The Pennsylvania State University

The Graduate School

College of Engineering

PHYSICS, FABRICATION AND CHARACTERIZATION OF III-V MULTI-GATE FETS FOR LOW POWER ELECTRONICS

A Dissertation in

Electrical Engineering

by

Arun V. Thathachary

© 2015 Arun V. Thathachary

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

December 2015

The dissertation of Arun V. Thathachary was reviewed and approved* by the following:

Suman Datta
Professor of Electrical Engineering and Computer Science
Dissertation Advisor
Co-Chair of Committee

Theresa Mayer
Professor of Electrical Engineering and Computer Science
Associate Dean for research and innovation, College of Engineering
Co-Chair of Committee

Nitin Samarth
Professor of Physics
George A. and Margaret M. Downsbrough Department Head

Weihua Guan
Assistant Professor of Electrical Engineering and Computer Science

Roman Engel-Herbert
Assistant Professor of Materials Science and Engineering

Kultegin Aydin
Professor of Electrical Engineering and Computer Science
Department Head

*Signatures are on file in the Graduate School

ABSTRACT

With transistor technology close to its limits for power constrained scaling and the simultaneous emergence of mobile devices as the dominant driver for new scaling, a pathway to significant reduction in transistor operating voltage to 0.5V or lower is urgently sought. This however implies a fundamental paradigm shift away from mature Silicon technology. III-V compound semiconductors hold great promise in this regard due to their vastly superior electron transport properties making them prime candidates to replace Silicon in the n-channel transistor. Among the plethora of binary and ternary compounds available in the III-V space, $In_xGa_{1-x}As$ alloys have attracted significant interest due to their excellent electron mobility, ideally placed bandgap and mature growth technology. Simultaneously, electrostatic control mandates multi-gate transistor designs such as the FinFET at extremely scaled nodes.

This dissertation describes the experimental realization of III-V FinFETs incorporating $In_XGa_{1-X}As$ heterostructure channels for high performance, low power logic applications. The chapters that follow present experimental demonstrations, simulations and analysis on the following aspects (a) motivation and key figures of merit driving material selection and design; (b) dielectric integration schemes for high-k metal-gate stack (HKMG) realization on $In_XGa_{1-X}As$, including surface clean and passivation techniques developed for high quality interfaces; (c) novel techniques for transport (mobility) characterization in nanoscale multi-gate FET architectures with experimental demonstration on $In_{0.7}Ga_{0.3}As$ nanowires; (d) Indium composition and quantum confined channel design for $In_XGa_{1-X}As$ FinFETs and (e) InAs heterostructure designs for high performance FinFETs. Each chapter also contains detailed benchmarking of results against state of the art demonstrations in Silicon and III-V material systems. The dissertation concludes by assessing the feasibility of $In_XGa_{1-X}As$ FinFET devices as n-channel Silicon replacement for low power logic technology scaling.

TABLE OF CONTENTS

List of Figures vi

List of Tables xii

Acknowledgements xiii

Chapter 1 $In_xGa_{1-x}As$ channels in the context of low power logic technology 1

- I. Motivation 3
    - Current in nanoscale MOSFETs 4
    - Quantum capacitance bottleneck 7
- II. Layer structure design and characterization 8
- III. FinFET device architecture 11

Chapter 2 High-k dielectric gate stack integration on $In_xGa_{1-x}As$ 13

- I. MOS capacitor fabrication and characterization: Thermal ALD 14
- II. Low power plasma surface clean and passivation 15
    - Bi-layer vs. single layer dielectric gate stack 20
- III. Conclusions 25

Chapter 3 Characterization of electronic transport in nanoscale multi-gate FETs 27

- I. Introduction 27
- II. Test structure design and fabrication 30
    - Device metrology and material characterization 31
- III. Results and discussion 33
    - Device characterization 33
    - Equivalent circuit model and mobility extraction 35
    - Validation of measured results 37
    - Identifying the cause for mobility degradation 39
- IV. Ballistic transport in nanoscale $In_xGa_{1-x}As$ structures 43
- V. Conclusions 46

Chapter 4 Channel architecture and composition engineering for high performance $In_xGa_{1-x}As$ FinFETs 47

- I. Channel architecture, device design and fabrication 47
- II. Results and discussion 50
    - Multi-fin split CV measurements 51
    - Long channel FinFET performance 53
    - Short channel FinFET performance 55
    - Benchmarking 58
    - Projected short channel performance 60
- III. Conclusions 61

Chapter 5 High performance InAs heterostructure FinFETs 63

- I. Motivation 63
- II. Device design and fabrication 64
  - Bandstructure effects 65
- III. Results and discussion 67
  - Long channel performance 67
  - Short channel benchmarking 70
- IV. Conclusions 73

Chapter 6 Conclusions and future work 75

- I. Summary and conclusions 75
  - Impact of confinement on transport : confined phonon modes 80
- II. Future work: High aspect ratio InAs channel FinFETs 81

Bibliography 86

- Appendix Process flow for III-V FinFET fabrication 96
  - E-beam litho markers + S/D pads 96
  - Mesa isolation 96
  - Gate recess etch 97
  - Fin Definition 97
  - Gate Stack: High-k deposition 98
  - Gate electrode definition 98

LIST OF FIGURES

Figure 1-1. Scaling trends of CPUs showing saturation in the clock frequency and an upper limit to power dissipation (source: [6]). 2

Figure 1-2. Schematic showing the potential performance gain from inserting high electron mobility III-V channel materials. Comparable drive currents may be achieved at 0.5V supply voltage. 3

Figure 1-3. Bulk electron and hole mobilities for III-V compound semiconductors plotted as a function of bandgap (from [3]). Ternary alloys of InGaAs have high electron mobility with a suitably high bandgap to enable scalable III-V logic technology. 4

Figure 1-4. Schematic representation of the top of the barrier injection model for nanoscale MOSFETs indicating the virtual source point. 5

Figure 1-5. Injection velocity as a function of gate length measured from planar III-V HEMT devices [4]. 6

Figure 1-6. Equivalent circuit representation of quantum effects in the inversion charge. Low effective mass in III-V semiconductors results in more pronounced quantum effects. 7

Figure 1-7. Electron mobility μ versus sheet electron density n_S in n-channel Si FETs and III-V QWFETs: Si MOSFETs, InGaAs/AlGaAs HFETs [12,13], lattice-matched InGaAs/InAlAs QWFETs [14,15], pseudomorphic InGaAs/InAlAs QWFETs [16,17], InAs/AlSb QWFETs [18,19]. 8

Figure 1-8. Schematic showing generic layer structure of quantum well channels investigated in this work. 9

Figure 1-9. (a) Device layer structure showing the remote δ-doping layer and (b) band diagram indicating the first two sub-bands and the Fermi level. 10

Figure 1-10. (a) Hall mobility measured as a function of temperature (symbols). Solid lines show fit obtained using a relaxation time approximation based scattering analysis. (b) Carrier density contribution from the first two sub-bands. 10

Figure 1-11. Scaling trends of drive current, fin width and fin height from 22nm node onwards. Values for the 10 and 7nm nodes are projected assuming simplistic scaling by 0.7x in dimensions per generation. 11

Figure 1-12. Representative schematic showing the architecture of the quantum well FinFET devices fabricated in this work. 12

Figure 2-1. Experimental measured capacitance (a) and conductance (b) data for a 1nm Al_2O_3/3nm HfO_2 stack grown by thermal ALD. 15

Figure 2-2. Measured capacitance (a) and conductance (b) data shown for MOS capacitor with bi-layer gate stack and low temperature (110C) plasma nitride surface passivation....16

Figure 2-3. Gate leakage measured for MOS capacitor shown in figure 2-2....17

Figure 2-4. Effect of FGA on bi-layer dielectric gate stack with high temperature plasma nitrogen clean and passivation. (a) and (b) show capacitance data before and after FGA while (c) and (d) show the corresponding conductance data. 18

Figure 2-5. Gate leakage comparison before and after FGA for bi-layer dielectric gate stack. 19

Figure 2-6. Cross-section transmission electron micrograph of the bi-layer dielectric gate stack with $250^{o}C$ plasma nitride surface passivation. 20

Figure 2-7. (a) and (b) show measured capacitance data (symbols) while (c) and (d) show measured conductance data (symbols) for bi-layer 1nm Al_2O_3/3nm HfO_2 and 4nm HfO_2 only stacks. Solid lines indicate fit from equivalent circuit model....21

Figure 2-8. (a) Extracted D_{IT} profiles and (b) estimated trap time constant for bi-layer and HfO_2 only gate stacks....22

Figure 2-9. (a) and (b) show measured capacitance data for pre-gate metal and post-gate metal forming gas anneals. (c) and (d) show the corresponding conductance data. Devices comprise of 4nm HfO_2 only gate stack with AlO_xN_y passivation. 23

Figure 2-10. Measured hysteresis of (a) pre-gate metal and (b) post-gate metal FGA. 24

Figure 2-11. (a) Capacitance and (b) conductance data for 3.5nm HfO_2 only gate stack with $250^{o}C$ plasma nitride surface passivation and FGA with frequency sweep down to 1KHz. 24

Figure 2-12. Gate leakage current measured for 3.5nm HfO_2 only gate stack. 25

Figure 2-13. Extracted DIT profiles for thermal ALD and plasma nitride surface passivated gate stacks....25

Figure 3-1. Schematic representation of various extrinsic scattering mechanisms that affect transport in multi-gate NWFET structures on III-V substrates....28

Figure 3-2. (a) SEM of multi-gate, $In_{0.7}Ga_{0.3}As$ NWFET with an array of five NWs of width 40nm. Additional probe electrodes are integrated to form a Hall bridge structure. Hall voltage (V_H) and longitudinal (V_L) voltage are measured as shown. (b) Schematic cross-section of the NW showing the layer structure of the substrate. (c) Schematic showing two representative NWs and the structure of the probe underneath the gate. The heavily doped n+ cap layer is retained on the probe electrode in the regions between the NWs and (d) perspective view of single NW. 31

Figure 3-3. Characterization of the top down patterned nanowire subsequent to plasma etching: (a) STEM image indicating location of EELS line scan and (b) EELS line scan showing possible formation of indium oxide interfacial layer. (c) EELS elemental map of NW cross-section confirming that top-down etch yields high quality, defect-free NW structure, while also highlighting the formation of native oxide layer on side walls. 32

Figure 3-4. (a) Transfer characteristics (I_D-V_G) of the multi-gate NWFET based Hall structure, with NW width of 40nm, showing good sub-threshold behavior. (b) Output characteristics (I_D-V_D) of the same device. 33

Figure 3-5. Experimentally measured Hall electron mobility for multi-gate NWFET of Figure 3-2 as a function of carrier concentration for different NW widths. (b) Mobility shows a monotonic roll-off with reducing NW dimensions. Peak mobility is observed at a carrier concentration of $2x10^{12}$ cm^{-2} for all NW widths. 35

Figure 3-6. Equivalent circuit model for interpretation of measured results. 37

Figure 3-7. Improving measurement accuracy: (a) simulated electrostatic potential through the NW array. (b) Simulated Hall potential profile plotted as a function of position along the Hall probe. The results show that each NW contributes a small Hall potential to the total V_H, which is measured across the device. 38

Figure 3-8. (a) Measured Hall mobility as a function of carrier concentration showing excellent agreement between single NW and NW array device. Results confirm the equivalent circuit based interpretation of Fig. 3-6. (b) Simulated Hall voltage as a function of NW number showing an increase in measured Hall voltage with number of NWs. Due to the averaging effect of the array, the error of the measured Hall voltage reduces as number of NWs increases. Error is less than 1% when the number of NWs is increased beyond 5...39

Figure 3-9. (a) Temperature dependent, un-gated Hall measurements showing mobility degradation as the InGaAs quantum well is patterned into NWs with width reducing from 1 μm (planar) down to 100nm. (b) Additional scattering extracted using Mathiessen's rule. This component is found to be independent of temperature clearly indicating that the scattering arises from roughness due to the side walls. 40

Figure 3-10. Calculated overlap integral values F_{nm} as a function of NW width at N_S = $2x10^{12}$. The effective electric field from the potential fluctuations of the sidewall roughness increases monotonically with reducing NW width. 42

Figure 3-11. Projection of mobility for 10nm NW width. The side-wall roughness scattering model is calibrated to the measured data. The gated Hall measurement results clearly indicate a mobility roll-off with decreasing nanowire width. Comparison of mobility between InGaAs and silicon NWs indicates over 10× higher mobility for InGaAs at a NW width of 10nm. 43

Figure 3-12. (a) Simulated potential distribution profile for a Hall cross structure in the diffusive regime. As current flows from contact 3 to 4 a positive bend resistance is measured. (b) Potential distribution in the ballistic regime. Electrons from contact 4 travel past the junction without scattering and start accumulating in contact 2 reversing the polarity of V_{21} giving $R_B < 0$. (c) Measured bend resistance as a function of magnetic field and temperature for $In_{0.7}Ga_{0.3}As$ NW based Hall cross structure. $R_B < 0$ is observed at room temperature confirming ballistic transport at room temperature. 46

Figure 4-1. Schematic showing the layer structure for devices A,B and C investigated in this work. all substrates are finally patterned into FinFET structures as shown in the bottom panel. 49

Figure 4-2. Cross-section transmission electron micrographs (TEM) showing (a) tight fin pitch with 10 fins/μm layout width (b) magnified image showing critical dimensions achieved with side wall slope of 75° (c) magnified false color image highlighting the high-k dielectric and metal gate. 49

Figure 4-3. Two dimensional Schrodinger-Poisson simulations show inherent volume inversion in the thin quantum well devices. Similar volume inversion densities in the thick InGaAs channel is achieved only at a highly scaled fin width of 8nm. 50

Figure 4-4. SEM of a long channel multi-fin device for measuring split-CV fin capacitance. 51

Figure 4-5. Measured multi-fin split CV capacitance (top panel) and conductance (bottom panel) for the thick $In_{0.53}Ga_{0.47}As$, thin $In_{0.53}Ga_{0.47}As$ and $In_{0.7}Ga_{0.3}As$ quantum wells in (a), (b) and (c) respectively. 52

Figure 4-6. Fit obtained between experimental capacitance and conductance data modeled using the equivalent circuit method. 52

Figure 4-7. Extracted interface state density (D_{it}) profile for the three multi-fin devices is shown and compared to planar MOS capacitor D_{it} profiles. 53

Figure 4-8. (a), (b) and (c) show measured long channel transfer characteristics (top panel) and output characteristics (bottom panel) for substrates A, B and C respectively. 54

Figure 4-9. Experimentally extracted field effect mobility for long channel FinFET devices. 55

Figure 4-10. (a) Schematic showing the structure of the short channel multi-fin FET device. (b) Top view SEM of fabricated device. 56

Figure 4-11. (a), (b) and (c) show measured transfer characteristics (top panel) and output characteristics (bottom panel) for the substrates A, B and C respectively. 57

Figure 4-12. Experimentally extracted transconductance (g_m) for the three short channel devices vs. gate overdrive. 58

Figure 4-13. Benchmarking plot showcasing the transconductance (g_m) for the three short channel devices as a function of sub-threshold slope (SS). 59

Figure 4-14. (a) Velocity profiles along the channel for the three short channel devices extracted using calibrated TCAD simulations. (b) Benchmarking of the injection velocity against state of the art silicon and $In_{0.7}Ga_{0.3}As$ HEMT devices[4]. 60

Figure 4-15. Projected performance for 7nm node FinFET at 0.5V supply voltage showing (a) drive currents and capacitance per fin for the three substrates. 61

Figure 5-1. Cross-section schematic of the three FinFET structures highlighting the single and dual InAs quantum wells. 64

Figure 5-2. (a) Schematic of fabricated FinFET devices (b) Process flow outlining the spacer technique for patterning fins (c) cross-section SEM of etched fins and (d) cross-section SEM of etched fin shown with residual hard mask. 65

Figure 5-3. One dimensional Schrodinger-Poisson simulations using nextnano showing reducing sub-band spacing going from the (a) InGaAs QW to the (b) single and (c) dual QW substrates. 66

Figure 5-4. Two dimensional Schrodinger-Poisson simulations using nextnano showing higher inversion charge concentration near the InAs regions in the single and dual quantum well structures. 66

Table 5-1. Effective perimeter of FinFET devices based on weighted average of carrier distributions. 67

Figure 5-5. Transfer characteristics (top panel) and output characteristics (bottom panel) are shown for FinFETs on the three substrates from figure 5-1 in column (a), (b) and (c) respectively. 68

Figure 5-6. (a) Long channel FinFET mobility extracted using inverse modeling in Sentaurus TCAD after calibration to experimental long channel transfer characteristics. (b) Long channel FinFET transconductance. 69

Figure 5-7. Benchmarking of experimental long channel FinFET g_m against published long channel $In_XGa_{1-X}As$/InAs devices [61,73]. 69

Figure 5-8. Short channel (L_G = 300nm) transfer characteristics shown for single (blue solid line) and dual (red solid line) FinFET devices. Right axis shows corresponding transconductance. 70

Figure 5-9. Fin cross-sections showing electron density profiles simulated at V_G = 0.5V (I_{OFF} = 100nA/µm) for scaled 8nm devices, (a) Si FinFET (b) InAs single QW FinFET and (c) InAs DQW FinFET. 71

Figure 5-10. Electron velocity profiles for FinFETs projected at 22nm node using Sentaurus TCAD. Transport parameters are calibrated to experimental short channel devices....72

Figure 5-11. Transfer characteristics (current per fin) of InAs SQW and DQW devices benchmarked against 22nm Si FinFET technology. 72

Table 5- 1 Benchmarking of projected short channel performance (at 22nm node) with state of the art 22nm silicon technology. InAs dual QW FinFET devices show promising on-current....73

Figure 6-1. External resistance for InAs QW FinFETs with raised source/drain architectures rises rapidly with stacking of InAs QW....77

Figure 6-2. Charge density estimated using two dimensional Poisson-Schrodinger simulations for different III-V architectures explored in this work. The higher DoS in silicon results in almost 2x higher charge....79

Figure 6-3. TCAD simulation of scaled devices comparing projected performance with 22nm Silicon FinFETs [20]....80

Figure 6-4. Linear grading of $In_XAl_{1-X}As$ buffer to accommodate thick pseudomorphic InAs channel (0-20nm). The lattice mismatch still limits maximum thickness in order to accommodate strain. Shown on the right is the corresponding band diagram....82

Figure 6-5. Band gaps and alignment of various Sb based buffer layers relative to InAs. 83

Figure 6-6. A thick AlSb metamorphic buffer growth on InP/InAlAs can provide a closely lattice matched system to InAs along with large conduction band offsets required for electron confinement. Shown on the right is the corresponding band diagram....84

Figure 6-7. Thick metamorphic AlSb growth followed by the growth of thick relaxed AlAsSb can provide a perfectly lattice matched buffer layer which also satisfies the criteria of large conduction band offset (good electron confinement). Shown on the right is the corresponding band diagram....85

LIST OF TABLES

Table 5- 1 Benchmarking of projected short channel performance (at 22nm node) with state of the art 22nm silicon technology. InAs dual QW FinFET devices show promising on-current. 73

ACKNOWLEDGEMENTS

I am indebted to all the people who have made my graduate studies a great experience and this dissertation possible. I would like to express my heartfelt gratitude to all of them.

First I would like to express my sincere gratitude to my advisor Prof. Suman Datta for giving me the opportunity to work on compound semiconductor devices. His guidance and motivation have kept me on edge and ensured I continue to grow intellectually and personally. I am very grateful for the sense of responsibility and commitment he has instilled in me.

I am also grateful to our sponsor, Samsung Electronics Co. Ltd. for funding this dissertation work. I would like to thank our collaborators Mirco Cantoro, K. K Bhuwalka, Y-C Heo, S. Maeda and the entire team from Samsung Logic technology development for their technical inputs and the many fruitful discussions we have had through the course of this dissertation. I would also like to thank my committee members Dr. Theresa Mayer, Dr. Nitin Samarth, Dr. Roman Engel-Herbert and Dr. Weihua Guan for reading

I would like to convey my special thanks and appreciation to the staff in the nanofabrication facility at Penn State, especially Guy Lavallee, Chad Eichfeld, Michael Labella, Bill Drawl, Bangzhi Liu and Jaime Reish for their help with training me at the Penn State Nanofab, as well as technical discussions on device fabrication and extensive process development during the course of my doctoral study.

I sincerely thank the Department of Electrical Engineering in selecting me for the Melvin P. Bloom Memorial Outstanding Doctoral Research Award 2015. I again thank my advisor Prof. Suman Datta for nominating me for this award.

Special thanks to my colleague and collaborator on the Samsung projects, Nidhi Agrawal. Her support with extensive simulations and data extractions have made keeping pace with the project and this dissertation possible. I would also like to thank Mike Barth and Matt Hollander

for their support in the clean room and the characterization labs. The many long late hours we have spent together fabricating devices in the nanofab is something I will remember for a long time. I also thank Rahul Pandey for the many fruitful discussions we have had and the entire Datta group for a great graduate experience.

No graduate student can complete a dissertation without support and a life outside the lab. I am very grateful for the great group of friends I have had the good fortune of knowing especially Nidhi Agrawal, Kadappan, Vishesh, Vivek, Amruta and Divya. Their presence, support and encouragement have enriched me many folds personally and professionally. The many fun hours we spent together like family is something I will cherish forever.

Finally, I owe a debt I cannot repay to my family. I thank my parents, V. T. Ranganathan and Sowmithri Ranganathan, for always leading by example and raising me with a strong sense of love, integrity, courage and resilience. They prepared me for facing the trials and tribulations of life. I will always remember my father's encouraging words and advice when I needed it the most. His gentle wisdom will remain my guiding beacon. I am also grateful to my sister Supriya for her constant words of encouragement and support. Last but not least, I am eternally grateful to my mother who has always stood by me and believed in me. But for her untold sacrifices I would not have come this far in life. I dedicate this thesis to my family and to the loving memory of my father.

Chapter 1

$In_xGa_{1-x}As$ channels in the context of low power logic technology

The metal oxide semiconductor field effect transistor or MOSFET has remained the workhorse of the electronics industry due to its almost miraculous property of increased performance with shrinking size. This simple rule has led to the staggering success of Moore's law, now in its 50th year, and continues to propel the microelectronics industry forward while ushering the world into an era of ubiquitous computing.

