulations than in the experiment. This is likely the result of not including interface states in the simulations, that are active in RT measurements. Interface states affect the slope of the drain current in the off regime as the gate control over the surface potential is degraded. No attempt was made to match measured and simulated $V_t$ because it is assumed that work function tuning of the gate metal would readily induce the necessary parallel shift along the $V_{gs}$ axis.

Fig. 6-11 (a) shows simulated contours of BTBT generation rate, $G$, at the edge of the gate near the drain contact for a $L_g$=500 nm device at $V_{ds}$=0.6 V and $V_{gs}$= -0.5 V. The selected bias corresponds to the point labeled with an “O” mark on the y-axis in Fig. 6-10 (b). Since the maximum E-field occurs at the top surface of the channel near the drain edge, BTBT takes place around this high-field region: electron generation prevails in the circled right side area of Fig. 6-11 (a) while hole generation mainly occurs in the channel surface near the gate edge (left
The BTBT generation rate depends strongly on E-field, or, in essence, \( V_{dg} \). It exhibits negligible dependence on gate length. The magnitude of the BTBT current per unit width, \( I_{BTBT} \), can be calculated by integrating the generation rate of either electrons or holes. Here hole generation (left circle) is used to estimate the BTBT current as shown in Fig. 6-11 (a). It is approximately \( 9 \times 10^{-11} \text{ A/\mu m} \), about three orders of magnitude smaller than \( I_d \) calculated at this bias point (Fig. 6-10b) or observed experimentally (Fig. 6-10a).

**Fig. 6-9:** (a) Experimental subthreshold characteristics of InGaAs MOSFETs for gate lengths between 80 and 500 nm, measured at \( V_{ds}=0.7 \text{ V} \) and at 200 K. (b) \( I_d^{-1} \) vs. \( L_g \) at fixed \( V_{ds} \) and \( V_{gs}-V_t \) under the same measurement conditions of (a).

**Fig. 6-10:** (a) Experimental room temperature (RT) subthreshold and gate current characteristics of long-channel InGaAs MOSFETs. (b) Simulated subthreshold characteristics with and without BTBT model suggesting that BTBT plays a role in the excess off-state current.
I_{bt} = q \int_{A} G \cdot dA \\
\approx 9 \times 10^{-11} \text{A/\mu m}

I_h(x_0) = 9.3 \times 10^{-11} \text{A/\mu m}

I_e(x_0) = 7.3 \times 10^{-8} \text{A/\mu m}

Fig. 6-11: (a) BTBT contours for electrons (right circle) and holes (left circle) at the drain edge of the channel. The channel consists of In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As = 1/2/5 nm (Ch. 2). Electron generation takes place in two regions: around the n+/undoped InGaAs junction as a result of the strong band bending and in the InAs core because of its narrow bandgap. (b) Electron and hole current density corresponding to the same situation as in (a). The currents per unit gate width at point x_0 are indicated on the right.

To understand this discrepancy, the simulated current components are further explored by plotting the electron and hole current magnitude contours in the channel. These are shown in Fig. 6-11 (b). Generated holes flow towards the source on the left, while electrons flow towards the drain on the right. Along the x-direction under the gate, both carrier currents are largely location independent. It is observed that the electron current is much higher than the hole current. The total electron (hole) current per unit width, I_e (I_h), is defined as the integration of the electron (hole) current magnitude along the y-axis. At x=x_0 in Fig. 6-11 (b) I_h of 9.3 \times 10^{-11} \text{A/\mu m} is obtained, matching the BTBT current in Fig. 6-11 (a). Meanwhile, I_e of 7.3 \times 10^{-8} \text{A/\mu m} matches the I_d calculated at this bias (label “O” in Fig. 6-10b). There is then a current gain in this situation of \beta = I_e/I_h = 780.
The experiment and simulation suggest an additional bipolar gain mechanism as in SOI MOSFET is indeed in action. In spite of this, at the moment, there is a lack of detailed physical understanding of the interaction between these relevant mechanisms. The following section will investigate the coupling of BTBT and bipolar gain effect by 2D device simulations.

6.4 Coupled BTBT and Bipolar Gain Effect

6.4.1 Simulation setup

A cross sectional schematic of the simplified InGaAs QW-MOSFET modeled in this study is shown in Fig. 6-12. This is a self-aligned device with a raised source and drain architecture similar to devices in which excess drain current has been observed in the OFF-state. In the modeled device, the gate dielectric is characterized by equivalent oxide thickness (EOT) =1 nm. Different from the device structure studied in the previous section, a uniform channel is used here for convergence reasons. The channel is 8-nm thick In_{x}Ga_{1-x}As with InAs composition of $x_{ch}$ on an InAlAs buffer layer lattice matched to InP. The cap that forms the raised SD is also In_{x}Ga_{1-x}As but with an InAs composition of $x_{cap}$. The doping level is $2 \times 10^{19}$ cm$^{-3}$, and no doping or δ-doping is present anywhere. The thickness of the cap and channel is denoted as $t_{cap}$ and $t_{ch}$ respectively. The space between the gate edge and the ohmic contact is known as the access region and has a length $L_{access}$. The baseline structure uses $x_{ch}=0.7$, $x_{cap}=0.53$, $t_{cap}=30$ nm, $t_{ch}=8$ nm, $L_{access}=0$ and $L_{g}=40$ nm. The device simulation environment, physics model, and assumption developed in previous section apply here.

