Design, Fabrication, and Characterization of High Performance III-V nMOSFETs for VLSI Beyond Si-CMOS Scaling Limit

A Dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical and Computer Engineering by Sanghoon Lee

Committee in Charge:
Professor Mark J. W. Rodwell, Chair
Professor Arthur C. Gossard
Professor John E. Bowers
Professor Umesh Mishra

October 2014
The Dissertation of Sanghoon Lee is approved.

Professor Arthur C. Gossard

Professor John E. Bowers

Professor Umesh Mishra

Professor Mark J. W. Rodwell, Committee Chair

October 2014
Design, Fabrication, and Characterization of High Performance III-V
nMOSFETs for VLSI Beyond Si-CMOS Scaling Limit

Copyright © 2014

by

Sanghoon Lee
This dissertation is dedicated to my parents

for their love and support.
Acknowledgements

First of all, I would like to express my deepest and sincere gratitude to my advisor Professor Mark Rodwell for giving me the opportunity to join his group and work on this exciting research topic of III-V MOSFETs. Whenever I faced difficulties in my research, his constant enthusiasm, passion, and inspiration motivated and guided me to the right direction and encouraged me to come up with new ideas. I have no doubt that the intellectual and personal growths fostered under his guidance would become one of the most invaluable assets for my future professional career and personal life. I would also like to thank my Ph. D. committee members, Professor Arthur Gossard, Professor John Bowers, and Professor Umesh Mishra for their critical and valuable comments and suggestions for the qualifier and the dissertation.

I am grateful to the predecessors of this project, Uttam Singisetti, Greg Burek, and Mark Wistey. Without their ground works and knowledge accumulated at the early stages of this project, it would have been very difficult for this work to be on the right track. I would also like to express my gratitude to the past and current FET teams. Andrew Carter was always willing to share his expertise and continued to inspire me in many ways with his great passion and exuberant energy. Jeremy Law and Doron Elias contributed to this dissertation by providing the epistuctures and regrowths and giving me helpful suggestions. Cheng-Ying Huang played a key role in this project by optimizing the essential epitaxial layers and providing most of the epistuctures and regrowths. Without his effort and devotion, our remarkable accomplishments could never have been achieved in such a short period of time. Hands-on experiences and fruitful discussions and adv-
vice on the experiments from Bill Mitchell and Brian Thibeault have also hugely contributed to enabling all these results.

I would like to thank the UCSB dielectric team, Varistha Chobpattana and Professor Susanne Stemmer for sharing their extraordinary high-\(k\) dielectric recipe. Without the close collaboration, such record-breaking results could not have been achieved. I would also like to thank Dr. Stephan Krämer for helping out with FIB and TEM in despite of his busy schedules. I would like to thank all of the nanofab staffs at UCSB who have provided continued support to well maintain the cleanroom tools.

I would also like to sincerely thank the past and current group members including, Evan Lobisser, Andrew Carter, Thomas Reed, Johann Rode, Han-Wei Chiang, Vibhor Jain, Ashish Baraskar, Rob Mauer, Jeremy Law, Doron Elias, Hyunchul Park, Seongkyun Kim, and Prateek Choudhary. I would like to send special thanks to Hyunchul Park for his moral support.

Last but not least, I would like to thank my parents who have raised me with unconditional support and strong sense of personal responsibility. I thank Jeong Kim for standing by me. Without her encouragement and unwavering love, I could have never overcome those difficult moments faced from time to time during the past four years.
Curriculum Vitae
Sanghoon Lee

Personal

Education
2014 Ph.D. in Electrical and Computer Engineering, University of California, Santa Barbara.
2009 M.S. in Electrical Engineering, Seoul National University, South Korea.
2007 B.S. in Electronics and Information Engineering, Kyunghee University, South Korea.

Awards
2014 UCSB graduate school division dissertation fellowship
Graduate Division at UC Santa Barbara, May, 2014.
2011 Best Student Paper Award
2009 Best Poster Paper Award at Mini-Colloquium on Nano-Scale Devices and Circuits
IEEE EDS Korea Chapter, 2009.
2008 Presidential Award for top honors
Kyunghae University, Aug. 2007.

Publications
[1] (Late News) 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/µm at VDD = 0.5 V and Ioff = 100 nA/µm) 25 nm-Gate-Length ZrO₂/InAs/InAlAs MOSFETs,” VLSI Symposia on VLSI Technology and Circuits, June, 2014.


[6] Sanghoon Lee, Cheng-Ying Huang, Doron C. Elias1, Andrew D. Carter, Jeremy J. M. Law, Varistha Chobpattana, Brian J. Thibeault, William Mitchell, Stephan Kramer, Susanne Stemmer, Arthur C. Gossard, and Mark J. W. Rodwell “High Transconductance (2.45 mS/$\mu m$ at $VDS = 0.5 V$) InAs/In$_{0.53}$Ga$_{0.47}$As channel MOSFETs using Source-Drain Regrowth and Digital Etching Technique,” TECHCON Conference, Sept., 2013

[7] (Late News) S. Lee, C.-Y. Huang, A. D. Carter, D. C. Elias, J. J. M. Law, V. Chobpattana, S. Krämer, B. J. Thibeault, W. Mitchell, S. Stemmer, A. C. Gossard, and M. J. W. Rodwell, “Record Extrinsic Transconductance (2.45 mS/$\mu m$ at $VDS = 0.5 V$) InAs/In$_{0.53}$Ga$_{0.47}$As Channel MOSFETs Using MOCVD Source-Drain Regrowth,” VLSI Symposia on VLSI Technology and Circuits, June, 2013


[13] C. Huang, S. Lee, V. Chobpattana, S. Stemmer, A. Gossard, B. Thibeault, W. Mitchell and M. Rodwell “Low Power III-V InGaAs MOSFETs Featuring InP Recessed Source/Drain Spacers with Ion=120 μA/μm at Ioff=1 nA/m and VDS=0.5V,” in Proc. IEEE IEDM, Dec. 2014. (To be presented)


InGaAs finFETs of 200 nm Height by Atomic Layer Epitaxy,” IEEE Device Research Conference, June 23-26, 2013 Notre Dame University South Bend, IN


Abstract

Design, Fabrication, and Characterization of High Performance III-V nMOSFETs for VLSI Beyond Si-CMOS Scaling Limit

by

Sanghoon Lee

The revolution of the silicon VLSI technology during the past several decades has been ultimately driven by the goal of miniaturization, which leads to an increase in switching speed as well as integration density and a reduced power consumption. As the device size in VLSI has nearly approached its physical limit in the last few years, the industry and academia have been actively evaluating some of the emerging technologies as an alternative to the classical Si-based metal-oxide-semiconductor field-effect-transistors (MOSFETs). Among them, III-V compound semiconductor based transistors are being considered as one of the viable candidates for the future VLSI at scaling generations beyond 7 nm-node. The low electron effective masses in III-V semiconductor materials (i.e. InGaAs) provide superior electron transport properties such as high electron velocity and mobility. According to ballistic transport calculation results, InGaAs based channel devices can potentially exhibit a 1.5 times higher drive current (> 2 mA/µm) even at a lower supply voltage ($V_{DD} < 0.7$ V) over the Si counterparts. Thus, for faster and smaller integrated circuits with reduced power consumption, III-V based transistors may be the solution to VLSI beyond the physical limitations encountered in scaling of the conventional Si-based MOSFETs.
In order to achieve device performances close to the idealized target, several critical requirements must be met. Firstly, the high-
$k$ gate dielectric must be ultra-thin (equivalent oxide thickness $< 0.5$ nm) and nearly defect-free (interfacial trap density $< 10^{12}$ /cm$^2$-eV). Hence, any high damage inducing process is not allowed, and surface passivation techniques must be carefully developed. Secondly, the epitaxial layer design should be optimized, especially since there is a trade-off between the on- and off-state performances associated with the channel thickness as well as indium content. Thirdly, S/D regions must be very heavily doped in order to avoid potential source starvation and to minimize the contact resistivity. Furthermore, this heavy doping must not extend more than 1-2 nm below the depth of the channel to avoid degradation of the short-channel characteristics. Lastly, the device must be highly scalable. To satisfy the tight integration density requirement in VLSI, the gate length and contact pitch should be less than 14 nm and 30 nm, respectively. To achieve this, S/D must be very close to the gate, i.e., self-aligned.

With the abovementioned key design considerations in mind, InGaAs based raised S/D quantum-well MOSFETs have been developed using S/D regrowth as well as the substitutional-gate (i.e., gate-last) scheme. By adopting this device structure, any process-induced damage at the semiconductor/dielectric interface is reduced, and heavily doped S/D is readily formed in a self-aligned manner. Recently, III-V MOSFETs with a record performance have been reported through this work, by implementing sub-1 nm EOT high-$k$ dielectrics with a low interface trap density and adopting an optimized device structure to suppress the off-state degradation at the short channel lengths. A device with a gate length of 18 nm
has shown a 3.0 mS/µm peak transconductance ($g_m$) at $V_{DS} = 0.5$ V, which is the highest peak $g_m$ from any reported field-effect transistor performance. A device with an ultra-thin channel and thick vertical spacer at a gate length of 25 nm exhibits an excellent performance in both the on-state and off-state, featuring 2.4 mS/µm peak $g_m$, 77 mV/decade minimum subthreshold swing at $V_{DS} = 0.5$ V, 76 mV/V drain-induced barrier lowering, and 500 µA/µm on-current at a fixed 100 nA/µm off-current and $V_{DD} = 0.5$ V. This is the highest on-current from any reported III-V-based MOSFETs and is comparable to state-of-the-art Si-Fin- and nanowire- FETs. In comparison with calculation results obtained from a ballistic FET model, it has been found that the fabricated devices with $L_g = 25$ nm are operating nearly in the ballistic limit.
# Contents

Acknowledgements ........................................ v
Curriculum Vitæ ........................................ vii

Abstract .................................................. xi

List of Figures ............................................ xvii

1 Introduction ........................................... 1
   1.1 Why III-V MOSFETs ................................. 3
   1.2 Outline ........................................... 7

2 III-V MOSFETs Design .................................. 9
   2.1 Theory for Nanoscale FETs .......................... 9
       2.1.1 MOS Electrostatics ........................... 10
       2.1.2 I-V in Ballistic FETs ......................... 14
       2.1.3 Subthreshold and 2-D Electrostatics ............ 17
   2.2 Key Design Considerations ......................... 19
       2.2.1 Channel Effective Mass ....................... 19
       2.2.2 Channel Thickness ............................ 21
       2.2.3 S/D Series Resistance ......................... 26
       2.2.4 Estimated Performance ......................... 29
   2.3 Potential Device Structures ........................ 31
   2.4 Conclusions ....................................... 34

3 1st Generation: the Gate-Last Process and Composite-Channel .... 36
   3.1 Gate-First vs. Gate-Last Process .................. 36
   3.2 Device Structure and Fabrication .................. 41
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Results and Discussion</td>
<td>45</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusions</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>2nd Generation: MOCVD Regrowth and Surface Digital Etching</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>MBE vs. MOCVD regrowth</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Surface Digital Etching</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>Device Structure and Fabrication</td>
<td>62</td>
</tr>
<tr>
<td>4.4</td>
<td>Results and Discussion</td>
<td>65</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>3rd Generation: Vertical Spacers and Barrier Optimization</td>
<td>72</td>
</tr>
<tr>
<td>5.1</td>
<td>Vertical Spacer</td>
<td>74</td>
</tr>
<tr>
<td>5.2</td>
<td>Other Leakage Mechanisms</td>
<td>80</td>
</tr>
<tr>
<td>5.3</td>
<td>Device Structure and Fabrication</td>
<td>83</td>
</tr>
<tr>
<td>5.4</td>
<td>Results and Discussion</td>
<td>86</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusions</td>
<td>92</td>
</tr>
<tr>
<td>6</td>
<td>4th Generation: Highly Scaled MOSFETs with ZrO&lt;sub&gt;2&lt;/sub&gt; High-k Dielectric</td>
<td>94</td>
</tr>
<tr>
<td>6.1</td>
<td>ZrO&lt;sub&gt;2&lt;/sub&gt; High-k Dielectric</td>
<td>96</td>
</tr>
<tr>
<td>6.2</td>
<td>Device Structure and Fabrication</td>
<td>97</td>
</tr>
<tr>
<td>6.3</td>
<td>Results and Discussion</td>
<td>100</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparison Study</td>
<td>110</td>
</tr>
<tr>
<td>6.4.1</td>
<td>5.0 nm- vs. 2.5 nm-thick InAs channel</td>
<td>110</td>
</tr>
<tr>
<td>6.4.2</td>
<td>HfO&lt;sub&gt;2&lt;/sub&gt; vs. ZrO&lt;sub&gt;2&lt;/sub&gt; high-k dielectric</td>
<td>117</td>
</tr>
<tr>
<td>6.4.3</td>
<td>In&lt;sub&gt;0.53&lt;/sub&gt;Ga&lt;sub&gt;0.47&lt;/sub&gt;As vs. InAs channel</td>
<td>121</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusions</td>
<td>124</td>
</tr>
<tr>
<td>7</td>
<td>Device Analysis: Ballisticity</td>
<td>125</td>
</tr>
<tr>
<td>7.1</td>
<td>Gate Capacitance</td>
<td>127</td>
</tr>
<tr>
<td>7.2</td>
<td>Subband Occupancy</td>
<td>129</td>
</tr>
<tr>
<td>7.3</td>
<td>Extrinsic $I_D$ Compared with Measured $I_D$</td>
<td>132</td>
</tr>
<tr>
<td>7.4</td>
<td>Current-gate Cut-off Frequency</td>
<td>137</td>
</tr>
<tr>
<td>7.5</td>
<td>conclusions</td>
<td>137</td>
</tr>
<tr>
<td>8</td>
<td>Conclusion and Future Works</td>
<td>138</td>
</tr>
<tr>
<td>8.1</td>
<td>Summary</td>
<td>138</td>
</tr>
<tr>
<td>8.2</td>
<td>Future Works</td>
<td>140</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Manufacturing Process Flow</td>
<td>140</td>
</tr>
<tr>
<td>8.2.2</td>
<td>III-V FinFET</td>
<td>141</td>
</tr>
</tbody>
</table>
8.2.3 mm-wave/THz device Application . . . . . . . . . . . . . . 143
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 145
# List of Figures

1.1 Trend of the increase in CPU power density in recent decades. 
1.2 Injection velocity of InGaAs-based FETs as a function of gate length compared with the state-of-the-art strained silicon MOSFETs. 
1.3 Key advantages offered by III-V MOSFETs. (a) Wide range of available alloys in III-V compound semiconductors. (b) Cross-sectional TEM image of a III-V MOSFETs with 2.7 nm-thick InAs channel. (c) Contact resistivities with respect to the active doping concentration for in-situ Mo on N-doped InAs. (d) Energy Bandgap alignments of silicon and InGaAs to high-\(k\) and cross-sectional TEM image of a dielectric stack in a silicon MOSFET.

2.1 Illustration of how electrons occupy at the top of the barrier under a high drain bias. The Fermi-levels at the source and drain are split by \(qV_{DS}\). Electrons from the source fill up the positive momentum states (blue), while electrons from the drain fill up the negative momentum states (red).
2.2 Simple equivalent circuit diagram of the gate capacitance overlaid with the energy band diagram under the gate toward the barrier.
2.3 Calculated injection velocity as a function of carrier density for various in-plane effective masses.
2.4 Equivalent circuit for gate capacitance including the effects of interfacial traps and drain capacitive coupling.
2.5 Drive current calculated using the ballistic FET model as a function of effective mass with various EOTs assuming 1 nm wave-function depth and 0.5 V \(V_{DD}\).
2.6 (a) The in-plane effective mass as a function of the well width. (b) The energy band diagrams with the quantized bandgaps of 4.5 nm/3 nm and 6.0 nm/3 nm InAs/In\(_{0.53}\)Ga\(_{0.47}\)As channel MOSFETs, which were simulated by a 1-D Poisson-Schrödinger solver.
2.7 (a) Drive current calculated using the ballistic FET model as a function of EOT with various wave-function depths assuming $m^*/m_0 = 0.07$ and $V_{DD} = 0.5$ V. (b) Comparison of wave-functions depths between 2.5 nm- and 5.0 nm-thick InAs channel MOSFETs overlaid on simulated conduction band energy diagrams.

2.8 Simulated eigenstates and wave-function for the first eigenstate of a 3 nm-thick In$_{0.53}$Ga$_{0.47}$As quantum well with an InP barrier.

2.9 Calculated S/D series resistance as a function of metal contact size for various specific contact resistivites.

2.10 Calculated (a) subthreshold ($\log I_D - V_{GS}$), (b) transfer ($I_D - V_{GS}$), and (c) output ($I_D - V_{DS}$) characteristics using the ballistic FET model under the assumptions that EOT = 0.55 nm, $R_{series} = 80 \Omega$-$\mu$m and $m^*/m_0 = 0.05$.

2.11 Illustration and cross-sectional TEM image of inversion-mode III-V MOSFETs with ion-implanted S/D.

2.12 Illustration and cross-sectional TEM image of MOSHEMTs.

2.13 Illustration and cross-sectional TEM image of trench-etched III-V MOSFETs.

2.14 Illustration of III-V MOSFETs with regrown S/D.

3.1 Process flow schematics of the gate-first process.

3.2 (a) Cross-sectional TEM image of a FET fabricated based on the gate-first process. The red dotted circles indicate ungated regions due to the sidewall spacers. (b) Output and (c) transfer characteristics for a 60 nm-$L_g$ gate-first FET.

3.3 C-V characteristics of MOSCAPs with different in-situ passivation methods.

3.4 Process flow schematics of the 1st generation gate-last MOSFETs with regrown S/D.

3.5 Key specifications for the 1st generation gate-last process.

3.6 (a) Output ($I_D - V_{DS}$) and (b) transfer ($I_D - V_{GS}$ and $g_m$) characteristics for a 2 nm In$_{0.53}$Ga$_{0.47}$As / 2.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As composite-channel device with 40 nm-$L_g$; (c) output and (d) transfer characteristics for a 2 nm In$_{0.53}$Ga$_{0.47}$As / 3.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As composite-channel device with 70 nm-$L_g$; and (e) output and (f) transfer characteristics for a 10 nm In$_{0.53}$Ga$_{0.47}$As-only channel device with 80 nm-$L_g$.

3.7 Cross-sectional STEM image of the source-channel junction of the InGaAs channel device. The insets show Ga, In, and As spatial distributions derived from an EDX analysis.
3.8  $\log(I_D)-V_{GS}$ plots of the composite channel devices with (a) 2.5 nm-, (b) 3.5 nm-, and (c) 4.5 nm-thick InAs.  

3.9  (a) Dependence of critical thickness on InAs growth temperature, where As/In flux ratio is 20. (b) Dependence of critical thickness on As/In flux ratio, where growth temperature is 400 °C.  

3.10  Off-state leakage behaviors of FET with a 4.5 nm-thick InAs layer grown at (a) 400 °C and (b) 450 °C.  

3.11  (a) Top-down SEM image and (b) cross-section schematic of the TLM structure. (c) Measured TLM of contacts to the regrown N+ InAs layer.  

3.12  On-resistance with respect to the gate length at $V_{GS}-V_T = 0.65$ V.  

4.1  Cross-sectional STEM image of a FET with S/D regions regrown by MBE. The junction of InGaAs channel/regrown N+ InAs is magnified on the right-hand side. A significant amount of defects at the regrowth interface and stacking faults near the junction are observed. Image courtesy of Stephan Krämer from UCSB Materials Department.  

4.2  (a) Process flow schematics of photo-resist(PR) planarization. (b) Top-down SEM image of a dummy gate post MBE S/D regrowth and subsequent PR planarization. The InAs debris still remains on the sidewalls. (c) SEM image of a FET active region with MBE regrown S/D post metallization. InAs particles are sitting on the channel. (d) SEM image of a dummy gate post MOCVD S/D regrowth. (e) SEM image of a FET active region with MOCVD regrown S/D post gate metallization.  

4.3  $C-V$ characteristics of MOSCAPs (a) without any exposure to capping/annealing cycle (control) and (b) with SiO$_2$ capping and 500 °C annealing in furnace.  

4.4  AFM surface and depth profiles of a test structure after 4 cycles of digital etching (15-minute UV ozone exposure + 1-minute dilute HCl (1:10 HCl:H$_2$O) dip). $\sim$5.2 nm of In$_{0.53}$Ga$_{0.47}$As is etched.  

4.5  Comparison of FET $I-V$ performances with and without surface digital etching (D.E.). (a) $I_D$- and $g_m$-$V_{GS}$ for a 75 nm-$L_g$ FET without D.E., (b) $\log(I_D)$-$V_{GS}$ for a 500 nm-$L_g$ FET without D.E., (c) $I_D$- and $g_m$-$V_{GS}$ for a 75 nm-$L_g$ FET with D.E., and (d) $\log(I_D)$-$V_{GS}$ for a 500 nm-$L_g$ FET with D.E.  

4.6  Process flow schematics for the 2nd generation devices.  

4.7  Key specifications for the 2nd generation devices.
4.8 Cross-sectional STEM of a 40 nm-\(L_g\) device. \(\sim3.6\) nm HfO\(_2\) and \(\sim0.5\) nm interfacial layer are present on a 5/3 nm InAs/In\(_{0.53}\)Ga\(_{0.47}\)As channel. Image courtesy of Stephan Krämer at UCSB Materials Department. .......................................................... 65

4.9 Energy band structure and charge distribution at \(V_{GS} = 0.5\) V by 1-D Poisson-Schrödinger simulation. .......................................................... 66

4.10 (a) Transfer characteristics of short channel devices (40, 70, and 90 nm-\(L_g\)), (b) subthreshold characteristics of short channel devices (40, 70, and 90 nm-\(L_g\)), (c) output characteristics of a 40 nm-\(L_g\) device, and (d) gate leakage current normalized by total gate overlap for a 40 nm-\(L_g\) device. .......................................................... 67

4.11 Subthreshold characteristics of a 510 nm-\(L_g\) channel device. The minimum \(SS\) is \(\sim93\) mV/decade at \(V_{DS} = 0.05\) V. .......................... 68

4.12 Subthreshold swing (SS) and threshold voltage obtained by linear extrapolation as a function of the gate length for \(V_{DS} = 0.05\) V and 0.5 V. .......................................................... 69

4.13 (a) TLM measurement for N+ regrown S/D. (b) On-resistance \((R_{on})\) as a function of gate length. \(R_{on}\) extrapolated at zero-\(L_g\) is 188 \(\Omega\cdot\mu\)m. .......................................................... 70

5.1 Comparison between a regrown S/D MOSFET and MOSHEMT. Cross-sectional TEM images of (a) regrown S/D MOSFETs and (b) MOSHEMTs. Device structures and energy band diagrams of (c) a regrown S/D MOSFET and (d) MOSHEMT. Subthreshold characteristics of (e) a regrown S/D MOSFET and (f) MOSHEMT. ............................ 74

5.2 (a) Band diagram showing how the accumulated holes may increase the off-leakage. (b) Forward-sweep \(\log(I_D)-V_{GS}\) from different starting \(V_{GS}\) of -0.5 V, -0.3 V and -0.1 V up to 0.5 V. (b) Reverse-sweep \(\log(I_D)-V_{GS}\) from \(V_{GS}\) of 0.5 V toward the ending \(V_{GS}\) of -0.5 V, -0.3 V, and -0.1 V. .......................................................... 77

5.3 Illustrations of device structures and corresponding energy band diagrams of raised S/D MOSFETs (a) without and (b) with a vertical spacer between N+ S/D and the channel. ............................... 79

5.4 (a) Subthreshold characteristics of a MOSFET suffering from high buffer leakage. (b) Result of secondary ion mass spectrometry (SIMS) for a wafer with epitaxial layers grown at UCSB showing a high Si peak. (c) Energy band diagram corresponding to the wafer in (b). .......................................................... 81
5.5 (a) Subthreshold characteristic of a MOSFET with a barrier with a P-doping at the back. (b) Energy band diagram corresponding to the MOSFET in shown (a). 82
5.6 Process flow schematics for the 3rd generation devices. 83
5.7 Key specifications for the 3rd generation devices. 86
5.8 (a) Transfer and (b) output characteristics of a FET with a 2 nm-thick vertical spacer and 18 nm-\(L_g\). 87
5.9 Transfer characteristics of a 40 nm-\(L_g\) FET with a 12 nm-thick spacer in (a) linear and (b) semi-log scales. (c) Common-source characteristic for a 40 nm-\(L_g\) device with a 12 nm-thick spacer. (d) Subthreshold and gate leakage characteristics of a 1 \(\mu\)m-\(L_g\) FET with a 12 nm-thick vertical spacer. 88
5.10 (a) \(SS_{min}\) as a function of \(L_g\) at \(V_{DS} = 0.1\) and 0.5 V for 2 nm-, 7 nm- and 12 nm-thick spacer devices. (b) DIBL as a function of \(L_g\) for 2 nm-, 7 nm- and 12 nm-thick spacer devices. 89
5.11 log(\(I_D\))-\(V_{GS}\) of 2 nm-, 7 nm- and 12 nm-thick spacer devices with \(\sim 1 \mu m-L_g\). 90
5.12 (a) Peak \(g_m\) as a function of \(L_g\) at \(V_{DS} = 0.5\) V, and (b) \(R_{on}\) at \(V_{GS} = 1.0\) V for or 2 nm-, 7 nm- and 12 nm-thick spacer devices. 91
5.13 Peak \(g_m\) with respect to \(SS_{min}\) at \(V_{DS} = 0.5\) V and (b) \(I_{on}\) with respect to \(L_g\) at \(I_{off} = 100\) nA/\(\mu m\) and \(V_{DD} = 0.5\) V, comparing with recently reported III-V MOSFETs. 92
6.1 Comparison of electrical characteristics of In\(_{0.53}\)Ga\(_{0.47}\)As MOSCAPs with \(\sim 4\) nm HfO\(_2\) and \(\sim 4\) ZrO\(_2\). \(C-V\) characteristics as a function of frequency for (a) HfO\(_2\) and (b) ZrO\(_2\). The insets in (a) and (b) show \(C-V\) curve at 1 MHz. Normalized parallel conductance maps for (c) HfO\(_2\) and (d) ZrO\(_2\). 96
6.2 Cross-section schematic of MOSFETs in the 4th generation. 98
6.3 Key specifications for the 4th generation devices. 99
6.4 The cross-sectional STEM of the fabricated device. \(\sim 1.0\) nm Al\(_2\)O\(_x\)N\(_y\) interfacial layer, \(\sim 2.5\) nm ZrO\(_2\) and \(\sim 2.7\) nm InAs channel are shown. Heavy elements look brighter. Image courtesy of Stephan Krämer at UCSB. 100
6.5 log(\(I_D\)) and log(\(I_G\)) as a function of \(V_{GS}\) for 1\(\mu\)m-\(L_g\) FET at \(V_{DS} = 0.1\) V and 0.5 V. 101
6.6 (a) Transfer characteristics (\(I_D\)-\(V_{GS}\) and \(g_m\)-\(V_{GS}\)) as a function of \(V_{GS}\), (b) subthreshold characteristic (log(\(I_D\))-\(V_{GS}\)), and (c) common-source characteristic (\(I_D\)-\(V_{DS}\)) of a FET with 25 nm-\(L_g\). 102
6.7 Illustration of which resistance components contribute to $R_{on}$ at zero-$L_g$. (b) $R_{on}$ as function of $L_g$ at $V_{GS} = 0.7$ V. (c) TLM measurement of N+ regrown S/D. (d) Calculated ballistic contact resistance as a function of $V_{GS}$ at $m^*/m_0$ range from 0.04 to 0.1.

