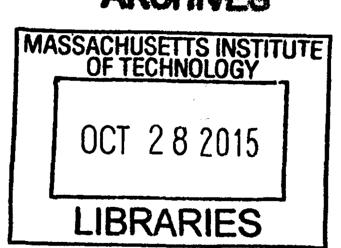
Interband Quantum Tunneling at the Band-Edges in III-V Semiconductor Heterojunctions for Low-Power Logic and Detectors

By

Ryan Iutzi

BASc Nanotechnology Engineering
University of Waterloo, 2010



Submitted to the Department of Materials Science and Engineering in Partial Fulfilment of the Requirements for the Degree of

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

September 2015

© 2015 Massachusetts Institute of Technology. All rights reserved

Signature of Author: Signature redacted
Department of Materials Science and Engineering
August 24, 2015

Certified by: Signature redacted
Eugene A. Fitzgerald
Merton C. Flemings-SMA Professor of Materials Science and Engineering
Thesis Supervisor

Accepted by: Signature redacted
Donald R. Sadoway
John F. Elliott Professor of Materials Science and Engineering
Chair, Departmental Committee on Graduate Students

Interband Quantum Tunneling at the Band-Edges in III-V Semiconductor Heterojunctions for Low-Power Logic and Detectors

By

Ryan Iutzi

Submitted to the Department of Materials Science and Engineering on August 24, 2015 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Materials Science and Engineering

ABSTRACT

This thesis explores interband tunneling in semiconductor heterojunctions where a density-of-states switching mechanism can be used to sharply modulate the junction conductance using small applied voltages via alignment and misalignment of the band-edges. Such a mechanism is useful for the pursuit of low-power logic devices as well as nonlinear analog components such as square-law detectors. In the former application, density-of-states switching can potentially allow for transistor subthreshold slopes steeper than the room-temperature thermal limit of 60 mV/decade in a device known as a tunnel field effect transistor (TFET), and in the latter application, it can allow for curvature coefficients greater than the thermal limit of 38.7 V^{-1}. However, predictions of stellar performance from simulation studies have not yet been matched experimentally, particularly in the area of TFETs, which have not produced a density-of-states switching device steeper than the thermal limit. To date however, the true steepness obtainable from density-of-states switching (band-edge steepness), and what affects it, remains unknown. This thesis fundamentally studies the band-edge steepness and how it depends on various factors including interfacial crystal defects, inhomogeneity, temperature, and band alignment. Using type-III band alignment epitaxially grown InAs/GaSb heterojunctions, as well as both type-II and type-III band alignment InGaAs/GaAsSb heterojunctions, it is identified that point defects gettered by dislocations at the interface as well as uneven distributions of point defects and composition can lead to both interfacial energy states and nonuiform band-alignment across the tunnel junction that result in poorer band-edge steepness. We identify a variety of techniques to improve the junction interface quality and demonstrate a corresponding improvement in band-edge steepness. We also determine that in our two-terminal devices designed to deconvolute TFET parasitics, the band-edge steepness does not depend on temperature, contrasting strongly with published TFET results and indicating that TFETs designed to utilize density-of-states switching are likely not actually harvesting interband tunneling and are instead dominated by thermally-activated parasitics. Using the various conclusions of our fundamental work as design guidelines, we develop and demonstrate an InGaAs/GaAsSb materials system integrated on an InP platform with a record 76 mV/decade band-edge steepness and 43 V^{-1} curvature coefficient near zero-bias.

Thesis Supervisor: Eugene Fitzgerald

Title: Merton C. Flemings-SMA Professor of Materials Science and Engineering

To Judith and Abby

Acknowledgements

*"Epitaxy is easy, predictable, controllable, and results can be expected quickly." – **Nobody Ever***

Research takes a very long time. It took 165 hours to write this thesis, five years to obtain the data in it, one year to figure out how to make a TEM sample without breaking it (one really long year), 13 years to learn the science necessary to do semiconductor research, and 27 years to learn all the other skills necessary to contribute this tiny piece to human knowledge. It also took 4 years to remember how to correctly spell "Stranski-Krastatanaovsdfakdjfk growth". And unlike that growth mode, I am not an island; I have relied on countless individuals to help me to this point, a subset of which are acknowledged below.

First and foremost, I would like to thank my thesis advisor Eugene Fitzgerald, for putting his patriotism aside and allowing a Canadian to join his group. I first met Gene in February 2010 at the MIT visit weekend, where he presented me with three potential projects: one on solar cells, one on LEDs, and one on this whacky idea involving quantum tunneling that is not really well understood but if you take it, you can go on free trips to California a lot. As it turned out, the Bay Area's surprisingly cold winter climate and collective 12 hours of round trip flying were not vacation-like enough to undo the emotional damage of trying to grow antimonides, but Gene's humour, intelligence, and enthusiasm kept me going with a smile on my face for five solid years. His group is a collection of intelligent and driven students to whom he gives just the right amount of autonomy so that they can develop their skills, identify directions, and follow them. I also had the fortune of overlapping with Mayank Bulsara, who was a senior scientist of our group and is now chief scientist of a much larger entity. He was an important mentor and a fun person to have around.

I would also like to thank my thesis committee, Polina Anikeeva, and Silvija Gradecak, both fantastic researchers with a clear vision and ambition who have already produced significantly and I imagine will continue to for decades. Their guidance in this work was incredibly helpful.

The most important aspect of grad school is being surrounded by people in your group you actually like. Unfortunately, I wasn't aware of that when I joined the Fitzgerald group. Luckily, however, this group turned out to be filled with fantastic people. In fact, I maintain that the Fitzgerald group has one of the strongest comraderies of any group at MIT. This is in part because everyday life is a constant battle between us and the reactor (our MOCVD system) which is constantly finding creative and unique ways to malfunction on us. It literally just started alarming as I am writing this paragraph, wow.

When I joined, the group consisted of Bai Yu, who warned me that my project will be much harder than I think (and in retrospect should have screamed it louder), Li Yang, who at the time was the only other transistor member in the group and her semiconductor physics knowledge inspired me. The group had just graduated Cheng-Wei Cheng, who would return multiple times as a scientist for IBM, and taught me a great deal about the various intricacies of

epitaxy. Prithu Sharma was also in the group at that time, was an important source of entertainment and humour, and later became a critical mentor in my career search. Nan Yang was also an important mentor. She taught me a lot about epitaxy, TEM, and fab, and was in fact the first person to show me how to use the reactor. I remember walking into the lab as a wide-eyed new recruit and asking in a gentle naïve voice "Nan, do you know what EVERY button does?" And she did. And then there was Adam Jandl, the youngest member above me when I joined, who taught our incoming batch everything from semiconductor physics to how to properly install a VCR gasket. I like to believe that I taught him some things too, like how Canada is a different country from Britain, although he keeps denying it.

Tim Milakovich and Kunal Mukherjee joined at the same time as me, and we went through this process together. They became two of my best friends. I knew I was in for a good five years when the three of us pioneered the idea of a perpetually-full beer fridge in the office, which came in quite handy during quals studying. Tim is strong, like muscle man strong: He graduated six weeks before me, and after he left I realized there are some things on our system he torqued so tight they will never be loosened. He also emphasized fitness in the FITgerald group, and forced me to sign up for a half marathon. I now like running. You read that correctly, I like it. Kunal is too smart for his own good. He knows too much and it's dangerous. I still don't understand how he can grow one structure on the reactor, have the growth abort halfway through, and still harvest 19 journal articles out of the data. With that said, I am a much better dancer than him.

Roger Jia joined after us and figured out how to learn and master epitaxy of superlattices despite his three immediate seniors being Tim, Kunal, and I. Rushabh Shah is creative and is envisioning some new clever ideas in lattice-mismatched epitaxy and I am excited to see him progress. Chris Heidelberger gets it, he is responsible and reliable but also a hilarious person to hang out with. He gets to continue with our grant solo, and all three of these guys are now tasked with keeping the reactor up and running, which is tough but I would choose them for the job. Brian Pearson sat in our office, but is not in the Fitzgerald group, so I won't acknowledge him, nor will I acknowledge his great personality and fantastic mustache, which is now shaved but never forgotten.

I also had the pleasure of working with some fantastic collaborators within our center, the NSF Center for Energy Efficiency Electronics, led by Eli Yablonovitch, whose insight and scientific genius was incredibly useful to our project, as was that of his graduate student Sapan Agarwal. I also thank Prof. Dimitri Antoniadis and Prof. Judy Hoyt, and their graduate students Jamie Teherani, Tao Yu, and Winston Chern, as well as Prof. Jesus del Alamo and his graduate student Xin Zhao. Additionally, there are a variety of technical staff at MIT who were instrumental to this thesis. In the Center for Materials Science and Engineering (CMSE), I must thank Libby Shaw, Scott Speakman, Charlie Settens, Yong Zhang, and Shiahn Chen. In the Microsystems Technology Laboratory, I received much help from Dave Terry, Dennis Ward, Eric Lim, and Paudely Zamora, among many others. Facilities was also helpful, especially our Building 13 custodian Johnny, who would always remind us that "Tomorrow's another day".

I was fortunate to have many other mentors prior to coming to MIT, beginning in secondary school when I first embraced science and mathematics under the guidance of David Masters, Kuldip Bachher, Margaret Ruck, and Joseph Gmach, among many others. At the University of Waterloo, I was inspired by many professors including Thorsten Hesjedal, Anthony Vannelli, Giuseppe Tenti, Ting Tsui, Anant Anantram, among many others. I also thank numerous scientific and engineering role models in internships, including George Liebermann, Marko Saban, and Santiago Faucher at Xerox Research Center of Canada, and especially Prof. Alan McGaughey of Carnegie Mellon University, who first brought me to the US (despite being a fellow Canuck) and introduced me to the world of academia.

There are a variety of friends I had the pleasure of spending my time here with. First on that list would have to be Michael Burek, who has actually been my friend since elementary school, continued with me to high school, went to the same undergrad program at Waterloo, then also came to Cambridge, MA to Harvard to do graduate school with me (he is clearly following me, or I'm following him, either way, kind of weird.) He is incredibly ambitious and his ambition rubbed off just enough onto me that I was able to succeed as well. I would not have made it here without him. Torrey Ah-Tye has kept me laughing and happy; this was especially important during the past several months where I simultaneously battled through the stress of trying to graduate and also finding a job. Thank you for bearing with me. My roommates Kevin Spencer, Audren Cloitre, and Daniel Dadon were among my best pals here, as was my main ski ally Neil Patel, and my quantum dot pals Jesse Keenan and Brandon Lehoux*. (*Habs fan)

Nobody ever said epitaxy is easy, or predictable, or controllable, and more importantly, nobody ever said that about life. And that is why I save for last my acknowledgement to my family. My Grandparents Ray and Josie Caravaggio have always encouraged me and are magnificent to be around. Ray (or Bumpy) is a great inspiration, having started a massively successful company in his day and running it from the helm until his retirement. They both know a thing or two about strength. My Grandparents Stan and Marg Iutzi put their plans on hold and moved into our house when tragedy struck our family, and I will forever be grateful to their efforts to remind us that the Iutzi family is a team and that we will get through together. Taylor, my brother, is even better looking than me, and more charming. I'm slightly better at math, but I don't think that really matters, he has all the skills necessary to rule the world. My father Mike Iutzi is the reason I am an engineer. He got me excited about science at age 3 when he taught me about space. We once woke up at 4am to watch a meteor shower on a beach in Florida. More importantly, he taught me about never giving up in the face of anything. He certainly didn't. He raised me and Taylor on his own from ages 15 and 11 on.

This thesis is dedicated to my mother, Judy Iutzi. She was beautiful inside and out. She believed in me, motivated me, and instilled in me a sense of confidence and purpose that I will carry for the rest of my life. Nothing in life brought her more pride than her children and she believed them to be capable of anything. She would not have understood a word of this thesis but would have nonetheless considered it to be the best thesis ever.

Table of Contents

List of Figures **12**
List of Tables **20**
Chapter 1: Introduction and Motivation **21**
1.1 Introduction 22
1.2 CMOS, Energy Efficiency, and Speed 22
1.3 Switching Using Quantum Tunneling 27
1.3.1 Barrier Width Modulation 28
1.3.2 Density of States Switching 31
1.4 Two Terminal Analog Applications of Density-of-States Switching 39
Chapter 2: Establishment of Theoretical Model and Experimental Plan **45**
2.1 Introduction 46
2.2 Bulk vs Interfacial Band-Edge Steepness 46
2.3 Factors that Affect Interfacial Band Edge Steepness 48
2.3.1 Interfacial Defect States 49
2.3.1 Defect and Inhomogeneity-Induced Nonuniform Band-Alignment 50
2.4 Direct Measurement of Band-Edge Steepness: Establishment of a Figure of Merit 55
Chapter 3: Experimental Methods **66**
3.1 Introduction 67
3.2 Epitaxial Growth: MOCVD 67
3.3 Device Fabrication and Characterization 71
3.4 Materials Characterization 73
3.4.1 Transmission Electron Microscopy (TEM) 73
3.4.2 X-Ray Diffraction 75
3.4.3 Atomic Force Microscopy 76
Chapter 4: Binary Growth of InAs/GaSb **77**
4.1 Introduction 78

4.2 Growth of InAs and GaSb Layers 78
4.3 Growth of InAs/GaSb Heterojunctions and Effect of Growth Order 81
4.4 Effect of Switching Sequence on Defect Formation 85
4.5 Effect of Growth Order on Defect Formation 87
4.6 Use of Strain to Suppress Defect Formation 90
4.7 Use of Growth Temperature to Suppress Defects 92
Chapter 5: Electrical Characterization of InAs/GaSb Heterojunctions 96
5.1 Introduction 97
5.2 Device Fabrication and I-V Characteristics 97
5.2.1 NDR Region Features 99
5.2.2 Junction Area 103
5.2.3 Series Resistance 105
5.2.4 Conductance Slope Measurement Procedure 106
5.3 Effect of Doping 108
5.4 Effect of Growth Order 112
5.5 Post-Growth Annealing, Point Defects, and Interface Uniformity 115
5.5 Strained-Underlayer Lattice Matching 118
Chapter 6: Low Temperature Study of Conductance Slope 124
6.1 Introduction 125
6.2 Comparison of I-V Curves 127
6.3 Peak Current 129
6.4 Excess Current 131
6.5 Conductance Slope 133
Chapter 7: Ternary Growth of InGaAs/GaAsSb Heterojunctions 138
7.1 Introduction 139
7.2 Composition Control of InGaAs and GaAsSb 141
7.2.1 InGaAs 141
7.2.2 GaAsSb 142
7.3 InGaAs/GaAsSb Heterojunctions 143
7.4 GaAsSb Graded Buffers 145

7.5 InAsP Graded Buffers 151
7.6 The GaAsSb/InGaAs Interface 154
Chapter 8: Electrical Characterization and Band-Alignment Control of InGaAs/GaAsSb Heterojunctions 162
8.1 Introduction 163
8.2 Shifting Composition for InGaAs/GaAsSb Heterojunctions Without Buffers 164
8.3 GaAsSb Graded Buffer Devices 168
8.4 Temperature Dependence of GaAsSb Buffer Device 172
8.5 InAsP Buffer Devices 174
Chapter 9: InGaAs/GaAsSb Heterojunctions for Millimeter Wave Detection 178
9.1 Introduction 179
9.2 Band Alignment Shifting 179
9.3 Thermal Curvature Limit vs. Thermal Conductance Slope Limit 181
9.4 Buffer Devices 184
9.5 Temperature Dependence 188
Chapter 10: Conclusions and Future Work 190
10.1 Summary of Results 191
10.2 Suggestions for Future Work 194
10.2.1 Vigorous Optimization of Growth and Processing Conditions 194
10.2.2 Utilizing Materials Less Electrically Affected by Defects 195
10.2.3 Utilizing Smaller Structures 195
10.2.5 New TFET Designs 196
Appendix A: Device Fabrication Process 197
References 199

List of Figures

Figure 1.1: Minimum feature size of a transistor (blue) and supply voltage (red) as a function of year. Data taken from ITRS roadmap 24

Figure 1.2: Clock speed of CPUs as a function of year. Taken from 25

Figure 1.3: (a) Transfer curve of a MOSFET showing a subthreshold slope limit of 60 mV/decade at room temperature. (b) Band diagram of a MOSFET along a source-channel-drain cross-section, showing a thermal distribution of electrons in the source and a gate voltage-modulated barrier from source to channel..27

Figure 1.4 (a) Band diagram of a junction doped such that the bands overlap and interband tunneling can occur. The triangular barrier is shown in green, and the total potential difference in blue. (b) Band diagram of the same structure under an increased external bias, showing a much thinner tunnel barrier as a result. (c) Transfer curve of such a device showing a subthreshold slope due to tunnel barrier width modulation. 29

Figure 1.5 a) Semiconductor heterojunction showing an effective band gap (E_G^{eff}) at the interface. This represents the off-state of a density-of-states switch, where there are no final states for current to tunnel to. b) Same junction biased to the on-state, where there is now band overlap, and tunneling is allowed. c) Transfer curve of a density-of-states switch, showing a subthreshold slope that results from the steepness of the band edges 33

Figure 1.6 (a) Type-III band-alignment in an InAs/GaSb heterojunction, (b) corresponding transfer curve that one would obtain from using density-of-states switching with this band alignment, (c) Type-II band alignment in an $In_{0.76}Ga_{0.23}As/GaAs_{0.3}Sb_{0.7}$ heterojunction, (d) corresponding transfer curve that one would obtain from using density-of-states switching with this band alignment. 34

Figure 1.7 (a) Minimum subthreshold slope vs temperature for a variety of published experimental results that have been complied together into this plot. The thermal limit is shown as the black line (b) Same data from (a), but every data point was normalized by the 300 K subthreshold slope of the set it was collected from. The expected behavior of a thermal device is shown as the grey dotted line...38

Figure 1.8: Atmospheric attenuation of the electromagnetic spectrum. The millimeter wave region of the spectrum is highlighted in green. 41

Figure 1.9: Example millimeter wave detection system 43

Figure 2.1 a) Absorption coefficient of InAs as a function of energy, showing an Urbach tail at low energy, which has been expressed as a slope of meV of energy per decade change in density absorption, making it analogous to a subthreshold slope. b) Schematic of measuring band-edge

steepness optically, where no bands are moved in energy, c) Schematic of measuring band-edge steepness electrically, where bands must be moved and an interface is required 47

Figure 2.2 (a) Heterojunction biased such that it is in off-state. A mechanism involving an interfacial defect is shown, allowing tunneling to still occur, (b) Density of states without interface defects, which would translate to a subthreshold slope without interface defects, (c) same as (b), but with a tail of defect states extending into the band gap 50

Figure 2.3 a) Band diagram of InAs/GaSb heterojunction. The arrow shows the direction of electron current due to tunneling, and defect states are also shown near the interface, which can also act as sites for tunneling. b) Band diagram of InAs along a direction parallel to the interface, showing a defect state and an example of band bending around the defect (the degree of band bending is not necessarily this large) (adapted from [62]) c) Same diagram as b) with the valence band energy of GaSb near the interface shown. Light blue region represents where tunneling can occur, and three reference points are marked as A, B, and C. d) Transfer curve showing sharpness of switching for current at each point A, B, and C, as well as a trap-assisted leakage current. The solid red curve shows the measured result, which is a weighted average of all of these conductance pathways, and the subthreshold slope is considerably less steep 51

Figure 2.4: Calculated transfer curve for the case of a perfect band edge (black), a band edge with a tail of states decreasing in density be three order of magnitude from 0 to 150 meV into the band gap, and for the case of nonuniform band alignment (green) resulting from regions varying up to 150 meV over an area that represents 1% of the junction area..54

Figure 2.5 Example TFET design 57

Figure 2.6 Design of a two-terminal device designed to probe band-edge steepness without convoluting with other three-terminal parasitics. The result of a current continuity finite element method simulation is overlaid, where the color corresponds to the potential in response to an externally applied 5 V bias, and the black represents select electric field lines...........................59

Figure 2.7 (a) Band diagram of an InAs/GaSb heterojunction in a TFET, showing the separate effect of the gate voltage and the drain voltage. (b) Same band diagram for the case of a two-terminal device, where only one applied bias controls everything...61

Figure 2.8 (a) Band diagram of two-terminal junction, with a graph of the $T_t N_C(E) N_V(E)$ term over energy and a graph of the $[F_C(E)-F_V(E)]$ term shown. (b) A graph of the $[F_C(E)-F_V(E)]$ term over energy, showing how it changes in shape with applied bias for low bias, (c) Same as (b) but for higher applied biases..64

Figure 3.1 (a) schematic of the MOCVD system, (b) labelled photograph of the system...........69

Figure 4.1 (a) Bulk phase diagram of InAs, (b) Bulk phase diagram of GaSb 79

Figure 4.2: Surface of GaSb layer at varying V/III ratios. (a) SEM secondary electron image, scale bar of 10 μm showing Ga-rich droplets. (b)AFM image with a z-scale of 2 nm showing a

featureless surface with a surface roughness of 0.143 nm rms. (c) AFM image with a z-scale of 50 nm showing Sb-rich precipitates on the surface 80

Figure 4.3 AFM image of GaSb surface grown at (a) 32 nm/min (z-scale: 7 nm), and (b) 16 nm/min (z-scale: 4 nm) 81

Figure 4.4: (a) Cross section TEM ([220] diffraction condition) of InAs grown on GaSb at 530°C showing dislocations originating from the interface. (b) Lattice-resolved TEM image of structure showing a dark region. Zoomed in portion shows a Burger's circuit drawn based on the identified lattice sites in the image........ 82

Figure 4.5: The Band-gap vs lattice constant diagram for III-V materials. InAs and GaSb are highlighted on the graph, and the filled in red region represents all possible compositions available from intermixing of the four elements. 84

Figure 4.6: An illustration of As-for-Sb exchange, where Sb displaces As from the top surface, resulting in a GaAs layer at the interface 84

Figure 4.7: [220]-diffraction condition TEM images of various switching sequences. 86

Figure 4.8: AFM images of top InAs surface for various pulsing sequences 87

Figure 4.9: Cross section TEM ([220] condition) of a) 80 nm thick GaSb grown on InAs at 530°C, b) 380 nm thick GaSb on InAs 89

Figure 4.10: [224] glancing exit reciprocal space maps of a) InAs/GaSb junction (82% relaxed) b) 80 nm GaSb / InAs (<5% relaxed) c) 380 nm GaSb/InAs (63% relaxed). Short dashed line shows the line of strain, and long dashed line shows the line of relaxation........ 89

Figure 4.11: TEM images of: a) InAs grown on 380 nm thick (63% relaxed) GaSb, b) InAs grown on 80 nm thick strained GaSb (images are at different magnifications to show all layers. Top InAs layer is the same thickness for both) 91

Figure 4.12: AFM image of GaSb surface (left) and InAs surface (right) for the case of a tensile-strained InAs layer and a lattice-matched InAs layer. 92

Figure 4.13: a) [220] diffraction condition cross section TEM image of InAs/GaSb structures for an InAs growth temperature of a) 530 °C, b) 490 °C, and c) 465 °C 94

Figure 4.14: [224] glancing exit reciprocal space maps of a) InAs/GaSb junction grown at 530°C (82% relaxed) b) InAs grown at 465 °C on GaSb (<1% relaxed) c) InAs grown at 465 °C and then ramped up to 530 °C while growing. Short dashed line shows the line of strain, and long dashed line shows the line of relaxation........ 95

Figure 5.1 a) SEM micrograph of InAs/GaSb device at a side-view, b) top-down optical micrograph of device, c) I-V curve of device........ 98

Figure 5.2 (a) Comparison of NDR region of an InAs/GaSb junction for two different values of series resistance, (b) equivalent circuit of this tunnel diode, (c) demonstration of load-line

intersection with I-V curve for different values of series resistance, showing stability for the case of $R_s < R_D$, and instability for the case of $R_s > R_D$ 100

Figure 5.3 (a) I-V curve showing instability region that is larger than the NDR region. The red curve shows the true I-V curve, and the blue curve shows the measured I-V curve due to the intersection of the load-line and oscillations. (b) A corresponding absolute conductance-voltage curve showing a significant underprediction of the true steepness of the conductance slope 101

Figure 5.4 (a) I-V curve showing instability region that is smaller than the NDR region. The red curve shows the true I-V curve, and the blue curve shows the measured I-V curve due to the intersection of the load-line and oscillations. (b) A corresponding absolute conductance-voltage curve, this time showing an accurate conductance slope that corresponds to the intrinsic one .. 103

Figure 5.5 (a) Peak current vs junction area based on mask dimensions for InAs/GaSb device, trend goes through origin. Inset shows optical micrograph of a 12 μm device (b) Peak current vs junction area based on mask dimensions for GaSb/InAs device, trend does not go through origin. Inset shows an optical micrograph of a device with significant undercut (c) Peak current vs. junction area for the same device with the optically measured diameters. Trend goes through origin 104

Figure 5.6: (a) I-V curve of a sample device showing the slope approaching a constant value at high and low bias, (b) differential resistance of the device, showing the curve asymptotically approaches the series resistance at high and low bias, and also showing the negative differential resistance in the bottom right quadrant for the NDR region, (c) I-V curve before and after correcting for series resistance, using the value extracted from (b) 106

Figure 5.7 (a) I-V curve (blue curve) of an InAs/GaSb device grown under non-optimized conditions. The curvature coefficient is also plotted (red curve) for the sake of using the peaks to identify the boundaries of the instability region, which is highlighted in light blue. (b) Absolute conductance-voltage curve of the device, showing a conductance slope of 265 mV/decade, which is continuous with the slope outside of the instability region 108

Figure 5.8 (a) [Si] as a function of depth in InAs measured by SIMS for a doping calibration. (b) same calibration for GaSb 110

Figure 5.9 (a) I-V curve for the samples from Table 5.1 with varying doping levels. (b) Corresponding absolute conductance-voltage curve for all samples, showing generally the same conductance slope for all samples 111

Figure 5.10: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve derived from a), for InAs/GaSb and strained GaSb/InAs devices 113

Figure 5.11: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve for 80 nm thick and 380 nm thick GaSb on InAs devices 114

Figure 5.12 Non-reconstructured dislocation core with dangling bonds vs. reconstructed core with no dangling bonds 115

Figure 5.13: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve InAs/GaSb devices before and after annealing at 600^{0}C. c) and d): Equivalent comparison for 80 nm thick GaSb/InAs devices 118

Figure 5.14: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve for InAs grown on 80 nm thick (strained) and 380 nm thick (63% relaxed) GaSb 120

Figure 5.15: a) I-V curve and b) Absolute conductance-voltage curve for (1) InAs grown at 465 ^{0}C, (2) 465 ^{0}C with a ramp to 530 ^{0}C, and (3) 465 ^{0}C and then the entire structure annealed at 600 ^{0}C. c) Comparing the annealed sample to growth at 530 ^{0}C with and without annealing ... 122

Figure 6.1 Two scenarios in which a thermal component would dominate the device characteristics. (a) a large conductance with a thermal dependence $G_{diode}(V)$ (expressed as a diode) in parallel with the tunnel junction $G_T(V)$ (expressed as a resistor with a voltage-dependent conductance) (b) a small conductance with a thermal dependence $G_{diode}(V)$ in series with the tunnel junction $G_T(V)$ 126

Figure 6.2 (a) I-V curve at varying temperatures for GaSb-on-InAs device and (b) I-V curves at varying temperatures for InAs-on-GaSb device. Inset shows a cross-sectional image on the [220] diffraction condition for each sample 129

Figure 6.3: (a) Peak current vs temperature for both devices, (b) Illustration showing effect of Fermi-tail on tunneling temperature dependence (c) Band diagrams for an InAs/GaSb heterojunction at 300 K and 4K 131

Figure 6.4: (a) Temperature dependence of the current at various bias points beyond the NDR region for GaSb/InAs device. (b) Band diagram for a device biased at 0.15 V, showing an example of a trap-assisted tunneling leakage route, in addition to a hole current (c) Valley current temperature dependence for both the GaSb/InAs and InAs/GaSb devices 133

Figure 6.5: (a) Absolute conductance-voltage curve at select temperatures for GaSb/InAs device. (b) Conductance slope as a function of temperature for both devices. For GaSb/InAs, a comparison of the results before and after correcting for series resistance is displayed. 134

Figure 6.6: Temperature-independent conductance slope measurements from this thesis, overlaid on Figure 1.7, which shows published TFET subthreshold slopes with a kT/q ln(10) temperature dependence 135

Figure 6.7: Band schematic of a heterojunction below a gate oxide and metal. r1 is a rate describing tunneling into gate oxide trap states. r2 is a rate describing generation into the conduction band from the trap states, which is thermally activated 136

Figure 7.1 %Ga composition in InGaAs measured by HRXRD as a function of TMGa precursor flow as a fraction of total group-III precursor flow 142

Figure 7.2: %As composition in InGaAs measured by HRXRD as a function of AsH_3 precursor flow as a fraction of total group-V precursor flow 143

Figure 7.3 (a) schematic of InGaAs/GaAsSb/GaSb structure, (b) cross section TEM image of structure on the [220] diffraction condition, (c) HRXRD [224] reciprocal space map of structure 144

Figure 7.4 (a) schematic of InGaAs/GaAsSb/GaAsSb graded buffer/GaSb structure, (b) cross section TEM image of structure on the [220] diffraction condition, (c) HRXRD [224] reciprocal space map of structure 146

Figure 7.5 (a) TEM image on the [220] diffraction condition of a GaAsSb buffer grown at 580 °C on GaSb, showing many defects. (b) a more zoomed in image of this sample, showing a very high concentration of defects originating right from the interface where the buffer was initiated: a rare observation for a graded buffer 148

Figure 7.6 (a) Schematic of "interlayer" As-Sb exchange in an InAs/GaSb growth, as originally shown in Figure 4.6 of Chapter 4. (b) schematic of "intralayer" As-Sb exchange within a GaAsSb layer 150

Figure 7.7 (a) GaAsSb/InGaAs device heterojunction on an InAsP buffer platform. (b) InGaAs/GaAsSb device heterojunction on the same platform 153

Figure 7.8 TEM image on the [220] diffraction condition for an InAsP buffer on InP, with an InGaAs cap. (b) [224] reciprocal space map of the structure. 154

Figure 7.9 (a) [220] TEM image of InGaAs/GaAsSb heterojunction grown on the InGaAs-capped InAsp buffer. (b) Corresponding AFM image of the top surface of the top InGaAs layer, indicating high roughness and likely island growth 155

Figure 7.10: (a) Schematic of InGaAs/GaAsSb/InGaAs/InP test structure. (b) [220] TEM image of structure (c) Nomarski reflection micrograph of top surface showing defects running along a [110] direction (d) Nomarski Reflectance of structure grown with only GaAsSb as the top layer, (e) Nomarski micrograph of structure with only the InGaAs layer grown on InP 157

Figure 7.11 Test growth conditions and corresponding Nomarski reflection micrographs 158

Figure 7.12 Illustration of the use of a special gas switching sequence to prevent intralayer As-Sb exchange. (a) Original As-rich surface, (b) surface after 30s H_2 bake with As desorbed, (c) surface after 30s TMSb flow with at least one monolayer of Sb on the surface (d) During growth, with Sb layer remaining on the surface 159

Figure 7.13: [220] TEM image of InGaAs/GaAsSb/InGaAs on an InAsP buffer using a 30s H_2 bake + 30s TMSb flow to initiate the GaAsSb layer. The dark circles are milling-induced defects. (b) AFM image of the top InGaAs showing evidence of step-flow growth. (c) [220] TEM image of a different region of the sample showing evidence of phase separation in the GaAsSb layer 161

Figure 8.1 Conduction band energies (yellow) and valence band energies (purple) for III-V semiconductors between 0.56 and 0.62 nm lattice constant, plotted on a scale where the electron

affinity of gold is defined as the zero point of energy. The three compositions are labelled. Taken from 165

Figure 8.2 (a) I-V curves of InGaAs/GaAsSb heterojunctions at three different band alignments. (b) Corresponding I/V-V curves, (c)-(e) Band diagram of heterojunction at each composition. 166

Figure 8.3: (a) I-V curve of an $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ with no buffer (blue curve) and with a GaAsSb buffer (green curve). (b) corresponding I/V-V curves 169

Figure 8.4 (a) Schematic of a top contact + back contact configuration for a device with a buffer. (b) Schematic for a top contact + circular side contact configuration. (c) Illustration showing a line-of-sight metal deposition where the top contact and undercut of the semiconductor layers below results in a self-aligned side contact. (d) SEM image showing the resulting device with the self-aligned side contact 171

Figure 8.5 I-V curve for $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ on GaAsSb buffer with a bottom contact (red curve) and a self-aligned side contact (blue curve). (b) corresponding I/V-V curve 172

Figure 8.6: I-V curve of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ on GaAsSb buffer for temperatures varying from 300 K down to 77 K. (b) corresponding I/V-V curves 173

Figure 8.7: (a) I-V curves of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction on no buffer (blue curve), a GaAsSb buffer (green curve), and an InAsP buffer (red curve). (b) corresponding I/V-V curves 175

Figure 8.8: (a) I/V-V curve of $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ heterojunction (red) and $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction (blue) on an InAsP/InP platform. (b) I/V-V curve of $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ heterojunction with an InAsP buffer (red) and without a buffer (blue) 176

Figure 9.1: (a) Curvature coefficient as a function of applied voltage for $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$, (blue) and $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ (red), both grown on GaAsSb layers with no buffer. (b) I-V curve of the two devices at low voltage, showing how the $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ does a much better job at accomplishing rectification of the forward current 180

Figure 9.2: "Curvature Enhancement Factor" – the second term of equation (6) – plotted as a function of voltage for the arbitrarily chosen case of a 100 mV/decade conductance slope 183

Figure 9.3 (a) Curvature coefficient of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction device with no buffer (blue), GaAsSb buffer (green), and InAsP buffer (red). (b) Corresponding I/V-V curves for the sake of comparison 185

Figure 9.4: (a) curvature coefficient of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ (red) and $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ (purple) on InAsP buffer, showing demonstration of beyond the thermal limit of curvature coefficient near the origin, (b) I-V curve of the two devices at low voltage 187

Figure 9.5: (a) Curvature coefficient of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction device on a GaAsSb buffer as a function of temperature. (b) Curvature coefficient at zero-bias for this device normalized by the value at 300 K, and compared to the thermal trend 189

List of Tables

Table 1.1 Prediction of Density-of-States Switching Performance from Simulation....................35

Table 1.2 Experimental Density-of-States Switching Performance 36

Table 3.1 MOCVD Growth Parameters..70

Table 4.1 Gas Switching Sequences Employed for InAs/GaSb Heterojunctions 85

Table 5.1 Acceptor (N_a) and Donor (N_d) Concentrations of Si-doped samples.........................110

Table 7.1: Growth Parameters for InAsP Buffer 152

Chapter 1: Introduction and Motivation

1.1 Introduction

This thesis is focused on developing and understanding the use of interband tunneling at the band edges in semiconductor heterojunctions. Specifically, its focus is on how the steepness of the band edges can be understood and controlled to obtain novel electronic devices in which the conductance sharply changes with applied voltage. Such a mechanism is very useful for a variety of applications, including transistors that can switch with very low voltage, or two-terminal devices with a high degree of nonlinearity for mixing and detection. This section will begin by highlighting how such a concept rose to prominence as a result of a search for lower energy transistors. However, this thesis is not motivated solely by low-energy transistors. Rather, this work is focused on understanding the fundamentals of interband tunneling in heterojunctions and how it depends on material properties and other aspects, and ultimately how it can be applied to various applications, including its original intended use in energy-efficient logic, and extending into the world of analog device applications as well.

1.2 CMOS, Energy Efficiency, and Speed

Silicon-based electronics constitute a significant proportion of all electronic devices in the market today. This has been true since the explosion of transistor devices for computing in the 1940's and then the development of integrated circuits (IC's) in the 1960's. The heart of the integrated circuit is the MOSFET (metal-oxide-semiconductor field effect transistor), and IC's typically use a combination of n-type and p-channel devices together in a logic configuration known as CMOS (Complementary MOSFET). Over the course of many decades, MOSFET devices in IC's have been shrinking very rapidly in accordance with Moore's Law[1]: while in 1970, the minimum feature size on a device was 10 μm, today the minimum feature size is

approaching 10 nm. This trend can be seen as the blue curve in Figure 1.1. However, this trend of miniaturization is approaching its limits, due to increased issues of parasitic resistances and capacitances with decreasing feature size [1]. The search to identify and develop new devices to replace the MOSFET and extend computing performance has already begun, and a variety of different candidates have been identified that use an extremely wide range of mechanisms. [2], [3]

While device size is important for the ability to fit many transistors on a single wafer, an additional metric has become extremely important in recent years: power consumption. There are two forms of power dissipation in an integrated circuit: dynamic and static dissipation. Dynamic power dissipation refers to the charge dissipated each time a transistor switches state and is given by:

$$P_{dynamic} = C_L V_{DD}^2 \tag{1}$$

where C_L is the load capacitance (resulting from the transistor and interconnects), and V_{DD} is the power supply voltage on the circuit. Even when not switching, transistors dissipate power due to leakage, and this is the cause of the second form of power consumption: static power dissipation. The two most dominant leakage mechanisms are current leakage through the gate oxide, and subthreshold leakage, which is due to the small fraction of hot carriers that are able to transport across the channel in the off state. Both of these currents are ultimately also driven by the supply voltage V_{DD}, and hence their power dissipation is proportional to V_{DD}^2, as is also the case for dynamic consumption. Hence, power dissipation in integrated circuits depends on the square of the supply voltage, and in order for power dissipation to decrease, the supply voltage must be decreased. In the earlier era of Moore's law, scaling of device dimensions was driven by constant-field scaling, where the supply voltage was in fact decreased, as can be seen

in the red curve of Figure 1.1 before 2001. However, after 2001, V_{DD} reached a plateau that has resulted in the supply voltage remaining around 1 V. [4] As will soon be described, the reason for this plateau is unavoidable in a MOSFET, and this puts an ultimate limit on how low power any CMOS integrated circuit can be.

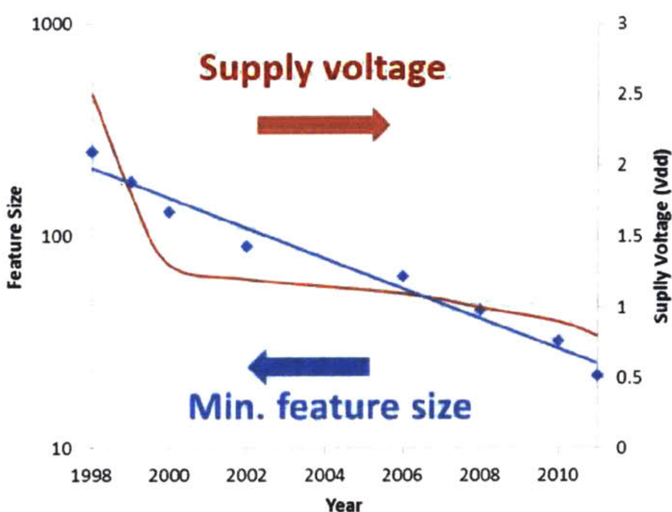


Figure 1.1: Minimum feature size of a transistor (blue) and supply voltage (red) as a function of year. Data taken from ITRS roadmap [5]

There are two reasons why it is important for integrated circuits to consume low power. The first is, of course, for the sake of energy efficiency. As computers and data centers are becoming more and more ubiquitous, a significant amount of the world's power is being consumed by data processing. In 2005, the data centers of the world alone (server farms, data storage centers, etc) consumed 152 billion kilowatt hours of power [6], putting them between Iran and Mexico in terms of total electricity consumption for the year. [7] In fact, every search query submitted to Google search engine costs 1 kJ of energy. [8] World energy consumption, however, is only part of the driving force in the interest for energy efficient electronic devices.

The second reason is that power dissipation is coupled to the maximum speed of an electronic device. This directly follows from the dynamic power dissipation in equation (1), which is the quantity dissipated each time the transistor switches. The higher the clockspeed on an IC, the more frequently transistors are switching, and hence, the more power is dissipated. This has led to microprocessor clock frequencies reaching a plateau in 2005, just shortly after the V_{DD} stopped decreasing in 2001, [4] as can be seen in Figure 1.2.

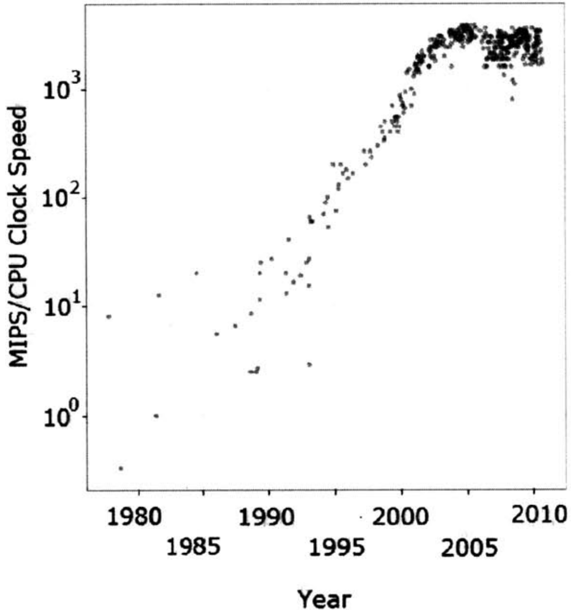


Figure 1.2: Clock speed of CPUs as a function of year. Taken from [5]

The reason for the inability to decrease V_{DD} further is ultimately due to the fact that the MOSFET has a fundamental limit to its subthreshold slope. The subthreshold slope defines the amount of gate voltage required to increase the drain current by an order of magnitude:

$$SS = \frac{\partial V_g}{\partial log I_D} \tag{2}$$

Figure 1.3a shows the subthreshold slope extracted from the transfer curve of a MOSFET. This essentially shows the output of the device (current collected at the drain) as a function of the input (gate voltage). Typically, many orders of magnitude difference between the on-state current and off-state current is required, which means a minimum amount of voltage swing must be supplied to turn the device on and off, based on the steepness of the subthreshold slope. Therefore, the supply voltage - and hence power dissipation - is directly limited by the subthreshold slope. The problem with this is that the subthreshold slope in a MOSFET has a fundamental limit, beyond which it cannot be steepened and therefore, beyond which energy efficiency cannot be enhanced. This fundamental limit results from the fact that the mechanism of turning on a MOSFET relies on thermal activation: There is a barrier from the source to the channel, and the barrier is decreased by increasing gate voltage, as shown in Figure 1.3b. Carriers in the source have a Fermi-Dirac distribution, and the hot carriers in the tail of the distribution are able to make it over the source-channel barrier. Hence, the current follows a thermally activated exponential form. This means that there is a limit on the steepest possible subthreshold slope given by the thermal voltage:

$$SS_{min} = \frac{kT}{q}\ln(10) \approx 60\ mV/decade\ \ (at\ room\ temperature) \tag{3}$$

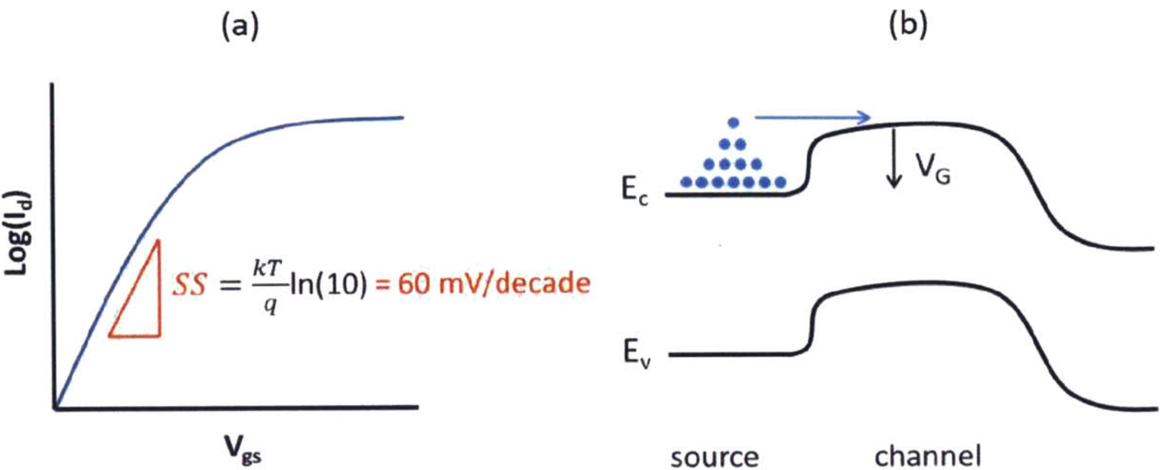


Figure 1.3: (a) Transfer curve of a MOSFET showing a subthreshold slope limit of 60 mV/decade at room temperature. (b) Band diagram of a MOSFET along a source-channel-drain cross-section, showing a thermal distribution of electrons in the source and a gate voltage-modulated barrier from source to channel

1.3 Switching Using Quantum Tunneling

The only manner in which to break this fundamental barrier is to utilize a device that is not thermally activated. In addition to this power problem, the transistor is already reaching the end of its scaling limits as described previously, and so the search is now on for a new device that, in addition to overcoming the scaling problems of the transistor, is also able to obtain a lower subthreshold slope. It is predicted that switching can occur at less than 60 mV/decade at room temperature in devices with similar structures to MOSFETs utilizing quantum tunneling [9]–[11], impact ionization[12]–[15], ferroelectric dielectrics[16], and mechanical switches [17]–[19]. The focus of this thesis is on devices that utilize tunneling, as they are not affected by the delay due to the positive-feedback mechanisms that lead to the low subthreshold slope in impact ionization, ferroelectric, and mechanical switching devices. [20]

There are essentially two mechanisms in which tunneling can be used to obtain a transistor: barrier width modulation and density of states switching.

1.3.1 Barrier Width Modulation

The original mechanism of switching in a TFET was barrier width modulation[21]. The concept is demonstrated in Figure 1.4. Comparing this back to the MOSFET example, if the source-channel junction in Figure 1.3b was doped heavily enough, then the valence band in the channel overlaps in energy with the conduction band in the source, as shown in Figure 1.4a. Now, instead of having to overcome the barrier, electrons and holes can simply tunnel across. This is known as interband tunneling. The tunnel barrier is approximately triangular as shown in Figure 1.4a. As bias is increased, the tunnel barrier becomes thinner, shown in Figure 1.4b. Since the tunnel probability depends exponentially on barrier thickness, there is an exponential turn-on, as shown in Figure 1.4c. This exponential turn-on is not bound by any thermal limits.

The tunnel current, I, is given by the following equation:

$$I = C_1 \int^{\Delta E} [F_C(E) - F_V(E)] T_t N_C(E) N_V(E) \boldsymbol{dE} \tag{4}$$

where $F_C(E)$ is the Fermi level on the n-side, $F_V(E)$ is the Fermi level on the p-side, T_t is the tunnel probability, $N_c(E)$ is the density of states of the conduction band on the n-side, and $N_v(E)$ is the density states of the valence band on the p-side. ΔE is the energy overlap of the bands, shown in Figure 1.4a. The basic premise of the equation is that three terms are integrated over energy over the extents of the band overlap. The first term (red) is the Fermi-level difference across the junction, which is essentially the term that pushes the current, given by the

source-drain voltage. The second term is the tunnel probability (green), which modulates the current, and is controlled by the gate, which changes the barrier thickness, as is shown in Figure 1.4. Finally, the third term is the joint density of states of the two overlapping bands (blue). While this term changes as well with the gate, it is not as important for barrier width modulation, as the tunnel probability term is what dominates the modulation.

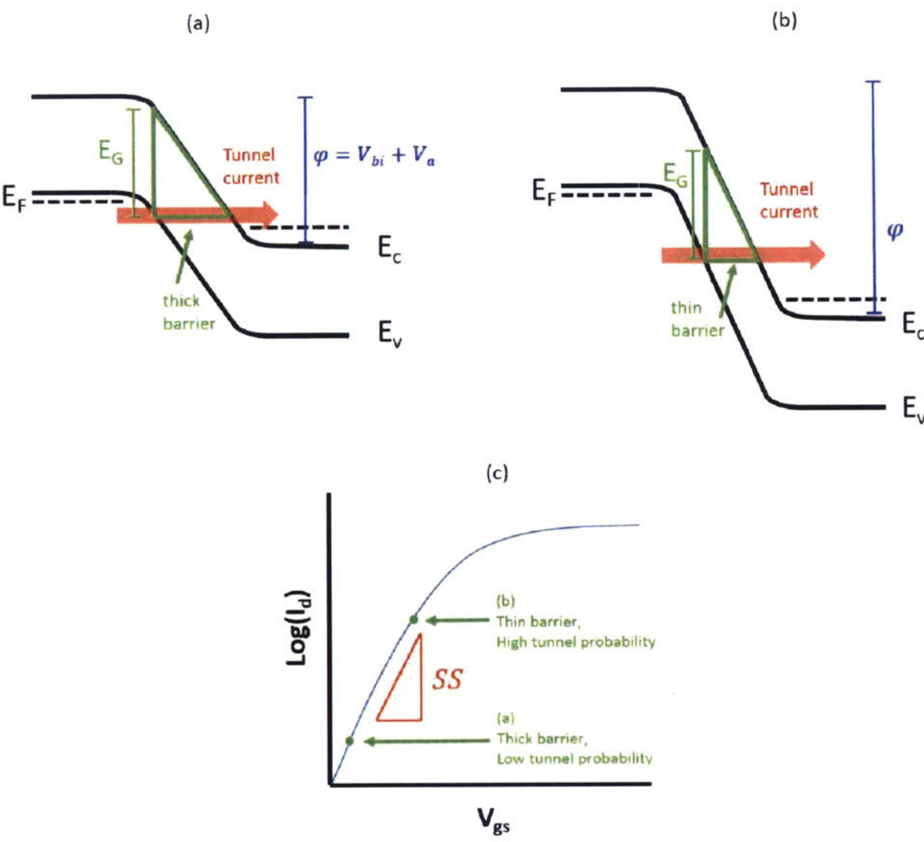


Figure 1.4 (a) Band diagram of a junction doped such that the bands overlap and interband tunneling can occur. The triangular barrier is shown in green, and the total potential difference in blue. (b) Band diagram of the same structure under an increased external bias, showing a much thinner tunnel barrier as a result. (c) Transfer curve of such a device showing a subthreshold slope due to tunnel barrier width modulation. Adapted from [21]

The first demonstration of a TFET using barrier width modulation was done in a GaAs device by Baba in 1992. [22]. Similar devices were fabricated in Si[23], Si on Insulator (SOI), [24] InGaAs, [25] and carbon nanotubes (CNT). [11] To date, only a few devices fabricated in this manner have obtained a subthreshold slope below 60 mV/decade, [11], [26]–[30] the best of which was 40 mV/decade in a CNT device in 2004 [11]. However, these device demonstrated on-currents in the pA/μm range, many orders of magnitude below MOSFET drive currents. As a result, these devices would not be fast enough to compete with MOSFET. Therefore, it is necessary to increase the tunnel probability.

The tunnel probability is given by the Kane model[31]:

$$T_t = \exp\left(\frac{-\pi (m^*)^{1/2} {E_G}^{3/2}}{2\sqrt{2}\hbar q F}\right) \tag{5}$$

where F is the field across the junction, E_G is the band gap, and m* is the effective mass for tunneling (given by the electron effective mass on the n-side and the hole effective mass on the p-side). It can be seen then, that the best way to increase the tunnel probability is with a lower band gap and a lower tunnel effective mass, both of which can be obtained with lower band gap materials. Indeed, SiGe [30] and InGaAs [25] devices have shown higher currents, consistent with this. However, such devices are still orders of magnitude lower in current than MOSFETs. One may argue then, that an even lower band gap material would be better. However, as the band gap is lowered further, a trade-off emerges where the steepest possible subthreshold slope decreases. This results from the fact that the subthreshold slope originates from how the tunnel barrier changes with applied bias.[32]:

$$\frac{1}{SS} = \frac{dlog(T_t)}{d\varphi} \tag{6}$$

Since the subthreshold slope, SS, is an inverse slope, a larger value of 1/SS gives a steeper subthreshold slope. Combining (5) and (6) ultimately gives

$$\frac{1}{SS} = \left|\log(T_t)\frac{dF(V)/dV}{F}\right| \tag{7}$$

In the best case scenario, (dF(V)/dV) /F = V, and so [32]:

$$\frac{1}{SS} = \left|\frac{\log(T_t)}{\varphi}\right| \tag{8}$$

Now the tradeoff can be seen: as tunnel probability increases, subthreshold swing decreases in steepness. It turns out that decreasing band gap will never achieve a steep enough subthreshold slope at high enough tunnel probabilities to be competitive with CMOS. For example, if a smaller band gap material such as InAs is used, it can be seen from Figure 1.4 that a φ of at least the band gap needs to be obtained. Hence, φ must be at least 350 mV. To obtain competitive CMOS-level currents, a tunnel probability of at least 1% would be required [32]. Putting these values into equation (8) gives a subthreshold slope of 175 mV/decade, far worse than the thermal limit.