Fig. 6-12: Cross sectional device structure used in simulations.
6.4.2 BTBT-induced source-channel barrier lowering

Fig. 6-13 (a) shows the simulated subthreshold $I_d$-$V_{gs}$ characteristics at $V_{ds}=0.6$ V for $L_g=40, 100,$ and $800$ nm MOSFETs with and without BTBT. For all devices, $I_d$ decreases monotonically as $V_{gs}$ is reduced when the BTBT model is not included, as expected in an ideal MOSFET. The subthreshold swing degradation with decreasing $L_g$ is due to short-channel effects (SCE). In contrast, when the BTBT model is included, the OFF-state drain current, $I_d$ ($V_{gs}<-0.2$ V, $V_{ds}=0.6$ V) reaches a minimum value and then increases as $V_{gs}$ decreases far below threshold. This level of $I_d$ is set by the coupling between BTBT, bipolar gain and gate voltage through the modulation of the height of the potential barrier, $\phi_b$, which exists between the source and the channel under the gate (defined in Fig. 6-8).

Fig. 6-13: (a) Simulated subthreshold $I_d$-$V_{gs}$ characteristics with and without BTBT for MOSFETs with gate lengths of 40, 100, and 800 nm. (b) Corresponding source-channel barrier height ($q\phi_b$) as a function of $V_{gs}$. In all cases, $V_{ds}=0.6$ V.

To better understand this, Fig. 6-14 plots the conduction band-edge ($E_c$) from source to drain in the longitudinal direction at a depth of 2 nm into the channel ($y=-2$ nm in Fig. 2) for two transistors with $L_g=800$ and 40 nm. With the BTBT model turned off (Fig. 6-14, left panels), as $V_{gs}$ decreases, $q\phi_b$ increases until normal hole accumulation in the channel becomes prevalent.
In contrast, with the BTBT model present (right panels), $q\phi_b$ increases only initially with decreasing $V_{gs}$ and then becomes independent of $V_{gs}$. This is what causes $I_d$ in the OFF-state to essentially saturate. Making $V_{gs}$ more negative flattens the top of the inverse-U shaped band profile in the channel. This is because a significant concentration of holes accumulates in the channel leading to a more effective screening of the longitudinal charge distribution.

The evolution of $q\phi_b$ with $V_{gs}$ for the simulations of Fig. 6-13 (a) is more clearly visible in Fig. 6-14 (b). In the absence of BTBT, $q\phi_b$ increases as $V_{gs}$ becomes more negative. Eventually, saturation takes place when the Fermi level at the surface approaches the valence
band edge and the channel goes into hole accumulation. Including BTBT introduces increased accumulation of holes that limits the rise of $q\phi_b$ at higher values of $V_{gs}$ as follows: Reduction in $V_{gs}$ gives rise to higher $V_{dg}$ and hence a higher electric field at the drain edge of the channel; this enhances the BTBT generated hole current injected into the channel ($I_{h,BTBT}$) which in-turn brings about a reduction in $q\phi_b$ counterbalancing the effect of $V_{gs}$.

6.4.3 Bipolar gain amplification and $L_g$ dependency

As shown in Fig. 6-13 (b), the source-channel barrier height in the OFF-state in the presence of BTBT is essentially the same for the $L_g=100$ and 800 nm devices which is understandable because the BTBT current ($I_{BTBT} = |I_{e,BTBT}| = |I_{h,BTBT}|$) only depends on the E-field at the drain-edge and is independent of gate length. This is indeed verified in Fig. 6-15 that plots $I_{BTBT}$ corresponding to the three gate lengths. Compared to the difference between total currents $I_{d,T}$, the $I_{BTBT}$ currents are very similar – almost the same for the two longer devices and within a factor of 3 including the shortest one.

Fig. 6-15: (a) Electron ($I_e$) and hole ($I_{BTBT}$) channel currents and total drain current ($I_{d,T}$) characteristics vs. $V_{gs}$ for MOSFETs with gate lengths of 40, 100, and 800 nm at $V_{ds}=0.6$ V. (b) Enlarged view for the box region in (a).
The bipolar current gain, $\beta$, on the other hand, does depend on $L_g$. $\beta$ can be defined as:

$$\beta = \frac{I_e}{I_{BTBT}}$$

In the subthreshold regime, electron transport through the channel takes place by diffusion and a shorter $L_g$ leads to a steeper electron gradient and higher electron current. Fig. 6-15 (a) also plots the electron channel current ($I_e$) and the terminal drain current ($I_{d,T}$). $I_{d,T}$ denotes the total current by source electron injection plus BTBT, $I_{d,T} = I_e + I_{BTBT}$. It is clear that in spite of a similar $I_{BTBT}$, the shorter devices have significantly higher $I_e$ and therefore $I_{d,T}$. Since $\beta$ is much larger than unity ($I_e \gg I_{BTBT}$), it is expected that $I_{d,T} \approx I_e$. The gain is reduced for longer devices. This is indeed the case for the longer device, $L_g=800$ nm, as shown in Fig. 6-15 (b).

The simple physical argument leads one to the postulation of an inverse linear relationship between $\beta$ and $L_g$. Indeed this is what is observed in the simulations at all values of $V_{gs}$ in Fig. 6-16 which also shows that, for a given $L_g$, $\beta$ is reduced as the gate voltage becomes more negative. The physical origin is discussed below.

Short-channel effects complicate this idealized picture in a number of ways. In a BJT, the quasi-neutral base is the portion of the base with nearly zero longitudinal E-field, through which injected carriers flow by diffusion. Due to the presence of the depletion regions associated with the pn junctions, the effective electrical thickness of the quasi-neutral base is smaller than that of the metallurgical base. In analogy with this, in a QW-MOSFET in the subthreshold regime, the portion of the channel that is under negligible lateral field is shorter than the gate length. This does not appreciably affect long-channel devices but it becomes significant in short-channel transistors. As Fig. 6-14 shows, for $L_g=800$ nm the conduction band under the gate is rather flat while for $L_g=40$ nm it exhibits an inverse-U shape. The effective channel length in this case is substantially shorter than $L_g$ leading to an increase in current gain.