6.8 Summary of the factors contributing to the total S/D parasitic series resistance.

6.9 (a) $SS_{min}$ as a function of $L_g$ and (b) DIBL as a function of $L_g$ compared to published III-V MOSFETs.

6.10 (a) Peak $g_m$ as a function of $L_g$ (b) $I_{on}$ as a function of $L_g$ at $V_{DD} = 0.5$ V and $I_{off} = 100nA/\mu m$, compared to other III-V results in literatures.

6.11 (a) I-V characteristics for (a) Intel 22 nm-node Si FinFETs and (b) IBM 22 nm-node Si nanowire FETs.

6.12 Comparison of I-V characteristics for $\sim 30$ nm-$L_g$ FETs with 5.0 nm- and 2.5 nm-thick InAs channel.

6.13 $SS_{min}$ as a function of $L_g$ for 2.5 nm- and 5.0 nm-thick InAs channel devices at $V_{DS} = 0.1$ and 0.5 V.

6.14 Comparisons on subthreshold characteristics and gate leakage for 2.5 nm- and 5.0 nm-thick InAs channel devices with $L_g = 500$ nm at $V_{DS} = 0.1$ and 0.5 V.

6.15 Peak $g_m$ as a function of $L_g$ for 2.5 nm- and 5.0 nm-thick InAs channel devices at $V_{DS} = 0.5$ V.

6.16 Low field effective mobility as a function of the carrier density for 2.5 nm- and 5.0 nm-thick InAs channel devices.

6.17 (a) Gate-channel capacitance and carrier density for 2.5 nm- and 5.0 nm-thick InAs channel devices from split-C-V measurements. (b) Comparison on 1$st$ eigenstate wave-functions of 2.5 nm- and 5.0 nm-thick InAs channel devices.

6.18 Comparison of I-V characteristics for $\sim 25$-$28$ nm-$L_g$ FETs with HfO$_2$ and ZrO$_2$ high-$k$ dielectrics.

6.19 (a) Peak $g_m$ as a function of $L_g$ for devices with $\sim 3$ nm HfO$_2$ and ZrO$_2$ high-$k$ dielectrics at $V_{DS} = 0.5$ V (b) $SS_{min}$ as a function of $L_g$ devices with $\sim 3$ nm HfO$_2$ and ZrO$_2$ high-$k$ dielectrics at $V_{DS} = 0.1$ (open) and 0.5 V (solid).

6.20 (a) Gate-channel capacitance, carrier density from split-C-V measurements and (b) normalized gate leakage for devices with $\sim 3$ nm HfO$_2$ and ZrO$_2$ high-$k$ dielectrics.

6.21 Comparison of I-V characteristics for $\sim 40$ nm-$L_g$ FETs with $\sim 3$ nm-thick In$_{0.53}$Ga$_{0.47}$As and InAs. Measurement data courtesy of Cheng-Ying Huang at UCSB.
6.22 (a) Peak $g_m$ as a function of $L_g$ at $V_{DS} = 0.5 \text{ V}$, and (b) $SS_{\text{min}}$ as a function of $L_g$ at $V_{DS} = 0.1$ (open) and 0.5 V (solid) for $\sim$40 nm-$L_g$ FETs with $\sim$3 nm-thick In$_{0.53}$Ga$_{0.47}$As and InAs. Measurement data courtesy of Cheng-Ying Huang at UCSB.

7.1 Drive current at $V_{GS} - V_T = 0.4 \text{ V}$ as a function of $1/L_g$, overlaid with the conventional mobility-limited long channel FET model (solid red line).

7.2 Calculated density of states capacitance as a function of $V_{GS}$ under equilibrium with $m^*/m_0$ ranging from 0.04 to 0.1.

7.3 (a) Calculated total gate capacitance as a function of $V_{GS}$ under equilibrium at $m^*/m_0$ ranging from 0.04 to 0.1. (b) Measured gate-channel capacitance (solid line) and carrier density (solid line + circular symbols) for a 2.5 nm-thick InAs channel FET.

7.4 Calculated $E_f-E_1$ as a function of $V_{GS}$ for $m^*/m_0$ ranging from 0.04 to 0.1 under the (a) equilibrium ($V_{DS} = 0 \text{ V}$) and (b) saturation ($V_{DS} = 0.5 \text{ V}$) conditions.

7.5 Computed eigenstates and the 1st state wave-function for a 2.5 nm-thick InAs channel FET using a 1-D Poisson-Schödinger solver.

7.6 Circuit diagram for an extrinsic FET including the S/D series resistances. The orange box represents the intrinsic FET.

7.7 Calculated extrinsic $I_D$ as a function of $V_{GS}$ in (a) semi-log and (b) linear scales at $V_{DS} = 0.5 \text{ V}$ and $m^*/m_0 = 0.07 \text{ V}$.

7.8 Comparison of (a) calculated and (b) measured $I_D$ at the same overdrive of 0.4 V.

7.9 Calculated extrinsic $I_D$ at $V_{DD} = 0.5 \text{ V}$ in terms of the source series resistance at the range of $m^*/m_0$ from 0.04 to 0.1.

8.1 Progression of device performance (peak $g_m$ and $SS_{\text{min}}$) through technology developments made in this dissertation.

8.2 Modified process flow to accommodate sidewall spacers as well as self-aligned metal-alloy contacts in consideration of manufacturing.

8.3 (a) Bird’s-eye view of schematic and (b) cross-sectional TEM of III-V FinFET with atomic layer epitaxy.

8.4 Calculated current-gain cutoff frequency as a function of $I_D$ at $m^*/m_0 = 0.7$ and $V_{DS} = 0.5 \text{ V}$ for 0.3 and 1 fF/µm parasitic capacitance.
Chapter 1

Introduction

The constant goal of VLSI logic technology has been to integrate as many switching devices with higher speed and lower power consumption as possible within a given chip area. In order to reduce power dissipation of each transistor in a microprocessor, the supply voltage of transistors has been scaled by a similar proportion as the transistor dimensions. The scaling trend shown over the past few decades has so far satisfied the Moore’s law [1]. However, since the number of transistors per chip has dramatically increased, the power consumption per unit area has continued to rise in spite of the supply voltage scaling. According to the microprocessor power density trend shown in Figure 1.1 [2], CPU’s power dissipation has had reached almost 100 W/cm² around the year of 2002, which is an order of magnitude higher than that of a typical hot plate. As the device scaling is limited by the power constraints, further increase in the packing density requires an aggressive scaling in supply voltage, thus resulting in no enhancement in switching speed.
To obtain a higher drive current under a given supply voltage, the threshold voltage needs to be lowered. However, unless the subthreshold slope becomes steeper, this will increase the off-state leakage, causing the static power consumption to be comparable to the dynamic power consumption. Alternatively, improving the carrier transport properties will boost the current without compromising the off-leakage. Technological breakthroughs such as implementation of high-$k$/metal gate and FinFET architectures have been made in the last decade [3–5]. Significant improvements in subthreshold performance driven from the aforementioned innovations have enabled achieving an excellent drive current with a moderately low supply voltage, featuring >1 mA/µm at 0.7-0.8 V $V_{DD}$ for 22 nm-node silicon FinFETs [4].

Future device scaling will require the supply voltage scaling below 0.7 V. At
such a small $V_{DD}$, transistors would operate near the threshold voltage, thus resulting in a lower drive current if the transconductance is not high enough. In this context, introduction of a new material, which yields a high transconductance due to its better carrier transport properties than silicon, would help maintain a high drive current even under aggressively scaled supply voltage to around 0.5 V.

1.1 Why III-V MOSFETs

![Figure 1.2: Injection velocity of InGaAs-based FETs as a function of gate length compared with the state-of-the-art strained silicon MOSFETs. [6]](image)

The InGaAs-based III-V compound semiconductor is one of the most potential candidates for the new channel material. The small effective mass in this material provides excellent electron transport properties. As opposed to other novel 2-D materials such as MoS$_2$ and graphene, III-Vs have already been widely
implemented in integrated circuits for RF communications and defense applications [7, 8]. This proves they have the manufacturing capability and no critical issues with reliability. Figure 1.2 compares the injection velocity of InGaAs-based III-V FETs with that of state-of-the-art strained Si MOSFETs [6]. The InGaAs-based III-V FETs have at least a $2 \times$ higher injection velocity than the strained Si FETs even at a half the $V_{DD}$. This strongly indicates that III-V FETs enable scaling the supply voltage without introducing any penalty in performance. Furthermore, the reduction of scattering probability originated from the small density of states as well as the large inter-valley separation in III-V FETs allows operation near the ballistic regime even with relatively long channel lengths.

Aside from the excellent carrier transport properties, III-V FETs have many other advantages over Si MOSFETs. First of all, the III-V material offers a variety in combinations of elements (i.e. alloys) as well as in their compositions as shown in Figure 1.3(a) [9], thus allowing to tune important electrical properties (e.g. bandgap, effective mass, degree of strain, and band offsets) based on target applications. Secondly, its mature epitaxial growth techniques allow controlling the layer thickness with single-atom precision. For instance, as shown in Figure 1.3(b), the channel thickness of a III-V MOSFET (equivalent to that of the silicon body in ultra-thin-body SOI devices) is only 2.7 nm. Thirdly, a very small specific contact resistivity ($\rho_c < 10^{-8} \Omega\cdot cm^2$) is achievable in III-V FETs. Since $\rho_c$ exponentially increases with respect to $\sqrt{m^*}$, the small effective mass in III-V material reduces the contact resistance arising from very small metal contact vias. Figure 1.3(c) shows experimental contact resistivities of a heavily doped InAs versus active doping concentrations overlaid with theoretical limits [10]. The
measured contact resistivities for N-type InAs lie within ~4:1 of the calculated resistivities, exhibiting as low as $6 \times 10^{-9} \, \Omega \cdot \text{cm}^2$ at a $10^{20} \, \text{cm}^{-3}$ electron concentration [11]. Finally, InGaAs-based III-V materials have a larger band offset to high-$k$ dielectrics than silicon as shown in Figure 1.3(d). Moreover, unlike Si MOSFETs, insertion of an interfacial layer with a low dielectric constant such as $\text{SiO}_x\text{N}_y$ is not necessary to implement a high-$k$ dielectric [12]. Therefore, dielectric scaling can be done more aggressively in III-V MOSFETs.
Figure 1.3: Other advantages of III-V MOSFETs. (a) Wide range of available alloys in III-V compound semiconductors. (b) Cross-sectional TEM image of a III-V MOSFETs with 2.7 nm-thick InAs channel. (c) contact resistivities with respect to the active doping concentration for in-situ Mo on N-doped InAs [10]. (d) Energy Bandgap alignments of silicon and InGaAs to high-\(k\) and cross-sectional TEM image of a dielectric stack in a silicon MOSFET. [13, 14]
1.2 Outline

Chapter 2 (III-V MOSFETs Design) investigates the theory of ballistic FETs, which is essential to understanding the nanometer-scaled MOSFETs. It discusses the critical design considerations for III-V MOSFETs. Based on the ballistic theory calculation performed with realistic design parameters, $I-V$ characteristic of an ideal III-V FET is estimated. This is used as a reference for the experimental FETs. In addition, prior to discussing the device results, the strengths and weaknesses of various device prototypes are evaluated.

The experimental results on III-V FETs consist of four generations (Chapters 3 through 6), which are identified according to the distinct key technology developments.

Chapter 3 (1st generation) discusses the initially performed baseline experiments. Unlike the conventional gate-first process that has been adopted in III-V MOSFETs previously studied at UCSB, the gate-last (i.e. substitutional-gate) process is used in this study. Hence, the benefits offered by the gate-last process are introduced. Moreover, discussions on the fabrication process and device results are followed. This chapter also examines preliminary device results of FETs with InAs/InGaAs composite-channel in comparison to $\text{In}_{0.53}\text{Ga}_{0.47}	ext{As}$-only channel.

Chapter 4 (2nd generation) investigates critical issues in process modules and integration that have caused poor device performances in the 1st generation devices. Key developments introduced in this chapter include MOCVD S/D regrowth method, which has enabled the FET process to be more reproducible and scalable, as well as the surface digital etching to control channel damage during the regrowth.
Chapter 5 (3rd generation) focuses on improving the off-state leakage from the device design point-of-view. Although in the 2nd generation, FETs with an excellent on-state performance are obtained, the off-state characteristics yet have room for improvement, exhibiting a much larger off-leakage than required by VLSI applications. Therefore, the key developments associated in this chapter include incorporation of the vertical spacers to reduce BTBT and impact ionization as well as P-doped barriers at the back to eliminate the buffer leakage.

Chapter 6 (4th generation) examines the most recent development in highly scaled (both horizontally and vertically) FETs and the implementation of ZrO$_2$ high-$k$ dielectrics. To further improve the off-leakage behavior and enhance the short channel effect immunity, an aggressive channel thinning is performed. Simultaneously, a new high-$k$ dielectric is adopted to further enhance the gate control. FETs in this generation show a record high on-current of 0.5 mA/$\mu$m at an off-current of 100 nA/$\mu$m, which is comparable to state-of-the-art silicon multi-gate MOSFETs. More importantly, this chapter also shows comparison studies on the important device metrics such as the channel thickness, high-$k$ dielectrics, and the composition of the channel material.

Chapter 7 (Ballisticity) discusses the ballisticity of the fabricated FETs. Based on the design parameters corresponding to the measured FETs discussed in Chapter 6, $I$-$V$ characteristics of a ballistic FET are computed using the ballistic FET model discussed in Chapter 2.1 and compared to measured FETs.

Chapter 8 (Conclusions) concludes the dissertation by summarizing the achievements and key device results obtained from this work. Also, the outlook for future performance improvements and process flows for manufacturing is discussed.
Chapter 2

III-V MOSFETs Design

2.1 Theory for Nanoscale FETs

Today, the target FET channel length for VLSI has nearly reached the sub-15 nm regime. In such a short channel length regime, the traditional diffusive transport model no longer reflects the real device performance. The drift-diffusion model assumes that the potential energy of carriers decreases along with the channel potential \([15, 16]\). It is valid if scattering events happen frequently while the carriers move from the source to drain along the channel, resulting in a decreased potential energy of the carriers. However, for FETs with extremely short channel lengths, the average scattering distance (\textit{i.e.} mean-free-path, \(\lambda\)) becomes much longer than the channel length itself. Hence, the carriers are swept toward the drain without losing their potential energy, while the high energy carriers relax upon their arrival at the highly doped drain region. It is noted that the effective channel length directly associated with scattering events is shorter than the phys-
ical channel length of the FET [17]. This is because the back-scattered carrier flow, which negates the total drain current, only happens in a small region near the top of the barrier. Thus, the short channel FETs can be considered more analogous to vacuum tubes than classical FETs. In practice, modern FETs operate between the drift-diffusion and the ballistic regimes. For state-of-the-art Si nanowire FETs with an effective channel length of 20-25 nm, the experimental ballisticity reaches \(~0.8\) and \(0.6\) in NFETs and PFETs, respectively [18,19]. Due to the much lower effective mass in III-V, it is expected that III-V FETs would likely exhibit a higher ballisticity at the same channel length compared to silicon FETs [20].

2.1.1 MOS Electrostatics

![Figure 2.1: Illustration of how electrons occupy at the top of the barrier under a high drain bias. The Fermi-levels at the source and drain are split by \(qV_{DS}\). Electrons from the source fill up the positive momentum states (blue), while electrons from the drain fill up the negative momentum states (red).](image_url)
Figure 2.1 shows the energy band (1st subband) diagram of a ballistic FET along the channel, from the source to drain, at a large applied drain voltage. Similar to a bipolar transistor with a short base, the current flow is controlled by the potential at the top of the barrier [21]. It is noted, however, that the barrier height is indirectly modulated by the gate potential in the case of FETs. The electron density is determined by the density of states as well as the position of the Fermi-level ($E_f$) relative to the 1st subband energy level at the top of the barrier ($E_1$) [22, 23]. Electrons populated in the positive momentum states (blue curve in Figure 2.1) contribute to the current from the source to drain ($I_{S\rightarrow D}$), and those in the negative momentum states (red curve in Figure 2.1) contribute to the current from the drain to source ($I_{D\rightarrow S}$) [22]. The total current is determined by the difference of the two current components described above. As the further applied drain bias drops the potential in the drain region by $qV_{DS}$, the number of electrons in the negative momentum states at the top of the barrier exponentially decreases, thus resulting in $I_{S\rightarrow D}$ dominating over $I_{D\rightarrow S}$. 
Figure 2.2: Simple equivalent circuit diagram of the gate capacitance overlaid with the energy band diagram under the gate toward the barrier.

Figure 2.2 illustrates the energy band diagram under the gate toward the buffer as well as a simplified equivalent circuit diagram of the gate capacitance in III-V FETs. The gate capacitance is modeled as a series combination of the oxide capacitance ($C_{OX}$), wave-function depth capacitance ($C_{Depth}$), and density of states capacitance ($C_{DoS}$) [24]. $C_{OX}$ is defined as

$$C_{OX} = \frac{\epsilon_{OX}}{T_{OX}},$$  \hspace{1cm} (2.1)$$

where $\epsilon_{OX}$ is the permittivity of the oxide and $T_{OX}$ is the oxide thickness. When the oxide is composed of two different layers, it can be simply expressed as two oxide capacitances in series. $C_{Depth}$ can be expressed as

$$C_{Depth} = \frac{\partial(-qN_s)}{\partial \left( \frac{E_1-E_c}{q} \right)},$$  \hspace{1cm} (2.2)$$

where $qN_s$ is the charge density. This capacitance represents the change in the total channel charge density with respect to the position of the subband energy
relative to the conduction band \([25]\). The \(E_1 - E_c\) depends on the change in surface potential \(\psi_s\) with changing gate potential. It can also be understood as a change in the position of the centroid of wave-function. For instance, a larger \(\psi_s\) at a higher \(V_{GS}\) will increase the gate-capacitive coupling, allowing for the wave-function to be closer to the surface. In general, in a thin quantum-well, it is located approximately at the center of the well. In turn, \(C_{\text{Depth}}\) can be express as

\[
C_{\text{Depth}} \simeq \frac{\epsilon_{\text{channel}}}{T_{\text{well}}/2},
\]

where \(T_{\text{well}}\) is the quantum-well thickness, and \(\epsilon_{\text{channel}}\) is the channel permittivity. \(C_{\text{DoS}}\) represents the change in the channel charge density with changing position of the Fermi-level \([25]\), thus is expressed as

\[
C_{\text{DoS}} = \frac{\partial(-qN_s)}{\partial \left(\frac{E_f - E_1}{q}\right)}.
\]

Therefore, it is manifested by the density of states (2-D density of states in the case of a FET channel) as well as the Fermi-Dirac statistics and can be re-written as

\[
C_{\text{DoS}} = \frac{g_v q^2 m^* \pi k^2}{1 + e^{\frac{E_f - E_1}{kT}}}.\]

Here, \(m^*\), \(k\), and \(g_v\) represent the in-plane effective mass in the quantum-well, the Boltzmann constant, and the valley degeneracy \((g_v = 1\) for InGaAs-based material\), respectively. In the case of high degeneracy, \((E_f - E_1 \gg \sim 3kT)\), \(C_{\text{DoS}}\) is nothing but the density of states itself: \(\frac{q^2 m^* \pi h^2}{1 + e^{\frac{E_f - E_1}{kT}}}\). The equation above is for the equilibrium state, where the drain voltage is zero or negligibly low. Therefore, both the positive and the negative momentum states are occupied. However, if
the drain voltage is applied, then the $C_{DoS}$ should be modified as

$$C_{DoS} = \frac{g_n q^2 m^*_{2\pi \hbar^2}}{1 + e^{-\frac{E_f - E_1}{kT}}} + \frac{g_n q^2 m^*_{2\pi \hbar^2}}{1 + e^{-\frac{E_f - E_1 - \phi_{DS}}{kT}}}.$$  \tag{2.6}

All of the capacitance components that determine the total gate-channel capacitance have been now identified. Assuming that $V_{GS}$, $E_f$, and $E_1$ are aligned under the flat band condition, $E_f - E_1$ can be expressed in terms of $V_{GS}$ as

$$\frac{(E_f - E_1)}{q} = \frac{C_{EET}}{C_{EET} + C_{DoS}} \cdot V_{GS},$$  \tag{2.7}

where the equivalent electrostatic capacitance, $C_{EET}$, consists of the oxide and the wave-function depth capacitances in series, expressed as $C_{EET} = \frac{C_{OX} \cdot C_{Depth}}{C_{OX} + C_{Depth}}$. Since $C_{DoS}$ in the nondegenerate regime is also a function of $E_f - E_1$, the $E_f - E_1$ term can be calculated for a full range of $V_{GS}$ using a numerical iteration method.

On the other hand, in a highly degenerate condition, $E_f - E_1$ is just linearly proportional to $V_{GS}$, since the $C_{DoS}$ term is a constant.

Also, the $E_f - E_1$ term can vary with $V_{DS}$ in ballistic FETs. Under the steady-state (i.e. at a fixed $V_{GS}$) and non-equilibrium (i.e. at a high $V_{DS}$) conditions, while the charge density at the top of the barrier is fixed, the total $C_{DoS}$, which includes both the positive and negative momentum states, decreases with increasing $V_{DS}$. Hence, $E_f - E_1$ must be raised to meet the space charge neutrality [26].

### 2.1.2 I-V in Ballistic FETs

The total current is determined by a competition between the two opposite electron flows that have originated from the source and drain contacts, which are denoted by $I_{S\rightarrow D}$ and $I_{D\rightarrow S}$. The $I_{S\rightarrow D}$ term is expressed as

$$I_{S\rightarrow D} = q \cdot N_{s+} \cdot v_{inj},$$  \tag{2.8}
where $N_{s+}$ is the number of electrons that occupy the positive momentum states. To find $N_{s+}$, the density of states capacitance is integrated over the energy in the positive half. To simplify the integral, the equation can be transformed using the Fermi-Dirac integral as

$$N_{s+} = \int_{E_1}^{\infty} \frac{q\hbar^2}{2\pi} \frac{dE_f}{1 + e^{-\frac{E_f-E_1}{kT}}}$$

(2.9)

$$= \frac{q\hbar^2}{2\pi} \cdot kT \cdot F_0 \left( \frac{E_f - E_1}{kT} \right).$$

(2.10)

Here, $F_0(\eta)$ is the Fermi-Dirac integral of an order of zero, which can be evaluated without a numerical calculation represented by $F_0(\eta) = \ln(1+e^\eta)$ [27]. In a highly degenerate condition ($E_f - E_1 \gg 3kT$), $N_{s+}$ is linearly proportional to $V_{GS}$ due to the equation 2.10. The last term, $v_{inj}$ is the so-called injection velocity, i.e., the average velocity at the top of the barrier [23]. It can be expressed as

$$v_{inj} = \sqrt{\frac{2kT}{\pi m^*}} \cdot \frac{F_{1/2}(\frac{E_f - E_1}{kT})}{\frac{E_f - E_1}{kT}}.$$  

(2.11)

Here, $\sqrt{\frac{2kT}{\pi m^*}}$ is the thermal velocity, and $F_{1/2}(\eta)$ is the Fermi-Dirac integral of an order of 1/2, which should be numerically evaluated [27]. It is expressed as

$$F_{1/2}(\eta) \equiv 2 \int_{0}^{\infty} \frac{e^{1/2} d\epsilon}{1 + e^{\epsilon-\eta}}.$$  

(2.12)

While $v_{inj}$ is a constant with the thermal velocity in the nondegenerate regime, it is proportional to $(V_{GS})^{1/2}$ in the degenerate regime. Figure 2.3 plots the injection velocity as a function of the carrier density for varied $m^*/m_0$ in InGaAs-based materials. It is confirmed that the thermal velocity increases with increasing in-plane effective mass. When the carriers become degenerate, $v_{inj}$ rapidly rises.
Figure 2.3: Calculated injection velocity as a function of carrier density for various in-plane effective masses.