1.3.2 Density of States Switching

One potential manner in which to break the tradeoff between current and steepness is to utilize a heterojunction. In such a case, the tunnel barrier becomes smaller, as there is an effective bandgap that is smaller than the actual band gap of the two semiconductors. This is shown in Figure 1.5. A much smaller φ can be obtained, allowing for a potentially steeper subthreshold slope.

However, once the band-alignment is such that the bands are close together, a new mechanism of switching dominates: density of state switching [32], or what is sometimes

referred to as energy filtering [21]. This can be seen from Figure 1.5 a and b. Under certain bias conditions, the bands overlap and tunneling is allowed. However, under other bias conditions, the bands to not overlap, and tunneling becomes forbidden. As a result, it is possible to use a small voltage to toggle between these states to obtain a subthreshold slope that is as steep as the steepness of the band edges.

In referring back to equation 4, the difference between barrier width modulation and density-of-states switching, is that barrier width modulation was relying on the change of the tunnel probability T_t in the tunnel current as gate voltage was swept, while density-of-states switching is relying on changing the density of states term $N_C(E)N_V(E)$ in the tunnel current while gate voltage is swept.

In the example from Figure 1.5, the heterojunction system has type-II band alignment, meaning that there is a small effective energy gap (E_G^{eff}) between the conduction band on one side and the valence band on the other side. It is also possible to utilize type-III band alignment, where the valence band on one side is actually higher in energy than the conduction band on the other side. In this case, tunneling would be allowed with zero applied bias, and the device would be "normally on". A bias is then applied to misalign the bands to turn it off. An example of both band alignments is shown in Figure 1.6. This new case of type-III can be seen on the left side of Figure 1.6 (a and b). The example of type II, as before, is shown on the right side in c and d.

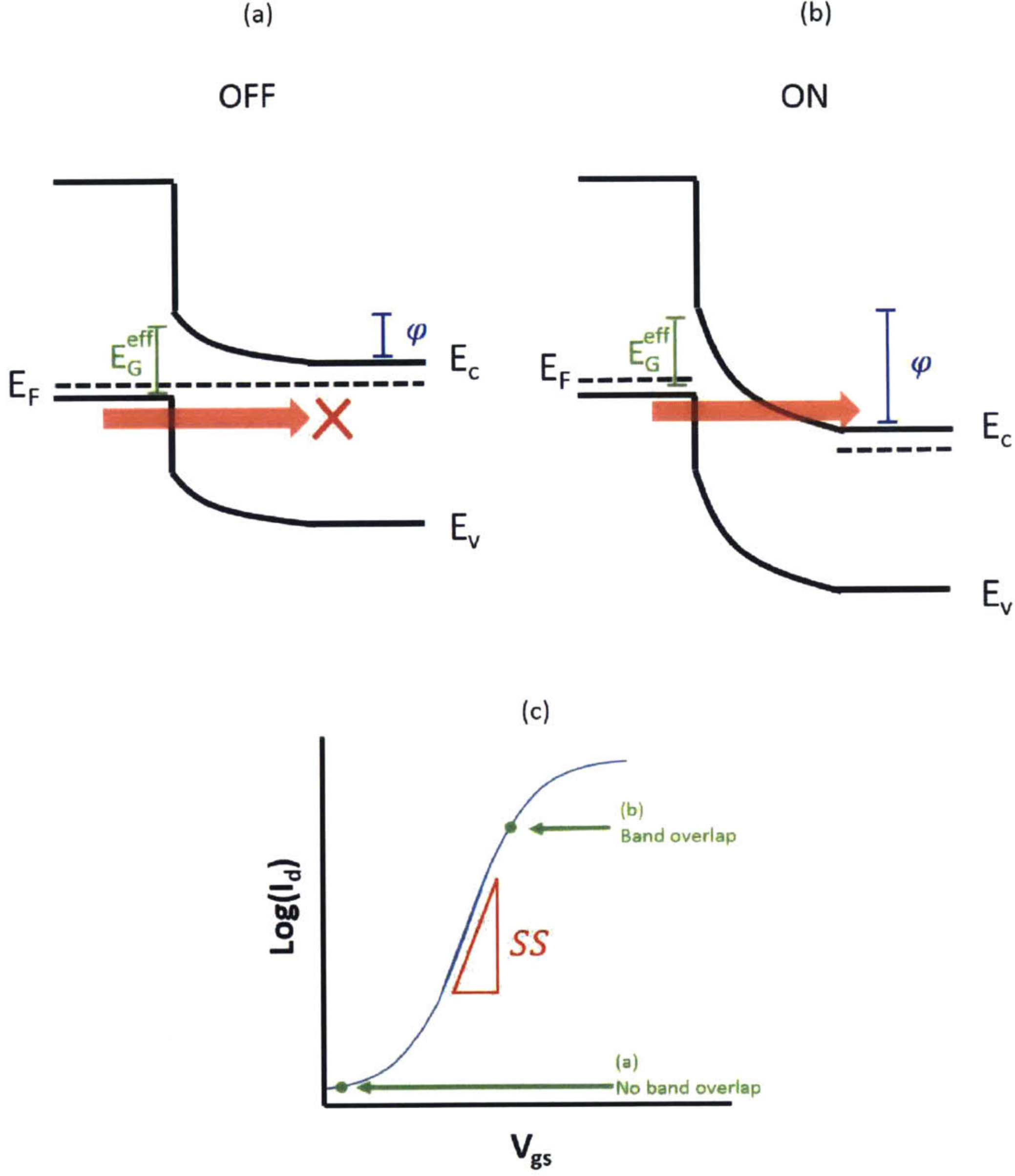


Figure 1.5 a) Semiconductor heterojunction showing an effective band gap (E_G^{eff}) at the interface. This represents the off-state of a density-of-states switch, where there are no final states for current to tunnel to. b) Same junction biased to the on-state, where there is now band overlap, and tunneling is allowed. c) Transfer curve of a density-of-states switch, showing a subthreshold slope that results from the steepness of the band edges [21]

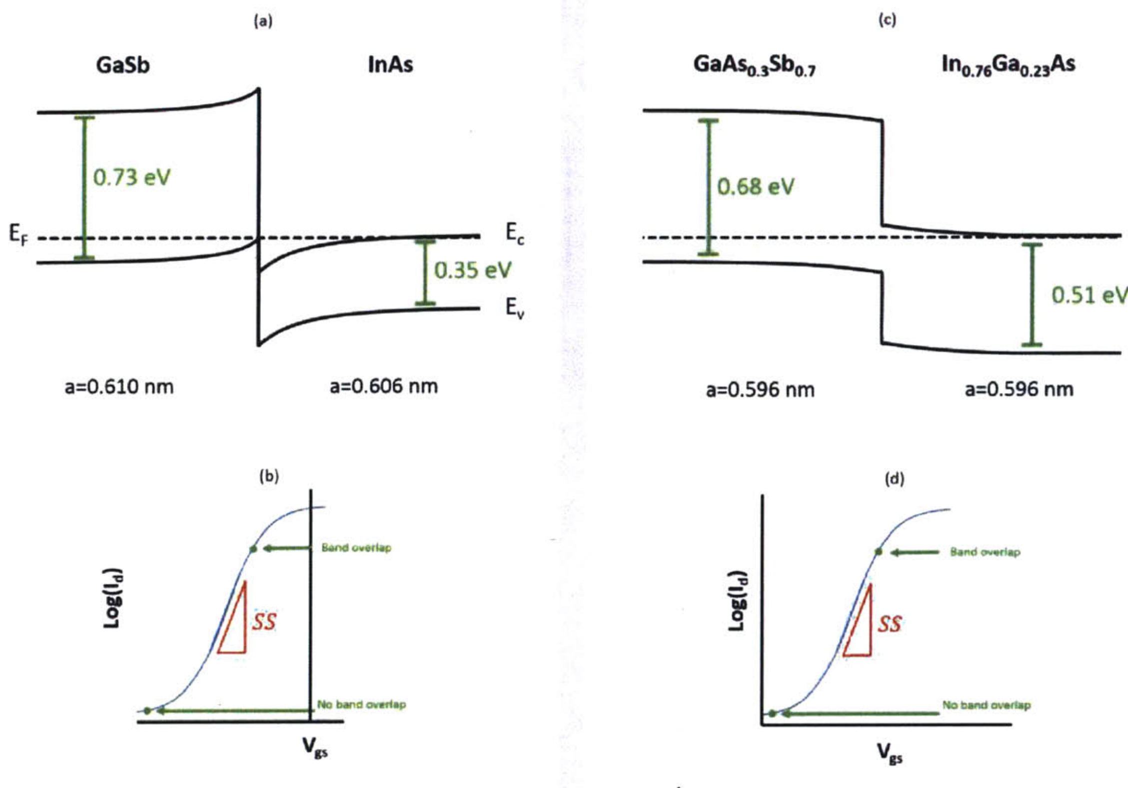


Figure 1.6 (a) Type-III band-alignment in an InAs/GaSb heterojunction, (b) corresponding transfer curve that one would obtain from using density-of-states switching with this band alignment, (c) Type-II band alignment in an $In_{0.76}Ga_{0.23}As/GaAs_{0.3}Sb_{0.7}$ heterojunction, (d) corresponding transfer curve that one would obtain from using density-of-states switching with this band alignment.

Type-III is commonly obtainable with InAs/GaSb [33], while type-II can be obtained by alloying Ga into InAs to obtain InGaAs, and As into GaSb to obtain GaAsSb [34]–[37]. It has also been obtained with other materials systems, including InAs/AlGaSb[38]–[40], InGaAs/InP [41], [42], and even Si/Ge[43] and SiGe/Si[10]. In this thesis, InAs/GaSb will be used for type-III, and various compositions of InGaAs/GaAsSb will be used for type-II.

Early simulation results on density-of-states switching began around 2009[44], and various researchers predicted sharp switching with currents high enough to compete with MOSFETs [38], [44]–[47]. These computational results showed room-temperature subthreshold slopes as low as 7 mV/decade [38] and in some designs, drive currents as high as 1.9 mA/μm at 0.4 V [45]. Table 1.1 summarizes a sampling of the predictions made.

Table 1.1 Prediction of Density-of-States Switching Performance from Simulation

Author, Year	Materials System	SS mV/decade	Id (on) (μA/μm)
Koswatta 2009[44]	CNT	**20**	851
Koswatta 2010 [45]	CNT	**21**	1900
Knoch 2010 [46]	InAs/AlGaSb	**33**	-
Lu 2012 [38]	InAs/AlGaSb	7	130

However, these simulations assumed perfect materials interfaces, and did not account for the effect of defects. It will be shown in Chapter 2, when preliminary theoretical calculations are presented, that we should expect materials defects, especially at the interface, to have an effect on the subthreshold slope due to the energy states they cause in the bandgap and the overall inhomogeneity of the band structure that they cause. Indeed, when experimental results were first published beginning in 2010-2012 [39], [48]–[51], gave considerably lower performance than predicted, with the steepest room-temperature subthreshold slope at 125 mV/decade, and highest obtained drive current of 180 μA/μm (microamps per micron of gate width) at 0.5 V [49]. Table 1.2 summarizes many of the experimental results on density-of-states switching TFETs. The SS (min) entry is for the minimum subthreshold slope, which is the steepest point of the transfer curves. For many, this steepness was only obtained over a small current range, and the effective subthreshold slope was much less steep.

Table 1.2 Experimental Density-of-States Switching Performance

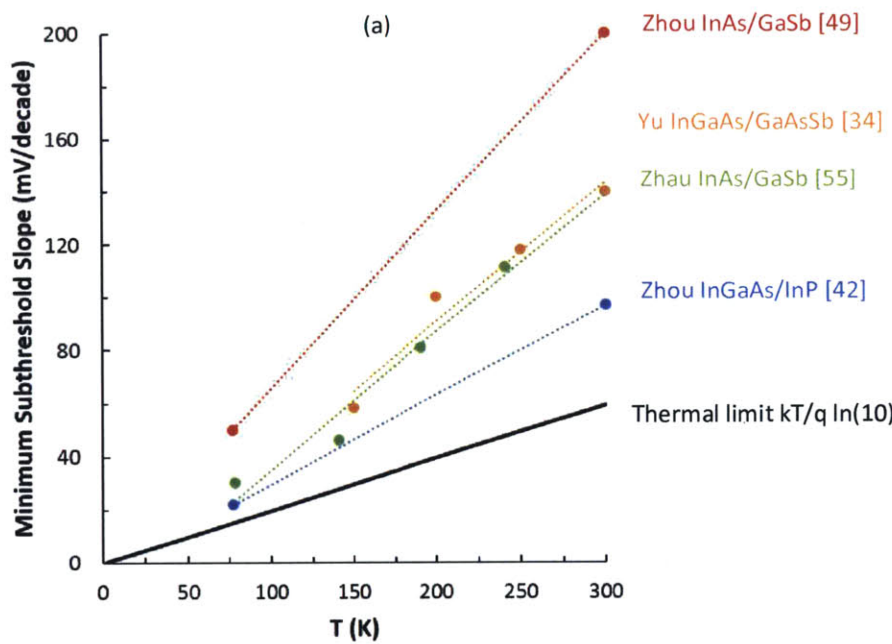
Year	Materials System	SS (min) mV/decade	Id (on) (μA/μm)
Zhou 2012 [49]	InAs/GaSb	**200**	380
Dey 2012 [50]	InAsSb/GaSb	**275**	110
Zhou 2012 [42]	InGaAs/InP	**93**	20
Mohata 2012 [52]	AlInAs/GaAsSb	**230**	135
Mohata 2011 [53]	InGaAs/GaAsSb	**750**	135
Li 2012 [40]	InAs/AlGaSb	**125**	78

While a good increase in current can be seen, all devices are considerably less steep than the thermal limit. In fact, the only result that reports even a sub-100 mV/decade slope is the InGaAs/InP device [42], which is by far the largest effective gap and may very likely still be using barrier-width modulation, as additionally evidenced by the much lower on current compared to the other devices.

The reasons for this discrepancy have been hypothesized to result from a variety of factors, such as poor gate oxide interface[39], [48]–[50], contact and other series resistance [39], [40], TFET geometry [39], [40], [50], as well as materials defects related to the tunnel junction [39], [48], [49], [51]. Indeed, all of these effects are important and likely affecting performance. Nevertheless, it remains to be seen what the true steepness of the band-edge is, and whether the concept of band-edge switching is truly feasible and capable of being steeper than thermal switching. Therefore, the goal of this thesis is more funeumental in its quest: it is desired to understand the steepness obtainable from the band-edges, what affects it, and how it can be controlled.

An additional peculiarity arises when the temperature dependence of experimental devices is examined. Equation (3) shows the steepest subthreshold slope for a thermal device

should depend linearly with temperature. A tunnel device, on the other hand, is not expected to have this temperature dependence, since the thermal tails are cut off by the band edges. However, experimental studies which examined the effect of temperature all found a marked improvement at lower temperature. Figure 1.7 compiles some of these results.



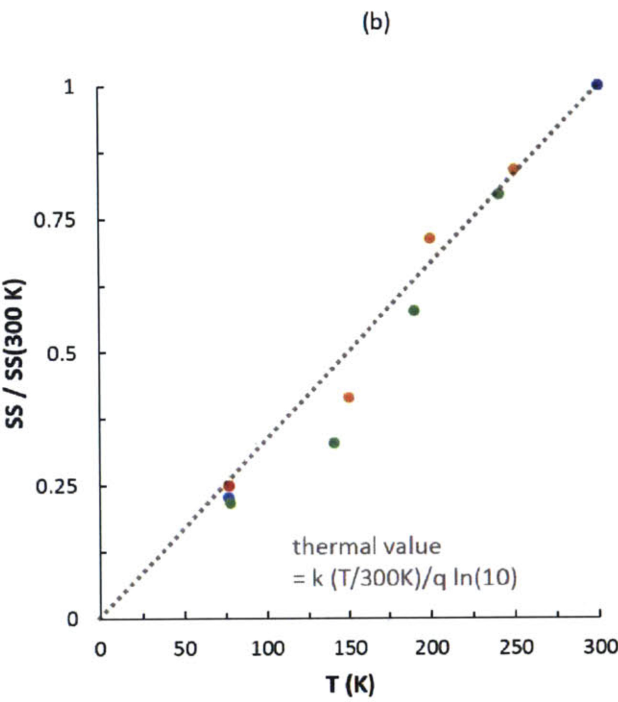


Figure 1.7 (a) Minimum subthreshold slope vs temperature for a variety of published experimental results that have been complied together into this plot. The thermal limit is shown as the black line (b) Same data from (a), but every data point was normalized by the 300 K subthreshold slope of the set it was collected from. The expected behavior of a thermal device is shown as the grey dotted line.

All experimentally reported heterojunctions using density-of-states switching show a linear dependence of subthreshold slope on temperature, as can be seen in Figure 1.7a. In fact, if all the data from all devices is normalized to their subthreshold slope at 300 K, they fall almost perfectly on a line defined by $\frac{kT}{q}\ln(10)$, which is exactly what would be expected of a thermal process involving carriers overcoming an energy barrier, not tunneling through it. This is a very strong indication that these devices are not observing density-of-states switching, and are likely not utilizing interband tunneling in any form. It is far more likely that they are dominated by a

parasitic thermal process. Indeed, some authors have suggested this same conclusion. [42], [54], [55] For example, Mookerja[54] proposed that in the region of steep-subthreshold switching, current is not true band-to-band tunneling, but tunneling into trap states at the gate-oxide/semiconductor interface and subsequent generation into the conduction band.

Hence, two important questions arise that must be answered:

1. Why are experimental density-of-states switching devices not obtaining subthreshold swings steeper than the thermal limit? (ie. What ultimately limits the steepness of band edges?)
2. Why do experimental density-of-states switching devices have a temperature dependence? (ie. Have any published devices actually demonstrated density-of-states switching?)

This thesis will answer both of these questions.

1.4 Two Terminal Analog Applications of Density-of-States Switching

So far, this thesis has discussed band-edges and density of states switching in the context of tunnel field effect transistors for low-energy electronics. However, the main goal of this thesis is to understand how sharp the band-edges truly can be, and what kind of performance can be obtained out of devices utilizing the band-edges for density of states-switching. While TFETs (a three terminal device) are one such application, there are additional applications for two terminal devices, namely in the area of analog devices, and most notably in the area of millimeter wave detection.

For a variety of analog applications such as mixing, amplification, and square-law detection, it is desirable to have a high curvature I-V curve. Much like a transistor has a thermal limit to subthreshold slope, a two terminal devices has a thermal limit of curvature for which tunneling offers a potential ability to overcome.

Millimeter waves are electromagnetic waves with a wavelength between 1 and 10 millimeters, and a frequency between 30-300 GHz. Figure 1.8a shows this frequency range and its atmospheric attenuation. What is very attractive about this frequency range is that it has a low attenuation by atmospheric obscurant and water vapor compared to infrared and visible light. As such, it is attractive for atmospheric and low visibility conditions imaging. However, because it has specific attenuation peaks corresponding to O_2 and H_2O, it is also useful for radio astronomy. Also, since clothing is transparent in some millimeter wave bands, it has found use for full-body scanning security screening. While active systems can easily scan for concealed weapons, a passive system can also be used, where natural blackbody emission of millimeter waves from colder metal can be distinguished from warmer human flesh. Passive systems such as this or for low-visibility condition imaging, however, require very high sensitivity and low-noise.

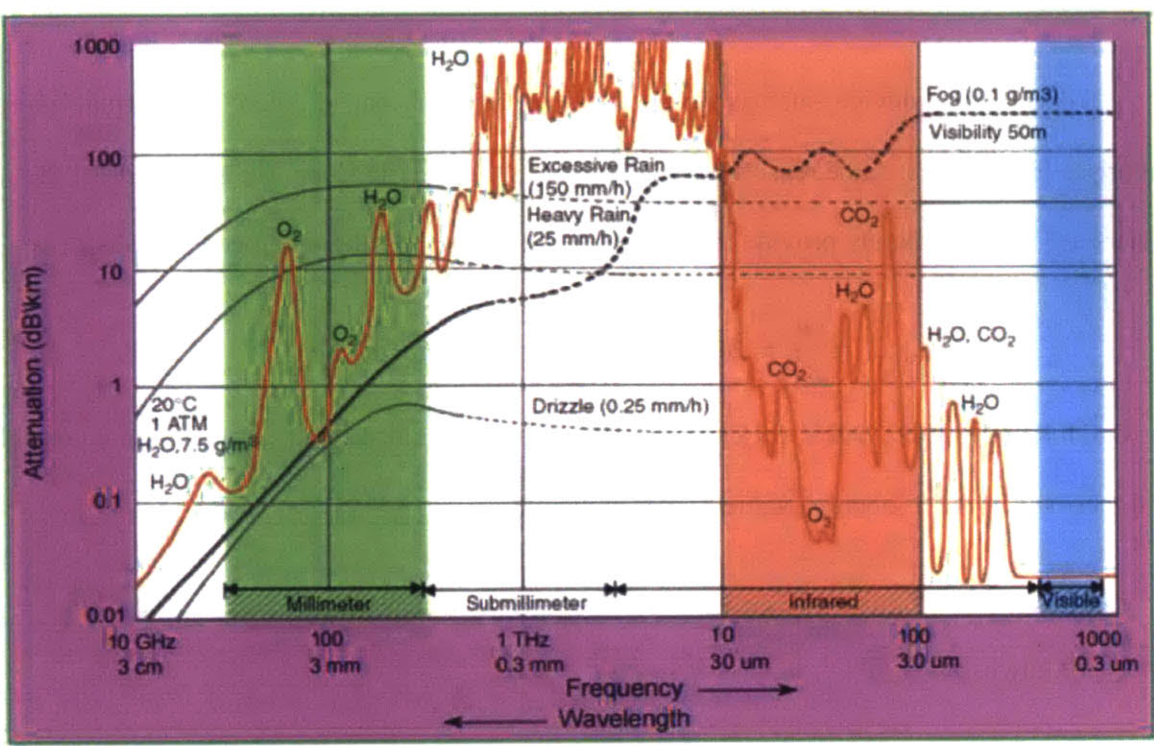


Figure 1.8: Atmospheric attenuation of the electromagnetic spectrum. The millimeter wave region of the spectrum is highlighted in green. [56]

An example millimeter wave detector system is shown in Figure 1.9. The detector is a square-law detector: a detector where the output voltage signal is proportional to the input power, or the square of the input voltage. Hence, the I-V curve of the device needs to have a high curvature. For this reason, the figure of merit commonly used for square-law detectors is curvature coefficient. A high curvature coefficient means the device will have a high sensitivity to the input. The curvature coefficient is defined as the second-derivative of the I-V curve normalized by the admittance level:

$$\gamma = \frac{d^2I}{dV^2} \Big/ \frac{dI}{dV} \tag{9}$$

For a thermal device, such as a p-n junction or Shottkey barrier, there is a thermal limit that arises for the exact same reason as in the case of subthreshold slope: the band-tails of the carriers are what ultimately provide the curvature. The limit for a thermal mechanism is:

$$|\gamma_{max}| = \frac{q}{kT} = 38.7\, V^{-1}\ (at\ room\ remperature) \tag{10}$$

This limit is obtained in the case of a perfect diode ideality factor (n=1). It is not possible for a thermal device to obtain a value higher in magnitude.

In the past, Schottky diodes have been used for these detectors. However, there are a variety of drawbacks for this:

1. As just described, they are limited to a maximum possible curvature limit of $38.7V^{-1}$
2. The region of curvature is not at zero-bias, so these devices need to be externally biased. This requires a biasing network, which results in a greater 1/f noise, making it difficult for passive applications
3. The curvature coefficient is very sensitive to temperature, which makes the device unreliable, especially given that many of the applications of such detectors involve them being subject to a wide temperature range.

A device utilizing quantum tunneling can overcome these problems because:

1. It is not bound by a thermal limit
2. The region of high curvature occurs where the bands cross/uncross. This can be tailored to occur right at zero bias if the band alignment can be brought right to the boundary of type-

II and type-III. Hence, the device does not need to be biased, allowing for lower 1/f noise, making it very useful for passive applications.

3. In theory, the curvature coefficient should be relatively independent of temperature, making it reliable in a wide range of environments. Of course, published density-of-states switching devices show a strong temperature dependence, but those results are suspicious and will be re-examined in Chapter 6.

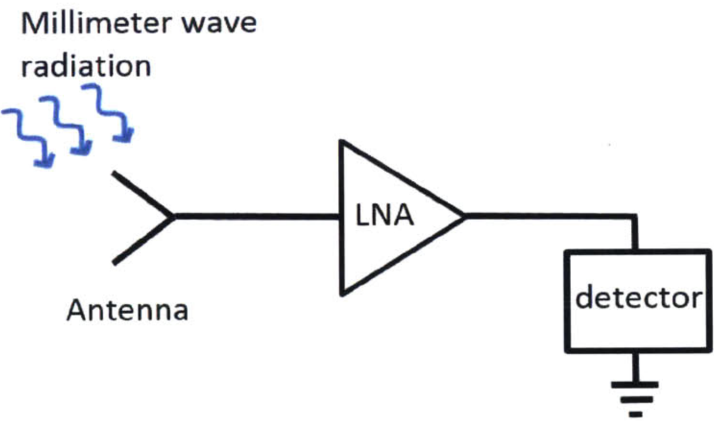


Figure 1.9: Example millimeter wave detection system

Indeed, devices using tunneling have begun to be published, with curvature greater than the thermal limit. A curvature coefficient of $-47V^{-1}$ was achieved with an AlGaSb/GaSb system[57]. InGaAs/GaAsSb has also been explored, and has achieved curvature coefficients of 49.4 V^{-1} as it is another method to obtain type-II/type-III alignment [58]. However, work to date has explored only InGaAs/GaAsSb at a composition lattice matched to InP. [41], [58], [59], and doping was used to bring the bands closer together.

This thesis will examine the effects of materials properties on band-edge steepness, and in the context of millimeter wave detection, this will offer considerable insight into how

materials defects limit curvature coefficient. Results of this may help guide future work in millimeter wave detectors. Furthermore, all published tunneling heterojunctions for millimeter wave detection have focused on lattice-match. In Chapter 7 and 8, the use of alloying to control band alignment is discussed. In the context of millimeter wave detection, this is incredibly novel, as heterojunctions that are not lattice matched to a binary substrate can be explored, via the use of graded buffers. This opens up the material space tremendously. For example, in Chapter 9, results will be shown where, instead of utilizing InGaAs/GaAsSb lattice matched to InP and then doping, the composition is graded and tweaked so that no doping is required to obtain the desired alignment of bands. The potential benefits of this are that a lower dopant density may lead to lower dopant states which could degrade performance, and that if doping is not required to control band-alignment, then doping profiles could be incorporated to change the shape of the bands instead, such as controlling how they deform in response to a potential. Ultimately, this adds an additional design degree of freedom.

Chapter 2: Establishment of Theoretical Model and Experimental Plan

2.1 Introduction

The previous chapter discussed the main applications of devices that obtain novel electrical characteristics via the band edges, be it either steep subthreshold swing switching in transistors, or high-curvature in millimeter wave or other types of square-law detectors. The fundamental question to be answered is, how steep are the band-edges, what limits their steepness, and how can it be controlled? The vast majority of this thesis is an experimental approach to answer this question. However, the experimental results of this thesis are also complimented by a set of theoretical results that form the framework in which experimental results can be obtained and interpreted. The goal of this chapter is to provide a theoretical background of the underlying physics, and to present theoretical results and models that were developed as part of this thesis work for the sake of analyzing and interpreting experimental data.

2.2 Bulk vs Interfacial Band-Edge Steepness

In terms of how "sharp" the band edges can be, there is already a precedent: The Urbach tail. When examining the optical absorption coefficient as a function of photon energy, there should, to first order, be an infinitely sharp drop at the band-gap energy, as photons below the energy of the band-gap cannot be absorbed. However, in reality, the drop-off in absorption coefficient has a slope associated with it, and this is a characterization of the steepness of the band-edges. This tail is due to a variety of factors, including phonons and sub band-edge electronic states. However, it is important to establish that the Urbach tail is actually **not** a good standalone approximation to how much steepness could be expected in an actual electronic device. For example, the Urbach tail of InAs is shown in Figure 2.1a.

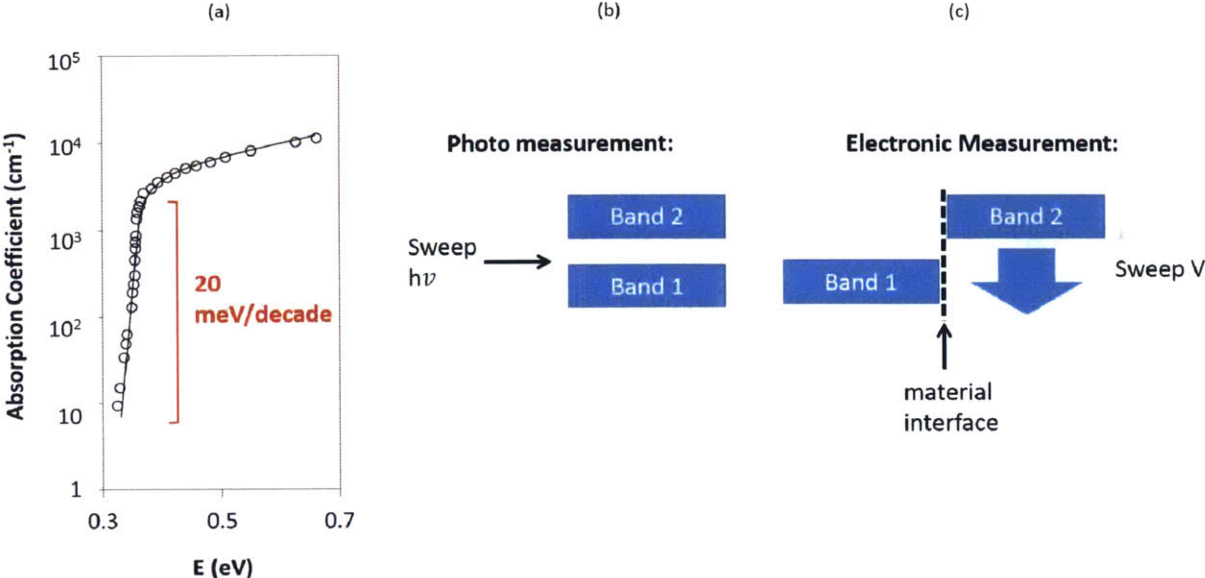


Figure 2.1 a) Absorption coefficient of InAs as a function of energy, showing an Urbach tail at low energy, which has been expressed as a slope of meV of energy per decade change in density absorption, making it analogous to a subthreshold slope. b) Schematic of measuring band-edge steepness optically, where no bands are moved in energy, c) Schematic of measuring band-edge steepness electrically, where bands must be moved and an interface is required

A band-edge steepness can be extracted from the plot giving 20 meV/decade. This number is the number of meV of energy over which the joint density of states decreases by an order of magnitude. It is cast in this way so that it can easily be compared to a density-of-states switch, where this would translate directly into a subthreshold slope under ideal conditions. This is because the subthreshold slope is directly measuring the change in joint density of states, as was described in section 1.3.2.

One might conclude then, that the true steepness obtainable from the InAs band-edges is 20 mV/decade. However, it is important to establish that measuring this optically is very

different than electrically. Optically measuring a band-edge steepness gives a bulk measurement. Because the input is a photon, it can easily be swept in energy, while bulk tunnel transitions (optical absorption) occur throughout the material, and the transmitted photon intensity is measured (Figure 2.1b). Electrically, however, the input is an electron, not a photon. Because of scattering, electrons cannot be "swept" in energy and inputted one energy at a time. Therefore, the only way to harvest band-edge steepness is to instead input electrons at all energies together, then sweep one band up and down using an applied bias and examine how many electrons get through (Figure 2.1c). In other words, measuring or harvesting band-edges in an electrical device requires transport. This means that, instead of occurring in the bulk, all of the tunnel transitions must occur across a two-dimensional interface. Since it is a heterojunction, this interface is likely to have considerably more defects than the bulk. Therefore, we expect the effect of defects to be more pronounced electrically. For this reason, a distinction will be made in this thesis between "bulk band-edge steepness", which is seen in optical measurements, and "interfacial band-edge steepness" which is the relevant metric for electrical devices that harvest the band-edges, and which has never been studied before.

2.3 Factors that Affect Interfacial Band Edge Steepness

There are conceivably a few physical aspects that should have an effect on how steep the band edges are at an interface. The goal of this thesis is to experimentally study such effects. Nevertheless, an estimate on their likelihood of significance can be done now in order to guide experiments.

2.3.1 Interfacial Defect States

The most obvious, and most likely to control the steepness, is defects. Interfaces are known to have defects, and particularly in epitaxy, we expect some degree of dislocations due to lattice mismatch. [60] Further, these dislocations can attract point defects [61] as will be described in Chapter 5. These defects have energy states associated with them that often tend to be in the band gap. Depending on the type of defect, they can be anywhere in the band gap, and can go quite deep, especially in the case of dislocations[61]. Figure 2.2a shows an example of a heterojunction biased such that there is no band overlap. In principle, there should be no tunneling and the current should be zero. However, in reality, there are defect states throughout, especially at the interface. Therefore, there is still a conductance path across as electrons can easily fall into defect states and tunnel across. Ultimately, these states will "blur" the band edges, causing a lower band-edge steepness. This is demonstrated in Figures 2.2 b and c. Again, the band-edge steepness is defined as how the density of states changes with energy, and in the application of a transistor, the subthreshold slope is just a measurement of how the density of states in both layers change with energy (via a sweeping of voltage). So Figures 2.2 b and c then demonstrate a band-edge steepness, which can also be thought of as an ideal subthreshold slope in the absence of other three-terminal parasitics. Figure 2.2b shows what the steepness might be without the interfacial defect states. In this example, it is shown as the same steepness as the Urbach tail from Figure 2.2a. However, when the defects are taken into account, we expect a tail of states extending into the band gap as shown in Figure 2.2c. As a result, the band-edge steepness is less steep

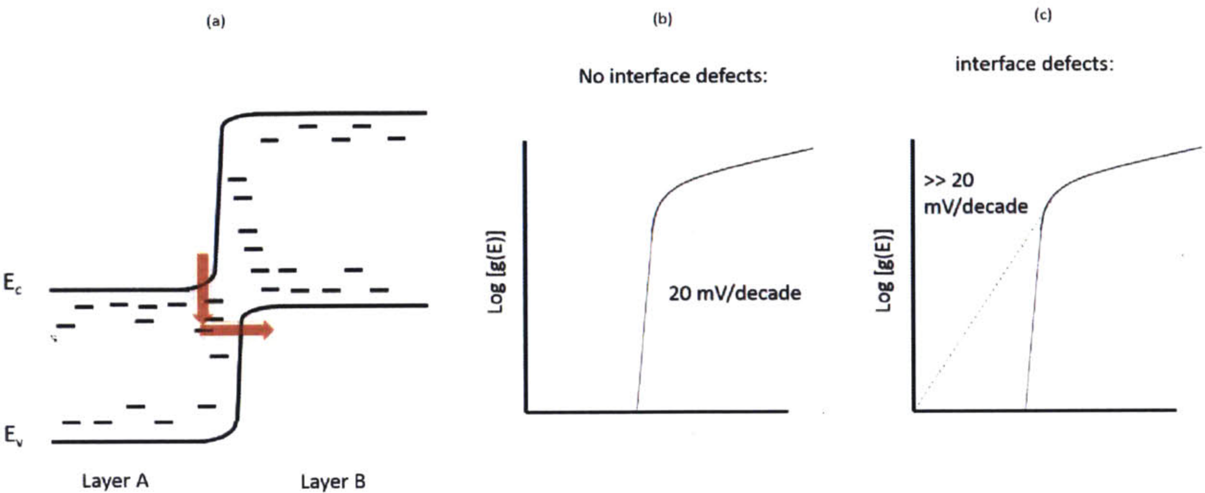


Figure 2.2 (a) Heterojunction biased such that it is in off-state. A mechanism involving an interfacial defect is shown, allowing tunneling to still occur, (b) Density of states without interface defects, which would translate to a subthreshold slope without interface defects, (c) same as (b), but with a tail of defect states extending into the band gap

2.3.1 Defect and Inhomogeneity-Induced Nonuniform Band-Alignment

In addition to this effect, we also expect defects at the interface to have a second effect on the band-edge steepness in an electrical device, typically at the very high and low ends of the current swing. Defects should be expected to lead to inhomogeneous band-alignment at the interface due to local band-bending in the region of defects as well as fluctuations in materials composition and strain. Figure 2.3a shows an example of a type-III InAs/GaSb junction, with example defect states that provide a band-edge blurring via the mechanism described previously. In Figure 2.3b, an energy-band diagram is envisioned along a direction parallel to the junction in the region of an example defect, in this case, an acceptor-type deep level trap [62], one that might typically result from a dislocation. It can be seen that, in addition to simply creating an

energy state, it has also resulted in a region of band-bending around it. This may result from the electric fields associated with the defect. It can also result from the strain the defects creates on the crystal around it, which results in a change in bands due to the deformation potentials. The defect need not be a deep-level trap, rather, any defect concentration that is not uniformly spaced will result in nonuniform band-edges. [63]

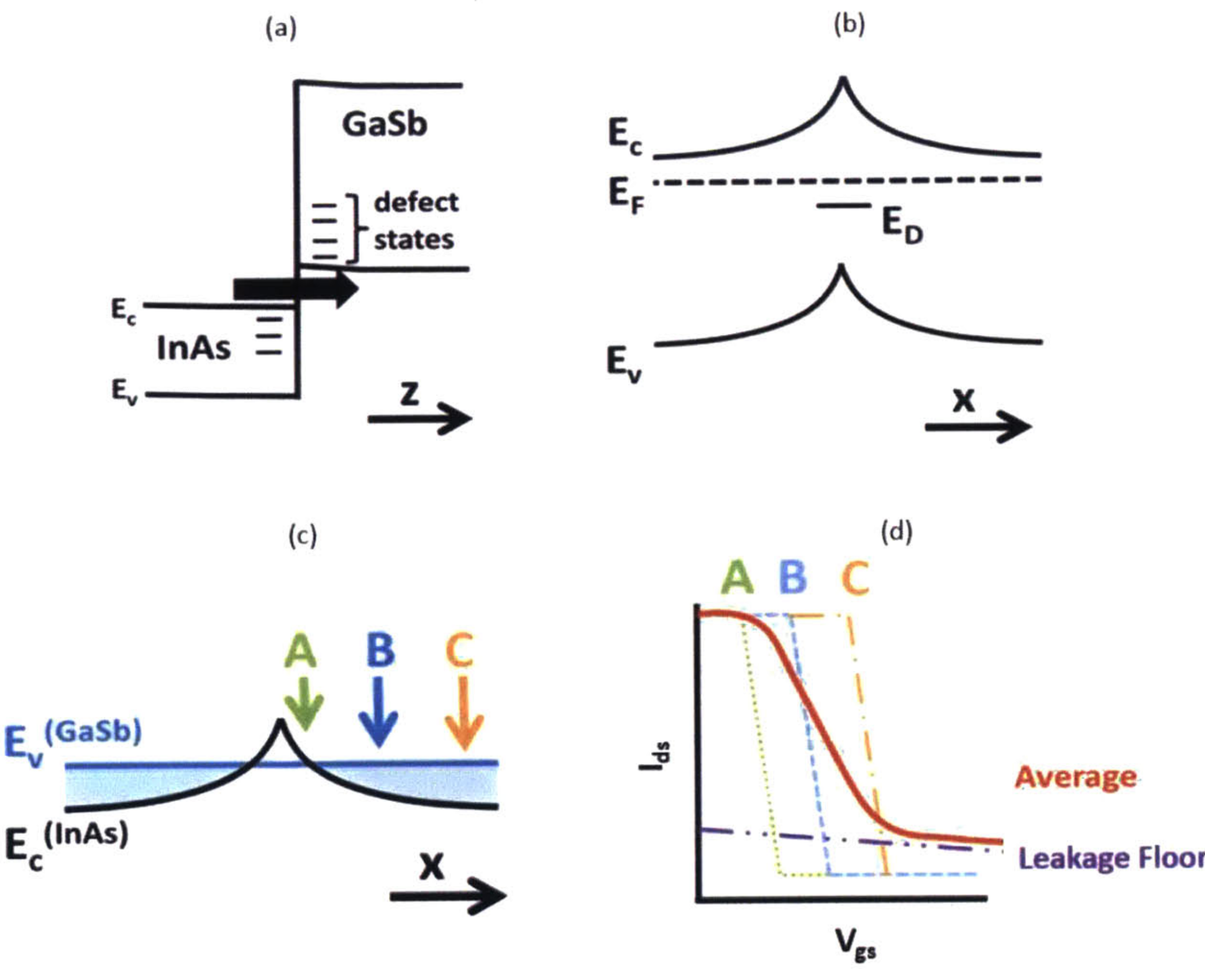


Figure 2.3 a) Band diagram of InAs/GaSb heterojunction. The arrow shows the direction of electron current due to tunneling, and defect states are also shown near the interface, which can also act as sites for tunneling. b) Band diagram of InAs along a direction parallel to the

interface, showing a defect state and an example of band bending around the defect (the degree of band bending is not necessarily this large) (adapted from [62]*) c) Same diagram as b) with the valence band energy of GaSb near the interface shown. Light blue region represents where tunneling can occur, and three reference points are marked as A, B, and C. d) Transfer curve showing sharpness of switching for current at each point A, B, and C, as well as a trap-assisted leakage current. The solid red curve shows the measured result, which is a weighted average of all of these conductance pathways, and the subthreshold slope is considerably less steep*

This means that such defects can alter the local band-alignment, as shown in Figure 2.3c, where the GaSb valence band right near the interface is shown and is assumed to not be affected by the defect for simplicity of illustration. Wherever the GaSb valence band lies below the InAs conduction band, tunneling can occur. This is illustrated as the light-blue region in Figure 2.3c. Three arbitrarily chosen points are labelled on this figure. As a voltage is applied to move the bands apart, tunneling will turn off at point A before point C. Figure 2.3d shows the result on subthreshold slope, where the slope is an average of the slope of every point. Even though the local steepness at each point may be steep, the effect of averaging the various shifted steep transitions results in a much less steep average.

The diagram also shows a leakage floor, which is the point at which current begins crossing the junction through some other mechanism. This will be examined more in detail in Chapter 6. If the leakage mechanism still involves trap states, than the magnitude of the leakage current will also depend on material quality.

Additionally, the term inhomogeneity is chosen to describe the material junction because this term encompasses not only crystal defects, but anything that result in an interface that is not

homogeneous. Particularly, this also encompasses the realistic degree of intermixing that should be expected as result of the high temperatures associated with epitaxy. If this intermixing were perfectly uniform, it would simply grade the bands slightly and result only in a slightly larger tunnel barrier. However, if there is any minor fluctuation throughout the plane of the interface, then it will also result in nonuniform band-alignment in the same manner as just described. Additionally, the intermixing may even respond to the strain fields of defects, resulting in different compositions near defects such as dislocations - an effect that was proven to exist recently.[64]

Figure 2.4 shows a relatively simple calculation that demonstrates the order of magnitude on which to expect defects to play a role. A relative transfer curve is calculated simply by using equation (4) in Chapter 1, assuming only the density of states term is being modulated by the gate voltage. The black shows an ideal switch (a perfect band edge). The red shows the steepness for a band-edge with a tail of defect states decaying by 3 orders of magnitude over 150 meV into the band gap. This results in a slope that is already only slightly steeper than the thermal limit. The green curve results from when band-alignment fluctuations are introduced. In this case, the bands are varied up to 150 meV in energy within regions that represent 1% of the material as a whole. The result is a significant blurring of the slope. Also, as can be seen, this causes a significant loss of steepness at high current, and a small loss at low current as well. This is because in the high-current regions – especially in the first order of magnitude of the turnoff – the current is being significantly averaged by all regions. Eventually, however, the regions with less band overlap have heavily turned off (such as region A in Figure 2.3c), and only the region with the most overlap is still observed. Around this regime, the slope approaches the same slope as the red curve without nonuniform band alignment because only one region is dominating the

slope. However, at low current, the region with the most band-overlap eventually reaches the same order of magnitude as the leakage floor, and averaging starts to become significant again. If the leakage floor is very high and the total swing is only an order of magnitude or two, it is possible for the entire region of subthreshold swing to be dominated by nonuniform band-alignment. It will be seen in Chapter 5 that this tends to happen for InAs/GaSb because of its high leakage floor.

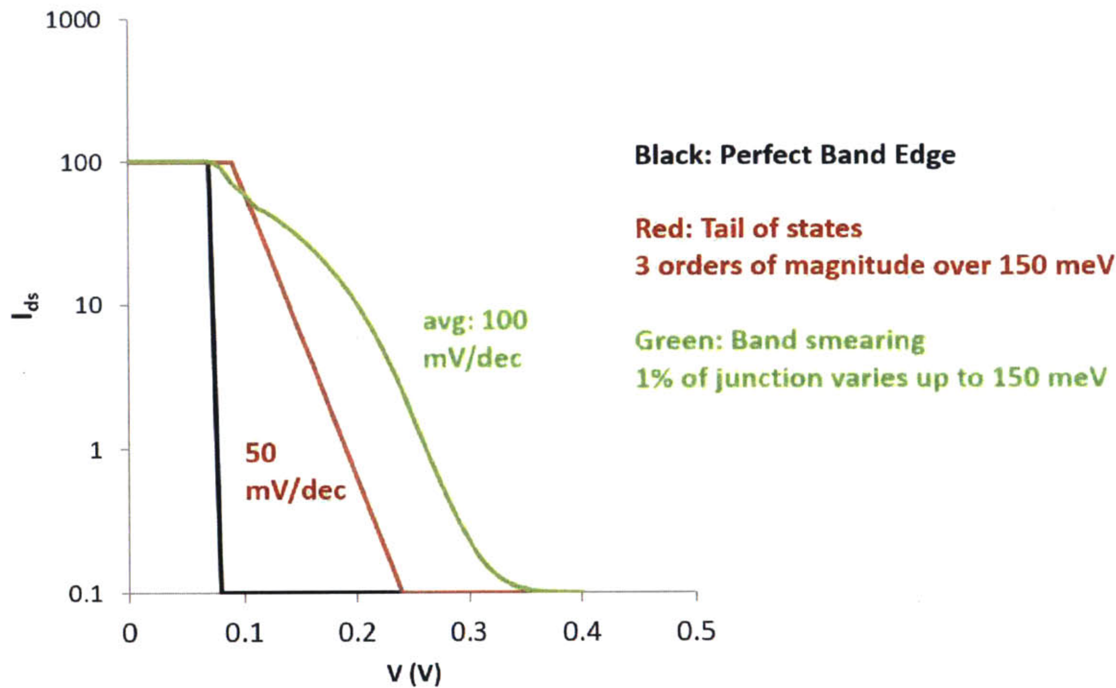


Figure 2.4: Calculated transfer curve for the case of a perfect band edge (black), a band edge with a tail of states decreasing in density be three order of magnitude from 0 to 150 meV into the band gap, and for the case of nonuniform band alignment (green) resulting from regions varying up to 150 meV over an area that represents 1% of the junction area

One additional interesting aspect of this calculation is that is shows a general shift towards the right on the I-V curve as defects are taken into consideration. For the case of shifting

from the black curve to the red curve, this is simply because there are more states extending into the band gap, so more voltage is required to turn them all off. When the effect of nonuniform band alignment is added, there is an additional shift to the right. This additional shift to the right implies a shift towards more type-III band alignment. This may seem odd, since defects can cause the band alignment to shift towards both type-II and type-III depending on which way they bend the bands. However, referring back to Figure 2.3c again, the most type-III region is always the last to turn off. Or, if sweeping in the other direction, it is always the first to turn on. Therefore, it tends to dominate the electrical characteristics, and a net shift towards type-III should be expected. It will be shown in Chapters 5 and 8 that, when comparing defective and nondefective samples, this is in fact exactly what is observed. The contribution of the regions that shift the other way can also be seen, but only slightly, as the small region at high current where the green curve dips slightly below the red curve.

The main conclusion of this calculation is that the density of states tail and nonuniform band alignment can cause an effect that is at a high enough order of magnitude that it may significantly affect the steepness obtainable in a device that utilizes the band-edges for switching or curvature. Of course, this is simply a rough estimate, and an experimental study is necessary, which will described in the following chapters.

2.4 Direct Measurement of Band-Edge Steepness: Establishment of a Figure of Merit

In Chapter 1, applications of band-edges were discussed, specifically tunnel field effect transistors and millimeter wave detectors. Both of these technologies are utilizing the band-edges and their steepness. The figure of merit for a TFET is the subthreshold slope, and the figure of merit for a millimeter wave detector is the curvature coefficient. However, the goal of this thesis

is to fundamentally study the band-edges and how steep they are, and then apply this to applications. Therefore, we do not necessarily want an application-specific figure of merit that convolutes other parameters. For that reason, it is necessary to establish a simple figure of merit that can easily be measured that captures the band-edge steepness in a pure manner.

Since the vast majority of study on band-edges in an electrical sense comes from TFETs, we will start by examining the TFET and its main figure of merit, and why this figure of merit in and of itself is not satisfactory. The problem is that the subthreshold slope – the TFET figure of merit – is a measure of a number of different variables. Overall, it is a measure of how the drain-source current changes in response to an applied gate bias. If that gate bias is being delivered from the gate to the junction perfectly and uniformly, and if the drain-source current is collected efficiently across the other two terminals perfectly, and without series resistance or parallel leakage paths, then the standard tunnel current equation (equation 4 from Chapter 1) accurately describes the situation. In such a case, the subthreshold slope is a pure measurement of the steepness of the band-edges. However, in reality, there are a variety of other factors that convolute this effect and may overpower it. In fact, it will be shown in Chapter 6 that the vast majority – if not all - of published TFETs are likely overpowered by leakage parasitics and are not measuring the band-edge steepness at all. Therefore a more pure figure of merit is desired.