In the simulations for a given layer design, the difference between physical gate length and effective channel length is weakly dependent of $V_{gs}$. To the first order, this dependency is neglected. Then this difference can be seen in Fig. 6-16 where the extrapolation of $\beta^{-1}$ to $\beta^{-1} = 0$
yields $\Delta L_g \approx 70$ nm. $\Delta L_g$ is related to the penetration of the electric field lines that emanate from the source and drain into the channel. This effectively shortens the effective channel length that electrons diffuse through on their way from source to drain.

Fig. 6-16 reveals that $\beta$ decreases as $V_{gs}$ is made more negative. As in an npn bipolar transistor, $\beta$ is related to the barrier that holes must overcome to get injected into the emitter (source, in this case). The lower the barrier, the easier the injection and the lower $\beta$ is. The reduction of $\beta$ as $V_{gs}$ becomes more negative is associated with the reduction of $\phi_b$ for negative $V_{gs}$, as illustrated in Fig. 6-13 (b) in this regime.

![Graph](image)

Fig. 6-16: $\beta^{-1}$ vs. $L_g$ at $V_{ds}=0.6$ V for $V_{gs}$ values between -0.45 and -0.25 V.

### 6.5 Transistor Design for $I_{off}$ Reduction

This section discusses device design approaches to reduce BTBT and bipolar gain. The degree to which such approaches might be employed to mitigate excess OFF-state $I_d$ in InGaAs QW-MOSFETs would depend on considerations of the overall device performance and specifications. A detailed study of this is outside the scope of this work.
6.5.1 Delta-doping and channel doping

Different doping profiles have been implemented in experimental InGaAs QW-MOSFETs. N-type δ-doping is typically added beneath the channel (indicated by dashed line in Fig. 6-8) to reduce ON-state access resistance and to maintain high channel mobility [26]–[28]. Uniform p-type doping of the channel region has also been used [148]. It is found that the doping type and profile strongly influence the BTBT rate and the bipolar gain characteristics, as well as many other important transistor parameters, and should be an important consideration in transistor design.

Fig. 6-17 shows simulated subthreshold $I_d-V_{gs}$ and $I_{BTBT}-V_{gs}$ characteristics at $V_{ds}=0.6$ V as a function of backside δ-doping type and level (left), and channel uniform p-type doping level (right) for $L_g=40$ nm transistors. The backside δ-doping is located in the InAlAs buffer 5 nm beneath the channel. In Fig. 6-17 (a), the black line shows the case for zero δ-doping (the baseline design) and the arrows indicate increasing magnitude of δ-doping in $0.25 \times 10^{12}$ cm$^{-2}$ increments. Red color with a left-pointing arrow refers to n-type and blue with a right-pointing arrow for p-type. In Fig. 6-17 (b) the arrows indicate channel p-doping level changes from 0 to $3.5 \times 10^{18}$ cm$^{-3}$.

Changing doping level and type affects the threshold voltage, which shifts positively with increasing p-type doping and negatively with increasing n-type doping. In addition, the saturated $I_d$ (in the OFF-state) increases as the n-type doping increases and it decreases as p-type doping increases though eventually this reduction saturates. Interestingly, the behavior of $I_{BTBT}$ is counter to this, it decreases as n-type doping increases and it increases as p-type doping increases.

Fig. 6-17 reveals the important role that $\beta$ plays in $I_{off}$ and its dependence on doping. This can be seen more clearly in Fig. 6-18 that plots $\beta$ at $V_{gs}=-0.4$ V and $V_{ds}=0.6$ V against the doping level in the δ-doped layer (top x axis) and channel p-type doping $N_A$ (bottom x axis).
Fig. 6-17: Subthreshold $I_d$-$V_{gs}$, and $I_{BTBT}$-$V_{gs}$ characteristics for $L_g=40$ nm for different (a) backside $\delta$-doping type and doping level, and (b) channel uniform p-type doping level. For both simulations, $V_{ds}$ is 0.6 V.

Fig. 6-18: $\beta$ vs. channel uniform p-doping ($N_A$ at the bottom x axis) and backside $\delta$-doping level ($N_\delta$ at the top x axis) at $V_{gs} = -0.4$ V and $V_{ds} = 0.6$ V. $L_g$ for this device is 40 nm. Negative $\delta$-doping values indicate n-type doping.
The large impact of channel doping on the bipolar gain can be understood by considering the role of the vertical electric-field across the channel. In the channel of a QW-MOSFET biased in the subthreshold regime, an electric field exists in the transverse direction (perpendicular to the surface.) Its presence can significantly influence the bipolar action that is partially responsible for the excess OFF-state current. To illustrate this, Fig. 6-19 sketches the energy band diagram in the transverse direction under the gate in a device biased in the OFF-state with some degree of hole accumulation in the channel. The presence of a transverse field splits the electron and hole distributions spatially with holes being located closer to the surface and the electrons placed against the channel-buffer interface. Effectively this also implies an energy separation between the electron and hole populations that, neglecting quantization, is narrower than the bandgap by an amount roughly equal to the band bending in the channel, or \( q\phi_c \) in Fig. 6-19. In analogy to the impact on \( \beta \) of bandgap narrowing in the base of a BJT, this should increase the current gain by an amount roughly given by \( \exp(q\phi_c/kT) \). As \( V_{gs} \) becomes more negative, \( \phi_c \) increases and \( \beta \) increases too. Similar behavior in the bipolar action of silicon MOSFET has been reported [149], [150]. Also, in silicon SOI MOSFETs, an increase in \( I_d \) is observed after irradiation [99], [141] which generates positive fixed charge in the buried oxide beneath the channel and leads to electron accumulation at the back interface.