In the context of integrated circuits, the MOSFET is used as a switch where its primary function is to switch between two logic states, viz. "0" (OFF) and "1" (ON). Further, with the imperative to make them smaller, faster and cheaper, the primary metrics driving the design of this switch are the area (footprint), switching speed and the energy dissipated during the switching operation [1-5]. Additionally, current integrated circuits employ MOSFETs implemented in a complementary fashion with n and p-channel transistors (based on the mobile charge carrier type) to enable complementary MOS (CMOS) technology. With the progression of Moore's law, the MOSFET has followed a geometric scaling rule enabling exponential gains in performance with every new generation. Increasing transistor density per chip and rising frequencies of operation led to dramatic increase in performance. This however also resulted in considerably high power dissipation and current designs strive to limit the power at 100W and frequencies of 2-3GHz, as seen in figure 1-1. Thus integrated electronics entered an era of power constrained scaling.

Further, with consumer trends and commerce rapidly embracing mobile technology, the electronics industry is predominantly focused on mobile computing platforms, fueling the ever increasing demand for ultra-low power computing technology. Development of this next generation low power technology will enable a plethora of new products such as ultra-small form factor tablets, more powerful smart phones and wearable computing devices that can interface to bio-sensors while potentially drawing power from energy harvesters. It is also critically important to emerging technologies such as the Internet of Things (IoT).

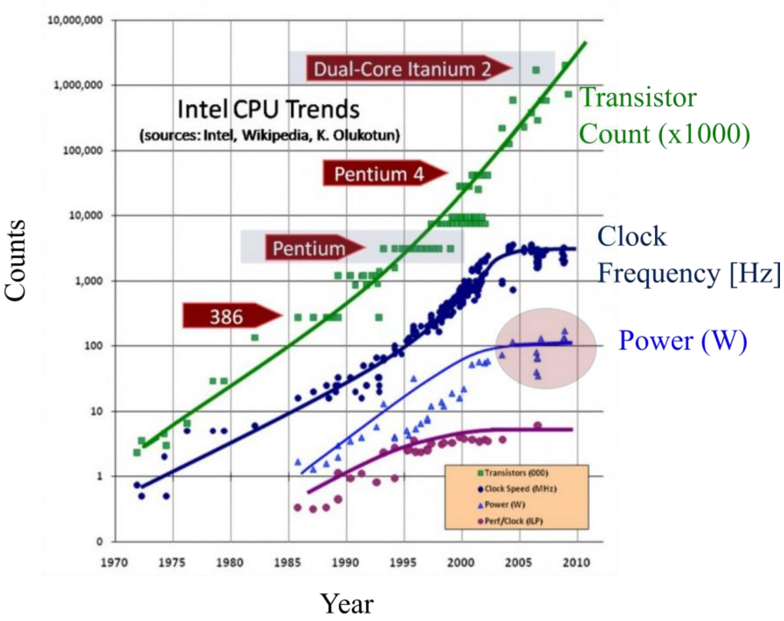


Figure 1-1. Scaling trends of CPUs showing saturation in the clock frequency and an upper limit to power dissipation (source: [6]).

A key requirement to achieve power constrained scaling is to reduce the supply voltage as the dynamic power scales geometrically with supply voltage. This places several stringent and competing requirements on the performance metrics of the transistor such as high drive-current at reduced supply voltage, low leakage current (I_{OFF}), high I_{ON}/I_{OFF} ratio and good sub-threshold slope with well controlled drain induced barrier lowering[2,3]. At reduced supply voltages silicon CMOS is limited by reduced drive currents. Enabling such low power integrated electronics therefore requires a new generation of nanoscale digital-logic transistors capable of operating at voltages of 500mV or lower; a very significant reduction from the present day values of close to 0.75V. One of the fundamental roadblocks preventing this is the non-scalability of threshold voltage (V_T) of the transistor as off-state leakage current trades off exponentially with V_T, as seen in figure 1-2. Therefore implementing a considerable reduction in supply voltage requires a paradigm shift in the design approach of the transistor which has traditionally relied on mature Silicon technology. In this context, interest in group III-V compound semiconductors has grown rapidly within the last decade. These materials have higher electron mobilities as compared to silicon which can be exploited to provide a viable path to reduced supply voltage operation as shown in figure 1-2.

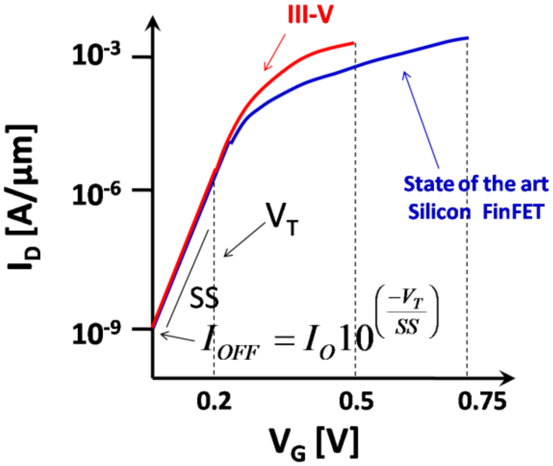


Figure 1-2. Schematic showing the potential performance gain from inserting high electron mobility III-V channel materials. Comparable drive currents may be achieved at 0.5V supply voltage.

I. Motivation

III-V semiconductors possess excellent transport properties in addition to spanning a wide range of bandgaps as seen in figure 1-3. It is further possible to engineer the properties of the material by creating ternary and even quaternary alloys of various binary compounds. Ternary $In_xGa_{1-x}As$ alloys with varying percentage of Indium have high electron mobility making them attractive replacements for the n-channel Silicon transistor. Additionally, a suitable bandgap between 0.35eV (InAs) and 0.75eV (lattice matched $In_{0.53}Ga_{0.47}As/InP$) may be achieved. A reasonable bandgap is critical to scaling the technology [3]. Large bandgaps prevent threshold voltage from scaling due to the absence of suitable work function metals and simultaneously, very small bandgaps can be detrimental to off-state leakage arising from both over the barrier (thermionic) and band-to-band tunneling currents. Room temperature bulk electron mobilities in excess of 10,000 $cm^2/V.sec$ have been measured for high quality lattice matched InGaAs substrates[4].

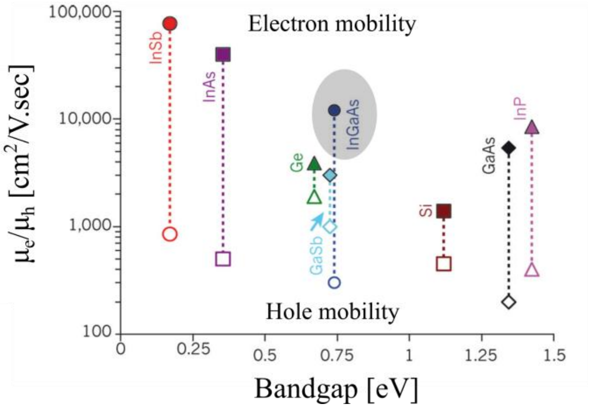


Figure 1-3. Bulk electron and hole mobilities for III-V compound semiconductors plotted as a function of bandgap (from [3]). Ternary alloys of InGaAs have high electron mobility with a suitably high bandgap to enable scalable III-V logic technology.

Current in nanoscale MOSFETs

For nanoscale MOSFETs, significant fraction of the current can propagate in a ballistic fashion from source to drain. At extremely scaled dimensions, a simplistic one dimensional scattering based transport model given by the Lundstrom-Natori model [7-8], captures the essential physics as depicted in figure 1-4. In this model the current is controlled by the inversion charge density and effective injection velocity at the top of the barrier near the source-channel junction, referred to as the virtual source.

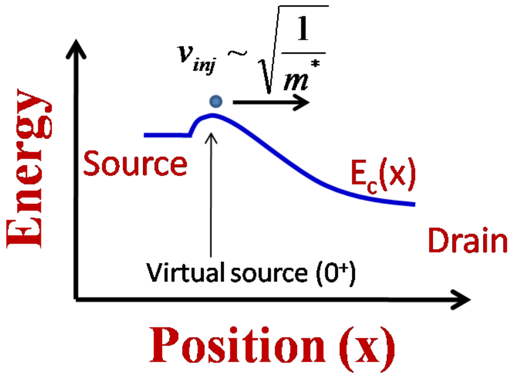


Figure 1-4. Schematic representation of the top of the barrier injection model for nanoscale MOSFETs indicating the virtual source point.

The current in the MOSFET is then given as

$$I_{ON} = WC_{ox}v_{inj}(V_{DD} - V_T)$$

Here, υ_{inj} (also called υ_{eff}) is the effective velocity of electrons at the top of the source which is

given as

$$\frac{1}{v_{inj}} = \frac{1}{v_{Bal}} + \frac{1}{\mu_{ch}E(0^+)} \quad ; v_{Bal} = \sqrt{\frac{2kT}{\pi m^*}}$$

Here, μ_{ch} refers to the channel mobility. Higher μ_{ch} allows the effective injection velocity v_{inj} to approach the ballistic limit denoted as v_{Bal}. It is important to note here that both the mobility as well as the ballistic injection velocity depend inversely on the channel effective mass. Thus introducing low effective mass materials will directly impact the injection velocity and hence on-current of the MOSFET.

Figure 1-5 summarizes injection velocity (υ_{inj}) for III-V semiconductors in comparison to silicon. It can be observed that the lower effective mass InGaAs channels provide significantly higher (2-3x) injection velocity at 0.5V supply as compared to silicon at close to 1V supply voltage. Additionally it is important that υ_{inj} is extracted and compared at the same DIBL conditions as this has a direct impact on the barrier height at the virtual source. Although at first it appears as if III-V devices have a clear advantage over Silicon in transport properties, it is important to look at the corresponding charge density at which υ_{inj} is extracted in these systems. Injection velocity is a function of the inversion charge density which is in turn determined by the gate bias (see reference 13 in [4]). It is observed that the injection velocity goes through a peak as a function of the gate voltage. The initial increase is due to the fact that the υ_{Bal} term in the injection velocity equation above is in fact weighted by Fermi functions (see [7]) dictated by the gate bias. The eventual decline however is to be expected due to factors such as surface roughness which reduces the diffusive term. Since the overall υ_{inj} is determined by the smaller of the two quantities, the diffusive term now dictates the velocity. But then it is important again to note that this trend changes for gate lengths smaller than 40nm. For example the 30nm gate length HEMT device exhibits a monotonic roll off in υ_{inj} as a function of gate bias.

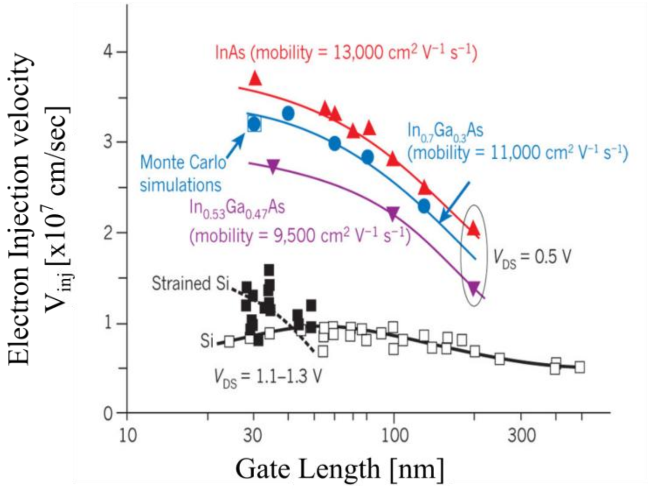


Figure 1-5. Injection velocity as a function of gate length measured from planar III-V HEMT devices [4].

Quantum capacitance bottleneck

An important consequence of introducing low effective mass channels however, is the emergence of the quantum capacitance[9-11]. Most of the electrons participating in transport in III-V semiconductors reside in the Γ-valley. Due to the spherical symmetry at the bottom of the Γ-valley there is a simple correlation between the transport and density of states (DoS) effective mass. Due to the low effective mass in these semiconductors, low DoS can limit the inversion charge. This is further exacerbated in the case of quantization where the conduction band splits into sub-bands with each having a constant two dimensional DoS given by $\frac{m_{i^*}}{\pi\hbar^2}$ where the index i refers to the ith sub-band. Quantization may be introduced by confinement druing heterostructure growth. It can also arise from band bending close to the gate oxide-semiconductor interface which is dictated by electrostatics in the on-state. As a consequence of these quantum effects the semiconductor capacitance reduces which in turn lowers the total gate capacitance at inversion. Figure 1-6 shows the two components of semiconductor capacitance which have to be taken into account in the quantum capacitance limit, viz. , the quantum capacitance arising from the finite 2-D DoS and the centroid capacitance which arises from the shape of the inversion layer charge distribution[11].

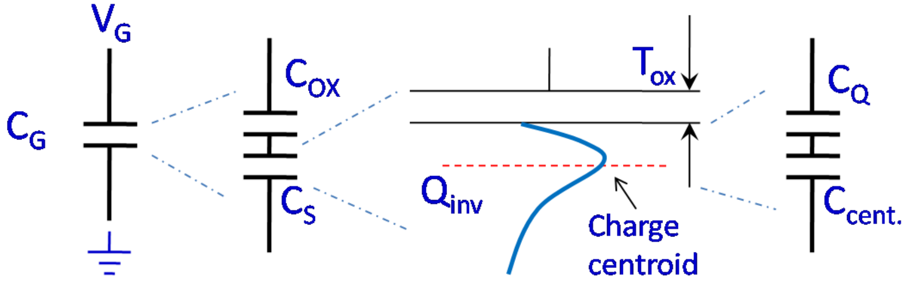


Figure 1-6. Equivalent circuit representation of quantum effects in the inversion charge. Low effective mass in III-V semiconductors results in more pronounced quantum effects.

Implementing highly scaled transistors dictates the need for ultra-thin quantum well (QW) structures to preserve the electrostatic integrity. However as outlined above this strong confinement leads to the quantum capacitance limit (QCL). High drive current requirement dictates that the new material system needs to provide high mobility at a reasonably high inversion charge density. Figure 1-7 shows the measured field effect mobility for various schottky gated $In_xGa_{1-x}As$ quantum well transistors as a function of the charge density in comparison to silicon MOSFETs. InGaAs provides considerably higher mobility (40x to 50x) at inversion charge density on the order of $5x10^{12}$ cm^{-2}, making it ideally suited for exploratory studies.

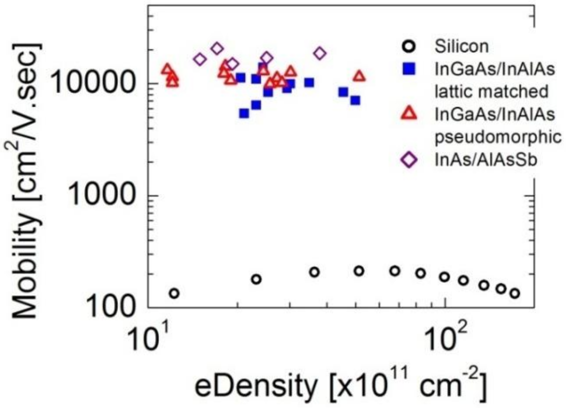


Figure 1-7. Electron mobility μ versus sheet electron density n_S in n-channel Si FETs and III-V QWFETs: Si MOSFETs, InGaAs/AlGaAs HFETs [12,13], lattice-matched InGaAs/InAlAs QWFETs [14,15], pseudomorphic InGaAs/InAlAs QWFETs [16,17], InAs/AlSb QWFETs [18,19].

II. Layer structure design and characterization

The general representative channel structure used throughout this dissertation work is shown figure 1-8. The device layer structure comprises of an $In_xGa_{1-x}As$ channel with varying Indium (In) percentage (x) grown pseudomorphically on a high bandgap $In_{0.52}Al_{0.48}As$ buffer layer. The wafers are grown by IQE inc. on 3-inch semi-insulating InP wafers using molecular beam epitaxy (MBE). Subsequent chapters explore the trade-offs inherent in engineering the channel in order to achieve higher performance.

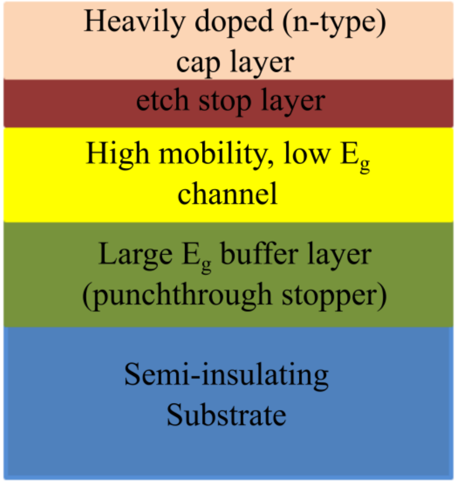
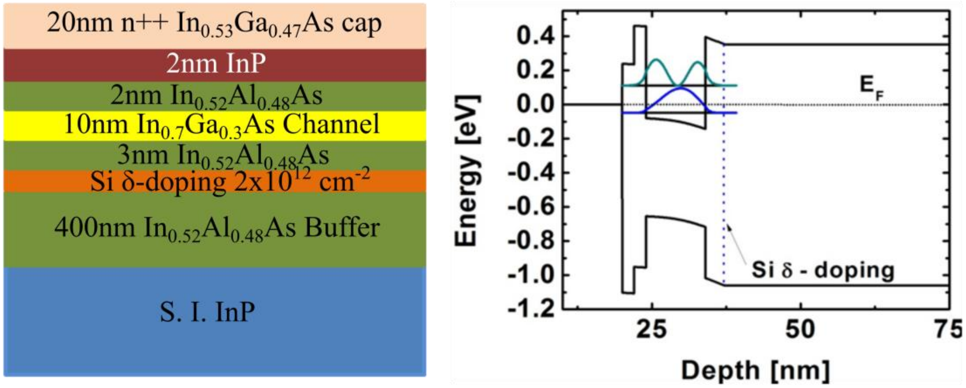

Heavily doped (n-type) cap layer
etch stop layer
High mobility, low E_g channel
Large E_g buffer layer (punchthrough stopper)
Semi-insulating Substrate

Figure 1-8. Schematic showing generic layer structure of quantum well channels investigated in this work.

Figure 1-9 (a) shows a particular example of layer structure used in this work. The channel comprises a 10nm $In_{0.7}Ga_{0.3}As$ quantum well (QW) grown pseudomorphically on the $In_{0.52}Al_{0.48}As$ buffer. The structure contains a silicon delta-doping layer to supply electrons to the channel and facilitate hall mobility measurements for characterizing the growth. A 20nm thick, heavily n-type doped ($2x10^{19}$ cm^{-3}) $In_{0.53}Ga_{0.47}As$ cap layer is grown in-situ to allow low resistance ohmic contact formation. The 2nm undoped InP layer serves as an etch stop layer during the recess etch of the heavily doped cap required to define the channel during device fabrication. Figure 1-9 (b) shows the corresponding band diagram obtained from self consistent 1-D Schrodinger-Poisson calculations using nextnano simulations.


20nm n++ In0.53Ga0.47As cap
2nm InP
2nm In0.52Al0.48As
10nm In0.7Ga0.3As Channel
3nm In0.52Al0.48As
Si δ-doping 2x10^12 cm^-2
400nm In0.52Al0.48As Buffer
S. I. InP
Energy [eV]
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
E_F
Si δ - doping
25
50
75
Depth [nm]

(a) (b)

Figure 1-9. (a) Device layer structure showing the remote δ-doping layer and (b) band diagram indicating the first two sub-bands and the Fermi level.

The presence of a delta-doping layer allows for easy characterization of the channel properties and growth quality through hall measurements. Figure 1-10 shows the temperature dependent hall measurement data for the $In_{0.7}$ $Ga_{0.3}As$ QW with room temperature mobility measured at 10,000cm^2/V.sec. Among the various scattering mechanisms modeled and quantified, polar optical phonon mechanism dominates at room temperature.

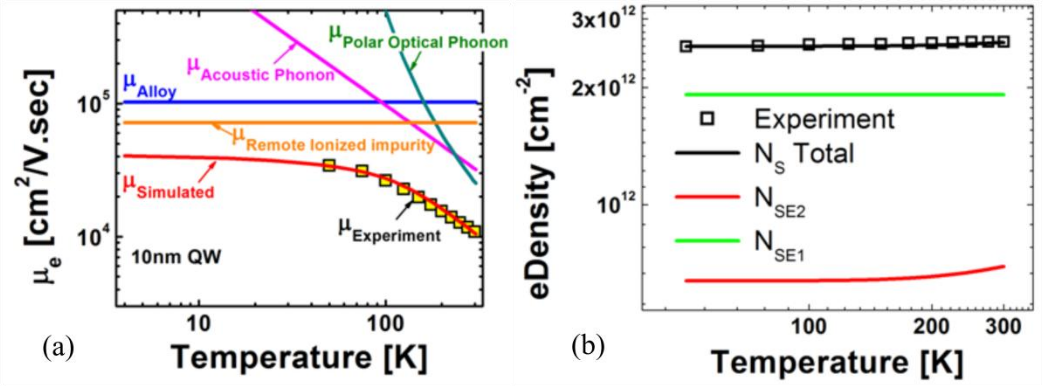


Figure 1-10. (a) Hall mobility measured as a function of temperature (symbols). Solid lines show fit obtained using a relaxation time approximation based scattering analysis. (b) Carrier density contribution from the first two sub-bands.

III. FinFET device architecture

Silicon CMOS technology scaling into the 22nm node and beyond requires advance multi-gate device architectures as planar devices do not provide the necessary short channel electrostatic control[20]. Several different architectures from silicon-on-insulator (SOI), double gate, tri-gate/FinFET to gate all around devices (GAA) have been demonstrated [20-22]. However in order to maintain feasibility from a manufacturing stand point industries have chosen to implement the 3-D tri-gate transistor or FinFET [20]. Figure 1-11 shows scaling trends for 22nm and 14nm technology nodes. Additionally, drive currents for present generation devices (14nm) are normalized to the layout width. With each successive generation, the fin pitch scales down while fin height scales up and thus FinFET architectures enable a new class of scaling where drive currents can approach close to 2mA/μm (normalized to layout width) at scaled nodes; values that are not attainable with traditional planar scaling.