Under saturation (i.e. high $V_{DS}$), the total current is only determined by $I_{S→D}$ due to the absence of $I_{D→S}$. For moderate $V_{DS}$, the total current is reduced by the magnitude of $I_{D→S}$. $I_{D→S}$ can be evaluated simply by substituting the Fermi-level term in the source ($E_f - E_1$) with the Fermi-level term in the drain ($E_f - E_1 - qV_{DS}$) in the abovementioned equations.
2.1.3 Subthreshold and 2-D Electrostatics

Based on the classical MOSFET theory, when an ideal FET with a fully depleted body (i.e. no change in depletion width) is operating in the subthreshold regime, any change in the gate potential is entirely transferred to altering the semiconductor surface potential. In this case, the subthreshold current should vary exponentially with the gate voltage as $I_D \propto e^{\frac{qV_{GS}}{kT}}$. In the case of real FETs, however, the interfacial traps and 2-D electrostatics, which respectively depend on the quality of the gate-stack and device geometry, may undesirably alter the surface potential. It is noted that the body effect does not exist in a well-designed quantum-well FET, since the substrate (or barrier) is fully depleted as in a SOI-based FET. Figure 2.4 illustrates a simple equivalent circuit that includes the effects from the 2-D electrostatics as well as the interfacial traps in a highly nondegenerate subthreshold regime. Since $C_{DoS}$ and $C_{Depth}$ are negligible
in low subthreshold, the surface potential $\psi_s$ can be expressed as

$$\psi_s = \frac{C_{OX}}{C_{\Sigma}} V_G + \frac{C_{gd}}{C_{\Sigma}} V_D + \frac{C_{Dit}}{C_{\Sigma}} V_S - \frac{q N_s}{C_{\Sigma}}. \quad (2.13)$$

Here, $C_{gd}$ is the gate-drain capacitance, which is closely associated with the channel length and device geometry. Also, $C_{Dit}$ is the interfacial trap capacitance, which is determined by the interface trap density as $q^2 D_{it}$. $C_{\Sigma}$ represents the sum of all the capacitances terms as $C_{\Sigma} = C_{OX} + C_{Dit} + C_{gd}$. Therefore, under the equilibrium ($V_{DS} \approx 0$) condition, $\psi_s$ varies with $V_{GS}$ by a factor of $\frac{C_{OX}}{C_{\Sigma}}$. Here, one could define $\frac{C_{OX}}{C_{\Sigma}}$ as $1/m$ or the body effect coefficient [16]. It can be now represented differently with the subthreshold slope term as

$$\frac{C_{OX}}{C_{\Sigma}} = \frac{2.3kT/q}{SS}. \quad (2.14)$$

From this equation, it is found that the subthreshold slope is governed not only by $D_{it}$ but also by $C_{gd}$, indicating that it will degrade in short channel devices. Next, the increase in the subthreshold current with changing applied drain voltage can be expressed as

$$\Delta \log I_D = \frac{\Delta V_{DS} C_{gd}}{2.3kT/q C_{\Sigma}}. \quad (2.15)$$

This equation suggests that if $C_{gd}$ starts to dominate over $C_{\Sigma}$, even a small change in $V_{DS}$ will significantly change $I_D$. Using the equation 2.14 and $SS = \frac{\Delta V_{GS}}{\log I_D}$, $C_{\Sigma}$ can be expressed in terms of the drain induced barrier lowering (DIBL) factor [16] as

$$\frac{C_{gd}}{C_{\Sigma}} = \frac{2.3kT DIBL}{q SS}. \quad (2.16)$$

In this subsection, the non-ideal effects from $C_{gd}$ and $C_{Dit}$ have been discussed only for the subthreshold characteristics, but in fact, these affect the on-state
performances as well. In a highly degenerate condition, if the border trap density associated with the quality of high-
$k$ dielectric is too large, the on-current and transconductance will substantially reduce due to the decreasing mobile carrier density [28]. In addition, a large gate-drain coupling, which is governed by the device aspect ratio, will increase the output conductance under saturation.

2.2 Key Design Considerations

2.2.1 Channel Effective Mass

The high $I_{on}$ in III-V nFETs arises mostly due to the high electron injection velocity from the small in-plane effective mass in III-V materials ($v_{inj} \propto \sqrt{1/m^*}$). Moreover, the small density of states in III-V FETs, which is also proportional to the in-plane effective mass ($N_{DoS} \propto m^*$), reduces the scattering probability and thus further increases $I_{on}$. However, this small density of states in III-V FETs may also cause a bottleneck effect on $I_D$ by limiting the available charge density ($N_s$) at the source-end of the channel [24, 29]. Therefore, the magnitude of $I_D$ is strongly dependent on the trade-off relationship between $v_{inj}$ and $N_s$. In this context, understanding this trade-off due to the small effective mass is extremely important to figure out the most optimum channel design, thus maximizing $I_D$.

Based on a ballistic FET model, the drive current ($I_D$) is calculated as a function of $m^*/m_0$ for EOT values of 0.4-1.0 nm at a supply voltage of 0.5 V as shown in Figure 2.5. EOT reflects solely the oxide thickness converted to SiO$_2$, while a 1 nm-thick wave-function depth is assumed for a 2 nm-thick quantum-well. Here, degeneracy ($g$) is the number of band minima contributing to the channel
transport. For [100] silicon channels, \( g \) is 2 [30,31], whereas it is 1 for conventional III-V channels. It is difficult to define a single deterministic value of effective mass in III-V FETs with a thin pseudomorphic quantum-well, because of the strong nonparabolicity effect as well as the bi-axial strain [32]. The effective mass of \( \text{In}_x\text{Ga}_{1-x}\text{As} \) based III-V FETs (blue box in Figure 2.5) can be determined within a \( m^*/m_0 \) range of 0.04-0.1, depending on the indium composition as well as the quantum-well thickness [32].

![Intrinsic \( I_D \) VS \( m^* \) varying EOT](image)

Figure 2.5: Drive current calculated using the ballistic FET model as a function of effective mass with various EOTs assuming 1 nm wave-function depth and 0.5 V \( V_{DD} \).

As calculated above, too small of an effective mass limits \( I_D \) due to the lack of the available charge density in the channel with a small density of states. On the other hand, too large of an effective mass causes a bottleneck effect on \( I_D \) due to
a slower electron transport. It is noted that the maximum $I_D$ in terms of $m^*/m_0$ varies with EOT. The optimum $m^*$ value increases with thinner EOT, since $C_{DoS}$ plays a dominant role in determining the total gate-channel capacitance when EOT is very thin.

In comparison to Si FETs with $m^* = 0.2m_0$, III-V FETs are expected to exhibit a $\sim10$-50% higher $I_D$. This result may be somewhat discouraging in consideration of the greater challenges associated with manufacturing III-V FETs. However, it should be noted that this calculation is based on the assumption of ballistic transport. In terms of ballisticity, III-V FETs possess a greater benefit over Si FETs as discussed previously. On the other hand, a large variation in $I_D$ at the given range of $m^*/m_0$ for III-V FETs implies that the engineering of the channel effective mass must be performed carefully with further consideration on any effects related to gate dielectrics as well as wave-function depths.

### 2.2.2 Channel Thickness

The thickness of quantum-well channels in III-V FETs plays a critical role in determining both the on- and off-state performances, since it is tightly associated with the short channel immunity, carrier confinement, and quantized bandgap.

The quantum-well thickness must be scaled as the gate length is shrunk in order to minimize the drain induced barrier lowering. In a well-designed III-V FET, the majority of electrons is confined within the quantum-well, and the barrier region is almost fully depleted. Therefore, III-V FETs with a hetero-barrier quantum-well structure can be considered analogous to fully depleted ultra-thin body SOI FETs. In SOI FETs, the minimum scalable channel length is dependent
on the silicon body thickness, gate dielectric thickness, and dielectric constants, as described in the equation below [33]:

\[ L_{\text{min}} \approx 4.5 \left( t_{Si} + \left( \frac{\epsilon_{Si}}{\epsilon_{ox}} \right) t_{ox} \right). \]  

(2.17)

Here, \( L_{\text{min}} \) represents the minimum scalable channel length, \( t_{Si} \) the silicon body thickness, \( t_{ox} \) the equivalent oxide thickness, \( \epsilon_{Si} \) the permittivity of silicon, and \( \epsilon_{ox} \) the permittivity of oxide. Although the constants may need to be modified, the overall trend should hold true for III-V FETs as well. Since \( \epsilon \) in In\(_{1-x}\)Ga\(_{x}\)As-based semiconductors is larger by \( \sim 25\% \) than in silicon, III-V FETs will likely suffer from increased short channel effects as the gate length is shrunk.

Figure 2.6: (a) The in-plane effective mass as a function of the well width [34]. (b) The energy band diagrams with the quantized bandgaps of 4.5 nm/3 nm (black solid line) and 6.0 nm/3 nm (red dotted line) InAs/In\(_{0.53}\)Ga\(_{0.47}\)As channel MOSFETs, which are simulated by a 1-D Poisson-Schrödinger solver (BANDPROF, heterojunction device simulator by W. R. Frensley).

Also, as the channel thickness is reduced, the transport (i.e. in-plane) and confinement (i.e. quantization) effective masses increase due to a high degree of
nonparabolicity of In$_{1-x}$Ga$_x$As material as well as the wave-function penetration into the barrier. Figure 2.6(a) shows in-plane effective mass of In$_{0.53}$Ga$_{0.47}$As/InP quantum-well as a function of quantum-well width [34]. When the quantum-well thickness is greater than 100 nm, the in-plane effective mass is close to that of a bulk. However, it rapidly increases when the quantum-well thickness becomes less than 100 nm, eventually reaching up to $\sim$0.066 in a 3 nm-thick quantum-well. Hence, the increased effective mass in thinner channels may substantially reduce the drive current as shown in Figure 2.5, which partly depends on EOT as well. This nonparabolic effect is expected to be more severe in high indium content channels ($x > 0.53$) [32]. Thinning of the quantum-well will simultaneously increase the quantized bandgap. For a simple model of an infinite quantum-well [35], the energy state, $E_n$, is defined by

$$E_n = \frac{\hbar^2}{2m^*_{\text{conf.}}} \left(\frac{n\pi}{d}\right)^2, \quad (2.18)$$

where $n$ represents the quantum number, $m^*_{\text{conf.}}$ the confinement effective mass, and $d$ the quantum-well thickness. A larger value of $E_n$ indicates that the first subband is more raised and has a larger separation with the second subband. As shown in the equation above, $E_n$ is inversely proportional to $d^2$ as well as to $m^*_{\text{conf.}}$. Figure 2.6(b) shows a decrease in the quantized bandgap by thinning the channel simulated by a 1-D Poisson-Schrödinger solver (BANDPROF heterojunction device simulator by W. R. Frensley). As an InAs/InGaAs composite-channel with an InAlAs barrier is thinned from 9 nm down to 7.5 nm, the quantized bandgap increases by 60 meV. Since a larger quantized bandgap is extremely beneficial to reducing the band-to-band tunneling and impact ionization at the off-state, FETs with thin channels are attractive for the low power VLSI operation. This will be
further discussed in Chapter 6.

Moreover, in a thinner channel, the centroid of electron wave-function is closer to the channel surface (i.e. the wave-function depth is smaller), thus resulting in an enhanced gate capacitive coupling. Figure 2.7(a) shows calculated ballistic $I_D$ as a function of EOT for various wave-function depths. As discussed in the previous section, since the total gate-channel capacitance ($C_{g-ch}$) is determined by $C_{DoS}$, $C_{OX}$, and $C_{depth}$ in series, the lowest capacitance component would dictate $C_{g-ch}$. Therefore, if the centroid of wave-function is too far from the surface, $C_{g-ch}$ will be limited by the small $C_{depth}$. In the calculation results shown in Figure 2.7, when the wave-function depth is 4 nm, the drive current is hardly affected by EOT scaling. EOT scaling, therefore, must be done along with channel thickness scaling. In general, the position of wave-function centroid is determined by the shape of the quantum-well, which varies with the gate potential. However, for a thinner quantum-well (< 5 nm-thick), it is located close to the center. Figure 2.7(b) depicts conduction band energy band diagrams and wave-functions for the 1st subband from the 1-D Poisson-Schrödinger solver for a 2.5 nm- and 5.0 nm-thick quantum-wells. In both cases, the wave-function centroids are placed near the center. Detailed discussion on this topic will be followed with experimental results in Chapter 6.
Figure 2.7: (a) Drive current calculated using the ballistic FET model as a function of EOT with various wave-function depths assuming \( m^*/m_0 = 0.07 \) and \( V_{DD} = 0.5 \, V \). (b) Comparison of wave-functions depths between 2.5 nm- and 5.0 nm-thick InAs channel MOSFETs overlaid on simulated conduction band energy diagrams.

The benefits offered by channel thickness scaling have been discussed so far. However, it is also important to recognize a few disadvantages associated with it. Firstly, as thinning a quantum-well, the eigenstates move up, resulting in a smaller effective conduction band offset to the barrier. As a result, the electron confinement inside the quantum-well is decreased; therefore, the gate controllability and average carrier injection velocity will reduce. Figure 2.8 shows the simulated eigenstates and wave-function of the 1\(^{st}\) eigenstate for a 3 nm-thick In\(_{0.53}\)Ga\(_{0.47}\)As quantum-well with an InP barrier. It is found that a substantial portion of 1\(^{st}\) wave-function penetrates into the barrier. In addition, the 2\(^{nd}\) eigenstate, which is separated from the 1\(^{st}\) eigenstate by \( \sim 50 \, \text{meV} \), is present in the barrier region. Secondly, the channel thickness scaling degrades the channel mobility due to the increased surface roughness scattering [36, 37]. For instance, 2.5 nm-thick InAs FETs only have an effective mobility of \( \sim 250 \, \text{cm}^2\text{V}^{-1}\text{s}^{-1} \), which is similar to that
of strained Si devices. Nonetheless, FETs beyond 7 nm-node would have a $\sim<15$ nm gate length; thus it is expected that the low effective mobility will not be so problematic as reaching toward the ballistic regime. This topic will be further discussed in Chapter 7.

![Simulated eigenstates and wave-function](image)

Figure 2.8: Simulated eigenstates and wave-function for the first eigenstate of a 3 nm-thick In$_{0.53}$Ga$_{0.47}$As quantum well with an InP barrier.

### 2.2.3 S/D Series Resistance

Another big challenge lies on minimizing the S/D series resistance. To obtain a high drive current, the on-resistance ($R_{on}$) of FETs must be low. As illustrated in the schematic shown in Figure 2.9, $R_{on}$ comprises of different resistive components:

$$ R_{on} = R_{series} + R_{channel} $$

$$ R_{series} = 2(R_{contact} + R_{sheet}), $$
where $R_{\text{series}}$ is the S/D series resistance, $R_{\text{contact}}$ the metal-semiconductor contact resistance, $R_{\text{sheet}}$ the resistance between the S/D contacts to the channel, and $R_{\text{channel}}$ the channel resistance including the ballistic contact resistance, which is $1/G_{DS}$ in the case of ballistic FETs. While intrinsic $I_D$ is only determined by $R_{\text{channel}}$, extrinsic $I_D$, or the actual current in real FETs, is governed by $R_{\text{on}}$.

According to the scaling scenario for III-V MOSFETs in the ITRS roadmap [38], its first implementation is expected to happen at the 15 nm-node, where the half pitch of the first metal vias is 15 nm, and the physical gate length is 14 nm. Consequently, the width of the first via is no larger than 15 nm. In addition, the requirement on $R_{\text{series}}$ in III-V FETs at the 15 nm-node is only 131 Ω-µm. In such a tight device pitch, $R_{\text{sheet}}$ is nearly negligible due to the small S/D-to-channel spacing in the order of only a few nanometers. On the other hand, $R_{\text{contact}}$ will be the dominant factor in total $R_{\text{series}}$. Using a simple distributed transmission line model (TLM), $R_{\text{contact}}$ can be expressed as [39]:

$$R_{\text{contact}} = \frac{R_{\text{sh}} L_{\text{tran.}}}{W} \coth \left( \frac{L_{\text{contact}}}{L_{\text{tran.}}} \right),$$

$$L_{\text{tran.}} = \sqrt{\frac{\rho_c}{R_{\text{sh}}}},$$

where $R_{\text{sh}}$ is the N+ S/D sheet resistance, $W$ the gate width, $\rho_c$ the specific contact resistivity, $L_{\text{contact}}$ the metal contact length, and $L_{\text{tran.}}$ the transfer length. At longer metal contact lengths, $R_{\text{contact}}$ has no large contribution to $R_{\text{series}}$. However, as $L_{\text{contact}}$ is scaled below $L_{\text{tran.}}$, $R_{\text{contact}}$ rapidly increases, making $\rho_c$ the sole dominant factor. Figure 2.9 plots $R_{\text{series}}$ as a function of metal contact width with various specific contact resistivities. Here, it is assumed that $R_{\text{sh}}$ is 50 Ω/square, which is readily achievable with 20 nm-thick N+ InAs layer, where the
active doping concentration is $\sim6\times10^{19}\text{cm}^{-3}$. As shown in the inset of the figure, $R_{\text{contact}}$ is 90:1 greater than $R_{\text{sheet}}$, which represents the resistance between the metal contact and channel edges. In order to meet $R_{\text{series}}$ of 131 $\Omega$-$\mu$m required for the 15 nm-node, $\rho_c$ is required to be lower than 1 $\Omega$-$\mu$m$^2$. Achieving such a low $\rho_c$ is non-trivial, thus it can trigger a bottleneck effect for ultra-high scaled integration. Ashish Baraskar, a former PhD student in Rodwell group at UCSB, had extensively studied this topic and demonstrated a record low $\rho_c$ on N+ InAs of $\sim0.6$ ($\pm$ 0.4) $\Omega$-$\mu$m$^2$ by implementing an in-situ molybdenum [11].

Figure 2.9: Calculated S/D series resistance as a function of metal contact size for various specific contact resistivites.
2.2.4 Estimated Performance

This section discusses the expected DC performance of III-V FETs based on the results obtained from ballistic FET modeling (See Chapter 2.1) by applying design parameters that can be comfortably realizable with currently available technology. Given the target application of high performance logic CMOS circuits, the supply voltage is set at 0.5 V, and the off-leakage must be below 100 nA/µm \(^{[38]}\). The calculation is performed using the Fermi-Dirac statistics at room temperature under several important assumptions. Firstly, it is assumed that the FET is working in the fully ballistic regime, thus there should be no back-scattering happening at the virtual source (\(i.e.\) the top of the barrier). This is a realistic assumption, considering the long mean free path in III-V materials \(^{[40]}\) as well as the target gate length of \(<\sim 20\) nm, which will be further discussed near the end of this dissertation. Secondly, it is assumed that the FETs have a 2 nm-thick quantum-well and in-plane effective mass \((m^*/m_0)\) of 0.05, which is chosen at the maximum \(I_D\) based on the simulation result shown in Figure 2.5. In this case, the wave-function centroid is expected to be positioned at the center of the well and that it does not change as a function of the gate potential. Thirdly, 2-D electrostatics and nonparabolicity are neglected for simplicity. Fourthly, the high-\(k\) dielectric is assumed to be consisting of 0.5 nm-thick \(\text{Al}_2\text{O}_3\) \((\epsilon_r = 9)\) and 2.0 nm-thick \(\text{ZrO}_2\) \((\epsilon_r = 23)\), resulting in EOT of 0.55 nm. These thickness values are determined based on the maximum gate leakage current with a tolerance of 1 A/cm\(^2\). Also, its interfacial trap density \((D_{it})\) is assumed to be less than \(10^{12}\) cm\(^{-2}\)eV\(^{-1}\). It is noted that such \(D_{it}\) is easily achieved, as shown in the recent experimental MOSCAP results obtained at UCSB \(^{[41,42]}\). Lastly, \(R_{series}\) is assumed
to be 80 Ω-µm from $\rho_c = 0.5 \Omega\mu$m$^2$ at a contact width of 15 nm, as estimated in Figure 2.9.

Figure 2.10: Calculated (a) subthreshold ($\log I_D-V_{GS}$), (b) transfer ($I_D-V_{GS}$), and (c) output ($I_D-V_{DS}$) characteristics using the ballistic FET model under the assumptions that EOT = 0.55 nm, $R_{series} = 80 \Omega$-µm and $m^*/m_0 = 0.05$.

Figure 2.10 shows $I-V$ characteristics obtained from the calculations. $V_T$ is negative, since the flat band condition in this calculation is defined as a condition in which the Fermi-level is aligned with the 1$^{st}$ eigenstate energy. In order to read the value of $V_{GS}$ at the on-state for a 0.5 V supply voltage, $\log(I_D)-V_{GS}$ is plotted in Figure 2.10(a). At $I_{off} = 100$ nA/µm and $V_{DD} = 0.5$ V, $V_{GS}$ for the on-state is
0.32 V. Also, it shows an ideal subthreshold slope of $\sim 60$ mV/decade as expected. Figure 2.10(a) and (b) show the calculated transfer and output characteristics, respectively. The FET calculation result exhibits a promising on-current of $\sim 1.1$ mA/$\mu$m at $V_{DD} = 0.5$ V, which is at least two times larger than that of the state-of-the-art Si-based multi-gate FETs reported in [43,44].

### 2.3 Potential Device Structures

To achieve the target device performance and simultaneously meet the device design rules for high level integration, it is very critical to select the right device architecture from multiple prototypes. Hence, this section will discuss merits and demerits of each prototype and select a device structure that will be the most appropriate for III-V based VLSI.

![Self-Aligned NiGeSi Contact](image1.png)

![Self-Aligned NiGeSi Contact](image2.png)

Figure 2.11: Illustration and cross-sectional TEM image of inversion-mode III-V MOSFETs with ion-implanted S/D [45].

Firstly, as shown in Figure 2.11, inversion-mode MOSFETs by S/D ion implantation are realizable based on the III-V material [45–47]. This structure is
identical to conventional planar Si-MOSFETs on a bulk silicon wafer, thus enabling a fully self-aligned S/D process in the nanometer dimensions. However, while the inverted channel design in this structure is relatively straightforward, it is much more inferior in terms of the carrier confinement as well as the electrostatic integrity when compared to a quantum-well channel with a heterobarrier. Furthermore, the S/D ion-implantation can possibly damage the channel due to the smaller binding energies in III-V materials [48]. Also, the maximum doping concentration achieved by ion-implantation is not as high as in-situ doping during the epitaxial growth, thus resulting in a larger S/D series resistance. Lastly, its finite S/D junction depth limits the gate length scaling.

Secondly, as illustrated in Figure 2.12, HEMT (or MOSHEMT) structures are well established based on III-V materials for the low density microwave ICs [49,50]. Such structures with very small parasitic capacitances offer a lot of benefits, not only in the high frequency response, but also in terms of their DC performance. Unlike inversion-mode FETs, 2-D electron gas is separated away from the doped

Figure 2.12: Illustration and cross-sectional TEM image of MOSHEMTs [49].
layer, resulting in an excellent electron mobility. Moreover, the large access regions between the gate and S/D regions enhance the 2-D electrostatic control and simultaneously suppress band-to-band tunneling near the channel-drain junction which accommodates for drain depletion. However, the large S/D spacings inevitably increase the device footprint, thus hindering high degree of device integration. In addition, the larger S/D access regions as well as the wide bandgap barrier underneath N+ S/D would increase the S/D series resistance.

Figure 2.13: Illustration and cross-sectional TEM image of trench-etched III-V MOSFETs [51].

Next, Figure 2.13 shows trench-etched MOSFETs [51]. In such structures, the channel and N+ S/D regions are formed using an epitaxial growth either by MOCVD or MBE. The gate region is formed by etching through the N+ S/D regions and stopping on an etch stop layer underneath, and finally the high-\textit{k} and metal gate-stacks are formed by ALD. As a test structure at the developing stage, it has realized an excellent transconductance and subthreshold slope. On the other hand, in consideration of the device dimensions and packing density for manufacturing, it can be very challenging to form millions of sub-15 nm grooves
in a reproducible fashion, either by wet-etching or reactive-ion etching.