The impact of channel doping on the bipolar gain seen in Fig. 6-18 can be easily understood in these terms. Fig. 6-20 (a) sketches energy band diagrams in the vertical direction in the channel under the gate at the same \( V_{gs}-V_t \) \((V_{gs}-V_t<0)\) for three different cases: p-type \( \delta \)-doping, no \( \delta \)-doping and n-type \( \delta \)-doping. For clarity, the band diagrams are matched at the oxide-semiconductor interface. With an n-type \( \delta \)-doping in the buffer, there is steeper band bending in the channel that results in increased current gain. Conversely, with increased p-type delta doping, \( \phi_c \) is reduced and \( \beta \) also shrinks. Similar reasoning applies for the uniform p-type channel doping design.

To confirm this understanding, Fig. 6-20 (b) shows \( \beta \) vs. \( \phi_c \) as extracted from the simulator for different \( \delta \)-doping designs. Those values are extracted at fixed \( V_{gs}-V_t=-0.35 \) V \((V_t \)}
is defined at $I_d=1 \ \mu A/\mu m$ and $V_{ds}=0.6 \ V$ from Fig. 6-17 (a). The expected exponential dependence of $\beta$ on $\phi_c$ is indeed obtained.

Fig. 6-19: Energy band diagram below the gate around the source in the direction perpendicular to the surface.

Fig. 6-20: (a) Simulated energy band diagrams in the perpendicular direction in the channel under the gate for three different types of backside $\delta$-doping (at the same $V_{gs}-V_t$). For clarity, the band diagrams are matched at the oxide-semiconductor interface. (b) $\beta$ vs. $q\phi_c$ at fixed $V_{gs}-V_t= -0.35 \ V$ and $V_{ds}=0.6 \ V$ in semilog scale indicates the exponential dependency of $\beta$ on $q\phi_c$. $\phi_c$ is extracted in the center of the channel ($z=0$).
This discussion reveals that the optimum design is the one that minimizes band bending in the channel in the transverse direction in the subthreshold regime. Consequently, device architectures which improve the gate control over the channel potential in the subthreshold regime, such as ultra-thin-body QW-MOSFET, double-gate FinFET, and Trigate MOSFET or gate-all-around nanowire design, will be an advantage.

6.5.2 InAs composition in channel and cap

So far, the studied device design has InAs composition in the channel of $x_{ch}=0.7$ and in the $n^+$-cap, $x_{cap}=0.53$. The use of a different InAs composition in the cap and channel results in a heterojunction at the cap/channel interface. This introduces an additional barrier to hole injection from the channel into the cap that greatly increases the bipolar current gain. In essence, this behaves as a heterojunction bipolar transistor (HBT) [115]. Fig. 6-21 illustrates this case by sketching an energy band diagram under the contacts in the vertical direction.

Therefore, the effect of varying $x_{cap}$ while keeping $x_{ch}$ constant has been studied. Fig. 6-22 graphs $I_d$, $I_{BTBT}$ and $\beta$ vs. $x_{cap}$ at a fixed $V_{gs}=-0.35$ V and $L_g=40$ nm for $x_{ch}=0.7$. It is clear that OFF-state $I_d$ decreases when $x_{cap}$ increases from 0.3 to 1. This is because $\beta$ decreases in a very significant way as the hole barrier at the channel-cap interface is reduced and eventually eliminated. In contrast, the BTBT current increases as $x_{cap}$ increases.

BTBT is inherently a non-local process (Fig. 6-11a). At the drain edge of the channel, electron generation mostly happens in the cap, and hole generation in the channel. Hence, the InAs composition of both the cap and the channel, through their respective bandgaps, affects the tunneling barrier. As the bandgap in the cap is reduced, the tunneling barrier for BTBT also shrinks and BTBT at the drain edge of the channel is enhanced. This is consistent with experimental observations in [133]. The increase in $I_{BTBT}$, however, is significantly weaker than the bipolar gain reduction and OFF-state leakage follows the trend of the latter. When $x_{cap}$ approaches 1, $\beta$ vanishes, and OFF-state leakage approaches $I_{BTBT}$.
Fig. 6-23 shows the impact of $x_{ch}$ for a fixed $x_{cap}=0.53$. OFF-state $I_d$ increases as $x_{ch}$ increases. In this case, $\beta$ and $I_{BTBT}$ are both enhanced. $\beta$ enhancement can be understood by the heterojunction effect and the BTBT current enhancement as a result of bandgap reduction in the channel.

This study reveals that the InAs composition in the channel and cap both play very important role in determining OFF-state $I_d$. All other things being equal, it is clear that reducing the valence band discontinuity ($\Delta E_v = E_{v,ch}-E_{v,cap}$) at the channel-cap interface constitutes an effective way to suppress OFF-state current.

![Energy band diagram](image)

Fig. 6-21: Energy band diagram under the contact along the perpendicular direction for $x_{cap}<x_{ch}$, ($E_{g,cap}>E_{g,ch}$).