To help understand this further, Figure 2.5 shows one possible design of a TFET device, using the common InAs/GaSb type-III material system. In terms of the dimensions and layout of the structure, the diagram is inspired by a previous report that gave one of the best results for a heterojunction [39]. There are a variety of parasitics that affect the subthreshold slope of this structure:

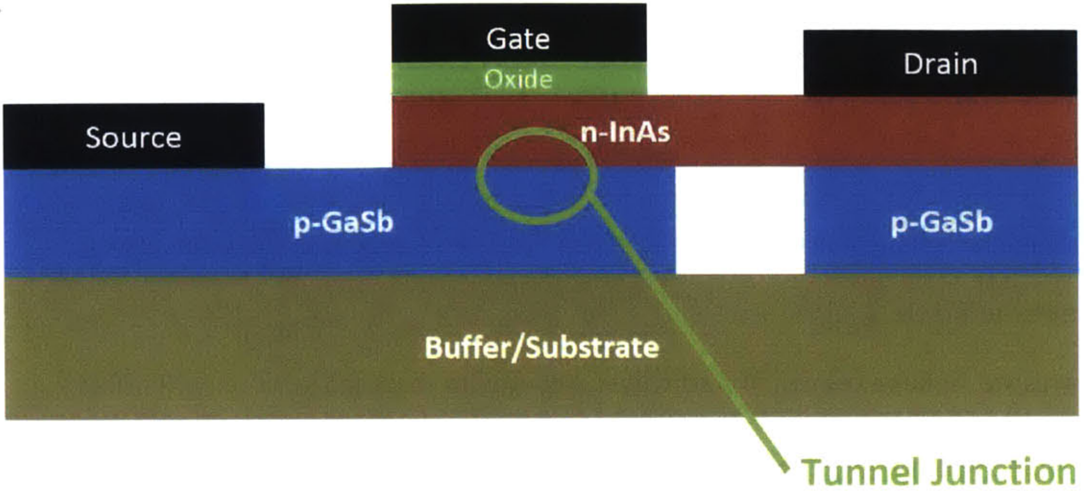


Figure 2.5 Example TFET design, adapted from [39]

1. **Oxide-InAs interface:** It is difficult to obtain a high quality gate-oxide interface with InAs, or for that matter, any material that is not silicon. As a result, there is always a high density of interface traps, which stretch out the capacitance-voltage curve, and result in a larger voltage being required to move the bands. This results in a less steep subthreshold slope.
2. **Lack of symmetry:** In the best designs, such as the one shown here, the tunneling is parallel to the field lines of the gate. However, since a drain must collect current, this requires conduction along a second dimension, which requires the symmetry of junction to be lifted. In other words, the current must "turn" at some point and conduct laterally. This results in the electrical field of the drain affecting the junction. This same effect occurs in MOSFETs and is known as drain-induced barrier lowering (DIBL). However, in a TFET, the result is even more catastrophic, because it can lead to diagonal field lines at the edges of the junction. [65], [66] This means that the potential is no longer delivered uniformly to the junction and the voltage drop across

the interface will vary with position, in a manner completely analogous to the nonuniform band alignment effect of defects from section 2.3.1. In the same regard, the device turns off nonuniformly, resulting in an averaging of the switching throughout, giving a net subthreshold slope that is less steep than the local slopes, similar to Figure 2.3d.

3. **Parasitic leakage (parallel) pathways:** Any way to shunt the junction will affect the subthreshold slope. In the example in Figure 2.5, any ability to leak into the substrate and then back into the drain allows the tunnel junction to be circumvented (it still must cross the tunnel junction, but under the drain where the drain field has kept it switched on). This has been proven to exist recently [34]. There is a much more important and universal parallel leakage pathway that will be described in Chapter 6 based on the observed results. If a leakage pathway can provide equal or greater current than the tunnel junction, then it will obstruct or even dominate the subthreshold slope.
4. **Parasitic Series Resistance:** Analogous to the effect of any parallel conduction pathway with a higher conductance, any series pathway with a lower conductance will obscure or dominate the device characteristics depending on whether it is on the same order of magnitude as the tunnel conductance or much lower. Again, this will be discussed in Chapter 6.

This thesis will isolate a few of these parasitcs as being dominant. However, another main goal of this thesis is outside of these parasitics: to understand the true band-edge steepness. In other words, a major goal of this thesis is to understand what is the steepest possible subthreshold slope of TFETs that could be obtained in the absence of these three-terminal parasitics. To do so, the effect of these parasitics must be deconvoluted. This can be realized by

developing a much simpler device with only two terminals, designed not to function as a TFET, but to serve as a vehicle to probe the tunneling at the band edges to understand band edge steepness. In Chapters 8 and 9, as the band alignment is changed to near the border of type-II/type-III, this simple structure effectively just becomes a millimeter wave detector. However, this structure can be used at all band-alignments to study the tunneling physics, without being specific to any one application such as TFETs or millimeter wave detectors. The device is shown in Figure 2.6

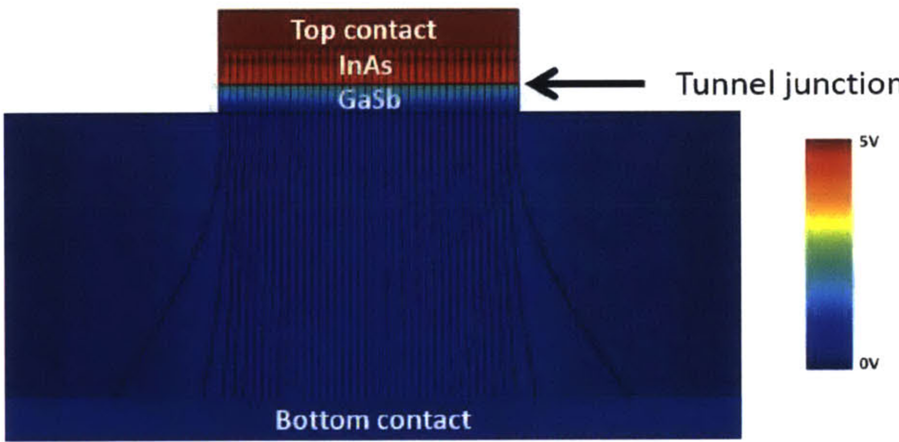


Figure 2.6 Design of a two-terminal device designed to probe band-edge steepness without convoluting with other three-terminal parasitics. The result of a current continuity finite element method simulation is overlaid, where the color corresponds to the potential in response to an externally applied 5 V bias, and the black represents select electric field lines.

In using only two terminals, the effects of a gate oxide (problem #1) are removed. The lack of symmetry (problem #2) is also fixed, as current does not have to "turn": tunneling occurs parallel to the field lines and the current effectively runs along these field lines as they cross the junction. There are no diagonal field lines until well below the junction when current begins to

spread throughout the substrate towards the back contact. This can be seen in the figure: the color represents the potential and black lines are the electric field lines. These values were calculated using a finite element method solution to the current continuity equation. With regards to problem #3, parallel leakage, the main issues related to lack of symmetry and the gate oxide have been removed. The only major parallel leakage is sidewall leakage: current leaking through the electronic states at the sidewall at the perimeter of the mesa. Fortunately, this scales with perimeter, while the main tunnel junction conductance scales with area. Hence, this parallel conductance can be defeated with geometry: making the mesa sufficiently large such that the sidewall leakage becomes insignificant solves problem #3. Finally, with regards to series resistance, we do expect there to still be series resistance resulting from both the contact resistance and the spreading resistance through the substrate. Fortunately, since this is simply a two terminal device, the value of the series resistance can easily be determined by extracting it from the series-resistance-limited regime of the I-V curve at high current, and then simply subtracting it out of the I-V curve. An explanation and demonstration of how this is done will be described in Chapter 5 in section 5.2.3. Therefore, problem #4 is solved, and we now have the ability to deliver a voltage difference directly to the tunnel interface, and examine the current that is collected as a result. In theory then, this should somehow be related to the band-edge steepness.

In one manner of thinking, this effectively accomplishes a three-terminal device without three-terminal parasitics, and so then extracting a subthreshold slope from this device would then give the band-edge steepness, since the subthreshold slope is simply the band-edge steepness in the absence of parasitics. The problem, however, is that the subthreshold steepness is a metric that is specific to a three-terminal device, not a two terminal device. This is demonstrated in

Figure 2.7. For the aid of explanation, the equation for tunnel current, which was originally shown in chapter 1 as equation 4, is shown again here:

$$I = C_1 \int^{\Delta E} [F_C(E) - F_V(E)] T_t N_C(E) N_V(E) \boldsymbol{dE} \tag{1}$$

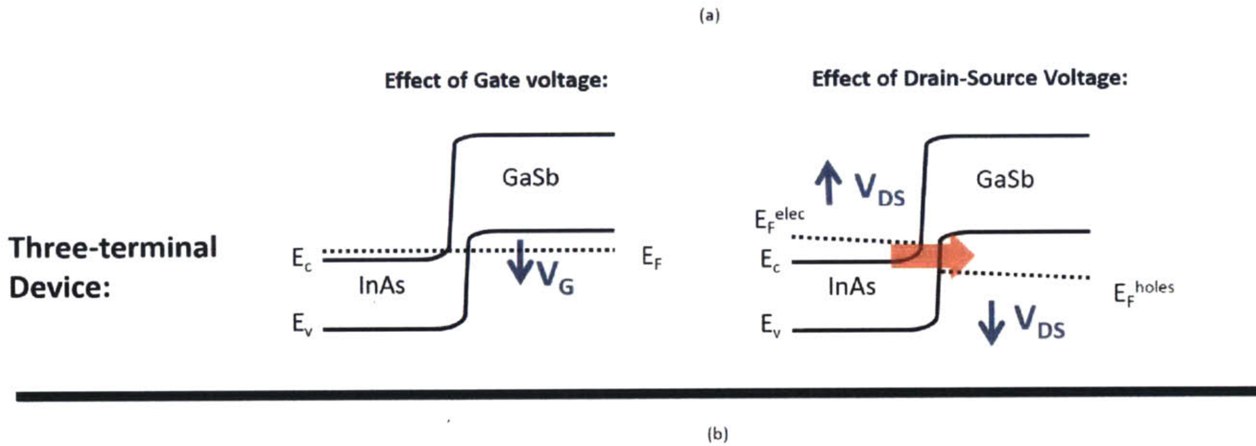


Figure 2.7 (a) Band diagram of an InAs/GaSb heterojunction in a TFET, showing the separate effect of the gate voltage and the drain voltage. (b) Same band diagram for the case of a two-terminal device, where only one applied bias controls everything

A three-terminal device, as shown in Figure 2.7a, has two separate voltages that are serving two separate functions. The gate voltage is the important **input**, which, in the absence of

three-terminal parasitics, is moving the band edges. Hence, it is changing the blue density of states $N_C(E)N_V(E)$ term. At the same time, the drain-source voltage is what collects the current, and hence works to support the **output**, which is the drain-source current. As such, it is changing the Fermi-level difference across the junction, which in the integrand of (1) is changing the Fermi-function difference at every value of energy: the red term $[F_C(E) - F_V(E)]$. The band-edge steepness is only a measure of how the integrated blue term $N_C(E)N_V(E)$ changes as voltage is swept. Hence, subthreshold slope accurately reflects the band-edge steepness when parasitics do not obscure or dominate it.

However, in the case of a two-terminal device, there is only one applied voltage that is changing both of these terms simultaneously. This is shown in Figure 2.7b. Hence, a "subthreshold slope" cannot be extracted from the I-V curve, as it would be meaningless. More importantly, such an extraction would not yield a value that depended only on the $N_C(E)N_V(E)$ term: it would yield a figure of merit that depended on both the $N_C(E)N_V(E)$ and the $[F_C(E) - F_V(E)]$ term. Hence, it is necessary to deconvolute these two terms, so that a figure of merit can be extracted from two terminals that depends only on the $N_C(E)N_V(E)$ term, making it characteristic of the band-edge steepness, and no other effect. This can be done by dividing out the $[F_C(E) - F_V(E)]$ term from outside the integral. The integral of $[F_C(E) - F_V(E)]$ over all energy must be equal to qV_a. Hence if the integral is divided by Va, then it is possible to remove the effect of the term. In this case, by diving by V, we are examining the absolute conductance. This was demonstrated by Agarwal [67].

$$\frac{I}{V} = \frac{I}{\int^{\Delta E}[F_C(E) - F_V(E)]dE} = C_1 \frac{\int^{\Delta E}[F_C(E) - F_V(E)]T_t N_C(E)N_V(E)\boldsymbol{dE}}{\int^{\Delta E}[F_C(E) - F_V(E)]dE} \quad (2)$$

This new expression is simply a weighted average of $T_t N_C(E) N_V(E)$, weighted by the thermal occupancy difference of each joint state:

$$\frac{I}{V} = C_1 \langle T_t N_C(E) N_V(E) \rangle \tag{3}$$

This expression looks good, because it has removed the $[F_C(E) - F_V(E)]$ term. Given that the T_t does not change significantly for density-of-states switches, we have a term that is characteristic of the band edges.

However, the $[F_C(E) - F_V(E)]$ is still present in the equation, as it acts as a weighting factor. Therefore, while its magnitude has been removed by normalizing by its integral, it could still affect the result if its shape changes and causes the terms to be weighed differently as applied bias changes. To examine this further, Figure 2.8. adapted from [67], shows a typical band-diagram for a tunnel junction, and shows the two terms side- by-side graphically: the first is the $T_t N_C(E) N_V(E)$ term. It is simply averaged, where the weighting factor is the second graph: $[F_C(E) - F_V(E)]$. If the applied potential is low, then the Fermi levels are close to each other on either side of the junction, and so the width in energy of the $[F_C(E) - F_V(E)]$ term results only from the shape of the Fermi distribution, leading to a width of about 4kT. As long as the shape does not change significantly with applied bias, then this weighting will not affect the result and the absolute conductance will depend only on the steepness of the band edges. Figure 2.8b shows the shape of this term for small biases. Up to 100 mV of applied bias, there is effectively no change in shape: the thermal width of the Fermi functions dominate, and the separation between the two is not noticed. Figure 2.8c shows how the term changes for higher biases. Eventually, the shape is lost: the function begins to flatten out and become wider. This is because the Fermi functions on either side of the junction have shifted far enough apart from each other in energy

that the difference can now be noticed. However, even at 400 mV applied bias, the weighting has changed only by a 4 factor at most. Referring again to Figure 2.8a, this expression is being integrated with the $T_t N_C(E) N_V(E)$ term, which is expected to change by many orders of magnitude with applied bias. Hence, even at an applied bias of 400 mV, a 4x factor should not affect the result as it is being integrated with a term that changes by orders of magnitude. The electrical measurements in this thesis never extend beyond this voltage range when a figure of merit is being taken.

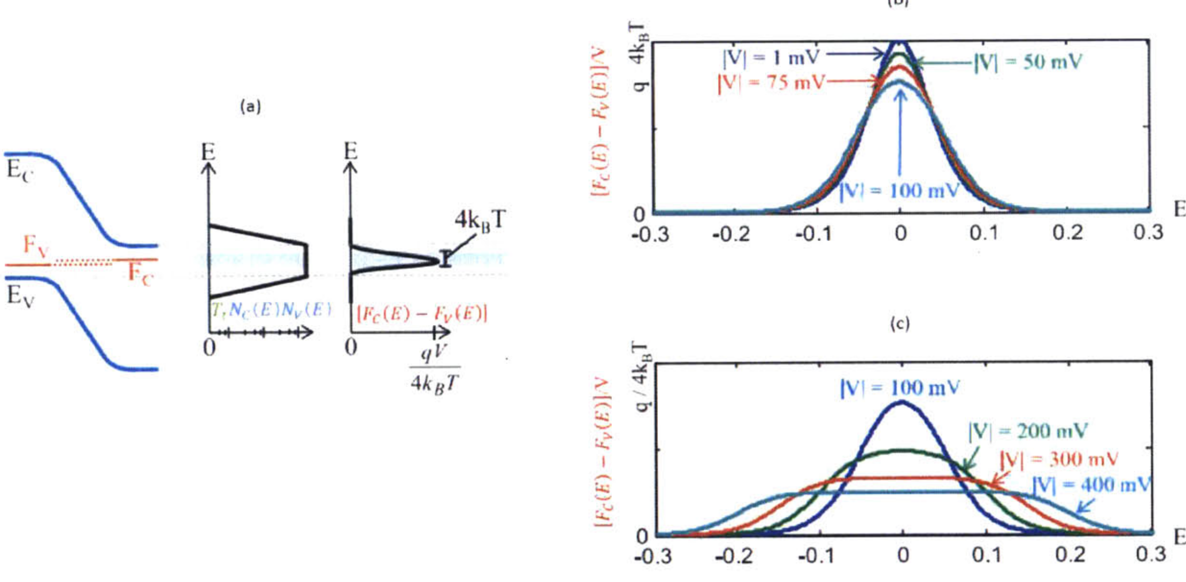


Figure 2.8 (a) Band diagram of two-terminal junction, with a graph of the $T_t N_C(E) N_V(E)$ term over energy and a graph of the $[F_C(E)-F_V(E)]$ term shown. (b) A graph of the $[F_C(E)-F_V(E)]$ term over energy, showing how it changes in shape with applied bias for low bias, (c) Same as (b) but for higher applied biases. Adapted from [67]

Therefore, we have found a function that depends on the band edges: the absolute conductance. The slope of how this changes with voltages gives the steepness of the band edges. Therefore, **the figure of merit of steepness for this thesis will be the conductance slope**:

$$conductance\ slope = \frac{dV_a}{dlog(I/V)} \tag{3}$$

By simply extracting this value from a two-terminal I-V curve, we obtain a figure of merit characteristic of how steep the band edges are. Of course, this must be done carefully while accounting for any parasitics. Chapter 5 will establish a robust procedure for reliably extracting this value.

This thesis will proceed by using the conductance slope as the primary figure of merit for the band-edge steepness, regardless of application. When considering the application, if a TFET is being considered, this figure of merit can be converted directly into a subthreshold slope, as the value of it is what a TFET would yield in the absence of three-terminal parasitics. For the purpose of a millimeter wave detector, this value can be converted into a curvature coefficient at small biases [68]

$$\gamma = \frac{-2\ln(10)}{conductance\ slope} \tag{4}$$

Of course, a millimeter wave detector is already a two-terminal device, and so the curvature coefficient can just as easily be extracted from the I-V curve directly.

Chapter 3: Experimental Methods

3.1 Introduction

This section describes the central methods that are used in this thesis to prepare and characterize material and devices. It is not exhaustive, as a variety of additional methods specific to certain analyses are also done and will be described as they come up in the results section.

3.2 Epitaxial Growth: MOCVD

The development and materials control of special III-V semiconductor heterojunctions is a cornerstone of this thesis. As mentioned in Chapter 1, this thesis focuses specifically on the growth of InAs/GaSb type-III heterojunctions, and InGaAs/GaAsSb type-II/type-III heterojunctions. All semiconductor heterojunctions were grown epitaxially using metalorganic chemical vapor deposition (MOCVD). Epitaxy requires adsorbed atoms (ad-atoms) to exist on a surface, diffuse to find the low energy sites, and incorporate into the crystal, resulting in a "grown" layer that has the same crystal structure as the underlying substrate. MOCVD delivers these ad-atoms chemically, via the controlled pyrolysis of chemical reactants (precursors) on the wafer surface. The group-III elements used in this thesis are Ga and In, and their precursors are trimethylgallium (TMGa) and trimethylindium (TMIn). They decompose to form organic byproducts (typically CH_4), leaving only the group-III atoms on the surface. The group V elements used in this study were P, As, and Sb. With the exception of Sb, group-V precursors are typically hydrides: for P and As, the precursors are phosphine (PH_3) and arsine (AsH_3). However, Sb cannot form a stable hydride, and so like group-III elements, it must be delivered in metalorganic form: its precursor is trimethylantimony (TMSb). A carrier gas must also be flown through the system to maintain a flow rate and pressure. This is typically either H_2 or N_2. While normally growth works in either carrier, we have found that antimonide layers can only be grown

in hydrogen, and hypothesize that this is because the Sb source is metalorganic instead of hydride. Therefore, since there are no hydrides involved, the only way to easily form a stable CH_4 product to remove the methyl groups on the precursors is to rely on the carrier H_2 gas as an additional source of hydrogen atoms. A comprehensive overview of MOCVD growth that goes into greater detail is given by Stringfellow.[69]

A custom-designed Thomas Swan/AIXTRON low pressure MOCVD reactor is used with a close-coupled showerhead. Figure 3.1a shows a schematic of the system. The close-coupled showerhead design prevents reactions in the gas-phase between group-III and group-V species, since the design prevents them from interacting until they reach the wafer surface. The showerhead consists of two separate gas plenums that feed into a series of holes on the showerhead ceiling designed to segregate group-III and group-V species. The substrates sit on a SiC-coated graphite susceptor that is heated resistively by a 3-zone graphite coil. The system is capable of growing on 2", 4", 6" and 8" wafers, but all wafers used in this study are 2". Depending on the sample, either an InAs, GaSb, or InP substrate is used. The pressure of the system is maintained at 100 torr using an Ebara dry mechanical vacuum pump and an MKS throttle valve. The reactor is enclosed in a N_2-purged glovebox which prevents oxygen and water contamination during loading and unloading of the reactor. Figure 3.1b shows a photograph of the reactor.

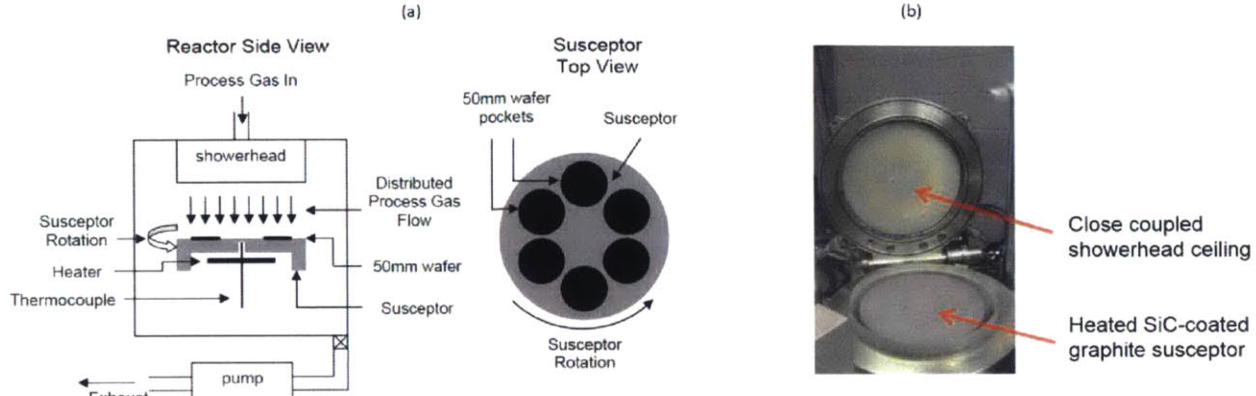


Figure 3.1 (a) schematic of the MOCVD system [70], *(b) labelled photograph of the system*

The susceptor temperature was measured using broadband optical pyrometry, which is accurate to within 1 °C. The temperature settings are calibrated on Si-coated quartzware with an uncoated 4" susceptor. The optical pyrometers are calibrated with a black-body temperature standard. The setpoint power level of the heater and the distribution of power to each of the three heater zones is calibrated to obtain an overall temperature within 1 °C of the desired temperature with no more than a 1 °C difference in temperature between any of the zones.

Table 3.1 summarizes the various important parameters for MOCVD growth and the settings used throughout this thesis. Temperature is arguably the most important parameter, as a variety of physical and chemical processes related to MOCVD are thermally activated, including precursor decomposition, surface diffusion, and desorption, as well as a variety of mechanisms within the crystal, including dislocation nucleation, interdiffusion, and other strain-relaxation and defect formation mechanisms. When growing ternary materials, as will be described in Chapter 7, the temperature control is very important, as the relative decomposition rates of all precursors depend on temperature, and a drift in temperature can shift the composition away from the target. Growth rate is also affected by temperature, primarily when growth conditions are in the surface-

reaction limited regime, where there is an exponential dependence of growth rate on temperature. This was found to be true for the growth of all antimonide layers. For the other layers, growth is typically in the mass-transport limited regime, where growth rate has a $T^{3/2}$ dependence.

Table 3.1 MOCVD Growth Parameters

Parameter	Value/Description
Films grown	Device layers: InAs, GaSb, InGaAs, GaAsSb Buffer layers: GaAsSb, InAsP Homoepitaxial layers: GaSb, InAs, InP
Substrates used	2" GaSb, InAs, InP, often cleaved into smaller pieces due to cost
Growth Temperature	Typical device layers: 530°C Device layer range: 465-580°C Homoepitaxial Layers: 530-650°C Buffer layers: 530-650°C
Precursors:	Group III: TMGa, TMIn Group V: PH_3, AsH_3, TMSb Dopant: Si_2H_6
V/III ratio	Arsenides and phosphides: 20-200 Antimonides: 0.8-2
Reactor pressure	100 torr
Wafer rotation	100 rpm

Another very important parameter is the V/III ratio. This describes the ratio of group-V precursors to group-III precursors delivered to the reaction chamber. This must be calibrated to ensure there is no excess of one species, which can lead to metallic droplets. Fortunately, group-V species tend to have a very high vapor pressure, and excess group-V species tend to spontaneously desorb instead of remaining on the surface. For this reason, arbitrarily high V/III ratios are chosen, to ensure the excess is always on the group-V side. Often, the only tradeoff of doing this is cost. Unfortunately, Sb is an exception as it has a low vapor pressure. This creates a

unique challenge for antimonide growth, as the V/III ratio has a narrow window that must be hit. This will be described in the next chapter.

Prior to loading a wafer, it is cleaned to remove a native oxide if the native oxide will not desorb easily during a bake. GaSb wafers were cleaned in HCl for 5 minutes, followed by two rinses in isopropanol. The oxide that is grown in the isopropanol then desorbs below the growth temperature. InAs substrates are cleaned in 1:1 HF:H_2O for 5 minutes followed by a rinse in water. The substrate is baked at 530°C for 10 minutes under protective AsH_3 flow to desorb the oxide put on the film by the water rinse. InP wafers do not require a clean. Heating up and cooling down always requires a protective group-V flow above 260 °C to replace group-V atoms that desorb. However, because of Sb's low vapor pressure, this is not done for heating up a GaSb substrate or cooling down with an antimonide layer as the top layer. Growth always begins with a homoepitaxial layer grown on the substrate, designed to bury any residual defects or oxide that may be on the surface. n-type doping of InAs and InGaAs is accomplished by Si using Si_2H_6 as a precursor, while p-type doping of GaSb was obtained with Si as well. All materials were often left unintentionally doped, and GaAsSb in particular, as at high enough As percentages there was no alternative, as Si would switch from p-type to n-type doping. Regardless of the structure, the top device layer was always an n+ or p+ contact layer designed to provide good contact resistance.

3.3 Device Fabrication and Characterization

The standard device used in this study is the two-terminal structure that was presented in Section 2.4. All two terminal devices were circular in shape from a top-down perspective and ranged from 12 μm to 40 μm in diameter. An important aspect of the device fabrication was to

ensure that the metal contact was not smaller in area than the mesa itself. This is because otherwise, current spreading occurs in the top device layers, resulting in a nonuniform potential being delivered to the tunnel interface. This results in nonuniform turn-on, which results in a loss of steepness similar to as described in section 2.3.1. In order for the field lines to all be perpendicular to the tunnel interface, as depicted in Figure 2.6 in Chapter 2, the metal then must be the same size as the mesa. This means that the two cannot be patterned separately, as then any misalignment would result in the metal shorting the sidewall of the mesa. To achieve this, a self-aligned process was used. First, the metal regions were patterned in AZ 5214E image reversal resist using i-line low vacuum contact photolithography. The tone of the resist was inverted using a 2.5 minute hotplate bake at 110 °C followed by a flood exposure of a dose > 200 mJ/cm^2. A top metal stack consisting of 120 nm Au/40 nm Pt/5 nm Ti was deposited via electron beam evaporation. The Ti layer acted as an adhesion layer and the Pt layer acted as a diffusion barrier. The resist was then lifted-off in acetone, leaving the metal regions. The top contact metal stack was then used as an etch mask to the device layers, allowing for a self-aligned mesa etch. InAs and InGaAs layers were etched in 2:1:2 $C_6H_8O_7$:H_2O_2:H_2O. GaSb and GaAsSb layers were etched in 1:5 NH_4OH:H_2O. In some cases, a SiO_2 inter-level dielectric layer was deposited, a contact hole opened in it, and a much larger contact pad was patterned using sputter deposition and subtractive photolithography. However this was only necessary for cryogenic measurements, as a room temperature probe station was able to easily probe down to the smallest mesa diameters. A complete fabrication process is given in Appendix A.

Current-voltage I-V curves were taken at room temperature on an Agilent B1500A Semiconductor parameter analyzer with Kelvin connections inside a Faraday cage. For low-temperature measurements, a Lakeshore Cryogenic probestation was used that was cooled

between 77 K and room temperature using liquid nitrogen, or between 4 K and room temperature using liquid helium.

3.4 Materials Characterization

3.4.1 Transmission Electron Microscopy (TEM)

Transmission electron microscopy (TEM) is one of the most important techniques of heterojunction materials characterization for this thesis. It is used to examine cross-section views of the heterojunctions and can highlight crystal defects and other important aspects of the materials junction and layers. A TEM consists of a high-energy electron source (in this work, 200 keV in energy) and a series of lenses that allow for a parallel beam of electrons incident on a thin sample. The beam interacts with the sample as it passes through, leading to contrast that is detected by the CCD detector below the sample. The contrast results from three mechanisms: absorption by the sample, elastic scattering (diffraction) and inelastic scattering. Diffraction is of particular importance because variations in the periodicity of the crystal due to defects will result in a difference in diffraction in that region. This results in contrast due to strain fields and deformation, making it possible to image crystal defects.

Cross section samples are prepared for TEM imaging by first gluing two small pieces of the same sample together such that the film sides are touching. The pieces are scribed along the [110] direction, and so the result is a cross section image with the horizontal axis along [110] and the vertical axis along [001]. The structure is then mounted to a pyrex stub with Crystalbond wax. The stub is loaded onto a Gatan 550 disc grinder. Wet grinding and polishing is performed on a series of progressively finer grits (500, 1200, 4000, followed by a 0.3 μm Al_2O_3 slurry). Grinding brings the thickness of the structure down to 10 μm or less in thickness. A copper grid

is then mounted to the structure with epoxy, and the structure is released from the stub by etching the Crystalbond in acetone. The samples are then milled with Ar+ ions at an accelerating voltage between 3 kV and 4 kV, depending on the material, and a beam current between 2.7 mA and 4 mA, using a Fishione Ion Mill. The milling angle is at 15° at the sample is rotated +/- 45°. Eventually, the milling produces a hole in the sample, and at the edges of this hole, the structure is thin enough to be electron transparent. If milled correctly, this hole will cross the interface line twice, and so there are two regions where the cross-section of the film is thin enough to image. Typically, the thickness for imaging is about 50-200 nm.

The samples are imaged on a JEOL 2011 TEM at 200 keV using a LaB_6 filament electron source. A double-tilt sample holder is used that allows the TEM sample to be tiled along two axes, making it possible to image on a number of diffraction conditions.

Typically, the sample is imaged either on the on pole or $\vec{g}$ = [220] diffraction condition. The on-pole condition is used for accurately measuring the thickness of layers as well as for high resolution TEM (HRTEM) where atomic resolution can be obtained. The $\vec{g}$ = [220] diffraction condition is used to image defects, particularly dislocations. Any defect that perturbs the lattice and leads to a strain that has some component along the diffraction vector direction will create contrast. For a dislocation, this means that the Burger's vector must have some component tangential to the diffraction vector. In other words, for a dislocation to generate contrast, it must be imaged on a diffraction condition such that:

$$\vec{g} \cdot \vec{b} \neq 0 \quad (1)$$

In a zincblende crystal with growth in the [110] direction, dislocations typically tend to be "60° dislocations" with Burgers vectors $\vec{b} = \frac{a}{2}[101]$, which gives a nonzero dot product with

the diffraction vector [220]. This condition also allows other crystal defects such as stacking faults, antiphase boundaries, and twins to be seen. The two-beam diffraction condition is set up by tilting the sample such that the beam strongly diffracts off of the (220) planes. Perturbations to spacing of those planes causes the electron beam to be scattered strongly. The objective aperture is then aligned so that those diffracted electrons are excluded from CCD which results in a dark region where the perturbation is, giving contrast.

3.4.2 X-Ray Diffraction

X-Ray diffraction (XRD) was used to determine composition and strain. There are a few types of scans done in this thesis. Firstly, symmetric scans were done, where there is no tilting of the wafer (ω), and so the diffraction vector is always perpendicular to the substrate. Since all wafers had a [001] surface orientation, it means that only plane spacing of planes parallel to (001) can be measured in a symmetric scan (but not all, since many are forbidden due to their structure factor). Typically for zincblende and diamond cubic, the [004] condition is studied, as it tends to give the strongest signal. This allows for the plane spacing of the (004) planes to be determined. An omega scan can be done as well, where ω is varied near a 2θ-peak. Since defects often lead to a slight tilting of the lattice, a defective film will tend to have a broader ω-distribution. A 2D map can also be generated where symmetric scans are done at varying ω-values, which is called a reciprocal space map. This allows the diffraction peaks to be seen, even if there is slight tilting of some of the layers. This is used for compositional analysis, specifically when compositional calibrations of ternary alloys are done. In a compositional calibration, various layers are grown on a substrate with varying flows of the three precursors. The films are grown thick enough so that the lattice constant relaxes to its equilibrium value. A reciprocal space map is generated, and a peak can be seen for each layer in the film, which gives the (004) plane spacing, which in turn gives the lattice constant. Hence, the composition of each layer can be determined.

XRD will also be used to measure strain, when an asymmetric scan is done. An asymmetric scan has a tilt on the wafer, so that the diffraction vector is no longer parallel to the wafer surface. Hence, planes such as (220) can be studied. When a film is tensile strained, the [004] diffraction will show the strain since the (004) spacing will be smaller via the Poisson ratio. The reverse scenario is true for compressive stress. However, if the composition is unknown, then not enough information is present to determine both the strain and composition. However, if an asymmetric scan is done, the lattice spacing of a plane with a component perpendicular to the normal of the wafer adds a second piece of information, and then both the composition and strain can be determined.

In this thesis, the scans described above were done on a Bruker triple-axis High Resolution X-Ray diffractometer.

3.4.3 Atomic Force Microscopy

Atomic force microscopy (AFM) is a very useful technique that can be used to probe the surface structure of a film. A Veeco Nanoscope IV AFM was used in tapping mode (the cantilever oscillates at a constant amplitude and height is adjusted). It is used to assess the surface roughness of samples, and to examine structures with significant structural defects (ie. holes in the film). It can easily see the atomic step structure on the surface, which reveals insights into how epitaxy occurred, and whether step-flow growth was maintained (where epitaxial growth proceeds as a movement of the steps), or whether a different growth-mode occurred, such as island growth, where the film grows as islands that may or may not later coalesce into a film.

Chapter 4: Binary Growth of InAs/GaSb

4.1 Introduction

A theoretical basis for measuring the band-edge steepness has been established, and a figure of merit – the conductance slope – has been selected. These next three chapters will begin by examining the simplest possible material system for tunneling: InAs/GaSb. Its simplicity is that it consists of binary materials, so composition control does not have to be carried out. At the same time, the two materials are very close in lattice constant. This chapter will focus on the MOCVD growth of these materials, understanding the types of defects that occur, and determining how they can be controlled. The subsequent chapter will then utilize various structures at different defect densities to extract conductance slopes and examine the band-edge steepness.

4.2 Growth of InAs and GaSb Layers

All layers were grown at 530 °C, as this has been shown in the past to be an ideal growth temperature for these materials. [71]–[74] Layer growth of both InAs and GaSb involves delivering the correct amount of group III and group V precursor to the epitaxial surface such that a binary compound is formed with the correct stoichiometry. Figure 4.1 a and b show the bulk phase-diagrams for InAs and GaSb. In both cases, the binary compound is a narrow region at 50% atomic concentration. For both materials, at the growth temperature of 530 °C, slightly more group-III concentration results in group-III-rich liquid droplets, while slightly high group-V concentration results in group-V rich metal. Since any metal on the surface would be extremely unfavorable for any epitaxy/device application, it is necessary to ensure the correct concentration of group-III and group-V species is obtained on the surface. Fortunately, this is easy for InAs, as the vapor pressure of As and related As-species is high enough, that excess As

on the surface simply desorbs, as was explained in section 3.2. Therefore, an arbitrarily high V/III ratio of 100 is chosen, meaning that the molar flow of arsine (AsH3) is 100 times higher than the molar flow of trimethylgallium (TMGa) into the reactor chamber. The excess As on the surface simply desorbs, and an InAs layer free of compositional defects is obtained.

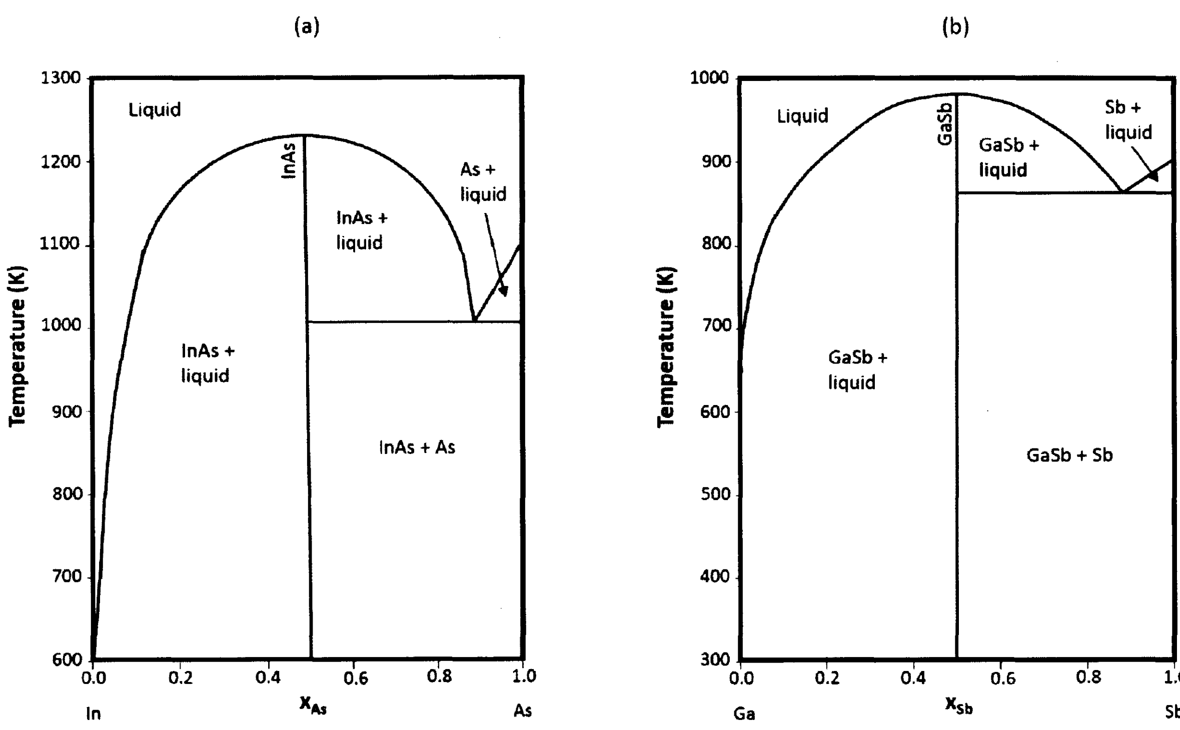


Figure 4.1 (a) Bulk phase diagram of InAs [75], *(b) Bulk phase diagram of GaSb* [76]

The situation is not as simple for GaSb, however, as Sb has a considerably lower vapor pressure, and will not desorb from the surface. As a result, a narrow window must be hit for V/III ratio, where excess group-III or group-V flow will result in excess group-III or group-V metal on the surface, respectively. Figure 4.2 shows the surface morphology of a GaSb film grown at 530 °C for three different V/III ratios. It was found that an ideal ratio is around 1.2zc : below this, Ga droplets form (Figure 4.2a), which also reject Sb from them as they cool and follow the

liquidus composition (as was shown on the left of Figure 4.1b), resulting in small pedestals on which the Ga droplets sit. At high V/III ratios (Figure 4.2c) nanoscale droplets and whisker-like features microns in length and diameter, and tens of nanometers in height form. At an ideal V/III ratio (Figure 4.1b), atomic force microscopy (AFM) reveals a compositionally uniform surface with atomic steps that can be observed on the surface.

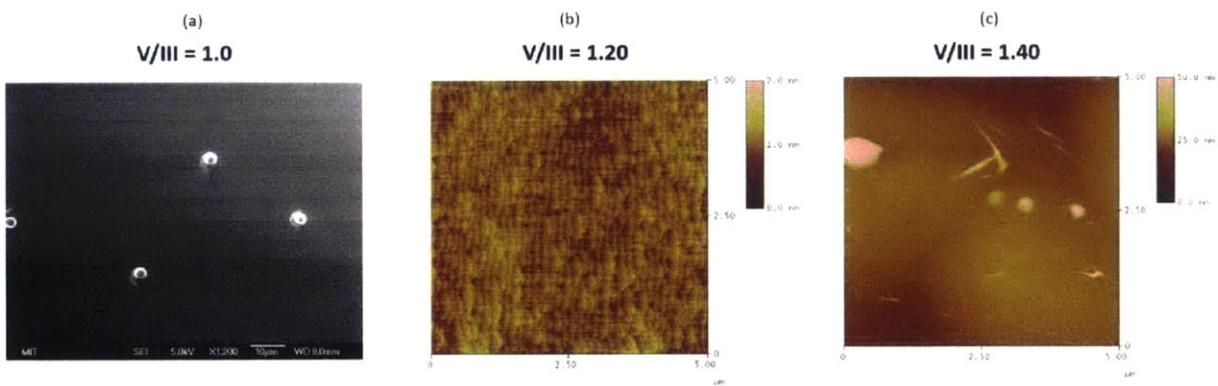


Figure 4.2: Surface of GaSb layer at varying V/III ratios. (a) SEM secondary electron image, scale bar of 10 μm showing Ga-rich droplets. (b)AFM image with a z-scale of 2 nm showing a featureless surface with a surface roughness of 0.143 nm rms. (c) AFM image with a z-scale of 50 nm showing Sb-rich precipitates on the surface

In addition to V/III ratio optimization, it was found that the GaSb growth rate also has an effect on surface morphology, in that at faster growth rates, compositional defects are still present even at optimized V/III ratios. A lower growth rate corrects this problem, and is presumed to be due to the fact that slower growth rates allow for enough time for all surface ad-atoms to diffuse and react, preventing any temporary regions where the V/III ratio of ad-atoms temporarily fluctuates from ideal. Figure 4.3 shows the effect of growth rate on surface quality.

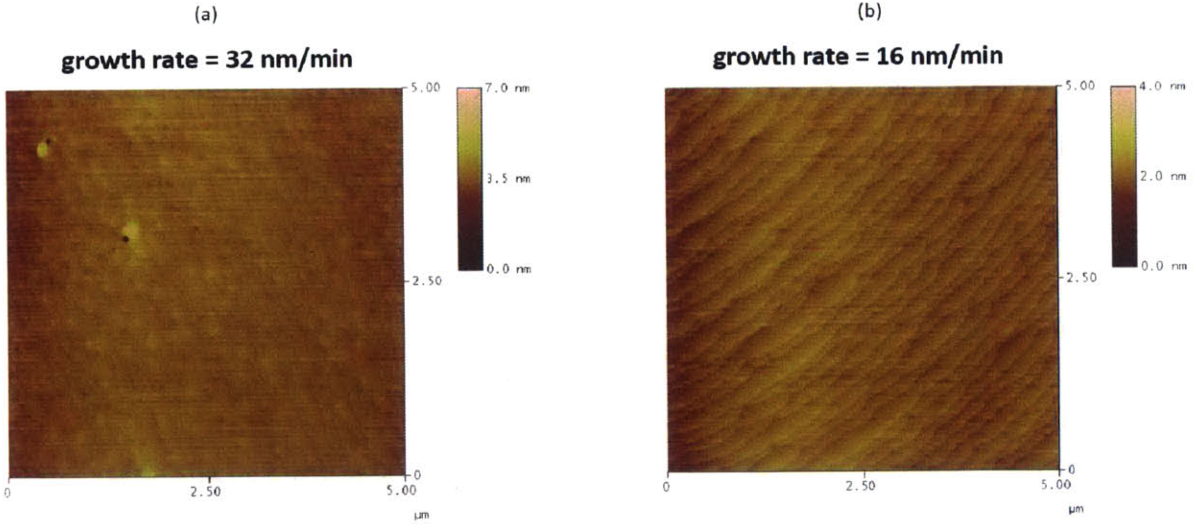


Figure 4.3 AFM image of GaSb surface grown at (a) 32 nm/min (z-scale: 7 nm), and (b) 16 nm/min (z-scale: 4 nm)

4.3 Growth of InAs/GaSb Heterojunctions and Effect of Growth Order

Even with optimized growth conditions for individual InAs and GaSb layers, growing heterojunctions of these materials was found to lead to many defects that ultimately result from the InAs-GaSb interface. Figure 4.4a shows cross sectional TEM of InAs grown on GaSb under the conditions described in the previous section. The interface appears to be very rough and numerous 60° threading dislocations appear to originate from the interface. This defect density is not characteristic of small-lattice-mismatch materials such as InAs and GaSb (0.6%), indicating that these dislocations are not created by the small tensile strain in the growing InAs layer. Figure 4.4b shows a high-resolution TEM (HRTEM) image, where dark regions can be seen, indicating a difference in strain or composition. The inset shows a Burger's circuit indicative of a 90° (edge) dislocation in this region with a Burger's vector, $\vec{b} = -\frac{a}{2}[110]$, which is indicative of very large lattice mismatch (typically more than 4%) and could be a result of strain-induced

island formation[60]. These observations are indicative that this relaxation is not due to the small lattice mismatch between InAs and GaSb, but rather due to much larger strain that must be present at the interface. Additional proof of this will be shown in section 4.7.

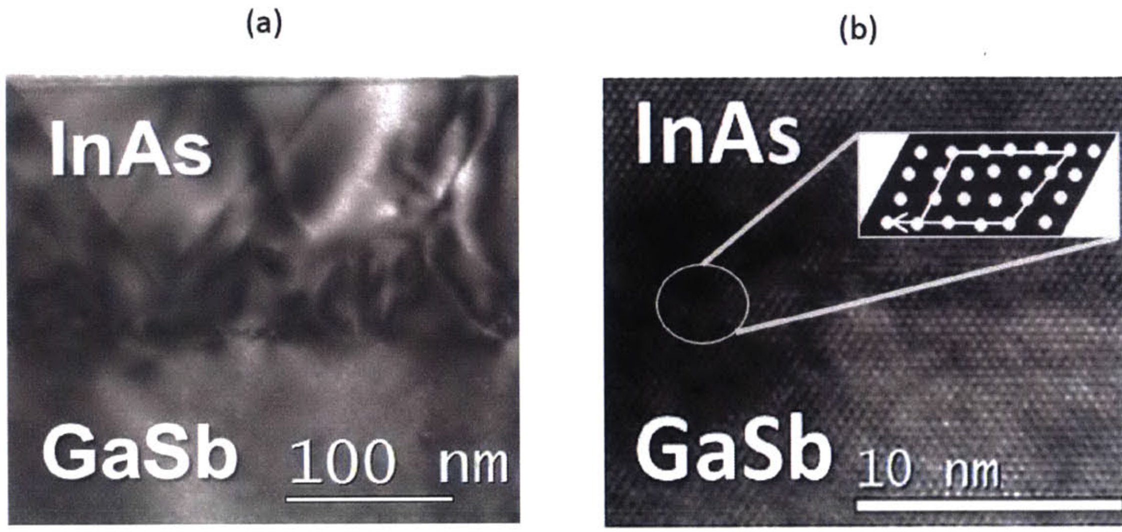


Figure 4.4: (a) Cross section TEM ([220] diffraction condition) of InAs grown on GaSb at 530°C showing dislocations originating from the interface. (b) Lattice-resolved TEM image of structure showing a dark region. Zoomed in portion shows a Burger's circuit drawn based on the identified lattice sites in the image.

It is likely that this strain at the interface results from intermixing. What is somewhat unique to this binary heterojunction system is that there is no common group-III or group-V element between the two materials. Because of this, a wide range of compositions can be obtained in an intermixing region of composition $Ga_xIn_{1-x}As_ySb_{1-y}$, spanning from the GaAs lattice-constant to the InAs lattice-constant, as shown by the red region in Figure 4.5. This could easily be a source of strain high enough to result in the defect levels seen in Figure 4.4. In theory,

bulk diffusion should avoid compositions resulting in large-strain, as the strain of these compositions would lower the diffusion potential. However, in the initial layers of growth, intermixing is more likely to occur via "exchange", where ad-atoms from the top layer can swap positions with a layer below. This process can be much faster than bulk-diffusion, but can only occur near the growth surface. Hence, exchange could lead to highly mismatched layers that are not thick enough yet to result in island-growth or dislocations. As the film grows thicker, the strain energy begins to grow, but since the intermixed region is now buried below the surface, only bulk diffusion can adjust the composition, which is much slower or even not possible. Therefore, islanding or dislocations will form. Figure 4.6 shows the most likely exchange mechanism, where Sb atoms in the GaSb swap with adsorbed As atoms initiating the InAs layer. This has been observed previously in MBE growth [51], [77], [78], and is due to the lower surface energy of Sb compared to As. The result is a GaAs layer. As this exchange can occur over a few layers, we expect a GaAs or As-rich $Ga_xIn_{1-x}As_ySb_{1-y}$ layer that will lead to large amounts of strain. Another mechanism that can occur at the surface is the Ga-carryover effect, which happens specifically in MOCVD growth with Ga in the lower layer[72], [74], [77], [79]–[81]. The mechanism is similar to the Sb-As exchange mechanism, except that the exchange reaction also involves the methyl groups in the precursor molecules, making it specific to MOCVD. This mechanism would lead to a Ga-rich $Ga_xIn_{1-x}As_ySb_{1-y}$ layer. Either mechanism alone would result in a much smaller lattice constant of the intermixed layer, and the mechanisms working together would cause the same effect: a Ga and As rich $Ga_xIn_{1-x}As_ySb_{1-y}$ which would have a much smaller lattice constant. Hence, these surface-mediated intermixing mechanisms are likely what lead to the strain.

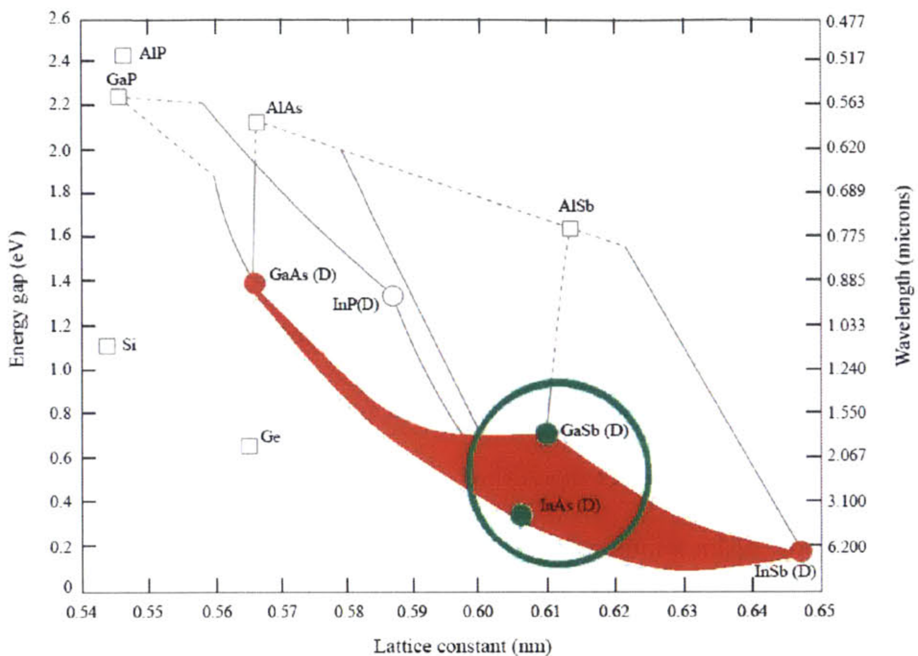


Figure 4.5: The Band-gap vs lattice constant diagram for III-V materials. InAs and GaSb are highlighted on the graph, and the filled in red region represents all possible compositions available from intermixing of the four elements.

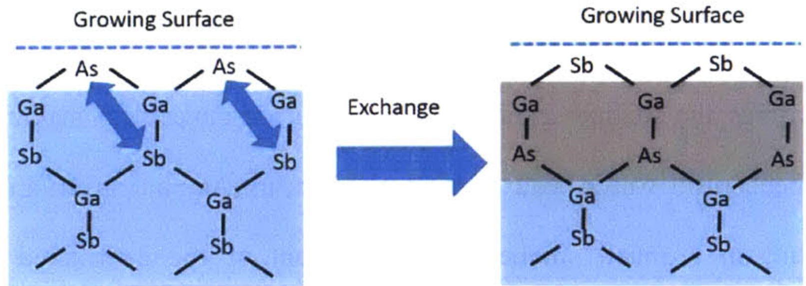


Figure 4.6: An illustration of As-for-Sb exchange, where Sb displaces As from the top surface, resulting in a GaAs layer at the interface

4.4 Effect of Switching Sequence on Defect Formation

The sequence of switching precursor gases at the interface can sometimes be used to control and suppress the exchange reactions, although we found this to be overall ineffective. Table 4.1 describes a variety of gas switching sequences that were attempted and their purpose. TEM images are compared for some of the sequences in Figure 4.7, and AFM images are compared in Figure 4.8.