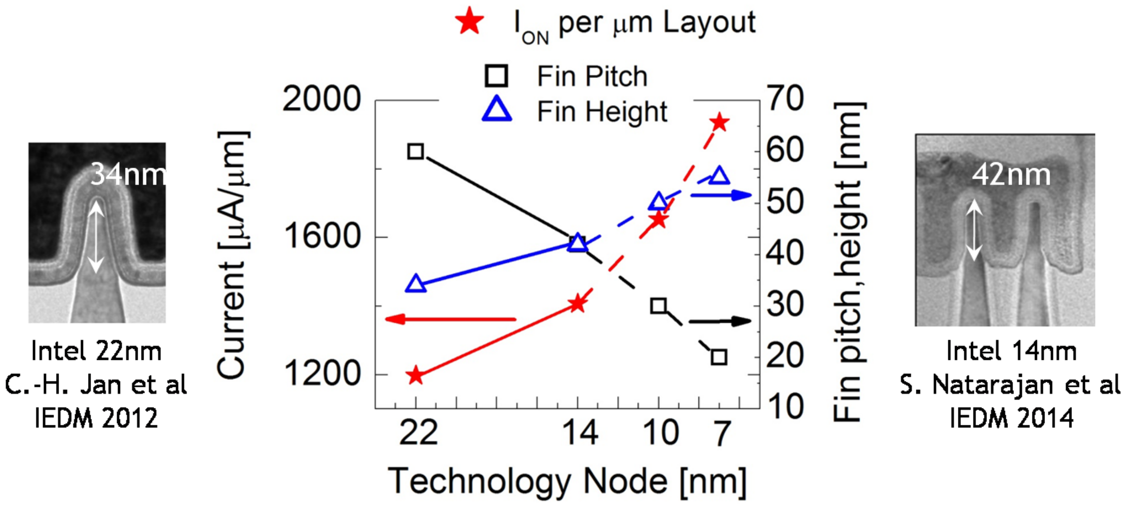


Figure 1-11. Scaling trends of drive current, fin width and fin height from 22nm node onwards. Values for the 10 and 7nm nodes are projected assuming simplistic scaling by 0.7x in dimensions per generation.

Thus in order to assess the feasibility of any new material system, it is essential to characterize and evaluate performance in multi-gate architectures. Results presented in this dissertation focus exclusively on the FinFET architecture as shown in the representative schematic of figure 1-12 for all experimental device demonstrations as well as simulations outlined in subsequent chapters. All devices are fabricated on layer structures grown by MBE on semi-insulating InP substrates as described in the previous section.

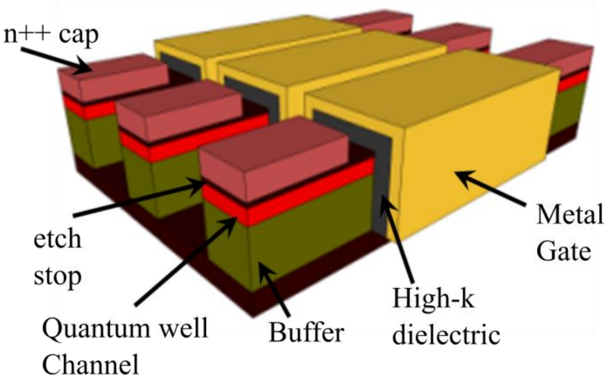

n++ cap
etch stop
Quantum well Channel
Buffer
High-k dielectric
Metal Gate

Figure 1-12. Representative schematic showing the architecture of the quantum well FinFET devices fabricated in this work.

Chapter 2

High-k dielectric gate stack integration on $In_xGa_{1-x}As$

Replacing silicon in the channel of the transistor is fraught with several difficult technical challenges. One of the most fundamental challenges, at the heart of transistor design, is the gate-stack which forms the control terminal of the device. While transistors have traditionally employed a Si/SiO_2 interface, the gate stack in modern transistors comprises of a very intricate atomic level interface between the channel semiconductor (silicon/non-silicon) material and high-k (high permittivity) dielectric based insulator. This is further followed by the deposition of a stack of metals which form the electrode. Different metals are used for the n-channel and p-channel devices in order to tune the workfunction suitably and achieve the desired threshold voltage for the device.

Demonstrating a high quality gate interface is necessary in order to achieve good electrostatic control and shut off the transistor. This is critical especially in low bandgap III-V semiconductors where it is necessary to control sub-threshold leakage current while simultaneously increasing on-current by exploiting higher electron mobilities. The quality of the high-k interface directly impacts the off-state leakage current and in-turn determines the viability of III-V materials as a new technology. Traditional MOSFETs which employ a Si/SiO_2 interface have several intrinsic advantages such as low interface state (trap) density, excellent thermodynamic stability and large band offsets to Silicon along with maturity of the technology[23].

However, demonstrating a high quality trap-state-free interface on non-silicon semiconductor materials such as III-V compound semiconductors with high-k dielectric is an extremely challenging problem[24-26]. III-V semiconductor surfaces in general exhibit more complicated reconstructions compared to Silicon. Further, GaAs surfaces suffer from gap states arising primarily from As-As dimers which have anti-bonding states in the bandgap just below the conduction band. However moving from GaAs to InGaAs and finally InAs, the bandgap reduces with lowering of the conduction band while the As-As anti-bonding states remain at the same energy. This could potentially provide some reprieve from the Fermi level pinning problem. However with changing group III composition, the interface states problem arises again this time due to uncoordinated group III (In/Ga) atoms and dangling bonds. Thus it is necessary to simultaneously suppress the formation of uncoordinated group III bonds as well As-As dimer states increasing the complexity of dielectric integration to InGaAs.

I. MOS capacitor fabrication and characterization: Thermal ALD

The first experimental demonstration of high-k dielectric integration on GaAs substrates implemented an Al_2O_3 layer grown using thermal atomic layer deposition[26] followed closely by studies with HfO_2 as well[27], due to their suitable band offsets and interface stability[28]. It was found that pulsing of the tri-methyl aluminum (TMA) precursor had a self-cleaning effect on the GaAs surface, by reducing the native oxides. Despite the demonstration of working FETs, the quality of surface passivation achieved was insufficient to demonstrate a scalable technology. However as indicated in the previous section by moving to higher Indium composition it might be possible to overcome limitations from Fermi level pinning to some extent due to the changing surface reconstruction and reduced impact of As-As dimers. In order to investigate this possibility MOS capacitors (MOSCAPs) were fabricated on $In_{0.53}Ga_{0.47}As$ (InP substrate). Samples were degreased in boiling Acetone and IPA followed by a 3 minute dip in 10:1 buffered HF acid (BOE). An Al_2O_3 passivation layer followed by HfO_2 layer was grown in order to scale the effective oxide thickness (EOT). EOT = $T_{High-K}*\varepsilon_{Si}/\ \varepsilon_{High-k}$, shows the equivalent thickness of SiO_2 required for the same capacitance density. Figure 2-1 shows experimental capacitance (CV) and conductance (GV) data for a bi-layer stack comprising 1nm Al_2O_3/3nm HfO_2 integrated on n-type $In_{0.53}Ga_{0.47}As$ (Si-doping; $5x10^{17}$ cm^{-3}) by thermal ALD. Platinum gate metal was deposited by e-beam evaporation followed by forming gas anneal (FGA) at 350°C for 20mins.

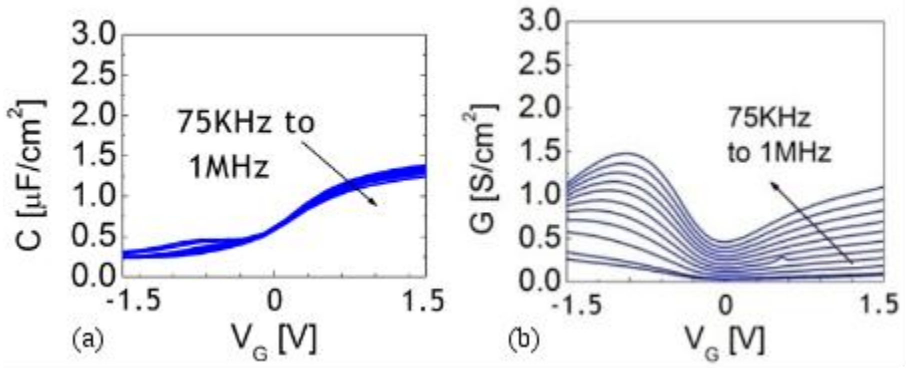
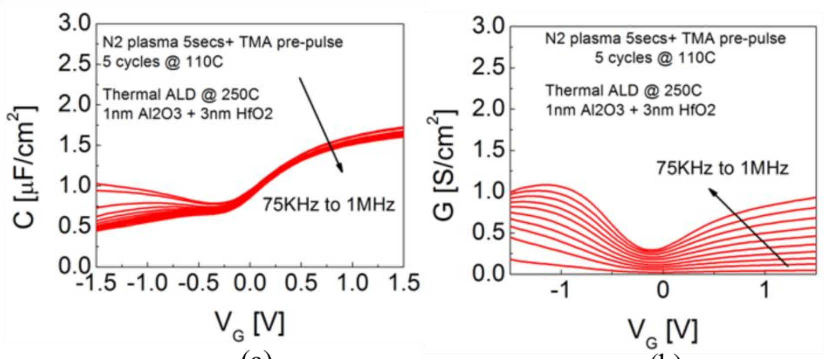


Figure 2-1. Experimental measured capacitance (a) and conductance (b) data for a 1nm Al_2O_3/3nm HfO_2 stack grown by thermal ALD.

The maximum capacitance density achieved using this bi-layer stack is limited to 1.4μF/cm^2. The large EOT is attributed to interfacial oxide growth during the anneal process caused by the presence of oxygen in the RTA furnace ambient. Additionally, large frequency dispersion is observed in the mid-gap and depletion regions. It should be noted further that the devices are characterized only down to a frequency of 75KHz. These results show that despite integrating Al_2O_3 as the passivation layer, there is significant frequency dispersion. Further the addition of HfO_2 did not result in scaled EOT.

II. Low power plasma surface clean and passivation

In order to improve the quality of the surface passivation low power plasma treatment was explored. Experiments were done on the Kurt Lesker LX 150 plasma ALD tool in the Penn State nanofab. Low power inductively coupled remote plasmas were generated by diluting Ar gas with various ratios of H_2 and N_2. All samples were degreased in boiling Acetone and then rinsed in IPA, followed by a 3 minute 10:1 BOE clean prior to loading into the loadlock of the ALD tool. Preliminary experiments indicated that H_2 plasma has a detrimental impact on the InGaAs surface resulting in poor CV modulation and large frequency dispersion. Whereas the primary function of the plasma clean is to reduce the native oxide, the H_2 plasma was found to induce surface damage hampering the subsequent nucleation of the ALD grown high-k layers. This surface damage was found to be present at all substrate temperatures from 100^oC to 250^oC. Subsequently N_2 plasma was investigated as outlined by Chobapattana et al in [29]. However in this work we explore highly diluted concentrations (5-10%) in an Ar ambient. The presence of Ar gas also facilitates quick plasma ignition at low powers facilitating short plasma pulse exposures. First set of N_2 plasma exposure experiments were carried out at relatively low temperature of 110^oC. Figure 2-2 shows the measured CV and GV characteristics for depletion-mode MOSCAPS fabricated with the above gate stack on n-type (Si-doping; $5x10^{17}cm^{-3}$) $In_{0.53}Ga_{0.47}As$.



(a) (b)

Figure 2-2. Measured capacitance (a) and conductance (b) data shown for MOS capacitor with bi-layer gate stack and low temperature (110C) plasma nitride surface passivation.

The surface is exposed to alternating pulses of N_2 plasma followed by TMA exposure. This was followed by thermal ALD growth of $1nmAl_2O_3$ and 3nm HfO_2. Additionally the gate metal was changed to thermally evaporated Nickel instead of e-beam Platinum. Special ceramic

(Alumina) coated Tungsten boats were used for the Nickel evaporation. Results immediately indicated better CV modulation along with reasonable frequency dispersion. The samples were annealed in FGA at 350C for 20mins. In order to avoid the presence of Oxygen in the annealing ambient the samples are annealed in the ALD chamber itself post metallization. The samples are transferred into the ALD chamber through the high vacuum loadlock into a H_2/Ar ambient which prevents unwanted oxidation of the surface. However the maximum capacitance density obtained was only around 1.5μF/cm^2 indicating that some native oxide still remains on the surface. This is also evidenced by the relatively large frequency dispersion in the mid-gap and depletion regions. Figure 2-3 shows the corresponding measured gate leakage current which is well controlled with the maximum current density value below 10mA/cm^2.

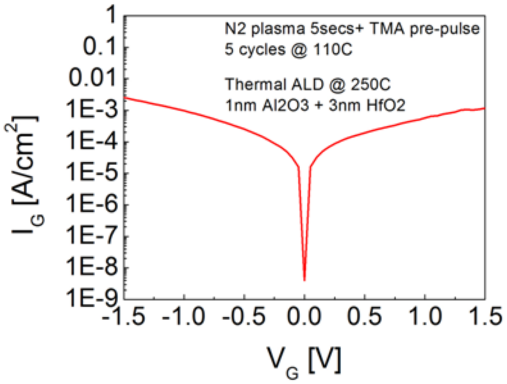


Figure 2-3. Gate leakage measured for MOS capacitor shown in figure 2-2.

The above results indicated that the N_2 plasma surface clean yielded reasonable CV characteristics thus providing a viable path for optimization. The reduced capacitance density and large frequency dispersion indicate the presence of residual native oxide and the absence of a pristine starting surface for nucleation of the high-k layers. In order to increase the effectiveness of the plasma clean one of the most critical parameters that needs to be modified is the substrate temperature. Various temperatures starting from 110°C upto 250°C were investigated for the N_2

plasma clean. Temperatures higher that 250°C were not investigated in order to limit the thermal budget and cooling/cycling times required for subsequent ALD high-k growth which was optimized at 250°C. The plasma surface treatment was performed as outlined before with alternating pulses of N_2 plasma followed by TMA exposure. The TMA pulse is shut off for the first cycle of exposure and is then followed by 5 cycles of the N_2 plasma/TMA sequence. Samples are metalized with thermally evaporated Nickel. In order to investigate the impact of the FGA CV was measured before and after post metal FGA. Figure 2-4 shows the CV and GV characteristics for MOSCAPs fabricated with the above surface clean performed at 250°C followed by thermal ALD of 1nm Al_2O_3 and 3nm HfO_2. The panels on the left show the corresponding characteristics for the samples without the FGA and on the right after FGA respectively.

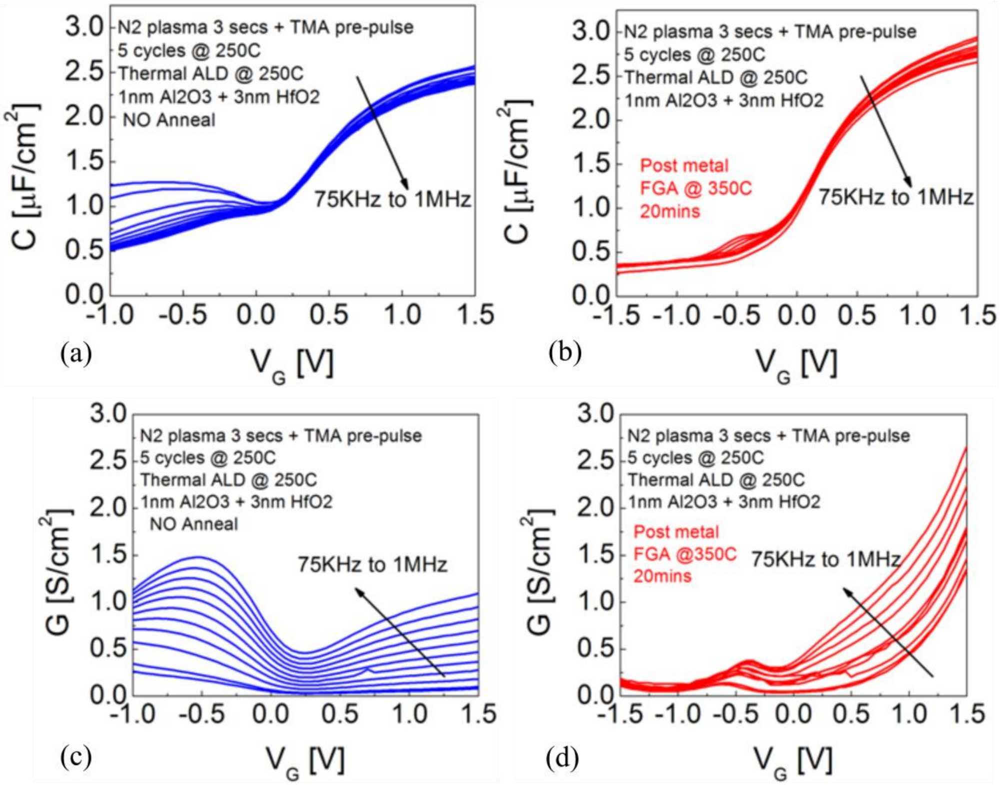


Figure 2-4. Effect of FGA on bi-layer dielectric gate stack with high temperature plasma nitrogen clean and passivation. (a) and (b) show capacitance data before and after FGA while (c) and (d) show the corresponding conductance data.

It is observed that the FGA has a significant impact on the frequency dispersion near mid-gap and depletion regions. Further the maximum capacitance density C_{max} also increases from 2.3 to 2.65μF/cm^2. The key results obtained with the high temperature (250°C) plasma clean in this experiment include the stabilization of the interface and higher C_{max} with simultaneously lower frequency dispersion. The capacitance equivalent thickness (CET) and equivalent oxide thickness (EOT) are measured at 1.3nm and 1nm respectively (corresponding to C_{max} of 2.65 μF/cm^2). The difference between the two thicknesses results from the finite thickness of the inversion layer which results in quantum capacitance. It is critical to account for this thickness especially in low mass systems as highlighted in chapter 1. Figure 2-5 shows the comparison of the gate leakage currents with and without FGA. The leakage current is observed to increase by nearly two orders post FGA. However the maximum current density was still measured to be less than 1A/cm^2 which is the maximum tolerance allowed for high performance CMOS devices.

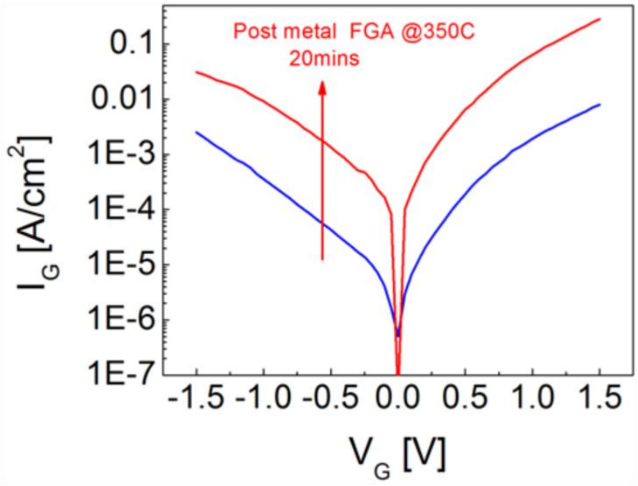


Figure 2-5. Gate leakage comparison before and after FGA for bi-layer dielectric gate stack.

In order to understand the nature of surface passivation and native oxide removal we further investigate the device using transmission electron microscopy. Figure 2-6 shows the cross-section TEM of the realized gate stack. It is observed that the alternating N_2 plasma/TMA sequence results in near complete removal of native oxide on the surface followed by growth of a 3A thick AlO_xN_y layer. This passivation layer also facilitates nucleation of the subsequent ALD oxide layers evidenced by the well controlled gate leakage which could be higher due to the presence of pinholes in the oxide layers and the Al_2O_3 layer in turn is used to facilitate nucleation of the HfO_2 layer.

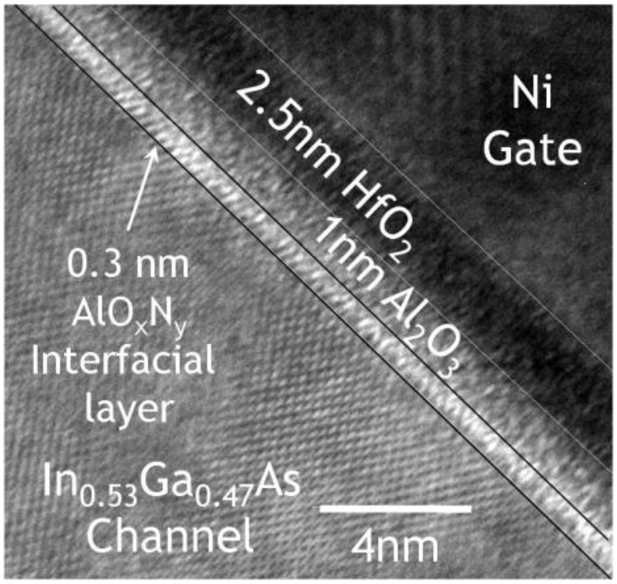


Figure 2-6. Cross-section transmission electron micrograph of the bi-layer dielectric gate stack with 250°C plasma nitride surface passivation.

Bi-layer vs. single layer dielectric gate stack

In order to fully exploit the benefits of a high mobility semiconductor channel it is necessary to increase the capacitance density of the gate stack so that higher inversion charge may be realized in the channel. Since the on current is directly proportional to the gate capacitance as $I_{ON} = \upsilon_{inj}C_G(V_G - V_T)$, a larger gate capacitance has a direct impact on the device performance. Further, for highly scaled devices it is important to maintain good electrostatic control for which a large C_G (C_{ox}) again helps to reduce the subthreshold swing (SS) which depends on the gate coupling ratio given as $\eta = (1+C_{dep}/C_{ox})$. Higher Cox may be realized by aggressively scaling the thickness of the high-k layers. However this will result in increased gate leakage. Instead, in our approach we migrate from the composite stack to a 4nm thick HfO_2 only gate stack. The thickness of HfO_2 is kept same as the composite gate stack in order to maintain the same order of gate leakage current. The passivation layer comprising of AlO_xN_y is retained and realized as described in the previous section. Figure 2-7 shows the comparison between the two resulting gate stacks. The panels on the left show CV and GV respectively for the composite (Al_2O_3/HfO_2) gate stack respectively whereas those on the right show the corresponding data for the HfO_2 only gate stack.