![Graphs](image)

Fig. 6-22: (a) Extracted $I_d$ and $I_{BTBT}$ and (b) $\beta$ vs. $x_{cap}$ at $V_{gs}=-0.35$ V and $V_{ds}=0.6$ V. $L_g$ for this device is 40 nm and $x_{ch}=0.7$. 
Fig. 6-23: (a) $I_d$ and $I_{BTBT}$ and (b) $\beta$ vs. $x_{ch}$ at $V_{gs} = -0.35$ V and $V_{ds} = 0.6$ V. $L_g$ for this device is 40 nm and $x_{cap} = 0.53$.

6.5.3 Carrier lifetime

In silicon SOI MOSFET, an effective approach to mitigate the parasitic bipolar effect is to introduce recombination centers that reduce the carrier lifetime [151]. Recombination of the BTBT-generated holes in the channel brings down their concentration and the need to inject them across the source-channel barrier.

This effect is studied by introducing a Shockley-Read-Hall type recombination process in the cap and channel that is characterized by a carrier lifetime, $\tau$ [145]. Fig. 6-24 shows the impact of $\tau$ on OFF-state $I_d$, $I_{BTBT}$, and $\beta$ at $V_{gs} = -0.45$ V and $V_{ds} = 0.6$ V for the baseline device. As expected, the BTBT current is independent of $\tau$. The $I_d$ behavior then follows the evolution of the bipolar gain. For $\tau > 1$ ns, $I_d$ and $\beta$ are almost independent of $\tau$. In this regime, the carrier diffusion length is much longer than the device dimensions so hole recombination mainly takes place at the contacts. As $\tau$ falls below 1 ns, $I_d$ and $\beta$ decrease rapidly as a result of hole recombination in the channel and cap.
In silicon MOSFETs, carrier lifetime reduction can be achieved by implantation of various impurities, e.g. germanium [151]. For InGaAs MOSFETs, lifetime reduction could be implemented by introducing recombination centers during source/drain regrowth or in the channel recess process. Of course, a concern in doing this is its impact on the transport characteristics.

![Fig. 6-24: (a) Extracted I_d and I_{BTBT} and (b) β vs. τ at V_{gs} = -0.45 V and V_{ds} = 0.6 V. L_g for this device is 40 nm.](image)

6.5.4 Dimensional design of access region

Fig. 6-25 shows the impact of access region design on OFF-state I_d. Increasing the access region length and the cap thickness both result in higher I_d. In both cases, I_{BTBT} is independent of L_{access} and t_{cap} as the access region is heavily doped and the electron concentration in the channel directly underneath is unaffected by changes in either of these two parameters. I_d then follows the bipolar gain β which exhibits a linear trend with both L_{access} and t_{cap}. This is expected from classic BJT behavior as in both cases hole injection is suppressed as the path that holes must diffuse through in the cap region before they recombine at the contact is lengthened [115]. The linear dependence of I_d on both L_{access} and t_{cap} is weaker than other dependencies studied above.
and therefore engineering the access regions offers limited opportunities for significant suppression of excess leakage in comparison with their relevance in the overall operation of the device.

![Graphs showing Id extraction at Vgs = -0.35 V and Vds = 0.6 V for varying Laccess and tcap.](image)

**Fig. 6-25**: $I_d$ extracted at $V_{gs} = -0.35$ V and $V_{ds} = 0.6$ V for (a) varying $L_{access}$ and (b) varying $t_{cap}$ in otherwise a baseline device design.

### 6.6 Chapter Summary

In this chapter, the origin of high OFF-state leakage in high-transconductance InGaAs MOSFETs has been investigated in a systematic way. The effects of gate leakage, buffer leakage, Fermi-level pinning and conventional short channel effects are first isolated. It is found that the OFF-state current bears the characteristic signature of GIDL. Temperature and bias dependence supports that BTBT is indeed the triggering mechanism for excess leakage. The fact that the OFF-state leakage scales inversely with gate length suggests an additional bipolar gain mechanism such as observed in SOI MOSFETs. 2D device simulations support this notion. Then the coupling behind the BTBT and bipolar gain effect are explored in 2D device simulations. Finally, approaches to mitigation of the excess OFF-state leakage current in scaled InGaAs QW-MOSFETs are discussed. The most effective approach consists of minimizing the current gain of
the parasitic bipolar transistor in these devices. Four different ways of accomplishing this are discussed: (1) Reduction of the vertical electric field in the channel by doping engineering; (2) Reduction of the heterojunction effect by engineering the InAs compositions in the cap and channel; (3) Reduction of lifetime of excess carriers generated by BTBT; and (4) Dimensional control of the extrinsic device.
Chapter 7: Conclusions and Suggestions for Further Research

7.1 Thesis Summary

This work has advanced the fabrication technology of III-V MOSFETs. It first identifies the critical challenges that must be addressed in current InGaAs FETs for CMOS applications. Then a new type of InGaAs MOSFET that takes into consideration scalability, performance, and CMOS-compatibility is designed and fabricated.

The device architectural design in this work emphasizes scalability and manufacturability by making extensive use of dry etching and Si-compatible materials. The fabrication sequence yields precise control of all critical transistor dimensions. This work achieved InGaAs MOSFETs with the shortest gate length ($L_g = 20$ nm), and MOSFET arrays with the smallest contact size ($L_c = 40$ nm) and smallest pitch size ($L_p = 150$ nm), at the time they were made. This work has achieved long channel InGaAs MOSFET with subthreshold swing of 69 mV/dec, one of the lowest values of any InGaAs MOSFET. This work was also the first to demonstrate pure HfO$_2$ as gate insulator for deeply scaled InGaAs MOSFETs, in which the EOT of the dielectric is only 0.5 nm. A novel gate recess technology that combines anisotropic III-V dry etch and digital etch have also been developed. Sub-1-nm thickness control is demonstrated. The III-V dry etch is optimized to reduce surface roughness, and to improve the trenching effect. Using a wafer bonding technique, the InGaAs QW-MOSFET is also integrated onto a silicon substrate.