Table 4.1 Gas Switching Sequences Employed for InAs/GaSb Heterojunctions

Sample:	Switching Sequence	Comments
A	Shut off TMGa, TMSb Immediately switch on TMIn, TMGa	Default sequence
B	Shut of TMSb Flow TMGa for 2 s at 8 sccm Switch off TMGa Flow AsH_3 for 2 s (130 sccm) Switch on TMIn	Force GaAs-like interface to prevent InSb-like interface
C	Shut off TMGa and TMSb Flow AsH3 for 2 s (130 sccm) Switch on TMIn	Same as B but prevent excess Ga
D	Switch of TMGa Continue to flow TMSb for 0.2 s (12 sccm) Switch off TMSb Flow AsH3 for 1 s (130 sccm) Switch on TMIn	Prevent Ga carryover by forcing GaSb surface to be Sb-rich. Allow a slight amount of As-Sb exchange
E	Switch off TMGa Continue to flow TMSb for 10 s (1.2 sccm) Switch on both TMIn and AsH_3	Prevent Ga exchange by making surface Sb-rich. Prevent As-Sb exchange by forcing an excess of Sb on the surface to prevent Sb from layers below from exchanging

TEM images indicate that all samples still contain dislocations originating from the interface, indicating that no gas-switching sequence was ultimately effective in preventing the exchange mechanisms that lead to the intermixing and strain. It is difficult to gauge from cross-

section TEM if any improvement is made, since different thickness of milled-samples lead to different concentrations of defects imaged. However, AFM provides a better view (Figure 4.8), as it can be seen that in all cases, the steps on the surface are not straight and uniform, as they were for the original GaSb layer, as can be seen back in Figure 4.3. This indicates that either growth of the InAs layer initiated as islands due to strain, or the high dislocation density has pinned the groups of the steps, disrupting the step-flow growth. In any case, it has led to roughness that likely results from the strain. It can be seen that switching sequences that introduce Sb at the interface seem to offer some improvement in surface roughness, which is consistent with the As-Sb exchange and Ga carryover models, as these sequences were designed to mitigate those two effects. Nevertheless, at no point is a step-flow growth seen in AFM, nor a dislocation-free interface seen in TEM. Therefore, while these switching sequences may offer some improvement, they are overall not fully effective in solving the problem of intermixing leading to strain buildup.

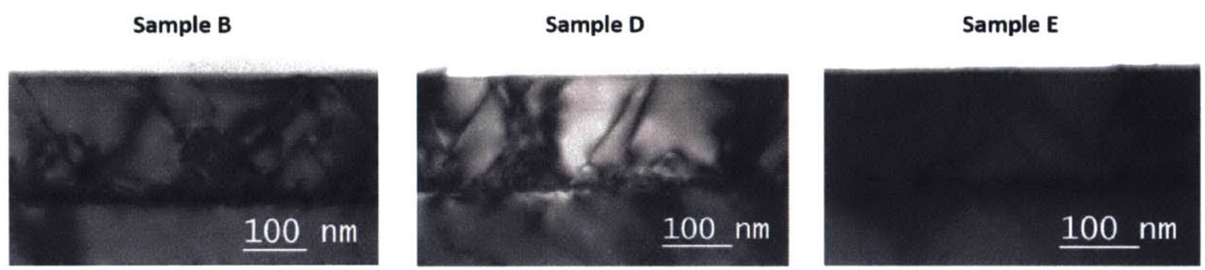


Figure 4.7: [220]-diffraction condition TEM images of various switching sequences.

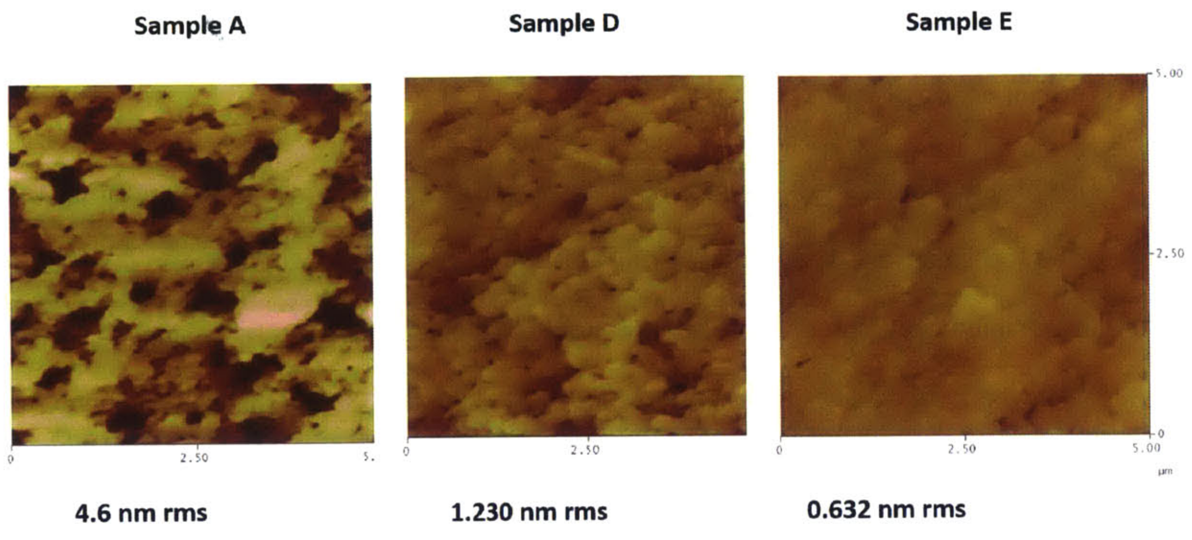


Figure 4.8: AFM images of top InAs surface for various pulsing sequences

4.5 Effect of Growth Order on Defect Formation

While designing gas switching sequences was largely unsuccessful in controlling the problem of exchange, there is the potential to solve the problem by changing the growth order. This is because both the As-for-Sb swap and Ga carryover processes are expected to be sensitive to the growth order of the interface, and hence should not occur when grown in the reverse order, with GaSb grown on top of InAs. This is because the As-for-Sb swap is driven heavily by the lower surface energy of Sb, and so it should only occur when Sb is in a lower layer in the structure and not the top layer. As for the Ga carryover effect, it is believed that the carryover occurs because Ga atoms in the lower layer react with the adsorbing TMIn to transfer methyl groups to the Ga atoms, which may allow it to become mobile and incorporate into the above film [82]. This methyl transfer is due to the stronger bond energy of the Ga-C bond than the In-C bond, and so the transfer should not occur between TMGa and In atoms in the scenario in which

InAs is being grown on GaSb. Therefore, it is expected that these effects should not occur if GaSb is grown on top of InAs.

Figure 4.9a shows cross-section TEM for a structure grown in the reverse order, with GaSb as the top layer, for layer thicknesses of 80 nm and 380 nm. In both cases, there are no longer any visible threading dislocations. The 80 nm thick GaSb layer on InAs shows no visible interface defects under these TEM conditions. For the case of the 380 nm thick layer, it is typical of relaxation in a low-mismatch interface: misfit dislocations can be seen at the interface, indicating that there has been relaxation, but without the heavily defective interface and thread density seen for the InAs/GaSb structure. Figure 4.10 shows a comparison of strain relaxation for the InAs/GaSb structure and the two GaSb/InAs structures, via glancing exit [224] reciprocal space maps. For InAs/GaSb in Figure 4.10a, the reciprocal space position of the InAs [224] peak indicates >80% relaxation. For the reverse order in Figure 4.10b, it can be seen that at 80 nm thickness, the GaSb remains mainly strained with about 5% relaxation, while the 380 nm thick layer (Figure 4.10c) shows about 63% relaxation. The fact that the thread density remains significantly lower than the InAs/GaSb sample, even after significant relaxation, and that the GaSb can be grown considerably thick before relaxing and still not reach the same relaxation level as for InAs on GaSb, indicates that the intermixing-driven dislocation formation observed for InAs/GaSb heterojunctions is not present when grown in the reverse order.

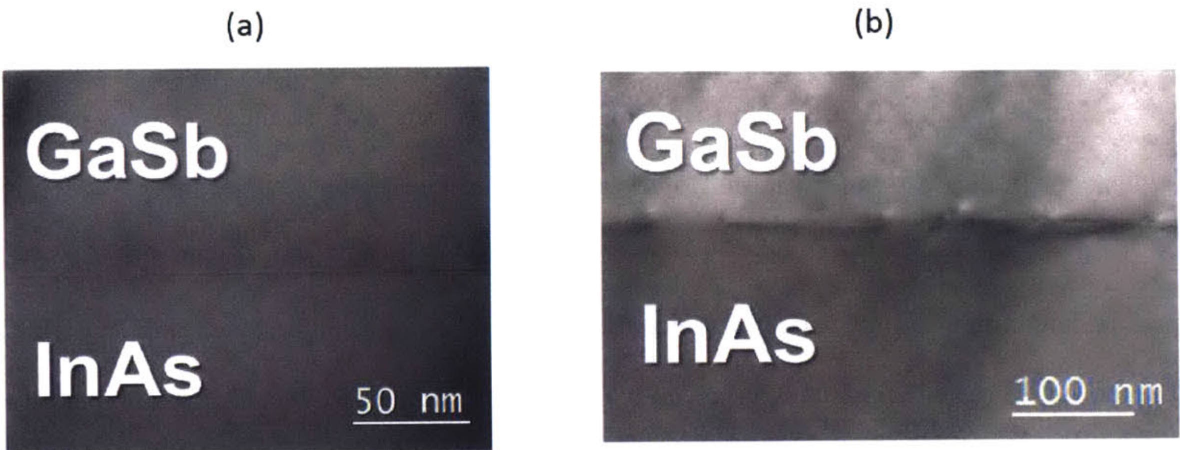


Figure 4.9: Cross section TEM ([220] condition) of a) 80 nm thick GaSb grown on InAs at 530°C, b) 380 nm thick GaSb on InAs

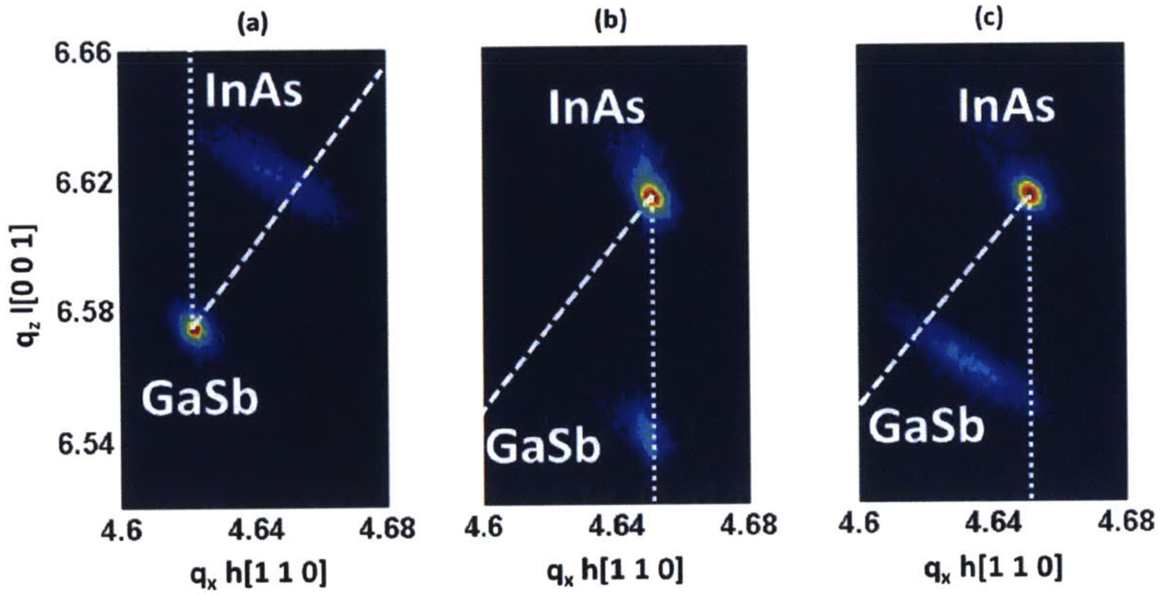


Figure 4.10: [224] glancing exit reciprocal space maps of a) InAs/GaSb junction (82% relaxed) b) 80 nm GaSb / InAs (<5% relaxed) c) 380 nm GaSb/InAs (63% relaxed). Short dashed line shows the line of strain, and long dashed line shows the line of relaxation

4.6 Use of Strain to Suppress Defect Formation

Thus far, it has been shown that GaSb can easily be grown strained on InAs, and this prevents misfit dislocation formation. However, at the normal growth temperature of 530 °C, it is not possible to grow the InAs strained, as it relaxes immediately due to the much larger strain of intermediate compositions. However, since the GaSb can be so easily strained, it is possible to "pseudo lattice match" the InAs to the GaSb to prevent relaxation by keeping the GaSb strained on the InAs lattice constant, and then growing the InAs layer on top of the GaSb layer without lattice mismatch. To test if this is effective, we grew InAs layers on both the strained and relaxed GaSb layers to determine if keeping the GaSb underlayer strained provides any improvement. Figure 4.11a shows a cross-section TEM image for InAs grown on the 380 nm relaxed GaSb layer, and is heavily dislocated in the same manner as when InAs is grown on GaSb substrates. However, as shown in Figure 4.11b, when InAs is grown on the strained GaSb layer, it has no observable defects in cross-section TEM. This seems to indicate that growing the InAs on GaSb that has been strained to the InAs lattice constant prevents either the intermixing that leads to high strains, or the relaxation of these high strains. It is possible that that intermixing is reduced because lattice-matching could prevent island formation and so the entire surface is quickly covered in InAs, making As-for-Sb exchange and Ga carryover more difficult. Alternatively, it is possible that the intermixing still occurs, but the film mismatch is required for relaxation. This is conceivable since the intermixing by exchange or carryover would have a shrinking driving force as strain begins to build up, and it is unlikely that it would continue beyond a point where enough strain energy was built up for relaxation mechanisms to activate. Hence, we would expect the diffusion to lead to unrelaxed strain. Additional strain would then be required to lead to relaxation and island formation. Hence, the film mismatch is needed to bring the strain energy

to the required minimum for strain relaxation to be energetically favorable, and removing the lattice mismatch between the GaSb and InAs prevents the strain relaxation from ever becoming energetically favorable.

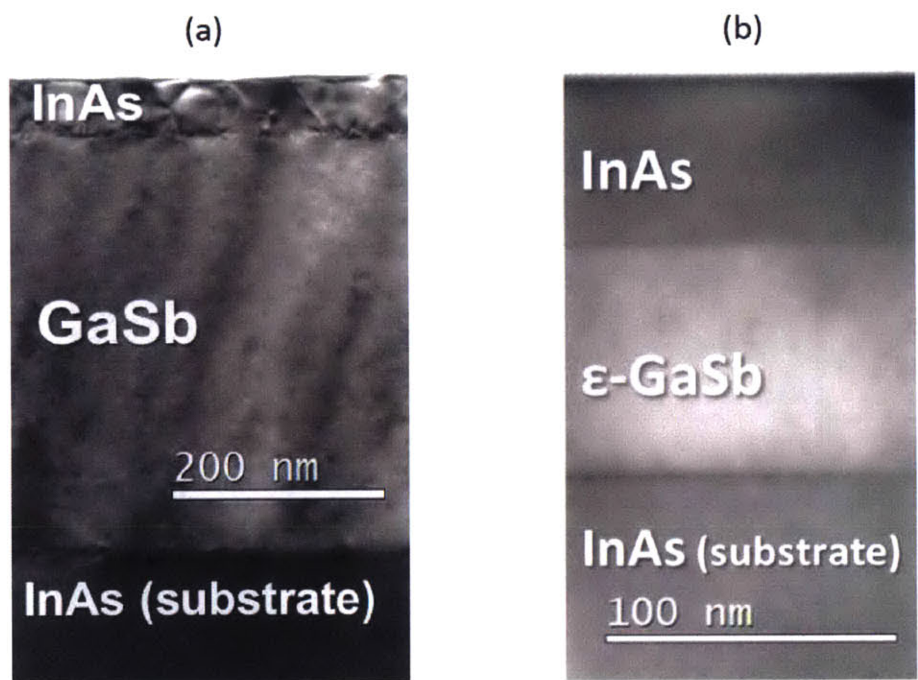


Figure 4.11: TEM images of: a) InAs grown on 380 nm thick (63% relaxed) GaSb, b) InAs grown on 80 nm thick strained GaSb (images are at different magnifications to show all layers. Top InAs layer is the same thickness for both)

AFM data of the top surface in each of these two structures (Figure 4.12) provides a view consistent with the model of intermixing leading to strain buildup. It can be seen that for both structures, the GaSb surface shows uniformly-spaced atomic steps, indicating that normal step-flow growth is occurring. For the strained GaSb (the sample in Figure 4.11 b), this step flow growth continues for the InAs layer, as atomic steps along one direction can still be seen,

indicating that InAs initiated as a single layer. However, for the case where GaSb is relaxed, and InAs is growing with tensile strain (the sample from Figure 4.11 a), uniformly-spaced steps are no longer seen on the surface, and a platelet-like system of steps is seen instead. Similar to as seen in section 4.4, this once again indicates that either growth of the InAs layer initiated as islands due to strain, or the high dislocation density has pinned the groups of the steps, disrupting the step-flow growth.

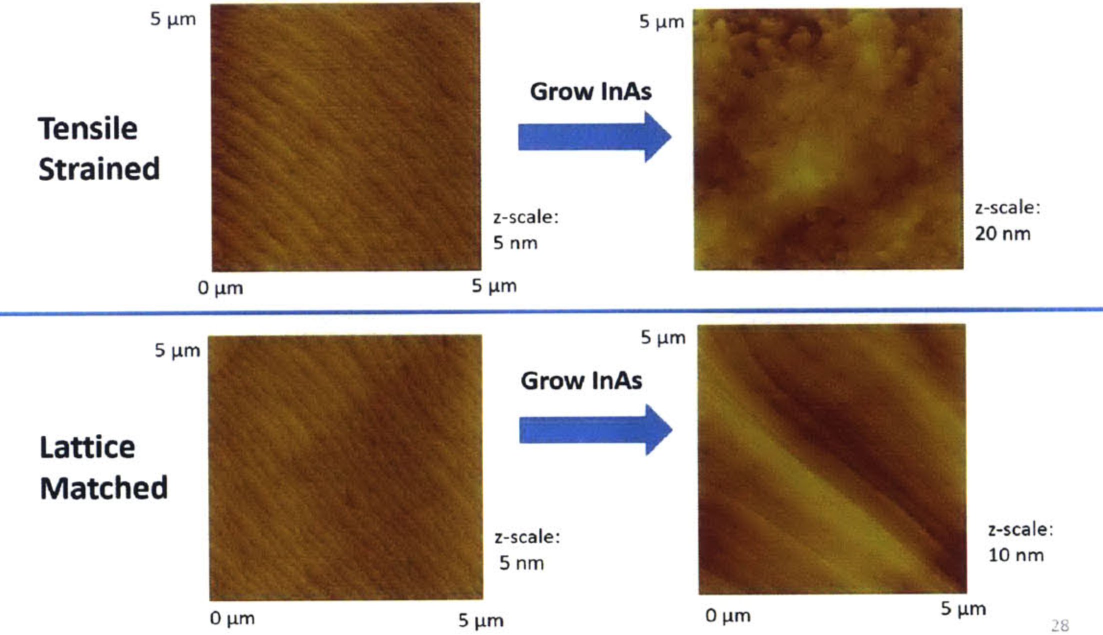


Figure 4.12: AFM image of GaSb surface (left) and InAs surface (right) for the case of a tensile-strained InAs layer and a lattice-matched InAs layer.

4.7 Use of Growth Temperature to Suppress Defects

While growing in reverse-order or with a strained GaSb underlayer prevents dislocation formation, these structures could lead to difficulty in later fabricating a three-terminal device. Another possibility, given that dislocations ultimately originate from intermixing, would be to lower the growth temperature of the InAs layer to prevent this intermixing. We attempted this by

reducing the growth temperature from 530°C after growing the GaSb layer, down to 490 °C and 465 °C for InAs growth. Cross section TEM results are shown in Figure 4.13 for three InAs growth temperatures. There is an improvement in interface quality and a drop in threading dislocation density as temperature decreases, indicating less intermixing and dislocation formation. Figure 4.14 a and b show a glancing exit [224] reciprocal space map for the InAs layer grown at 530 °C and 465°C. The layer grown at 465 °C is <1% relaxed.

The results of this study indicate that higher quality interfaces can be obtained by lowering the temperature of the InAs layer growth to prevent intermixing by either As-for-Sb exchange, or Ga carryover, or both. This may be due to the fact that the lower temperature provides less kinetic energy to surmount the activation energy barriers associated with swap and carryover. Lowering the temperature can also allow the InAs layer to coalesce earlier in the growth, leading to InAs/InAs growth sooner and a 2D growth mode sooner. It has been reported previously that growing InAs on GaSb in MBE above 460°C resulted in island formation due to the strain of an intermediate Ga-rich layer(s).[78] It is possible that at higher temperatures, the Ga-rich intermediate material causes enough strain that the InAs layer forms islands, which then leaves the GaSb layer underneath exposed in some areas for a longer period of time, allowing further exchange to occur between the Sb in the GaSb layer, and the incoming As flux. By preventing island formation, there would be no exposed GaSb regions after growth initiated, and the exchange process would become self-limiting.

To verify that the lower dislocation density at lower temperature was a result of decreased intermixing, and not simply due to increased barriers to dislocation formation, we grew a sample where a 40 nm thick InAs layer was grown at 465 °C, followed by an additional 50 nm grown while ramping the temperature from 465 °C to 530 °C, and then 50 nm of growth at

530 °C. Figure 4.14c shows the [224] reciprocal space map, showing a heavily strained InAs layer with little relaxation. This indicates that intermixing has been suppressed. Had the lower temperature only suppressed dislocation formation, it is likely that ramping back up to 530 °C would have led to dislocation formation and relaxation. Instead, we observed a similar structure to the 465 °C structure, implying that the lower temperature likely suppresses the exchange and/or carryover. Then, when the temperature is ramped back up, this cannot enhance the exchange and carryover processes, since the interface is already buried, and both of these processes require a free surface to mediate the intermixing.

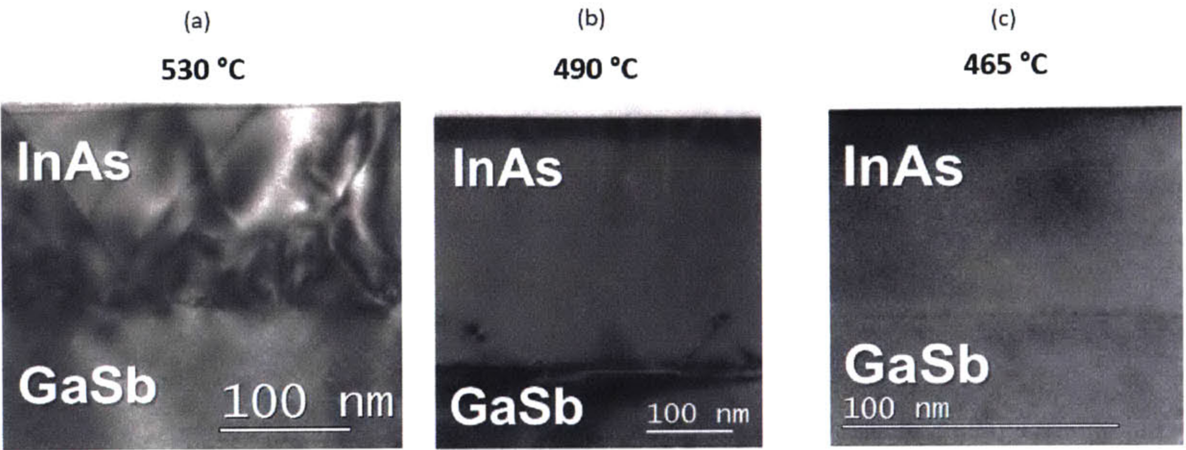


Figure 4.13: a) [220] diffraction condition cross section TEM image of InAs/GaSb structures for an InAs growth temperature of a) 530 °C, b) 490 °C, and c) 465 °C

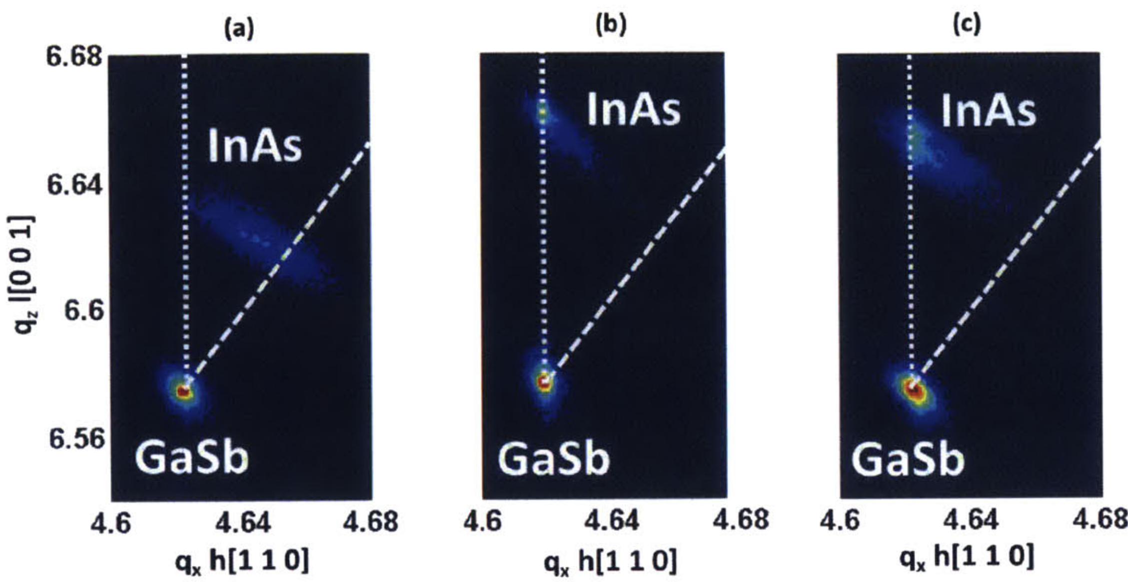


Figure 4.14: [224] glancing exit reciprocal space maps of a) InAs/GaSb junction grown at 530 ℃ (82% relaxed) b) InAs grown at 465 ℃ on GaSb (<1% relaxed) c) InAs grown at 465 ℃ and then ramped up to 530 ℃ while growing. Short dashed line shows the line of strain, and long dashed line shows the line of relaxation

Chapter 5: Electrical Characterization of InAs/GaSb Heterojunctions

5.1 Introduction

The previous chapter established a model for defect formation in InAs/GaSb heterojunctions based on intermixing leading to strain buildup. This model was used to develop techniques to epitaxially grow such heterojunctions at lower defect densities. This provides the opportunity to – for the first time – use samples of varying defect densities to study the linkage between materials defects and steepness. However, before this can be done, an electronic device platform must be established and optimized to avoid parasitics. It must be determined how to properly interpret the data, and discern between differences in materials structure and differences in other effects, such as doping. This chapter focuses on first developing an electronic platform to extract conductance-slope reliably, and then uses this platform to understand how conductance-slope depends on materials defects. In later chapters, devices with differing band alignments will be presented, where the region of conductance slope steepness moves towards zero applied bias. In this case, applications in analog devices arise in addition to digital, in which case the curvature coefficient will also be used as a figure of merit. However, this chapter focuses solely on conductance slope as a figure of merit, as these type-III band alignment of InAs/GaSb would not be appropriate for analog applications. Therefore, this chapter utilizes the InAs/GaSb system as model system to study how steepness depends on materials defects.

5.2 Device Fabrication and I-V Characteristics

The fabrication of the devices was outline in section 3.3. Figure 5.1a shows an SEM image of the side of an InAs/GaSb device after completion. Figure 5.1b shows a top-down optical micrograph of a 13.7 μm diameter InAs/GaSb device, grown at 530 °C with unoptimized conditions (defective interface), and with no doping added beyond unintentional doping. Figure

5.1c shows the current-voltage (I-V) characteristic. The three regions of the characteristic are highlighted: in reverse bias, and up to 0.15 V forward bias, the current is tunneling through the bands. Above this voltage, the bands are uncrossed, and there is a distinct negative differential resistance. Beyond this, the device is in the excess current regime, where current is crossing the interface through some other mechanism (the origin of this mechanism will be elucidated in the subsequent chapter).

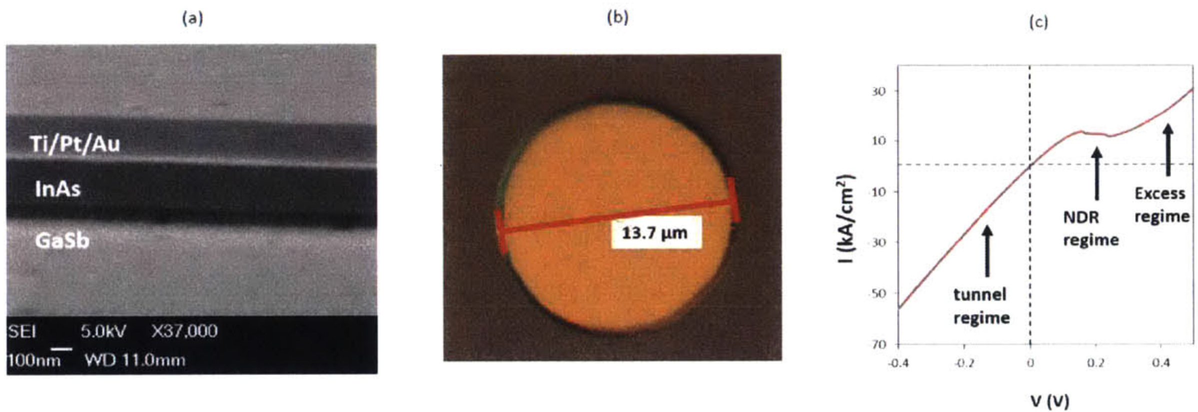


Figure 5.1 a) SEM micrograph of InAs/GaSb device at a side-view, b) top-down optical micrograph of device, c) I-V curve of device

While it is tempting to begin extracting conductance slope data from the I-V curve in Figure 5.1c, there are a variety of questions that need to be answered before this can be done reliably, namely:

1. Why are there two distinct drops in the NDR region instead of only one?
2. What is the true area of the junction?
3. What is the potential that is actually being dropped across the junction (ie. what is the series resistance?)

5.2.1 NDR Region Features

To begin with question 1, Figure 5.2a shows a zoomed in portion of the NDR region for a device with a higher peak-to-valley ratio. Two distinct drops can be seen. The blue curve shows the same device with a series resistor added. Now, there is only one drop, indicating that the double-drop feature is a circuit artefact. This double drop feature in the past has been suggested as potentially being due to discrete quantum states[83]. However, it is more likely to be due to a circuit instability, consistent with its dependence on series resistance. Figure 5.2 b shows an equivalent circuit of a typical tunnel diode. The DC current through the diode is ultimately determined by the intersection of the pure tunnel-diode current and the load-line resulting from the series resistance. Figure 5.2c demonstrates the intersection for two different series resistances. As can be seen, if the load-line is less steep than the conductance of the negative resistance regime, then the load-line intersects the curve at multiple points, leading to instability. Under the correct conditions, the circuit will rapidly oscillate between these metastable points. Hence, the first drop in the NDR regime is when the circuit switches to an oscillation, and the measured DC current is the average of the oscillation. The second drop results from the circuit switching back to a single operating point. At even higher series resistance, the stability conditions likely change, making it more difficult for an oscillation to grow, and the result is that the circuit switches only once: from a high current condition to a low-current condition.

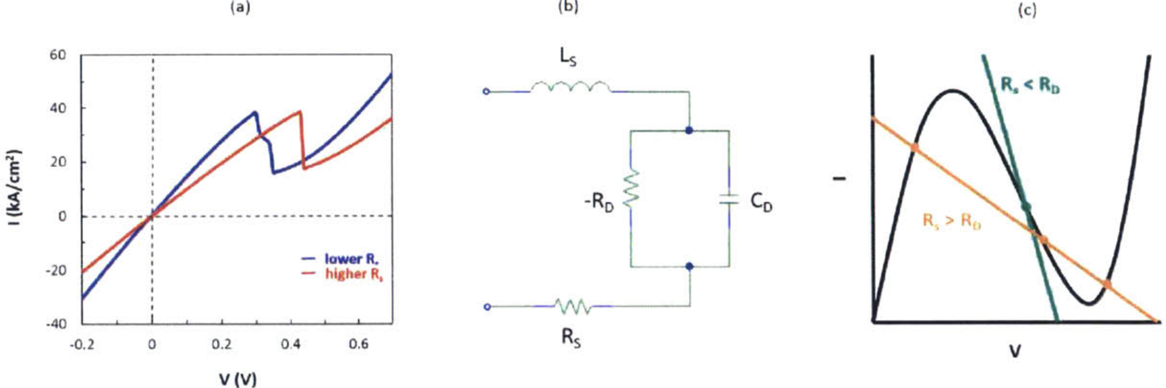


Figure 5.2 (a) Comparison of NDR region of an InAs/GaSb junction for two different values of series resistance, (b) equivalent circuit of this tunnel diode, (c) demonstration of load-line intersection with I-V curve for different values of series resistance, showing stability for the case of $R_s < R_D$, and instability for the case of $R_s > R_D$

Therefore, the data inside the unstable region does not reflect the true nature of the tunneling and cannot be used in extraction of the conductance slope. However, the endpoints of the region can be used to obtain an average conductance slope. This has been done previously in literature[67]. Analysis must be done carefully, however. Figure 5.3a shows an example of an I-V curve of a hypothetical device with instability in blue, and what the true I-V curve might look like without instability in red. The instability region is highlighted. Figure 5.3b shows the conductance slope for each case. This is a highly exaggerated scenario for the sake of illustration, but it can be seen that with the instability, taking the average conductance slope across the instability region leads to an underprediction of slope.

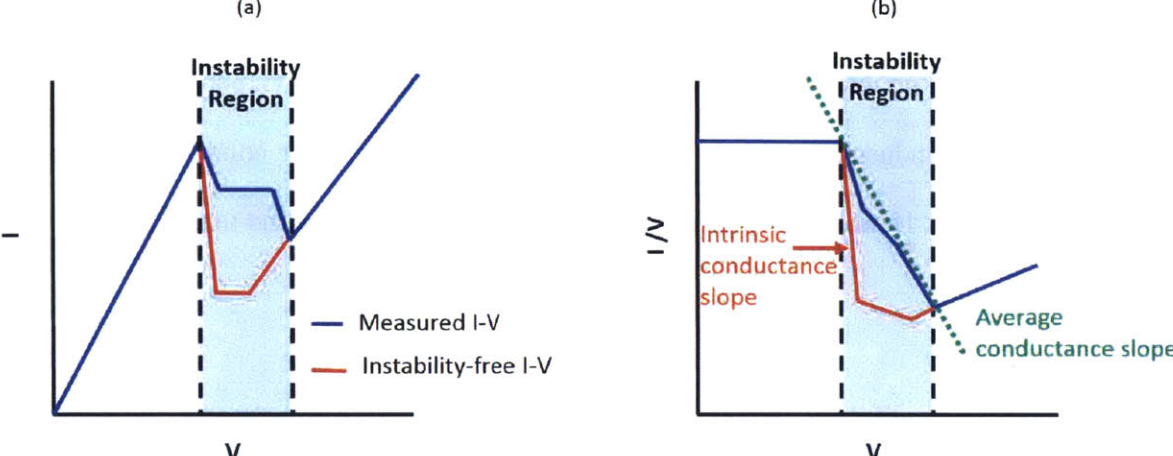


Figure 5.3 (a) I-V curve showing instability region that is larger than the NDR region. The red curve shows the true I-V curve, and the blue curve shows the measured I-V curve due to the intersection of the load-line and oscillations. (b) A corresponding absolute conductance-voltage curve showing a significant underprediction of the true steepness of the conductance slope

The reason this scenario is exaggerated is because the region of conductance slope steepness is entirely confined to the instability region. As will be seen soon when actual measured conductance-voltage curves are displayed, the region of conductance slope steepness is rarely confined to the instability region. In fact, it typically extends beyond the negative differential resistance (NDR) region altogether. While the NDR region is often attributed to where the band uncross, this is not necessarily true. The bands can uncross at lower voltages, and often, there may even be no NDR. Negative differential resistance occurs when the following condition is satisfied:

$$\frac{dI}{dV} = G + V\frac{dG}{dV} < 0 \tag{1}$$

The statement $dI/dV < 0$ is obviously the condition for negative resistance. This derivative can be broken into two components, the absolute conductance, G, and the rate of change of absolute conductance with respect to voltage (comparable to the conductance slope) multiplied by voltage. Hence, for negative differential resistance, this second term must be greater:

$$V \left| \frac{dG}{dV} \right| > G \tag{2}$$

Hence, negative resistance only occurs in the region where the conductance slope is steep enough and/or the applied bias is high enough, that the product of the two is greater than the conductance slope itself at that bias point. This condition is rather arbitrary and does not correspond to anything fundamentally physical about the band-edges. In other words, it is merely a circuit condition. Therefore, we do not expect region where we extract a conductance slope to be confined to the NDR region, let alone the instability portion of the NDR region. Figure 5.4 presents a more likely scenario, in which the bands are already uncrossing at voltages lower than the NDR region. When a conductance-voltage curve is generated, it can be seen that the average slope across the instability region is equal to the slope before or after. Hence, the measurement is accurate.

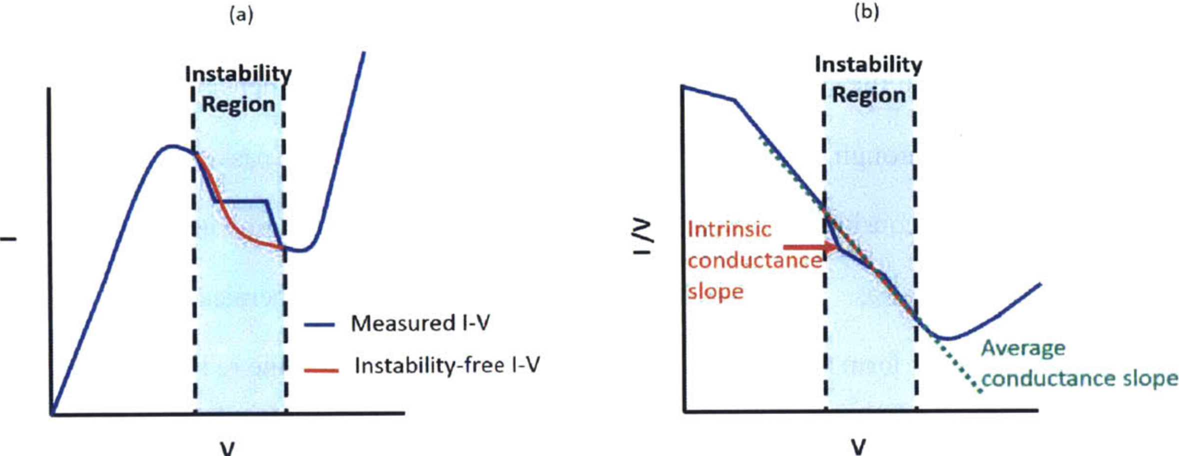


Figure 5.4 (a) I-V curve showing instability region that is smaller than the NDR region. The red curve shows the true I-V curve, and the blue curve shows the measured I-V curve due to the intersection of the load-line and oscillations. (b) A corresponding absolute conductance-voltage curve, this time showing an accurate conductance slope that corresponds to the intrinsic one

5.2.2 Junction Area

With regards to device area for question 2, Figure 5.1a reveals that the GaSb etch has undercut the GaSb. Therefore, we expect the true area of the junction to be slightly smaller than the mask dimensions. Figure 5.1b shows no visible undercut from above, and so it is unlikely to be very large. One simple method to determine this is to compare the peak current at varying junction sizes, since peak current scales with the area of the junction. This is shown for an InAs/GaSb device in Figure 5.5a. As can be seen, the peak current scales with junction area calculated from the photomask dimensions, and the trend passes through the origin, indicating that the change in undercut is negligible. However, it is interesting to see that this is not the case for a GaSb/InAs device (Figure 5.2b). In this case, while current still scales linearly with area, the trend does not pass through the origin when the mask dimensions are used. Furthermore, the

inset shows a clear smaller inner circle visible, indicating that significant undercut has occurred. When the dimensions of the smaller circle are used instead of the mask dimensions (figure 5.3c), the trend then passes through the origin. This indicates that when GaSb is masked directly by the contact metal, there is considerably more undercut, and this must be taken into account when comparing between devices. The origin of this undercut is likely electrochemical. The gold surface and GaSb layer form an electrochemical half-cell that accelerates the reduction-oxidation reaction between GaSb and the NH_4OH-based etchant underneath the metal, as has been observed before.[84]

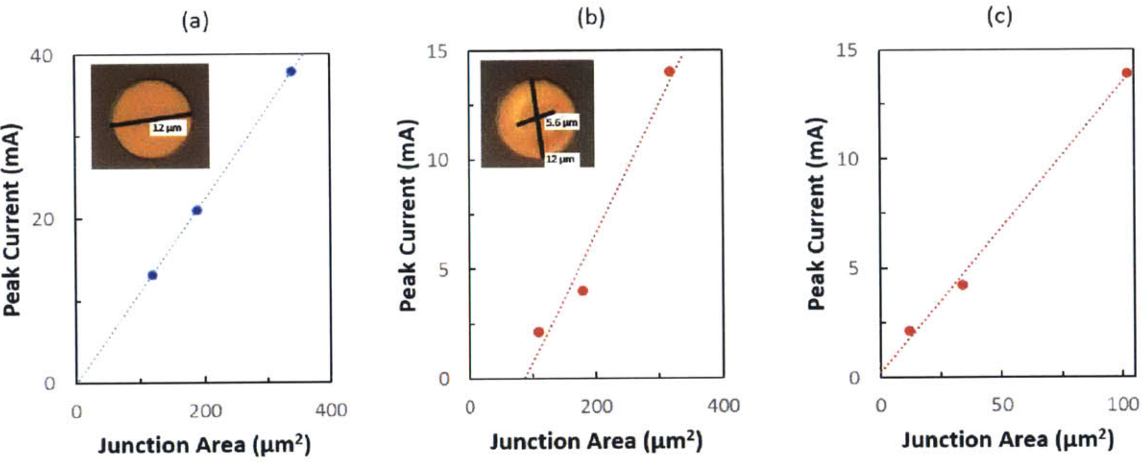


Figure 5.5 (a) Peak current vs junction area based on mask dimensions for InAs/GaSb device, trend goes through origin. Inset shows optical micrograph of a 12 μm device (b) Peak current vs junction area based on mask dimensions for GaSb/InAs device, trend does not go through origin. Inset shows an optical micrograph of a device with significant undercut (c) Peak current vs. junction area for the same device with the optically measured diameters. Trend goes through origin

5.2.3 Series Resistance

Series resistance is another important aspect that must be controlled as the series resistance directly affects how much of the applied potential is delivered to the junction. Figure 5.2b showed an equivalent circuit diagram of the device and test setup. At DC, the inductor is short and the capacitor is open, and so the tunnel diode and series resistance effectively form a voltage divider. The series resistance comes from a few sources: 1. The contact resistance of the top contact, 2. The probe resistance of the probe and wires, and 3. The spreading resistance of current through the substrate. The wires of the test setup do not contribute as a four-point probe setup is used. The series resistance can be quite easily extracted by examining the high bias portions of the I-V curve. At forward bias, the device is effectively a p-n junction and will hence reach a series resistance limited regime, where the series resistance is given by the inverse slope of the I-V curve. At high negative bias, current tunnels, and as the tunnel probability increases, the current eventually becomes limited by the series resistance again. Figure 5.6a shows an I-V curve of an example device with high and low bias. A corresponding differential resistance defined by 1/ (dI/dV) is shown in Figure 5.6b. The negative differential resistance portion can easily be seen on the curve where the differential resistance goes negative. At high and low bias, the differential resistance asymptotically approaches the series resistance, which is shown as the blue line. For this example device, the series resistance is about 3.5 Ω. Based on this resistance, the I-V curve can be corrected to remove the series resistance (R_S) by determining the true voltage applied to the junction (V_j) at a given bias point (V_a) via:

$$V_j = V_a - IR_s$$

Figure 5.6c shows an example of the device I-V curve before and after correcting for the 3.5 Ω series resistance. As the current becomes high, the series-corrected curve deviates heavily

from the raw curve. However, in the region of interest – the negative resistance region and the slightly higher and lower bias regions around it – there is not an enormous difference between the two. Still, with this correction, we now have an accurate method of determining how much potential is being dropped across the junction for accurate measurement of the devices' conductance slope.

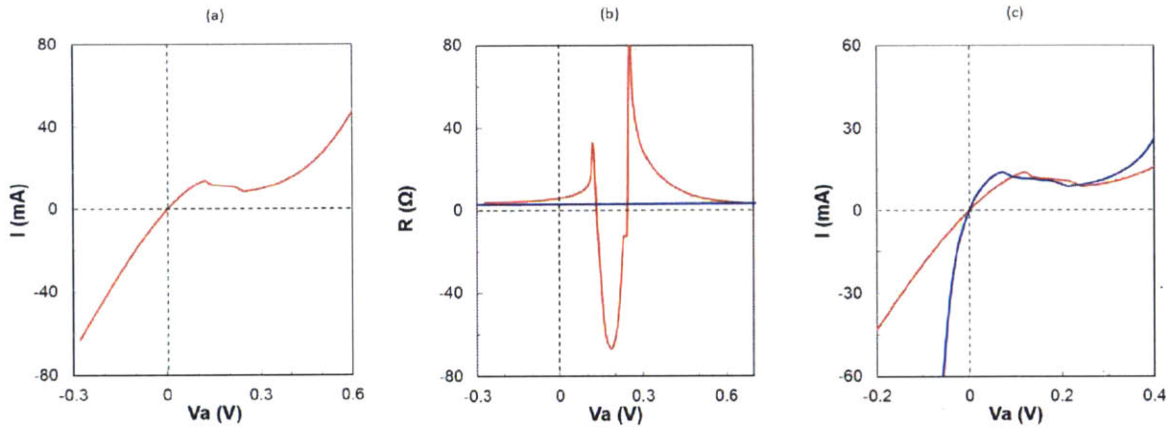


Figure 5.6: (a) I-V curve of a sample device showing the slope approaching a constant value at high and low bias, (b) differential resistance of the device, showing the curve asymptotically approaches the series resistance at high and low bias, and also showing the negative differential resistance in the bottom right quadrant for the NDR region, (c) I-V curve before and after correcting for series resistance, using the value extracted from (b)

5.2.4 Conductance Slope Measurement Procedure

With the various subtleties of the device and its electric characteristics under control, a working procedure is now in place to extract a reliable conductance slope:

1. Determine Area of junction
2. Measure I-V curve

3. Extract series resistance from series limited region in forward bias
4. Subtract series resistance from data to obtain true I-V curve of device
5. Convert I in I-V curve to I/V to obtain absolute conductance-voltage curve
6. Measure slope of absolute conductance as an average across the NDR region

Figure 5.7 shows the final results, with an I-V curve of an InAs/GaSb device in Figure 5.7a and the corresponding absolute conductance-voltage curve in Figure 7.5b. The unstable portion of the NDR curve is highlighted in both figures, and is most easily identified by observing the curvature coefficient, which is also shown in the red curve in Figure 5.7a. The curvature coefficient, the second derivative of the I-V curve normalized by the conductance, was defined in Chapter 1 and is a good figure of merit for curvature – an important characteristic for analog applications. This particular device would not be useful for such applications as the curvature is not high without an applied bias. In Chapter 9, the curvature coefficient will be used as an actual figure of merit for devices with useful analog applications. In this case however, the curvature coefficient is useful simply because it easily shows a pair of peaks and a crossing of the x-axis at every point where an instability occurs. It can be seen that a variety of peaks occur in this NDR region, highlighting the fact that there is much instability. Figure 5b shows the absolute conductance-voltage curve, where, as described in section 5.2.1, an average slope is taken across the region of instability. It can be seen that the average slope is continuous with the slope outside of the instability region, making it analogous to Figure 5.4b as opposed to Figure 5.3b, indicating that the average slope is likely not underpredicting the true slope. The value of the conductance slope for this device is 265 mV/decade – considerably higher than the thermal limit of 60 mV/decade, and comparable to published subthreshold slope results. [40], [49], [52]

Hence, a reliable method of extracting a conductance slope has been established and comparisons can now be made between differing samples.

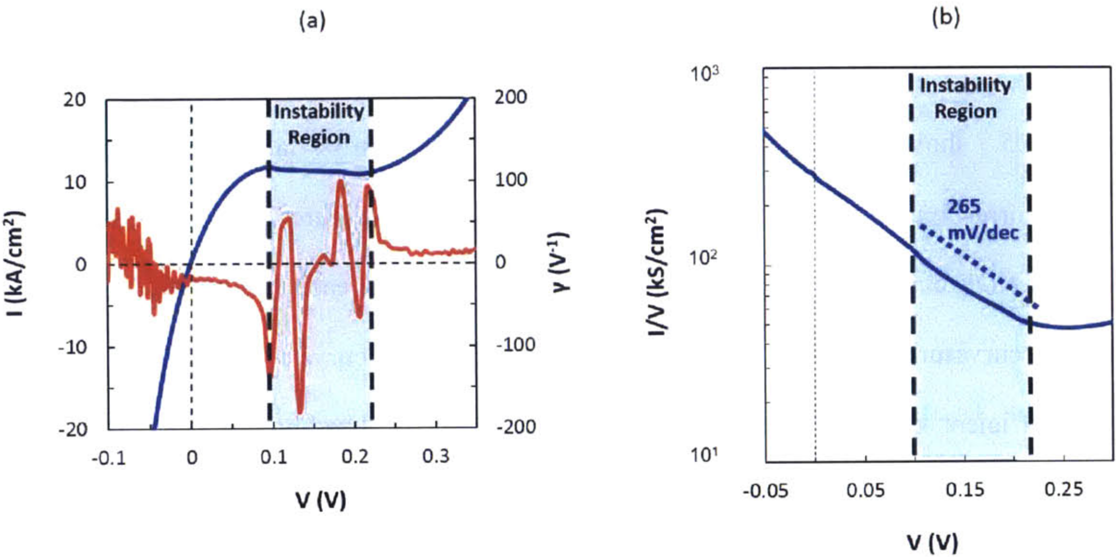


Figure 5.7 (a) I-V curve (blue curve) of an InAs/GaSb device grown under non-optimized conditions. The curvature coefficient is also plotted (red curve) for the sake of using the peaks to identify the boundaries of the instability region, which is highlighted in light blue. (b) Absolute conductance-voltage curve of the device, showing a conductance slope of 265 mV/decade, which is continuous with the slope outside of the instability region

5.3 Effect of Doping

Doping is the first important metric to study, as it is the standard method of control of bands and carrier concentration. Before heterojunctions, it was the only method to create band overlap for interband tunneling. However, dopants lead to states near the band edge, which may cause a loss of steepness as described in section 2.3.1. One advantage of heterojunctions is that electron affinity, instead of doping, can be used to create the band overlap. Hence, in theory, the

band overlap could be obtained without all of the dopant states. Of course, we realistically imagine that there will still be states at the interface due to the defects that we expect to form there. The question then becomes, if energy states do matter, which states dominate?