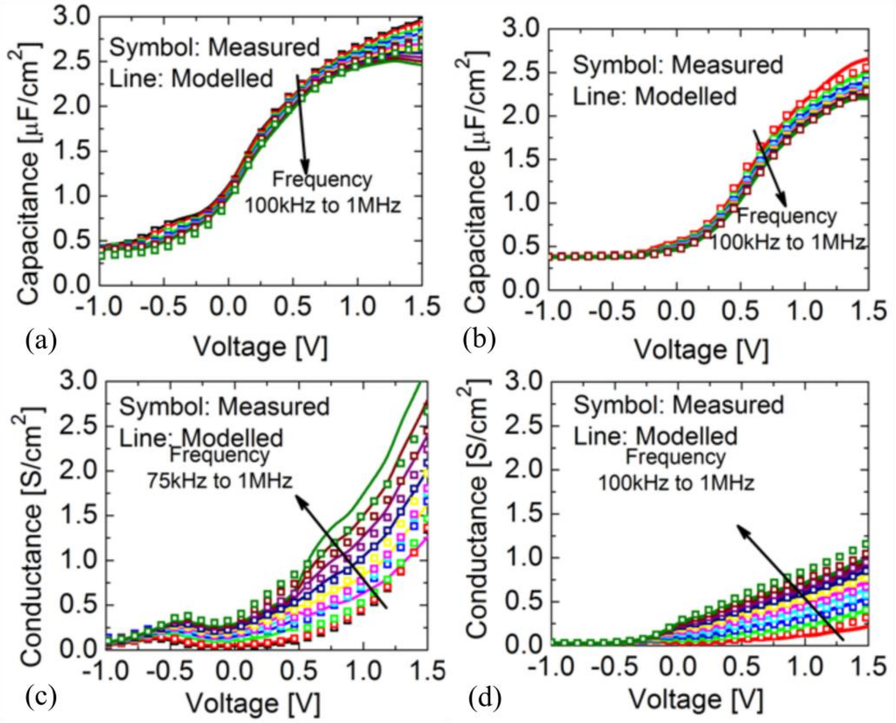


Figure 2-7. (a) and (b) show measured capacitance data (symbols) while (c) and (d) show measured conductance data (symbols) for bi-layer 1nm Al_2O_3/3nm HfO_2 and 4nm HfO_2 only stacks. Solid lines indicate fit from equivalent circuit model.

It is immediately observable that the AlO_xN_y passivation + HfO_2-only gate stack shows superior frequency dispersion characteristics in the mid-gap and depletion regions without significant change in C_{max}. This is further evidenced from the conductance (GV) characteristics of the devices. In order to compare and quantify the improvement in the passivation characteristics extract the interface state density (D_{IT}) for both the HfO_2-only as well as composite gate stacks. This is extracted using the modified conductance technique described in [30] where both the CV and GV characteristics are simultaneously modeled while also accounting for the gate leakage currents. The corresponding fit is also shown in figure 2-7 where the symbols show measured data points and solid lines indicate the output of the model. Good fit is obtained giving confidence in the extraction technique. Figure 2-8 shows the resulting D_{IT} and trap time constant profiles. It is observed that the HfO_2 only gate stack shows D_{IT} density around $1x10^{12}$ $cm^{-2}eV^{-1}$ starting close to mid-gap up to the conduction band (0V - 1.5V), which is almost order of magnitude lower than the composite gate stack. However both gate stacks show an increase in D_{IT} towards the conduction band and again from mid-gap onwards going into the lower half of the band gap.

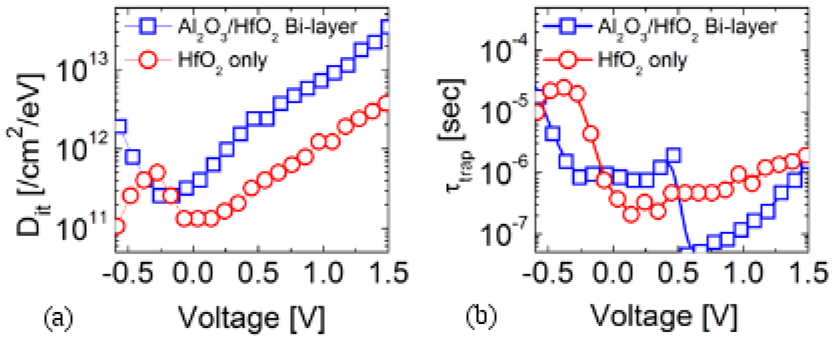


Figure 2-8. (a) Extracted D_{IT} profiles and (b) estimated trap time constant for bi-layer and HfO_2 only gate stacks.

From the above results it is clear that HfO_2 demonstrates better CV characteristics as compared to the composite Al_2O_3/HfO_2 stack. Further we note that the FGA has a significant impact in improving the CV characteristics irrespective of the dielectric integrated. However all FGA steps thus far were carried out post gate metal deposition. It is thus necessary to also investigate the impact of pre-gate metal anneal to see if CV characteristics can be improved. Figure 2-9 shows the comparison of pre-gate metal and post gate metal anneal on a AlO_xN_y/4nm HfO_2 gate stack. It can be clearly observed that while both devices show better characteristics compared to un-annealed samples (figure 2-4), post gate metal anneal displays lowest frequency dispersion in the mid-gap region evidenced by the much smaller bump in the CV characteristics. Negligible difference is observed between the C_{max} (2.6μF/cm^2) values as well.

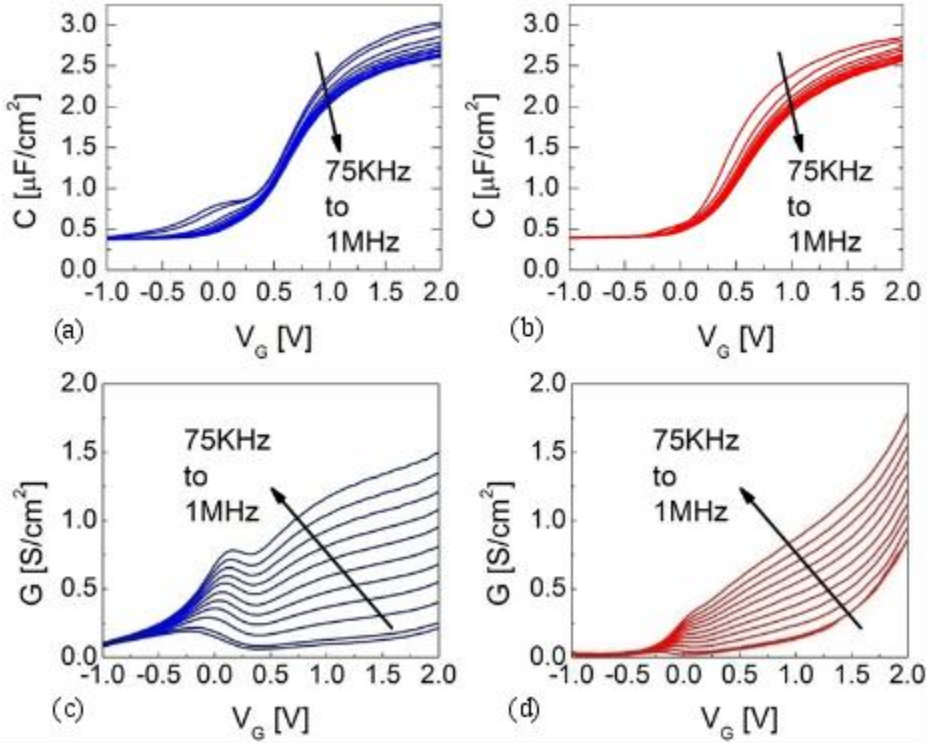


Figure 2-9. (a) and (b) show measured capacitance data for pre-gate metal and post-gate metal forming gas anneals. (c) and (d) show the corresponding conductance data. Devices comprise of 4nm HfO_2 only gate stack with AlO_xN_y passivation.

Apart from CV and GV characteristics, hysteresis was also measured for the two different anneals. Figure 2-10 shows the hysteresis measured with a forward-backward frequency sweep at 1MHz. Contrary to observations from CV characteristics we see that the pre-gate metal annealed samples display lower hysteresis of ΔV_{FB} = 100mV compared to 200mV for the post gate metal annealed sample. However because lower frequency dispersion takes precedence over reduced hysteresis, we continue to adopt the post-gate metal anneal strategy.

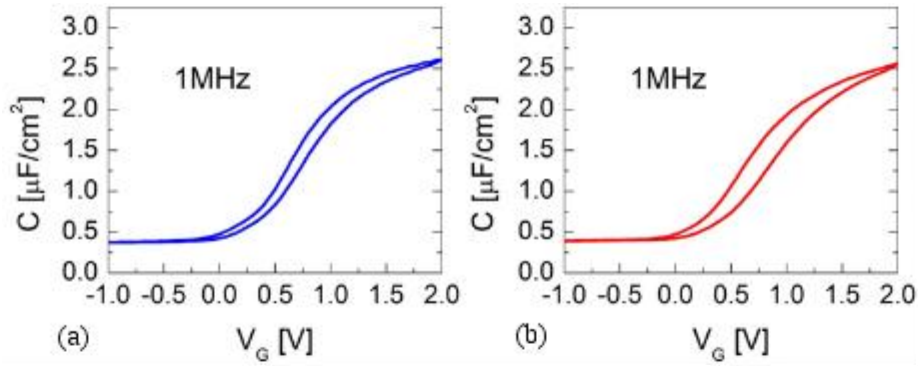


Figure 2-10. Measured hysteresis of (a) pre-gate metal and (b) post-gate metal FGA.

With the transition to HfO_2 only gate stack, the possibility of further scaling the oxide thickness in order to improve the C_{max} exists. Experimental splits were designed to investigate 3.5nm, 3nm and 2.5nm HfO_2 respectively. The AlO_xN_y interfacial layer remains unchanged as described in previous sections followed by thermal ALD growth of 3.5nm HfO_2. Whereas high gate leakage was observed for both the 3nm and 2.5nm thick HfO_2, the 3.5nm HfO_2 gate stack yielded good C_{max} along with well controlled frequency dispersion. Figure 2-11 shows the corresponding CV and GV characteristics. C_{max} of 3μF/cm^2 is obtained yielding CET of 1.12nm and EOT of 0.8nm, with well controlled frequency dispersion to 1KHz. A slight bump is observed near mid-gap which is exacerbated at 1KHz, indicating states within the bandgap.

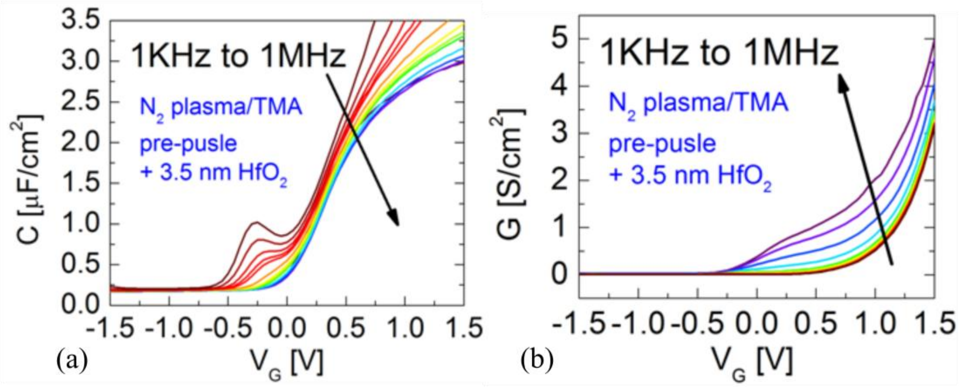


Figure 2-11. (a) Capacitance and (b) conductance data for 3.5nm HfO_2 only gate stack with 250°C plasma nitride surface passivation and FGA with frequency sweep down to 1KHz.

Gate leakage was measured as shown in figure 2-12. Maximum current density is limited to 300mA/cm^2 which is well within the limit of 1A/cm^2 allowed for state of the art CMOS.

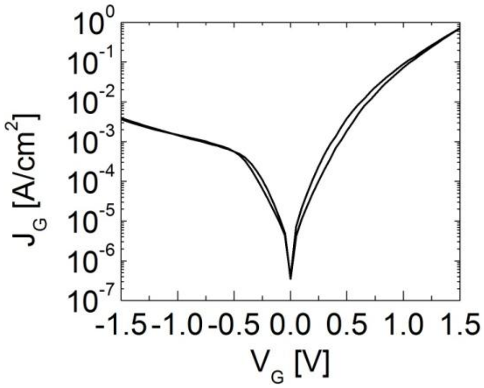


Figure 2-12. Gate leakage current measured for 3.5nm HfO$_2$ only gate stack.

III. Conclusions

As a final comparison we quantify the improvement obtained using the plasma nitride passivation with HfO$_2$ high-k dielectric as compared to the composite bi-layer gate stack integrated without the passivation layer. Figure 2-13 shows the comparison of DIT between the two schemes showing the significant impact from plasma nitride passivation.

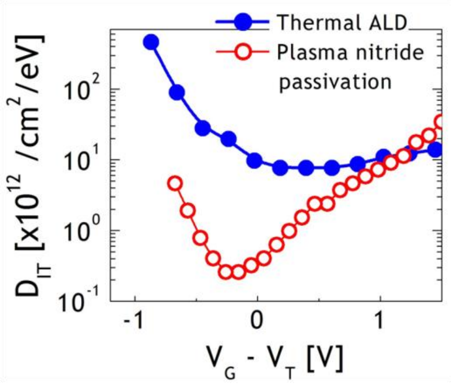


Figure 2-13. Extracted DIT profiles for thermal ALD and plasma nitride surface passivated gate stacks.

Although D_{IT} concentration approaches $1x10^{13}cm^{-2}eV^{-1}$ towards the band edges, there is close to two orders of magnitude reduction compared to the device with a thermal ALD-only gate stack not incorporating the passivation interlayer. Thus the new scheme of passivation followed by thermal ALD based high-k dielectric growth, developed in this chapter, is adopted as the baseline gate stack integration scheme for $In_XGa_{1-X}As$ devices demonstrated through the remainder of this thesis.

Chapter 3

Characterization of electronic transport in nanoscale multi-gate FETs

I. Introduction

In order to take advantage of high mobility materials while maintaining a high I_{ON}/I_{OFF} ratio in highly scaled transistors, multi–gate architectures have been introduced to achieve better electrostatic integrity [31,32]. Although the excellent transport properties of $In_XGa_{1-X}As$ (x = 0.53 – 0.70) is well established, top-down patterning of quantum well structures into strongly confined quasi-one dimensional (1-D) nanowire geometries can potentially degrade the transport properties. In this chapter we explore the effects of such patterning and confinement in InGaAs quantum wells on the electron transport properties through quantitative mobility measurements. A novel technique is developed for characterizing electron transport in tri-gate III-V nanowire FETs (NWFETs). We demonstrate NWFETs integrated with additional probe electrodes in Hall Bridge geometry to enable four–point measurements of both longitudinal and transverse resistance. This allows, for the first time, accurate extraction of Hall mobility and its dependence on carrier concentration in III-V NWFETs. It is shown that by implementing parallel arrays of nanowires, it is possible to enhance the signal to noise ratio of the measurement, enabling more reliable measurement of Hall voltage (carrier concentration) and mobility. Mobility is measured for various nanowire widths down to 40nm and a monotonic reduction in mobility is observable compared to planar devices. However, the NWFET mobility is shown to outperform state of the art strained silicon FETs. Finally, evidence of room temperature ballistic transport is also demonstrated in strongly confined III-V nanowire junctions through magneto-transport measurements in nanoscale Hall-cross structures.

Figure 3-1 shows a schematic illustration of the different extrinsic scattering mechanisms that can arise in such structures. Among the various factors that influence the mobility, the dominant mechanism is likely to be the scattering from the side walls [33,34]. It is therefore critical to understand whether the advantage of good electrostatic control in nanowire structures will come at the expense of reduced mobility, negating the benefits of using high mobility channel materials.

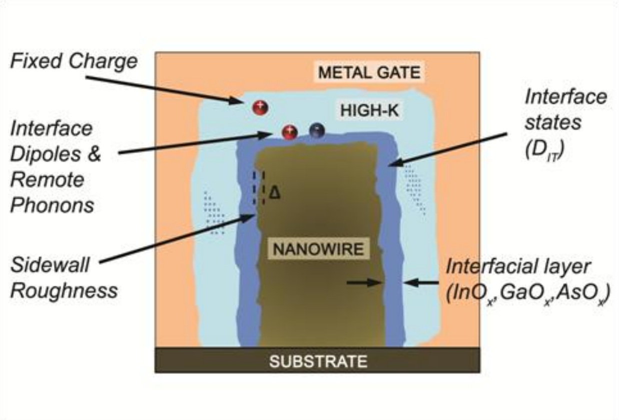

Fixed Charge
METAL GATE
HIGH-K
Interface Dipoles & Remote Phonons
Δ
Sidewall Roughness
NANOWIRE
SUBSTRATE
Interface states (D_{IT})
Interfacial layer (InO_x, GaO_x, AsO_x)

Figure 3-1. Schematic representation of various extrinsic scattering mechanisms that affect transport in multi-gate NWFET structures on III-V substrates.

Despite recent demonstrations of high performance NW based multi-gate devices including FETs and inverters [35,36] quantitative measurements of the true carrier concentration and mobility in such confined structures is lacking. Field effect based mobility estimates have been reported for III-V NW and planar devices [34,37]. This technique does not discriminate between the true mobile charge and contributions from charge occupying and emptying the interface trap states (D_{IT}). This inability to distinguish between the two types of charge typically results in an overestimation of channel charge leading to significant error in the estimated mobility. This is partially addressed by using the two FET method, as employed by Gunawan *et al* in [38]. However, quantitatively accurate measurements of the true mobile charge concentration may only be obtained through the use of Hall effect measurements as charges residing in interface traps do not respond to the applied transverse magnetic field [39,40]. Although this technique allows for accurate estimation of the true mobile charge concentration, reliable measurement of the Hall voltage remains challenging in the case of NWs due to their prohibitive geometry. These challenges have become apparent in several recent studies on Hall mobility in NW geometries for various semiconductors[41-43]. Storm *et al* [41] measure the Hall mobility in a single InP core-shell NW while simultaneously mapping the spatial distribution of carriers along the NW using cathodoluminescence. The dependence on the gate field (carrier concentration), however, was not studied in this case. Bloomers *et al* [42] measure the Hall effect mobility for a surface inversion layer in InAs NWs and were able to vary the carrier concentration by applying a gate voltage through a SiO_2 back gate. In both these studies, however, accurate measurement depends critically on the highly accurate lithographic placement of directly opposing Hall probe electrodes, which makes it prohibitive to study very small NWs (sub-100nm diameter). By realizing offset probes fabricated using angled contact deposition scheme DeGrave*et al* [43] were able to address this problem. Although this technique can be extended to ultra-thin NW like geometries, the need to contact the side walls with metal electrodes as well as the requirement for a thick insulator layer on top of the NW prevents the integration of scaled dielectric layers that is required to study technologically relevant semiconductor multi–gate NWFET architectures.

II. Test structure design and fabrication

In order to address the above mentioned shortcomings a novel test structure is designed. In this test structure, long channel multi-gate NWFETs are fabricated using NW arrays while simultaneously integrating Hall probe electrodes that extend between the NWs. The final structure forms a Hall bridge with two pairs of opposing Hall probes contacting each NW. Figure 3-2a shows a false color scanning electron micrograph (SEM) of the completed device for a NW width of 40nm. The channel comprises10nm thick $In_{0.7}Ga_{0.3}As$ quantum well (QW) with 4nm thick InP etch stop layer and 20nm heavily doped n-type $In_{0.53}Ga_{0.47}As$ cap layer to allow ohmic contacts to the metal electrodes. The n+ cap layer is first recessed, using a wet etch comprising citric acid and H_2O_2, in order to define the channel region. The etch mask is created by opening windows, using e-beam lithography, in diluted ZEP 520A resist to define the recess etch openings over the regions where nanowires will subsequently be patterned. The spacing between consecutive NWs in the array is 500nm. Accounting for the 50nm undercut from each side resulting from the wet etch, 400nm of heavily doped cap layer remains on the Hall probes in the regions between the NWs as seen in Figure 3-2b. The n+ cap layer in these regions maintains low resistance on the probes. Subsequently the entire structure comprising of the source/drain electrodes, NW array and Hall probe electrodes is patterned using e-beam lithography followed by dry etching in a chlorine (Cl) based plasma. This is followed by atomic layer deposition (ALD) of 1nm Al_2O_3 and 3nm HfO_2 high-k dielectric stack followed by Ti/Au gate metal electrodes, patterned using evaporation and lift-off process. The cross section of the NW is shown schematically in Figure 3-2c. Finally, a second metal stack comprising of Ti/Au is deposited on the n+ cap using lift-off to form the source/drain pads. Figure 3-2d provides a three dimensional perspective view of the gated nanowire region.

Device metrology and material characterization

Devices are fabricated with various NW widths starting from 1μm down to 40nm where we have quasi 1-D confinement. The gate lengths for all the test structures are kept constant at 2.5μm while maintaining a source to drain separation of 4 μm and Hall probe separation of 1.5μm.