Finally, Figure 2.14 shows MOSFETs with regrown N+ S/D [52–55]. In this structure, S/D regions are formed by an epitaxial regrowth either before or after the gate-stack formation. The regrown S/D is self-aligned to the gate, which is hugely advantageous for achieving a very small S/D contact pitch. Also, it does not form any heterobarrier in the current path, thus eliminating any current choke or increase in the S/D series resistance. The aforementioned issues associated with implanted S/D can be readily resolved by adopting regrown S/D in a MOSFET structure. Thus, this device structure has been chosen at UCSB, as it is attractive for fabricating III-V FETs for VLSI applications.

2.4 Conclusions

In this chapter, an essential FET theory for understanding nanoscale FETs has been investigated. MOS electrostatics and I-V characteristics have been discussed in view of the capacitance components associated with the quantum confinement and ballistic carrier transport. Next, key design parameters in III-V MOSFETs
have been discussed regarding the channel effective mass and thickness as well as the limitation in the S/D parasitic series resistance. Moreover, in order to specify a goal for the DC performance, $I-V$ characteristics of ideal FETs have been evaluated based on the ballistic FET model with realizable device parameters. Lastly, potential device structures have been assessed by comparing the main merits and demerits for each structure.
Chapter 3

1st Generation: the Gate-Last Process and Composite-Channel

3.1 Gate-First vs. Gate-Last Process

In the early stages of III-V CMOS development at UCSB, the gate-first process scheme had been adopted, which is illustrated in Figure 3.1 [52]. In this process, the high-\(k\) dielectric and gate metal stack are defined by dry etching, followed by the formation of S/D regions. There are two main justifications for implementing this process for fabricating III-V CMOS. Firstly, the arsenic capping method, which is one of the most reliable solutions to avoiding III-V surface oxidation, is applicable only to the gate-first process [56]. The III-V channel surface is very sensitive to surface oxidation, which may increase the interface trap density due to the formation of dangling bonds. Therefore, it is essential to prevent the InGaAs (or InAs) channel surface from being exposed to air prior to the high-\(k\) dielectric
deposition. The arsenic capping method allows for the high vacuum condition to be sustained at the initiation of high-

k deposition, thus achieving high quality high-

k/InGaAs interfaces. Secondly, the gate-first process allows for the formation of in-situ S/D metal contacts immediately after the S/D region regrowth, which yields lower S/D contact resistances [52].

![Process flow schematics of the gate-first process.](image)

However, the gate-first process includes a few high damage-inducing steps. First of all, a refractory gate metal (i.e. tungsten) is required for the high tem-
perature (500 °C) S/D regrowth process. Sputtering and subsequent dry etching of the refractory metal cause irreversible damages to the FET channel and high-$k$ dielectric [58]. In addition, the gate-first process requires implementing sidewall spacers (shown in Figure 3.2(a)) to prevent shorting of S/D-to-gate regions during regrowth. A sidewall spacer (which is typically 20 nm-thick) thicker than the junction diffusion length will result in increasing the FET access resistance. This is because the regrown S/D structure forms a perfectly abrupt doping profile at the junction to the channel region, as opposed to the case of an implanted S/D. In order to compensate for this bottleneck effect of the sidewall spacers on high access resistance, the pulse doping concentration in the back barrier must be very high, in the order of $\sim 4 \times 10^{12}$ cm$^{-2}$. However, adopting such a high pulse doping introduces detrimental side effects to both the on- and off-state performances: (1) a decrease in the threshold voltage, (2) a lower total gate capacitance due to the retreated electron wave-function, (3) an increase in the off-state leakage due to the lower barrier height, and (4) a reduction in the quantized bandgap due to the strong quantum-confined stark effect. In Figure 3.2(a) and (b) [59], a 60 nm-$L_g$ device fabricated using the gate-first process shows poor $g_m$ and subthreshold characteristics as well as a large off-state leakage. More comprehensive analyses on this topic are described in the previous studies done at UCSB [57,59].
Figure 3.2: (a) Cross-sectional TEM image of a FET fabricated based on the gate-first process. The red dotted circles indicate ungated regions due to the sidewall spacers. (b) Output and (c) transfer characteristics for a 60 nm-$L_g$ gate-first FET. [59]

On the contrary, the gate-last process does not incorporate steps that induce process-related damages on the channel/high-$k$ interface. In this process, a thermally evaporated gate metal is used rather than the sputtered refractory metal, which is highly damage-inducing as described previously. Since the dry etching of dielectric dummy gates is a far more gentle process compared to that of a refractory metal in terms of the ICP plasma power, the use of the gate-last process results in reduced process-induced damages. The arsenic capping method used in the gate-first process, however, is not compatible with the gate-last process. This demands an alternative, yet effective, surface passivation technique for the surfaces of air-exposed InGaAs (or InAs) channels. Hence, an in-situ passivation method has been developed using cyclical exposures of hydrogen plasma and Trimethyl Aluminum (TMA) to restore the air-exposed InGaAs (or InAs) channel surface [60]. Figure 3.3 compares three different MOCAP samples: (1) one without any pretreatment, (2) one only with the hydrogen ($H_2$) plasma strike, and (3) one with both the $H_2$ plasma and TMA exposures [59]. Upon adding $H_2$ plasma
and TMA dosing steps, the false inversion humps at low frequency start to decrease. This implies that the H$_2$ plasma/TMA pretreatment effectively passivates electrically activated interfacial traps near the mid-gap. In-depth studies on this topic are describe in [59].

Moreover, the gate-last process has a shorter turnaround time than the gate-first process. While the gate-first process readily implements the sidewall spacers as well as the self-aligned S/D metal contacts, which are desired features in CMOS VLSI, its process turnaround takes at least a month in a university research setting. On the other hand, the gate-late process omits several process steps that are necessary for the aforementioned features, allowing more experiments to be run in the same amount of time.
3.2 Device Structure and Fabrication

Figure 3.1 illustrates the overall sequence of the fabrication process for the 1st generation gate-last MOSFETs [61]. A solid-source-based molecular beam epitaxy (MBE) is used to grow the desired layers. On an InP (100) Fe-doped semi-insulating (S. I.) substrate, the epitaxial layers are grown as follows: a 300 nm unintentionally doped (UID) In$_{0.52}$Al$_{0.48}$As barrier and buffer layer, a 3 nm Si-doped (1.8×10$^{12}$ /cm$^2$) In$_{0.52}$Al$_{0.48}$As pulse doping layer, a 3 nm In$_{0.52}$Al$_{0.48}$As (UID) setback layer, and the channel layer. Four different channel designs are prepared: (1) 2 nm In$_{0.53}$Ga$_{0.47}$As / 2.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As, (2) 2 nm In$_{0.53}$Ga$_{0.47}$As / 3.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As, (3) 2 nm In$_{0.53}$Ga$_{0.47}$As / 4.5
nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As, and (4) 10 nm In$_{0.53}$Ga$_{0.47}$As (UID). It is noted that the fourth sample has no setback layer between the channel and the pulse doping layer.

On these wafers, 300 nm of SiO$_2$ is deposited at 250 °C by plasma-enhanced chemical vapor deposition (PECVD). A 20 nm-thick chromium (Cr) is subsequently deposited by e-beam evaporation, which will serve as an etch mask to define the SiO$_2$ dummy gates. For an MBE regrowth, the dummy gates must be compatible to a photoresist planarization process [62], thus requiring a 300 nm-thick SiO$_2$. A detailed discussion on this issue will be discussed in Chapter 4. Dummy gate patterns with gate lengths ranging from 50 nm to 1 µm are patterned using e-beam lithography with ma-N 2403 resist. Subsequently, they are transferred to the Cr hard mask by ICP dry etching based on Cl$_2$/O$_2$ and finally to the SiO$_2$ layer by ICP dry etching based on SF$_6$/Ar. This two-step dry etching process reduces any “foot” left at the bottom edge of the dummy gates, which tends to leave ungated access regions. In order to obtain a consistent regrowth condition near the dummy gate edge, the Cr hard mask left on SiO$_2$ dummy gates is removed later by photoresist planarization and dry etching.

Prior to the S/D regrowth in the MBE chamber, samples are intentionally oxidized by a 30-minute UV ozone exposure and then subsequently dipped in dilute HCl (1:10 HCl:H$_2$O) for 1 minute to remove the surface oxides. Inside the MBE chamber at a pressure less than 1×10$^{-9}$ Torr, samples are heated at 325 °C for 1 hour. Then, they are further heated up to 420 °C and treated with thermally cracked hydrogen (∼1×10$^{-6}$ Torr) for 40 minutes before the regrowth. Finally, approximately 60 nm of 6×10$^{19}$ cm$^{-3}$ Si-doped InAs is grown relaxed on
the sample surfaces non-selectively. Due to this non-selective nature of the MBE regrowth, amorphous InAs debris is left on the top as well as on the sidewalls of the dummy gates. The debris on the top surface of the dummy gates is removed by the photoresist planarization process. The devices are then mesa-isolated by wet etching in a mixture of 50 ml H$_2$O:50 g anhydrous citric acid:75 ml H$_2$O$_2$. The mesa height is approximately 180 nm from the top of the regrown S/D stacks. The SiO$_2$ dummy gate is removed using a buffered oxide etch (BOE) with a dilute concentration of an ethylene oxide/propylene oxide copolymer surfactant to suppress possible deposition of the debris on the exposed channel surfaces.

Immediately prior to the gate dielectric deposition, samples are dipped in dilute HCl (1:10 HCl:H$_2$O) to remove any native oxide on the channel surface. The samples are then transferred into an Oxford Instruments FlexAL ALD system. After a surface pretreatment based on multiple-cycle TMA and H$_2$ plasma exposures [60], a 30-cycle Al$_2$O$_3$ (∼3.3 nm) is deposited at 300 °C, followed by a 15-cycle HfO$_2$ (∼1.5 nm) deposition at 300 °C. This HfO$_2$ layer protects Al$_2$O$_3$ from etching in the photoresist developer. The samples are then transferred in air to a rapid thermal annealer and annealed at 400 °C for 1 hour with forming gas (10% H$_2$/90% N$_2$).

The gate metallization is done by a photoresist lift-off process. To avoid damaging the channel regions from any X-ray exposure or physical damage from the gate metal sputtering, a thermal evaporation method is used to deposit 90 nm of Ni as the gate electrodes [63]. The source/drain electrode regions are defined by optical lithography, and the high-$k$ dielectric on the regions is removed with BOE. Finally, S/D electrodes (20 nm Ti/20 nm Pd/130 nm Au) are defined by e-beam
evaporation and subsequent lift-off. The epitaxial layers and process specifications are summarized in Figure 3.5 below.

<table>
<thead>
<tr>
<th>Epitaxial Layers (By MBE)</th>
<th>Channel:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1) 2 nm In_{0.53}Ga_{0.47}As / 2.5 nm InAs / 3 nm In_{0.53}Ga_{0.47}As</td>
<td></td>
</tr>
<tr>
<td>Sample 2) 2 nm In_{0.53}Ga_{0.47}As / 3.5 nm InAs / 3 nm In_{0.53}Ga_{0.47}As</td>
<td></td>
</tr>
<tr>
<td>Sample 3) 2 nm In_{0.53}Ga_{0.47}As / 4.5 nm InAs / 3 nm In_{0.53}Ga_{0.47}As</td>
<td></td>
</tr>
<tr>
<td>Sample 4) 10 nm In_{0.52}Al_{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Setback: 3 nm, U.I.D, In_{0.52}Al_{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Pulse doping: 3 nm, 1.8×10^{12} cm^{-2} Si-doped, In_{0.52}Al_{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Buffer/barrier: 300 nm, U.I.D, In_{0.52}Al_{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Substrate: Semi-insulating InP</td>
<td></td>
</tr>
</tbody>
</table>

| Dummy Gate | 300 nm PECVD SiO_2 and dry-etched by SF_6/Ar based ICP |
| Regrowth | 60 nm, 6.0×10^{19} cm^{-3} Si-doped, InAs by MBE |
| High-k dielectric | 3.3 nm Al_2O_3 by TMA/H_2 plasma passivation / 1.5 nm HfO_2 |

Figure 3.5: Key specifications for the 1st generation gate-last process.
3.3 Results and Discussion

Figure 3.6: (a) Output ($I_D-V_{DS}$) and (b) transfer ($I_D-V_{GS}$ and $g_m$) characteristics for a 2 nm In$_{0.53}$Ga$_{0.47}$As / 2.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As composite-channel device with 40 nm-$L_g$; (c) output and (d) transfer characteristics for a 2 nm In$_{0.53}$Ga$_{0.47}$As / 3.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As composite-channel device with 70 nm-$L_g$; and (e) output and (f) transfer characteristics for a 10 nm In$_{0.53}$Ga$_{0.47}$As-only channel device with 80 nm-$L_g$. 
The output ($I_D-V_{DS}$) and transfer ($I_D-V_{GS}$ and $g_m-V_{GS}$) characteristics of composite-channel FETs are shown in Figure 3.6(a), (b), (c) and (d), and a 10 nm-thick channel FET are shown Figure 3.6(e) and (f) [61]. The composite-channel device with 2 nm In$_{0.53}$Ga$_{0.47}$As / 3.5 nm InAs / 3 nm In$_{0.53}$Ga$_{0.47}$As exhibits a maximum drain current density of $\sim$0.5 mA/$\mu$m at $V_{GS}-V_T = 0.7$ V and $V_{DS} = 0.4$ V and a peak transconductance ($g_m$) of 0.76 mS/$\mu$m at $V_{DS} = 0.4$ V. Its extracted threshold voltage ($V_T$) by linear extrapolation is approximately 0.36 V. When comparing composite-channel FETs with a 2.5 nm- and 3.5 nm-thick InAs layers, the FET with a 3.5 nm-thick InAs layer shows $>10\%$ larger peak $g_m$ than that with a 2.5 nm-thick InAs layer, even at a longer gate length of 70 nm. It is also observed that the composite channel FET with a thinner InAs has a larger $V_T$, which is attributed to the increase in the quantized bandgap.

It is also important to understand the differences between the In$_{0.53}$Ga$_{0.47}$As-only channel FET and the composite-channel FETs. The composite-channel devices show significantly larger $I_{DS}$ and $g_m$ than the InGaAs-only channel device. Several possible explanations for this observation can be proposed. Firstly, gallium outmigration during MBE regrowth may transform an InGaAs channel effectively to an InAs channel under the regrown S/D regions, as shown in the energy dispersive X-ray spectroscopy (EDX) analysis in Figure 3.7 [61]. Consequently, an InAs/In$_{0.53}$Ga$_{0.47}$As hetero-barrier would exist in the source-to-channel junction of the InGaAs-only channel device. In the composite-channel FETs, the average indium composition is higher, hence the barrier is rather suppressed. Secondly, the lower bound-state effective mass in the composite-channel should result in an increase in the carrier injection velocity as well as the reduction in the scattering
probability due to its smaller density of states. Lastly, the mean depth of the electron wave-function is greater for the composite-channel device, which should reduce surface scattering. Although the deeper wave-function depth should lower the total gate capacitance, it is important to note that the relatively small oxide capacitance in the experimental FETs is likely the most dominant limiting factor for the total gate capacitance. Yet, at this point, it is difficult to distinguish between these effects with available data.

Figure 3.7: Cross-sectional STEM image of the source-channel junction of the InGaAs channel device. The insets show Ga, In, and As spatial distributions derived from an EDX analysis.

Figure 3.8 shows log($I_D$)-$V_{DS}$ plots of the composite-channel FETs with varied InAs layer thicknesses. Figure 3.8(a), (b), and (c) represent devices with 2.5 nm-, 3.5 nm-, and 4.5 nm-thick InAs layer, respectively. For all devices, a hysteresis of approximately 60 mV between forward and reverse $V_{GS}$ sweeps is observed in the subthreshold region. The device with the 2.5 nm-thick InAs layer (Figure 3.6(a)) shows a subthreshold swing of 130 mV/decade at $V_{DS} = 0.1$ V and a drain induced barrier lowering of 400 mV/V. All devices show a high off-state leakage under
high drain bias, which indicates band-to-band tunneling occurring due to both the high drain field near the channel-drain junction and the small bandgap of the InAs-containing composite-channel design. With thicker InAs in the composite-channel, increasing source-drain leakage and decreasing drain breakdown voltage are observed. The increase in leakage may be attributed to misfit dislocations acting as a structural donor source [64].

The strain due to the lattice mismatch is likely relaxed by the generation of misfit dislocations, since the InAs critical thickness is expected to be less than 4.5 nm based on the growth condition used for this particular study. Figure 3.9(a) and (b) from [65] show the dependence of the InAs critical thickness on growth temperature and As/In flux ratio, respectively. It has been found that the InAs
critical thickness rapidly decreases around 450 °C, and it becomes only ~2 nm at 460 °C. On the other hand, the critical thickness decreases with decreasing As/In flux ratio. All samples studied in these studies are grown at ~460 °C and with As/In flux ratio of 30. Therefore, the devices with thicker InAs likely consist of a relaxed InAs layer in the channel. Figure 3.10 compares two composite channel devices containing a 4.5 nm-thick InAs layer grown at (a) 400 °C and (b) 450 °C. The device grown at 400 °C shows an off-state leakage current which is less than one-tenth of that grown at 450 °C.

Figure 3.9: (a) Dependence of critical thickness on InAs growth temperature, where As/In flux ratio is 20. (b) Dependence of critical thickness on As/In flux ratio, where growth temperature is 400 °C. [65]
Figure 3.10: Off-state leakage behaviors of FET with a 4.5 nm-thick InAs layer grown at (a) 400 °C and (b) 450 °C.

Figure 3.11 shows transmission line method (TLM) measurements on the S/D metal-semiconductor contact resistivity and the sheet resistance of the N+ regrown S/D. TLM structures on the samples have varied spacings ranging from 1 µm to 25 µm. A top-down SEM image and cross-sectional schematic of the TLM structure are shown in Figure 3.11(a) and (b). From Figure 3.11(c), a metal-to-regrown N+ layer contact resistivity of $\sim$1 Ω·µm² and an N+ InAs regrown S/D layer sheet resistance of $\sim$18 Ω/square are determined. In consideration of the $\sim$1.2 µm spacing between the edge of S/D and channel in the FETs, the total S/D access resistance from N+ S/D is calculated to be $\sim$45 Ω·µm.
Figure 3.11: (a) Top-down SEM image and (b) cross-section schematic of the TLM structure. (c) Measured TLM of contacts to the regrown N+ InAs layer.

Figure 3.12 shows a plot of $R_{on}$ as a function of $L_g$ for the device with a 3.5 nm-thick InAs layer, which is measured at $V_{GS} - V_T = 0.65$ V. $R_{on}$ extrapolated at a zero-$L_g$ is 300 $\Omega \cdot \mu$m. This value includes $\sim 45$ $\Omega \cdot \mu$m of the S/D access resistance from N+ S/D and 50-60 $\Omega \cdot \mu$m ballistic resistance from the quantum limit in the 2-D quantum well.
3.4 Conclusions

This chapter has outlined the baseline experiments in this dissertation. It includes the process flow and device results of the 1st generation gate-last III-V MOSFETs with regrown S/D. In contrast with the previous studies done at UCSB, this work has adopted the gate-last process scheme, since it minimizes process-induced damages and reduces the process turnaround. To adopt the gate-last process, an in-situ ALD passivation technique on the air-exposed InGaAs surface has been implemented. The device fabricated by the gate-last process has exhibited a significantly improved device performance over the gate-first FETs from the previous work. Moreover, preliminary results on the InAs/InGaAs composite-channel devices have been discussed in detail.
Chapter 4

$2^{nd}$ Generation: MOCVD

Regrowth and Surface Digital Etching

In the $1^{st}$ generation devices, an *in-situ* ALD passivation technique on the air-exposed InGaAs surface has been implemented along with the gate-last process scheme. This has resulted in a significantly improved device performance in comparison to the previous FETs with the gate-first process. Also, further improvements in the transconductance and on-current have been achieved by incorporating an InAs layer into an InGaAs-only channel to form a composite-channel. However, the on-state characteristics of $1^{st}$ generation FETs have not matched the level of performance expectation mentioned in Chapter 2 as well as the performance of Si-based counterparts. Therefore, the work involved in the $2^{nd}$ generation devices is aimed to understand issues related to process modules and integration that may
cause performance degradation. Key technological developments pursued in this generation are the following: (1) selective S/D regrowth using metal-organic chemical vapor deposition (MOCVD) and (2) digital etching which removes damaged surface layers post S/D regrowth.

4.1 MBE vs. MOCVD regrowth

In the 2nd generation devices, the most significant change made in the process module development is the replacement of the MBE InAs S/D regrowth with the MOCVD InGaAs S/D regrowth. MBE S/D regrowth requires a substantial adatom mobility to ensure a smooth epitaxial growth without forming voids near the edges of dummy gates [57]; therefore, the use of InAs instead of InGaAs is necessary for regrowth. Since InAs is not lattice-matched to InGaAs, this regrowth forms threading dislocations at the channel/regrowth interface as shown in the TEM image in Figure 4.1. Moreover, the conduction band offset between InAs/InGaAs reduces the number of available source electrons to be supplied into the channel.
Figure 4.1: Cross-sectional STEM image of a FET with S/D regions regrown by MBE. The junction of InGaAs channel/regrown N+ InAs is magnified on the right-hand side. A significant amount of defects at the regrowth interface and stacking faults near the junction are observed. Image courtesy of Stephan Krämer from UCSB Materials Department.

More importantly, the solid-source based MBE regrowth is not fully selective between the surfaces of the semiconductor and dielectric-based dummy gates. Therefore, InAs is not only deposited on the InGaAs surface during regrowth but also on the dummy gates in an amorphous or polycrystalline form. To remove this InAs debris, a photo-resist planarization step must be incorporated as illustrated in Figure 4.2(a). Unfortunately, even after this step, some InAs debris remain on the sidewalls of the dummy gates, which tend to fall down to the InGaAs surface during the removal of the dummy gates via wet etch. Figure 4.2(b) shows an SEM image of InAs debris on the sidewalls of a dummy gate post photo-resist planarization, and Figure 4.2(c) shows an image of the channel region with fallen InAs debris. These InAs particles sitting on the FET active region presumably in-
troduce defects, thus causing a degradation in the device performance. Moreover, the photo-resist planarization process requires dummy gates with a high-aspect ratio, therefore limiting the gate length scaling. It also lengthens the fabrication turn-around time.

On the other hand, MOCVD utilizes chemical vapor phase precursors that provide a perfectly selective regrowth between the surfaces of the semiconductor and dielectric dummy gates. Therefore, the formation of any InAs debris is completely suppressed, and smooth edges of regrown layers are achieved simultaneously. Figure 4.2(d) shows an SEM image of a dummy gate after S/D regrowth by MOCVD, which has very clean and well-defined edges. Figure 4.2(e) shows an image of a FET active region with MOCVD regrown S/D regions post gate metal deposition. It is clearly shown in these images that InGaAs S/D regrowth by MOCVD prevents any formation of crystal defects associated with lattice-mismatch. It also eliminates the InAs/InGaAs conduction band offset, which would serve as a barrier to electrons at the source. Most importantly, it substantially simplifies the overall process of the III-V MOSFETs with a shorter process turn-around.
Figure 4.2: (a) Process flow schematics of photo-resist (PR) planarization. (b) Top-down SEM image of a dummy gate post MBE S/D regrowth and subsequent PR planarization. The InAs debris still remains on the sidewalls. (c) SEM image of a FET active region with MBE regrown S/D post metallization. InAs particles are sitting on the channel. (d) SEM image of a dummy gate post MOCVD S/D regrowth. (e) SEM image of a FET active region with MOCVD regrown S/D post gate metallization.
4.2 Surface Digital Etching

In the 1st generation devices, a decent MOSCAP $C-V$ characteristic has been obtained as shown in Figure 3.3(c) by applying an in-situ surface passivation technique. Using the conductance method [66], it has been found that the interfacial trap density of the MOSCAP is below $5 \times 10^{12} \text{eV}^{-1}\text{cm}^{-2}$, which is a value corresponding to $SS$ of $< 80 \text{mV/decade}$ in long-channel FETs. However, much higher $SS$ of $> 200 \text{mV/decade}$ has been measured in the long-channel FETs fabricated so far. Upon comparing the two devices, it is noted that the fabrication process of MOSCAPs lacks the regrowth step which is essential in that of FETs. If the regrowth process which is performed at a relatively high temperature of $500 \, ^\circ \text{C}$ degrades the quality of the InGaAs channel surface, a higher $D_{it}$ may be present at the high-$k$/InGaAs channel interface.