It would be desirable to obtain as low doping as possible while still maintaining some p-type doping in the GaSb and some n-type doping in the InAs. The reason for this is to ensure that there is a decent concentration of holes in the valence band of GaSb and electrons in the conduction band of InAs so that tunneling can occur. The lowest doping attainable would be the unintentional doping (UID) level, which is the doping that results from defects or contamination during the epitaxial growth. Previous reports have indicated that MOCVD-grown InAs tends to have n-type UID doping [85], while GaSb tends to have p-type UID doping [86]. This is a favorable condition and may be ideal.

To test the effect of doping, a variety of structures with differing doping levels were grown, as outlined in Table 5.1. Sample A was unintentionally doped, and so only the regular group III and group V precursors were flown through the chamber during growth. The remaining samples B-D were doped with Si for both the InAs and GaSb layers, which dopes InAs n-type, and GaSb p-type. This was accomplished by flowing Si_2H_6 in the chamber during layer growth. The effect of Si_2H_6 flowrate on Si dopant incorporation was measured using secondary ion mass spectrometry (SIMS) and was used a calibration for doping. The relationship is shown in Figure 5.8 a and b for both InAs and GaSb.

Table 5.1 Acceptor (N_a) and Donor (N_d) Concentrations of Si-doped samples

Sample #	GaSb N_a (cm^{-3})	InAs N_d (cm^{-3})
A	UID (expected 4×10^{16}) [87]	UID (expected 1×10^{16}) [85]
B	1×10^{17}	1×10^{17}
C	2×10^{19}	1×10^{18}
D	5×10^{18}	1×10^{19}

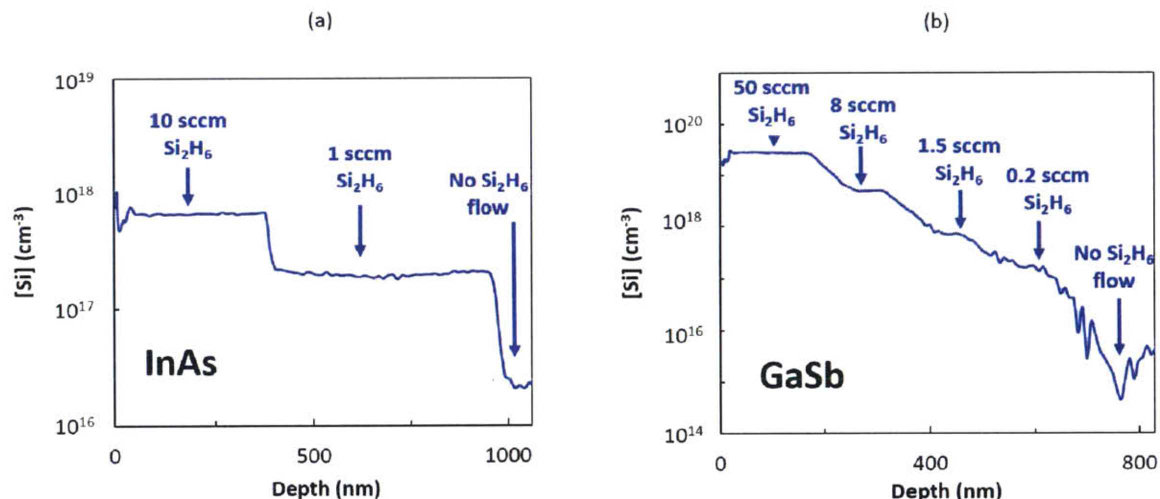


Figure 5.8 (a) [Si] as a function of depth in InAs measured by SIMS for a doping calibration. (b) same calibration for GaSb

Figure 5.9a shows the I-V curve for all samples. There is a marked difference between them in that the tunnel conductance, which can be seen as the slope of the I-V curve through the origin, increases for the more heavily doped samples. In a related sense, the peak current (the highest current value before negative differential resistance) also increases with doping. This increase is likely due to a decrease in the tunnel barrier that results from heavier doping and was observed in another study for MBE-grown material. [37]

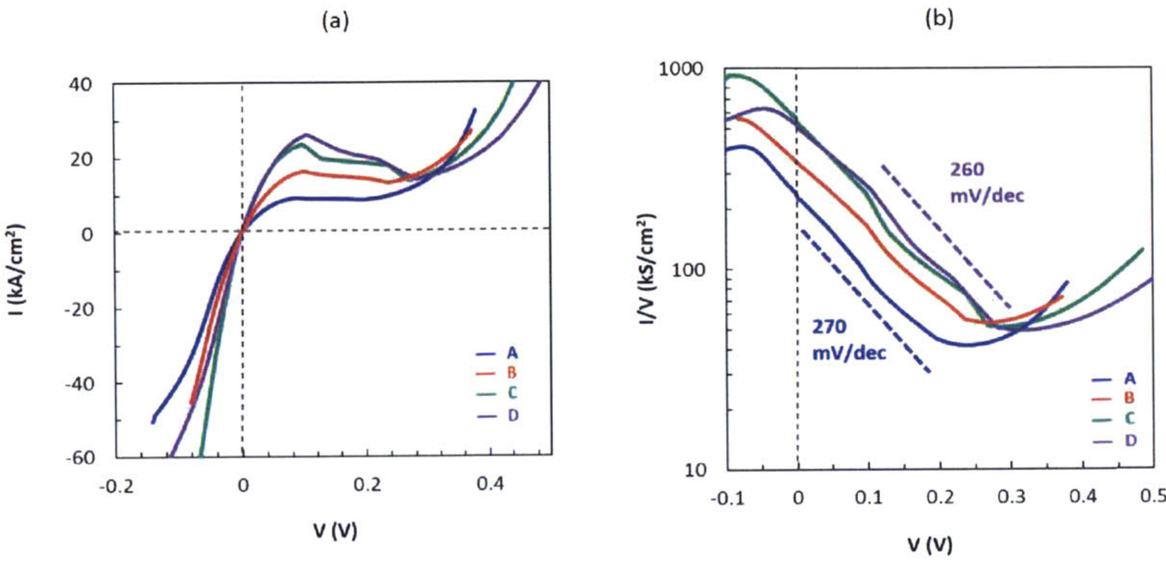


Figure 5.9 (a) I-V curve for the samples from Table 5.1 with varying doping levels. (b) Corresponding absolute conductance-voltage curve for all samples, showing generally the same conductance slope for all samples

Interestingly, while there is a considerable change visible in the I-V curve as doping is changed, when observing the absolute conductance-voltage characteristic, this translates only to a shift in the curves to higher conductance as doping level increases. The slope, however, remains essentially the same for all device. This is the first important result in terms of what affects the steepness of these heterojunctions: In these structures, doping does not limit the steepness. This is a rather unexpected result, as it is known that dopants lead to energy states near the band edge. The implication of this is that these states must not be limiting the steepness. The question becomes, what is? At this point, the two most likely possibilities are either states resulting from materials defects, or phonons. In the subsequent sections, the effect of materials defects on the conductance slope will be assessed.

5.4 Effect of Growth Order

As was shown in Chapter 4, growing InAs on GaSb results in a very defective interface, while growing GaSb on InAs results in a seemingly perfect interface. Hence, this provides an ideal comparison for how the conductance slope depends on materials defects. The series-resistance-corrected I-V characteristics of the InAs/GaSb structure compared to the 80 nm thick GaSb/InAs structure is shown in Figure 5.10a, and the conductance-voltage characteristics are shown in Figure 5.10b. The GaSb/InAs device shows a much steeper conductance slope in the region of the bands uncrossing. This indicates that the heavily defective InAs/GaSb interface has caused a blurring of the turn-off, consistent with what was described in section 2.3. A less steep conductance slope can be seen, consistent with less uniformity, as well as a higher current beyond the steep portion, which appears consistent with more trap-assisted leakage, although the diode current becomes significant at higher forward biases, which can overshadow the leakage current. Further, while deep level traps would lead to a leakage current, leakage through more shallow-level traps, if concentrated enough, would lower the slope throughout the entire region of band-uncrossing, not just at lower currents. Therefore, the main characteristic to extract from this data is the less-steep slope in the region in which the bands are uncrossing, which is attributable to a variety of energy states due to defects, and the local band-bending they cause. Hence, this presents the first evidence that the steepness of tunneling in heterojunctions is affected by materials defects.

This difference in slope between these samples was seen consistently. For a t-test on a sample size of 15 devices of each type grown across three separate wafers, a $p < 0.0001$ was obtained, indicating statistical significance.

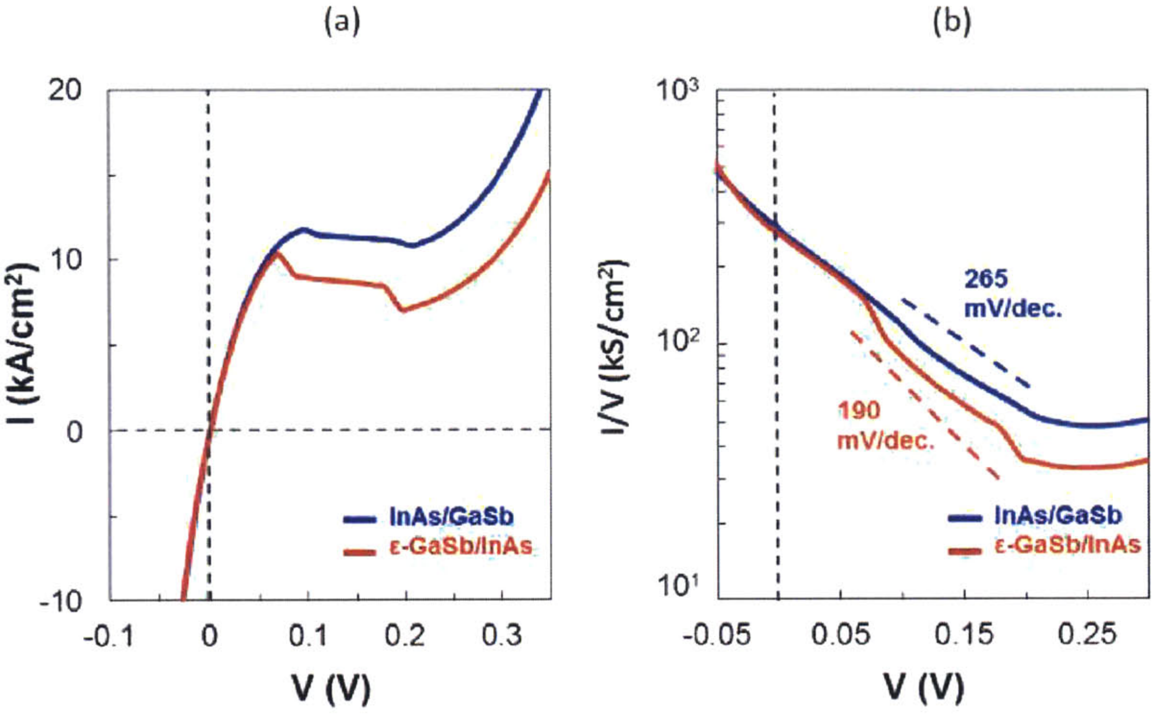


Figure 5.10: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve derived from a), for InAs/GaSb and strained GaSb/InAs devices

This result does not indicate, however, which defects specifically cause this decrease in conductance slope steepness, as the InAs/GaSb interface has 60° and 90° misfit dislocations and many threading dislocations. Therefore, it is useful to compare the 80 nm GaSb/InAs device to the case where the GaSb layer is grown to 380 nm thick and undergoes significant relaxation. In the latter case, the interface contains 60° misfit dislocations with a low threading dislocation density, as was seen in Figure 4.9b, but without many threads and edge dislocations as in the InAs/GaSb device. Consistent with the TEM image, we can estimate from the [224] RSM in Figure 4.10c that with 63% relaxation, the misfit dislocations are spaced by about 60 nm on average. Whereas, for the minimal relaxation of the 80 nm GaSb film, the RSM indicates a spacing of $> 1\,\mu m$. Hence, these two samples have very different dislocation densities. Figure

5.11 shows the result: the conductance slope is significantly less steep and about the same as the InAs/GaSb device, indicating that the loss of steepness is correlated to the presence of misfit dislocations.

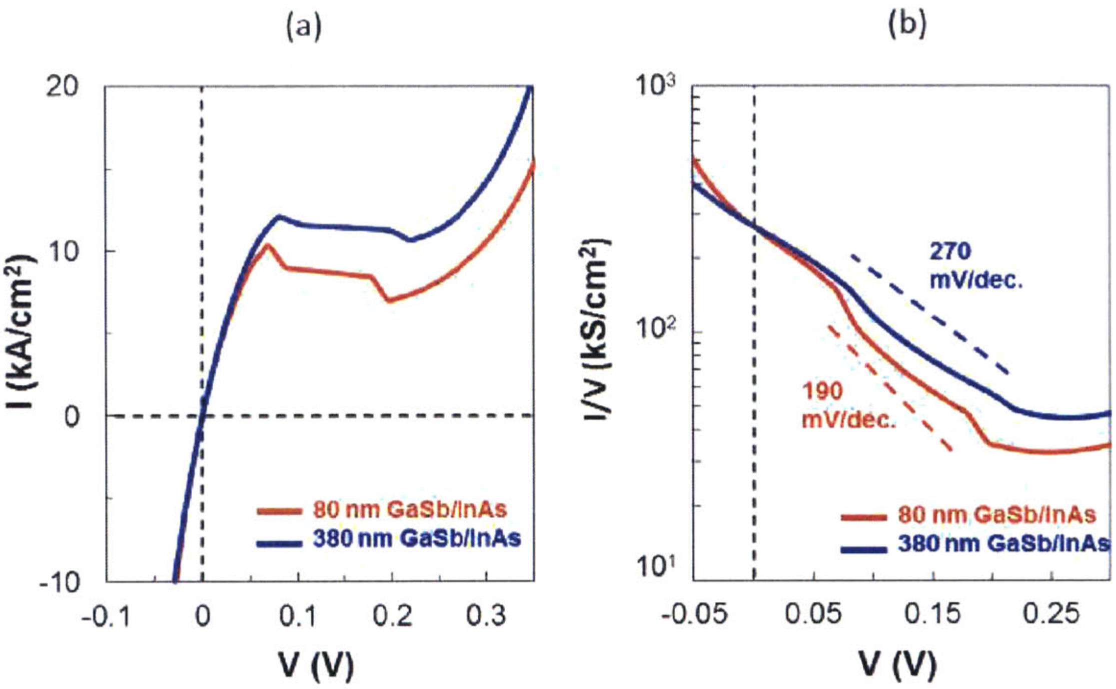


Figure 5.11: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve for 80 nm thick and 380 nm thick GaSb on InAs devices

Theoretically, it can be proposed that the misfit dislocations are leading to energy states and local band bending in their region due either to 1) states associated specifically with the core of the dislocations, 2) states resulting from the local deformation of the lattice in the region of the dislocations or 3) states associated with point-defects, which tend to accumulate in the region of dislocations due to the difference in strain and charge. However, in reality, it is more likely to be a combination of 2) and 3), and mainly 3), as it has been shown previously that dislocations tend to getter many point defects which lead to states and lattice deformation which dominate the

electrical characteristics[61], and that the dangling bonds in the core of most dislocations can reconstruct to passivate and eliminate the defect states that would be expected of a dislocation core[88], as is shown in Figure 5.12. This may explain why the InAs/GaSb device, which in addition to having 60° dislocations also contained pure-edge 90° dislocations, gave about the same conductance slope steepness: while a true 90° dislocation can have a different density of energy states than a 60° dislocation, both types attract point defects, and the effect of point defects overwhelms the much lower concentration of dangling bonds and reconstructed states.

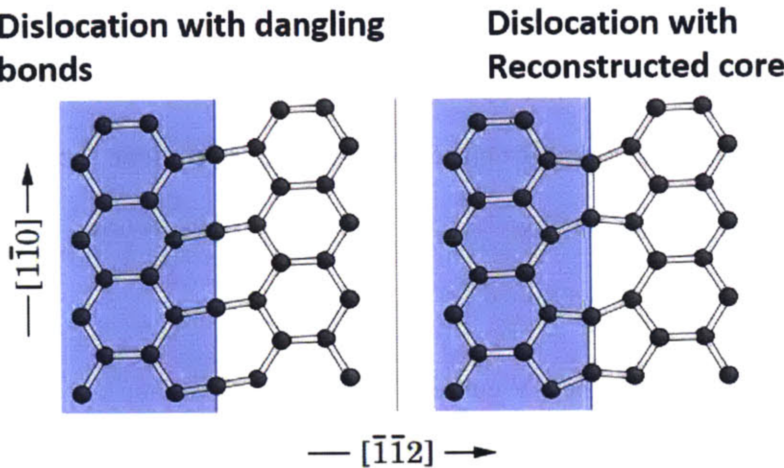


Figure 5.12 Non-reconstructured dislocation core with dangling bonds vs. reconstructed core with no dangling bonds [89]

5.5 Post-Growth Annealing, Point Defects, and Interface Uniformity

Given that the point defects gettered by dislocations and the resulting lattice-deformation are expected to lower the conductance slope steepness, it stands to reason that the steepness may be improvable via thermal annealing. This is because annealing can cause migration of point defects, which can allow for their annihilation. This migration can also act to average out the

concentration to create a more uniform distribution. This would lead to fewer regions of largely different band alignment and hence, steeper switching in devices utilizing this interface. Further, as also described in section 5.3.2, this blurring also can result from nonuniform mixing, and so it is likely that annealing would help to increase steepness by averaging out any nonuniformity resulting from both uneven intermixing as well as uneven/elevated point defect concentrations.

We annealed both the high dislocation density InAs/GaSb and low dislocation density 80 nm thick GaSb/InAs at 600°C for 7.5 minutes. This was done in a rapid thermal annealer, with PECVD-deposited SiO_2 protective layers added to the top and bottom surface to prevent degassing, which was subsequently removed in a buffered oxide etch after annealing. Devices were then fabricated on the annealed structures. We kept the anneal time short to avoid significant intermixing between the layers above what had already occurred during growth, and to avoid any relaxation in the still strained GaSb/InAs device.

Figure 5.13a and b show the I-V curve and conductance-voltage curve of the InAs/GaSb device with and without the annealing step. It can be seen that there is a noticeable increase in the steepness of the conductance slope after annealing, which is indicative that in both cases, point defects are being eliminated and/or more evenly distributed. It is clear, however, that the annealed device still has many electrically active defects, as the steepness after annealing is still not as good as the device without dislocations. We do not expect that there was much intermixing between the layers during the annealing, as this would lead to a noticeable decrease in current in reverse bias due to a lower tunnel probability, and would likely lead to a lower conductance slope as well. But it can be seen that the current and conductance in reverse bias, where tunneling is occurring, has not changed after annealing.

Even without a substantial concentration of misfit dislocations, there is a significant improvement in conductance slope, as is shown in Figure 5.13c and 5.13d, for the annealing of the GaSb/InAs structure which is almost completely strained. This is likely an indication that all samples contain point defects and nonuniform composition fluctuations. The presence of the misfit dislocations at the interface tends to getter these point defects and increase their concentration. Hence, the conductance slope is made steeper by annealing dislocated samples, enhanced further by preventing dislocations, and enhanced even further by annealing undislocated samples. Each enhancement is due to fewer point defect states and more interface uniformity. It can be seen that the enhancement for the undislocated device is about the same as for the dislocated device. This is likely an indication that many point-defects are still present in the dislocated samples: they can only be annealed out to a degree, as the strain fields of the dislocations create a driving force to maintain their concentration at high amounts, more so than an undislocated sample. Furthermore, since there are two separate effects – point defect concentrations and nonuniform intermixing – it is difficult to compare the percent improvement with annealing for between dislocated and undislocated sample, as the relative magnitude of either effect cannot be distinguished.

Again in this case, the electrical results do not indicate increased intermixing. Furthermore, we confirmed via HRXRD that there was no additional relaxation of the GaSb layer after annealing, consistent with the much steeper slope.

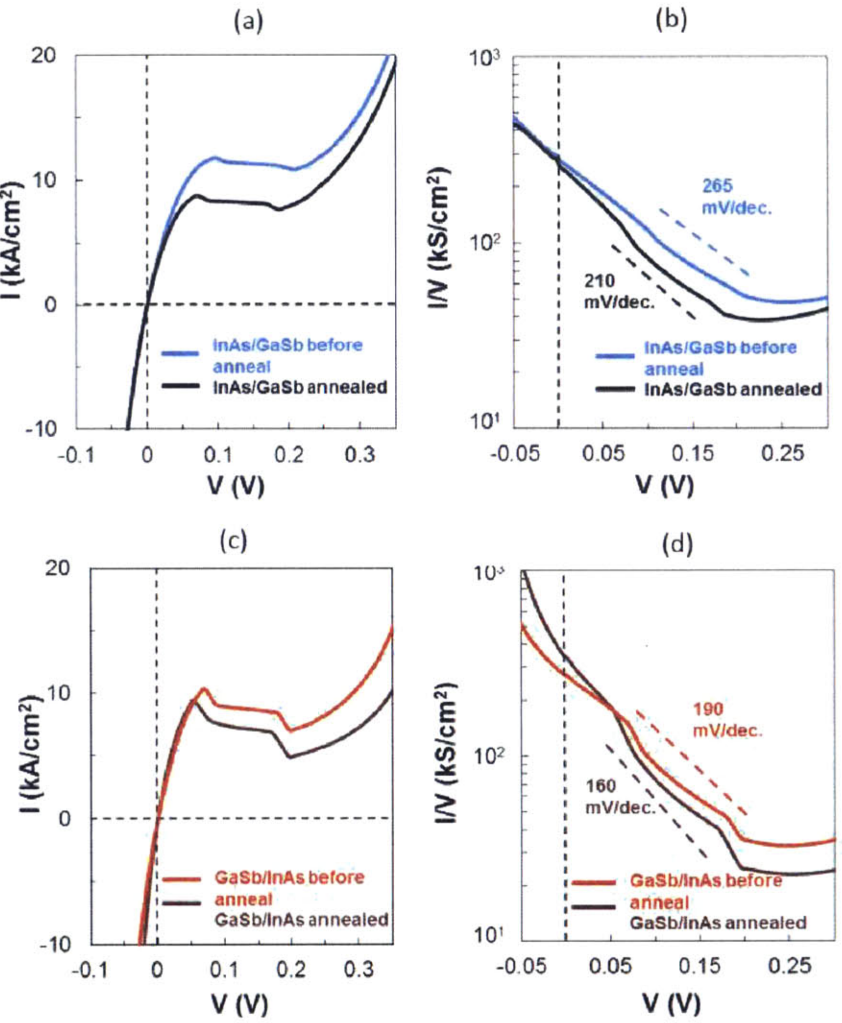


Figure 5.13: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve InAs/GaSb devices before and after annealing at 600℃. c) and d): Equivalent comparison for 80 nm thick GaSb/InAs devices

5.5 Strained-Underlayer Lattice Matching

It was shown in section 4.6 of the previous chapter that quality InAs layers can still be obtained provided the GaSb underlayer is strained so that InAs is growing on a "pseudo-lattice matched" GaSb layer. This provides an additional opportunity to test the effect of dislocations, as

this structure is a truly lattice-matched structure, where growing InAs layer has effectively no driving force for relaxation, whereas the GaSb/InAs structure had no driving force from intermixing, but still had a potential driving force from lattice mismatch. Devices were fabricated on the two samples shown in Figure 4.11 from chapter 4: InAs on 80 nm-thick (strained) GaSb and InAs on 380 nm (63% relaxed) GaSb. However, since there are now two interfaces (as can be seen in Figure 4.11) - the main InAs on GaSb tunnel interface, as well as the lower interface for the GaSb layer on the InAs substrate - great care was taken to prevent the lower interface from affecting the I-V curve. This was done by growing n+-InAs on the InAs substrate, and growing the first half of the GaSb film as Si p+ doped. This makes the lower tunnel interface much more conductive. Furthermore, the GaSb was not etched all the way through the layer, to allow for room for current spreading in the GaSb layer, making the effective area of the lower junction larger than that of the main tunnel junction. Figure 5.14a shows the I-V curve for both cases, and Figure 5.14b shows the conductance-voltage curve. The results are consistent with the rest of the study, in that the dislocated InAs on relaxed-GaSb shows the same less steep conductance slope (refer back to Figure 5.10b for a comparison). The InAs/strained-GaSb device, on the other hand, shows a sharper conductance slope, due to its lack of dislocations. It is worth noting that for these plots, while the series resistance was removed in the standard manner of using the slope at high voltage, the variable resistance of the lower interface tunnel interface cannot be removed this way, since it is small at high voltage and becomes larger at small forward bias and increasing reverse bias. This is likely the origin of the loss of steepness through the origin for both devices in Figure 5.14b. Since this variable resistance was the same for both devices, the comparison is still fair.

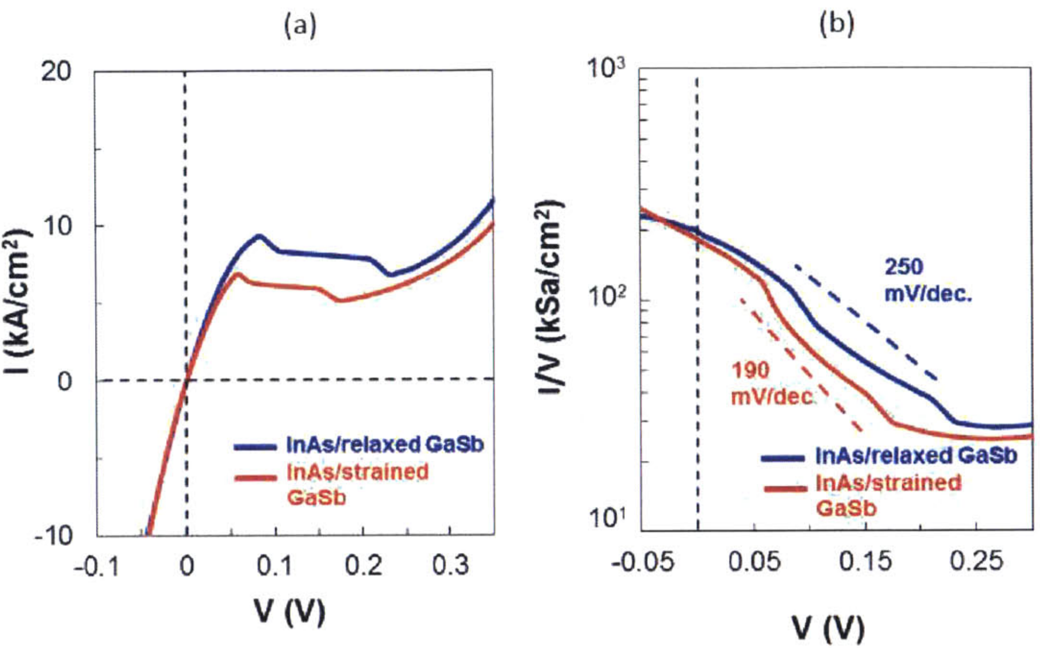


Figure 5.14: a) Series resistance-corrected I-V curve and b) Absolute conductance-voltage curve for InAs grown on 80 nm thick (strained) and 380 nm thick (63% relaxed) GaSb

5.6. Electrical Effect of Varying Growth Temperature

Finally, the use of growth temperature to control defect densities was described in section 4.7, and it was found that good quality InAs layers on GaSb could be obtained by lowering the InAs growth temperature, which suppressed exchange. Figure 5.15a and 5.15b compare the electrical results of three devices: 1) a device with the InAs layer grown entirely at 465 °C, 2) a device with 40 nm InAs grown at 465 °C and then ramped up to 530 °C while growing, and 3) a device with the InAs layer grown at 465 °C and then the entire epitaxial structure annealed at 600 °C for 7.5 minutes, using the anneal procedure described previously. Device (1) shows a change in conductance, but with extremely small steepness which levels off at a fairly flat slope

afterwards, indicating that significant tunneling is still occurring either through defect states or regions where the band overlap is still large. For device (2), growing with a ramp in growth temperature to 530 °C provides some improvement, but minor. However, for device (3), when the structure is annealed at 600 °C, the conductance slope steepness improves greatly, giving a conductance slope about the same as the annealed dislocation-free GaSb/InAs device, consistent with the fact that this device indeed is also a dislocation-free annealed device. The phenomenon here appears to be the same as that for annealing the GaSb/InAs structure, in that the annihilation of point defects and/or averaging of fluctuations in band alignment have increased the steepness of the conductance slope. However, it appears in this case that the steepness was much weaker before annealing. This makes sense since we expect a growth temperature below the optimized growth temperature of 530 °C to introduce more point defects into the layer, due to a variety of possible mechanisms such as lack of surface diffusion during growth resulting in defect concentrations above the equilibrium value, greater carbon-incorporation due to incomplete precursor decomposition, increased impurity incorporation due to lower growth rate, and a shift in the equilibrium stoichiometry of the compound via a change in partial pressure of the In and As, via a change in the precursor decomposition rate. Also, since after the GaSb layer is grown, the temperature must be ramped down to 465 °C before the InAs can be grown, there is the opportunity to incorporate impurities at the surface. Despite all of these issues, however, dropping the InAs growth temperature to such a low value is still beneficial, since the issues it introduces can be resolved by annealing, and the net result is a major improvement over growing at 530 °C. Again, we expect minimal intermixing due to annealing given the lack of change in conductance in the tunnel region (reverse bias), and confirmed with HRXRD that annealing led to a negligible amount of relaxation.

Figure 5.15c compares this annealed device to the device grown at 530 °C annealed and unannealed. This agrees with and reiterates the concepts established in section 5.5: there is improvement in steepness of conductance slope in annealing dislocated samples, further improvement in a sample with no dislocations, and even further improvement in an annealed sample with no dislocations. Furthermore, Figure 5.15c shows a general shift of the curve to the left as the defect density is lowered. This was predicted in Chapter 2 by the calculation that was shown in Figure 2.4. This will be observed again in later chapters.

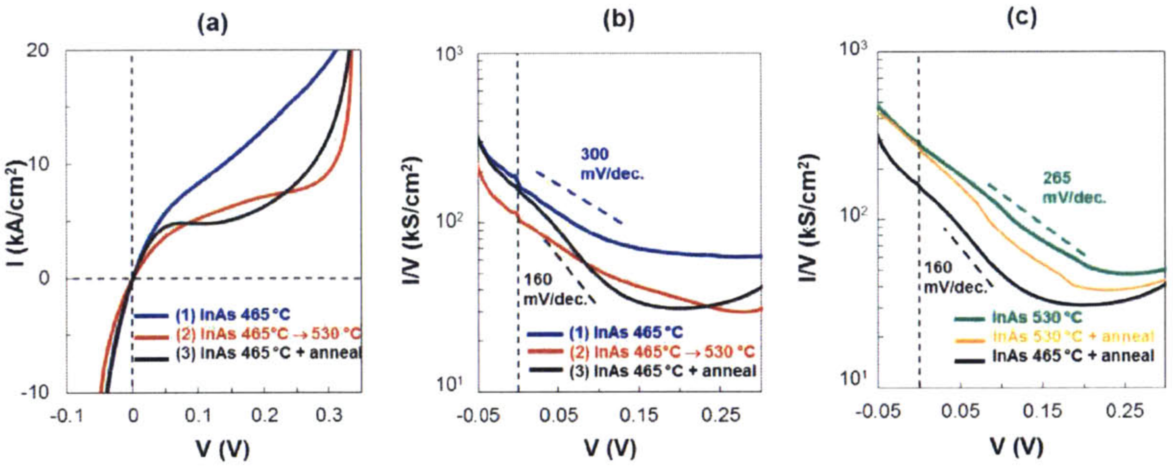


Figure 5.15: a) I-V curve and b) Absolute conductance-voltage curve for (1) InAs grown at 465 °C, (2) 465 °C with a ramp to 530 °C, and (3) 465 °C and then the entire structure annealed at 600 °C. c) Comparing the annealed sample to growth at 530 °C with and without annealing

5.7 Conclusions

We have identified point and line defects, as well as nonuniform intermixing as having a major effect on the steepness of the conductance slope of InAs/GaSb devices. For line defects, misfit dislocations dominate and lead to a major decrease in conductance slope steepness. This is

presumed to be due to gettering of point defects leading to energy states and bend bending that allow for trap-assisted leakage and nonuniform band alignment across the interface. This is unavoidable for InAs/GaSb structures grown at 530°C, since the material relaxes instantly due to the buildup of strain from intermediate compositions from intermixing. This can be circumvented by growing in the reverse order, to prevent the intermixing, or growing on strained-GaSb, so that the InAs layer is lattice-matched. It can also be avoided by lowering the growth temperature of the InAs layer to prevent intermixing.

Even without dislocations, point defects and nonuniform composition still affect the performance of the devices, and all devices respond positively to annealing, as it lowers the concentration of these defects.

Chapter 6: Low Temperature Study of Conductance Slope

6.1 Introduction

The previous chapter established a clear correlation between defect levels and conductance slope steepness, indicating that inhomogeneity, namely point defects and uneven composition, lead to a lower steepness. This chapter focuses on further studying the physics of tunneling from the band-edges by assessing how these phenomena depend on temperature, still using our model type-III band-alignment InAs/GaSb materials system with different levels of inhomogeneity at the interface. Ultimately, the work of this chapter seeks out to answer two important questions:

1. Is the steepness of the best devices seen so far still limited by inhomogeneity, or have they entered a new regime where they are limited by phonons?
2. In the absence of a gate and other three-terminal parasitics, will we still observe a strong temperature dependence of conductance slope?

Question 1 is essentially a question formed from analyzing the results of the previous chapter. A linkage between inhomogeneity has now been established, and the goal now is to compare the temperature dependence at different defect levels to see if further improvement is possible, or if we have hit a limit where defects no longer limit steepness.

Question 2, on the other hand, is a new question, and is based on the literature published to date. As was described in the introduction of this thesis, published TFET devices, particularly those using heterojunctions and density-of-states switching, have a tremendous temperature dependence in which the measured subthreshold slope increases with temperature. Figure 1.7 in Chapter 1 showed a summary of various reported results collected by various research groups at various temperatures for density-of-states switching devices. Not only was the existence of a

temperature dependence strange, but the fact that it seems to have a linear dependence is especially strange, as it implies an Arrhenius behavior. This essentially indicates that there is likely a thermal barrier present somewhere in the primary current path in all published devices. As shown in Figure 6.1, this could either be due to a strong series resistance component with a thermal barrier after the tunnel interface (Figure 6.1a), or a parallel leakage pathway with a thermal barrier (Figure 6.1b). For the case of a parallel pathway, it would have to be a leakage pathway that, at all temperatures studied, provides a higher conductance than the interband tunneling, hence dominating the equivalent circuit.

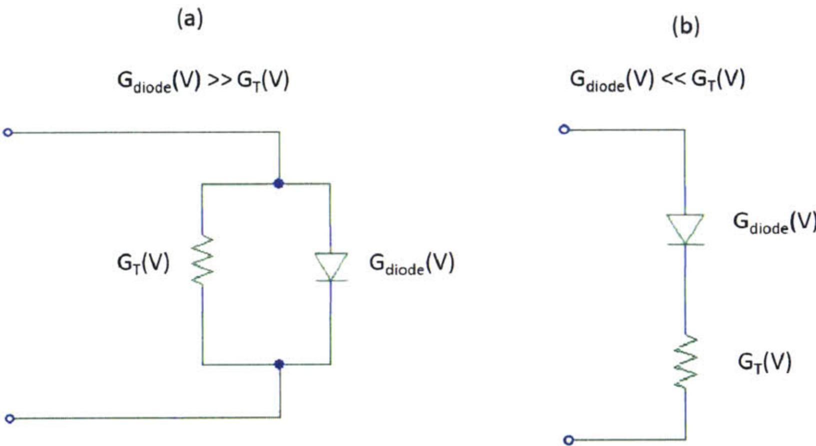


Figure 6.1 Two scenarios in which a thermal component would dominate the device characteristics. (a) a large conductance with a thermal dependence $G_{diode}(V)$ (expressed as a diode) in parallel with the tunnel junction $G_T(V)$ (expressed as a resistor with a voltage-dependent conductance) (b) a small conductance with a thermal dependence $G_{diode}(V)$ in series with the tunnel junction $G_T(V)$

Indeed, many authors who have observed this temperature dependence are aware of its potential implications that it is not likely due to interband tunneling directly[34], [42], [49], [54],

[55], [90]. While some proposals have been made that it could still be intrinsic to interband tunneling but via a trap assisted-mechanism involving interface defects[42], [49], or phonons,[42] others have proposed a true parasitic leakage pathways [34], [54], [55], [90]. Most notably, Mookerja proposed that gate oxide trap states may be assisting the tunneling. [54]

The unique advantage of a two terminal device is that it allows the interband tunneling to be measured without three terminal parasitics. Therefore, an explanation of the commonly seen temperature dependence can be obtained: If a temperature dependence is still seen, it is likely intrinsic to the tunnel interface and is likely due to trap states at the interface being at such a high density that they dominate true interband tunneling. This is somewhat plausible since it has already been seen in the previous chapter that defect densities seem to play a major role. However, as the interface became less defective with specialized growth techniques, we did not observe a decrease in conductance or peak current, only an increase in slope. This is a hint that perhaps the trap-assisted tunneling was not dominating the true interband tunneling, and perhaps then we should not expect to see a temperature dependence in wto-terminals. Indeed, if there is no temperature dependence in two terminals, then the published TFET temperature dependencies would certainly have to be due to three-terminal parasitics.

In addition, this chapter seeks to understand the origin of some of the other metrics of tunneling, namely to confirm tunneling in the on-state, and determine the origin of the off-state current.

6.2 Comparison of I-V Curves

The I-V curves of an InAs/GaSb device and a GaSb/InAs device were measured at temperatures between room temperature down to 4 K. Figures 6.2a and 6.2b display I-V curves

measured at select temperatures for GaSb-on-InAs and InAs-on-GaSb devices, respectively. The inset shows cross sectional TEM images of the samples. As was shown in Chapter 3, InAs films grown on GaSb are heavily defective with many threading dislocations visible in TEM , while GaSb films grown on InAs are relatively less defective, with no visible threading dislocations in cross section. As temperature is decreased for both devices, there is a visible increase in peak current, an increase in conductance at zero-bias, and a decrease in excess-current beyond the negative differential resistance (NDR) region (which is the off-state current in a TFET).

An increase in peak current implies an increase in the number of carriers tunneling, and the increase in conductance at zero-bias is also a result of this. However, it can be seen that the conductance through the origin continues to increase even after the peak current has stopped increasing. This is because the conductance also depends on series resistance, and we expect an increase in substrate mobility at lower temperatures. Since this mobility effect is separate from the tunneling, and since the peak current is independent of this effect, we choose to focus on the peak current effect since it provides a clearer look at the tunnel current. Therefore, the nature of tunneling in the on-state can be confirmed by looking at the effect of peak current on temperature. Hence, in examining these curves, the three important aspects on which to study the temperature dependence is the peak current (on-state), excess current (off state), and the conductance slope.

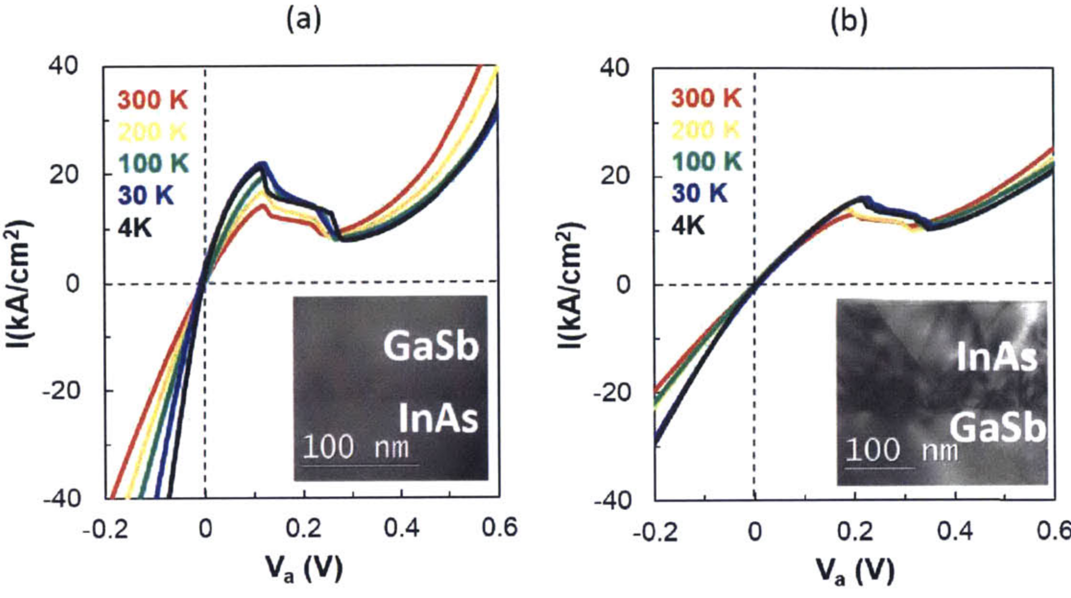


Figure 6.2 (a) I-V curve at varying temperatures for GaSb-on-InAs device and (b) I-V curves at varying temperatures for InAs-on-GaSb device. Inset shows a cross-sectional image on the [220] diffraction condition for each sample

6.3 Peak Current

The peak current of both devices as a function of temperature is shown in Figure 6.3a for all temperatures measured. The increase in peak current as temperature is decreased has been reported previously in InAs/GaSb[91] and similar materials systems[92], and is attributed to smaller thermal tails as the temperature drops, allowing a greater percentage of all carriers to have low enough energy to be within the region of band overlap to tunnel, as illustrated in Figure 6.3b. This is the reverse of the typical scenario, in which carriers must overcome a barrier, and in which the current then increases with temperature. Hence, this inverse temperature dependence is

a signature of tunneling and serves as additional evidence that the on-state current is dominated by interband tunneling.

However, as the temperature decreases below about 60 K, we have consistently observed all devices on both samples show an inversion of the trend, and the peak current begins to decrease with decreasing temperature. Below this temperature, the vast majority of carriers are low enough energy to tunnel, but there are now fewer total carriers due to the fact that the Fermi-level increase with decreasing temperature has resulted in less of an accumulation region near the interface. To illustrate this, Figure 6.3c shows a comparison of the band-edges near the interface at 300 K and 4 K, calculated from a solution of the Poisson and Schrodinger equations, with the temperature-dependence of band gaps included via Varshni parameters. At 300 K, it is likely that there is quantum confinement and a very high density 2D electron and hole gas present at either side of the interface, since wells have formed in the InAs conduction band on one side, and in the GaSb valence band on the other side. As the temperature drops, this confinement is lost, and at 4 K, the confinement is minimal and the accumulation region is not deep, which likely leads to the decrease in peak current at these low temperatures.

Interestingly, the temperature dependence is considerably more pronounced for the higher material quality GaSb/InAs devices. Again, differences in series-resistance between the two samples can be ruled out as the cause, since series resistance does not affect the peak current. Differences in doping are also unlikely, since both samples are doped identically. Therefore, it is likely that the interface itself has caused this difference, likely due to differences in band-alignment that cause a shift in the fraction of carriers that are originally low-enough energy to tunnel, or by a more pronounced-loss of confinement and accumulation at lower temperatures for the more defective interfaces.

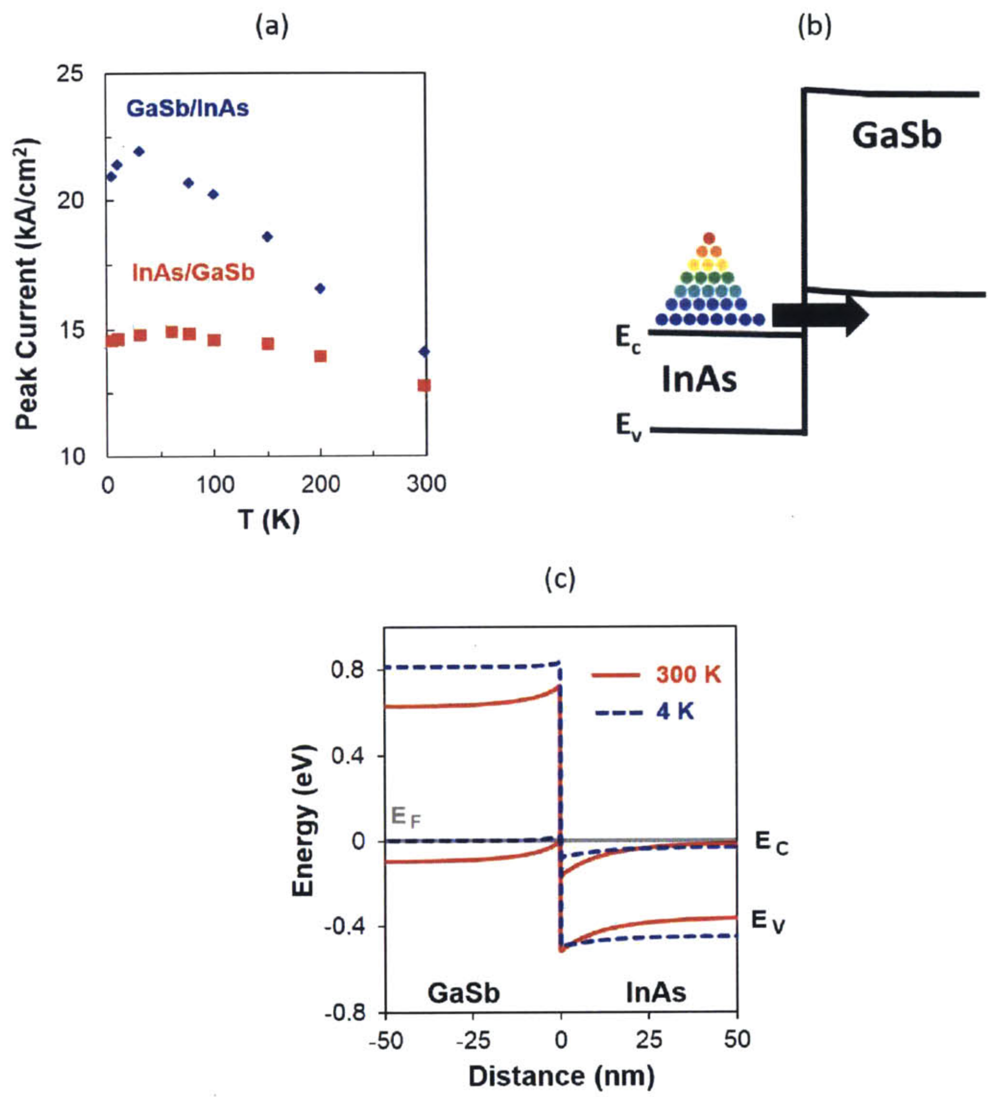


Figure 6.3: (a) Peak current vs temperature for both devices, (b) Illustration showing effect of Fermi-tail on tunneling temperature dependence (c) Band diagrams for an InAs/GaSb heterojunction at 300 K and 4K

6.4 Excess Current

In contrast, when tunneling has stopped, the valley and excess current show a much weaker temperature dependence. Figure 6.4a shows the current of a GaSb/InAs device as a

function of inverse temperature at various voltages, beginning near the valley voltage. The temperature dependence is stronger at higher temperature and then reaches a plateau. There is a slight increase in current at higher bias and lower temperature, which we attribute to the large increase in substrate mobility at very low temperatures. The temperature dependence with a plateau is similar to what was reported previously for resonant tunnel diodes in this material system[93], and the trend was attributed to hole current: thermionic hole emission at high temperatures and hole Fowler-Nordheim tunneling at lower temperatures/higher bias, as illustrated in Figure 6.4b. We observe higher currents and a general shift of the curve's features to higher temperature for our data, and we attribute this to the absence of an AlSb tunnel barrier, which results in Fowler-Nordheim tunneling being favored over thermionic emission up until a higher temperature, due to the higher tunnel probability. However, at low temperature and bias, our currents are still considerably high to be explained by any hole process given the low energy of the hole distribution. Such an anomaly was also observed in the resonant tunnel diode[93] and was hypothesized as potentially being due to dislocations. In Figure 6.4(c), we compare the valley currents of both samples, and see that while the shape of the temperature dependence is essentially the same, the valley current is lower for the dislocation-free device, and this is consistently observed across all devices on both samples, and consistent with the results of the previous chapter. This may be a confirmation that trap-assisted tunneling from the dislocations and other defects (as shown in Figure 6.4b) is occurring. The difference in current is quite small[94], and this may be an indication that other defect states are still present in the GaSb/InAs device.

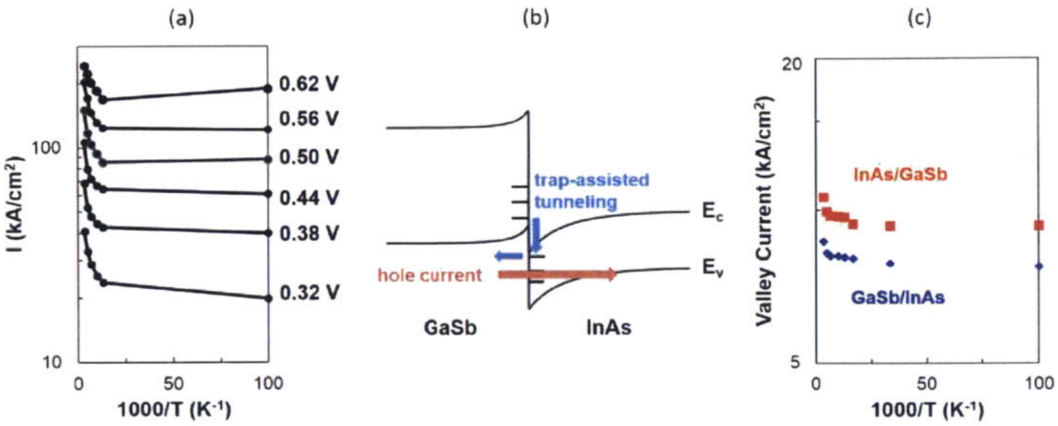


Figure 6.4: (a) Temperature dependence of the current at various bias points beyond the NDR region for GaSb/InAs device. (b) Band diagram for a device biased at 0.15 V, showing an example of a trap-assisted tunneling leakage route, in addition to a hole current (c) Valley current temperature dependence for both the GaSb/InAs and InAs/GaSb devices

6.5 Conductance Slope

Finally, we can extract a conductance slope from both devices as a function of temperature. Figure 6.5a shows the absolute conductance vs voltage curve for the GaSb/InAs device at varying temperatures. Figure 6.5b shows the conductance slope extracted as a function of temperature. As can be seen, the temperature dependence is extremely weak, and after correcting for the decrease in series resistance as temperature decreases, the temperature dependence essentially disappears. When we compare to the conductance slope values for the more defective InAs/GaSb device, we see they are less steep, and also have no temperature dependence. We have shown previously that the more defective devices have less steep conductance slope due to more defect states and less-unifiorm band-alignment.[95] We also showed that all devices gained improvement with annealing, which was attributed to a reduction

in point defect concentrations and better smoothing of intermixing and point defect concentrations. Such a result indicated that even the dislocation-free samples were still limited by point defects and material inhomogeneity, and this is consistent with the valley current result, where the removal of dislocations gives only a slight decrease. The lack of temperature dependence of conductance slope is consistent with this, as it indicates that the steepness for both devices is still limited by defects and material inhomogeneity, even when the interface appears structurally perfect in TEM. Such inhomogeneity and defects overpower any thermal-blurring effects such as band-edge blurring due to the deformation potentials caused by phonons, or phonon-assisted tunneling after the bands are misaligned. Therefore, the first question of this chapter has been answered: even for structures appearing defect-free in cross section TEM, material inhomogeneity still limits the steepness of the band edges.