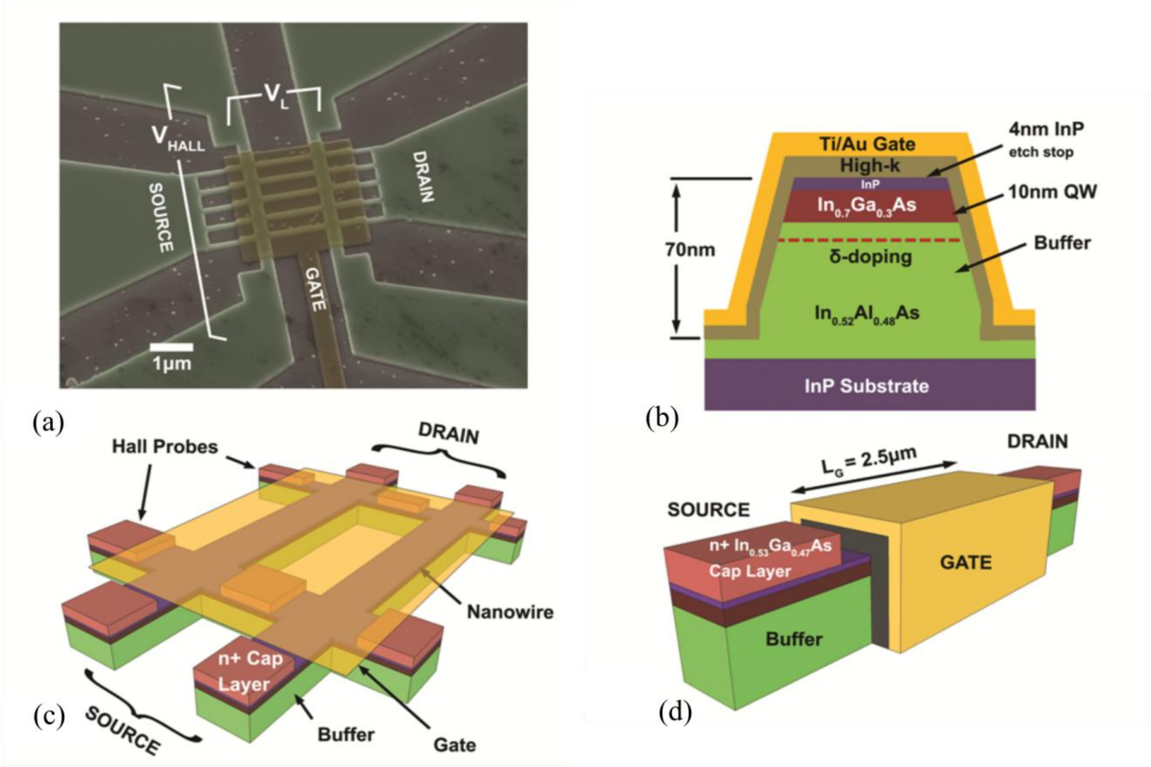

V_L
V_HALL
SOURCE
DRAIN
GATE
1μm
(a)
Ti/Au Gate
High-k
InP
In0.7Ga0.3As
δ-doping
In0.52Al0.48As
InP Substrate
4nm InP etch stop
10nm QW
Buffer
70nm
(b)
Hall Probes
DRAIN
Nanowire
n+ Cap Layer
Buffer
Gate
SOURCE
(c)
L_G = 2.5μm
DRAIN
SOURCE
n+ In0.53Ga0.47As Cap Layer
GATE
Buffer
(d)

Figure 3-2. (a) SEM of multi-gate, $In_{0.7}Ga_{0.3}As$ NWFET with an array of five NWs of width 40nm. Additional probe electrodes are integrated to form a Hall bridge structure. Hall voltage (V_H) and longitudinal (V_L) voltage are measured as shown. (b) Schematic cross-section of the NW showing the layer structure of the substrate. (c) Schematic showing two representative NWs and the structure of the probe underneath the gate. The heavily doped n+ cap layer is retained on the probe electrode in the regions between the NWs and (d) perspective view of single NW.

The quality and morphology of the side walls of the top-down etched structures is important due to the direct impact on the transport properties as explained in figure 3-1. We analyze the cross–

section of the etched NWs through transmission electron microscopy (TEM) and electron energy loss spectroscopy (EELS) as shown in figure 3-3. The side wall shows a smooth etch profile without any visible damage to the NW. A thin (~1nm) indium rich native oxide layer is left behind after etching. This layer is etched subsequently using a dilute Sulfuric acid treatment prior to deposition of high-k dielectric.

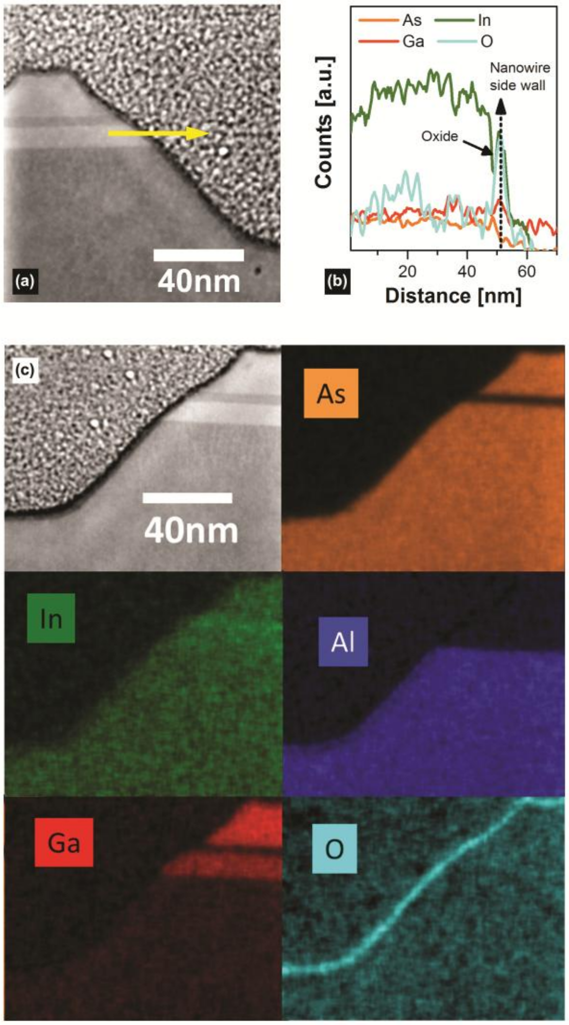


Figure 3-3. Characterization of the top down patterned nanowire subsequent to plasma etching: (a) STEM image indicating location of EELS line scan and (b) EELS line scan showing possible formation of indium oxide interfacial layer. (c) EELS elemental map of NW cross-section confirming that top-down etch yields high quality, defect-free NW structure, while also highlighting the formation of native oxide layer on side walls.

III. Results and discussion

We demonstrate an interface with good electrical properties as evidenced by NWFET characteristics shown in figure 3-4. The measured transfer characteristics for the long channel array device with 40nm NW width shows a sub-threshold slope of 85mV/dec. Nearly zero drain induced barrier lowering (DIBL) indicates good electrical isolation between source and drain pads for the$In_{0.7}Ga_{0.3}As$ quantum well structure, aided by the presence of InAlAs barrier layer. Output characteristics show good saturation with a peak drive current of around 250μA/μm.

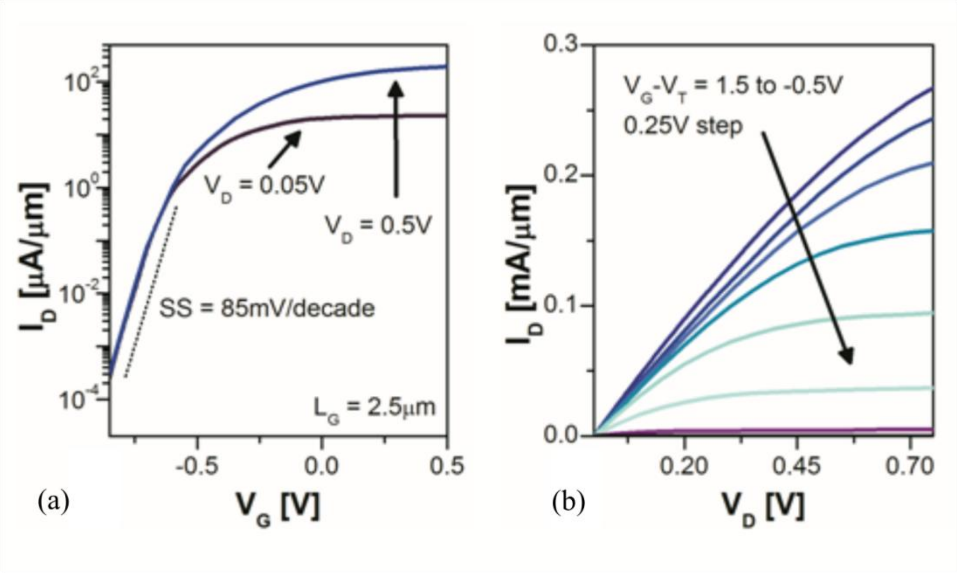


Figure 3-4. (a) Transfer characteristics (I_D-V_G) of the multi-gate NWFET based Hall structure, with NW width of 40nm, showing good sub-threshold behavior. (b) Output characteristics (I_D-V_D) of the same device.

Device characterization

Hall measurements are performed in a Lakeshore TTP6 cryogenic probe station equipped with a superconducting magnet that can generate magnetic fields upto 2.5T. The magnetic field is oriented perpendicular to the substrate for all the measurements carried out in this study. We point out here that since the test structures are FETs constructed in a Hall bridge geometry, it is possible to modulate the carrier concentration in a well-controlled fashion using the gate. The gate voltage (V_{gs}) is varied from -0.5V to 1V in steps of 50mV, while applying a fixed source to drain bias (V_{ds}) of 50mV. The Hall voltage (V_H) is then recorded at magnetic field values of ± 0.5T and ± 0.8T for each gate bias, and the measurements are then repeated after reversing the polarity of V_{ds}. Thus the Hall voltage from each pair of probes is written as

$$V_H = (V_{H1} + V_{H2})/2$$

where,

$$V_{H1} = \frac{1}{2}\left[\frac{V_T(+B_1) - V_T(-B_1)}{2} + \frac{V_T(+B_2) - V_T(-B_2)}{2}\right] \text{at } V_{ds} > 0$$

$$V_{H2} = \frac{1}{2}\left[\frac{V_T(+B_1) - V_T(-B_1)}{2} + \frac{V_T(+B_2) - V_T(-B_2)}{2}\right] \text{at } V_{ds} < 0$$

The entire set of measurements is then repeated using the second pair of Hall probes. The results are thus averaged over the set of 16 measurements to yield each data point in order to ensure reliable estimates as well as rule out measurement artifacts arising from geometrical asymmetries such as unintentional offsets between Hall probes.

The measurements are performed using a Keithley 4200 semiconductor characterization system with five source measure units (SMU). Two SMUs are configured as voltage sources, connected to the gate and the drain respectively while the source terminal is grounded. These SMUs determine the bias conditions for the FET while the three remaining SMUs are configured as nano-voltmeters to measure the Hall and longitudinal voltages respectively. The gate voltage dependence, as shown in Figure 3-5a, shows a roll off with increasing carrier density (gate electric field) in the channel. Figure 3-5b shows that the mobility reduces monotonically as the NW dimensions are scaled down, indicating that scattering increases as a result of confinement.

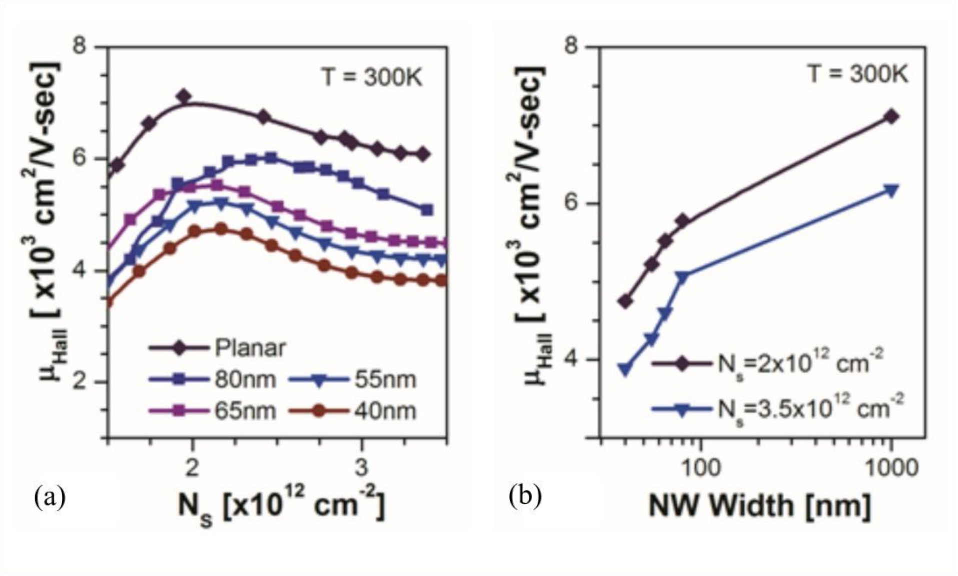


Figure 3-5. Experimentally measured Hall electron mobility for multi-gate NWFET of Figure 3-2 as a function of carrier concentration for different NW widths. (b) Mobility shows a monotonic roll-off with reducing NW dimensions. Peak mobility is observed at a carrier concentration of $2x10^{12}$ cm^{-2} for all NW widths.

Equivalent circuit model and mobility extraction

Devices with nanoscale channels are particularly challenging in the context of Hall measurements due to the small currents involved, which results in low values of Hall voltage for practically accessible magnetic field strengths. We show that by increasing the number of NWs in parallel, it is possible to generate larger Hall voltages as the contribution from each NW add in series.

In order to correctly interpret the observed potential and charge distribution, we construct an equivalent circuit model as shown in Figure 3-6. The Hall voltage produced by each NW is represented as a gate voltage dependent voltage source since the carrier concentration (Hall voltage) is controlled by the gate voltage, and the longitudinal resistance of each NW is represented by a corresponding voltage dependent resistor.

Since the Hall voltage is measured using a high impedance voltmeter, current does not flow perpendicular to the NWs ruling out resistive voltage drops along the Hall probe. The total Hall voltage may then be written as the addition of individual Hall voltages produced from each NW. This is in turn a function of the current (which in turn depends on gate bias) and the magnetic field. For a uniform array of NWs, assuming that the dimensions for each NW are the same, the average Hall voltage per NW maybe expressed as

$$V_H = \frac{1}{N}\sum_{i=<N>} V_{Hi}$$

where, N is the number of NWs in parallel. Next, in order to measure longitudinal resistance the magnetic field is turned off. The SMUs configured as high impedance voltmeters are now connected to a pair of probes spaced along the NW. The resulting structure then forms a parallel resistor network so that the total longitudinal resistance for each NW may be expressed, in units of ohm/square, as

$$\frac{1}{R_{SH}} = \frac{1}{N}\frac{L}{W_{NW}}\sum_{i=<N>}\frac{1}{R_i}$$

where the effective width is taken to be the active perimeter of the NW which includes the top width and side walls, i.e., $W_{NW} = W_{Top} + 2H_{SW}$. It should be noted that here L refers to the distance between the hall probes. For a given hall voltage the sheet carrier density may then be evaluated, in units of cm^{-2}, as

$$n_S = \frac{IB}{qV_H}$$

ultimately yielding the average mobility for an individual NW in units of cm^2/V-sec as

$$\mu = \frac{1}{qR_{SH}n_S}$$

The mobility trends observed in figure 3-5a and b for the NW geometries are estimated using the above procedure and clearly depict a mobility roll-off with reducing NW width.

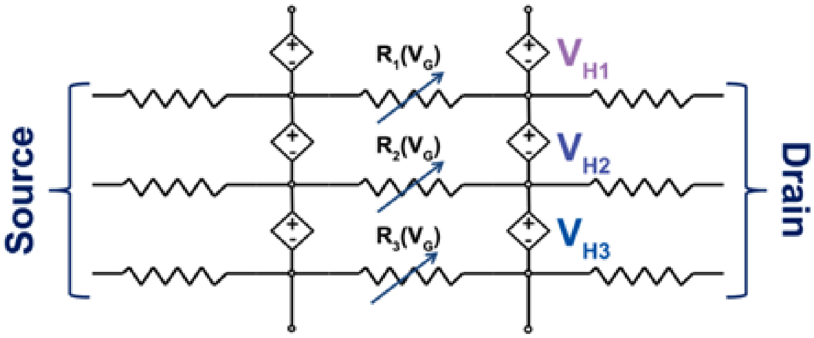
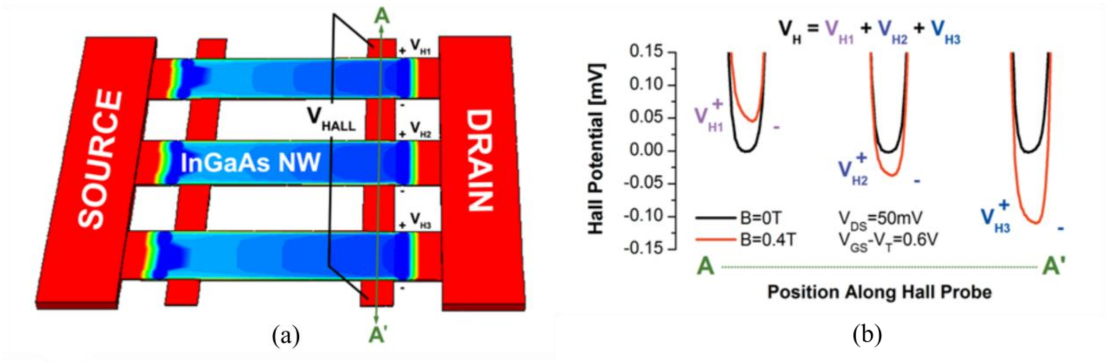


Figure 3-6. Equivalent circuit model for interpretation of measured results.

Validation of measured results

In order to understand the validity of the above method, we simulate the entire structure using Sentaurus TCAD [44]. Three dimensional drift diffusion simulations are performed while incorporating a density gradient approximation model to capture the quantization effects. The simulations are first calibrated to the measured I_D-V_G characteristics of Figure 3-4. We subsequently add the Magnetic field dependence in the simulation using the built-in drift-diffusion-based model for current densities augmented by magnetic-field-dependent terms corresponding to the Lorentz force on the motion of the carriers [44]. Figure 3-7a shows the potential profile in the structure with an applied magnetic field of 0.8T. Figure 3-7b shows that the Hall potentials generated from each NW add in series, resulting in a larger V_H across the entire structure.



(a) (b)

Figure 3-7. Improving measurement accuracy: (a) simulated electrostatic potential through the NW array. (b) Simulated Hall potential profile plotted as a function of position along the Hall probe. The results show that each NW contributes a small Hall potential to the total V_H, which is measured across the device.

In addition to the numerical validation, we validate results experimentally by comparing the mobility between single NW and multi–NW devices as shown in Figure 3-8a. We compare the results for a single NW and array device at a NW width of width of 120nm so that reliable measurements of Hall voltage may be obtained even for a single NW device. The results agree closely thus validating the equivalent circuit model based interpretation. Figure 3-8b compares the error between the expected total Hall voltage and that calculated by summing the values from individual NWs. We note that, as the number of NWs is increased, the percentage error in the estimated Hall voltage reduces and the accuracy of mobility extraction increases. The experimental structures presented in this study implement NW arrays with five NWs in parallel for which the estimated error is lower than 1%. The excellent agreement between the results from different methods, extracted independently of each other, sufficiently validates the technique presented in this work for gated Hall measurement in NWs.

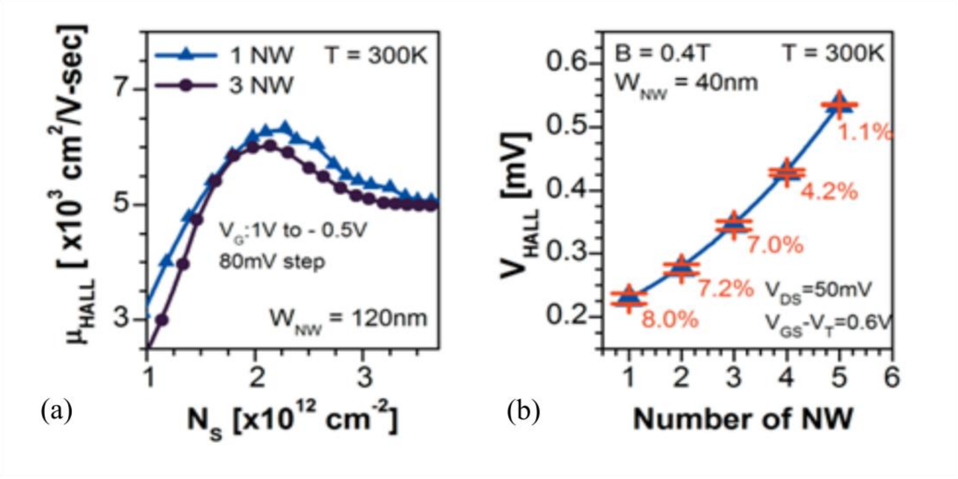


Figure 3-8. (a) Measured Hall mobility as a function of carrier concentration showing excellent agreement between single NW and NW array device. Results confirm the equivalent circuit based interpretation of Fig. 3-6. (b) Simulated Hall voltage as a function of NW number showing an increase in measured Hall voltage with number of NWs. Due to the averaging effect of the array, the error of the measured Hall voltage reduces as number of NWs increases. Error is less than 1% when the number of NWs is increased beyond 5.

Identifying the cause for mobility degradation

An important question however, is the value of mobility that can be expected at smaller NW widths close to or below 10nm which is of significant technological interest. Simulation studies for silicon NWs [45,46] show that the additional scattering present in thin NWs can be effectively modeled by potential fluctuations resulting from surface roughness. This type of scattering is a temperature independent phenomenon. Experimental evidence for such surface roughness based scattering has also been observed in ballistic InAs NWFETs[47]. Figure 3-9a shows the results of un-gated (V_G = 0)Hall measurements for top-down patterned NW devices. These measurements are possible as the devices operate in the depletion mode with normally-on. Mobility reduction from reducing the NW width is extracted by Mathiessen's rule using the mobility of the planar structure as the reference. Thus we write $\frac{1}{\mu_{SW}} = \frac{1}{\mu_{Planar}} - \frac{1}{\mu_{NW}}$ where, μ_{SW} refers to the side wall roughness limited mobility. Figure 3-9b shows clearly that the scattering limited mobility exhibits temperature independence. Similar mobility trends have also been observed for vapor-transport grown InGaAs NWs[48] measured using the field effect technique which suffers from drawbacks highlighted before.

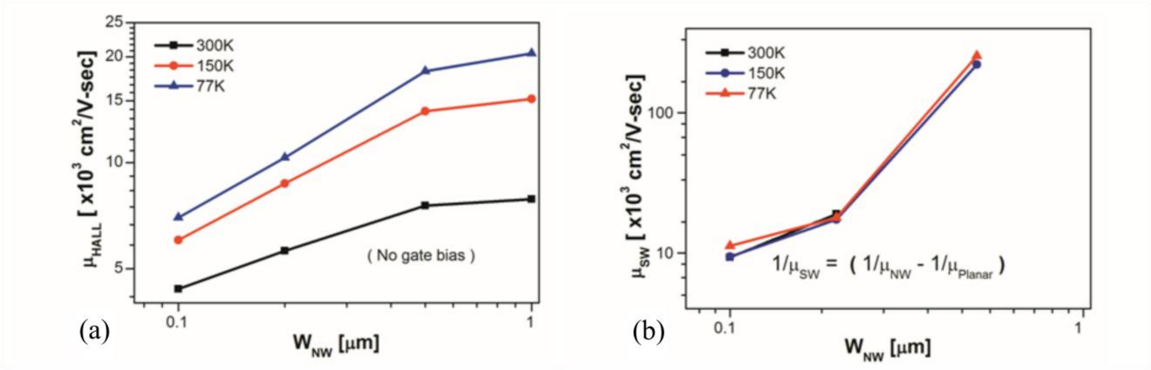


Figure 3-9. (a) Temperature dependent, un-gated Hall measurements showing mobility degradation as the InGaAs quantum well is patterned into NWs with width reducing from 1 μm (planar) down to 100nm. (b) Additional scattering extracted using Mathiessen's rule. This component is found to be independent of temperature clearly indicating that the scattering arises from roughness due to the side walls.