In order to verify this postulate, the regrowth process is emulated during a MOSCAP fabrication by employing a $\text{SiO}_2$ deposition and subsequent annealing at $500 \, ^\circ \text{C}$ in furnace [59]. In Figure 4.3 [59], a sample that has undergone the emulated regrowth cycle is compared to a control sample that has not sustained such a thermal cycle. The $C-V$ result of the MOSCAP post the emulated regrowth shows a very strong frequency dispersion when biased toward depletion. This indicates a substantial interface degradation, which typically results in a large increase in the interface trap density near the midgap. Hence, the surfaces of the semiconductor channel are likely contaminated or damaged during regrowth. The next key technological improvement involves the removal of the likely damaged channel surfaces post S/D regrowth.
In order to address this issue, a few nanometers of sacrificial InGaAs cap are grown on the channel during the epitaxial layer growth step. After S/D regrowth and dummy gate etching, the sacrificial cap on the top of the channel, which is likely damaged, is removed in a self-limited manner. Since the cap is of the same material as the channel, it is essential to control the etch depth with nanometer-scale precision. Thus, a digital etching technique is adopted and optimized for this study \[67\]. A single etching cycle consists of a 15-minute UV ozone exposure followed by 1-minute dip in dilute HCl (1:10 HCl:H\textsubscript{2}O) and 1-
minute rinse in DI water. The UV ozone exposure is a very slow and gentle oxidation process when compared to dilute H$_2$O$_2$ and dry O$_2$ plasma, allowing for well-controlled etch depth, profile, and uniformity. The obtained etch rate is 1.3-1.5 nm per cycle, which is determined by averaging the etch depths scanned using atomic force microscopy (AFM) on multiple test samples with a 10 nm-thick hard mask. Figure 4.4 shows an etch profile after 4 cycles of digital etching. The RMS roughness of the etched InGaAs surface is ∼0.16 nm, which is comparable to the as-grown surface.

![Image](Etch profile from AFM (4 cycles))

Figure 4.4: AFM surface and depth profiles of a test structure after 4 cycles of digital etching (15-minute UV ozone exposure + 1-minute dilute HCl (1:10 HCl:H$_2$O) dip). ∼5.2 nm of In$_{0.53}$Ga$_{0.47}$As is etched.

Figure 4.5 compares $g_m$ and $SS$ of 10 nm-thick In$_{0.53}$Ga$_{0.47}$As channel FETs with and without surface digital etching. With surface digital etching, a 75 nm-$L_g$ FET shown in Figure 4.5(c) exhibits ∼75% higher $g_m$, and a 500 nm-$L_g$ FET in Figure 4.5(d) shows ∼200% reduced $SS$. Such dramatic improvements serve as
strong evidence that the digital etching effectively reduces process-induced border traps as well as interfacial traps that are likely located at energy levels accessed by the range of $V_{GS}$ used during the FET operation.

Figure 4.5: Comparison of FET $I-V$ performances with and without surface digital etching (D.E.). (a) $I_D$- and $g_m$-$V_{GS}$ for a 75 nm-$L_g$ FET without D.E., (b) log($I_D$)-$V_{GS}$ for a 500 nm-$L_g$ FET without D.E., (c) $I_D$-and $g_m$-$V_{GS}$ for a 75 nm-$L_g$ FET with D.E., and (d) log($I_D$)-$V_{GS}$ for a 500 nm-$L_g$ FET with D.E.
4.3 Device Structure and Fabrication

Figure 4.6 summarizes the sequence of the fabrication process. The epitaxial layers, grown on a semi-insulating InP substrate by solid-source MBE, consist of the following [55]: a 400 nm unintentionally doped $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ buffer/barrier layer, a 3 nm Si-doped ($3.9 \times 10^{12}$ cm$^{-2}$) $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ pulse doping layer, a 3 nm $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ bottom cladding layer, a 5 nm InAs channel (strained), a 3 nm $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ upper cladding and a 5 nm Si-doped ($4-5 \times 10^{19}$ cm$^{-3}$) $\text{N+ In}_{0.53}\text{Ga}_{0.47}\text{As}$ cap. In order to avoid strain relaxation of the InAs layer in the composite-channel design, as discussed in detail in Chapter 3.3, the As/In flux ratio is set to 10, and the InAs and the InGaAs upper cladding are grown at a lower temperature of 400 °C [68].
A 50 nm-thick SiO$_2$ is deposited by PECVD and subsequently patterned as dummy gates by e-beam lithography and Cl$_2$/O$_2$ based ICP dry etch. As opposed to the 1st generation process scheme, the photo-resist planarization step is now eliminated due to the implementation of MOCVD S/D regrowth. This enables a substantial reduction of dummy gate heights; therefore, the dummy gate width, which determines the gate length, is further scaled down to 30 nm. The samples are oxidized by 30-minute UV ozone exposure and then etched by 1-minute dilute 1:10 HCl:H$_2$O prior to being transferred to the MOCVD. Inside the chamber, the samples are heated up to 600 °C for several minutes to remove any remaining native oxide, then 60 nm Si-doped (4×10$^{19}$ cm$^{-3}$) In$_{0.53}$Ga$_{0.47}$As is selectively grown on the N+ cap layer at 500 °C. Then device mesas are defined by wet-etching, and the dummy gates are removed in BOE using the identical process steps as in the 1st generation devices. The exposed N+ cap and upper cladding layer are digitally etched with 5 cycles of (1) oxidation by 15-minute UV ozone exposure and (2) removal of surface oxide by 1-minute dilute 1:10 HCl:H$_2$O dip [55]. Immediately after removing the oxidized InGaAs surface layer from the last cycle of the UV ozone exposure in BOE, the samples are transferred into the ALD chamber. They are pre-cleaned and passivated by a cyclic nitrogen plasma and TMA in-situ treatment; subsequently, a ∼3.5 nm HfO$_2$ gate dielectric is deposited [41]. The samples are then annealed for 15 minutes at 400 °C in forming gas. A 20 nm Ni/100 nm Au stack is thermally deposited as the gate electrode. Lastly, a 20 nm Ti/20 nm Pd/130 nm Au stack is lifted off to form S/D electrodes. The epitaxial layers and main process specifications are listed in Figure 4.7.
Epitaxial Layers  
(By MBE)  
Cap : 5 nm, 4-5.0×10¹⁹ cm⁻³ Si-doped, In₀.₅₃Ga₀.₄₇As  
Channel : 3 nm In₀.₅₃Ga₀.₄₇As / 5 nm InAs / 3 nm In₀.₅₃Ga₀.₄₇As  
Pulse doping : 3 nm, 3.9×10¹² cm⁻³ Si-doped, In₀.₅₂Al₀.₄₈As  
Buffer/barrier: 400 nm, U.I.D, In₀.₅₂Al₀.₄₈As  
Substrate : Semi-insulating InP

<table>
<thead>
<tr>
<th>Dummy Gate</th>
<th>50 nm PECVD SiO₂ and dry-etched by SF₆/Ar based ICP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regrowth</td>
<td>60 nm, 4×10¹⁹ cm⁻³ Si-doped, In₀.₅₃Ga₀.₄₇As by MOCVD</td>
</tr>
<tr>
<td>High-k dielectric</td>
<td>0.5 nm Al₂O₃,Nₓ by TMA/N₂ plasma passivation / 3.6 nm HfO₂</td>
</tr>
</tbody>
</table>

Figure 4.7: Key specifications for the 2nd generation devices.

Figure 4.8 shows a cross-sectional STEM image for a 40 nm-\(L_g\) device. Because of the lattice-match to the N+ In₀.₅₃Ga₀.₄₇As cap, In₀.₅₃Ga₀.₄₇As regrowth by MOCVD is free of growth-related defects that occur in InAs regrowth by MBE. The HR-TEM image shown in the inset of Figure 4.8 shows that an \(\sim0.5\) nm-thick interfacial layer is formed by N₂ plasma and TMA pretreatment in ALD. The thickness of HfO₂ high-\(k\) dielectric is approximately 3.5 nm. It is shown that the gate dielectric stack is deposited on a channel with 5 nm-thick strained InAs and 3 nm-thick In₀.₅₃Ga₀.₄₇As. The 3 nm-thick InGaAs upper cladding of the composite-channel is removed by digital etching as expected.
Figure 4.8: Cross-sectional STEM of a 40 nm-$L_g$ device. $\sim$3.6 nm HfO$_2$ and $\sim$0.5 nm interfacial layer are present on a 5/3 nm InAs/In$_{0.53}$Ga$_{0.47}$As channel. Image courtesy of Stephan Krämer at UCSB Materials Department.

4.4 Results and Discussion

Figure 4.9 shows an energy band structure and charge density at $V_{GS} = 0.5$ V simulated by a 1-D Poisson-Schrödinger solver. The centroid of the 2-D electron gas is located approximately at the center of the InAs/InGaAs composite-channel, $\sim$4 nm away from the gate dielectric. Its quantized bandgap is calculated to be $\sim$0.5 eV.
Figure 4.9: Energy band structure and charge distribution at \( V_{GS} = 0.5 \) V by 1-D Poisson-Schrödinger simulation.

The transfer characteristics \( (I_D - V_{GS} \text{ and } g_m - V_{GS}) \) of short channel devices with varied \( L_g \)'s are shown in Figure 4.10(a). The devices with 40, 70, and 90 nm-\( L_g \) show peak extrinsic \( g_m \) of 2.45, 2.40, and 2.16 mS/\( \mu \)m at \( V_{DS} = 0.5 \) V, respectively. All devices on the test samples (30 dies) show a similar performance (< 10% variability). Figure 4.10(b) shows subthreshold characteristics of the short channel devices. \( SS \) of devices with 40, 70, and 90 nm-\( L_g \) are \( \sim 155, 115 \) and 110 mV/decade at \( V_{DS} = 0.05 \) V and \( \sim 400, 235 \), and 190 mV/decade at \( V_{DS} = 0.5 \) V, respectively. Figure 4.10(c) shows output characteristics of a 40 nm-\( L_g \) device. Its maximum drain current density is 1.95 mA/\( \mu \)m at \( V_{GS} = 1.4 \) V and \( V_{DS} = 0.5 \) V, and its on-resistance is 214 \( \Omega \)-\( \mu \)m. The gate leakage current normalized by the total gate metal overlap is negligible; it is less than \( 10^{-3} \) A/cm\(^2\) at all measured gate biases as shown in Figure 4.10(d).
Figure 4.10: (a) Transfer characteristics of short channel devices (40, 70, and 90 nm-$L_g$), (b) subthreshold characteristics of short channel devices (40, 70, and 90 nm-$L_g$), (c) output characteristics of a 40 nm-$L_g$ device, and (d) gate leakage current normalized by total gate overlap for a 40 nm-$L_g$ device.

The transfer and subthreshold characteristics of a long channel ($L_g = 510$ nm) device are shown in Figure 4.11. A peak extrinsic $g_m$ of 1.05 mS/$\mu$m at $V_{DS} = 0.5$ V and minimum $SS$ of $\sim 93$ mV/decade at $V_{DS} = 0.05$ V are measured. The relatively high minimum $SS$ value measured at such low $V_{DS}$, despite the longer...
channel length of the device, can be attributed to both the interfacial traps and high off-state leakage current.

Figure 4.11: Subthreshold characteristics of a 510 nm-$L_g$ channel device. The minimum $SS$ is $\sim 93$ mV/decade at $V_{DS} = 0.05$ V.

Figure 4.12 shows the dependency of $SS$ as well as linearly extrapolated threshold voltage on the device gate length. $SS$ rapidly increases with decreasing gate length, especially at a high $V_{DS}$ of 0.5 V. This is not only because of the classical short channel effect from poor 2-D electrostatics, but it reflects a combined effect of multiple factors such as a high buffer leakage and BTBT at the channel-drain junction. It is difficult to distinguish between these effects based on the available data. This topic will be further discussed in detail in the next chapter.
Figure 4.12: Subthreshold swing (SS) and threshold voltage obtained by linear extrapolation as a function of the gate length for $V_{DS} = 0.05$ V and 0.5 V.

From transmission line method (TLM) measurements, shown in Figure 4.13, a S/D metal contact resistivity of 4.7 $\Omega\cdot\mu$m$^2$ and sheet resistance of $\sim$25 $\Omega$/square are extracted from the regrown S/D contact layers. Due to the lower doping concentration of MOCVD regrown InGaAs S/D regions as well as the lower Fermi-level pinning position with respect to the conduction band edge of InGaAs, the contact resistivity and sheet resistance are $\sim$5:1 and $\sim$1.3:1 greater compared to MBE regrown InAs S/D used in the previous generation. However, since the FETs have large S/D contact areas, the abovementioned differences between MOCVD and MBE regrown layers become negligible for the total access resistance. Based on the gap distance between S/D metal contact and channel as shown in inset of Figure 4.13(a), the S/D access resistance is calculated to be 82 $\Omega\cdot\mu$m, which corresponds to a degradation in the intrinsic $g_m$ by $\sim$8%.

Finally, Figure 4.13(b) shows $R_{on}$ as a function of the gate length. Despite the
increased S/D access resistance, the zero-$L_g$ extrapolated $R_{on}$ of 200 $\Omega$-$\mu$m (fitting error is $\pm$ 1.77 $\Omega$-$\mu$m) is significantly less than that of the FETs fabricated in the 1st generation (Figure 3.12). This observation may be attributed to an absence of defects at the regrowth interface, which would reduce the junction resistance between the channel and regrown N+ S/D.

\[
Y = 21.8 + 25.1X \\
\rho_c = 4.7 \, \Omega\text{-}\mu\text{m}^2 \\
R_{SD} = 82 \, \Omega\text{-}\mu\text{m}
\]

(a)

(b)

Figure 4.13: (a) TLM measurement for N+ regrown S/D. (b) On-resistance ($R_{on}$) as a function of gate length. $R_{on}$ extrapolated at zero-$L_g$ is 188 $\Omega$-$\mu$m.

4.5 Conclusions

This chapter has outlined two of the main developments since the previous generation: MOCVD S/D regrowth and surface digital etching. Firstly, it has been shown that MOCVD S/D regrowth eliminates issues regarding formations of amorphous (or poly-crystalline) InAs (or InGaAs) debris due to its selective growth to the dielectric dummy gates, thus resulting in an improved device performance and variability. In addition, the omission of the planarization process
due to MOCVD S/D regrowth has allowed to further scale down the dummy gate widths (gate lengths) and to reduce process turnaround. Secondly, it has been found that surface digital etching plays a pivotal role in improving the quality of high-$k$/III-V semiconductor interface. By adopting this technique, damaged InGaAs surface layers post the high-temperature regrowth can be removed in a self-limited manner. The control experiments have confirmed that surface digital etching results in a significantly improved device performance, showing a $\sim 75\%$ increased $g_m$ and $\sim 200\%$ lowered $SS$. The device fabricated with these developments shows a very high peak $g_m$ of 2.4 mS/$\mu$m at 0.5 V $V_{DS}$ and 40 nm-$L_g$. 
Chapter 5

3rd Generation: Vertical Spacers and Barrier Optimization

The previous generations have primarily focused on enhancing the device on-state performance. Developing a low-damage process along with a faster turn-around has resulted in a significantly improved $g_m$ of 2.4 mS/$\mu$m at $V_{DS} = 0.5$ V, which had been a record result among all reported III-V MOSFETs at the time it was reported. However, their off-leakage and subthreshold swing as shown in Figure 4.10(b) have not yet met the specifications required for VLSI applications. The subsequent device development outlined in this generation, therefore, concentrates on improving the subthreshold swing and off-leakage.

According to the classical Si-based MOSFET theory, short-channel effects arise primarily from degraded 2-D electrostatics due to the reduced aspect ratio, i.e., the ratio of the gate length to the sum of channel thickness and dielectric EOT. It is noted, however, that the FETs presented in the previous chapters have a satis-
factory aspect ratio with a \(\sim10\) nm-thick channel, \(\sim1\) nm-thick EOT dielectrics, and \(\sim40\) nm-long gate length; hence only the classical short channel effects due to electrostatics cannot explain such a poor off-state performance. Therefore, in consideration of off-leakage behavior and device structure, dominant leakage mechanisms in III-V MOSFETs are carefully examined and addressed in this chapter.
5.1 Vertical Spacer

Figure 5.1: Comparison between a regrown S/D MOSFET and MOSHEMT. Cross-sectional TEM images of (a) a regrown S/D MOSFET and (b) MOSHEMT [50]. Device structures and energy band diagrams of (c) a regrown S/D MOSFET and (d) MOSHEMT. Subthreshold characteristics of (e) a regrown S/D MOSFET and (f) MOSHEMT [50].
In the previous generations, an InAs/In$_{0.53}$Ga$_{0.47}$As composite-channel has been chosen over an In$_{0.53}$Ga$_{0.47}$As-only channel in order to further improve the on-state performance. InAs-based materials have a small quantized bandgap, e.g., $\sim 0.5$ eV for a 5 nm InAs/3 nm InGaAs composite-channel FET (Figure 4.9). Therefore, a large leakage may arise from BTBT and/or from impact ionization in the high-field gate-drain region, resulting in degraded off-state characteristics and $SS$.

FETs with raised N+ S/D by epitaxial regrowth have self-aligned N+ S/D regions within a few nanometers from the gate edges as shown in Figure 5.1(a) [55]. This enables achieving a tight S/D contact pitch that is necessary for VLSI. However, as illustrated in Figure 5.1(b), such a short gate-drain separation makes the drain field extremely high especially under a high drain bias. This causes an increase in both the leakage and short-channel effects at a given $L_g$.

As opposed to regrown S/D III-V FETs, HEMT-like III-V MOSFETs shown in Figure 5.1(b) have the N+ source and drain regions situated away from the gate edges by a large distance of typically 50-200 nm [50]. In this case, the supply of electrons in the channel is usually established by a pulse-doped layer placed in the barrier layer either above or below the channel. Under the high drain bias, the electrons in the channel deplete laterally away from the gate-drain edge, allowing the gate-drain potential to distribute itself over a relatively large distance of 50-200 nm as illustrated in Figure 5.1(d). This reduces the peak drain field; thus, the leakage currents associated with BTBT and impact ionization are reduced as well. Therefore, HEMT-like III-V MOSFETs have a low leakage due to the implementation of a lateral spacer within the gate-drain high-field region.
In VLSI, however, the S/D contact pitch must be small, thus increased pitch distances by employing such wide lateral spacers cannot be tolerated.

Figure 5.1(e) and (f) compare off-state characteristics of a regrown S/D MOSFET and HEMT-like MOSFET with a similar gate length of 60-70 nm. Due to BTBT, the regrown S/D MOSFET exhibits a much higher leakage current at $V_{DS} = 0.5$ V. Also, $SS$ of the device further degrades as the gate-drain voltage increases, resulting in an abnormally high off-state leakage of $>10^{-3}$ mA/µm. It is noted that this value is at least an order of magnitude greater than the barrier limited leakage represented with a black dotted line in Figure 5.1. The HEMT-like MOSFET shows much a lower leakage current at similar bias conditions. Also, the $SS$ at $V_{DS} = 0.5$ V of this device stays nearly constant up to a drain current of $10^{-4}$ mA/µm, which indicates BTBT is adequately suppressed [50]. Hence, the goal of this generation is to match the low leakage characteristics of HEMT-like MOSFETs with the regrown S/D MOSFET design, which has a distinct advantage of a shorter S/D pitch contact.
Figure 5.2: (a) Band diagram showing how the accumulated holes may increase the off-leakage. (b) Forward-sweep log($I_D$)-$V_{GS}$ from different starting $V_{GS}$ of -0.5 V, -0.3 V and -0.1 V up to 0.5 V. (c) Reverse-sweep log($I_D$)-$V_{GS}$ from $V_{GS}$ of 0.5 V toward the ending $V_{GS}$ of -0.5 V, -0.3 V, and -0.1 V.

The high off-leakage currents so far observed in regrown S/D III-V MOSFETs have not originated solely from the excess electrons, but also from the excess holes generated by BTBT and/or impact ionization. III-V FETs based on quantum wells, which can be considered analogous to floating body silicon-on-insulator (SOI) FETs, have no contact to the quantum well (i.e. body) [69]. The generated holes, therefore, would pile up if the recombination rate is not as high as the
generation rate. Here, it is noted that the diffusion length of the InGaAs (or InAs) is much greater than the channel length, which implies the recombination in the channel is almost negligible [70]. As a result, the accumulated holes can effectively lower the potential barrier for the electron injection at the source, resulting in a substantially increased off-current as illustrated in Figure 5.2(a). This is well known as the floating body bipolar effect [71].

For verification, \( \log(I_D) - V_{GS} \) at different starting \( V_{GS} \) values of -0.5 V, -0.3 V and -0.1 V are measured and compared in Figure 5.2(b) and (c), where each value respectively represents a substantial, moderate, and weak BTBT case. In the case of a forward sweep from a negative to positive \( V_{GS} \) as shown in Figure 5.2(b), the subthreshold slope is further degraded with a more negative starting value of \( V_{GS} \), which amplifies the hole accumulation. On the other hand, when measured by a reverse sweep from a positive to negative \( V_{GS} \) as shown in Figure 5.2(c), all curves have essentially the same subthreshold slope due to the absence of hole accumulation. Therefore, this verifies that the excess holes accumulate and thus effectively lower the S-to-D barrier, resulting in a further degradation in off-leakage.
Figure 5.3: Illustrations of device structures and corresponding energy band diagrams of raised S/D MOSFETs (a) without and (b) with a vertical spacer between N+ S/D and the channel.

To address the high leakage at high $V_{DS}$ and short $L_g$, a lightly doped (i.e. unintentionally doped) InGaAs vertical spacer is grown on the channel, which consists of a narrow quantized bandgap, prior to depositing N+ S/D regions. As illustrated in Figure 5.3, the introduction of a spacer layer reduces both BTBT and impact ionization by decreasing the electric field within the gate-drain high-field region. Because the spacer layer is vertically stacked on the channel, the S/D contact pitch does not increase. Moreover, since the sidewalls of the spacer are gated, the source access resistance can be readily controlled without introducing a delta doping or surface fermi-level pinning used in the case of other III-V MOSFETs with spacers [72]. Furthermore, in consideration of 2-D electrostatics, the addition of a spacer increases the effective channel length, consequently resulting in an improved $DIBL$ especially at shorter gate lengths. The impact of
varied spacer thicknesses on both the on- and off-state device performances will be discussed in the following sections.

5.2 Other Leakage Mechanisms

To further suppress the excess off-leakage, a P-doped back barrier is employed. As shown in Figure 5.4(a), some devices fabricated in the previous generations suffer from an abnormally high off-state leakage, which is not modulated by the gate but linearly increase with the applied drain voltage. Such an ohmic conduction can happen when the Fermi-level is pinned near the conduction band edge because of either (1) a large defect density from an imperfect growth or (2) a presence of unintentionally N-doped region in the barrier. In the 2\textsuperscript{nd} generation, the leakage due to InAs relaxation in the composite-channel has been resolved by adjusting the growth conditions to increase the InAs critical thickness. Hence, to confirm the possibility of a conduction through an unintentionally N-doped barrier, secondary ion mass spectrometry (SIMS) is performed on a sample with a high off-state leakage as shown in figure 5.4(b) [59]. A very high peak concentration ($>10^{19}$ cm$^{-3}$) of silicon, which is an N-type dopant for InAlAs, is observed; thus, it would likely pin the Fermi-level at the InP/InAlAs interface, consequently lowering the barrier height of the entire back barrier region.
To compensate for this heavily N-type doped InP/InAlAs interface, a 250 nm-thick, $1.0 \times 10^{17}$ cm$^{-3}$ P-type doped layer is implemented near the back of the In$_{0.52}$Al$_{0.48}$As barrier. Its doping concentration is determined by the fact that an order of $10^{17}$ P-type doping in In$_{0.52}$Al$_{0.48}$As is possible using MBE. Figure 5.5(b) illustrates the energy band diagram with an added P-doped barrier. With the P-doping in the back, electrons are nearly depleted throughout the entire back
barrier, thus suppressing any conduction through the barrier. Figure 5.5(a) shows \( \log(I_D) - V_{GS} \) for a FET with a P-type doped barrier. With this design improvement, a significantly reduced leakage of \(< 10^{-5} \text{ mA/\mu m}\) is measured, which satisfies the off-current limit for high performance applications in VLSI.

Figure 5.5: (a) Subthreshold characteristic of a MOSFET with a barrier with a P-doping at the back. (b) Energy band diagram corresponding to the MOSFET in shown (a).
5.3 Device Structure and Fabrication

In addition to FET design enhancements introduced in the previous sections, this generation includes a couple of important process module developments regarding scaling of the FETs. Firstly, hydrogen silesequioxane (HSQ) is used instead of PECVD SiO₂ to form dummy gates. HSQ is a very high resolution negative tone electron beam resist, and its composition can be cross-linked to SiOₓ by e-beam radiation. Therefore, very fine features as small as ~10 nm can be obtained with HSQ, which obsoletes the dummy gate formation process by a
blanket deposition of SiO$_2$ and a subsequent dry etching. Moreover, HSQ-based process reduces the fabrication turn-around even further. Secondly, a 2 nm-thick InGaAs cap layer is used instead of a 5 nm-thick InGaAs cap in this generation. A separate experiment on thinning the cap has confirmed that a 2 nm-thick InGaAs cap is enough to negate the surface damage during regrowth. Furthermore, adopting a thinner cap allows reducing the increase in gate lengths (∼2 times the cap thickness) induced by the isotropic nature of surface digital etching.