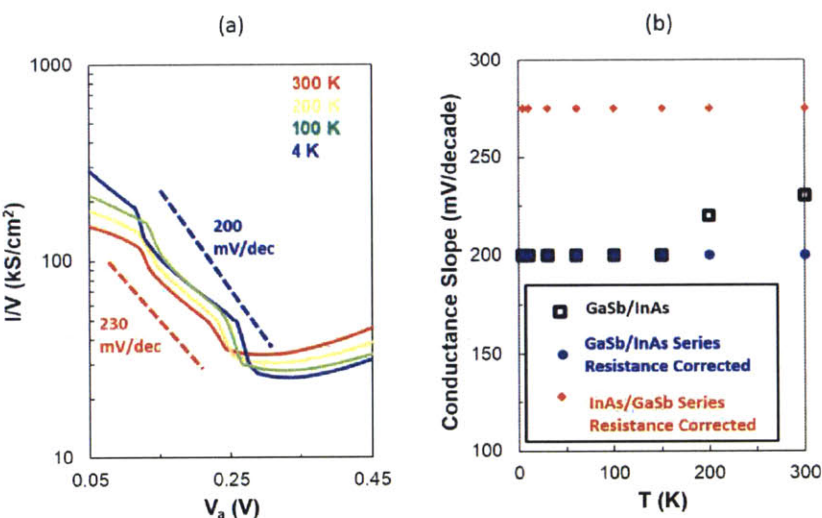


Figure 6.5: (a) Absolute conductance-voltage curve at select temperatures for GaSb/InAs device. (b) Conductance slope as a function of temperature for both devices. For GaSb/InAs, a comparison of the results before and after correcting for series resistance is displayed.

Question 2 posed by this chapter is now also answered: three-terminal paraitsics have been removed, and now for the first time ever, there is no temperature dependence seen. Figure 6.6 compares these results to the published results that were shown in Chapter 1.

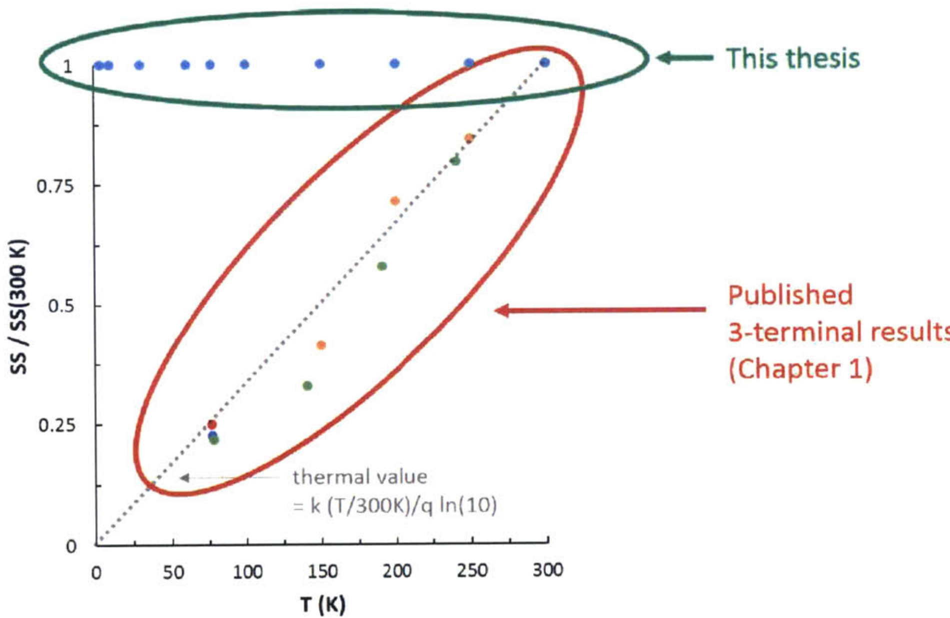


Figure 6.6: Temperature-independent conductance slope measurements from this thesis, overlaid on Figure 1.7, which shows published TFET subthreshold slopes with a kT/q ln(10) temperature dependence

Mookerja[54] proposed that in the region of steep-subthreshold switching, current is not true band-to-band tunneling, but tunneling into trap states at the gate-oxide/semiconductor interface and subsequent generation into the conduction band. Our results support this hypothesis, since two terminal measurements were chosen specifically to avoid a gate oxide, and in the absence of it, there is no temperature dependence. Figure 6.7 shows an example of how this could work. Because of the small effective band gap of the heterojunction and the fact that

the channel thickness is usually quite small (<20 nm typically), there can easily be a high tunnel probability to tunnel into the gate-oxide states when the bands are misaligned. From here, electrons can simply generate into the conduction band in the channel and be collected by the drain. It is possible for the generation step to be rate limiting, in which case, a strong temperature dependence is present.

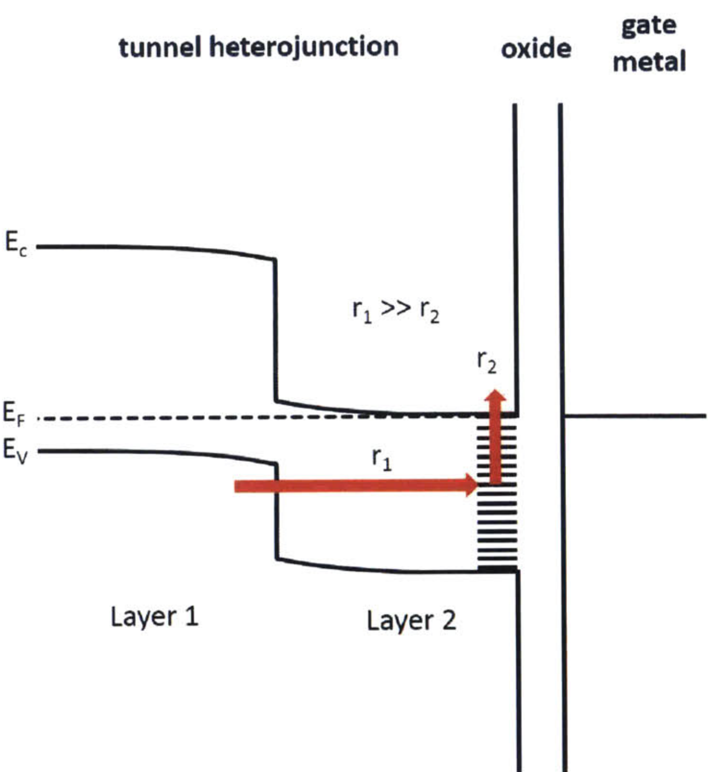


Figure 6.7: Band schematic of a heterojunction below a gate oxide and metal. r1 is a rate describing tunneling into gate oxide trap states. r2 is a rate describing generation into the conduction band from the trap states, which is thermally activated

In terms of the two possible parasitic mechanisms presented at the beginning of this chapter, this mechanism represents a parallel conductance as was shown in Figure 6.1a. But it can also be possible to have a series resistance as was shown in Figure 6.1b. Keeping with Figure 6.7, once carriers have reached the conduction band, either through this mechanism or the desired interband tunneling, their concentration depends on gate voltage, as the gate will shift the Fermi energy. In other words, the channel has a concentration of carriers in the same manner as a basic MOSFET. Therefore, if conduction through the channel is more resistive than the resistance of the tunnel junction, this MOSFET-like characteristic will dominate the I-V curve, and so the device will behave as though it has a resistive MOSFET in series. Once again, a strong thermal dependence will be seen.

Both of these possible effects are absent in the devices in this thesis, and so the lack of temperature dependence in this work confirms that it is indeed a three-terminal parasitic that causes the temperature dependence seen in published devices, possible one of the two described here, that is dominating most or all published TFET devices.

Chapter 7: Ternary Growth of InGaAs/GaAsSb Heterojunctions

7.1 Introduction

The previous chapters established growth and electrical results of the binary system InAs/GaSb with type-III band alignment. The following chapters will assess the effect of the alloyed materials system, InGaAs/GaAsSb as a means of changing the band alignment. Chapter 1 showed how this system results in a shift in band alignment from type-III towards type-II. While it has clearly been seen that materials defects limit performance, it is possible that improvement will be gained from changing the band-alignment for constant defect density. Alternatively, it may be possible to obtain lower defect densities with ternary compounds, particularly since the lattice constant can be controlled and so lattice matching is relatively simple. The list below summarizes all important aspects of ternary compounds and potential advantages, all of which will be explained in more detail over the next few chapters.

- Shifting the band-alignment to type-II allows for devices to function as normally-off transistors instead of normally-on
- Bringing the band-alignment right to the border of type-II and type-III will bring the switching point to zero-bias:
    - In a transistor configuration, this means that the least possible amount of voltage will be required to switch the device, and hence, the smallest potential amount of field will be perturbing the bands, which may increase steepness
    - In a two-terminal configuration, this causes the maximum curvature to occur at zero-bias, allowing for low-voltage, high curvature analog devices without the need for a biasing network
- Extracting a conductance slope becomes much simpler and useful because:

    - As the band-overlap of type-III alignment decreases, the negative differential resistance decreases and eventually disappears, which removes the instability in the I-V curve. As a result, the true nature of the conductance slope at all bias points can be seen without having to average
    - As type-II band alignment is reached, an effective band gap at the interface forms, which leads to a lower tunnel probability, and hence a lower junction conductance (but still high enough to be competitive with CMOS). This ultimately results in the series resistance becoming less significant. It will be shown that I-V curves can be assessed in their raw form without any series resistance correction
- Potential for lower defect densities since lattice-matching can be done

With that said, however, there is the considerable challenge of perfecting the materials growth of ternary materials:

- The ability to calibrate the flowrates of the MOCVD precursors to reliably obtain a specific composition and lattice constant is paramount.
- The system is not lattice-matched to a binary substrate, and so it is necessary to develop a method to grade the lattice constant to the correct composition from a binary substrate
- It will be seen soon that this alloyed system is even more affected by interface phenomena beyond that of binary growth, and additional ingenuity becomes necessary to prevent defect formation.

These growth challenges and how they are overcome is the focus of this chapter. The subsequent chapter will then focus on assessing electrical characteristics of tunneling in these structures and determining if the potential benefits postulated here have been realized.

7.2 Composition Control of InGaAs and GaAsSb

7.2.1 InGaAs

InGaAs growth is accomplished by flowing both TMIn and TMGa into the chamber simultaneously at the correct ratio in order to obtain the desired composition. Typically, the growth temperature of 530°C that was used for the binary materials was also used for both InGaAs and GaAsSb. The main challenge with InGaAs growth is that the TMIn precursor is the only precursor that is actually a solid-state source. Carrier H_2 gas is flown through a cylinder with TMIn at its vapor pressure which results in a specific incorporation rate of TMIn in the flow stream. However, due to the much lower vapor pressure, the TMIn molar flow rate can be difficult to control, especially since condensation of TMIn on the walls of the cylinder and piping is possible. While typically the precursor molar flow rate is linear with the total volumetric flow rate through the cylinder, because of these solid-source issues, TMIn tends to have some nonlinearity in molar flow rate. Therefore, it is necessary to fix the flow rate of TMIn and change the TMGa flow to control the group-III precursor flow ratio, and change the AsH_3 flow accordingly to maintain a constant V/III ratio. In this work, the TMIn flow was fixed at 206 sccm, and the V/III ratio was fixed at 90. Figure 7.1 shows the % Ga incorporation in the film as a function of the ratio of TMGa flow (in sccm) to total group V flow (TMGa + TMIn). Composition was determined using [004] and [224] XRD reciprocal space maps as was described in section 3.4.2.

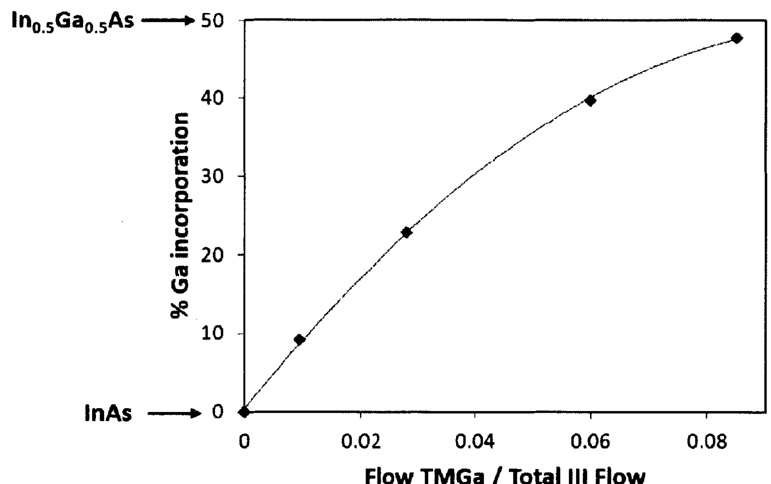


Figure 7.1 %Ga composition in InGaAs measured by HRXRD as a function of TMGa precursor flow as a fraction of total group-III precursor flow

7.2.2 GaAsSb

GaAsSb differs from InGaAs in that it is a mixed group-V ternary instead of a mixed group-III ternary. Hence, lattice constant and composition is controlled by controlling the ratio of group-V precursors. The challenge in this case is that, although As is being alloyed in now, GaAsSb is still an antimonide, and so the challenges related to Sb's low vapor pressure as described in section 4.1 still apply. Hence, the AsH_3 flow fraction must be controlled while still maintaining a V/III ratio of 1.2 to prevent excess Sb on the surface. This is accomplished by fixing the TMSb flow at 12.7 sccm and applying very small AsH_3 flows in order to meet the ratio required and still keep the V/III ratio low. The required flows for AsH_3 turn out to be in the range of 0.01-1 sccm. Such small flow is accomplished using a double dilution setup, where the AsH_3 flow is initially diluted with H_2 carrier gas, and the majority of the resultant mixture flow is dumped while a small fraction is mixed again with more H_2 carrier gas. The group-III flow (TMGa) is adjusted to maintain a constant V/III ratio of 1.2. The As % incorporation in the film

as a function of AsH_3 flow fraction is shown in Figure 7.2. It will be shown later that, unlike InGaAs, this relationship is only a bulk average. Due to the tendency of Sb and As to exchange (described in section 4.2), there is actually likely a gradient of composition at the beginning of layer growth in which the film is likely slightly As-rich. This turns out to have a major effect on heterojunction growth, which will be discussed later in this chapter.

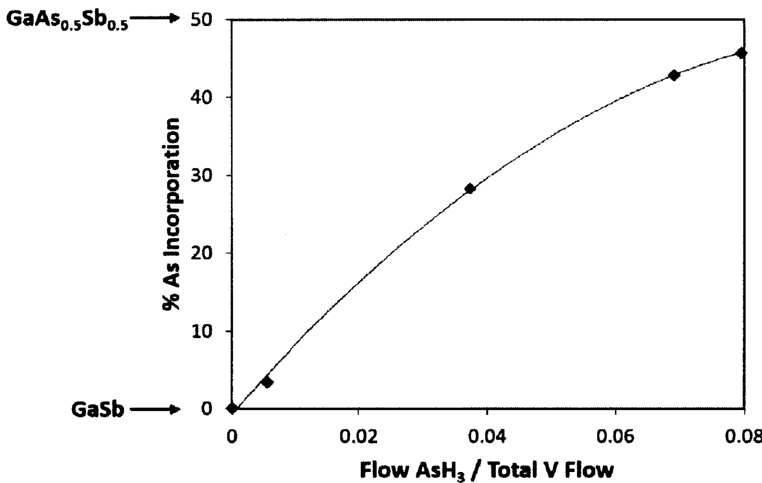


Figure 7.2: %As composition in InGaAs measured by HRXRD as a function of AsH_3 precursor flow as a fraction of total group-V precursor flow

7.3 InGaAs/GaAsSb Heterojunctions

We will start by assessing InGaAs/GaAsSb heterojunctions. Later, we will examine the reverse growth order, which in the case of the binary compounds was superior. However, it was also shown in section 4.4 that when the GaSb underlayer was pseudomorphically strained, the pseudo-lattice match resulted in a non-defective InAs layer. This implies that if InGaAs and

GaAsSb are properly lattice matched to each other based on the calibrations in Figures 7.1 and 7.2, then it may be possible to obtain InGaAs/GaAsSb heterojunctions free of dislocations.

Figure 7.3a shows the simplest possible manner of growing this structure, which begins with a GaSb substrate, followed by growth of the GaAsSb layer at the chosen composition. The InGaAs layer is grown above this. For the specific case of a lattice matched $In_{0.76}Ga_{0.24}As/GaAs_{0.3}Sb_{0.7}$ structure, a cross-section TEM image on the [220] diffraction condition is shown in Figure 7.3b, and a [224] reciprocal space map is shown in Figure 7.3c.

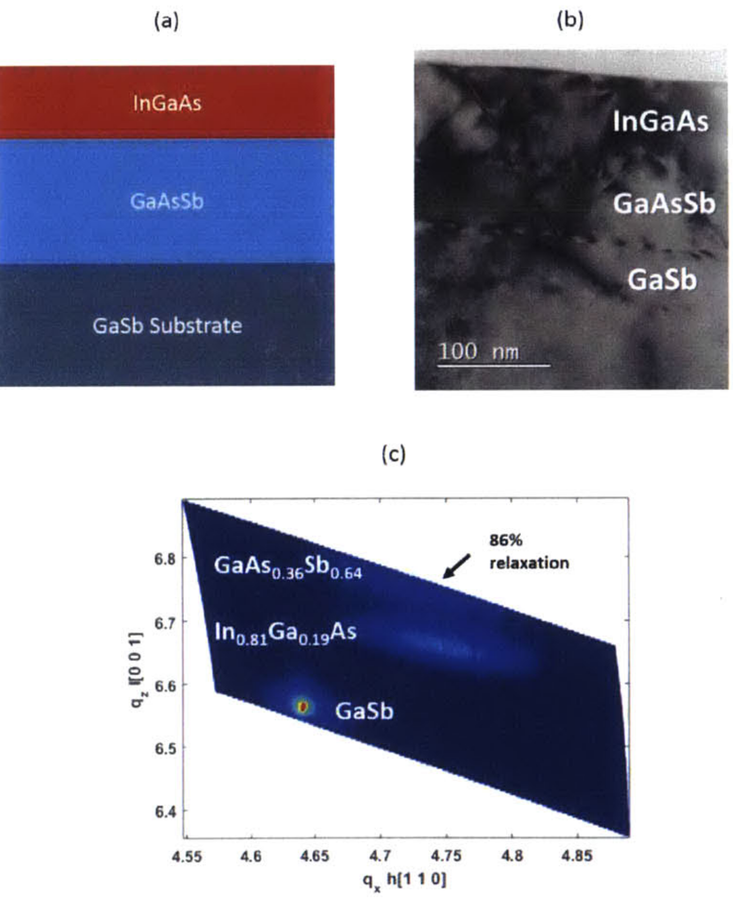
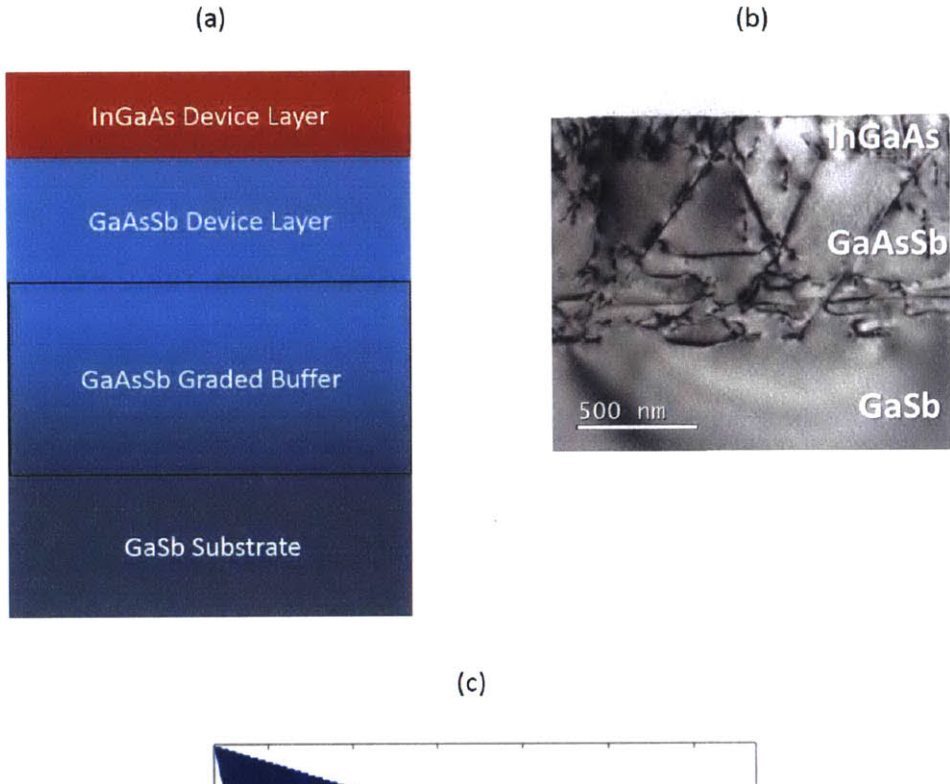


Figure 7.3 (a) schematic of InGaAs/GaAsSb/GaSb structure, (b) cross section TEM image of structure on the [220] diffraction condition, (c) HRXRD [224] reciprocal space map of structure

As can be seen, the structure is defective. The GaAsSb layer is expected to relax its strain, via dislocations, and so the dislocations seen originating at the GaAsSb/GaSb interface is to be expected. The threads then cross the interface, however, which is problematic. But the more serious problem is that dislocations also originate at the InGaAs/GaAsSb interface, which indicates that there are misfit dislocations present at the tunnel junction. This is known from the results of Chapter 5 to seriously degrade steepness. The origin of these dislocations is for two reasons, but primarily the second. The first is that the composition of the layers is not perfectly at the target composition. This is always the case as there is always some drift in calibration. This creates a slight lattice mismatch. But the more important reason can be seen in the [224] RSM in Figure 7.3c: the GaAsSb layer is only 80% relaxed. Hence, even if the layers were calibrated correctly, the InGaAs layer would have to be matched to 80% relaxed $GaAs_{30}Sb_{70}$, not the true lattice-constant of $GaAs_{30}Sb_{70}$. Hence, the main source of mismatch comes from the fact that there is still residual strain in the GaAsSb layer that then translates into lattice-mismatch at the InGaAs/GaAsSb interface.

7.4 GaAsSb Graded Buffers

The logical extension of the basic structure to include a buffer is to simply grade the composition of the GaAsSb layer instead of immediately growing it at the device composition. A schematic of such a structure is shown in Figure 7.4a. GaAsSb buffers were grown initially at 530°C to a thickness of 800 nm, to a composition target of 64% Sb giving a strain grade rate of 3.3 % strain/μm. This is a relatively high strain grade rate, due to the fact that GaAsSb cannot be grown quickly (for the reasons described in section 4.1).



(c)

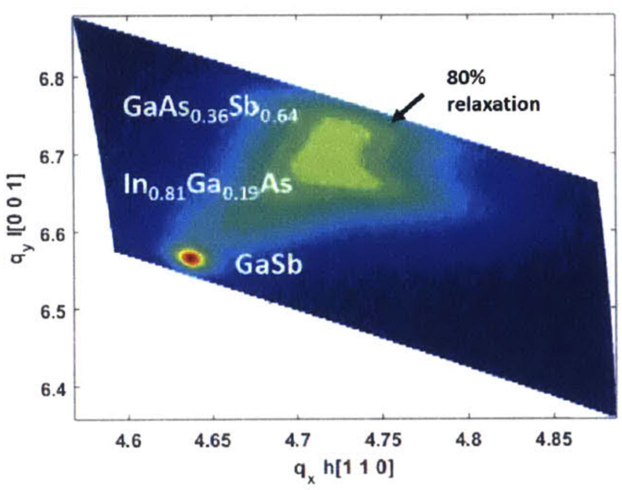


Figure 7.4 (a) schematic of InGaAs/GaAsSb/GaAsSb graded buffer/GaSb structure, (b) cross section TEM image of structure on the [220] diffraction condition, (c) HRXRD [224] reciprocal space map of structure

It can be seen from the TEM image in Figure 7.4b that there is clear glide of misfit dislocations relieving the strain, especially in the deeper portions of the buffer. However, there

are threading dislocations, and potentially stacking faults visible, indicating that the buffer is not effectively recycling threading dislocations and relieving the strain effectively. Furthermore, the [224] RSM plot in Figure 7.4b reveals that still only 80% relaxation has been achieved. For the same reason as seen in the previous section, additional dislocations originate from the tunnel interface.

One potential solution to all of these problems would be to lower the strain grade rate. However, this is somewhat impractical because of the slow growth rate of GaAsSb at 530°C. Another option is to raise the temperature. The benefits of this are twofold: 1) it can allow for a higher growth rate if the growth is in the temperature-limited regime, and 2) it can lower the barriers to dislocation glide making it easier to recycle threads. For this reason, a GaAsSb buffer layer was grown at 580°C. Previous results by Yang [96] have shown good quality GaAsSb buffers grown by MOCVD can be obtained at approximately this temperature (575°C) when grading from the GaAs lattice constant to the InP lattice constant. It was found that going higher than this temperature led to a decrease in material quality.

Figure 7.5a shows a TEM image of this structure, on the [220] diffraction condition as always. The thickness of the structure is now 2.3 μm, confirming that the growth was in the temperature-limited regime, and indicating an increase in growth rate by a factor of about 2.9. This results in a strain grade rate of 1.1%, at rate at which Yang [96] reported threading dislocations of 5×10^{6} cm^{-2}. However, the buffer layer in the TEM image is considerably more defective than that and shows an incredibly high dislocation density. This indicates that there is a profound difference when grading from the GaSb lattice constant towards the InP lattice constant when compared to grading from the GaAs lattice constant. Furthermore, it seems that increasing the temperature and decreasing the grade rate has actually resulted in a considerably worse film.

One potential explanation for this comes from examining the very beginning of the buffer. For the 530°C film in Figure 7.4a, all that is seen are a few misfit dislocations at the very beginning of the grade. However, the interface where the grade begins for the 580°C film is a much more distinct interface, and appears more defective than the rest of the film. Figure 7.5b shows a more magnified image of the interface. There are clear defects, likely stacking faults, originating from the interface. This is highly indicative of a large misfit strain between the start of the GaAsSb buffer and the underlying GaSb substrate. This is uncharacteristic of a graded buffer, as there should be no large step in strain about an interface if the strain is slowly being graded. This implies that some other effect has occurred to create this strain.

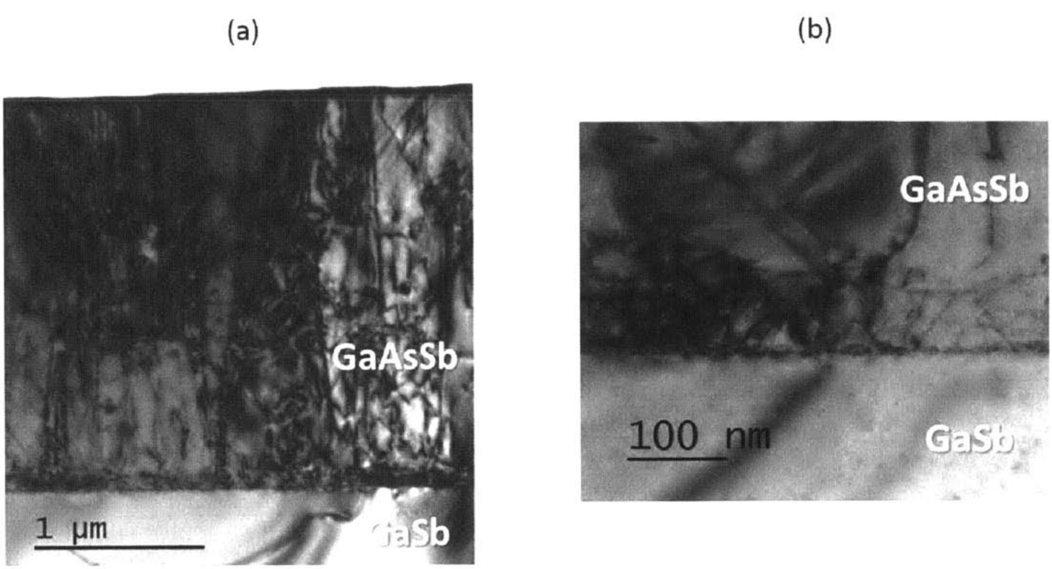


Figure 7.5 (a) TEM image on the [220] diffraction condition of a GaAsSb buffer grown at 580 °C on GaSb, showing many defects. (b) a more zoomed in image of this sample, showing a very high concentration of defects originating right from the interface where the buffer was initiated: a rare observation for a graded buffer

One potential explanation for this is that Sb-As exchange is occurring again, similar to with InAs on GaSb growth, as was described in section 4.3. For the case of InAs on GaSb growth, the Sb was confined to one layer, while the As was confined to the other. The situation is shown again in Figure 7.6a: Sb from the underlying GaSb layer exchanges with the As from the above layer. The problem was solved simply by growing in the reverse order so that Sb remained on top. However, this cannot be done for the case of GaAsSb growth, as As and Sb are now in the same layer. Figure 7.6b shows the scenario now. The first group-V layer of GaAsSb is originally a random mixture of As and Sb. However, as subsequent group-V layers are grown, As from higher layers exchanges with Sb from lower layers, leading to As-rich GaAsSb near the interface. In fact, in the case of complete exchange, it can be seen in Figure 7.6 that the final result is exactly the same in both scenarios.

This scenario is plausible as it does derive from the concept of exchange in the binary materials, which was explicitly observed in Chapter 4. Furthermore, this model is consistent with the two key findings so far for GaAsSb buffers:

1. The fact that it becomes worse at higher temperature: higher temperatures would enable for more exchange since more thermal energy is available to overcome the activation energy barrier of exchange
2. The fact that this was only observed in grading from the GaSb lattice constant, and not from the GaAs lattice constant, even when growing on the same MOCVD system with the same conditions [96]. When starting on GaAs and grading in the reverse direction, the effect helps the grade rate instead of harming it. The movement of As down into the film and Sb upwards would result in more As-rich GaAsSb, but since it is growing on GaAs, this would be less strained than the GaAsSb in the case of no exchange. As this effect wears off over many

layers, the result would be a slower grade rate than targeted, which is typically associated with better material quality.

This problem of intra-layer As-Sb exchange as temperature is increased puts a difficult constraint on GaAsSb buffers, since temperature cannot be increased to improve the buffer quality, and the buffer quality is not good enough to obtain low threading dislocation densities or full relaxation at 530 °C. Therefore, it becomes necessary to examine an alternate materials system that would not be affected by this.

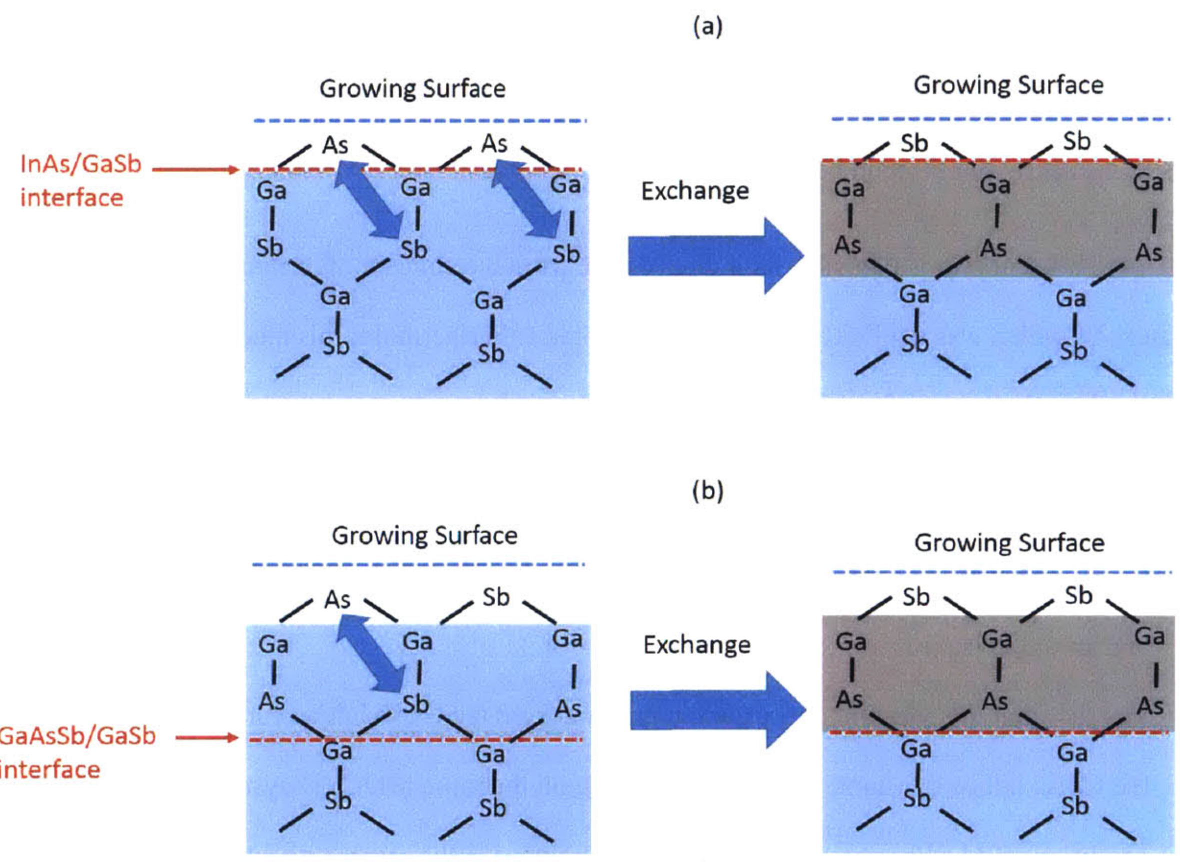


Figure 7.6 (a) Schematic of "interlayer" As-Sb exchange in an InAs/GaSb growth, as originally shown in Figure 4.6 of Chapter 4. (b) schematic of "intralayer" As-Sb exchange within a GaAsSb layer

7.5 InAsP Graded Buffers

For a variety of reasons, InAsP graded buffers are likely a better candidate for a graded buffer materials system as a means to achieve the required device lattice constant. It was explained in the previous section that intralayer As-Sb exchange is likely what plagues the GaAsSb buffer system on a GaSb platform. So the first major advantage of the InAsP system is that it will not be affected by this, since it does not involve Sb. An additional advantage is that, because of its lack of Sb, InAsP can be grown at much higher growth rates, allowing for much lower strain grade rates.

There is an additional benefit of grading with InAsP in terms of economics and ultimate applications: InAsP grading can begin on an InP substrate, compared to GaAsSb grading, which must be grown on a GaSb substrate. This is a major advantage, as InP substrates are considerably less expensive than GaSb. Allowing these devices to exist on an InP platform would be very advantageous for a variety of reasons:

- For the case of TFETs, it could allow for integration with a variety of other device technologies on the InP lattice constant, such as InGaAs MOSFETs. This could allow for TFET-CMOS heterointegration, which has been postulated to be advantageous [97]
- For the case of high curvature analog devices, it could allow for integration with photonics on an InP platform, which is a commonly used platform for photonics.
- Work has been done by our group and others to grade from Si to GaAs [98]–[101]and from GaAs to InP. [102] Hence, for the case of both TFETs and analog devices, integration on InP ultimately paves the way for integration on Si.

InAsP buffer capability has already been accomplished in MOCVD within our group, beginning at the InP lattice constant and moving to higher lattice constant, [103] although not as far out as is required. Nevertheless, these previous optimized conditions were used as a starting point. The following parameters were used to grow a buffer from InP to the lattice constant of $GaAs_{0.3}Sb_{0.7}$:

Table 7.1: Growth Parameters for InAsP Buffer

Temperature:	650°C
Grade Rate:	0.25 % strain/μm
Buffer thickness:	5 μm
Initial Strain jump:	0.3% ($InAs_{0.09}P_{0.91}$)
Buffer cap:	150 nm InAsP, then 150 nm InGaAs

The initial strain jump refers to a ~250 nm layer of $InAs_{0.09}P_{0.91}$ being grown on InP, then grading from there to the final composition of the buffer. The reason for doing this instead of grading directly from 0% As is that a small initial amount of strain is needed to act as a driving force to nucleate misfit dislocations that are spatially uniform and have high glide velocities which ultimately lead to low defect densities. If not enough driving force is initially present, lower-activation energy dislocations are nucleated instead, which tend to be from wafer edges or particles, and typically do not have as high glide velocities and result in higher defect densities. This is referred to as three-phase nucleation [103]. The buffers are capped with a layer of InAsP at the final composition ($InAs_{0.49}P_{0.51}$) in order to add enough strain energy to fully relax the buffer. This is then capped by a lattice matched $In_{0.76}Ga_{0.24}As$ layer which can either serve as the lower device layer if we are making a GaAsSb/InGaAs device (Figure 7.7a) or serve as an intermediate layer for GaAsSb to grow on if we are making an InGaAs/GaAsSb device (Figure 7.7b). For the latter case, the reason that InGaAs would be used as an intermediate layer instead

of growing GaAsSb directly on the InAsP would be to prevent an even more complicated exchange interaction between the Sb and P. This would likely occur and be even more pronounced, since P is an even smaller atom compared to As, has an even higher surface free-energy, and would hence have an even greater driving force to exchange. Therefore, the buffer need be capped with InGaAs regardless of the device in mind.

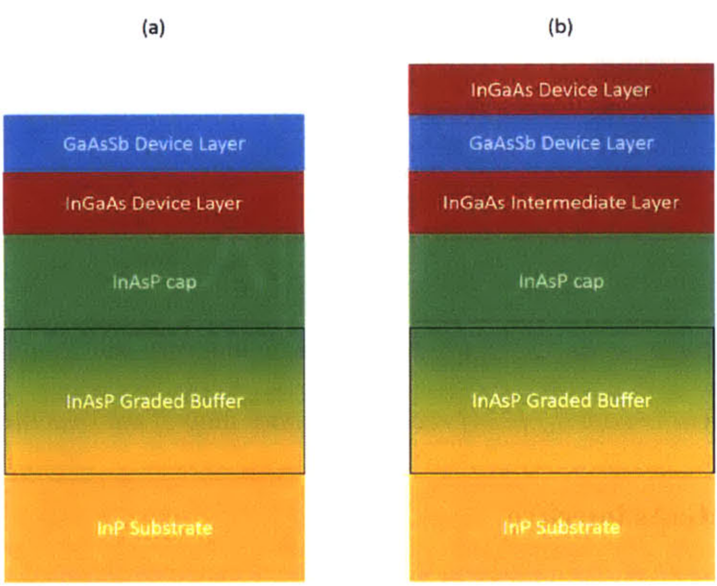


Figure 7.7 (a) GaAsSb/InGaAs device heterojunction on an InAsP buffer platform. (b) InGaAs/GaAsSb device heterojunction on the same platform

The results of these buffer parameters is shown in Figure 7.8. From the TEM image in Figure 7.8a, it can be seen that the buffer is of considerably better quality than for GaAsSb buffers: misfit dislocations can be seen relieving strain throughout the entire structure, with no threading dislocations visible at the top of the buffer or in the top InGaAs layer. The [224] RSM reveals a much higher degree of relaxation, 93%, has been achieved. Hence, overall, this is a considerably better platform for InGaAs/GaAsSb tunnel devices.

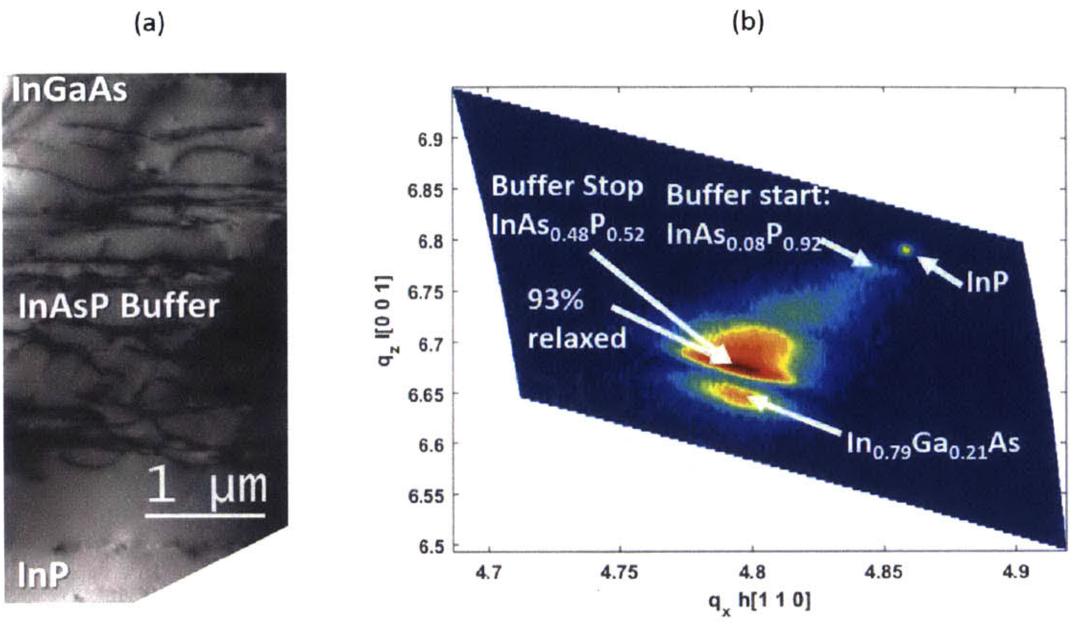


Figure 7.8 TEM image on the [220] diffraction condition for an InAsP buffer on InP, with an InGaAs cap. (b) [224] reciprocal space map of the structure.

7.6 The GaAsSb/InGaAs Interface

Now that a working buffer platform exists, the next step is to reintroduce GaAsSb to the structure. It was shown previously in section 7.4 that at 580°C, GaAsSb suffered from a likely intra-layer exchange that degraded quality. However, at 530°C, this problem was not pronounced. Hence, it is obvious that GaAsSb layer growth should be kept at 530°C. This was done simply by ramping the temperature from 650°C down to 530°C after the InGaAs cap was grown. An additional 100 nm of InGaAs was then grown at 530°C, and then the GaAsSb layer was grown. Following this, an InGaAs layer was grown on top, so that both GaAsSb/InGaAs and GaAsSb/InGaAs interfaces could be examined (refer back to the structure in Figure 7.7b for a full schematic). The [220] TEM image of the $In_{0.76}Ga_{0.24}As$ and $GaAs_{0.3}Sb_{0.7}$ layers is shown in Figure 7.9a, and an AFM of the top surface is shown in Figure 7.9b, indicating island growth.

Both the InGaAs and GaAsSb layers are considerably defective. The defects seem to originate from the GaAsSb/InGaAs interface, where numerous stacking faults can be seen, analogous to the 580°C GaAsSb/GaAs interface seen in Figure 7.5b. This is an indication that a similar effect may be occurring where intralayer As-Sb exchange is occurring. While this was not seen before at 530°C for the GaAsSb/GaSb interface, it is possible that now that the layer is being initiated at a higher As composition (30% As), there is sufficiently more As present to swap with Sb so that the effect is noticeable. In theory, there should be a composition at which this effect is maximal: As As% in the film is increased, more exchange occurs, although as the lattice constant is moved closer to GaAs, the mismatch resulting from this exchange becomes less. It is possible and likely then that $GaAs_{0.3}Sb_{0.7}$ is closer to this maximum composition compared to when GaAsSb first begins grading from the GaSb lattice constant at effectively 0% As.

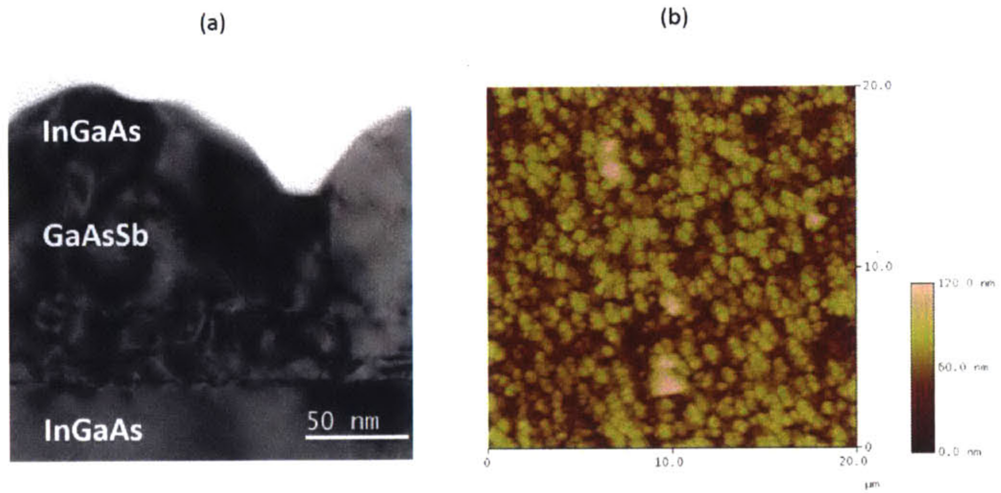


Figure 7.9 (a) [220] TEM image of InGaAs/GaAsSb heterojunction grown on the InGaAs-capped InAsp buffer. (b) Corresponding AFM image of the top surface of the top InGaAs layer, indicating high roughness and likely island growth

To test that this observation is due to the exchange process a variety of experiments were done to rule out other phenomenon. First and foremost, to ensure that this is not somehow due to the strain fields of dislocations in the buffer, a lattice matched InGaAs/GaAsSb/InGaAs/InP structure was grown, at the lattice constant of InP, with no buffer present, as shown in the schematic in Figure 7.10a.

As can be seen in Figure 7.10b, the film is less defective, but similar features can still be seen, with numerous stacking faults at the GaAsSb interface, pointed out by the white arrows. It is likely that the same effect is occurring, and that the composition is simply now past the composition point where this effect would be maximal. Figure 7.10c shows a Nomarski reflection micrograph of the surface, where features can be seen on the surface running along one of the two [110] surface directions. These features are likely correlated to the stacking faults seen in TEM. A similar structure was grown without the top InGaAs layer, and the Nomarski image appears to have the same linear features, as shown in Figure 7.10d. An additional structure, without the GaAsSb layer, shown in Figure 7.10e, shows a lack of these surface features, indicating they do in fact arise from the GaAsSb layer on InGaAs. The simplicity of this material structure and the ease of being able to see this effect with Nomarski makes it an excellent platform to study this presumed exchange effect and how it can be controlled.

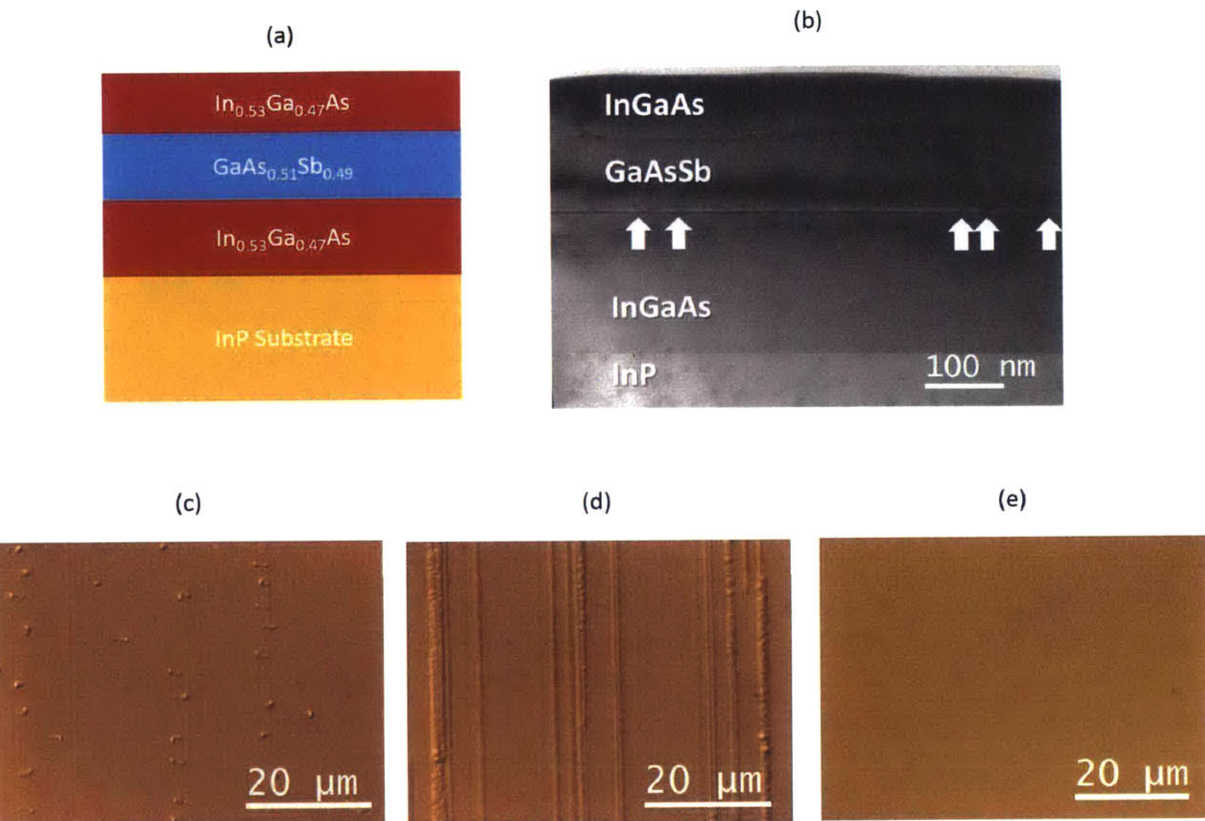


Figure 7.10: (a) Schematic of InGaAs/GaAsSb/InGaAs/InP test structure. (b) [220] TEM image of structure (c) Nomarski reflection micrograph of top surface showing defects running along a [110] direction (d) Nomarski Reflectance of structure grown with only GaAsSb as the top layer, (e) Nomarski micrograph of structure with only the InGaAs layer grown on InP

To test the origin of these features and how the quality can ultimately be controlled, a variety of test structures were grown with GaAsSb/InGaAs/InP to easily probe the GaAsSb/InGaAs interface with Nomarski reflectance microscopy. They are summarized in Figure 7.11, with their Nomarksi results included.

Sample	Description	Purpose	Nomarski
A	Shift GaAsSb composition by +1.5% Sb	Test if GaAsSb was off calibration or if compressive strain helps prevent defect formation	20 μm
B	Shift GaAsSb composition by -1.5% Sb	Test if GaAsSb was off calibration or if tensile strain helps prevent defect formation	20 μm
C	Interface sequence: 30s H_2 bake + 30s TMSb flow (12.7 sccm)	Test if it is due to interlayer As-Sb exchange and if it can be surpressed by adding Sb surface layer	20 μm

Figure 7.11 Test growth conditions and corresponding Nomarski reflection micrographs

Indeed, adding a specially-design interface gas switching sequence to the GaAsSb/InGaAs successfully removes these features, indicating it likely is removing the TEM defects. The design of this switching sequence is explained in Figure 7.12. The goal is ultimately to obtain at least one-monolayer of Sb coverage on the surface of InGaAs. Ideally, this Sb layer can "ride" the surface as a surfactant. Since the surface will be Sb-rich the entire time, there will be no driving force for exchange; In other words, there is no need to pull Sb atoms up from lower within the layer via exchange, since Sb atoms were already put on the surface via the pulsing sequence. As illustrated in Figure 7.12b, the H_2 bake is designed to force the As to desorb, leading to a group-III rich surface. The Sb flow (Figure 7.12c) then allows Sb to bond to the In

and Ga, forming at least one monolayer of Sb. The flow is kept low to prevent excess Sb from forming on the surface, although the material has some tolerance for this, as it can remain on the surface and slowly incorporate into the film. This Sb monolayer can remain on the surface (Figure 7.12d) preventing exchange.