It is clear that the mobility degradation trends observed in this work can be attributed to surface roughness induced scattering resulting from the side walls. In order to quantify the impact of this scattering we assume an exponentially correlated roughness and calculate the associated scattering rate using the approach in[47]. The matrix element (overlap integral) however, is estimated using the model in [45] which accounts for the dependence of wavefunction spread and effective electric field on the NW size The theoretical model of surface roughness (SR) scattering is based on the expression used by S. Chuang et. al. [47]. SR is assumed to have an exponentially decaying autocorrelation function [45], and the collision rate due to SR scattering is expressed as

$$\Gamma_{nm}^{sr} = \frac{4\sqrt{2}e^2 D(E)}{\hbar}\frac{\Delta^2\Lambda}{2+{q_x}^2\Lambda^2}|F_{nm}|^2(1-cos\theta)$$

where D(E) is the one-dimensional final density of states for the scattered electrons. Here Δ is the rms height and Λ is the correlation length for the roughness. The difference between initial and final wave vectors is given as$q_x = (k_x \pm k_x^{1})$, assuming that the nanowire is oriented along the x-axis. For quasi 1-D transport assume that θ = π, since only backscattering is allowed [47].

However, for calculating the overlap integral,F_{nm}, we use the form reported in [45] which captures the dependence of the overlap integral on the dimensions of the nanowire. Thus we write

$$F_{nm} = \iint dydz \begin{bmatrix} -\dfrac{\hbar^2}{eWm_y}\psi_m(y,z)\dfrac{\partial^2\psi_n(y,z)}{\partial y^2} + \psi_n(y,z)\epsilon_y(y,z)\left(1-\dfrac{y}{W}\right)\psi_m(y,z) \\ +\psi_n(y,z)\left(\dfrac{\varepsilon_m-\varepsilon_n}{e}\right)\left(1-\dfrac{y}{W}\right)\dfrac{\partial\psi_n(y,z)}{\partial y} \end{bmatrix}$$

where, W is the width of the nanowire. The wavefunctions $\psi_i(y,z)$ as well as electric field profile along the width of the wire, $\epsilon_y(y,z)$, are obtained through a self-consistent Poisson-Schrodinger solver, nextnano, for various nanowire widths[49]. The first term in the overlap integral depends only on the confinement and thus accounts for roughness scattering even at low gate fields, while the second and third terms account for scattering at higher transverse fields from the gate. Here we ignore the third term since we are dealing with low field transport and the energy separation between the first two sub-bands itself is on the order of 100meV, as verified from the nextnano simulations. Therefore we account only for intra-subband scattering. The electric field (potential) profile and electron wavefunctions are extracted at the appropriate gate bias required to realize inversion charge of 2x10^{12} cm^{-2}, for each nanowire width. The evaluated overlap integral F_{nm}, for each NW width is summarized in Figure 3-10.

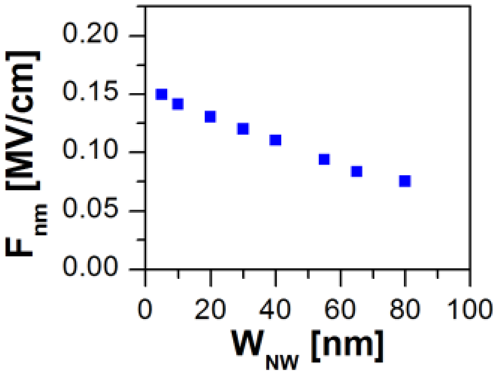


Figure 3-10. Calculated overlap integral values F_{nm} as a function of NW width at N_S = 2x10^{12}. The effective electric field from the potential fluctuations of the sidewall roughness increases monotonically with reducing NW width.

The scattering rate equation lumps together the remaining terms into a single parameter, which is fitted to the experimental data. Here, we assume that the scattering rate from the top surface is independent of NW width since this roughness corresponds to the MBE-grown interface and should have the same scattering dynamics across all NW widths. Thus, we can model the total contribution from roughness based scattering and extract the NW mobility, using Mathiessen's rule as

$$\frac{1}{\mu_{NW}} = \frac{1}{\mu_{Planar}} + \frac{1}{\mu_{SW}}$$

where,μ_{Planar} is the mobility corresponding to the planar device and is assumed to be the starting mobility value.

Since the top surface is capped with the InP layer, it is reasonable to assume that the scattering contribution from this MBE grown interface remains constant with NW width while that from the side walls scales commensurately.Figure8 projects the mobility values for NWs at

10nm width after calibrating the model to the measured results. We note that these values are significantly higher compared to what is projected for silicon NW transistors with similar dimensions [33].

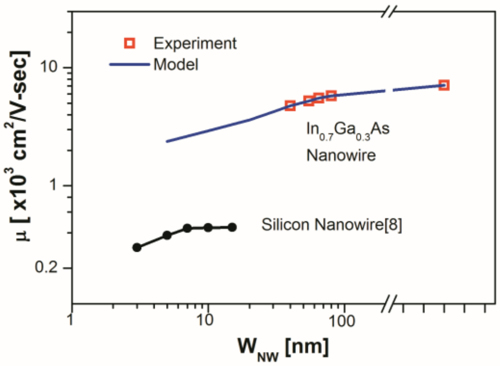


Figure 3-11. Projection of mobility for 10nm NW width. The side-wall roughness scattering model is calibrated to the measured data. The gated Hall measurement results clearly indicate a mobility roll-off with decreasing nanowire width. Comparison of mobility between InGaAs and silicon NWs indicates over 10× higher mobility for InGaAs at a NW width of 10nm.

IV. Ballistic transport in nanoscale $In_xGa_{1-x}As$ structures

In the context of low power digital CMOS technology at scaled gate lengths however, it is important to investigate if ballistic transport is achievable at room temperature in such top down patterned III-V NWs. Ballistic transport is conducive to higher drive currents at low supply voltages[50]. Ballistic transport is observed in two dimensional electron gases in the seminal works published on the Quantum and Fractional Quantum Hall effects[51][52]. These effects however, are observed only in the presence of strong magnetic fields. Zero longitudinal resistance has been demonstrated however, in a NW type of structure by Picciotto *et al* [53]. A confined nanowire type geometry is created by a cleaved edge overgrowth technique, while simultaneously connecting to the planar 2-dimensional electron gas (2-DEG). In order to realize contacts, probes are defined by depleting the planar 2-DEG in selected regions using gate electrodes, enabling the demonstration of zero longitudinal resistance through a four-point measurement. However measurements are performed at very low (mK) temperatures. Despite these experimental demonstrations of ballistic transport, little direct evidence exists for observing such phenomena at room temperature. At room temperature phonon scattering dominates transport, significantly reducing the mobility, quenching all observable effects. Scattering free transport at room temperature may perhaps be accessible with more exotic materials such as topological insulators[54].

For semiconductor NWFET structures ballistic transport has been demonstrated through quantized conductance measurements [47][55] where conductance plateaus are observed using a traditional two probe configuration. These results have been observed at significantly higher temperatures of upto 190K for NWs with diameters smaller than 25nm [47]. At 120K and L_G = 60nm, significant fraction of carriers participating in the transport are ballistic with the corresponding carrier mean free path (λ) of 170nm. For NWs demonstrated here, accounting for the reduction in mobility with NW width, we project that for a width of 10nm the mobility is close to 3000 cm^2/V-sec at room temperature. This corresponds to a mean free path of λ = 60nm, calculated as $\lambda = \upsilon_F \tau$, where we assume υ_F to be the Fermi velocity and τ is the life time estimated from mobility. Thus at room temperature, it may be possible to observe ballistic transport only over very short length scales.

We then investigate ballistic transport in the top-down patterned NWs by constructing a nanoscale Hall cross structure as shown schematically in Figure 3-12. This structure allows four point resistance measurement. However, instead of measuring traditional longitudinal four point resistance we measure the so called bend resistance [56,57] in these structures as a function of both magnetic field and temperature. The bend resistance is defined as $R_B = V_{21}/I_{34}$ as shown in Figure 3-12b. If the transport in the junction is predominantly diffusive, the electrons undergo scattering within the junction and the voltages measured between probe 1 and 2 simply correspond to the resistive drop as current flows diffusively from probe 3 to 4, producing $V_{21} > 0$. In the case of ballistic transport, as seen in Figure 3-12b, electrons leaving terminal 4 do not scatter within the junction. This allows them to conserve momentum, travel past probe 3, and accumulate in probe 2. A negative potential then builds up in probe 2 relative to probe 1 thus reversing the polarity of V_{21}. By definition, R_B is negative thus giving rise to negative bend resistance. In the presence of a magnetic field the electrons are forced to curl back into contacts 3 or 1, thus restoring a positive value for V_{21}. In such a configuration, ballistic transport may be observed more easily than in a longitudinal resistance measurement as the behavior of the carriers is probed over a very small length scale. Negative bend resistance provides a clear signature for ballistic transport through the nanowire junction.

Figure 3-12c shows the experimental results in NW Hall cross structure. The device consists of a junction of two perpendicularly oriented NWs each of width of 100nm which is comparable to the mean free path at the smallest NW widths. A current of 100nA is forced from probe 3 to 4 while measuring the voltage difference between probes 2 and 1. As seen the resistance shows a negative peak at zero magnetic field and diminishes as the magnetic field is ramped in either direction. Further, we note that the magnitude of the peak reduces with increasing temperature due to the onset of phonon scattering which pushes the carriers into the diffusive regime. We observe negative bend resistance is present up to room temperature clearly indicating that a significant fraction of the carriers exhibit ballistic behavior at room temperature. We conclude that despite the significant deterioration resulting from side-wall roughness scattering in these NWs, the mobility is still significantly high so that a large fraction of electrons participating in transport are ballistic over short lengths at room temperature. These results indicate that III-V NWs are likely to behave as ballistic channels over short lengths which will allow high drive currents in short channel NWFETs at reduced supply voltages.

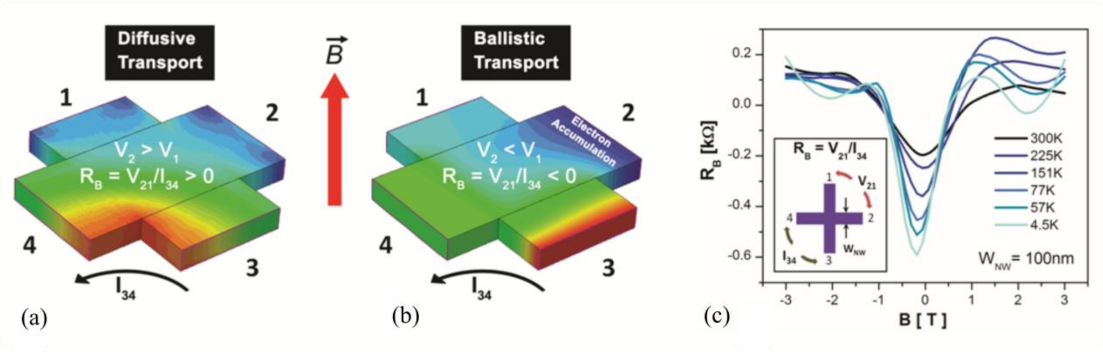


Figure 3-12. (a) Simulated potential distribution profile for a Hall cross structure in the diffusive regime. As current flows from contact 3 to 4 a positive bend resistance is measured. (b) Potential distribution in the ballistic regime. Electrons from contact 4 travel past the junction without scattering and start accumulating in contact 2 reversing the polarity of V_{21} giving $R_B < 0$. (c) Measured bend resistance as a function of magnetic field and temperature for $In_{0.7}Ga_{0.3}As$ NW based Hall cross structure. $R_B < 0$ is observed at room temperature confirming ballistic transport at room temperature.

V. Conclusions

In this chapter we demonstrate a novel technique for accurate measurement of charge and mobility in III-V NWFETs. Gated Hall measurements accurately measure mobility as a function of gate bias modulated carrier concentration for $In_{0.7}Ga_{0.3}As$ NWFETs for widths down to 40nm. Although the method is demonstrated for the specific example of InGaAs Tri-gate NWFETs it is applicable to other NWFET architectures and materials. We show that reducing the nanowire width results in mobility degradation due to scattering from the side wall roughness. Calibration of scattering models to measured experimental data allows projection of results to smaller nanowire dimensions. For ultra-narrow NW dimensions close to 10nm, expected mobility is approximately 3000cm^2/V-sec which significantly better than state of the art silicon devices. Finally, we demonstrate direct observation of room temperature ballistic transport for the top-down NWs in this work. These results indicate that III-V semiconductor NWFETs can provide a viable path to low power CMOS logic technology at nanoscale channel lengths of 14nm or smaller.

Chapter 4

Channel architecture and composition engineering for high performance $In_xGa_{1-x}As$ FinFETs

The previous chapter demonstrates evidence that III-V materials are promising for scaled n-channel FETs due to their excellent low voltage transport properties especially high electron mobility. Further possibilities are opened by varying the Indium percentage composition (x) in $In_xGa_{1-x}As$ channel which can span a wide range of bandgap and effective mass. Additionally, tuning of quantum confinement by designing ultrathin channels confined between high bandgap buffer layers with optimized band offsets can further enhance transport properties. Scalability to the 7nm or 5nm technology node, which is the likely point of introduction for III-V channels, demands excellent electrostatic integrity. Hence multi-gate architectures such as the FinFET[31,32] need to be investigated.

This chapter systematically explores performance in quantum well (QW) channel FinFET devices with varying indium composition and quantum confinement. Three different channel architectures are investigated [58,29], viz. the thick $In_{0.53}Ga_{0.47}As$ channel with body thickness T_{Body}= 40nm (also referred to as bulk), the $In_{0.53}Ga_{0.47}As$ QW channel with T_{Body} = 10nm and the $In_{0.7}Ga_{0.3}As$ QW channel with T_{Body} = 10nm. The above three channel architectures are henceforth referred to as structures A, B and C respectively for the remainder of this chapter.

I. Channel architecture, device design and fabrication

Figure 4-1 shows the epitaxial structures for devices A,B and C respectively. The layer structures are grown by IQE Inc. using molecular beam epitaxy (MBE) on 3-in semi-insulating

InP substrates. All three channels are realized on $In_{0.52}Al_{0.48}As$ buffers on InP substrates. The channels with 53% Indium composition are lattice matched to the buffer and substrate, while the 70% Indium QW channel is compressively strained with respect to the $In_{0.52}Al_{0.48}As$ buffer layer. The structures also incorporate a heavily doped cap layer with a 5nm thick higher Indium percentage $In_{0.7}Ga_{0.3}As$ sub-cap to facilitate formation of ohmic contacts with low contact resistivity.

All devices are patterned into quantum well QW FinFET devices with a raised source/drain architecture as shown in the bottom panel of figure 4-1. The first step in device fabrication is gate area recess etch of the cap layer by a citric acid/H_2O_2 solution based etch. A well-controlled over etch is included to remove the 2nm InP etch stop layer to allow formation of gate stack directly on the $In_xGa_{1-x}As$ channel. This is followed by the formation of fins within the recessed region. The fin patterns are defined using ZEP electron beam resist. The pattern is first transferred to atomic layer deposited (ALD) Al_2O_3 hard mask through a dry etch step. Subsequently, the fin etch is performed using a Cl_2/N_2 plasma based dry etch. Nickel is then directly deposited on the heavily doped cap regions using a lift-off process to form the source/drain ohmic contacts. The gate stack is integrated next using the nitrogen plasma passivation technique outlined in chapter 2. Ultrathin AlO_xN_y passivation layer is grown using alternating cycles of nitrogen plasma exposure with tri-methyl aluminum (TMA) pre-pulsing followed by thermal ALD growth of 1nm Al_2O_3 and 2.5nm HfO_2 at 250°C. This is followed by thermal evaporation of Nickel and lift-off to form the gate electrode and a forming gas anneal at 350°C for 15 minutes. All lithographic patterns are defined using direct writing with e-beam lithography.

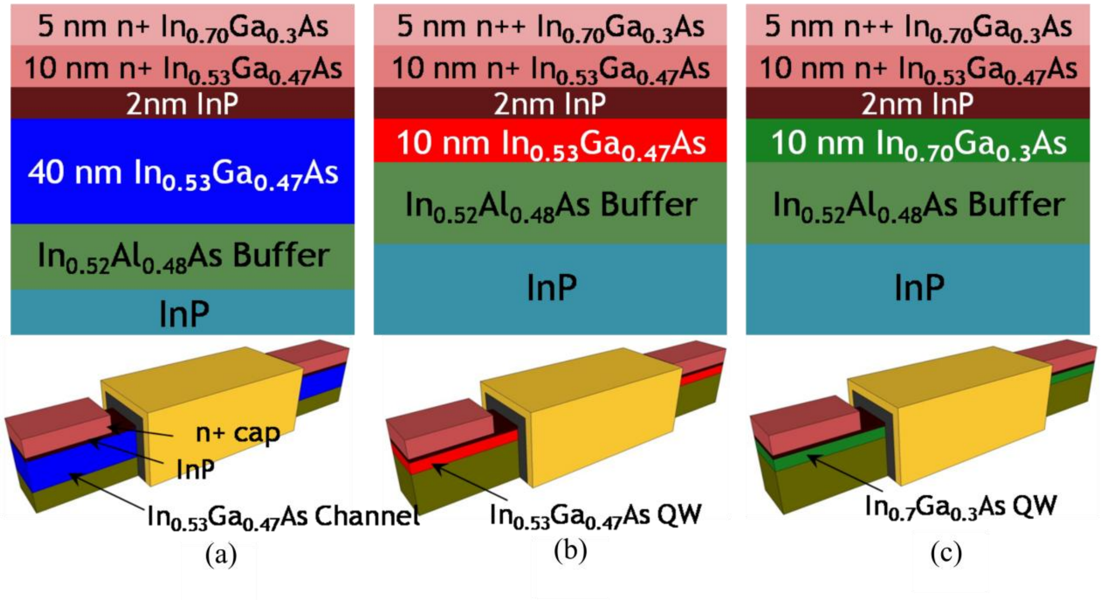

5 nm n+ In0.70Ga0.3As
10 nm n+ In0.53Ga0.47As
2nm InP
40 nm In0.53Ga0.47As
In0.52Al0.48As Buffer
InP
5 nm n++ In0.70Ga0.3As
10 nm n+ In0.53Ga0.47As
2nm InP
10 nm In0.53Ga0.47As
In0.52Al0.48As Buffer
InP
5 nm n++ In0.70Ga0.3As
10 nm n+ In0.53Ga0.47As
2nm InP
10 nm In0.70Ga0.3As
In0.52Al0.48As Buffer
InP
n+ cap
InP
In0.53Ga0.47As Channel
In0.53Ga0.47As QW
In0.7Ga0.3As QW
(a)
(b)
(c)

Figure 4-1. Schematic showing the layer structure for devices A,B and C investigated in this work. all substrates are finally patterned into FinFET structures as shown in the bottom panel.

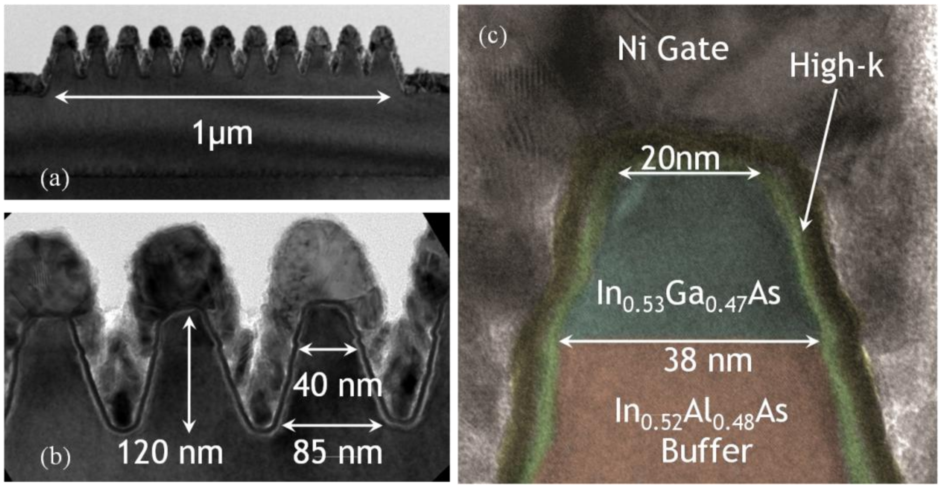

1μm
(a)
40 nm
120 nm
85 nm
(b)
(c)
Ni Gate
High-k
20nm
In0.53Ga0.47As
38 nm
In0.52Al0.48As
Buffer

Figure 4-2. Cross-section transmission electron micrographs (TEM) showing (a) tight fin pitch with 10 fins/μm layout width (b) magnified image showing critical dimensions achieved with side wall slope of 75° (c) magnified false color image highlighting the high-k dielectric and metal gate.

Figure 4-2a shows transmission electron micrographs (TEM) of the experimental device for structure A. We realize fin pitch of 100nm allowing 10 fins per μm of layout width. Figure 4-

2b shows the critical dimensions of the fin and figure 4-2c highlights the gate stack comprised of the bi-layer high-κ dielectric and nickel gate metal. Fin width of 20nm (top of fin) is realized with a sidewall taper angle of 71° designed to allow gate metal deposition on the FinFET sidewalls.

II. Results and discussion

The effect of quantum confinement in these structures is studied using two dimensional Schrodinger-Poisson simulations using nextnano. Figure 4-3 shows the inversion charge density at high gate overdrive of V_G - V_T = 0.5V for two different fin widths. It can be observed that at fin width of 40nm, structure A has maximum inversion charge concentration near the surface of the fins, making it more susceptible to surface roughness scattering from large sidewall perimeter. On the other hand structures B and C are inherently less susceptible due to volume inversion from strong confinement in the z-direction. In order to take similar advantage of volume inversion, structure A has to be scaled aggressively to fin widths to 8nm or smaller.