The fabricated transistors show the potential of InGaAs to yield devices with well-balanced electron transport, electrostatic integrity and parasitic resistance. A device design optimized for the ON-state exhibits a transconductance of 3.1 mS/µm, a record value for III-V
FETs of any kind in 2014. This device also exceeded the transconductance of HEMT for the first time in history. The field effective mobility for a buried channel InGaAs MOSFET represents the highest values of III-V MOSFETs. The subthreshold swing characteristics and ON-current benchmarking in the scaled InGaAs MOSFETs in this work had been showing the highest values at the time when they were made.

A three-layer resistive-network model is developed to understand the operation of a fabricated MOSFET array. From this analysis, a contact resistivity from Mo to $n^+$ InGaAs of $8 \times 10^{-9} \Omega \cdot \text{cm}^2$, and from $n^+$ InGaAs to channel of $2 \times 10^{-9} \Omega \cdot \text{cm}^2$ is extracted. When benchmarked with other ohmic contact technologies proposed for n-type InGaAs MOSFETs, the refractory metal contact in this work represents the lowest film resistivity and is among the lowest for contact resistivity.

The precise fabrication technology has also enabled a detailed study of the impact of critical transistor design on its performance. Two critical transistor designs are discussed. The first one is the impact of channel thickness scaling. InGaAs MOSFETs of different channel thickness with gate lengths that span from 40 nm to 5 $\mu$m is fabricated. This addresses the issues of both carrier transport and SCE control. It is found that the intrinsic transconductance follows a classic gate length scaling behavior. Surface channel designs with a relatively thick channel provide the best transconductance characteristics. Buried-channel designs feature lower transconductance presumably due to a loss of gate capacitance. Thin channels suffer from severe transconductance degradation. From a study of channel mobility it identifies the InAs composition that electrons experience in the channel as a key factor affecting transport characteristics. In contrast with their excellent ON-state performance, the OFF-state characteristics of thick channel devices are significantly compromised. It is found that the subthreshold swing and DIBL follow classic scaling behavior similar to that of Si FDSOI MOSFETs. The current InGaAs QW-MOSFET is at the limit of scaling around $L_g=50$ nm.

Access region design is found to play important role in overall device performance. A fabrication method is developed to control the sheet carrier and length in the access region.
through which the balance of ON-state performance vs. short-channel effects is explored. The short-access MOSFETs demonstrate the highest $g_m$ and low $R_{sd}$. Long-access MOSFETs have outstanding short-channel effects and high ON-current. When carrier concentration is low in the access region, source starvation starts to emerge.

The scaled III-V device architecture achieved in this work has also enabled new device physics studies relevant for the application of InGaAs transistors for future logic. A particularly important one is OFF-state leakage. For the first time, this work has unambiguously identified BTBT amplified by a parasitic bipolar effect as the cause of excess OFF-state leakage current in these transistors. This finding has important implications for future device design. Mitigating approaches to excess OFF-state leakage current in scaled InGaAs QW-MOSFETs have also discussed. The most effective approach consists of minimizing the current gain of the parasitic bipolar transistor in these devices. Approaches for accomplishing this have also been investigated.

7.2 Suggested Further Research

7.2.1 Transistor redesign and 3D device architectures

As benchmarked in Ch. 3, the recessed-gate InGaAs QW-MOSFET developed in this work represents the state-of-the-art deeply-scaled III-V MOSFET technology to date. Following analysis has indicated several possible intrinsic and extrinsic engineering approaches for further performance enhancement. In particular, important design parameters to investigate include the channel doping for $I_{off}$ suppression and source doping for ON-state enhancement. These studies can guide the next generation of transistor redesign.

In the quest for high device density, footprint scaling demands scaling of all transistor dimensions. Short channel effects become a serious concern due to gate length scaling. Results in
Ch. 3 has shown that current InGaAs QW-MOSFET is at the limit of scaling around $L_g=50$ nm. It suggests the need for advanced 3D device architectures to deliver further progress.

In order to mitigate short channel effects, enhanced electrostatic gate control to the channel is required. To realize this, transistor design must evolve, from a planar MOSFET to advanced 3D device structures that include FinFET, Tri-gate MOSFET and Gate-All-Around Nanowire MOSFET. All these advanced 3D device structures have been demonstrated in the InGaAs system [152]. However, a 3D InGaAs transistor displaying well-balanced electron transport, electrostatic integrity and parasitic resistance have yet to be demonstrated. The challenges are discussed next.

The fabrication of InGaAs MOSFET with 3D architectures is a great challenge. To fabricate III-V Fin and nanowire structure, extensive III-V RIE is required. The semiconductor in the intrinsic region can be severely deteriorated. RIE damage becomes an important concern. In this thesis, the processes have extensively use RIE to define the intrinsic gate region. Despite multiple RIE processes being used, the low subthreshold swing, high transconductance and high ON-drive-current (Ch. 2 and Ch. 3) in device fabricated in this work suggest that the III-V interface remains superior. Extending the device technology in this work toward a 3D MOSFET structure can help to address those problems mentioned above.

Another reason for inferior ON-state performance of 3D InGaAs MOSFETs is related to the strong quantization in the channel. This is required to mitigate short channel effects. But it will also degrade the transport properties of the channel carrier due to various scattering mechanisms and the non-parabolic effective mass. In this regard, it is important to study the fundamental trade-off between OFF- and ON-state performances in the context of 3D device architectures.