Figure 5.6 illustrates the sequence of the fabrication process [73]. The epitaxial layers for all samples, grown on semi-insulating InP substrates by solid source molecular beam epitaxy, consist of the following: a 50 nm unintentionally doped (UID) InAlAs buffer, a 250 nm $1.0 \times 10^{17}$ cm$^{-3}$ beryllium-doped P-type InAlAs barrier, a 100 nm UID InAlAs barrier, a 2 nm $1.0 \times 10^{12}$ cm$^{-2}$ N-type doped InAlAs pulse-doping layer, a 5 nm UID InAlAs setback, a 6 nm strained InAs, and a 2 nm UID In$_{0.53}$Ga$_{0.47}$As cap layer. Prior to the dummy gate formation, ∼1 nm of Al$_2$O$_3$ is deposited in ALD for promoting the adhesion of hydrogen silesequioxane (HSQ) to the InGaAs surface. The portions of this adhesion layer not covered by the dummy gates are etched away in the HSQ developer. To define the dummy gates, HSQ is spun to ∼40 nm and patterned by e-beam lithography. To realize ≤ 20 nm gate lengths, a reduced beam current of 500 pA is used with the thin HSQ. Also, to maximize the resolution, the HSQ developing process based on sodium chloride and sodium hydroxide has been carefully optimized [74]. Immediately after etching off the native oxide in dilute HCl, samples are transferred into the MOCVD reactor to form the vertical spacers and N+ S/D regions. Here, three samples are selectively regrown, with a 0 nm-, 5 nm-, and 10 nm-thick UID
In$_{0.53}$Ga$_{0.47}$As spacer ($\sim 1.2 \times 10^{15}$ cm$^{-3}$ N-type), followed by the deposition of a 60 nm-thick Si-doped ($4 \times 10^{19}$ cm$^{-3}$) In$_{0.53}$Ga$_{0.47}$As N+ S/D regions. The device mesas are then isolated using the identical process from the 1$^{st}$ generation, and the dummy gates are removed in buffered oxide etch. By 2 cycles of digital etching, the 2 nm cap layer and $\sim$1 nm of the InAs channel are isotropically removed. The samples are immediately transferred into the ALD chamber and passivated by 9 cycles of N$_2$ plasma and tri-methyl-aluminum. Subsequently, 3 nm-thick HfO$_2$ gate dielectric is deposited. After a 400 $^\circ$C post-deposition annealing in forming gas, 20 nm Ni/80 nm Au thermally-evaporated gate metal stacks and 20 nm Ti/20 nm Pd/130 nm Au e-beam-evaporated S/D contacts are lifted off subsequently. All samples have a composite-channel with a $\sim$5 nm-thick InAs. The gate lengths range from 18 nm to 1 $\mu$m on the samples with 2 nm- and 7 nm-thick vertical spacers and from 30 nm to 1 $\mu$m on the sample with a 12 nm-thick vertical spacer. For the sample with a 12-nm thick vertical spacer, the sub-30 nm-$L_g$ FETs have resulted in a low yield. Here, it is noted that the total thickness of vertical spacers includes the 2 nm-thick InGaAs cap underneath. The epitaxial layers and key process specifications for the 3$^{rd}$ generation are listed in Figure 5.7.
Table 5.7: Key specifications for the 3rd generation devices.

<table>
<thead>
<tr>
<th>Epitaxial Layers (By MBE)</th>
<th>Epitaxial Layers (By MBE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap : 2 nm, U.I.D, In_{0.53}Ga_{0.47}As</td>
<td>Cap : 2 nm, U.I.D, In_{0.53}Ga_{0.47}As</td>
</tr>
<tr>
<td>Channel : 6 nm Strained InAs</td>
<td>Channel : 6 nm Strained InAs</td>
</tr>
<tr>
<td>Setback : 5 nm, U.I.D, In_{0.53}Al_{0.47}As</td>
<td>Setback : 5 nm, U.I.D, In_{0.53}Al_{0.47}As</td>
</tr>
<tr>
<td>Pulse doping : 2 nm, 2.0×10^{12} cm^{−2} Si-doped, In_{0.53}Al_{0.47}As</td>
<td>Pulse doping : 2 nm, 2.0×10^{12} cm^{−2} Si-doped, In_{0.53}Al_{0.47}As</td>
</tr>
<tr>
<td>Barrier : 100 nm, U.I.D, In_{0.53}Al_{0.47}As</td>
<td>Barrier : 100 nm, U.I.D, In_{0.53}Al_{0.47}As</td>
</tr>
<tr>
<td>P-doped barrier : 250 nm, 1.0×10^{17} cm^{−3} Be-doped, In_{0.52}Al_{0.48}As</td>
<td>P-doped barrier : 250 nm, 1.0×10^{17} cm^{−3} Be-doped, In_{0.52}Al_{0.48}As</td>
</tr>
<tr>
<td>Buffer : 50 nm, U.I.D, In_{0.52}Al_{0.48}As</td>
<td>Buffer : 50 nm, U.I.D, In_{0.52}Al_{0.48}As</td>
</tr>
<tr>
<td>Substrate : Semi-insulating InP</td>
<td>Substrate : Semi-insulating InP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummy Gate</th>
<th>Dummy Gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>~30 nm HSQ patterned by e-beam lithography</td>
<td>~30 nm HSQ patterned by e-beam lithography</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regrowth (By MOCVD)</th>
<th>Regrowth (By MOCVD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N+ S/D : 60 nm, 4×10^{19} cm^{−3} Si-doped, In_{0.53}Ga_{0.47}As</td>
<td>N+ S/D : 60 nm, 4×10^{19} cm^{−3} Si-doped, In_{0.53}Ga_{0.47}As</td>
</tr>
<tr>
<td>Vertical Spacer : 0, 5, or 10 nm, U.I.D, In_{0.53}Ga_{0.47}As</td>
<td>Vertical Spacer : 0, 5, or 10 nm, U.I.D, In_{0.53}Ga_{0.47}As</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High-k dielectric</th>
<th>High-k dielectric</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 nm Al_{2}O_{x}N_{y} by TMA/N_{2} plasma passivation / 3.0 nm HfO_{2}</td>
<td>1 nm Al_{2}O_{x}N_{y} by TMA/N_{2} plasma passivation / 3.0 nm HfO_{2}</td>
</tr>
</tbody>
</table>

5.4 Results and Discussion

The transfer and output characteristics of a FET with a 2 nm-thick vertical spacer and 18 nm-$L_g$ are shown in Figure 5.8(a) and (b). The device shows a peak $g_m$ of 3.0 mS/μm at $V_{DS} = 0.5$ V, which is the highest transconductance reported so far from any type of FETs at the given $V_{DS}$. However, its minimum $SS$ is 196 mV/decade at $V_{DS} = 0.5$ V, and the off-leakage is far above the leakage limit of $10^{-7}$ A/μm for high performance VLSI applications. This is likely due to both the high BTBT through the relatively thin (2 nm) vertical spacer and the bad 2-D electrostatic control at a shorter effective gate length. On the other hand, it shows a record low on-resistance of $\sim$200 Ω·μm at $V_{GS} = 1.0$ V.
Figure 5.8: (a) Transfer and (b) output characteristics of a FET with a 2 nm-thick vertical spacer and 18 nm-$L_g$.

Figure 5.9(a) and (b) show transfer characteristics of a 40 nm-$L_g$ device with a 12 nm-thick spacer in linear and semi-log scales. At $V_{DS} = 0.5$ V, a peak $g_m$ of 2.5 mS/µm and $SS_{min}$ of 86 mV/decade are measured. Its threshold voltage and $DIBL$ defined at $I_D = 1$ µA/µm are 35 mV and 83 mV/V, respectively. Using a criteria of $I_{off} = 100$ nA/µm and $V_{DD} = 0.5$ V, the device has $I_{on}$ of 482 µA/µm. Figure 5.9(c) shows an output characteristic of a 40 nm-$L_g$ device with a 12 nm-thick spacer. $R_{on}$ is 289 Ω·µm at $V_{GS} = 1.0$ V. Figure 5.9(d) plots log($I_D$) vs. $V_{GS}$ as well as the gate leakage of a FET with a 12 nm-thick spacer and 1 µm-$L_g$. The $SS_{min}$ at $V_{DS} = 0.1$ V is 66 mV/decade. Assuming that the dielectric constants of Al$_2$O$_x$N$_y$ and HfO$_2$ are 9 and 17, respectively, the equivalent oxide thickness (EOT) is $\sim$1.1 nm. Then, $D_{it}$ is calculated to be $\sim$2.2×10$^{12}$ cm$^{-2}$·eV$^{-1}$ from $D_{it} = C_{EOT}/q\cdot(SS/2.3kT-1/q)$ [39]. The measured gate leakage is negligible.
Figure 5.9: Transfer characteristics of a 40 nm-$L_g$ FET with a 12 nm-thick vertical spacer in (a) linear and (b) semi-log scales. (c) Common-source characteristic for a 40 nm-$L_g$ device with a 12 nm-thick spacer. (d) Subthreshold and gate leakage characteristics of a 1 $\mu$m-$L_g$ FET with a 12 nm-thick vertical spacer.

Figure 5.10(a) compares the minimum subthreshold slope of samples with 2, 7, and 12 nm-thick vertical spacers as a function of $L_g$ both at $V_{DS} = 0.1$ and 0.5 V. At $V_{DS} = 0.5$ V, $SS_{min}$ at all gate lengths shows a huge improvement as the spacer thickness is increased. However, only a minor improvement in $SS_{min}$ is observed
at $V_{DS} = 0.1$ V. This suggests that the spacer improves the FET electrostatics and reduces the leakage originating from the high-field region between the gate and the N+ drain. However, the current set of data is not sufficient to fully distinguish between the relative contributions of BTBT and impact ionization to this leakage current. Figure 5.10(b) compares $DIBL$ as a function of $L_g$. The 12 nm-thick spacer devices exhibit significantly lower $DIBL$ than the 2 nm- and 7 nm-thick spacer devices.

![Figure 5.10: (a) SS$_{min}$ as a function of $L_g$ at $V_{DS} = 0.1$ and 0.5 V for 2 nm-, 7 nm- and 12 nm-thick spacer devices. (b) DIBL as a function of $L_g$ for 2 nm-, 7 nm- and 12 nm-thick spacer devices.](image)

Figure 5.11(c) compares the off-state leakage vs. $V_{GS}$ and $V_{DS}$ for all devices with $L_g = 1$ $\mu$m. In principle, the increase in the effective channel length due to the added spacer thickness should be negligible in these long-channel devices. However, the 12 nm-thick spacer devices show a $\sim$8:1 to 9:1 lower minimum leakage at $V_{DS} = 0.7$ V than the 2 nm-thick spacer devices. Hence, improvements seen in the minimum off-state leakage as well as in $DIBL$ (shown in Figure 5.10(b)) from
12 nm-thick spacer devices with longer gate lengths suggest that the reduced leakage should not be attributed solely to the increase in the effective channel length because of a thicker spacer.

![Image of a graph showing log(I_D)-V_{GS} for 2 nm-, 7 nm- and 12 nm-thick spacer devices with \(~1 \mu m-L_g\).](image)

Figure 5.11: log(I_D)-V_{GS} of 2 nm-, 7 nm- and 12 nm-thick spacer devices with \(~1 \mu m-L_g\).

Figure 5.12(a) shows the peak \(g_m\) as a function of \(L_g\). 12 nm-thick spacer devices exhibit \(~7\)-10% reduction in the peak \(g_m\) at shorter \(L_g\)'s. This can be attributed to the increased effective channel length, which can be also interpreted as an added resistance from the thicker spacer. Figure 5.12(b) compares \(R_{on}\) vs. \(L_g\). \(R_{on}\) of 2 nm-, 7 nm-, and 12 nm-thick spacer devices, extrapolated to zero \(L_g\), are 183, 212, and 242 \(\Omega\cdot\mu m\), respectively; hence a 12 nm-thick spacer adds 59 \(\Omega\cdot\mu m\) to \(R_{on}\) compared to a 2-nm-thick spacer. It is noted that \(R_{on}\) also includes \(~85\ \Omega\cdot\mu m\) S/D access resistance, as computed from the regrown N+ S/D sheet resistance and 1.2 \(\mu m\)-wide spacing between the channel and the S/D contacts.
Figure 5.12: (a) Peak $g_m$ as a function of $L_g$ at $V_{DS} = 0.5$ V, and (b) $R_{on}$ at $V_{GS} = 1.0$ V for 2 nm-, 7 nm- and 12 nm-thick spacer devices.

Figure 5.13 benchmarks recently published results on III-V MOSFETs. Firstly, Figure 5.13(a) compares the peak $g_m$ with respect to $SS_{min}$. A device with 2 nm-thick spacer and 18 nm-$L_g$ shows a peak $g_m$ of 3.0 mS/µm at $V_{DS} = 0.5$ V, which is the highest among all types of MOSFETs published to date [12, 36, 50, 51, 54, 72, 75–78]. The device with a 12 nm-thick spacer exhibits an excellent peak $g_m$ of 2.5 mS/µm as well as $SS_{min}$ of 86 mS/µm, in spite of a short $L_g$ of 40 nm and the planar ultra-thin-body architecture. Secondly, Figure 5.13(b) compares $I_{on}$ at a fixed $I_{off} = 100$ nA/µm and $V_{DD} = 0.5$ V. The 2 nm-thick spacer devices are not benchmarked here, since they have a significantly lower $I_{on}$ at a fixed $I_{off}$ due to the high off-state leakage at high $V_{DS}$, despite their superior transconductance. On the other hand, the 12 nm-thick spacer devices have $> 450$ µA/µm at sub-50 nm-$L_g$, which is the highest reported $I_{on}$ among all III-V MOSFETs to date.
5.5 Conclusions

This chapter has discussed in detail improvements made on the overall off-state performance. Firstly, to reduce the off-state leakage related to BTBT at the high drain-field region as well as to improve 2-D electrostatic control, undoped vertical spacers have been incorporated between the channel and the N+ S/D. Without increasing the S/D contact pitch, vertical spacers reduce the drain field and increase the effective channel length, therefore significantly reducing the off-state leakage at a high drain bias as well as the short channel effect. Next, to suppress the buffer/back barrier leakage, a P-doped barrier layer has been added, which has enabled suppressing the off-leakage below $10^{-5}$ mA/µm. Comparisons among the devices with 2 nm-, 7 nm- and 12 nm-thick vertical spacers suggest that increasing the vertical spacer thickness up to 12 nm substantially improves
the subthreshold characteristics such as $SS$, $DIBL$, and the off-leakage, with only a marginal degradation in $g_m$ ($\sim 10\%$).
In the previous chapter, significant reductions in the off-state leakage as well as in $SS$ have been achieved by implementing a vertical spacer and P-doped back barrier. Thus, further improved FETs have been fabricated, with a peak $g_m$ of $> 2.4$ mS/µm at $V_{DS} = 0.5$ V, off-state leakage of $< 10^{-7}$ A/µm, and minimum $SS$ of $< 90$ mV/decade at a short gate length of 40 nm, as shown previously in figure 5.9. However, the subthreshold slope roll-off observed near the off-state $V_{GS}$ at $V_{DS} = 0.5$ V shown in figure 5.9(b) is a remaining obstacle to obtaining a better on-current at a fixed off-current, which is one of the most critical figures of merit for CMOS logic devices.

As discussed in the previous chapter, the off-state leakage is mainly attributed
to BTBT and/or impact ionization, where the primary mechanism for inducing leakage is identical in both cases. Therefore, adopting a thicker vertical spacer as well as a larger bandgap channel becomes attractive in this regard. However, it is found that further increasing the vertical spacer thickness substantially degrades the on-state performance by adding an unacceptably high parasitic S/D resistance. Therefore, the focus of this generation lies on vertically scaling the composite-channel design, in which the larger quantized bandgap should suppress leakage due to BTBT or impact ionization by increasing the tunneling barrier height. Also, the enhanced aspect ratio of the channel thickness to length should further improve the 2-D electrostatics. However, it is important to perform lateral scaling of the channel simultaneously with vertical scaling. Otherwise, degradation of channel mobility induced by surface scattering may compromise the on-state performance. Furthermore, increasing the gate capacitance should enhance both the on- and off-state performances. In this regard, this particular generation closely examines how FETs behave with highly scaled design parameters such as $\sim 20$ nm-$L_g$, < 1 nm EOT, and $\sim 2.5$ nm-thick channel. Also, this chapter will convey comparison studies on some of the important design parameters such as ZrO$_2$ vs. HfO$_2$ high-$k$ dielectrics, 2.5 nm- vs. 5.0 nm-thick channel, and finally InAs- vs. In$_{0.53}$Ga$_{0.47}$As-based channel.
6.1 ZrO$_2$ High-$k$ Dielectric

![Comparison of electrical characteristics of In$_{0.53}$Ga$_{0.47}$As MOSCAPs with ~4 nm HfO$_2$ and ~4 ZrO$_2$. C-V characteristics as a function of frequency for (a) HfO$_2$ and (b) ZrO$_2$. The insets in (a) and (b) show C-V curve at 1 MHz. Normalized parallel conductance maps for (c) HfO$_2$ and (d) ZrO$_2$. [41, 42]

The dielectrics researchers in Stemmer group from the Materials Department at UCSB have developed a new high-$k$ dielectric based on ZrO$_2$, which has a
wider bandgap and larger dielectric constant than HfO₂, along with an optimized nitrogen plasma-based surface passivation technique that allows to maintain the same low $D_{it}$ as HfO₂ [42]. Figure 6.1(a) and (c) show electrical characteristics of MOSCAPs with $\sim$4 nm-thick HfO₂, and Figure 6.1(b) and (d) show those of MOSCAPs with $\sim$4 nm-thick ZrO₂. As shown in the $C-V$ plot in Figure 6.1(b), the accumulation capacitance density of ZrO₂ is $\sim$3.5 µF/cm² at 1 MHz, which is $\sim$15% larger than that of HfO₂. It is also the highest capacitance density among all of the III-V gate dielectrics reported to date. It is noted that the measured accumulation capacitance displays artifacts from the gate leakage at low frequencies, which is expected due to its small physical thickness. In addition to the high capacitance density measured from ZrO₂, the small frequency dispersion as well as a steep slope and small midgap $D_{it}$ hump at negative biases indicate $D_{it}$ is low and comparable to that of HfO₂. Quantitatively, its $D_{it}$ is in the $10^{12}$ cm$^{-2}$eV$^{-1}$ range near the midgap, which is calculated from the conductance map in Figure 6.1(d). It is noted that the data presented here are measured from MOSCAPs, thus measurement results from MOSFETs will be discussed later.

6.2 Device Structure and Fabrication

The device structure and process flow for this generation are mostly identical to those discussed in the 3rd generation. As mentioned above, the high-$k$ dielectric now consists of a 2.5 nm-thick ZrO₂ dielectric. To address a yield issue related to collapses of sub-20 nm-$L_g$ dummy gates, HSQ process is further optimized using an even smaller beam current of 100 pA and a thinner HSQ of $\sim$20 nm. More
importantly, the epitaxial InAs layer thickness in the composite-channel is reduced down to 3.5 nm from 6 nm in order to improve the off-state performance.

![Cross-section schematic of MOSFETs in the 4th generation.](image)

Figure 6.2: Cross-section schematic of MOSFETs in the 4th generation. [79]

The epitaxial layer structure, grown on a semi-insulating InP substrate by solid-source MBE, has a 50 nm UID InAlAs buffer, a 250 nm 1.0×10^{17} cm^{-3} P-type doped InAlAs barrier, a 100 nm UID InAlAs barrier, a 2 nm 1.0×10^{12} cm^{-2} N-type InAlAs pulse-doped layer, a 5 nm UID InAlAs setback, a 3.5 nm InAs (strained) channel and 2 nm of the UID In_{0.53}Ga_{0.47} spacer [79]. Prior to spinning HSQ, 1 nm-thick Al_{2}O_{3} is deposited using ALD to promote adhesion. To form dummy gates with lengths ranging from 12 nm to 1000 nm, ~20 nm of HSQ is spun and subsequently patterned by e-beam lithography. The vertical spacers and N+ S/D regions are formed by regrowing 10 nm UID (~1.2×10^{15} cm^{-3}) and 60 nm Si-doped (4.0×10^{19} cm^{-3}) In_{0.53}Ga_{0.47}As using MOCVD. The device mesas are isolated by wet etching, and the dummy gates are stripped off in buffered HF. Using a 2-cycle isotropic digital etching, ~2 nm of the In_{0.53}Ga_{0.47}As
cap and ~1 nm of InAs channel are removed from the gate regions, leaving a 2.5 nm-thick InAs channel. The samples are then immediately loaded into the ALD chamber. After an in-situ N\textsubscript{2} plasma/TMA pretreatment during which ~1 nm of Al\textsubscript{2}O\textsubscript{x}N\textsubscript{y} is formed, a ~2.5 nm ZrO\textsubscript{2} gate dielectric is deposited [42]. The samples are subsequently annealed in forming gas at 400 °C. Finally, Ni/Au gate and Ti/Pd/Au S/D metal contacts are deposited using thermal and e-beam evaporation technique, respectively. Figure 6.2 shows a cross-sectional schematic of a 4\textsuperscript{th} generation device. Its epitaxial layers and main process specifications are listed in Figure 6.3.

<table>
<thead>
<tr>
<th>Epitaxial Layers (By MBE)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap : 2 nm, U.I.D, In\textsubscript{0.53}Ga\textsubscript{0.47}As</td>
<td></td>
</tr>
<tr>
<td>Channel : 3.5 nm Strained InAs</td>
<td></td>
</tr>
<tr>
<td>Setback : 5 nm, U.I.D, In\textsubscript{0.52}Al\textsubscript{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Pulse doping : 2 nm, 2.0\times10\textsuperscript{19} cm\textsuperscript{-2} Si-doped, In\textsubscript{0.52}Al\textsubscript{0.48}As</td>
<td></td>
</tr>
<tr>
<td>P-doped barrier : 250 nm, 1.0\times10\textsuperscript{17} cm\textsuperscript{-3} Be-doped, In\textsubscript{0.52}Al\textsubscript{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Buffer : 50 nm, U.I.D, In\textsubscript{0.52}Al\textsubscript{0.48}As</td>
<td></td>
</tr>
<tr>
<td>Substrate : Semi-insulating InP</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummy Gate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>~20 nm HSQ patterned by e-beam lithography</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regrowth (By MOCVD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N\textsuperscript{+} S/D : 60 nm, 4\times10\textsuperscript{19} cm\textsuperscript{-3} Si-doped, In\textsubscript{0.53}Ga\textsubscript{0.47}As</td>
<td></td>
</tr>
<tr>
<td>Vertical Spacer : 10 nm, U.I.D, In\textsubscript{0.53}Ga\textsubscript{0.47}As</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High-k dielectric</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 nm Al\textsubscript{2}O\textsubscript{x}N\textsubscript{y} by TMA/N\textsubscript{2} plasma passivation / 2.5 nm ZrO\textsubscript{2}</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.3: Key specifications for the 4\textsuperscript{th} generation devices.

The FETs have $L_g$ ranging from 18 nm to 1 \( \mu \)m, where $L_g$ is defined as the spacing between the edges of the regrown layers. The sidewalls of the vertical spacers are also gated, thus the effective gate length is ~35 nm greater than $L_g$ itself, which is an approximated value based on the given spacer thickness and regrowth angle. In the production of VLSI devices, the increase in the minimum S/D contact pitch is determined by the thicknesses of the spacer and N+ regions...
as well as the regrowth angle; if the regrown regions can be completely vertical, the presence of the spacers should not increase the minimum S/D pitch.

### 6.3 Results and Discussion

![Cross-sectional STEM image of a FET with a ZrO$_2$ high-$k$.](image)

Figure 6.4: The cross-sectional STEM of the fabricated device. ~1.0 nm Al$_2$O$_y$N$_y$ interfacial layer, ~2.5 nm ZrO$_2$ and ~2.7 nm InAs channel are shown. Heavy elements look brighter. Image courtesy of Stephan Krämer at UCSB.

Figure 6.4 shows a cross-sectional STEM image of a FET with a ZrO$_2$ high-$k$. The STEM shows a 2.7 nm-thick InAs channel thickness, which is in agreement with the expected thickness of ~2.5 nm based on the calibrated etch rate obtained from digital etching as well as InAs growth rate by MBE. The high resolution images shown on the right-hand side of Figure 6.4 confirm that the gate dielectric
stack consists of a 1.0 nm-thick Al$_2$O$_x$N$_y$, which is likely formed during the N$_2$ plasma pretreatment, and a 2.5 nm-thick ZrO$_2$. There is no sign of defects at the interface between the MBE grown InGaAs cap and the MOCVD regrown N+ InGaAs S/D, as expected.

Figure 6.5: log($I_D$) and log($I_G$) as a function of $V_{GS}$ for 1$\mu$m-$L_g$ FET at $V_{DS}$ = 0.1 V and 0.5 V.