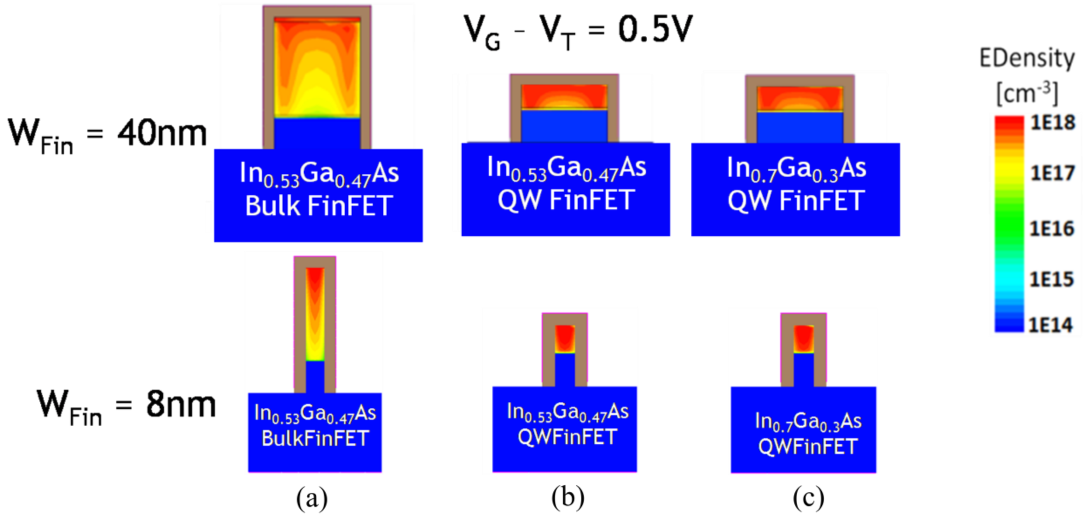

$V_G - V_T = 0.5V$
W_{Fin} = 40nm
W_{Fin} = 8nm
$In_{0.53}Ga_{0.47}As$ Bulk FinFET
$In_{0.53}Ga_{0.47}As$ QW FinFET
$In_{0.7}Ga_{0.3}As$ QW FinFET
$In_{0.53}Ga_{0.47}As$ BulkFinFET
$In_{0.53}Ga_{0.47}As$ QWFinFET
$In_{0.7}Ga_{0.3}As$ QWFinFET
EDensity
[cm^{-3}]
1E18
1E17
1E16
1E15
1E14
(a)
(b)
(c)

Figure 4-3. Two dimensional Schrodinger-Poisson simulations show inherent volume inversion in the thin quantum well devices. Similar volume inversion densities in the thick InGaAs channel is achieved only at a highly scaled fin width of 8nm.

Multi-fin split CV measurements

In order to quantify the relative performance of the three channel structures it is necessary to quantify the experimental long channel FinFET mobility. However in order to do that we need to estimate the charge using split-CV measurements. Figure 4-4 shows a long channel (L_G = 10μm) multi-fin FET device with 100 fins to enable reliable measurement of capacitance.

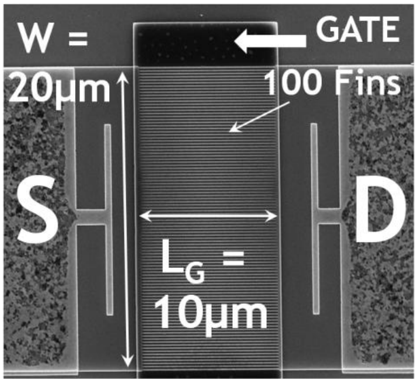
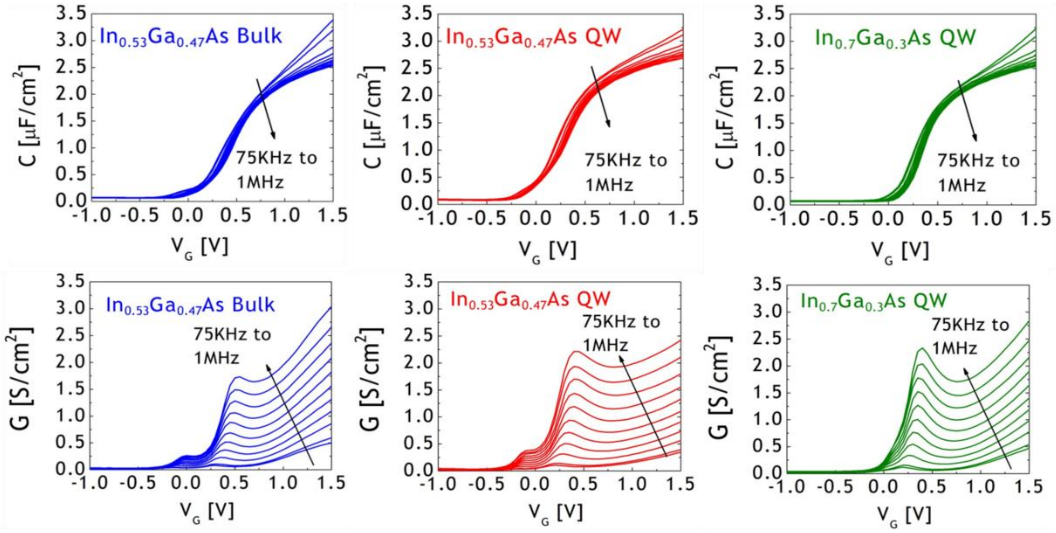


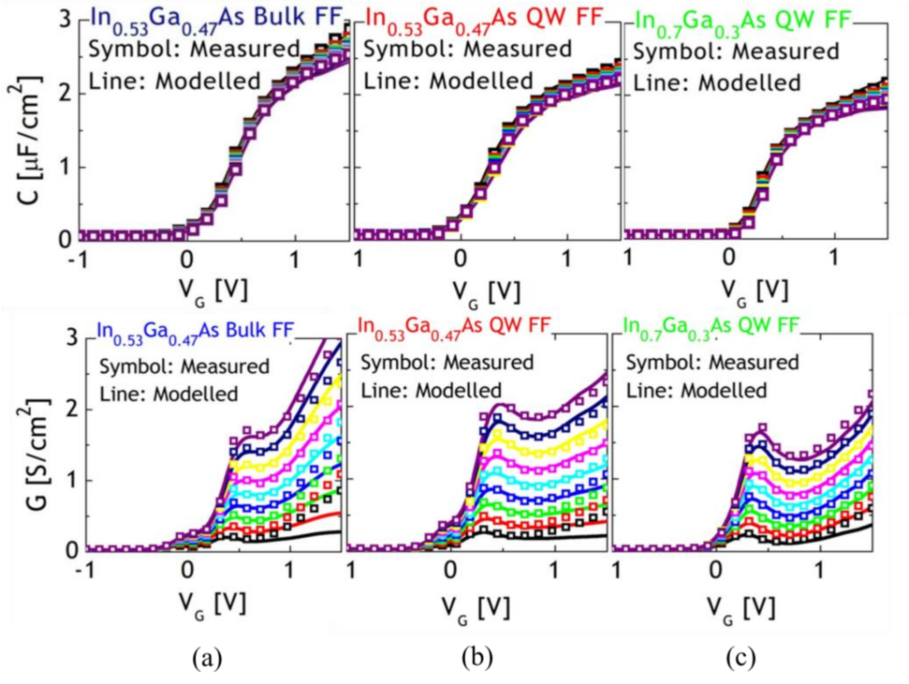
Figure 4-4. SEM of a long channel multi-fin device for measuring split-CV fin capacitance.

Figure 4-5 shows the measured capacitance (CV) and conductance (GV) characteristics for structures A, B and C over a frequency range from 75KHz to 1MHz and voltage range of -1V to 1.5V. All three structures exhibit a peak capacitance close to 2.5μF/cm^2 yielding a capacitance equivalent oxide thickness (CET) of 1.4nm. Since the three structures have varying percentage of Indium as well as different side wall perimeters, it is important to quantify the relative effectiveness of the surface passivation scheme. The quality of the passivation scheme may be estimated from interface state density (D_{it}) in the three devices. The D_{it} concentrations are extracted using the equivalent circuit method[30], by modeling the measured capacitance simultaneously along with conductance. Figure 4-6 shows the excellent fit obtained between the experimental data (symbols) and the model (lines).



(a) (b) (c)

Figure 4-5. Measured multi-fin split CV capacitance (top panel) and conductance (bottom panel) for the thick $In_{0.53}Ga_{0.47}As$, thin $In_{0.53}Ga_{0.47}As$ and $In_{0.7}Ga_{0.3}As$ quantum wells in (a), (b) and (c) respectively.



(a) (b) (c)

Figure 4-6. Fit obtained between experimental capacitance and conductance data modeled using the equivalent circuit method.

Figure 4-7 shows the extracted D_{it} as a function of the gate overdrive voltage for the three structures in comparison to planar MOSCAP devices. While the planar MOSCAP has the lowest D_{it}, it can be observed that the FinFET structures exhibit higher D_{it} near threshold with concentrations increasing further into the conduction band. This has implications on both the sub-threshold region as well as the on-state of the device. It is also observed that structure A has the highest D_{it} concentrations among the three structures which is commensurate with the higher side wall interface exposure.

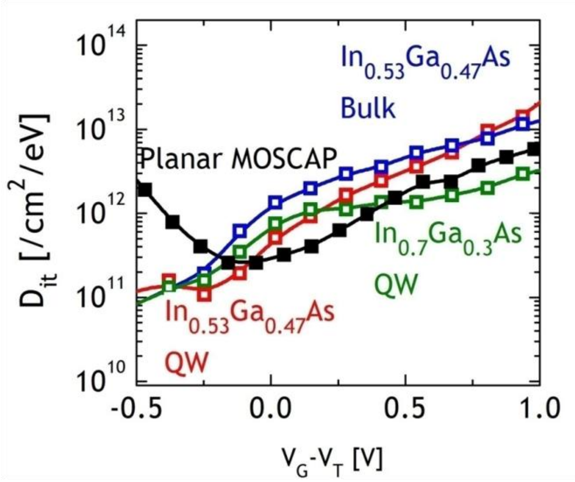
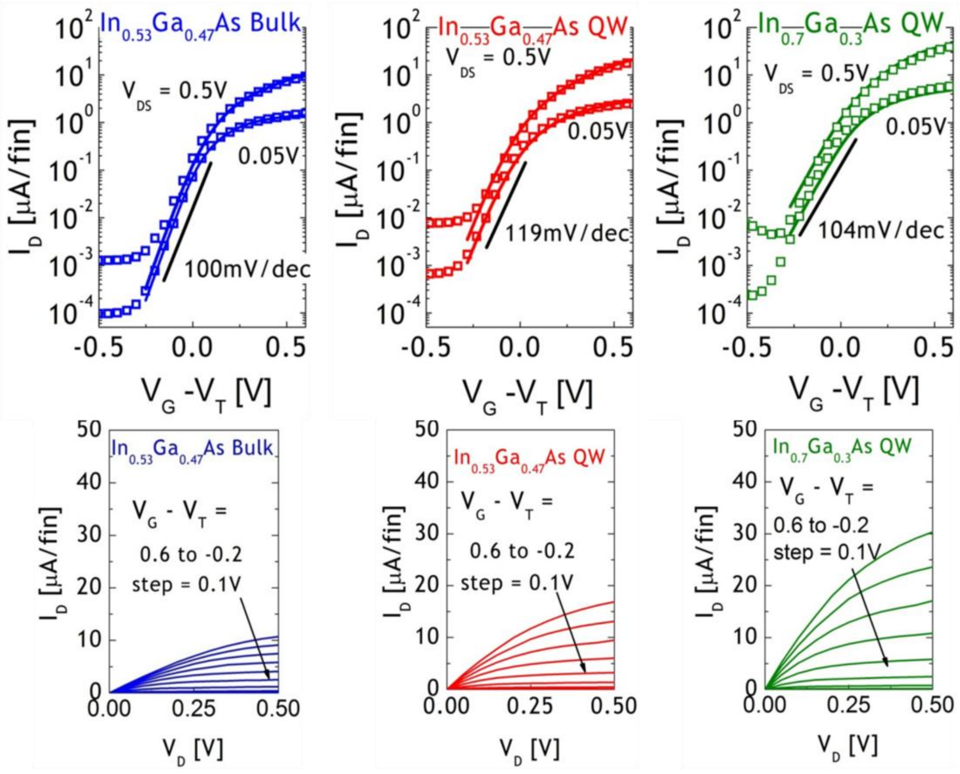


Figure 4-7. Extracted interface state density (D_{it}) profile for the three multi-fin devices is shown and compared to planar MOS capacitor D_{it} profiles.

Long channel FinFET performance

Figure 4-8 shows the per-fin measured transfer (top panel) and output (bottom panel) characteristics respectively for long channel FinFET devices with L_G = 1µm. The plots shown in blue, red and green correspond to FinFET structure A (thick $In_{0.53}Ga_{0.47}As$), B ($In_{0.53}Ga_{0.47}As$ QW) and C ($In_{0.7}Ga_{0.3}As$ QW) respectively. All devices are fabricated with tight fin pitch of 10fins/µm as shown in figure 4-2. Devices A, B and C show well behaved transfer characteristics with $I_{ON}/I_{OFF} > 10^3$ at high drain bias, V_{DS} = 0.5V. Threshold voltage of 0.24V, 0.23V and 0.31V was extracted for A, B and C respectively using the peak g_m method. Further, external parasitic resistances R_{EXT} of 314 Ω-µm, 276Ω-µm and 283 Ω-µm was estimated using the L_G extrapolation method for FinFET structures A, B and C respectively. The devices exhibit close to zero DIBL with sub-threshold slope in the linear region (V_{DS} = 0.05V; SS_{lin}) is measured at 100, 119 and 104mV/dec for device A, B and C respectively. No degradation is observed in SS for high drain bias (SS_{sat}) for these devices. From the output characteristics it is clearly observed that the 70% In, QW channel shows highest drive current close to 30µA/fin at V_G-V_T = 0.6 and V_{DS} = 0.5V. This confirms the advantage of inherent volume inversion in the QW channel, along with reduce sidewall roughness exposure and lower effective mass for the higher Indium percentage channel.



(a) (b) (c)

Figure 4-8. (a), (b) and (c) show measured long channel transfer characteristics (top panel) and output characteristics (bottom panel) for substrates A, B and C respectively.

From the measured long channel data of figure 4-8 and the measured split-CV data of figure 4-5, we extract experimental field effect mobility as shown in figure 4-9. The peak mobility is measured at 1040, 2085 and 3480 cm^2/Vs for FinFET structures A, B and C respectively. At higher carrier concentration of $n_s = 5e12cm^{-2}$ the structures show mobility of 732, 1780 and 3015 cm^2/Vs, respectively. This behavior is commensurate with the increased sidewall roughness scattering at higher gate overdrive where the charge centroid is pulled closer to the sidewall surface. These trends show the relative mobility advantage for a thin quantum well channel (structures B and C) over the thicker channel (structure A).

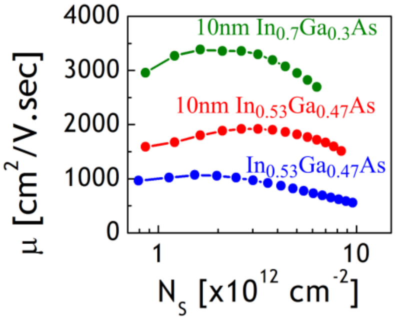
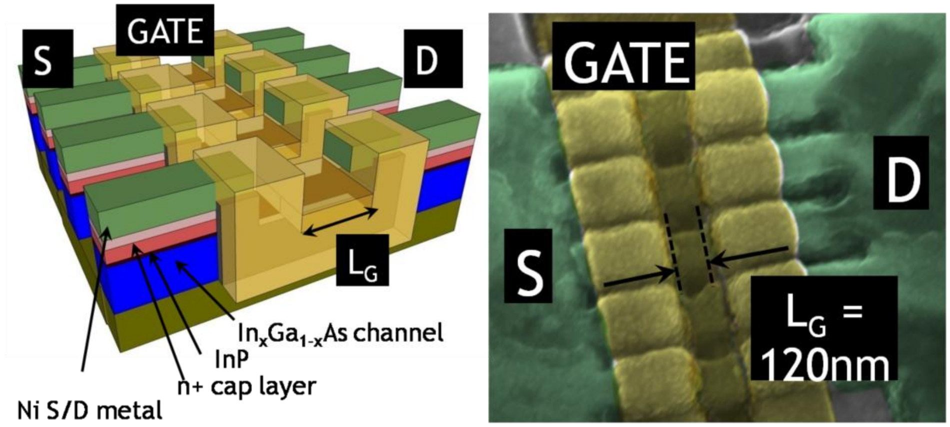


Figure 4-9. Experimentally extracted field effect mobility for long channel FinFET devices.

Short channel FinFET performance

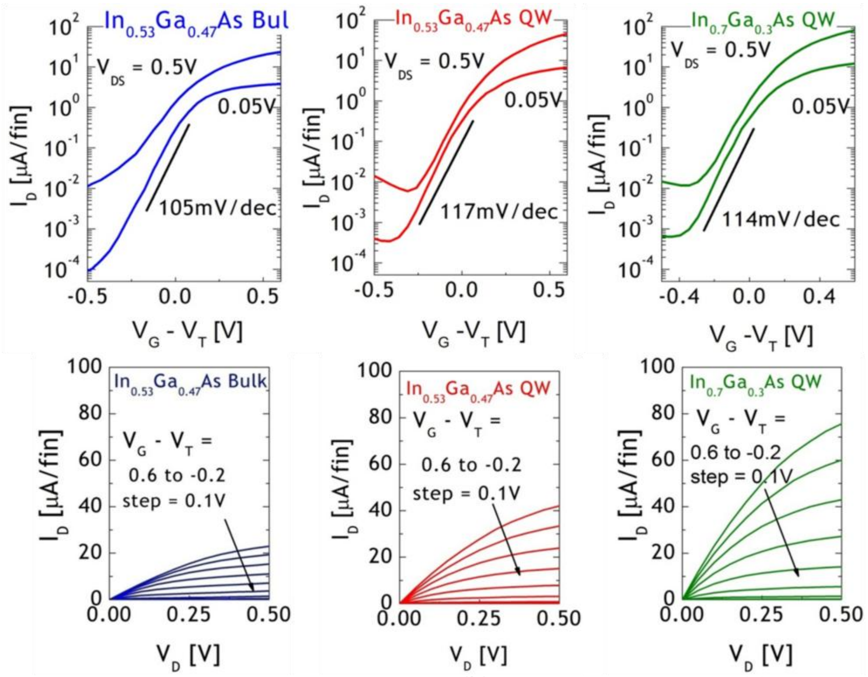
Short channel devices were fabricated at the same fin pitch as shown in figure 4-10. As shown in the schematic, the channel length (L_G) is defined by the width of the gate recess trench. In order to minimize parasitic series resistance effects, the source/drain Nickel metallization is extended as close to the channel recess as possible. Since the gate electrode is defined by lithography and lift-off, sufficient gate to source/drain overlap is provided in order to avoid misalignment as seen in the SEM in figure 4-10. Working devices with L_G down to 120nm were successfully demonstrated on all three structures.



(a) (b)

Figure 4-10. (a) Schematic showing the structure of the short channel multi-fin FET device. (b) Top view SEM of fabricated device.

Figure 4-11 summarizes the short channel transfer (top panel) and output (bottom panel) characteristics of all three structures. SS_{lin} (V_{DS} = 0.05V) of 105mV/dec, 117mV/dec and 114m/dec was obtained for devices A, B and C respectively. Whereas devices B and C do not show much degradation in SS_{sat}, device A shows large degradation. This is attributed to the large side wall perimeter and relatively large taper angle (figure 4-2) which results in poor electrostatic coupling at the bottom of the fin in device A. Raw drive currents (I_{ON}) for the as-is device structures (with 10fins/μm) is measured at 250, 400 and 760μA for structures A, B and C respectively, at V_G - V_T = 0.6V and V_{DS}=0.5V. This translates to drive currents of 25, 40 and 76 μA/fin respectively. All devices exhibit an I_{ON}/I_{OFF} ration close to 10^4 at both low and high V_{DS} bias. DIBL for QW channels B and C is comparable at 103 mV/V whereas the bulk FinFET shows a DIBL of 200mV/V again attributed to conduction through the bottom of the tapered fin which is not electrostatically well controlled. Again it is observable that device C with highest Indium percentage and quantum confinement demonstrates highest relative performance gain of 90% which is consistent with the long channel device results.



(a) (b) (c)

Figure 4-11. (a), (b) and (c) show measured transfer characteristics (top panel) and output characteristics (bottom panel) for the substrates A, B and C respectively.

Finally we summarize the transconductance (g_m) for short channel FinFET devices as shown in figure 4-12. The left axis shows the g_m per-fin (μS/fin). We now normalize the measured g_m to the gated circumference of the fin ($2H_{Fin} + W_{Fin}$) in order to facilitate easy benchmarking of the devices subsequently. We report peak g_m of 0.445, 0.92 and 1.62 mS/μm for FinFET structures A, B and C respectively. From the TEM of figure 4-2c, it is observed that W_{Fin} = 20nm and H_{Fin} = 45nm, which is the extent of the entire gated perimeter of the sidewall, resulting in an effective perimeter of 110nm per fin, for all three structures. As observed the gm data provides further confirmation that structure C with highest Indium percentage and quantum confinement provides the highest performance with a relative gain of 3.6x over the thick lattice matched channel of structure A.

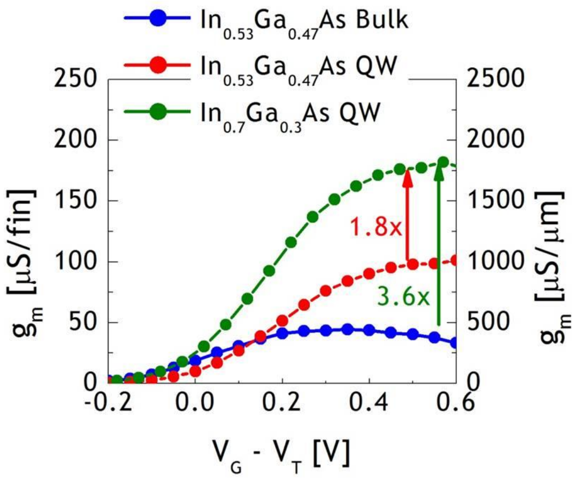


Figure 4-12. Experimentally extracted transconductance (g_m) for the three short channel devices vs. gate overdrive.