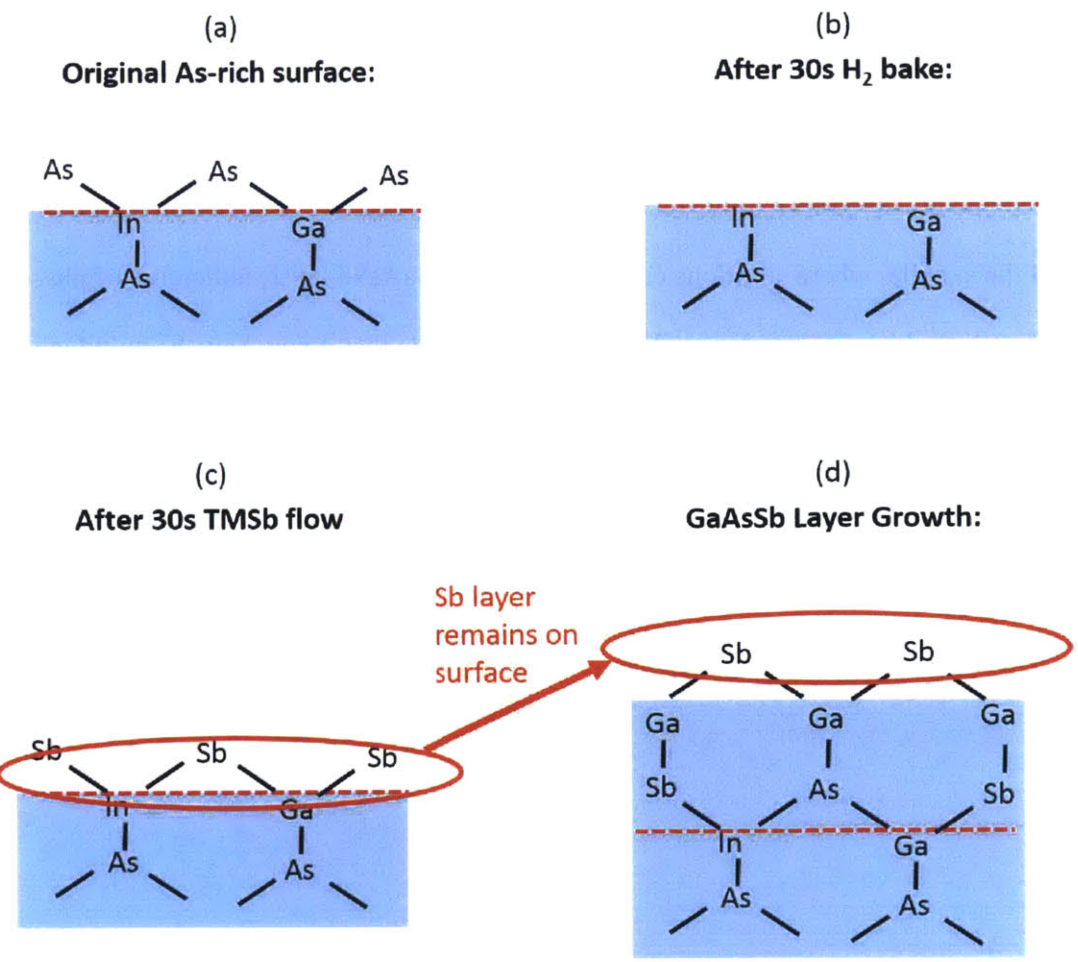


Figure 7.12 Illustration of the use of a special gas switching sequence to prevent intralayer As-Sb exchange. (a) Original As-rich surface, (b) surface after 30s H_2 bake with As desorbed, (c) surface after 30s TMSb flow with at least one monolayer of Sb on the surface (d) During growth, with Sb layer remaining on the surface

The next step is to apply this switching sequence to the device structure at the correct composition, as was originally attempted unsuccessfully in Figure 7.9 without this designed interface switching sequence. Figure 7.13 shows the result now with the switching sequence: a clean interface free of defects. In this particular image, the only visible feature is milling damage brought about by the TEM sample preparation process. The InGaAs layer above is also without defects. Additionally, the AFM image in Figure 7.13b, which can also be compared back to Figure 7.9b, is now showing a regular step structure, indicative that step-flow growth was maintained and no island formation occurred. However, Figure 7.13c shows an image of another region of the sample, where striations can be seen in the GaAsSb layer, indicative of phase separation. Since this was not observed before, it is possible that it may have been induced by the Sb on the surface. Nevertheless, a platform for growing good quality InGaAs/GaAsSb layers has now been established. The next chapter will focus on measuring these structures electrically.

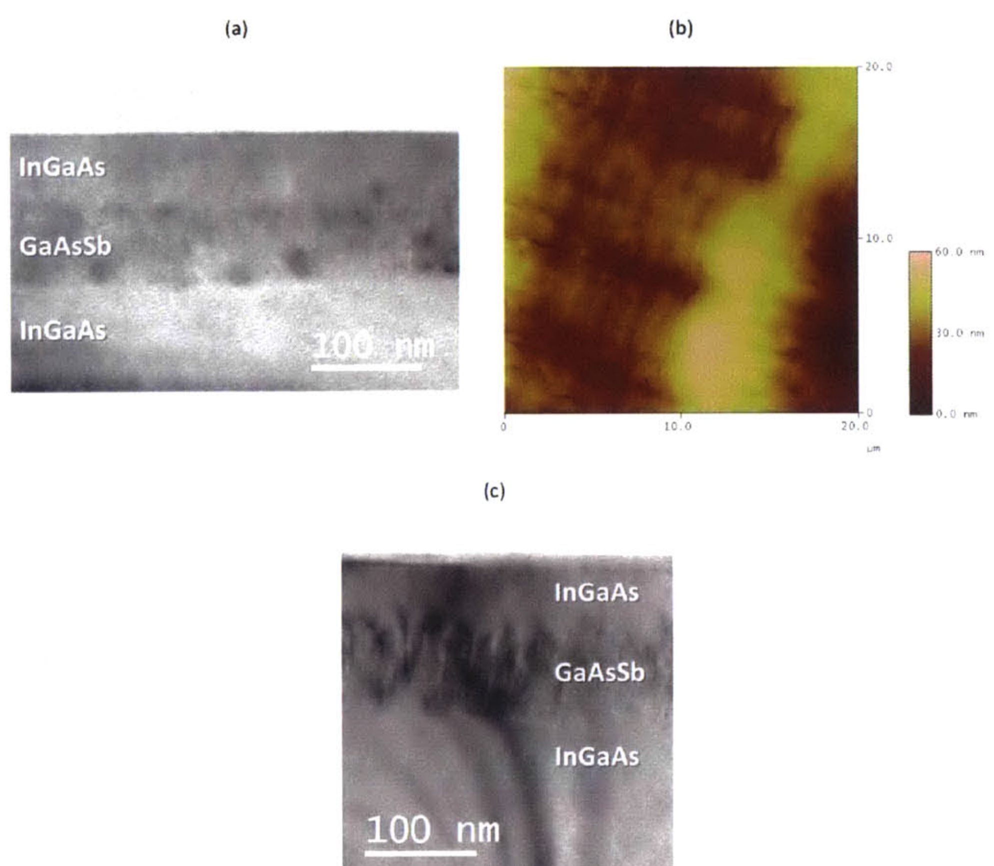


Figure 7.13: [220] TEM image of InGaAs/GaAsSb/InGaAs on an InAsP buffer using a 30s H_2 bake + 30s TMSb flow to initiate the GaAsSb layer. The dark circles are milling-induced defects. (b) AFM image of the top InGaAs showing evidence of step-flow growth. (c) [220] TEM image of a different region of the sample showing evidence of phase separation in the GaAsSb layer

Chapter 8: Electrical Characterization and Band-Alignment Control of InGaAs/GaAsSb Heterojunctions

8.1 Introduction

The previous chapter established the important aspects of growth of ternary InGaAs/GaAsSb structures. This chapter focuses on examining these structures electrically. The following questions will be answered:

- How does the band-alignment shift as we alloy the materials?
- How do other features, such as on and off-current change with band alignment?
- How does conductance slope depend on band-alignment?
    - This immediately determines how TFETs would perform at different band alignments. In chapter 9, the curvature coefficient will be examined, and analog applications will be assessed
- How do all these aspects depend on materials defects?

The InGaAs/GaAsSb structures also provide an opportunity to confirm the results of the previous few chapters, particularly because a conductance slope can easily be extracted from them raw without the special processing that was described in Chapter 5. Namely, the series resistance becomes insignificant because, as will be seen, the conductance of the junctions are lower, and so no correcting of the I-V curve is required. Also, as the alignment shifts away from type-III, the negative differential resistance becomes much smaller and eventually disappears. As a result, there are no instabilities in the I-V curve, and so no averaging of the conductance slope is required: The full extent and shape of the conductance swing can be observed.

8.2 Shifting Composition for InGaAs/GaAsSb Heterojunctions Without Buffers

Heterojunctions were grown initially using the InGaAs/GaAsSb/GaSb structure described in Chapter 7 and displayed in Figure 7.3a. Three different compositions of InGaAs/GaAsSb heterojunctions were grown in this study: InAs/ $GaAs_{0.1}Sb_{0.9}$, $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$, and $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ allowing for three different band alignments to be seen. Figure 8.1 summarizes the three compositions on the band-alignment lattice constant diagram. The samples fall on either side of the crossing point between type-II and type-III band alignment.

The I-V curves of the three heterojunctions at different compositions are shown in Figure 8.2. The corresponding conductance-voltage curves are shown in Figure 8.2b. Simulated band-diagrams for each device are shown in Figure 8.2c-e.

There is a noticeable shift of the conductance swing to the left, towards lower and more negative voltage, as the heterojunction is alloyed towards smaller lattice constant (Figure 8.2b). The red curve begins its conductance decrease right at zero applied bias and continues into forward bias. The blue curve has its conductance decrease begin in reverse bias and proceed into forward bias slightly. This is consistent with the band alignment shift between these structures. For example, Figure 8.2c shows the band diagram for the red curve: it is clearly still type-III alignment, with the GaSb valence band higher in energy than the InAs conductance band at the interface. As such, its I-V curve has negative differential resistance and the conductance swing is in forward bias. The band diagram for the blue curve shows type-II band alignment. As such, there should be no conductance at zero applied bias, and there should be a turn-on in conductance as increasing reverse bias is applied when the bands eventually overlap. Interestingly, however, the blue curve appears to not be shifted enough, in that the swing is not

entirely in reverse bias, and there is still some conductance at zero applied bias. Hence, there must still be some band-overlap, and so the structure is still electrically behaving with some type-III character, at the very least on the border of type-II and type-III. With regards to the conductance slope, a second important observation is that the conductance slope of this swing becomes less steep for the alloyed samples.

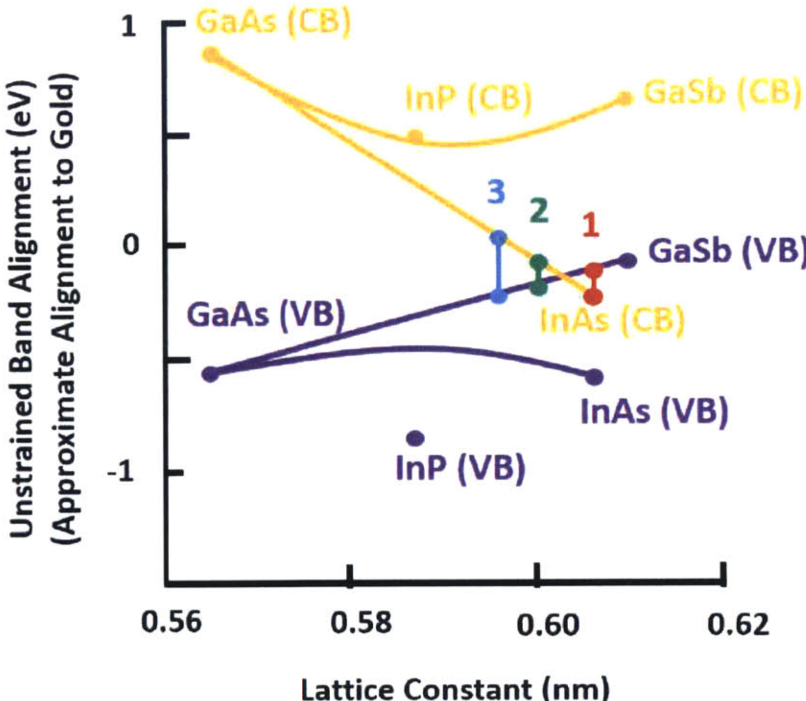


Figure 8.1 Conduction band energies (yellow) and valence band energies (purple) for III-V semiconductors between 0.56 and 0.62 nm lattice constant, plotted on a scale where the electron affinity of gold is defined as the zero point of energy. The three compositions are labelled. Taken from [104]

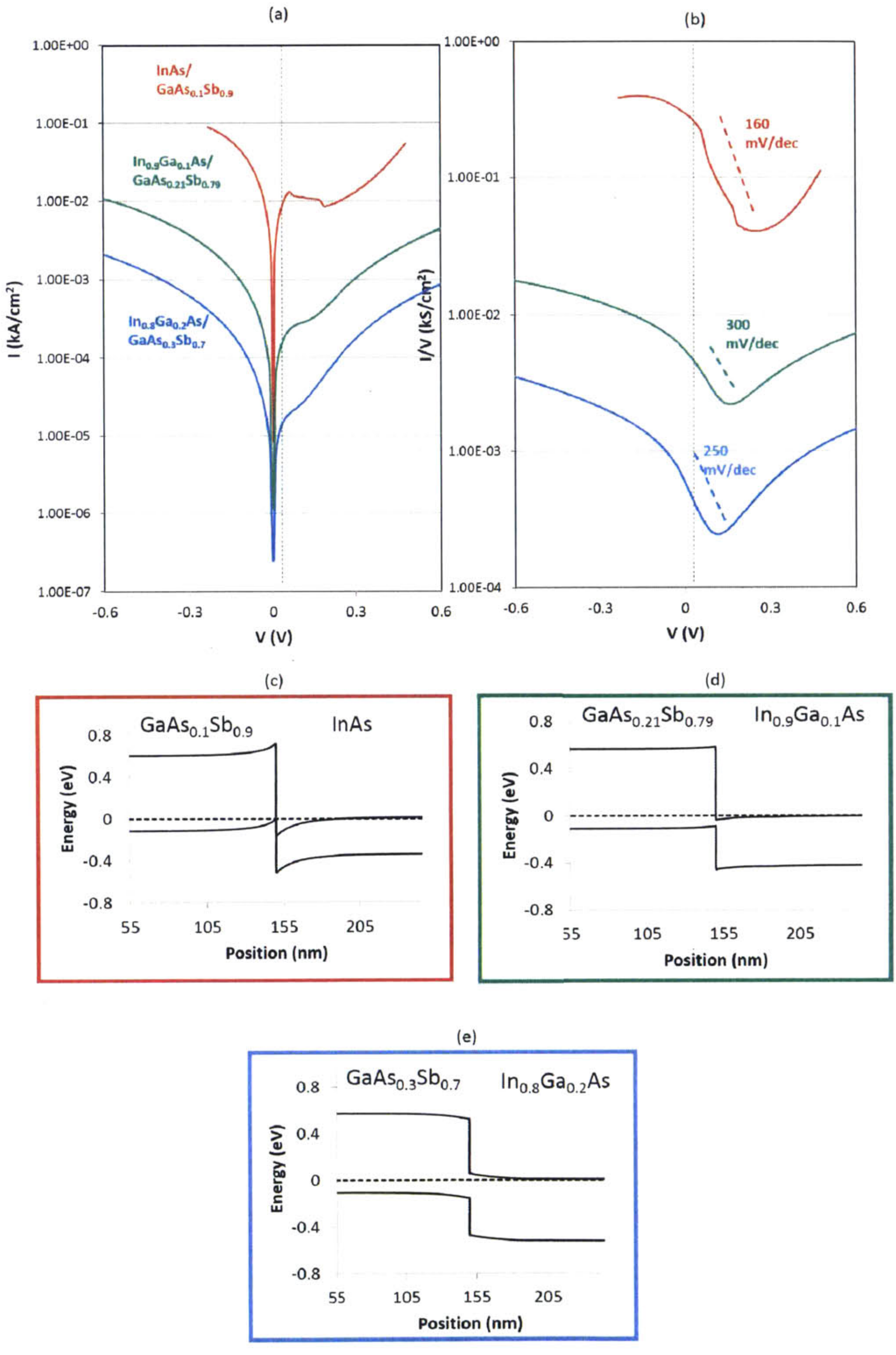


Figure 8.2 (a) I-V curves of InGaAs/GaAsSb heterojunctions at three different band alignments. (b) Corresponding I/V-V curves, (c)-(e) Band diagram of heterojunction at each composition

Therefore, there are two major observations to be made with regards to the conductance swing:

1. While the general trend of the conductance swing shifting to the left with alloying is consistent with theory, the shift is not as much as expected.
2. The conductance slope becomes less steep for the alloyed samples

This is indicative of defects. Indeed, we know from the previous chapter that these structures, grown without buffers, are considerably defective. It is also known from Chapter 5 that defects lead to a decrease in conductance slope, so observation #1 can be easily explained. Observation #2 can also be explained by this, as it was demonstrated in Chapter 2 how defects, as a result of their band bending, shift the band alignment in both directions in a spatially varying manner across the interface. This was also seen when examining defective versus non-defective type-III structures in Chapter 5. It was explained in Chapter 2 that regions shifted towards type-III dominate the current since they turn off last when being turned off and turn on first when being turned on. Therefore, spatially varying band alignment due to defects will always result in a general shift towards type-III band alignment. This concept will be revisited soon when we start comparing the effective band alignment of less defective structures with buffers at the same lattice constant.

There is also a major decrease in current in all regimes as band-alignment is shifted to type-II. In examining the band-diagrams in Figures 8.2c-d, at constant doping, there are accumulation regions near the interface that ultimately result from the type-III band alignment. However, for the more type-II-like structure, there are depletion regions instead. As the bands are made to overlap with applied bias, these depletion region become part of the tunnel barrier,

resulting in a larger effective tunnel barrier, and a lower current. This also explains the lower-off state current, as it was shown in Chapter 6 that hold field emission can become a major component of excess current. In Figure 8.2c, the triangular barrier from the GaSb valence band to the InAs valence band can be seen, which becomes thinner as bias is applied. However, because of the depletion regions that form with alloying in Figure 8.2e, there is no band-bending to create a triangle barrier that can be tunneled through, only a step barrier that must be overcome in energy thermally. As bias is applied, the GaSb valence band will eventually bend to allow for a triangle barrier to tunnel through, but it will be much larger, and will be much more difficult for holes to tunnel through.

8.3 GaAsSb Graded Buffer Devices

The GaAsSb buffer, described in the previous chapter, is imperfect, but likely is providing a lower defect density, as many of the misfit segments are being recycled to relieve strain. Figure 8.3 a and b show the I-V curve and G-V curve for a device of composition $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ on a GaAsSb buffer. For comparison, the electrical results of the same junction without the buffer is shown.

The conductance swing becomes considerably steeper for the buffer device, consistent with what has been seen previously where lower defect densities lead to steeper conductance slopes. Furthermore, the swing has moved left on the curve: the conductance drops more than a decade in reverse bias and continues to drop more at low forward bias, compared to the structure without a buffer, which drops mainly in forward bias. This is indicative of a net shift in band alignment. It was already hypothesized in the previous chapter that the band-bending due to defects may be causing a net shift of band alignment towards type-III. The fact that a less

defective structure is closer to type-II band alignment is consistent with that. This topic will be revisited once again when assessing structures that do not show any defects in cross-section TEM.

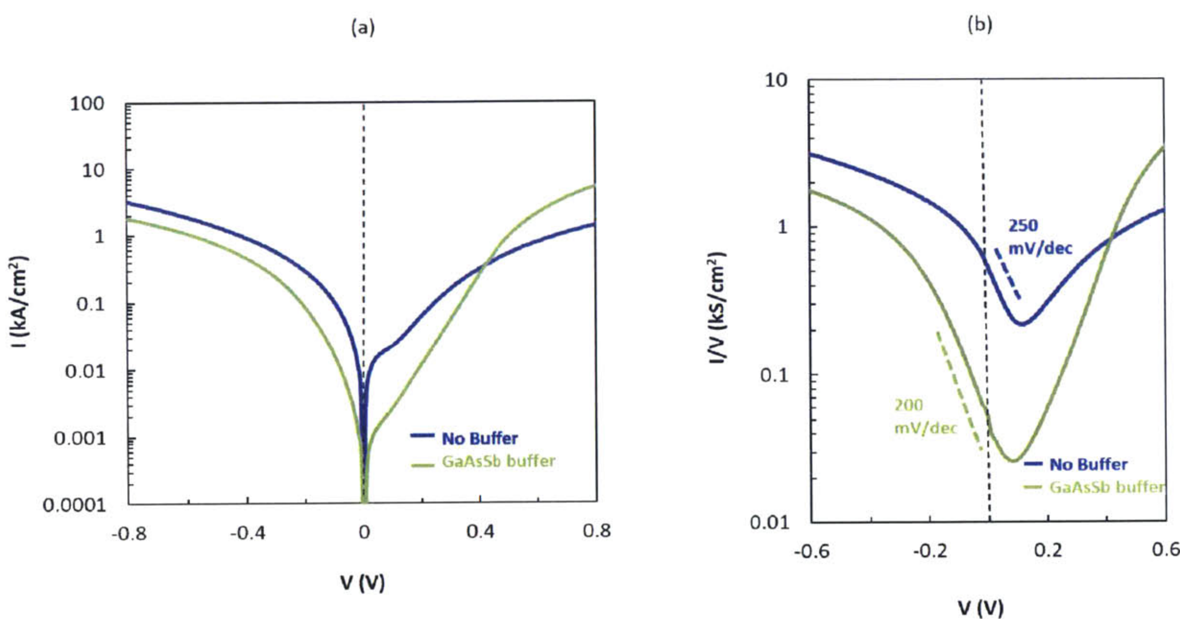


Figure 8.3: (a) I-V curve of an $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ with no buffer (blue curve) and with a GaAsSb buffer (green curve). (b) corresponding I/V-V curves

There is also a much lower off-current in the device with a buffer, causing the overall difference in on-state and off-state current to be much larger. This is an indication that the defects in the buffer-less sample are likely contributing to the current. It was shown in Chapter 6 that the binary structures likely had a current floor given by trap-assisted tunneling. It is possible that this effect still dominates these alloyed devices.

This buffer device is unique compared to the other ternary devices thus far, in that it contains features in the epitaxial stack that are present solely for structural reasons and not for

electrical reasons: the buffer itself. Hence, it becomes questionable whether a top and bottom contact setup can still be used, given that current now has to transport through the buffer, which may be resistive, or add some form of unwanted curvature to the total I-V curve. Figure 8.4a shows the schematic of a top-bottom contact structure, which is the structure used for the data just presented. The GaAsSb device layer is etched only 20 nm. This allows for as much room as possible for current to spread before reaching the buffer, making the effective cross-sectional area of the buffer much larger than the junction area, allowing the opportunity for the tunnel junction to dominate the I-V characteristic. In order to test if the buffer is affecting the I-V curve, a second structure was utilized, shown in Figure 8.4b. This structure uses a self-aligned side-contact. Figure 8.4c demonstrates how such a structure is fabricated: The fabrication takes advantage of the undercut that occurs during wet-etching, for the Sb-containing layers, as was described in section 5.2.2. Electron beam evaporation of a Au/Ti metal stack, which deposits as line-of-sight, results in a self-aligned side-contact, where the top contact metal masks the deposition, resulting in the side contact being extremely close to the mesa, minimizing any series resistance. Figure 8.4d shows an SEM image of an example devices fabricated this way.

A comparison of the effect of contacting from the bottom versus the side is shown in Figure 8.5. It can be seen that the only effect of the buffer is that it adds some series resistance, but only enough that it is observable at higher currents. In the range of conductance-slope switching, the measured conductance slope is the same for both devices. Hence, conductance-slopes can be confidently extracted raw from top-bottom contact structures without having to correct for any series resistance. This side-contact scheme was used to test all buffer devices, including the InAsP buffers that will be discussed shortly, to ensure that no devices are being affected parasitically by the buffer layers below.

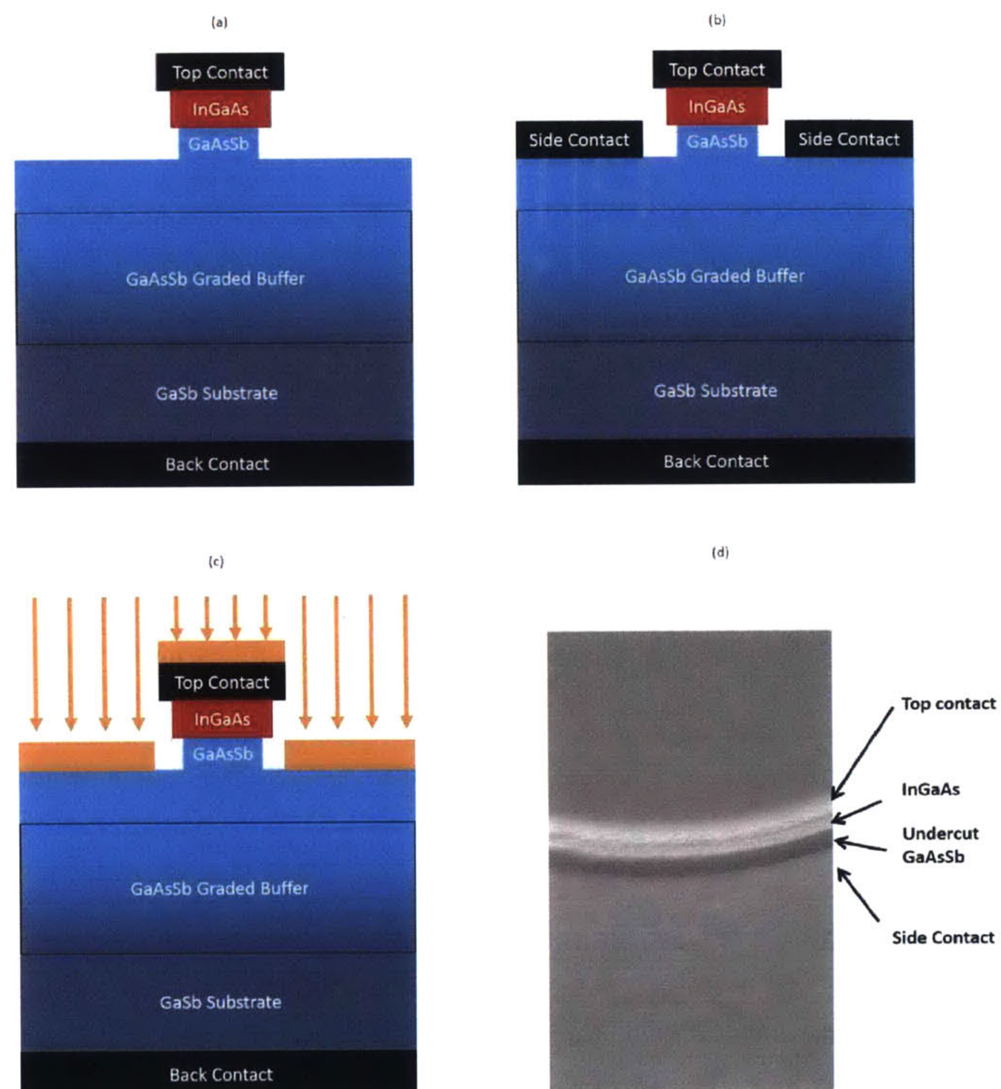


Figure 8.4 (a) Schematic of a top contact + back contact configuration for a device with a buffer. (b) Schematic for a top contact + circular side contact configuration. (c) Illustration showing a line-of-sight metal deposition where the top contact and undercut of the semiconductor layers below results in a self-aligned side contact. (d) SEM image showing the resulting device with the self-aligned side contact

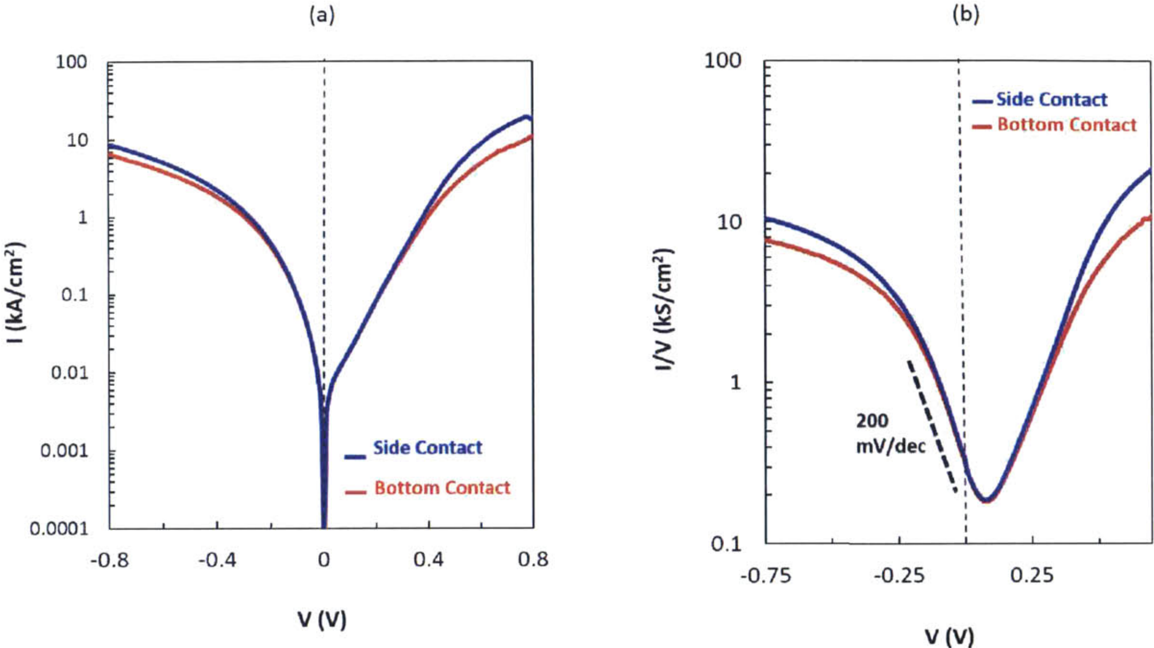


Figure 8.5 I-V curve for $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ on GaAsSb buffer with a bottom contact (red curve) and a self-aligned side contact (blue curve). (b) corresponding I/V-V curve

8.4 Temperature Dependence of GaAsSb Buffer Device

The temperature dependence of the GaAsSb buffer device was measured. This was done for a few reasons:

- While type-III band alignment devices typically show negative differential resistance, it disappears as the band alignment shifts towards type-II and becomes type-II. Without negative differential resistance, it is difficult to prove that the electrical characteristics are due to tunneling, and not simply a thermal mechanism, which would appear almost the same electrically. The easiest way to distinguish is to examine the temperature dependence since any thermal mechanism would have an Arrhenius dependence. Before proceeding any further

with additional devices, it is important to determine this way that we are indeed still utilizing interband tunneling.

- It is desirable to test the conclusions of Chapter 6 at a different band alignment, to see if they are true for various band alignments, or specific to type-III. Most notably, we want to test at this band alignment the conclusion that there is no temperature dependence of conductance slope.

The device was measured from room temperature down to 77 K. The I-V and I/V-V curves are shown in Figure 8.6.

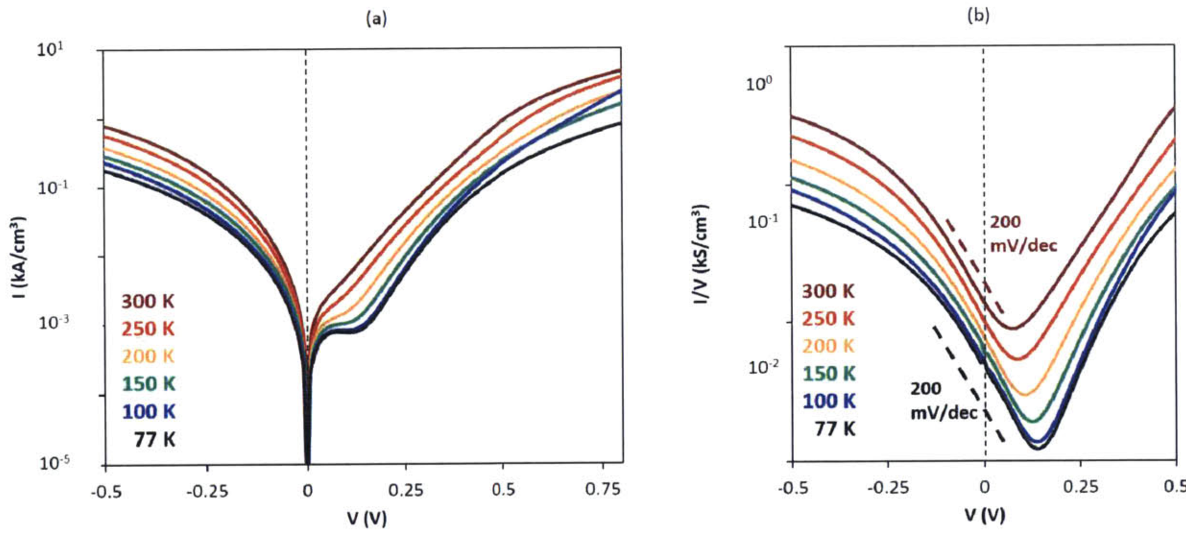


Figure 8.6: I-V curve of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ on GaAsSb buffer for temperatures varying from 300 K down to 77 K. (b) corresponding I/V-V curves

As the temperature is decreased, a negative differential resistance begins to develop again, confirming tunneling. Consistent with this emergence of NDR, a shift in the I/V-V curve towards type-III band alignment is observed. This is likely due to shifts in band-alignment with temperature and strain brought on by lattice expansion and the deformation potentials. The off-

current decreases with temperature as did the binary materials in Chapter 6, but it is still not exponential, indicating that it may still be a trap-assisted process that gives the off-current.

Most importantly, the trend of conductance slope remains the same: There is still no temperature dependence. Hence, the main conclusions of Chapter 6 can be expanded to this material system with a more type-II band alignment, indicating that this lack of temperature-dependence is likely not band-alignment specific.

8.5 InAsP Buffer Devices

As was shown in the previous chapter, utilizing an InAsP buffer with the correct switching sequence results in high quality layers with no observable crystal defects in cross-section TEM. Based on the results of this thesis so far, these devices are expected to provide the highest conductance slope of all the ternary devices thus far, since it is known that materials defects lower steepness. Indeed, this turns out to be the case. Figure 8.7 compares an $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ device on the InAsP buffer to the other two platforms. There is a pronounced increase in conductance slope to 100 mV/decade, the best so far in this thesis. Additionally, consistent with the trend that was seen before, there is an additional shift of the location of the swing to the left, into reverse bias. In fact, the structure is now behaving with entirely type-II band alignment, consistent with the band diagram that was calculated in Figure 8.2e: The entire swing is in reverse bias, with only the regular p-n junction diode swing in forward bias.

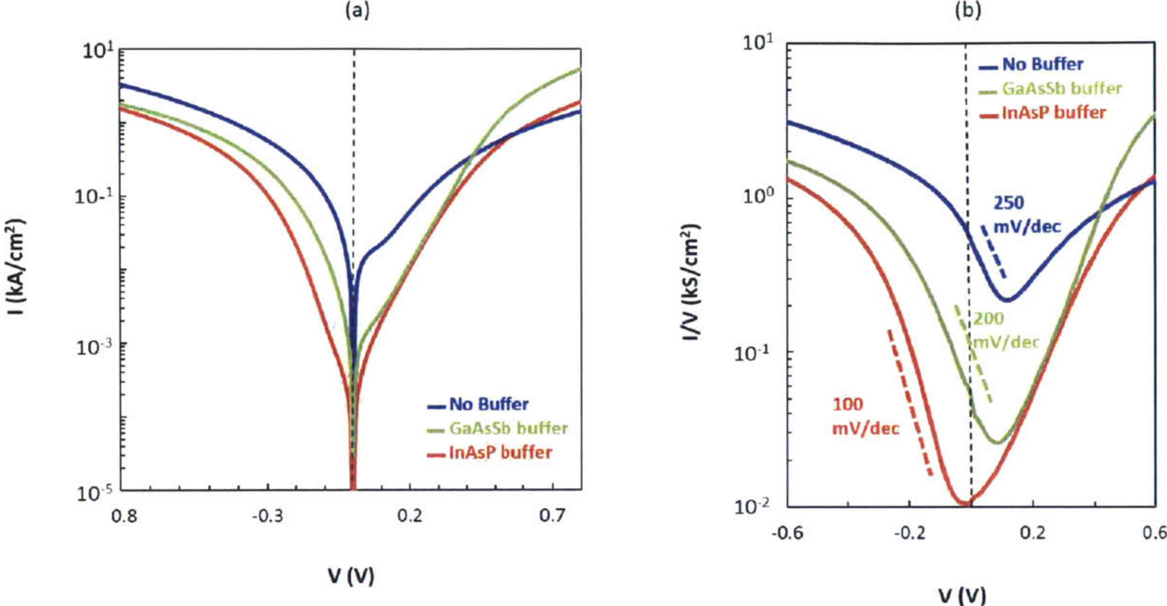


Figure 8.7: (a) I-V curves of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction on no buffer (blue curve), a GaAsSb buffer (green curve), and an InAsP buffer (red curve). (b) corresponding I/V-V curves

This improvement is promising. It is possible that this device is simply the lowest defect density to date, giving it the best possible performance. However, it was established in section in Figure 7.13 in the previous chapter that it actually has a degree of phase separation, indicating that it may not be ideal. In such a case, why is it so much steeper? One possibility is that it is because it is very close to the border of type-II/type-III band alignment, and perhaps right at the border the slope becomes the steepest, since the smallest possible amount of voltage is required to toggle it on and off, leading to the smallest possible voltage-induced deformation of the bands.

To test this, this same InAsP buffer platform was used to alloy the device further towards type-III, at the $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ composition to bring it as close as possible to the border of type-II/type-III band alignment. There are two reasons to believe such a device may be better:

1. If the steepness benefits from being on the border of type-II/type-III, then this composition would be better as it should be closer to the border
2. It is less likely to have phase separation, as it has been alloyed to a composition where phase separation is no longer favorable

The result in alloying further towards type-III is shown in Figure 8.8a. There is a remarkable increase in conductance slope that is almost equal to the thermal limit – 76 mV/decade. For comparison, the device is compared to the same composition without a buffer in Figure 8.9b, where an enormous difference in conductance slope is seen and the same shift of the swing to the left.

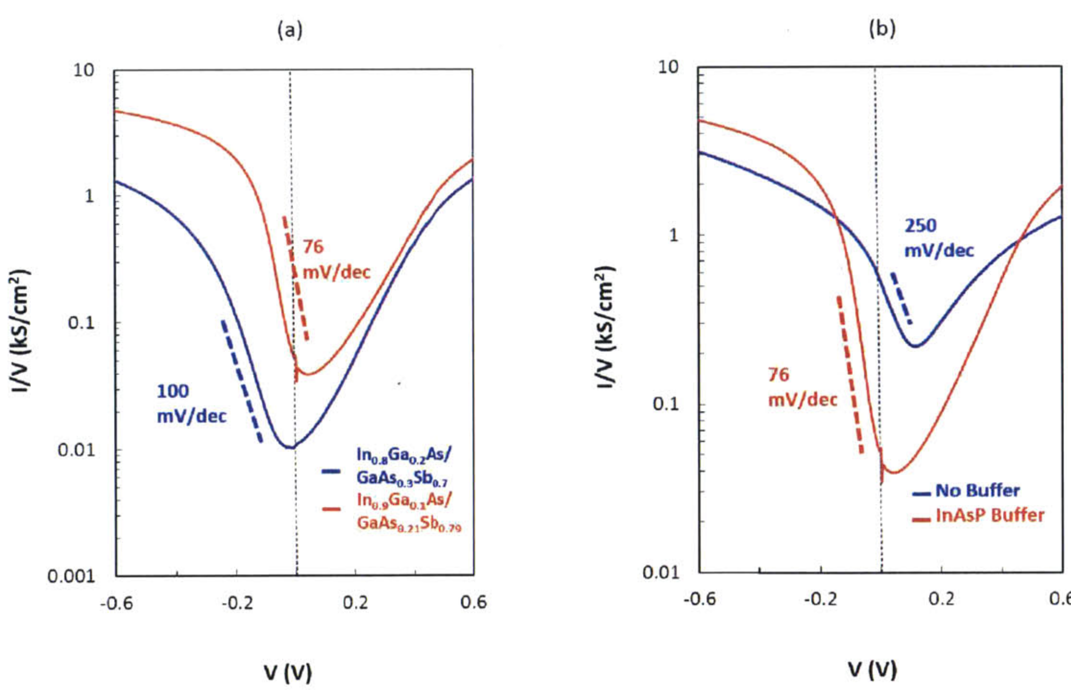


Figure 8.8: (a) I/V-V curve of $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ heterojunction (red) and $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction (blue) on an InAsP/InP platform. (b) I/V-V curve of $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ heterojunction with an InAsP buffer (red) and without a buffer (blue)

The steepness of this device appears to be the steepest seen to date, and is a good indication that with the correct control of growth conditions and materials parameters, it is possible to obtain a steepness that approaches the thermal limit. This result is encouraging, as it indicates that with further optimization of some of the finer parameters of the growth, it may be possible to eventually observe steeper than 60 mV/decade switching. In fact, at 76 mV/decade, such a heterojunction already has potential for switching at comparable subthreshold slope and drive currents to MOSFETs, but without the temperature dependence, making it superior in terms of its ability to withstand fluctuations in temperature, which may be useful in specialty applications.

A steepness this close to the thermal limit may raise questions about whether or not this is simply a thermal barrier that is interfering with the result. However, the results of the next chapter will show that that is not possible, and that this must be due to tunneling.

Chapter 9: InGaAs/GaAsSb Heterojunctions for Millimeter Wave Detection

9.1 Introduction

The previous chapter demonstrated the ability to control the band-alignment to shift the region of conductance-swing. Further, through an optimized materials structure, conductance slopes approaching the thermal limit were obtained. Hence, the applications for transistors were extracted, as the conductance slope is simply the subthreshold slope that would be expected in the absence of three-terminal parasitics. It remains to be seen, however, what the curvature coefficient of these devices are, and hence how useful they would be in the application of millimeter wave detection. This chapter focuses on exploring this.

9.2 Band Alignment Shifting

Starting with the same buffer-less samples that were examined in Figure 8.2 in the previous chapter, Figure 9.1a shows the curvature coefficient, γ, as a function of applied bias for $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$, and $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$. A peak curvature coefficient of a little over -30 V^{-1} is obtained in forward bias. This is already a very high curvature, and quite close to the thermal limit of 38.7 V^{-1}. Of course, as described in section 1.4, it is desirable to have the peak curvature at zero-bias, as it is not desirable to introduce a biasing network, which results in higher noise. Therefore, the peak curvature needs to be shifted. It can be seen that this was somewhat accomplished in alloying from $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ to $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$. However, the peak remains on the right side of the origin.

In theory, the peak should hit the origin right at the border of type-II/type-III band alignment, as that is where the curvature at the origin would be highest. This is because a reverse bias would result in tunneling since the bands overlap, while a forward bias would result in little current, since the bands are not aligned. This would give much curvature to the I-V curve. For

this reason, another way to characterize a good square-law detector is by how well it rectifies current (allowing it only in reverse bias). Figure 9.1b shows the I-V curve of both devices at low voltage. Indeed, the $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ device, with the band-alignment being less type-III, and the maximum curvature shifted closer to the origin, does a better job of rectifying the forward current while allowing reverse current.

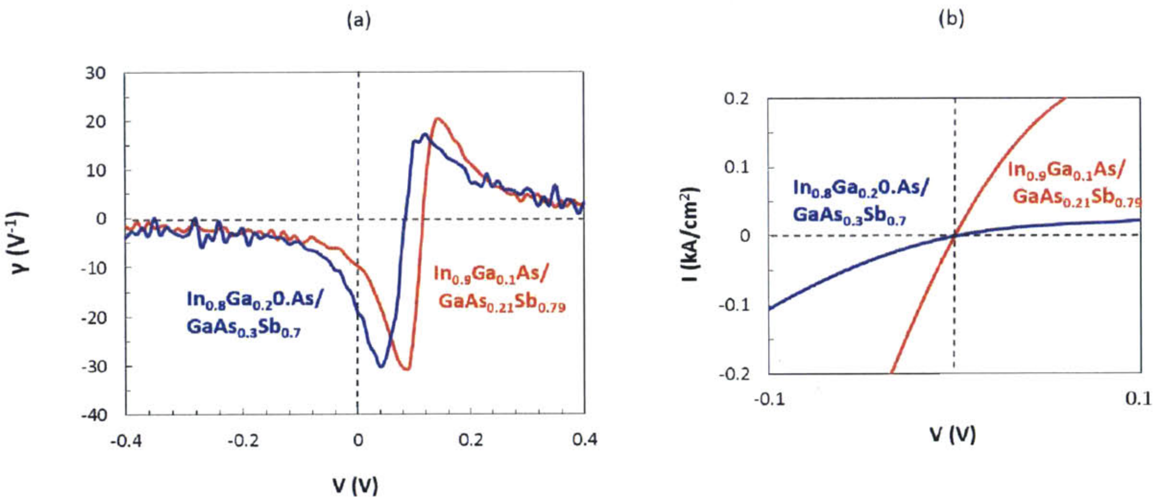


Figure 9.1: (a) Curvature coefficient as a function of applied voltage for $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$, (blue) and $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ (red), both grown on GaAsSb layers with no buffer. (b) I-V curve of the two devices at low voltage, showing how the $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ does a much better job at accomplishing rectification of the forward current

Once again, it seems that the peak curvature coefficient is further to the right than expected. This is no doubt the same phenomenon as seen in previous chapters, where the more defective structures result in a shift to toward type-III band alignment.

9.3 Thermal Curvature Limit vs. Thermal Conductance Slope Limit

It may seem strange that these heterojunctions, which gave conductance slopes of 250-300 mV/decade - very far from the thermal limit of **conductance slope** - are already quite close to the thermal limit of **curvature**. This seems to imply that it may be easier to beat the thermal curvature limit than it is to beat the thermal subthreshold slope limit. Upon close inspection, this is indeed the case. This may seem counter-intuitive, since both limits result from the same underlying physical mechanism. However, part of the reason conductance slope was chosen as the main fundamental figure of merit of band-edge steepness for this thesis was because the curvature coefficient is not an exclusive probe of the band-edge steepness, it is an application-specific figure of merit that is actually convoluted with something else. An explanation of why begins with the origin of the thermal limit from the ideal diode equation, and taking the second derivative of it to obtain the curvature:

$$I = I_0\left(e^{qV_a/kT} - 1\right) \tag{1}$$

$$\frac{d^2I}{dV^2} = I_0\frac{q^2}{(kT)^2}e^{qV_a/kT} \tag{2}$$

This second derivative term is what ultimately gives the thermal limit, and the curvature coefficient is obtained by diving by the first derivative ($I_0\frac{q}{(kT)}e^{qV_a/kT}$) to give:

$$\gamma \le \frac{q}{kT} = 38.7\ V^{-1} \tag{3}$$

For the case of a tunnel device, the equation for tunnel current is shown again:

$$I = C_1\int^{\Delta E}[F_C(E) - F_V(E)]T_tN_C(E)N_V(E)\boldsymbol{dE} \tag{4}$$

For the case of trying to probe the band-edge steepness the $[F_C(E) - F_V(E)]$ was removed by diving by V. However, since we are now taking a device figure of merit, this term remains, as it is part of the current. This is the key difference: the curvature coefficient is also affected by this term, gets "help" from it in that differentiating with the V term adds curvature. This result can be best seen by recasting equation (4) in terms of conductance slope (CS), assuming we are in a region of constant conductance slope:

$$I = C_2 V e^{V \cdot ln(10) / CS} \tag{5}$$

The key difference between this expression and the ideal diode equation (1) is the presence of the V term in front of the exponential. Calculating a curvature coefficient from this expression gives:

$$\gamma = \frac{-ln10}{CS} \times \frac{2 - V\frac{ln(10)}{CS}}{1 - V\frac{ln(10)}{CS}} \tag{6}$$

This is a very important result. The curvature coefficient is the product of two terms. The first, -ln(10)/CS, depends only on the conductance slope and if the total expression consisted only of this term, then there would be a simple one-to-one conversion between the conductance slope and the curvature coefficient. Plugging in a value of 60 mV/decade (thermal limit) into the first term, a curvature coefficient of 38.7 V^{-1} is obtained (thermal curvature coefficient). Hence, if the expression consisted of only this term, then the two thermal limits would be equivalent. However, resulting from the V term in front of the exponential in equation (5), a second term results that depends on both CS and V. This term is dimensionless and can be thought of as providing an enhancement factor to the curvature coefficient. Figure 9.2 shows the value of this term as a function of voltage, for the arbitrary case of a 100 mV/decade conductance slope.

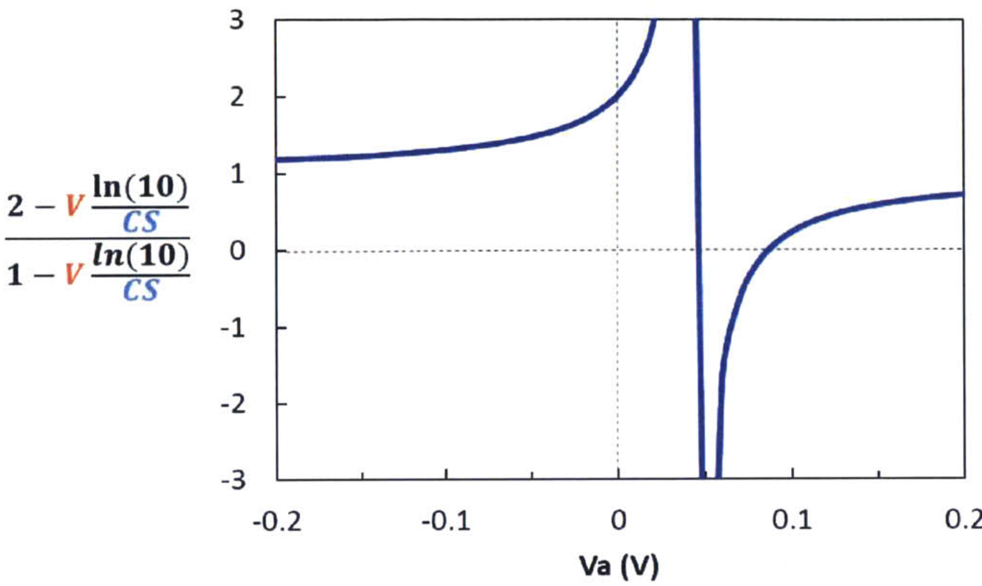


Figure 9.2: "Curvature Enhancement Factor" – the second term of equation (6) – plotted as a function of voltage for the arbitrarily chosen case of a 100 mV/decade conductance slope

At both high forward and high reverse bias, the term converges to one, and hence there is a simple one-to-one relationship in these regions between curvature coefficient and conductance slope, where γ=- ln10/CS and the thermal limits of both curvature coefficient and conductance slope are equivalent. However, as the origin is approached from reverse bias, the term begins to increase, reaching a value of exactly two at the origin. This is an important result, meaning that at the origin, there is a factor of two improvement that ultimately results from the extra curvature gained from the V term, as was described in the explanation following equation (4). This means that, for zero-bias curvature, it is not necessary for the band-edges to be steeper than the thermal limit, they must only be steeper than half the steepness of the thermal limit. In other words, plugging in a value of 38.7 V^{-1} to equation (6) returns a conductance slope of 119 mV/decade: It is only necessary to beat 119 mV/decade conductance slope steepness (which was already

demonstrated in the previous chapter), in order to beat the thermal limit of curvature. Overall, this result indicates that high-curvature two-terminal devices and their applications have a greater chance of succeeding utilizing tunneling near the band-edges than does the application area of transistors.

It is worth noting that as the enhancement factor moves into forward bias, at some point it diverges. This is due to negative resistance, as can be seen by solving for the condition of divergence of the enhancement factor (setting the denominator to zero):

$$1 - \frac{Vln(10)}{CS} = 0 \tag{7}$$

$$\frac{1}{Vln(10)} = \frac{1}{CS} = \frac{dlog(G)}{dV} = \frac{1}{Gln(10)}\frac{dG}{dV} \tag{8}$$

$$V\left|\frac{dG}{dV}\right| = G \tag{9}$$

which is the exact condition for negative resistance in terms of conductance, as was originally derived as equation (2) in Chapter 5. While it may be tempting to design a device such that it is biased to this NDR point to obtain maximal curvature, doing so would not be ideal, as we ideally want a device operating at zero-bias. Therefore, a major goal is to control the band alignment via alloying such that the curvature coefficient is maximal at zero-bias, and the enhancement factor of 2 can be utilized.