### 7.2.2 Modeling of the transport and charge control

The prospects of using alternative III-V channel as performance drivers for future CMOS should be examined by benchmarking it with silicon technology using appropriate metrics. The
low effective mass of InGaAs is beneficial to achieving a high virtual source velocity. However, also a result of low effective mass, there is increasing concern about the low density of states, hence, low effective gate capacitance for InGaAs MOSFETs. Simulation studies have alerted the possibility of losing ON-current due to low density of states in InGaAs channel transistors [153], [154]. At this point, this remains as theoretical study that is based on microscopic physics models. The high-performance device fabricated in this work can serve as an ideal technology platform to calibrate theoretical model. Experimental studies on high-performance InGaAs MOSFETs of different quantum-well thickness (different quantization) and with gate lengths that span from the drift-diffusion regime to the ballistic transport regime should be fabricated, and analyzed using advanced device models [155], [156] to achieve this goal.
## Appendix: Process Integration for InGaAs MOSFET

<table>
<thead>
<tr>
<th>Module</th>
<th>Step</th>
<th>Spec</th>
<th>Rep</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohmic</td>
<td>Clean</td>
<td>HCl:H₂O (1:3) 30 sec</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>Mo 20 nm</td>
<td>linjq_Mo_gun3_Run</td>
<td>AJATRL</td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>W 10 nm</td>
<td>linjq_W_gun3_Run</td>
<td>AJATRL</td>
</tr>
<tr>
<td></td>
<td>Oxide deposition</td>
<td>SiO₂ 50 nm</td>
<td>CVD teos zero stress 350C</td>
<td>Oxford</td>
</tr>
<tr>
<td>Alignment mark</td>
<td>EBL</td>
<td>PMMA A8</td>
<td>Coater</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exposure</td>
<td>Elionix</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIBK:IPA (1:3) 1 min</td>
<td>Photohood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>Ti/Au (10/80 nm)</td>
<td>EbeamFP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiftOff</td>
<td>Acetone (1 h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-III-V recess</td>
<td>EBL</td>
<td>Zep:Anisole</td>
<td>Coater</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exposure</td>
<td>Elionix</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Xylene 1 min</td>
<td>Photohood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Etching</td>
<td>SiO₂ etch</td>
<td>JQSIO2</td>
<td>PQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mo/W etch</td>
<td>JQW1</td>
<td>PQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mo/W pull in</td>
<td>JQMoTrim</td>
<td>PQ</td>
</tr>
<tr>
<td></td>
<td>Resist strip</td>
<td>NMP (1 h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesa</td>
<td>Positive photo</td>
<td>PR OCG825</td>
<td>Coater</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-bake: 80°C 3 min</td>
<td>Hotplate</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA6 Exposural 10 sec</td>
<td>MA6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Develop OCG934 ~90 sec</td>
<td>Photohood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Etching</td>
<td>SiO₂ etch</td>
<td>JQSIO2</td>
<td>PQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mo/W etch</td>
<td>JQW1</td>
<td>PQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III-V etch</td>
<td>InGaAsJQ</td>
<td>Samco</td>
</tr>
<tr>
<td></td>
<td>Resist strip</td>
<td>NMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module</td>
<td>Step</td>
<td>Spec</td>
<td>Rcp</td>
<td>Tool</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>III-V recess</td>
<td>Etching</td>
<td>III-V etch</td>
<td>InGaAsJQ</td>
<td>Samco</td>
</tr>
<tr>
<td></td>
<td>Damage anneal</td>
<td>Damage anneal</td>
<td>340°C N₂ 15 min</td>
<td>Savannah</td>
</tr>
<tr>
<td></td>
<td>Digital etch</td>
<td>DE</td>
<td>Acid / Asher</td>
<td></td>
</tr>
<tr>
<td>Gate stack</td>
<td>ALD</td>
<td>HfO₂ 2.5 nm 250°C</td>
<td>Savannah</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>Mo 50 nm no rotation</td>
<td>EbeamAu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EBL</td>
<td>PMMA A8</td>
<td>Coater</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exposure</td>
<td>Eionix</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIBK:IPA (1:3) 1 min</td>
<td>Photohood</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>Ti/Au (10/100 nm)</td>
<td>EbeamFP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liftoff</td>
<td>Acetone (1 h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Etching</td>
<td>Mo etch</td>
<td>JQW2</td>
<td>PQ</td>
</tr>
<tr>
<td>Back-end</td>
<td>EBL</td>
<td>Zep/PMGI</td>
<td>Coater</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exposure</td>
<td>Eionix</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xylene 1 min + CD₂₆ 15 sec</td>
<td>Photohood</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Etching</td>
<td>HfO₂ etch (RT)</td>
<td>High-K ETCH</td>
<td>Samco</td>
</tr>
<tr>
<td></td>
<td>SiO₂ etch</td>
<td>JQSIO2</td>
<td>PQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>Ti/Au (10/100 nm)</td>
<td>EbeamFP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liftoff</td>
<td>PG remover</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative photo</td>
<td>PR AZ5214</td>
<td>Coater</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pre-bake: 80°C 3 min</td>
<td>Hotplate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA6 Exposural 10 sec</td>
<td>MA6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-bake: 110°C 2 min</td>
<td>Hotplate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Develop AZ422 ~ 2 min</td>
<td>Photohood</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metal deposition</td>
<td>Ti/Au (10/200 nm)</td>
<td>EbeamFP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liftoff</td>
<td>Acetone</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Recipes in the process