Figure 6.5 shows a subthreshold characteristic and normalized gate leakage for a 1 $\mu$m-$L_g$ device. Its subthreshold slope is 61 mV/decade at $V_{DS}$ = 0.1 V, which is the lowest reported value from any III-V MOSFETs. Since 2-D electrostatics and BTBT related leakage are negligible in long channel devices, any degradation observed in the subthreshold slope can be attributed to the existence of interface traps. Using its EOT of $\sim$0.85 nm, the $D_{it}$ is calculated to be $\sim$7$\times$10$^{11}$ cm$^{-2}$·eV$^{-1}$ by using $D_{it} = C_{EOT}/q\cdot(\text{SS}/2.3kT/1/q)$, where $C_{EOT}$ is the gate oxide capacitance density. This calculated $D_{it}$ agrees with $D_{it}$ estimated using the conductance
method on MOSCAPs discussed in the previous section. In addition to the low SS, the negligible DC hysteresis implies a high-quality gate dielectric. The gate leakage normalized by the gate metal overlap area is < 1 A/cm² at all measured bias conditions.

Figure 6.6: (a) Transfer characteristics ($I_D$-$V_{GS}$ and $g_m$-$V_{GS}$) as a function of $V_{GS}$, (b) subthreshold characteristic (log($I_D$)-$V_{GS}$), and (c) common-source characteristic ($I_D$-$V_{DS}$) of a FET with 25 nm-$L_g$. 

\[ L_g = 25 \text{ nm} \]
\[ I_m = 500 \mu A/\mu m \]
\[ V_{GS} = 0.1 \text{ to } 0.7 \text{ V} \]
\[ 0.2 \text{ V increment} \]
\[ DIBL = 76 \text{ mV/V} \]
\[ \text{at } V_{GS} = 0.1 \text{ V} \]
\[ SS_{min} \approx 72 \text{ mV/dec. (at } V_{GS} = 0.1 \text{ V) } \]
\[ \text{SS}_{min} \approx 77 \text{ mV/dec. (at } V_{GS} = 0.5 \text{ V) } \]
\[ V_{DS} = 0.1 \text{ to } 0.7 \text{ V} \]
\[ 0.2 \text{ V increment} \]
\[ R_{on} = 303 \Omega \mu m \]
\[ \text{at } V_{GS} = 0.7 \text{ V} \]
\[ V_{GS} = -0.4 \text{ V to } 0.7 \text{ V} \]
\[ 0.1 \text{ V increment} \]

102
Figure 6.6(a) shows transfer characteristics of a 25 nm-$L_g$ FET at a varied range of applied $V_{DS}$. It shows a peak $g_m$ of 2.38 mS/µm at $V_{DS} = 0.5$ V and an $I_{on}$ of 0.5 mA/µm at $I_{off} = 100$ nA/µm and $V_{DD} = 0.5$ V. Its subthreshold characteristics (Figure 6.6(b)) show $SS_{min}$ of 72 mV/decade at $V_{DS} = 0.1$ V and 77 mV/decade at $V_{DS} = 0.5$ V. Since BTBT related leakage is well suppressed by adopting a thinner channel, the subthreshold slope for $V_{DS} = 0.5$ V stays constant nearly all the way down to 100 nA/µm off-current limit for high performance applications. Defined at $I_D = 1$ µA/µm, its threshold voltage is -85 mV and drain induced barrier lowing ($DIBL$) is 76 mV/V. Common-source characteristics (Figure 6.6(c)) show $\sim 300$ Ω-µm on-resistance at $V_{GS} = 0.7$ V. On the other hand, the current kink observed near a relatively small applied $V_{DS}$ of 0.5 V confirms that electron-hole pair generation from either BTBT or impact ionization is happening in the small bandgap channel, as mentioned in the previous chapter. However, since the target $V_{DD}$ for III-V FETs is as small as 0.5 V, such non-ideal effects are adequately managed in FETs from this generation.
Figure 6.7: (a) Illustration of which resistance components contribute to $R_{on}$ at zero-$L_g$. (b) $R_{on}$ as function of $L_g$ at $V_{GS} = 0.7$ V. (c) TLM measurement of N+ regrown S/D. (d) Calculated ballistic contact resistance as a function of $V_{GS}$ at $m^*/m_0$ range from 0.04 to 0.1.

In Figure 6.7 analyzes the S/D series resistance in the FETs. Figure 6.7(a) shows normalized output resistances ($R_{on}$) as a function of $L_g$. $R_{on}$ extrapolated to zero-$L_g$ is $\sim220 \ \Omega-\mu$m (± 4.7 $\Omega-\mu$m standard deviation from the Y-intercept of the fitted line) at $V_{GS} = 0.7$ V. Using a lumped element model, it can be simplified...
to four resistance components: (1) metal-semiconductor resistance \( R_{contact} \), (2) access resistance for N+ S/D \( R_{N+S/D} \), (3) resistance from the undoped vertical spacer \( R_{spacer} \), and (4) ballistic contact resistance \( R_{ballistic} \), as illustrated in Figure 6.7(b). Figure 6.7(c) shows measured TLM for N+ S/D; the sheet resistance of \( \sim60 \) nm-thick N+ S/D is \( \sim25 \) \( \Omega \)-square, and metal-semiconductor specific contact resistivity is \( \sim5 \) \( \Omega\cdot\mu m^2 \). Considering the spacing of approximately \( 1.2 \mu m \) between the S/D metal contact and the channel edge, the S/D access resistance, which includes \( R_{N+S/D} \) and \( R_{contact} \), is determined to be \( \sim85 \) \( \Omega\cdot\mu m \) for both the source and drain sides. As discussed in Chapter 2.1, based on the ballistic FET theory, the ballistic contact resistance is estimated with an isotropic effective mass \( m^* / m_0 \) in the range of 0.02-0.1 (for quantum-well based on InGaAs) as shown in Figure 6.7(d). At a highly degenerate regime of \( V_{GS} > 0.5 \) V, it is calculated to be 55-65 \( \Omega\cdot\mu m \). Three resistance components, \( R_{contact} \), \( R_{N+S/D} \), and \( R_{ballistic} \) out of the four, are now determined from either measurement or numerical calculation; thus \( R_{spacer} \) is estimated to be \( \sim75 \) \( \Omega\cdot\mu m \). Here, it should be noted that the increase in resistance due to any quantum mechanical reflection at the In\(_{0.53}\)Ga\(_{0.47}\)As spacer/InAs channel heterointerface is not taken into account. However, in such a thin InAs quantum-well channel, actual conduction band discontinuity between the InGaAs S/D and the InAs channel will not be severe. This is because the 1st quantized state in the InAs channel is positioned \( \sim0.3 \) eV above the conduction band edge (See Figure 6.17(b)) while InGaAs S/D is not quantized. Figure 6.8 summarizes the contributions from each resistance to the total \( R_{on} \) at zero-\( L_g \).
Figure 6.8: Summary of the factors contributing to the total S/D parasitic series resistance.

Figure 6.10(a) shows $SS_{\text{min}}$ at $V_{DS} = 0.1$ and 0.5 V as a function of $L_g$ and compares it to other published III-V MOSFETs [12, 36, 50, 51, 54, 75–78]. Open and solid symbols represent $V_{DS} = 0.1$ and 0.5 V, respectively. For gate lengths down to 30 nm, $SS_{\text{min}}$ at the high $V_{DS}$ is smaller than 70 mV/decade and similar with that at low $V_{DS}$, indicating an excellent short channel immunity. This can be attributed to the increased aspect ratio by adopting a thinner channel, increased effective gate length by incorporating a 12 nm-thick spacer, and an adequately suppressed BTBT by increasing the quantized bandgap. Comparing with results from the published III-V literatures, $SS_{\text{min}}$ measured from this work is the lowest at all $L_g$. Figure 6.10(b) shows $DIBL$ as a function of $L_g$ from this work as well as from other reported III-V MOSFETs with both the planar and non-planar architectures. The FETs from this work exhibit $\sim 100$ mV/V at 20 nm-$L_g$, which is the lowest $DIBL$ among any planar MOSFETs.
Figure 6.9: (a) $SS_{\text{min}}$ as a function of $L_g$ and (b) $DIBL$ as a function of $L_g$ compared to published III-V MOSFETs.

Figure 6.10(a) shows peak $g_m$ at $V_{DS} = 0.5$ V as function of $L_g$ compared with other published III-V MOSFETs [12, 36, 50, 51, 54, 72, 75–78]. Despite their low mobility in the ultra-thin channel, sub-20 nm-$L_g$ devices exhibit a peak $g_m$ of $>2.4$ mS/µm at $V_{DS} = 0.5$ V, which is the highest peak $g_m$ for any FETs with a similar $L_g$. This is likely due to the high gate-channel capacitance due to the thin dielectric and channel. The results obtained on the mobility and gate-capacitance of these FETs will be discussed in the following section. In Figure 6.10(b), on-current ($I_{on}$), defined at 100 nA/µm off-current ($I_{off}$) and 0.5 V supply voltage, is benchmarked with other published III-V MOSFETs [12, 36, 50, 51, 75–78]. At $\sim$25 nm-$L_g$, FETs have $I_{on}$ of 0.5 mA/µm, which is the highest $I_{on}$ and $>2:1$ larger than the previous best device results obtained at VLSI-relevant gate lengths (sub-30 nm). At very short gate lengths, the degradation in the subthreshold characteristics such as $SS$ and $DIBL$ becomes substantial, but $g_m$ hardly increases by the gate length scaling, hence resulting in lower $I_{on}$ for $L_g < 25$ nm.
Figure 6.10: (a) Peak $g_m$ as a function of $L_g$ (b) $I_{on}$ as a function of $L_g$ at $V_{DD} = 0.5$ V and $I_{off} = 100$ nA/µm, compared to other III-V results in literatures.

Lastly, in Figure 6.11, III-V FETs from this work are benchmarked with the state-of-the-art Si MOSFETs with 3-D architectures. Intel 22 nm-node Si FinFET (Figure 6.11(a)) for high performance applications shows $\sim 0.5$ mA/µm $I_{on}$ at $I_{off} = 100$ nA/µm, $V_{GS} = 0.5$ V and $V_{DS} = 0.75$ V. IBM 22 nm-node Si nanowire FET (figure 6.11(b)) shows $\sim 0.4$ mA/µm $I_{on}$ at $I_{off} = 100$ nA/µm, $V_{GS} = 0.5$ V and $V_{DS} = 0.75$ V. Therefore, FET results obtained in this work is the first demonstration of III-V FET with a comparable performance with the state-of-the-art Si multi-gate FETs [43,44].
Figure 6.11: (a) $I-V$ characteristics for (a) Intel 22 nm-node Si FinFETs [43] and (b) IBM 22 nm-node Si nanowire FETs [44].
6.4 Comparison Study

6.4.1 5.0 nm- vs. 2.5 nm-thick InAs channel

Figure 6.12: Comparison of \( I-V \) characteristics for \( \sim 30 \text{ nm-} L_g \) FETs with 5.0 nm- and 2.5 nm-thick InAs channel.
This subsection compares FETs with 5.0 nm- and 2.5 nm-thick InAs channel. FETs with a 2.5 nm-thick InAs channel are identical to FETs introduced so far in this generation. FETs with a 5.0 nm-thick InAs channel have been fabricated using the exactly same process flows in regards to the epitaxial layer design (except for the channel), spacer thickness, and high-$k$ dielectric. Figure 6.12 shows plots of $I_D-V_{GS}$, $\log(I_D)-V_{GS}$, and $I_D-V_D$ for 5.0 nm- and 2.5 nm-thick InAs channel FETs at $\sim$30 nm-$L_g$. Since the thinner channel FET has a more positive threshold voltage, it is noted that comparisons of parameters at certain $V_{GS}$ become less meaningful. The 2.5 nm InAs FET shows $\sim$10% lower peak $g_m$ of $\sim$2.2 mS/µm. On the other hand, when comparing $\log(I_D)-V_{GS}$, the 2.5 nm-thick InAs channel FET shows significantly better off-state characteristics including $SS$, $DIBL$, and $I_{off}$. For the 5.0 nm-thick InAs FET, it is observed that $SS$ is rolling off at a far earlier stage of the off-state, resulting in $I_{off}$ that does not reach 100 nA/µm.

![Figure 6.13: $SS_{min}$ as a function of $L_g$ for 2.5 nm- and 5.0 nm-thick InAs channel devices at $V_{DS} = 0.1$ and 0.5 V.](image)
Figure 6.13 shows $SS_{\text{min}}$ at $V_{DS} = 0.1$ and 0.5 V as a function of $L_g$, for 5.0 nm- and 2.5 nm-thick InAs FETs. Open and solid symbols represent $V_{DS} = 0.1$ and 0.5 V, respectively. In the case of the 5.0 nm-thick InAs FET, even at longer gate lengths where 2-D electrostatic effect is nearly negligible, $SS_{\text{min}}$ at high $V_{DS}$ is much larger than that at low $V_{DS}$. This degradation in $SS_{\text{min}}$ can be attributed to an increase in BTBT-related leakage due to the smaller bandgap in the thicker channel. As opposed to the 5.0 nm-thick InAs FET, the 2.5 nm-thick InAs FET shows a similar $SS_{\text{min}}$ at both low and high $V_{DS}$ all the way down to $\sim 30$ nm-$L_g$. Such a dramatic reduction in $SS_{\text{min}}$ seen in the 2.5 nm-thick InAs FET at shorter gate lengths can be attributed to reduced BTBT as well as improved 2-D electrostatics from the larger channel length/thickness ratio.

![Figure 6.13: SS vs Lg for InAs FETs](image)

Figure 6.14: Comparisons on subthreshold characteristics and gate leakage for 2.5 nm- and 5.0 nm-thick InAs channel devices with $L_g = 500$ nm at $V_{DS} = 0.1$ and 0.5 V.

Figure 6.14 compares $\log(I_D)-V_{GS}$ and gate leakage of 2.5 nm- and 5.0 nm-
thick InAs FETs with 500 nm-$L_g$. While the 2.5 nm-thick InAs FET shows a \(~\sim\)10:1 reduction in the minimum off-state leakage, its gate leakage is \(~\sim\)5:1 larger than the 5.0 nm-thick InAs FET. This is likely due to the raised eigenstate in the thinner quantum well lowering the effective tunneling barrier to gate leakage.

![Figure 6.15: Peak $g_m$ as a function of $L_g$ for 2.5 nm- and 5.0 nm-thick InAs channel devices at $V_{DS} = 0.5$ V.](image)

Figure 6.15 compares peak $g_m$ at $V_{DS} = 0.5$ V as a function of $L_g$. As expected, for long gate lengths ($>100$ nm), 5.0 nm-thick InAs FETs exhibit 1.5-3 times larger peak $g_m$ compared to 2.5 nm-thick InAs FETs. This is primarily due to the lower channel mobility resulting from strong surface scattering in the thinner channel FETs. To quantify this effect, effective channel mobilities are extracted as shown in Figure 6.16. At a carrier density of $\sim3\times10^{12}$ cm$^{-2}$, the effective channel mobility of 5.0 nm-thick InAs FETs is about $\sim1100$ cm$^2$/s$\cdot$V$^{-1}$, which is 5:1 larger than that of 2.5 nm-thick InAs FETs.
Figure 6.16: Low field effective mobility as a function of the carrier density for 2.5 nm- and 5.0 nm-thick InAs channel devices.

The effective mobility is extracted from a split-C-V measurement for large active area FETs (FATFETs) with a 21 $\mu$m $\times$ 25 $\mu$m active area (Figure 6.17(a)) and output conductance measurement at low $V_{DS}$ [80]. A few important underlying assumptions should be pointed out. Firstly, in the split-C-V measurement, the oscillation frequency should be carefully chosen, considering the $RC$ charging delay of the tested FETs. In essence, the measured capacitance becomes no longer valid if the oscillation frequency is greater than the FET $RC$ delay. Due to the long gate lengths of the measured FATFETs, the channel resistance is very large, resulting in a large $RC$ delay. In this light, the oscillation frequency of 200 kHz is picked for this measurement. Secondly, it should be noted that the effective mobility can be under-estimated by the presence of $D_{it}$. In the previous section, however, it has been proven that $D_{it}$ effect is negligible, since $SS$ at long gate lengths nearly approaches 60 mV/decade. Lastly, the gate metal source-drain overlap is assumed to be negligible compared to the channel area. This is a valid
assumption, because the overlap is in the order of 200 to 400 nm, as shown in the TEM of Figure 6.4, while the gate length of FATFETs tested is $\sim 21 \ \mu m$.

As observed in Figure 6.15, the peak $g_m$ from 2.5 nm-thick InAs FETs becomes almost comparable to that of 5.0 nm-thick InAs FETs as scaling $L_g$ down to sub-30 nm. It can be inferred that at such a short gate length regime, the gate capacitance density and source injection velocity affect the on-state performance much more significantly than the effective mobility. Figure 6.17 compares the gate capacitance as a function of $V_{GS}$. The accumulation capacitance density for 2.5 nm-thick InAs FETs is $\sim 10\%$ larger than that for 5.0 nm-thick InAs FETs, since the wave-function in the thinner quantum-well is closer to the surface than in a thicker well. Figure 6.17(b) compares the energy band diagram and wave-function profiles simulated using a 1-D poisson-schrödinger solver. Under the same condition of $E_F - E_1 = 0.18$ eV, their wave-functions are located approximately at the center of the quantum-well. The change in total gate capacitance from the wave-function depth should become more significant as thinning down the high-$k$ dielectric and as adopting channel materials with a larger density of states effective mass.
Figure 6.17: (a) Gate-channel capacitance and carrier density for 2.5 nm- and 5.0 nm-thick InAs channel devices from split-$C-V$ measurements. (b) Comparison on 1st eigenstate wave-functions of 2.5 nm- and 5.0 nm-thick InAs channel devices.
6.4.2 HfO$_2$ vs. ZrO$_2$ high-$k$ dielectric

Figure 6.18: Comparison of $I$-$V$ characteristics for $\sim$25-28 nm-$L_g$ FETs with HfO$_2$ and ZrO$_2$ high-$k$ dielectrics.
This subsection compares FETs with two different high-$k$ dielectrics: HfO$_2$ and ZrO$_2$. The two FETs have been fabricated on the same wafers with the identical layer design, same process flows, and design parameters, which are summarized in Figures 6.2 and 6.3. The pretreatment condition in ALD has also been identical, except for the choice material for the high-$k$ itself. $I_D-V_{GS}$, log$(I_D)$-V$_{GS}$, and $I_D-$V$_D$ for $\sim$25 nm-$L_g$ FETs with ZrO$_2$ and HfO$_2$ high-$k$ dielectrics are shown in Figure 6.18. For the on-state performance, the ZrO$_2$ FET show slightly a larger peak $g_m$ and $I_{on}$. Considering a small positive $V_T$ shift in the HfO$_2$ FET, the off-state leakages are similar for both cases. As for the subthreshold slope, the ZrO$_2$ FETs show a slightly better $SS_{min}$ at high $V_{DS}$. Given the variability of measured FET samples, which is $\sim$10%, it is difficult to conclude superiority between the two dielectrics in terms of the device performance.

Figure 6.19: (a) Peak $g_m$ as a function of $L_g$ for devices with $\sim$3 nm HfO$_2$ and ZrO$_2$ high-$k$ dielectrics at $V_{DS} = 0.5$ V (b) $SS_{min}$ as a function of $L_g$ devices with $\sim$3 nm HfO$_2$ and ZrO$_2$ high-$k$ dielectrics at $V_{DS} = 0.1$ (open) and 0.5 V (solid).
The plot of peak $g_m$ against $L_g$ shown in Figure 6.19(a) re-confirms there is no noticeable difference in the on-state performance between ZrO$_2$ shown by blue circles and HfO$_2$ shown by green triangles. Figure 6.19(b) compares $SS_{min}$ as a function of $L_g$. HfO$_2$ FETs show a slightly larger $SS_{min}$ at $V_{DS} = 0.5$ V. This can be attributed to the higher dielectric constant in ZrO$_2$, which enhances the 2-D electrostatic control. As mentioned in Chapter 2.1.3, in a subthreshold condition, the 2-D electrostatics are only affected by the oxide capacitance ($C_{ox}$), interfacial trap capacitance ($C_{it}$) (if it is non-negligible), and gate-drain capacitance($C_{GD}$); thus a larger dielectric constant should result in improving the 2-D electrostatics regardless of $C_{DoS}$ and $C_{Depth}$. However, the observed differences may just be due to the variability in devices, so the above statements are not yet conclusive.

Figure 6.20: (a) Gate-channel capacitance, carrier density from split-C-V measurements and (b) normalized gate leakage for devices with ~3 nm HfO$_2$ and ZrO$_2$ high- $k$ dielectrics.

Figure 6.20(a) shows the gate capacitances as a function of $V_{GS}$, which are measured at 200 kHz on the 21 $\mu$m $\times$ 25 $\mu$m large area devices. Considering the
positive $V_T$ shift in the HfO$_2$ FETs, the gate capacitance is nearly the same as ZrO$_2$. According to a literature [42], the dielectric constants of ZrO$_2$ and HfO$_2$ are 23 and 17, respectively. Therefore, a $\sim$26% increase in the oxide capacitance can be expected by adopting ZrO$_2$ instead of HfO$_2$. On the other hand, the total gate capacitance ($C_{g-c}$) of FETs under accumulation is determined by the (1) oxide capacitance including both the interlayer and high-$k$ dielectric, (2) wavefunction capacitance, as well as (3) density of state capacitance in series. From a calculation with appropriate device metrics, switching HfO$_2$ to ZrO$_2$ high-$k$ is estimated to result in an increase of only $\sim$8% in $C_{g-c}$. Lastly, Figure 6.20(b) compares the normalized gate leakages, where the ZrO$_2$ FET shows a $\sim$5:1 larger gate leakage.

As a side note, in Figure 6.19(a) and (b), the red squares represent FETs with 3 nm-thick ZrO$_2$, which have been introduced in the previous sections, and the blue circles represent “reproduced” FETs that are fabricated based on exactly the same epitaxial layer design (but grown at a different lot), process flow, and design parameters. The similar peak $g_m$ and $SS_{min}$ measured throughout all gate lengths suggest that the epitaxial growth and FET processes done at UCSB are highly reproducible.
6.4.3 $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ vs. InAs channel

Figure 6.21: Comparison of $I$-$V$ characteristics for $\sim 40$ nm-$L_g$ FETs with $\sim 3$ nm-thick $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ and InAs. Measurement data courtesy of Cheng-Ying Huang at UCSB.
In order to evaluate device performances in terms of the channel composition, two samples with 3 nm-thick In$_{0.53}$Ga$_{0.47}$As and 3 nm-thick InAs channel are simultaneously fabricated using the most updated process flow introduced in this chapter. Both samples have 3 nm-thick HfO$_2$ high-$k$ and 13 nm-thick vertical spacers. For this lot, the samples have been fabricated and measured by Cheng-Ying Huang. Note that in the 1$^{st}$ generation, similar studies have already taken place with InAs/InGaAs composite-channel FETs with various InAs thicknesses. Yet, those FETs have been fabricated using immature process flows and device design; hence they suffer from a $>10^{-3}$ mA/µm high off-leakage, which complicates the analysis.

Figure 6.21 compares $I_D$-$V_{GS}$, log($I_D$)-$V_{GS}$, and $I_D$-$V_D$ for 40 nm-$L_g$ FETs with a 3 nm-thick InGaAs and a 3 nm-thick InAs channel. Regarding the off-leakage, both FETs exhibit excellent subthreshold behaviors and a very low minimum leakage current of $<10$ nA/µm. In particular, the minimum leakage in the InGaAs channel FET is as small as 1 nA/µm at $V_{DS} = 0.5$ V, which meets the leakage requirement of the general purpose VLSI applications. On the other hand, InAs channel FET shows a much higher peak $g_m$ of $\sim2.0$ mS/µm when compared to the InGaAs channel FET. According to [32], the rate of increase in the in-plane effective mass as decreasing the width of the quantum-well is highly dependent on the indium composition of the quantum-well due to the change in nonparabolicity. Therefore, the in-plane effective mass of an InAs quantum-well would become comparable to that of InGaAs when its quantum-well width decreases down to $\sim2$ nm. Therefore, such a large discrepancy in peak $g_m$ observed between InAs and InGaAs FETs with a thin channel thickness of 3 nm is yet to be explained.
Figure 6.22: (a) Peak $g_m$ as a function of $L_g$ at $V_{DS} = 0.5$ V, and (b) $SS_{min}$ as a function of $L_g$ at $V_{DS} = 0.1$ (open) and 0.5 V (solid) for ~40 nm-$L_g$ FETs with ~3 nm-thick In$_{0.53}$Ga$_{0.47}$As and InAs. Measurement data courtesy of Cheng-Ying Huang at UCSB.