9.4 Buffer Devices

Figure 9.3a shows the curvature coefficient for an $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ with no buffer, a GaAsSb buffer, and an InAsP buffer. Figure 9.3b shows the conductance slope data: the same figure as Figure 8.7b from the previous chapter, so that a comparison can be made.

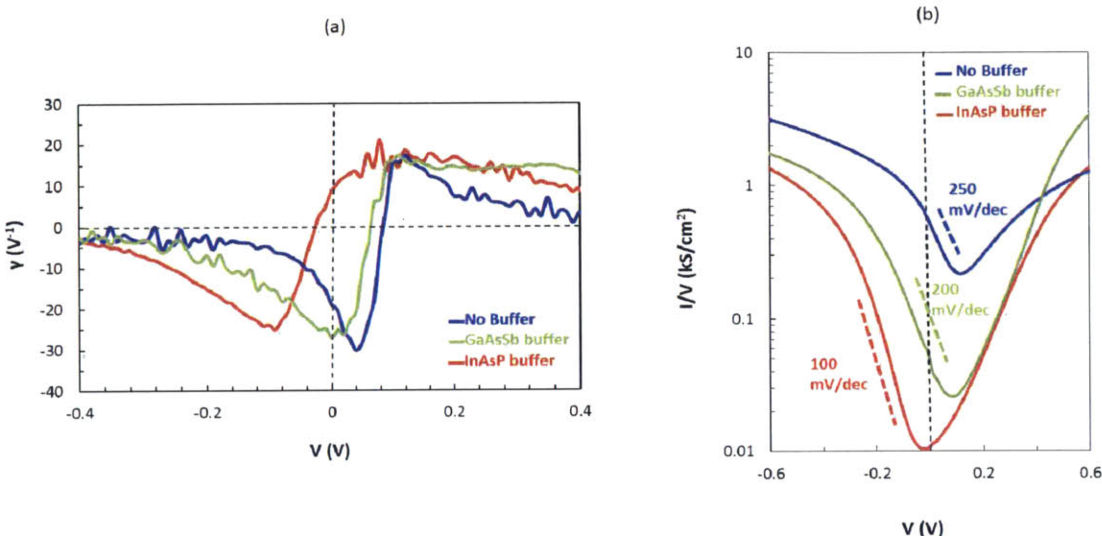


Figure 9.3 (a) Curvature coefficient of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction device with no buffer (blue), GaAsSb buffer (green), and InAsP buffer (red). (b) Corresponding I/V-V curves for the sake of comparison

What is the same is that there is a general shift to the left as defect density lowers. However, an important difference is that, while the conductance slope became steeper for the less defective samples, the curvature coefficient did not. This is quite interesting, as it preliminarily implies the curvature coefficient might not depend on defects, or at least may not depend as much. However, the shift of the curve to the left affects the curvature coefficient. This is once again due to the enhancement factor described in the previous section. In referring back to Figure 9.2, it can be seen that near the origin, the enhancement factor decreases as voltage is shifted from forward bias to the left towards reverse bias and continues decreasing past the origin into reverse bias. Hence, as the band alignment shifts to the left, the peak curvature coefficient shifts to the left and is modulated by a smaller enhancement factor, causing a lower peak curvature coefficient.

Given that, it is likely that the InAsP buffer sample shown in the red curve actually has the potential of having a much steeper curvature, but is alloyed too far towards type-II. Figure 9.4a compares the result when the composition is tweaked, and the device consists of an $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ heterojunction on InAsP.

As can be seen, with this tweak in composition, the peak has shifted to almost zero-bias, and the peak curvature coefficient has grown dramatically and has **exceeded the thermal limit,** with a peak value of 43 V^{-1}. In fact, given that the peak is still slightly to the left of the origin, then for the case of a peak right at the origin (slightly tweaked composition), the full factor of 2 enhancement factor would provide a curvature coefficient of 60 V^{-1}. Hence, it is very likely that this platform is capable of 60 V^{-1} curvature, making it the best heterojunction platform.

Figure 9.4b also shows its much better rectification. Hence, a demonstration of a platform has been accomplished where, without doping, the curvature coefficient can exceed the thermal limit and can be brought effectively to zero bias. This is very promising for millimeter wave detection. Further, from a technological perspective, this was accomplished on an InP platform, paving the way for heterointegration of such devices with other analog and digital components.

This also confirms that the curvature coefficient does depend on materials defects and inhomogeneity, in a manner similar to the conductance slope, but that this dependence is convoluted with band alignment in that a shift to type-II lowers the peak curvature coefficient.

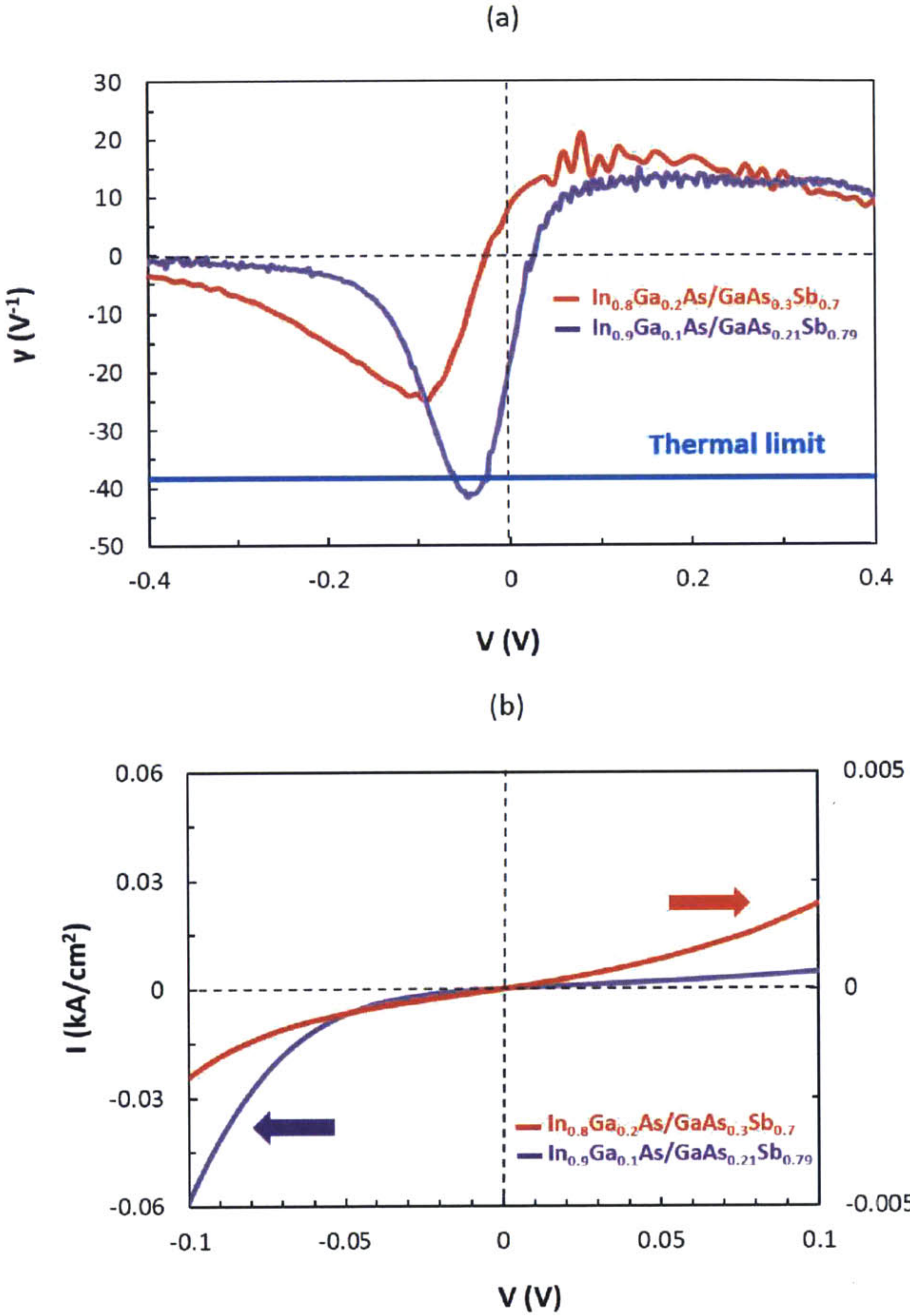


Figure 9.4: (a) curvature coefficient of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ (red) and $In_{0.9}Ga_{0.1}As/GaAs_{0.21}Sb_{0.79}$ (purple) on InAsP buffer, showing demonstration of beyond the thermal limit of curvature coefficient near the origin, (b) I-V curve of the two devices at low voltage

9.5 Temperature Dependence

The $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction device on a GaAsSb buffer was measured at temperatures from 300 K down to 77 K. The result of curvature coefficient is shown in Figure 9.5a. The dependence is much more complicated than for conductance slope, which had no temperature dependence in the region of switching. Again, the reason is simply because of the shift in band alignment that occurs for this heterojunction as temperature changes, as was first observed in section 8.4 of the previous chapter, and the fact that the curvature coefficient depends on band alignment. What can be seen, then, is a general shift to the right, and a corresponding increase in peak curvature as the device is cooled and the band alignment shifts towards type-III. Eventually, the type-III band alignment becomes so pronounced that NDR occurs at a large enough magnitude to cause oscillations to occur from bistability. This can be seen as the multiple large peaks for the 100 K and 77 K samples, analogous to the peaks shown in Figure 5.7, when the curvature coefficient was used to identify the boundaries of the instability region.

Still, even with this change due to band-alignment, the change in curvature coefficient at zero bias is quite minor. Figure 9.5b shows the curvature coefficient at zero bias as a function of temperature for this device, normalized by the value at 300 K. When comparing to the thermal trend, it can be seen that this device is indeed quite stable to temperature fluctuations, making it very useful for field environments where temperature will fluctuate.

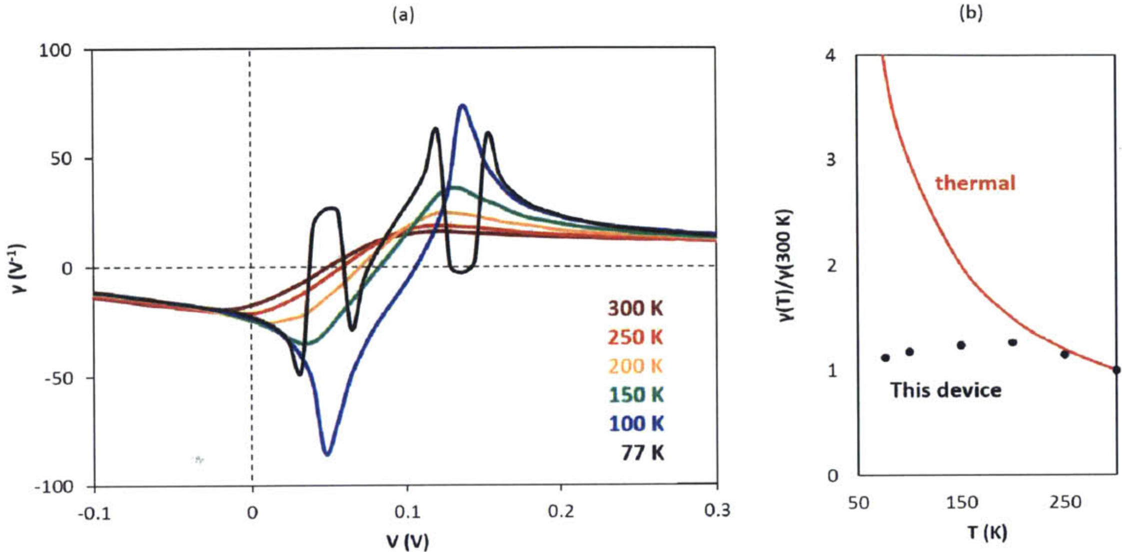


Figure 9.5: (a) Curvature coefficient of $In_{0.8}Ga_{0.2}As/GaAs_{0.3}Sb_{0.7}$ heterojunction device on a GaAsSb buffer as a function of temperature. (b) Curvature coefficient at zero-bias for this device normalized by the value at 300 K, and compared to the thermal trend

Chapter 10: Conclusions and Future Work

10.1 Summary of Results

The primary goal of this thesis was to understand what the true steepness of band-edges are for a device in which quantum tunneling is used to obtain a sharp change in conductance with voltage. This was accomplished by first establishing a figure of merit and developing a special device that can be used to extract it.

Beginning with the simple case of type-II binary InAs/GaSb heterojunctions, it was found that they are prone to defects due to the intermixing that occurs between layers and the resulting strain the builds up as a result. A variety of techniques were established to control defect density First, the growth order was changed so that the antimonide layer was grown on top, preventing a driving force for As-Sb swap and Ga carryover. The result was an interface that had no observable defects in cross-section TEM. Other techniques included straining the GaSb underlayer to create a pseudo-lattice match, and growing the InAs layer at lower temperatures to prevent exchange, which incorporates more point defects or impurities, but which can later be annealed out to restore the quality.

Using the growth techniques for these binary heterojunctions, it was possible to obtain a set of heterojunctions at varying degrees of defects and inhomogeneity. This allowed for conductance slopes to be extracted for the first time as a function of material quality. It was found that defects indeed lead to a less steep conductance-slope, and specifically misfit dislocations are the most dominant defect. It was postulated that the gettering of point defects and nonuniform composition related to such dislocations likely cause the formation of states and nonuniformity, and consistent with this, it was found that annealing improves the conductance slope. Further, since point defects and inhomogeneity are expected to be present in all

heterojunctions, even those without a high dislocation density, it was found that all devices respond positively to annealing.

Low temperature measurements were then carried out to answer two important questions: 1. Are these optimized structures still limited by defects? And 2. Given that all published three-terminal results show a temperature dependence, will this be seen in two-terminals in the absence of three-terminal parasitics? Indeed, there was no temperature dependence, answering these two questions. 1. The structures appear to still be limited by defects, and 2. The published TFET devices are almost certainly dominated by a thermally activated three-terminal parasitic. A model was described where current may be tunneling through oxide trap states and then generating into the conduction band. It was also shown that a series MOSFET could also be causing this effect. The observed absence of a temperature dependence in two-terminals is consistent with both of these possible explanations.

The InAs/GaSb material system was then alloyed: Ga alloyed into InAs to move the conduction band up in energy, and As into GaSb to move the valence band down. The result was InGaAs/GaAsSb heterojunctions at varying compositions which were demonstrated electrically to have varying band alignments including type-II and type-III. Buffers were utilized to obtain the best quality, and it was found that the best results were obtained on an InAsP buffer on an InP substrate: a very commercially friendly platform for cost and future heterointegration. The main device layer growth challenge was found to be intralayer As-Sb exchange, analogous to the As-Sb exchange in the InAs/GaSb heterojunctions, except this time confined to the GaAsSb layer alone. The solution was found to be a 30 s H2 bake + 30 s TMSb flow before GaAsSb initiation, designed to put an Sb layer on the surface that would prevent exchange. High quality structures were obtained this way.

Using optimized growth conditions, including the GaAsP buffer, the Sb gas switching sequence, and a tailored composition of InGaAs/GaAsSb, a record conductance slope of 76 mV/decade was obtained. Curvature coefficient measurements proved that this must be from tunneling. A slope this close to the therml limit is a strong indication that with further tweaking and optimization, the potential may be present to obtain band-edges with steepness greater than the thermal limit of 60 mV/decade. However, this thesis has shown that doing so is a great challenge and that materials defects always play an important role that limit steepness. Nevertheless, the potential already exists with these results for logic devices competitive with MOSFETs that can operate reliably within a much larger temperature range, which could find novel applications.

It was derived from theory that it is not necessary to beat the thermal limit of conductance slope in order to beat the thermal limit of curvature, due to the fact that the curvature limit comes in two terminals. Based on this, and using these same optimized growth conditions as before, an example millimeter wave detector structure was produced, giving a curvature coefficient that exceeded the thermal limit, with a peak curvature of 43 V^{-1}, near zero bias, and demonstrated to be controllable in its position on the voltage axis by alloying. It was also shown that such a device is considerably resistant to changes in temperature, making it very promising for applications in passive millimeter wave imaging.

Overall, this thesis established the important principles that affect the steepness of the band edges for tunneling devices, determined how they can be controlled, and optimized them to give a demonstration of a competitive transistor technology and millimeter wave detector.

10.2 Suggestions for Future Work

This thesis identified two major challenges. The first is the difficulty in obtaining steepness of the band edges due to materials defects. The second is the fact that three-terminal device are plagued by thermally-activated parasitics that dominate the transfer characteristic. For the sake of the TFET, this second problem must be solved before the field can advance. For the sake of other applications, this second problem does not need to be solved, and the first problem can be focused on. Regardless, the first problem is of more fundamental nature, and must be applied to all possible applications of utilizing tunneling at the band-edges for steepness. Below are a collection of possible future directions for research in this area.

10.2.1 Vigorous Optimization of Growth and Processing Conditions

The first obvious approach would be to attempt to further optimize the heterojunctions to obtain the correct composition and the lowest possible defect levels. This will likely be a tedious process, as there will always be some concentration of defects that are thermodynamically favorable and unavoidable, and once this level is reached, the challenge becomes more that of quality control and reproducibility, which results in a challenge more appropriate for industry. Nevertheless, there are still techniques that could be considered for improvement, including an extensive study on annealing to determine how far one can go before activating degradation mechanisms. Most likely, every material system, and every composition within each system may or may not have an annealing temperature window within which defect and composition averaging mechanisms improve performance, with an upper limit bound by where some other mechanism becomes activated that degrades performance. Tedious studies could identify these windows to push performance. It may also be possible to kinetically limit the thermodynamically

favorable defect formation processes using low temperature techniques, such as atomic layer deposition.

10.2.2 Utilizing Materials Less Electrically Affected by Defects

The interface is very susceptible to interface states, strain, and nonuniform composition. One potential alternative is to use layered chalcogenides. Because these materials are two-dimensional, their interface has no dangling bonds. Hence, in forming a heterojunction, there is no need for a perfect interface to be formed that passivates dangling bonds. Therefore, epitaxy is not required, since they can be layer transferred. This prevents intermixing which comes from high temperatures. There is also no strain, since the two layers are both totally relaxed. This usually comes at the expense of an incoherent interface, but in this case, interface coherence is less important because there are no dangling bonds. Work in this area has already begun.[105], [106] However, there are still many challenges, as controlled processing on these structures is difficult, It also remains to be seen how the layers can be made to be perfectly flat without ripples, how the band-edge steepness will be affected by the incoherent interface with regards to the intermolecular forces, and whether or not such a heterojunction could be reliable controlled by a gate voltage that can change the bands of the structure.

10.2.3 Utilizing Smaller Structures

Another school of thought is to make the device very small, the hope being that the length scale of the device becomes smaller than the length scale of the inhomogeneity. If this were to succeed, it would most likely result in large device-to-device variability, so the challenge would then become to find ways to either control that variability, isolate devices that fall outside of an accepted range, or adapt circuit design to embrace this variability.

As the device dimensions become very small, eventually quantum confinement will occur. It is possible that eventually, the confinement will lift the sub-band edges above the true band-edges, and may move them to an energy region where there are fewer defect states, which may give an improvement. At this point, however, the band edges would take on a shape that reflects the degree-of-perfection of the nanostructure, and variations in its diameter or surface quality would affect the conductance slope. Hence, this would most likely shift the band-edge steepness into a regime where it is no longer limited by interface defects but instead by defects and/or dimensional variations in the nanostructure.

10.2.5 New TFET Designs

Finally, with regards to the parasitic problem specific to TFETs, future work clearly needs to focus on alleviating this problem. If the trap assisted oxide leakage pathway is in fact dominating, then the structures needs to be designed such that there is a feature (such as a special barrier) that suppresses the tunneling. Or the device needs to operate in a completely different electrostatic regime where this effect cannot occur.

Appendix A: Device Fabrication Process

This is the fabrication process submitted to the MIT Microsystems Technology Laboratory, to be done in the Technology Research Laboratory. All machine names are included.

	Description	Device Processing	Lab	Machine
1	Degrease/clean	Acetone/IPA/Water/N2 blow dry	TRL	Photo-wet
2	Photolithography mask 1:	HMDS Spin-coat OCG at 3krpm for 30s Soft-bake 90C 5 min Expose PR 9 s low-vac Develop in OCG 934 3:2 for 5 min. Hard bake 120C for 5 min	TRL	HMDS Coater Hotplate (coater) MA-6 Photo-wet Hotplate 1 or 2
3	Mesa Etch	2g:2mL:1mL citric acid:H2O:H2O2 5 min 1:5 NH4OH:H2O 2 min 2g:2mL:1mL citric acid:H2O:H2O2 5 min	TRL	Acidhood
4	Strip Resist	Acetone in sonicator 20 min Asher 30 min	TRL	Photo-wet Asher
	Inspection	Dektak Profilometer SEM	TRL ICL	Dektak semZeiss
5	Clean top side	1:1 HCl:H2O, 30s	TRL	Acidhood
6	SiO2 blanket deposition	250-500 nm SiO2	TRL	STSCVD
7	Degrease/clean	Acetone/IPA/Water/N2 blow dry	TRL	Photo-wet
8	Photolithography mask 2:	HMDS Spin-coat OCG at 3krpm for 30s Soft-bake 90C 5 min Expose PR 9 s low-vac Develop in OCG 934 3:2 for 5 min. Hard bake 120C for 5 min	TRL	HMDS Coater Hotplate (coater) MA-6 Photo-wet Hotplate 1 or 2
9	Etch contact window	7:1 BOE, 5 min 2g:2mL:1mL citric acid:H2O:H2O2 5 min	TRL	Acidhood
10	Inspection (optional)	Dektak Profilometer	TRL	Dektak
11	Deposit Top electrode	5 nm Ti/40 nm Pt/120 nm Au	TRL	ebeamFP
12	Liftoff resist and remove residue	Acetone in sonicator 20 min Asher 30 min	TRL	Photo-wet Asher
13	Inspection (optional)	Dektak Profilometer SEM	TRL ICL	Dektak semZeiss
14	Sputter deposit contact pad	20 nm Ti, 250-500 nm Au OR 20 nm Ti, 250-500 nm Al, 100 nm Au	TRL	AJA-TRL
15	Degrease/clean	Acetone/IPA/Water/N2 blow dry	TRL	Photo-wet
16	Photolithography mask 3:	HMDS AZ5214 at 4krpm 30s Soft-bake 90C 5 min Expose PR 9 s low-vac Image reversal bake 105C 2 min Flood expose 97.8 s	TRL	HMDS Coater Hotplate (coater) MA-6 Hotplate 1 or 2 MA-6

		Develop in AZ422		Photo-wet
17	Etch contact pad	Transene gold etch 5 min 1:2 NH4OH:H2O2 1 min	TRL	Acidhood
18	Strip Resist	Acetone in sonicator 20 min Asher 30 min	TRL	Photo-wet Asher
19	Inspection (optional)	Dektak Profilometer SEM	TRL ICL	Dektak semZeiss
20	Deposit resist to protect top-side	Spin OCG 4krpm 30s Soft-bake 90C 5 min	TRL	Coater Hotplate (coater)
21	Deposit bottom contact	5 nm Ti/40 nm Pt/120 nm Au	TRL	ebeamFP
22	Strip resist on top side	Acetone in sonicator 20 min Asher 30 min	TRL	Photo-wet Asher
23	RTA contacts	300C, 90s	TRL	rta35

References

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

[2] P. M. Solomon, "Device Proposals beyond Silicon CMOS," *Futur. Trends Microelectron.*, vol. 24963, pp. 127–140, 2010.

[3] K. Galatsis, A. Khitun, R. Ostroumov, K. L. Wang, W. R. Dichtel, E. Plummer, J. F. Stoddart, J. I. Zink, J. Y. Lee, and Y. H. Xie, "Alternate state variables for emerging nanoelectronic devices," *Nanotechnology, IEEE Trans.*, vol. 8, no. 1, pp. 66–75, 2009.

[4] T. N. Theis and P. M. Solomon, "It's Time to Reinvent the Transistor!," *Science (80-.).*, vol. 327, no. 5973, p. 1600, 2010.

[5] ITRS, "ITRS 2006 Roadmap, High Performance," 2006.

[6] J. G. Koomey, "The Environmental Cost of Cloud Computing: Assessing Power Use and Impacts," in *Green:Net*, 2009.

[7] "Key World Energy Statistics," International Energy Agency, 2007.

[8] U. Hölzle, "Powering a Google Search," *The Official Google Blog*. 2009.

[9] P. F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel, and W. Hansch, "Complementary tunneling transistor for low power application," *Solid. State. Electron.*, vol. 48, no. 12, pp. 2281–2286, 2004.

[10] K. K. Bhuwalka, J. Schulze, and I. Eisele, "Performance Enhancement of Vertical Tunnel Field-Effect Transistor with SiGe in the $\ delta $ p+ Layer," *Japanese J. Appl. physics. Pt. 1, Regul. Pap. short notes*, vol. 43, no. 7, pp. 4073–4078, 2004.

[11] J. Appenzeller, Y. M. Lin, J. Knoch, and P. Avouris, "Band-to-band tunneling in carbon nanotube field-effect transistors," *Phys. Rev. Lett.*, vol. 93, no. 19, p. 196805, 2004.

[12] K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer, "I-MOS: A novel semiconductor device with a subthreshold slope lower than kT/q," in *Electron Devices Meeting (IEDM)*, 2002, pp. 289–292.

[13] W. Y. Choi, J. D. Lee, and B. G. Park, "Integration process of impact-ionization metal-oxide-semiconductor devices with tunneling field-effect-transistors and metal-oxide-semiconductor field-effect transistors," *Jpn. J. Appl. Phys.*, vol. 46, p. 122, 2007.

[14] C. W. Lee, A. N. Nazarov, I. Ferain, N. D. Akhavan, R. Yan, P. Razavi, R. Yu, R. T. Doria, and J. P. Colinge, “Low subthreshold slope in junctionless multigate transistors,” *Appl. Phys. Lett.*, vol. 96, p. 102106, 2010.

[15] E. H. Toh, G. H. Wang, L. Chan, D. Weeks, M. Bauer, J. Spear, S. G. Thomas, G. Samudra, and Y. C. Yeo, “Cointegration of In Situ Doped Silicon–Carbon Source and Silicon–Carbon I-Region in P-Channel Silicon Nanowire Impact-Ionization Transistor,” *Electron Device Lett. IEEE*, vol. 29, no. 7, pp. 731–733, 2008.

[16] S. Salahuddin and S. Datta, “Use of negative capacitance to provide voltage amplification for low power nanoscale devices,” *Nano Lett.*, vol. 8, no. 2, pp. 405–410, 2008.

[17] H. C. Nathanson, W. E. Newell, R. A. Wickstrom, and J. R. Davis Jr, “The resonant gate transistor,” *Electron Devices, IEEE Trans.*, vol. 14, no. 3, pp. 117–133, 1967.

[18] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, “Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture,” in *International Symposium on Quality Electronic Design (ISQED)*, 2002, pp. 496–501.

[19] H. Kam, D. T. Lee, R. T. Howe, and T. J. King, “A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics,” in *Electron Devices Meeting (IEDM)*, 2005, pp. 463–466.

[20] A. C. Seabaugh and Q. Zhang, “Low-voltage tunnel transistors for beyond CMOS logic,” *Proc. IEEE*, vol. 98, no. 12, pp. 2095–2110, 2010.

[21] S. Agarwal, E. Yablonovitch, and C. Sciences, “Fundamental Tradeoff between Conductance and Subthreshold Swing Voltage for Barrier Thickness Modulation in Tunnel Field Effect Transistors,” UC Berkeley Technical Report # UCB/EECS-2013-247 2014.

[22] T. Baba, “ Proposal for Surface Tunnel Transistors ,” *Japanese Journal of Applied Physics*, vol. 31, no. Part 2, No. 4B. pp. L455–L457, 1992.

[23] W. M. Reddick and G. A. J. Amaratunga, “Silicon surface tunnel transistor,” *Appl. Phys. Lett.*, vol. 67, p. 494, 1995.

[24] Y. Omura, “Negative Conductance Properties in Extremely Thin Silicon-on-Insulator (SOI) Insulated-Gate pn-Junction Devices (SOI Surface Tunnel Transistors),” *JAPANESE J. Appl. Phys. PART 2 Lett.*, vol. 35, pp. 1401–1403, 1996.

[25] T. Uemura and T. Baba, “Large enhancement of interband tunneling current densities of over 105 A/cm2 in In0. 53Ga0. 47As-based surface tunnel transistors,” *Electron Device Lett. IEEE*, vol. 18, no. 5, pp. 225–227, 1997.

[26] W. Y. Choi, B. G. Park, J. D. Lee, and T. J. K. Liu, "Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec," *Electron Device Lett. IEEE*, vol. 28, no. 8, pp. 743–745, 2007.

[27] F. Mayer, C. Le Royer, J. F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si1-xGexOI and GeOI substrates on CMOS compatible Tunnel FET performance," in *Electron Devices Meeting (IEDM)*, 2008, pp. 1–5.

[28] K. Jeon, W. Y. Loh, P. Patel, C. Y. Kang, J. Oh, A. Bowonder, C. Park, C. S. Park, C. Smith, and P. Majhi, "Si tunnel transistors with a novel silicided source and 46mV/dec swing," in *Symposium on VLSI Technology*, 2010, pp. 121–122.

[29] D. Leonelli, A. Vandooren, R. Rooyackers, A. S. Verhulst, S. De Gendt, M. M. Heyns, and G. Groeseneken, "Performance enhancement in multi gate tunneling field effect transistors by scaling the fin-width," *Jpn J Appl Phys*, vol. 49, no. 4, pp. 04DC10–04DC10–5, 2010.

[30] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and < 60mV/dec subthreshold slope," in *Electron Devices Meeting (IEDM)*, 2008, pp. 1–3.

[31] E. O. Kane, "Theory of tunneling," *J. Appl. Phys.*, vol. 32, no. 1, pp. 83–91, 1961.

[32] S. Agarwal and E. Yablonovitch, "Designing a Low Voltage , High Current Tunneling Transistor Density of States Switching," in *CMOS and Beyond*, 2015, pp. 79–116.

[33] K. Taira, I. Hase, and H. Kawai, "Negative differential resistance in InAs/GaSb single-barrier heterostructures," *Electron. Lett.*, vol. 25, no. 25, pp. 1708–1709, 1989.

[34] T. Yu, J. T. Teherani, D. A. Antoniadis, and J. L. Hoyt, "Effects of substrate leakage and drain-side thermal barriers in In0.53Ga0.47As/GaAs0.5Sb0.5 quantum-well tunneling field-effect transistors," *Appl. Phys. Express*, vol. 7, p. 094201, 1882.

[35] T. Yu, S. Member, J. T. Teherani, D. A. Antoniadis, J. L. Hoyt, and A. Vertical, "In0.53Ga0.47As/GaAs0.5Sb0.5 Quantum-Well Tunnel-FETs with Tunable Backward Diode Characteristics," *Electron Device Lett. IEEE*, vol. 34, no. 12, pp. 1503–1505, 2013.

[36] B. Romanczyk, P. Thomas, D. Pawlik, S. L. Rommel, W.-Y. Loh, M. H. Wong, K. Majumdar, W.-E. Wang, and P. D. Kirsch, "Benchmarking current density in staggered gap In0.53Ga0.47As/GaAs0.5Sb0.5 heterojunction Esaki tunnel diodes," *Appl. Phys. Lett.*, vol. 102, no. 21, p. 213504, 2013.

[37] B. Romanczyk, P. Thomas, D. Pawlik, S. L. Rommel, W. Y. Loh, M. H. Wong, K. Majumdar, W. E. Wang, and P. D. Kirsch, "Benchmarking current density in staggered

gap In0. 53Ga0. 47As/GaAs0. 5Sb0. 5 heterojunction Esaki tunnel diodes," *Appl. Phys. Lett.*, vol. 102, p. 213504, 2013.

[38] Y. Lu, G. Zhou, R. Li, Q. Liu, Q. Zhang, T. Vasen, S. D. Chae, T. Kosel, M. Wistey, and H. Xing, "Performance of AlGaSb/InAs TFETs with gate electric field and tunneling direction aligned," *Electron Device Lett. IEEE*, vol. 33, no. 5, pp. 655–657, 2012.

[39] R. Li, Y. Lu, G. Zhou, Q. Liu, S. D. Chae, T. Vasen, W. S. Hwang, Q. Zhang, P. Fay, and T. Kosel, "AlGaSb/InAs Tunnel Field-Effect Transistor With On-Current of 78< formula formulatype=," *Electron Device Lett. IEEE*, vol. 33, no. 3, pp. 363–365, 2012.

[40] R. Li, Y. Lu, S. D. Chae, G. Zhou, Q. Liu, C. Chen, M. Shahriar Rahman, T. Vasen, Q. Zhang, and P. Fay, "InAs/ AlGaSb heterojunction tunnel field- effect transistor with tunnelling in- line with the gate field," *Phys. status solidi*, vol. 9, no. 2, pp. 389–392, 2012.

[41] T. Takahashi, M. Sato, Y. Nakasha, and N. Hara, "Lattice-Matched p-GaAsSb/n-InP Backward Diodes Operating at Zero Bias for Millimeter-Wave Applications," *Appl. Phys. express*, vol. 5, no. 9, p. 4201, 2012.

[42] G. Zhou, Y. Lu, R. Li, Q. Zhang, Q. Liu, T. Vasen, H. Zhu, J. Kuo, T. Kosel, M. Wistey, P. Fay, S. Member, A. Seabaugh, and H. Xing, "InGaAs / InP Tunnel FETs With a Subthreshold Swing of 93 mV / dec and I ON / I OFF Ratio Near 10 6," *IEEE Electron Device Lett.*, vol. 33, no. 6, pp. 782–784, 2012.

[43] O. M. Nayfeh, C. N. Chleirigh, J. Hennessy, L. Gomez, J. L. Hoyt, and D. A. Antoniadis, "Design of tunneling field-effect transistors using strained-silicon/strained-germanium type-II staggered heterojunctions," *Electron Device Lett. IEEE*, vol. 29, no. 9, pp. 1074–1077, 2008.

[44] S. O. Koswatta, S. J. Koester, and W. Haensch, "1D broken-gap tunnel transistor with MOSFET-like on-currents and sub-60mV/dec subthreshold swing," in *Electron Devices Meeting (IEDM)*, 2009, pp. 1–4.

[45] S. O. Koswatta, S. J. Koester, and W. Haensch, "On the possibility of obtaining MOSFET-like performance and sub-60-mV/dec swing in 1-D broken-gap tunnel transistors," *Electron Devices, IEEE Trans.*, vol. 57, no. 12, pp. 3222–3230, 2010.

[46] J. Knoch and J. Appenzeller, "Modeling of high-performance p-type III–V heterojunction tunnel FETs," *Electron Device Lett. IEEE*, vol. 31, no. 4, pp. 305–307, 2010.

[47] L. Wang, E. Yu, Y. Taur, and P. Asbeck, "Design of Tunneling Field-Effect Transistors Based on Staggered Heterojunctions for Ultralow-Power Applications," *Electron Device Lett. IEEE*, vol. 31, no. 5, pp. 431–433, 2010.

[48] B. M. Borg, K. A. Dick, B. Ganjipour, M. E. Pistol, L. E. Wernersson, and C. Thelander, "InAs/GaSb heterostructure nanowires for tunnel field-effect transistors," *Nano Lett.*, 2010.

[49] G. Zhou, R. Li, T. Vasen, M. Qi, S. Chae, Y. Lu, Q. Zhang, H. Zhu, J. M. Kuo, and T. Kosel, "Novel gate-recessed vertical InAs/GaSb TFETs with record high ION of 180 μA/μm at VDS= 0.5 V," in *Electron Devices Meeting (IEDM)*, 2012, pp. 32.6.1 – 32.6.4.

[50] A. W. Dey, B. M. Borg, B. Ganjipour, M. Ek, K. A. Dick, E. Lind, P. Nilsson, C. Thelander, and L.-E. Wernersson, "High current density InAsSb/GaSb tunnel field effect transistors," in *Device Research Conference (DRC), 2012 70th Annual*, 2012, pp. 205–206.

[51] Y. Zhu, N. Jain, S. Vijayaraghavan, D. K. Mohata, S. Datta, D. Lubyshev, J. M. Fastenau, W. K. Liu, N. Monsegue, and M. K. Hudait, "Role of InAs and GaAs terminated heterointerfaces at source/channel on the mixed As-Sb staggered gap tunnel field effect transistor structures grown by molecular beam epitaxy," *J. Appl. Phys.*, vol. 112, no. 2, pp. 24306–24316, 2012.

[52] D. K. Mohata, R. Bijesh, Y. Zhu, M. K. Hudait, R. Southwick, Z. Chbili, D. Gundlach, J. Suehle, J. M. Fastenau, D. Loubychev, a. K. Liu, T. S. Mayer, V. Narayanan, and S. Datta, "Demonstration of improved heteroepitaxy, scaled gate stack and reduced interface states enabling heterojunction tunnel FETs with high drive current and high on-off ratio," *Dig. Tech. Pap. - Symp. VLSI Technol.*, pp. 53–54, 2012.

[53] D. K. Mohata, R. Bijesh, S. Mujumdar, C. Eaton, R. Engel-Herbert, T. Mayer, V. Narayanan, J. M. Fastenau, D. Loubychev, a. K. Liu, and S. Datta, "Demonstration of MOSFET-like on-current performance in arsenide/antimonide tunnel FETs with staggered hetero-junctions for 300mV logic applications," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 5, pp. 781–784, 2011.

[54] S. Mookerjea and D. Mohata, "Temperature-Dependent IV Characteristics of a Vertical In0. 53Ga0. 47As Tunnel FET," *IEEE Electron Device Lett.*, vol. 31, no. 6, pp. 564–566, 2010.

[55] X. Zhao, A. Vardi, and J. A. Alamo, "InGaAs / InAs Heterojunction Vertical Nanowire Tunnel FETs Fabricated by a Top-down Approach," *Electron Devices Meet.*, pp. 590–593, 2014.

[56] P. Y. Aghdam, "Sb-Heterostructure Backward Diode for Millimetre-Wave Detection," Chalmers University of Technology, 2012.

[57] M. D. With, Z. Zhang, R. Rajavel, P. Deelman, P. Fay, S. Member, A. Inas, and A. Algasb, "Sub-Micron Area Heterojunction Backward Diode Millimeter-Wave Detectors with 0.18 pW/Hz^1/2 Noise Equivalent Power," *IEEE Microw. Wirel. Components Lett.*, vol. 21, no. 5, pp. 267–269, 2011.

[58] T. Takahashi, M. Sato, Y. Nakasha, and N. Hara, "Improvement in nonlinear characteristics of zero bias GaAsSb-based backward diodes," in *Indium Phosphide and Related Materials (IPRM), 2013 International Conference on*, 2013, pp. 1–2.

[59] T. Takahashi, M. Sato, Y. Nakasha, and N. Hara, "Lattice-matched p+-GaAsSb/i-InAlAs/n-InGaAs zero-bias backward diodes for millimeter-wave detectors and mixers," in *Indium Phosphide and Related Materials (IPRM), 2012 International Conference on*, 2012, pp. 95–98.

[60] E. A. Fitzgerald, "Dislocations in strained-layer epitaxy: theory, experiment, and applications," *Mater. Sci. reports*, vol. 7, no. 3, pp. 87–140, 1991.

[61] C. Kisielowski, J. Palm, B. Bollig, and H. Alexander, "Inhomogeneities in plastically deformed silicon single crystals. I. ESR and photo-ESR investigations of p- and n-doped silicon," *Phys. Rev. B*, vol. 44, no. 4, pp. 1588–1599, 1991.

[62] P. R. Wilshaw and G. R. Booker, "New results and an interpretation for SEM EBIC contrast arising from individual dislocations in silicon," in *Inst Phys Conf Ser*, 1985, vol. 76, pp. 329–336.

[63] D. Redfield, "Electric fields of defects in solids," *Phys. Rev.*, vol. 130, no. 3, p. 914, 1963.

[64] K. Mukherjee, D. a. Beaton, a. Mascarenhas, M. T. Bulsara, and E. a. Fitzgerald, "Effects of dislocation strain on the epitaxy of lattice-mismatched AlGaInP layers," *J. Cryst. Growth*, vol. 392, pp. 74–80, Apr. 2014.

[65] J. T. Teherani, T. Yu, D. Antoniadis, and J. L. Hoyt, "Electrostatic design of vertical tunneling field-effect transistors," in *Energy Efficient Electronic Systems (E3S), 2013 Third Berkeley Symposium on*, 2013, pp. 1–2.

[66] W. Hsu, J. Mantey, L. F. Register, and S. K. Banerjee, "On the Electrostatic Control of Gate-Normal-Tunneling Field-Effect Transistors," *IEEE Trans. Electron Devices*, vol. 62, no. 7, 2015.

[67] S. Agarwal and E. Yablonovitch, "Band-Edge Steepness Obtained From Esaki / Backward Diode Current – Voltage Characteristics," *IEEE Trans. Electron Devices*, vol. 61, no. 5, pp. 1488–1493, 2014.

[68] S. Agarwal and E. Yablonovitch, "Pronounced Effect of pn-Junction Dimensionality on Tunnel Switch Sharpness," *Arxiv Prepr. arXiv1109.0096*, 2011.

[69] G. B. Stringfellow, *Organometallic Vapor-Phase Epitaxy: Theory and Practice*. Elsevier Science, 1998.

[70] P. Sharma, "Integration of GaAsP alloys on SiGe virtual substrates for Si-based dual-junction solar cells." Massachusetts Institute of Technology, 2013.

[71] Y. Huang, J.-H. Ryou, R. D. Dupuis, V. R. D'Costa, E. H. Steenbergen, J. Fan, Y.-H. Zhang, a. Petschke, M. Mandl, and S.-L. Chuang, "Epitaxial growth and characterization of InAs/GaSb and InAs/InAsSb type-II superlattices on GaSb substrates by metalorganic chemical vapor deposition for long wavelength infrared photodetectors," *J. Cryst. Growth*, vol. 314, no. 1, pp. 92–96, Jan. 2011.

[72] X. B. Zhang, J. H. Ryou, R. D. Dupuis, C. Xu, S. Mou, A. Petschke, K. C. Hsieh, and S. L. Chuang, "Improved surface and structural properties of InAs/GaSb superlattices on (001) GaSb substrate by introducing an InAsSb layer at interfaces," *Appl. Phys. Lett.*, vol. 90, p. 131110, 2007.

[73] X. B. Zhang, J. H. Ryou, R. D. Dupuis, A. Petschke, S. Mou, S. L. Chuang, C. Xu, and K. C. Hsieh, "Metalorganic chemical vapor deposition growth of high-quality InAs/GaSb type II superlattices on (001) GaAs substrates," *Appl. Phys. Lett.*, vol. 88, p. 72104, 2006.

[74] Y. Huang, J. H. Ryou, R. D. Dupuis, D. Zuo, B. Kesler, S. L. Chuang, H. Hu, K. H. Kim, Y. T. Lu, and K. C. Hsieh, "Strain-balanced InAs/GaSb type-II superlattice structures and photodiodes grown on InAs substrates by metalorganic chemical vapor deposition," *Appl. Phys. Lett.*, vol. 99, p. 11109, 2011.

[75] O. Oda, *Compound semiconductor bulk materials and characterizations*, vol. 2. World Scientific, 2007.

[76] T. L. Ngai, R. C. Sharma, and Y. a. Chang, "The Ga-Sb (Gallium-Antimony) system," *Bull. Alloy Phase Diagrams*, vol. 9, no. 5, pp. 586–591, 1988.

[77] G. R. Booker, P. C. Klipstein, M. Lakrimi, S. Lyapin, N. J. Mason, R. J. Nicholas, T. Y. Seong, D. M. Symons, T. A. Vaughan, and P. J. Walker, "Growth of InAs/GaSb strained layer superlattices. I," *J. Cryst. Growth*, vol. 145, no. 1–4, pp. 778–785, 1994.

[78] Q. Xie, J. Van Nostrand, J. E. Van Nostrand, J. L. Brown, and C. E. Stutz, "Arsenic for antimony exchange on GaSb, its impacts on surface morphology, and interface structure," *J. Appl. Phys.*, vol. 329, no. 1, p. 329, 1999.

[79] Y. K. Su, C. L. Lin, S. M. Chen, J. R. Chang, and D. H. Jaw, "Optical and structural characterization of InAs/GaSb superlattices," *J. Appl. Phys.*, vol. 81, p. 7529, 1997.

[80] M. Lakrimi, R. W. Martin, N. J. Mason, R. J. Nicholas, and P. J. Walker, "GaSb/InAs heterojunctions grown by MOVPE: Effect of gas switching sequences on interface quality," *J. Cryst. Growth*, vol. 110, no. 4, pp. 677–682, 1991.

[81] D. Lackner, O. J. Pitts, S. Najmi, P. Sandhu, K. L. Kavanagh, A. Yang, M. Steger, M. L. W. Thewalt, Y. Wang, and D. W. McComb, "Growth of InAsSb/InAs MQWs on GaSb for mid-IR photodetector applications," *J. Cryst. Growth*, vol. 311, no. 14, pp. 3563–3567, 2009.

[82] V. M. Donnelly and J. A. McCaulley, "Methyl exchange reaction of trimethylindium on GaAs (100) and the preferential etching of gallium," *Surf. Sci.*, vol. 235, no. 2, pp. L333–L337, 1990.

[83] Y. Zeng, C. I. Kuo, R. Kapadia, C. Y. Hsu, A. Javey, and C. Hu, "Two-dimensional to three-dimensional tunneling in InAs/AlSb/GaSb quantum well heterojunctions," *J. Appl. Phys.*, vol. 114, no. 2, pp. 2–5, 2013.

[84] N. Su, Y. Tang, Z. Zhang, T. F. Kuech, and P. Fay, "Observation and control of electrochemical etching effects in the fabrication of InAs/AlSb/GaSb heterostructure devices," *J. Vac. Sci. Technol. B Microelectron. Nanom. Struct.*, vol. 26, no. 3, p. 1025, 2008.

[85] C. von Eichel-Streiber, M. Behet, M. Heuken, and K. Heime, "Doping of InAs, GaSb and InPSb by low pressure MOVPE," *J. Cryst. Growth*, vol. 170, no. 1–4, pp. 783–787, 1997.

[86] S. K. Haywood, A. B. Henriques, N. J. Mason, R. J. Nicholas, and P. J. Walker, "Growth of GaSb by MOVPE," *Semicond. Sci. Technol.*, vol. 3, p. 315, 1988.

[87] S. K. Haywood, R. W. Martin, N. J. Mason, R. J. Nicholas, and P. J. Walker, "GaSb/InAs heterojunctions grown by MOVPE," *J. Cryst. Growth*, vol. 107, no. 1, pp. 422–427, 1991.

[88] J. Hornstra, "Dislocations in the diamond lattice," *J. Phys. Chem. Solids*, vol. 5, no. 1, pp. 129–141, 1958.

[89] A. T. Blumenau, "The Modelling of Dislocations in Semiconductor Crystals," in *EUROMAT*, 2005.

[90] Y. Liu, J. Yan, H. Wang, and G. Han, "Temperature dependent I DS – V GS characteristics of an N-channel Si tunneling field-effect transistor with a germanium source on Si(110) substrate," *J. Semicond.*, vol. 35, no. 2, p. 024001, Feb. 2014.

[91] J. F. Chen, M. C. Wu, L. Yang, and A. Y. Cho, "InAs / AlSb / GaSb singlebarrier interband tunneling diodes with high peakto valley ratios at room temperature," *J. Appl. Phys.*, vol. 68, 1990.

[92] R. G. Meyers, P. Fay, J. N. Schulman, S. Thomas III, D. H. Chow, J. Zinck, Y. K. Boegeman, and P. Deelman, "Bias and temperature dependence of Sb-based heterostructure millimeter-wave detectors with improved sensitivity," *Electron Device Lett. IEEE*, vol. 25, no. 1, pp. 4–6, 2004.

[93] J. Shen, "Mechanisms of valley currents in InAs/AlSb/GaSb resonant interband tunneling diodes," *J. Appl. Phys.*, vol. 78, no. 10, p. 6220, 1995.

[94] K. Majumdar, P. Thomas, W. Loh, P. Hung, K. Matthews, D. Pawlik, B. Romanczyk, M. Filmer, A. Gaur, R. Droopad, S. L. Rommel, C. Hobbs, and P. D. Kirsch, "Mapping

Defect Density in MBE Grown Valley Characteristics," *IEEE Trans. Electron Devices*, vol. 61, no. 6, pp. 2049–2055, 2014.

[95] R. M. Iutzi and E. A. Fitzgerald, "Microstructure and conductance-slope of InAs/GaSb tunnel diodes," *J. Appl. Phys.*, vol. 115, no. 23, p. 234503, Jun. 2014.

[96] L. Yang, M. T. Bulsara, K. E. Lee, and E. a. Fitzgerald, "Compositionally-graded InGaAsInGaP alloys and GaAsSb alloys for metamorphic InP on GaAs," *J. Cryst. Growth*, vol. 324, no. 1, pp. 103–109, 2011.

[97] V. Saripalli, G. Sun, and A. Mishra, "Exploiting heterogeneity for energy efficiency in chip multiprocessors," *IEEE Trans. Emerg. Sel. Top. Circuits Syst.*, vol. 1, no. 2, pp. 109–119, 2011.

[98] M. E. Groenert, C. W. Leitz, A. J. Pitera, V. Yang, H. Lee, R. J. Ram, and E. A. Fitzgerald, "Monolithic integration of room-temperature cw GaAs/AlGaAs lasers on Si substrates via relaxed graded GeSi buffer layers," *J. Appl. Phys.*, vol. 93, no. 1, pp. 362–367, 2003.

[99] E. A. Fitzgerald, Y. Xie, M. L. Green, D. Brasen, A. R. Kortan, J. Michel, Y. Mii, and B. E. Weir, "Totally relaxed GexSi1− x layers with low threading dislocation densities grown on Si substrates," *Appl. Phys. Lett.*, vol. 59, no. 7, pp. 811–813, 1991.

[100] M. T. Currie, S. B. Samavedam, T. A. Langdo, C. W. Leitz, and E. A. Fitzgerald, "Controlling threading dislocation densities in Ge on Si using graded SiGe layers and chemical-mechanical polishing," *Appl. Phys. Lett.*, vol. 72, no. 14, pp. 1718–1720, 1998.

[101] J. A. Carlin, S. A. Ringel, E. A. Fitzgerald, M. Bulsara, and B. M. Keyes, "Impact of GaAs buffer thickness on electronic quality of GaAs grown on graded Ge/GeSi/Si substrates," *Appl. Phys. Lett.*, vol. 76, no. 14, pp. 1884–1886, 2000.

[102] N. J. Quitoriano and E. A. Fitzgerald, "Relaxed, high-quality InP on GaAs by using InGaAs and InGaP graded buffers to avoid phase separation," *J. Appl. Phys.*, vol. 102, no. 3, p. 33511, 2007.

[103] A. Jandl, M. T. Bulsara, and E. a. Fitzgerald, "Materials properties and dislocation dynamics in InAsP compositionally graded buffers on InP substrates," *J. Appl. Phys.*, vol. 115, no. 15, 2014.

[104] R. A. Kiehl, "Complementary Heterostructure FET Integrated Circuits," *Semicond. Semimetals*, vol. 41, pp. 219–290, 1994.

[105] M. O. Li, D. Esseni, J. J. Nahas, D. Jena, and H. G. Xing, "Two-Dimensional Heterojunction Interlayer Tunneling Field Effect Transistors (Thin-TFETs)," *J. Electron Devices Soc.*, vol. 3, no. 3, 2015.

[106] T. Roy, M. Tosun, X. Cao, H. Fang, D. Lien, P. Zhao, and R. O. Y. E. T. Al, "Dual-Gated MoS 2 /WSe 2 van der Waals Tunnel Diodes and Transistors," *ACS Nano*, vol. 9, no. 2, pp. 2071–2079, 2015.