#### AJA Mo deposition
Recipe name: linjq_Mo-gun3_Run
Loadlock pressure <1E-5 Torr
Main chamber base pressure <1E-6 Torr
   - Plasma strike power: 150 W
   - Deposition power: 100 W
Rate: 0.9 A/sec

#### AJA W deposition
Recipe name: linjq_W-gun3_Run
Loadlock pressure <1E-5 Torr
Main chamber base pressure <1E-6 Torr
   - Plasma strike at 150 W
   - Deposition power 120 W
Rate: 1 A/sec

#### PQ SiO2 etching
Recipe name: JQSIO2
   - w/ gas stabilization and plasma strike
   - Temperature: RT
   - Process pressure: 20 mTorr
   - ECR: 200 W
   - DC bias: 100 V
   - CF₄ flow: 28 sccm
   - H₂ flow: 4 sccm
Rate: ~10 nm/min (calibration needed)

#### PQ W and Mo etching
Recipe name: JQW1
   - w/ gas stabilization and plasma strike
   - Temperature: RT
   - Process pressure: 14 mTorr
   - ECR: 250 W
   - DC bias: 50 V
   - SF₆ flow: 90 sccm
   - O₂ flow: 10 sccm
Rate: ~20 nm/min (calibration needed)

#### PQ W and Mo pull in
Recipe name: JQMoTrim
   - w/ gas stabilization and plasma strike
   - Temperature: RT
   - Process pressure: 50 mTorr
   - ECR: 300 W
   - DC bias: 0 V
   - CF₄ flow: 20 sccm
   - O₂ flow: 80 sccm
Rate: ~30 nm/min (calibration needed)
**Samco III-V recess etching**
Recipe name: rcp67-InGaAsJQ  
  w/ gas stabilization and plasma strike  
  Temperature: 120 °C  
  Process pressure: 0.2 Pa  
  Bias: 50 W (~120 V)  
  ICP: 20 W  
  Cl₂ flow: 10 sccm  
  N₂ flow: 3 sccm  
  Rate: ~15 nm/min (calibration needed)

**Samco high-k etching**
  w/ gas stabilization and plasma strike  
  Temperature: RT  
  Process pressure: 0.2 Pa  
  Bias: 75 W  
  ICP: 75 W  
  BCl₃ flow: 10 sccm  
  Ar flow: 3 sccm  
  Rate: ~20 nm/min (calibration needed)

**Digital Etch (DE)**
  Native oxide formation  
      TRL-Asher  
      Power: 800 W  
      Time: 150 sec  
  Oxide stripping  
      H₂SO₄:H₂O =1:1 (cool to RT)  
      Ultrasonic for the first 3 sec  
      (Model 250D, power level 2)

**InGaAs cap wet**
  Citri acid:H₂O₂=20:1  
  Ultrasonic for the first 3 sec  
      (Model 250D, power level 2)  
  Rate: ~2 nm/sec
Bibliography


[22] T.-W. Kim, D.-H. Kim, and J. A. del Alamo, “30 nm In0.7Ga0.3As Inverted-Type HEMTs with reduced gate leakage current for logic applications,” in *IEDM Tech. Dig.*, 2009, pp. 483–486.


[24] D.-H. Kim, B. Brar, and J. A. del Alamo, “fT = 688 GHz and fmax = 800 GHz in Lg = 40 nm In0.7Ga0.3As MHEMTs with gm_max=2.7 mS/um,” in *IEDM Tech. Dig.*, 2011, pp. 13.6.1–13.6.4.


[31] Y. Yonai, T. Kanazawa, S. Ikeda, and Y. Miyamoto, “High drain current (>2A/mm) InGaAs channel MOSFET at VD=0.5V with shrinkage of channel length by InP anisotropic etching,” in *IEDM Tech. Dig.*, 2011, pp. 13.3.1–13.3.4.


[33] D.-H. Kim, P. Hundal, A. Papavasiliou, P. Chen, C. King, J. Paniagua, M. Urteaga, B. Brar, Y. G. Kim, J.-M. Kuo, J. Li, P. Pinsukanjana, and Y. C. Kao, “E-mode planar $L_g=35$ nm In$_0.7$Ga$_{0.3}$As MOSFETs with InP/Al$_2$O$_3$/HfO$_2$ (EOT=0.8 nm) composite insulator,” in IEDM Tech. Dig., 2012, pp. 32.2.1–32.2.4.


[57] X. Zhang, H. X. Guo, X. Gong, C. Guo, and Y.-C. Yeo, “Multiple-Gate In0.53Ga0.47As Channel n-MOSFETs with Self-Aligned Ni-InGaAs Contacts,” *ECS Trans.*, vol. 45, no. 4, pp. 209–216, Apr. 2012.


tri-gate MOSFETs with Al2O3/HfO2 (EOT=1 nm) for low-power logic applications,” in *IEDM Tech. Dig.*, 2013, pp. 16.3.1–16.3.4.

[101] S. Lee, V. Chobpattana, C.-Y. Huang, B. J. Thibeault, W. Mitchell, S. Stemmer, A. C. Gossard, and M. J. W. Rodwell, “Record Ion (0.50 mA/um at VDD = 0.5 V and Ioff = 100 nA/um) 25 nm-gate-length ZrO2/InAs/InAlAs MOSFETs,” in *Proc. VLSI Tech.*, 2014, pp. 54–55.


[109] X. Zhou, Q. Li, C. W. Tang, and K. M. Lau, “30nm enhancement-mode In0.53Ga0.47As MOSFETs on Si substrates grown by MOCVD exhibiting high transconductance and low on-resistance,” in *IEDM Tech. Dig.*, 2012, pp. 32.5.1–32.5.4.