Figure 6.22(a) shows peak $g_m$ as a function of $L_g$ at $V_{DS} = 0.5$ V. InAs FETs exhibit $> 1.8 \times$ higher $g_m$ than InGaAs FETs throughout all gate lengths. Figure 6.22(b) compares $SS_{min}$ as a function of $L_g$ at $V_{DS} = 0.1$ V (open) and 0.5 V (solid). It is found that both InGaAs and InAs FETs have nearly identical $SS_{min}$. These results suggest that adopting a lower indium content channel offers no benefit in the subthreshold characteristics, but it only degrades the on-state performance. Yet, it is noted that this comparison is targeting for high performance application, where the off-leakage limit is 100 nA/µm. For general purpose or low power applications, InAs FETs will exhibit worse $I_{on}$ at a fixed $I_{off}$, since its subthreshold slope starts to degrade near 100 nA/µm $I_{off}$ and its off-leakage does not drop below 10 nA/µm $I_{off}$ as shown in Figure 6.21.
6.5 Conclusions

This chapter has discussed a device with a vertically scaled channel design and a ZrO$_2$ high-$k$ dielectric. Adopting a thinner channel helps suppress the BTBT-related off-leakage by enlarging the quantized bandgap and simultaneously enhance the short channel immunity by increasing the aspect ratio. Implementing a high capacitance density ZrO$_2$ high-$k$ dielectric with a very low $D_{it}$ of $<\sim 10^{12}$ cm$^{-2}$ has enabled achieving a nearly ideal SS in the long-channel devices. A device with a 2.5 nm-thick InAs channel and a 2.5 nm-thick ZrO$_2$ high-$k$ at a gate length of 25 nm has shown record-setting performances in both the on-state and off-state, featuring a 2.4 mS/$\mu$m peak $g_m$, 77 mV/decade minimum subthreshold swing at $V_{DS} = 0.5$ V, and 500 $\mu$A/$\mu$m on-current at a fixed 100 nA/$\mu$m off-current and $V_{DD} = 0.5$ V. In addition, this chapter has also investigated the impact of critical device metrics such as the channel thickness, high-$k$ dielectrics, and composition of the channel material on $I$-$V$ characteristics.
Chapter 7

Device Analysis: Ballisticity

By ruling out the critical non-ideal effects due to BTBT-related off-leakage and degradation in $SS$ from interfacial traps, FETs in the 4th generation exhibit “textbook-like” $I$-$V$ characteristics at VLSI-compatible gate lengths and a target supply voltage of $<0.7$ V. Hence, calculations based on the ballistic FET model can be reliably executed using the design parameters of these high-performance FETs. These III-V FETs have achieved a comparable value of $I_{on}$ at a fixed $I_{off}$ to state-of-the-art silicon multi-gate FETs, which are reportedly operating within 60-80% ballistic regime [18, 19]. Therefore, it would be meaningful to calculate and compare the ballisticity of the III-V FETs with their Si counterparts.
Figure 7.1 plots the current density at $V_{GS} - V_T = 0.4$ V as a function of $1/L_g$ for a 2.5 nm-thick InAs FET as described in the 4th generation. Here, $V_T$ is defined by the linear extrapolation method. The details on $V_T$ extraction and determination of the gate overdrive voltage ($V_{GS} - V_T$) will be discussed later in this chapter. The FETs with a long gate length of > 200 nm are well fitted with the mobility-limited transport FET model (solid red line) represented by $\frac{1}{2}\mu_{eff}C_{g-c}(V_{GS} - V_T)^2$ [16], where the effective mobility ($\mu_{eff}$) and total gate-channel gate capacitance ($C_{g-c}$) are determined to be 250 cm$^2$/V·s$^{-1}$ and 2.5 μF/cm$^2$, respectively, by the split- $C$-$V$ shown in Figure 6.17 and output conductance at a low $V_{DS}$. As $L_g$ is shrunk to less than 200 nm, the current density starts to deviate from the mobility-limited current model and eventually saturates around $L_g < 25$ nm, which strongly
indicates that the FETs with $L_g < 25$ nm are approaching the ballistic transport regime.

7.1 Gate Capacitance

As discussed in Chapter 2.1, in order to produce a valid $I$-$V$ calculation using the ballistic FET model, it is essential to know a correct value of the effective mass of the channel. This is mainly because the effective mass plays a dominant role in determining the density of states capacitance ($C_{DoS}$) as well as the injection velocity ($v_{inj}$). The electron effective mass in III-V is typically 2-5 times smaller than in Si. Also, the lowest band degeneracy in III-V FETs is 1, which leads to a small $C_{DoS}$. As shown in Figure 7.2, $C_{DoS}$ as a function of $V_{GS}$ is calculated using the room temperature Fermi-Dirac statistics for a reasonable range of $m^*/m_0$ for an $\text{In}_x\text{Ga}_{1-x}\text{As}$ quantum-well. In the given range of effective mass values, $C_{DoS}$ is
varying from \(-2.5\text{-}6.5 \, \mu F/cm^2\), which confirms that it is comparable with the oxide capacitance from \(<\sim 1 \, \text{nm EOT}\) as well as the wave-function depth capacitance of FETs with 2.5 nm-thick channel.

![Diagram](image)

**Figure 7.3:** (a) Calculated total gate capacitance as a function of \(V_{GS}\) under equilibrium at \(\frac{m^*}{m_0}\) ranging from 0.04 to 0.1. (b) Measured gate-channel capacitance (solid line) and carrier density (solid line + circular symbols) for a 2.5 nm-thick InAs channel FET.

Figure 7.3(a) plots the calculated \(C_{g-c}\) with \(\frac{m^*}{m_0}\) ranging from 0.04 to 0.1 with the following assumptions. Firstly, it is assumed that the centroid of electron wave-function is fixed at the center of the quantum-well regardless of \(V_{GS}\). This assumption is validated by 1-D Poisson-Schrödinger simulation results, which has shown that the shape of the quantum-well does not change much in the given range of \(V_{GS}\) (See Figure 7.5). Secondly, the dielectric thicknesses of 1 nm/2.5 nm Al₂O₃/ZrO₃ have been applied as shown in the TEM image in Figure 6.4. Also, the dielectric constants of 9/23 for Al₂O₃/ZrO₃ are assumed as determined from a thickness series study on MOSCAPs [42]. Thirdly, the interfacial and border traps are assumed to be negligibly small. This is a reasonable assumption
in consideration of the facts that (1) $SS_{min}$ at long channel FETs is close to 60 mV/decade and (2) there is no apparent frequency dispersion in the accumulation regime. Fourthly, for the simplicity of the calculation, nonparabolicity effects are not taken into account. Lastly, most electrons are assumed to be populating in the first subband, thus the population in the second subband becomes negligible. This will be discussed further in the next section.

By comparing the calculated total gate capacitance in Figure 7.3(a) with the measured shown in Figure 7.3(b), it is confirmed that the measured $C_{g-c}$ lies within the calculated $C_{g-c}$ at $m^*/m_0$ ranging from 0.07 to 0.1. However, the measured $C_{g-c}$ under the accumulation regime is not flat, as opposed to the calculated $C_{g-c}$. Thus, it is difficult to define a fixed $m^*/m_0$ from this calculation based on the ballistic FET model. The discrepancy is likely due to the strong quantum confinement as well as the nonparabolicity of the conduction band of the 2.5 nm-thick InAs quantum-well, which should lead to an increase in $C_{DoS}$ as the Fermi-level is moving up from the bottom of the first eigenstate [81]. Although this error in the determination of the effective mass can be reduced by performing a rigorous calculation using such as tight-binding or semiclassical nonparabolic models [82,83], that would be out of scope of this work.

7.2 Subband Occupancy

One of the important assumptions made in this calculation is that all electrons populate only the first subband. In order to validate this, two factors must be known: (1) the separation between the 1st and 2nd subbands and (2) the position
of the Fermi-level ($E_f$) relative to the first eigen-state ($E_1$) in the given supply voltage range. Figure 7.4 shows calculated $E_f - E_1$ as a function of $V_{GS}$ for $m^*/m_0$ ranging from 0.04 to 0.1. Here, the zero gate potential is set at $E_f - E_1 = 0$. Under the non-degenerate condition ($E_f - E_1 < 0$), the Fermi-energy changes at the same rate as the change in $V_{GS}$, because the electron density in the channel is negligibly low. As $E_f - E_1$ becomes positive, electrons start to accumulate in the channel, resulting in the Fermi-level moving more slowly than the increase in applied $V_{GS}$. The rate of change in the Fermi-level under the degenerate condition ($E_f - E_1 > 0$) is governed by the density of states; in other words, a quantum-well with a smaller $m^*$ is more highly degenerate at the same $V_{GS}$ compared to a quantum-well with a larger $m^*$.

![Image](image.png)

Figure 7.4: Calculated $E_f - E_1$ as a function of $V_{GS}$ for $m^*/m_0$ ranging from 0.04 to 0.1 under the (a) equilibrium ($V_{DS} = 0$ V) and (b) saturation ($V_{DS} = 0.5$ V) conditions.

In ballistic FETs, the drain voltage also affects $E_f - E_1$ at the top of the barrier. While electrons barely occupy the negative momentum states ($k_{S\leftarrow D}$) as the drain potential is lowered, the space charge neutrality must be held at the top of the barrier under the steady-state condition. Therefore, $E_f - E_1$ increases
in order to maintain the same amount of charges as in the case of the equilibrium \((V_{DS} = 0 \text{ V})\). Figure 7.4(a) and (b) compare \(E_f - E_1\) under equilibrium \((V_{DS} = 0 \text{ V})\) and saturation conditions \((V_{DS} = 0.5 \text{ V})\). At \(V_{GS} = 0.3 \text{ V}\) and the on-state voltage of \(V_{DS} = 0.5 \text{ V}\) \((i.e.\ under\ saturation)\) as indicated by the dotted line in Figure 7.4(b), \(E_f - E_1\) is \(\sim 0.2 \text{ eV}\) for \(m^*/m_0 = 0.07\), which is greater than that under equilibrium by \(\sim 0.05 \text{ eV}\).

![Computed eigenstates and the 1st state wave-function for a 2.5 nm-thick InAs channel FET using a 1-D Poisson-Schödinger solver.](image)

Figure 7.5: Computed eigenstates and the 1st state wave-function for a 2.5 nm-thick InAs channel FET using a 1-D Poisson-Schödinger solver.

From the above analysis, it has been found that the maximum difference between the positions of \(E_f\) and \(E_1\) is 0.2 eV \((i.e.\ under\ saturation)\). Thus, if the energy separation between the 1st and 2nd subbands is at least a few \(kT\) higher than 0.2 eV, the initial assumption of all electrons occupying the first subband...
becomes valid. In Figure 7.5, the electron eigenstates for the given 2.5 nm-thick InAs quantum-well FET are calculated using the Poisson-Schrödinger solver. Only the first subband is existent within the quantum-well, which is $\sim 0.3$ eV from the bottom of the conduction band. The second higher state, which is positioned in the back barrier, is $> 0.3$ eV above from the first subband. Hence, all of these results suggest that a 2.5 nm-thick InAs quantum-well with an InAlAs back barrier provides only $\sim 0.3$ V electron barrier for carrier confinement. However, it should be noted that this value is high enough for the given low supply voltage of 0.5 V.

### 7.3 Extrinsic $I_D$ Compared with Measured $I_D$

The ballistic FET model depicted in Figure 7.6 can yield $I_D$ in terms of intrinsic voltages ($V_{DS_{int}}$ and $V_{GS_{int}}$), as marked with an orange box, assuming no voltage drops at the source and drain. On the other hand, any real device inevitably includes some series resistances, thus actual voltages applied to S/D terminals are larger than the voltages dropping across the intrinsic FET. The voltage drop in the gate terminal can be ignored, since the current through the gate terminal is negligibly low. The total source and drain series resistances can be extracted by subtracting the ballistic contact resistance from the on-resistance extrapolated at zero-$L_g$ as shown in the table in Figure 6.8. Since the FETs have a symmetric S/D structure, the source and drain series resistances ($R_S$ and $R_D$) are determined to be 80 $\Omega$-$\mu$m each (i.e. a half of the total series resistance). To calculate $I_D$ in terms of the extrinsic voltages ($V_{DS_{ext}}$ and $V_{GS_{ext}}$), a simple circuit analysis is performed with $R_S$ and $R_D$ using the equations below. The equations are
numerically solved using a nonlinear Newton-Raphson method.

\[ V_{GS_{int.}} = V_{GS_{ext.}} - I_D(V_{GS_{int.}}, V_{DS_{int.}}) \cdot R_s \]  
\[ V_{DS_{int.}} = V_{DS_{ext.}} - I_D(V_{GS_{int.}}, V_{DS_{int.}}) \cdot 2R_s. \]  

\( \text{Rs} = 80 \, \Omega\cdot\mu\text{m} \)
\( \text{RD} = 80 \, \Omega\cdot\mu\text{m} \)

Figure 7.6: Circuit diagram for an extrinsic FET including the S/D series resistances. The orange box represents the intrinsic FET.

Figure 7.7 plots the calculated \( I_D \) as a function of \( V_{GS_{ext.}} \) both in (a) semi-log and (b) linear scales at \( V_{DS_{ext.}} = 0.5 \, \text{V} \), assuming \( m^*/m_0 \) is 0.7. The subthreshold behavior is shown to be ideal (60 mV/decade), since the interfacial traps and 2-D electrostatics have not been taken into account. Based on the target supply voltage of 0.5 V and off-current limit of 100 nA/\( \mu\text{m} \), the on-state current is determined to be 0.8 mA/\( \mu\text{m} \) at \( V_{GS_{ext.}} = 0.33 \, \text{V} \) as shown in Figure 7.7(b). Comparing with the measured \( I-V \) in Figure 6.6, the calculated ballistic FET outperforms over the measured FET by \( \sim 60\% \). Such a large discrepancy is likely due to both the non-ideal subthreshold characteristics and quasi-ballistic transport in the measured FET. Although this comparison suggests how close the measured
FET is to an “ideal” FET, it does not distinguish the dominant factor for the on-current degradation at the fixed off-current.

Figure 7.7: Calculated extrinsic $I_D$ as a function of $V_{GS}$ in (a) semi-log and (b) linear scales at $V_{DS} = 0.5$ V and $m^*/m_0 = 0.07$ V.

134
In order to find out how closely the measure FET is working in the ballistic limit, the on-current should be evaluated at the same overdrive voltage \((V_{GS} - V_T)\) rather than at the same off-current. When \(V_T\) is defined using a linear extrapolation method in linear \(I_D-V_{GS}\) plots, the subthreshold characteristics do not
affect the determination of the on-current. $I_D$ in a ballistic FET is proportional to $(V_{GS} - V_T)^{\frac{3}{2}}$ (See Chapter 2.1). Thus, $I_D^{\frac{3}{2}}$ is supposed to be linear against $V_{GS} - V_T$, as shown in the dotted line of Figure 7.8(a). Therefore, $V_T$ is nothing but the X-intercept of the linearly fitted line from $I_D^{\frac{3}{2}}$. Figure 7.8(b) plots $I_D$-$V_{GS}$ for the measured FET. Its $V_T$ is determined to be $\sim$0.03 V, and $I_D$ at $V_{GS} - V_T = 0.4$ V is $\sim$ 0.7 mA/µm. Here, it should be noted that the overdrive voltage of 0.4 V is relevant for the supply voltage of 0.5 V, since $V_T$ is $\sim$0.1 V and $V_{GS}$ at the on-state is $\sim$0.33 V as shown in Figure 7.8(a). $I_D$ of the measure FET is only $\sim$15% lower than of that the ballistic FET, indicating that the measured FET operates nearly in the ballistic regime.

As mentioned earlier, the ballistic $I$-$V$ shown above has been calculated under the assumption that $m^*/m_0$ is 0.7. Since nonparabolic effect has been ignored, a large error may arise if the calculated $I_D$ is very sensitive to small changes in $m^*/m_0$ within the range relevant to an InGaAs quantum-well. Hence, how $I_D$ varies with respect to $m^*/m_0$ should be evaluated. In addition, an error made in the determination of $R_S$ can result in a large variation in the current calculation. Figure 7.9 plots the extrinsic $I_D$ as a function of $R_S$, which represents the source series resistance only from the source side, at $m^*/m_0$ range of 0.04 to 0.1. If $R_S$ varies by $\sim$10% and $m^*/m_0$ from 0.04 to 0.1, the calculated current would vary from 0.7 to 0.9 mA/µm. From this, the ballisticity of the measured FET is extracted to be 0.8 to 1.
Figure 7.9: Calculated extrinsic $I_D$ at $V_{DD} = 0.5$ V in terms of the source series resistance at the range of $m^*/m_0$ from 0.04 to 0.1.

### 7.4 Current-gate Cut-off Frequency

### 7.5 conclusions

This chapter has investigated how close the device fabricated in the 4th generation operates to the ballistic limit. Based on the design parameters of the measured device, $I-V$ characteristics of a ballistic FET have been calculated using the theory introduced in Chapter 2 and compared with the measured one. It is found that the measured FETs operate within $\sim 80\%$ of the ballistic regime, with the considerations of possible errors made in determining the channel effective mass and S/D series resistance.
Chapter 8

Conclusion and Future Works

8.1 Summary

In this dissertation, essential FET device physics has been examined to understand the nanoscale FETs, and key considerations for designing III-V MOSFETs for the high performance VLSI application have been investigated. The research has mainly focused on fabrication and characterization of InGaAs-based raised S/D quantum-well MOSFETs with S/D regrowth via the gate-last process scheme. In comparison with the results obtained from previous studies over the past several years, a significant improvement in device performance (e.g. > 5:1 lowered $SS$ and > 3:1 increased $g_m$ compared to the initial results) has been achieved by implementing new process modules and renovating the device design. The progression of the device performance (peak $g_m$ and $SS_{min}$) throughout this thesis, based on key technology improvements, are summarized in Figure 8.1. The device from the most recent development described in Chapter 6 exhibits a 500 $\mu$A/$\mu$m
on-current at a fixed 100 nA/µm off-current and $V_{DD} = 0.5$ V at a VLSI-relevant gate length of 25 nm, which is the highest on-current from any reported III-V MOSFETs and comparable to state-of-the-art silicon Fin- and nanowire-FETs. Furthermore, in comparison with results from the ballistic FET model calculations, it is estimated that the device with 25 nm-$L_g$ is operating within 80-100% of the ballistic regime.

![Figure 8.1: Progression of device performance (peak $g_m$ and $SS_{min}$) through technology developments made in this dissertation.](image-url)
8.2 Future Works

8.2.1 Manufacturing Process Flow

The future work beyond this thesis should aim to make further improvements on the device fabrication and design. Especially in consideration of compatibility to manufacturing, two important additions are required: (1) sidewall spacers and (2) self-aligned metal-alloy contacts. Adding the sidewall spacers will help mitigate the increase in the total input capacitance induced by the large gate-drain overlap as well as the fringing effects, resulting in a reduced intrinsic gate delay. To implement these into the already established process flow, the undoped
sidewall spacers and N+ S/D regions must be regrown separately as illustrated in Figure 8.2, since the sidewall spacers need to be gated but not the N+ S/D regions. Incorporating self-aligned metal-alloy contacts and more heavily doped S/D will further reduce $\rho_c$ for vias with a smaller size. Since FETs in this work have large S/D metal contacts ($>1 \times 1 \ \mu m^2$), a relatively large $\rho_c$ of $\sim 5 \ \Omega-\mu m^2$ has not limited the device performance. However, for manufacturing purposes, it should be reduced down below $1 \ \Omega-\mu m^2$ by adopting the alloyed contacts [84,85] as well as by implementing more heavily doped S/D regions as mentioned in Chapter 2.2.

8.2.2 III-V FinFET

Another key direction for the future lies toward FinFETs. Despite the fact that the fabricated FET is estimated to operate near the ballistic limit as discussed in Chapter 7, its on-current at a fixed off-current (0.5 mA/$\mu$m) is still much lower than that of a ballistic FET (0.8 mA/$\mu$m). Such a large discrepancy is likely due to the non-ideal subthreshold characteristics of the fabricated FET. Considering the nearly straight subthreshold slope at $\sim 60 \ \text{mV/decade}$ all the way down to the target off-current in the long-channel devices, it implies that the poor 2-D electrostatic integrity at short channels plays a dominant role in degrading the subthreshold characteristics more strongly than the $D_{it}$ effect and BTBT-related leakage. The fabricated FET with 25 nm-$L_g$ is almost reaching the scaling limits of planar ultra-thin body FETs, given the EOT of $\sim 0.8 \ \text{nm}$. Although the poor electrostatics at such short channels can be further improved by adopting thicker vertical spacers, it will degrade the on-current, resulting in larger series resistance.
In this context, transiting to a multi-gate device architecture would help maintain an adequate level of electrostatic control with thinner spacers.

There are several challenges in fabricating III-V FinFETs. To obtain better electrostatics and maintain the on-current and integration density, fins of $< 6$ nm width and $> 100$ nm height must be realized. Moreover, the surface of the fins must be nearly defect-free. Regarding these critical requirements, dry-etching of fins introduces difficulty in the control of etched sidewall slopes and in minimizing the surface damage. Instead, UCSB has taken a different approach to the fin formation using a sidewall epitaxial growth on InP dummy templates [86]. In this process, the InP growth template is defined by a facet-selective wet-etching, thus obtaining perfectly vertical and atomically flat sidewalls. The fins are grown, by atomic-layer epitaxy on the template sidewalls, and then the template is selectively removed by wet-etching, leaving few-nm-thick InGaAs fins as shown in Figure 8.3. Such a bottom-up approach allows to avoid any dry-etch damage and to control the fin width in sub-nm precision.

![Figure 8.3: (a) Bird’s-eye view of schematic and (b) cross-sectional TEM of III-V FinFET with atomic layer epitaxy [86]](image-url)
8.2.3 mm-wave/THz device Application

For CMOS logic applications that require a high DC on/off ratio and small device footprint, the device geometry has been optimized to enhance subthreshold characteristics instead of maximizing the transconductance. In this sense, the channel thickness has been aggressively scaled down and a thicker vertical spacer has been adopted in the most recent development discussed in Chapter 4. In spite of the significantly improved $I_{on}$ at a fixed $I_{off}$ from such design changes, $g_m$ is inevitably compromised to a certain extent because of the degradation of carrier transport properties in such a thin channel and the increase in the S/D series resistance from thick spacers.

On the other hand, for monolithic microwave integration circuit applications where the high frequency response is more emphasized over the DC performance, the device design must be re-optimized in order to minimize the parasitic capacitance and maximize the transconductance. In this context, thin vertical spacers as well as a thick and high indium content channel should be implemented, thus a higher current-gain cutoff frequency ($f_T$) and maximum frequency of oscillation ($f_{max}$) can be achieved. As shown in Figure 5.8(a), the FET with a relatively thick channel of 5 nm and a very thin vertical spacer of 2 nm exhibits a record high peak $g_m$ of 3.0 mS/µm at $V_{DS} = 0.5$ V and $L_g = 18$ nm, which is the best reported $g_m$ out of any type of transistors including III-V HEMTs. It is noted that such a high $g_m$ has been achieved despite the MOS-based channel design, which inherently suffers from large surface scattering. This shows strong promise toward achieving a very high $f_T$ if the S/D geometry is modified (i.e. by reducing the gate-to-S/D overlap) such that the parasitic capacitance is substantially reduced.
Conclusion and Future Works Chapter 8

Figure 8.4 shows the current-gain cutoff frequency \( f_T \) as a function of the drive current for a 25 nm-\( L_g \) device at \( m^*/m_0 = 0.7 \) and \( V_{DS} = 0.5 \text{ V} \) with two different (1 fF/\( \mu \text{m} \) vs. 0.3 fF/\( \mu \text{m} \)) parasitic capacitance values. The gate-capacitance \( (C_{g-ch}) \) and injection velocity \( (v_{inj}) \) are calculated based on the ballistic FET theory introduced in Chapter 2, and \( f_T \) is computed as \( \frac{g_m}{2\pi(C_{g-ch}+C_{par})} \), where \( C_{par} \) is the parasitic capacitance.

Here, it is assumed that a VLSI-like FET as shown in Figure 8.2 with thick sidewall spacers has a parasitic capacitance value in the order of 1 fF/\( \mu \text{m} \). This corresponds to a computed \( f_T \) of \( \sim 500-600 \text{ GHz} \) at 0.8-1.0 mA/\( \mu \text{m} \). However, by increasing the drain-to-gate offset and adopting a T-gate structure, it is possible to reduce the parasitic capacitance to \( \sim 0.3 \text{ fF/\( \mu \text{m} \)} \), which results in a very high \( f_T \) of > 1 THz at the same current level.

**Current-gain Cutoff Frequency**

![Figure 8.4: Calculated current-gain cutoff frequency as a function of \( I_D \) at \( m^*/m_0 = 0.7 \) and \( V_{DS} = 0.5 \text{ V} \) for 0.3 and 1 fF/\( \mu \text{m} \) parasitic capacitance.](image_url)
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


[57] U. Singisetti, “In0.53Ga0.47As MOSFETs with 5 nm channel and self-aligned source/drain by MBE regrowth,” Ph.D Dissertation.


REFERENCES


REFERENCES


[72] S. W. Chang, X. Li, R. S. Oxland, and W. Wang, “InAs N-MOSFETs with record performance of $I_{ON}=600$ A/m at $I_{OFF}=100$ nA/m ($V_{D}=0.5$ V),” *... of Technical Digest...*, 2013.


[75] *E-mode planar Lg & $\#x003D; 35$ nm $In0.7Ga0.3As$ MOSFETs with $InP/Al2O3/HfO2$ (EOT $\#x003D; 0.8$ nm) composite insulator*. IEEE, 2012.


REFERENCES


nm width ingaas finfets of 200 nm height by atomic layer epitaxy,” in Device Research Conference (DRC), 2013 71st Annual. IEEE, 2013, pp. 1–2.