Benchmarking

Comparison between the three structures shows the enhancement in I_{ON} and g_m with increasing indium percentage consistent with lowering of effective mass which significantly enhances transport properties. Additionally, I_{ON} further improves with quantum confinement in the channel as compared to the thick channel structure. As seen in the previous chapter, significant degradation in mobility can be expected in FinFETs compared to planar structures due increased exposure to sidewall roughness scattering. Quantum confined channels improve transport due to the enhanced volume inversion and simultaneously reduced side wall exposure. Trends observed in this work are consistent with this observation thereby providing guidance for III-V FinFET channel design towards increasing indium percentage together with quantum confinement. It is however important to compare and benchmark this performance relative to state of the art silicon devices and other III-V FET demonstrations (planar and multi-gate).

Benchmarking of results discussed in this chapter is done using the Q-factor metric as defined by Doornbos et. al. [57]. The Q-factor metric provides a convenient benchmark for comparing III-V devices with widely varying channel material composition and architecture. Figure 4-13 shows that the QW $In_{0.7}Ga_{0.3}As$ FinFETs show among the highest g_m reported for III-V multi-gate (tri-gate, gate all around) devices along with a Q-factor close to 15. It is observed however that multi-gate devices with superior sub-threshold slopes (SS) have been demonstrated. Further planar III-V MOSFETs with higher g_m [60,61] are also demonstrated. These devices however report binary InAs channels with significantly shorter gate lengths. Additionally it can be seen that III-V devices show significant g_m at lower supply voltage of 0.5V as compared to state of the art silicon FinFET devices operating at 0.75V (shown as black dot). Further improvement in g_{mmax} is also expected for devices demonstrate in this chapter with scaling of L_G below 120nm, indicating significant room for improvement at V_{DS} = 0.5V. These additional factors indicate that there exists a window for further optimizing III-V FinFET performance so as to compete with Si FinFET, by improving SS and further increasing the Indium percentage approaching binary InAs.

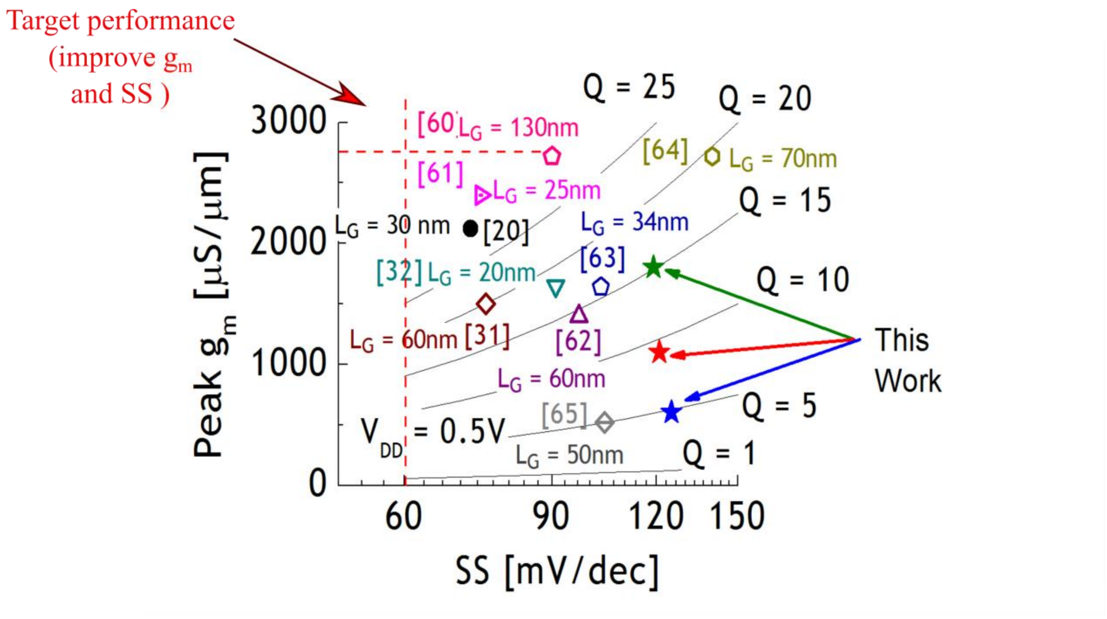
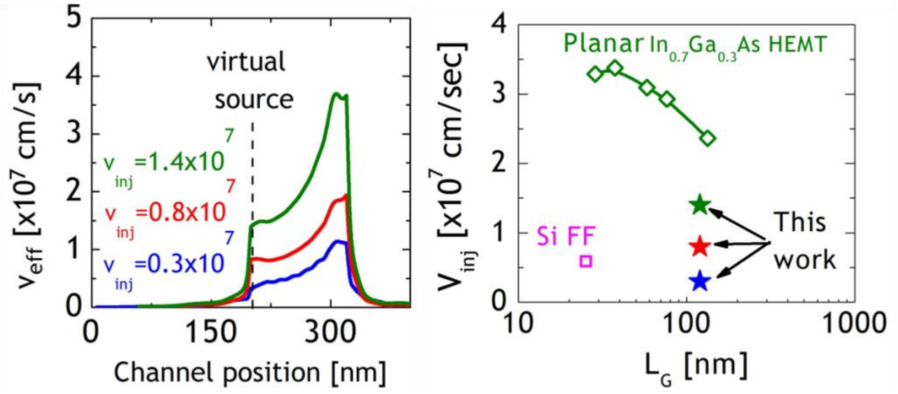


Figure 4-13. Benchmarking plot showcasing the transconductance (g_m) for the three short channel devices as a function of sub-threshold slope (SS).

Projected short channel performance

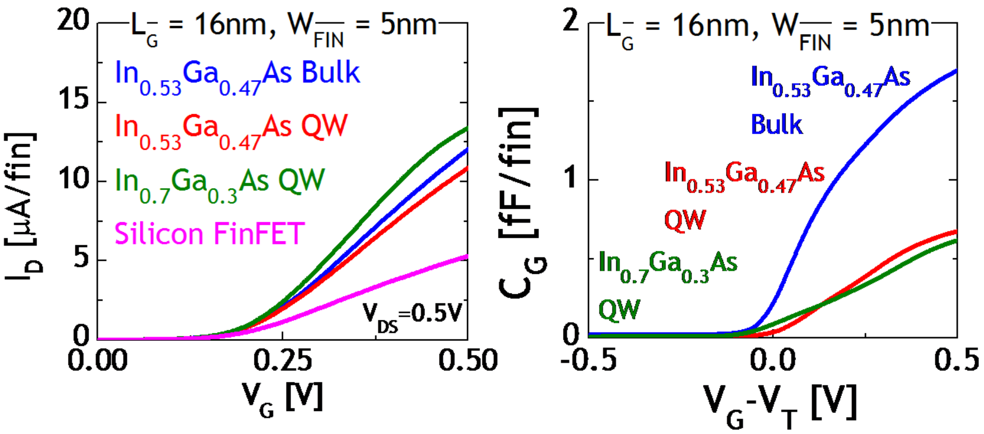
True performance for short channel devices is characterized by the injection velocity as outlined in chapter 1. In order to extract the injection velocity we create simulation models in Sentaurus TCAD for all three device structures and calibrate the models to the experimental results obtained above. Figure 4-14a shows the extracted velocity profiles along the channel for short channel devices A, B and C from figure 4-11. Highest v_{inj} of $1.4x10^7$ cm/sec is obtained for device C commensurate with the device performance. Figure 4-14b compares the extracted injection velocities to $In_{0.7}Ga_{0.3}As$ HEMT devices. The relative degradation in injection velocity for the FinFET devices is expected because of the side wall roughness which impacts low field mobility and effective injection velocity. Further enhancements in v_{inj} can be obtained by improving the fin etch to reduce side wall roughness along with reducing the channel length.



(a) (b)

Figure 4-14. (a) Velocity profiles along the channel for the three short channel devices extracted using calibrated TCAD simulations. (b) Benchmarking of the injection velocity against state of the art silicon and $In_{0.7}Ga_{0.3}As$ HEMT devices[4].

III-V channel materials are likely to be introduced at the 7nm node or beyond and in order to assess their feasibility for these technologies it is important to benchmark performance with scaled devices. The simulation models created above for extracting injection velocity are scaled to device dimensions for 7nm node. Figure 4-15a shows the drive currents simulated for devices A, B and C in comparison to a scaled Silicon FinFET device which is calibrated to the 22nm silicon FinFET data[20]. As seen structure C provides the best drive currents with over 2x improvement as compared to a scaled Si FinFET. Figure 4-14b shows the capacitance per fin for all three scaled structures. Structure A provides highest capacitance (charge) commensurate with the higher fin perimeter. However the higher drive currents in structure C come from the significantly higher injection velocity which compensates for the relative capacitance penalty.



(a) (b)

Figure 4-15. Projected performance for 7nm node FinFET at 0.5V supply voltage showing (a) drive currents and capacitance per fin for the three substrates.

III. Conclusions

Experimental demonstrations of channel engineered $In_xGa_{1-x}As$ FinFETs show that in order to enhance performance, higher indium percentage and quantum confinement are necessary. An advanced plasma nitride passivation scheme with a bi-layer high-k dielectric stack yields excellent sub-threshold and on-state characteristics. High peak transconductance is demonstrated and significantly enhanced transport properties are observed for a thin QW $In_{0.7}Ga_{0.3}As$ channel device compared to a thick $In_{0.53}Ga_{0.47}As$ channel architecture. High field effect mobility in excess of 3000 cm^2/V-sec is measured from long channel multi-fin split-CV measurements. We obtain g_{mmax} = 1.62mS/μm for QW $In_{0.7}Ga_{0.3}As$ FinFETs at L_G = 120nm, which is among the highest g_{mmax} reported for multigate $In_xGa_{1-x}As$ devices. Benchmarking exercise indicates that further improvements in FinFET performance can be achieved with enhancements to the channel to include even higher Indium percentage approaching binary InAs. Well characterized etch chemistry to pattern fin side walls with reduced roughness will allow more well controlled interface formation with the dielectric stack and enable demonstration of competitive devices which can potentially outperform state of the art Si FinFET devices.

Chapter 5

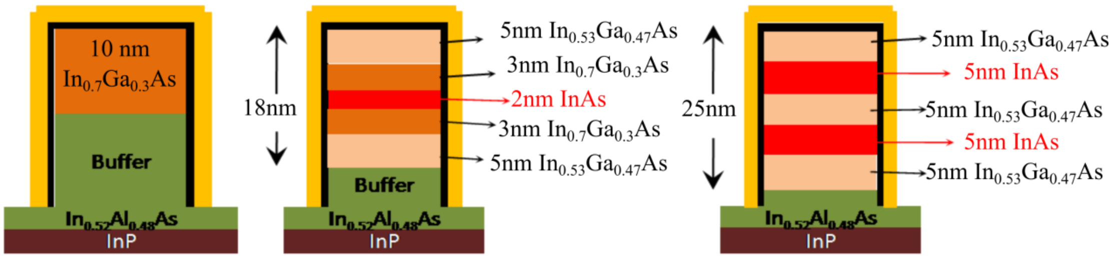
High performance InAs heterostructure FinFETs

I. Motivation

The previous chapter outlined the exploration of the trade-offs inherent in engineering the channel to include higher In% and quantum confinement. Results showed that further increasing the In% to approach binary InAs will yield the ultimate performance. Designing an InAs channel requires careful consideration of the factors affecting heteroepitaxial growth . InAs has a 3.2% lattice mismatch to InP [67] and thus lattice matched growth requires an Antimony (Sb) based buffer design. Recent demonstrations have been made on STI trench growth of AlGaSb buffer layers with upto 10nm relaxed InAs channels[60]. Other demonstrations have also been made with extremely thin body -on insulator (ETB-OI) enabled by wafer bonding[63]. Arsenic based buffer designs on the other hand are more attractive due to their well established growth and processing procedures. However, pseudomorphic epitaxy of strained InAs is limited to layers thinner than 3-4nm (10 mono-layers) [68, 69]. Still recent HEMT demonstration have included channels with up to 5nm thick InAs cores designed with top/bottom $In_{0.53}Ga_{0.47}As$ cladding layers[70,71]. Thus recent device demonstrations have focused on composite InGaAs/InAs channels. Although ultra thin pseudomorphic layers of InAs are suitable for planar HEMT devices, a fundamentally different approach is required for demonstrating scalable InAs FinFET technology. In order to take advantage of the large sidewall area it is necessary to incorporate thicker InAs cores. However due to the limitations of pseudomorphic growth the only viable growth approach is by including composite channels with multiple quantum wells.

II. Device design and fabrication

In this chapter we focus on the design and demonstration of FinFET devices with composite InAs channels. Specifically we include a single quantum well (SQW) with a 2nm InAs core and a dual quantum (DQW) structure with two 5nm thick InAs cores. Additionally we benchmark the performance of these devices against the highest performance $In_{0.7}Ga_{0.3}As$ QW FinFETs shown in the previous chapter from [72]. Figure 5-1 shows the cross-section schematic of all three devices. The SQW structure is sandwiched in a symmetric cladding comprising 3nm $In_{0.7}Ga_{0.3}As$/5nm $In_{0.53}Ga_{0.47}As$ whereas the cladding in the DQW structure comprises only of 5nm $In_{0.53}Ga_{0.47}As$. The DQW channel is designed to allow the maximum possible InAs thickness similar to that reported in [70].


10 nm
$In_{0.7}Ga_{0.3}As$
Buffer
$In_{0.52}Al_{0.48}As$
InP
18nm
5nm $In_{0.53}Ga_{0.47}As$
3nm $In_{0.7}Ga_{0.3}As$
2nm InAs
3nm $In_{0.7}Ga_{0.3}As$
5nm $In_{0.53}Ga_{0.47}As$
Buffer
$In_{0.52}Al_{0.48}As$
InP
25nm
5nm $In_{0.53}Ga_{0.47}As$
5nm InAs
5nm $In_{0.53}Ga_{0.47}As$
5nm InAs
5nm $In_{0.53}Ga_{0.47}As$
$In_{0.52}Al_{0.48}As$
InP

(a) (b) (c)

Figure 5-1. Cross-section schematic of the three FinFET structures highlighting the single and dual InAs quantum wells.

The channels are then patterned into FinFET devices following a process flow similar to that used in the previous chapter with the devices having the same fin layout density of 10fins/μm. Figure 5-2a shows the representative schematic of the device. In order to improve the uniformity and quality of the etched side walls in these fins, a new modified side wall image transfer process was developed. Figure 5-2b shows the process flow used for fin patterning. The devices are first patterned with dummy mandrels of ZEP e-beam resist. Low temperature atomic layer deposition (ALD) of Al_2O_3 is then carried out to define the thickness of the fin. This is followed by anisotropic etch back of the Al_2O_3 in a fluorine based chemistry which is selective to the underlying substrate. The ZEP mandrels are then ashed away in a O_2 based plasma leaving behind Al_2O_3 fin mask patterns. The fins are then transferred into the InGaAs substrate using a high temperature chlorine based plasma etch process. Figure 5-2c,d show the cross section SEMs of fabricated fins along with the residual hard mask. As seen good uniformity and side wall smoothness is achieved.

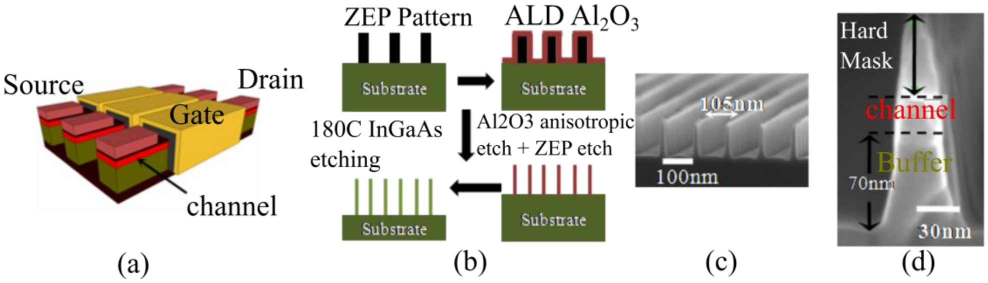
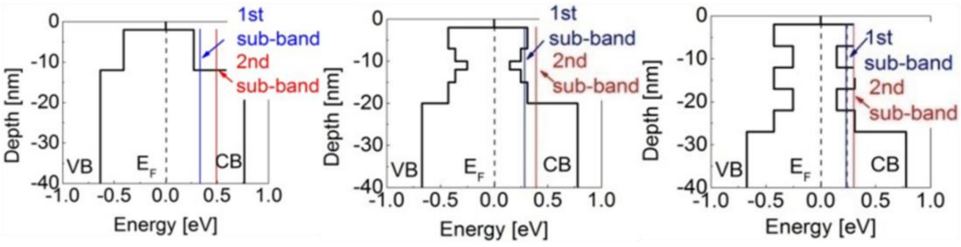

Source
Drain
Gate
channel
ZEP Pattern
ALD Al_2O_3
Substrate
Substrate
180C InGaAs etching
Al2O3 anisotropic etch + ZEP etch
Substrate
Substrate
105nm
100nm
Hard Mask
channel
Buffer
70nm
30nm
(a)
(b)
(c)
(d)

Figure 5-2. (a) Schematic of fabricated FinFET devices (b) Process flow outlining the spacer technique for patterning fins (c) cross-section SEM of etched fins and (d) cross-section SEM of etched fin shown with residual hard mask.

Bandstructure effects

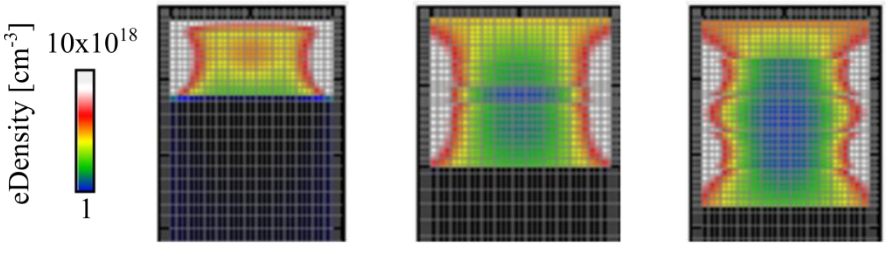
The primary objective of including multiple QWs of InAs is to increase the amount of charge available over the fin cross-section. Performance of the heterostructure FinFETs can be maximized by increasing the amount of charge confined in the high mobility InAs layers. Figure 5-3 shows the one dimensional self-consistent Schrodinger-Poisson bandstructure calculations with the position of the first two sub-bands. The effective mass approximation is used to perform the simulations as it provides a good first order estimate for thick channels. First we observe that due to the favorable conduction band offset between InAs/InGaAs inversion charge will be concentrated in the high mobility InAs layers. Further due to the increasing body thickness (fin height H_{Fin}) the composite wavefunction for the entire heterostructure causes the sub-band separation to decrease. This is advantageous for obtaining higher inversion charge densities thus providing additional benefit along with the increased gated sidewall area. The subband separation is simulated to be 180meV, 130meV and 30meV for structures a, b and c respectively.



(a) (b) (c)

Figure 5-3. One dimensional Schrodinger-Poisson simulations using nextnano showing reducing sub-band spacing going from the (a) InGaAs QW to the (b) single and (c) dual QW substrates.

In order to verify the charge distribution in the FinFET geometry we perform two dimensional SP simulations as shown in figure 5-4. Simulations are performed for a 40nm wide fin cross-section including a high-k oxide layer. We see that highest inversion charge density is achieved in the high mobility InAs and $In_{0.7}Ga_{0.3}As$ cladding layers for the single QW structure whereas most of the charge is exclusively concentrated in the InAs layers for the dual QW structure. It is also observed that inversion charge density is higher on the sidewalls as compared to top surface again attributed to the higher conduction band offset between InAs and high-k dielectric.



(a) (b) (c)

Figure 5-4. Two dimensional Schrodinger-Poisson simulations using nextnano showing higher inversion charge concentration near the InAs regions in the single and dual quantum well structures.

In order to allow proper comparison of the relative performance of the devices, it is necessary to identify the effective conduction perimeter of the fin carefully. By defining an effective perimeter based on a weighted average of the perimeter from the individual layers of the fin we can account for the varying charge distributions within the fin cross-section. Table 5-1 summarizes the effective perimeters for the three structures using this procedure.

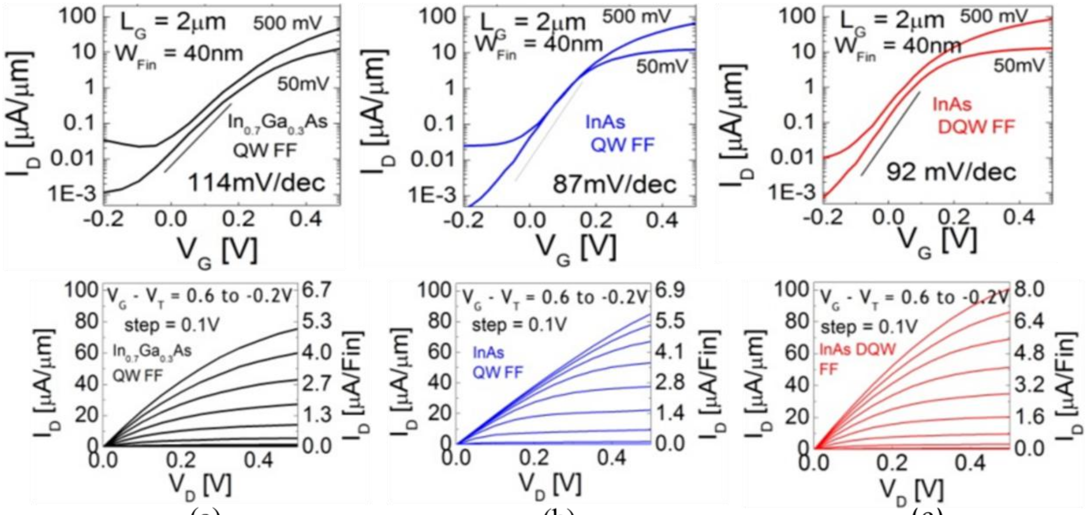
FinFET	$W_{eff} = W_{Tot} \times \frac{\sum N_S W}{W_{Tot}.N_{S\_MAX}}$
$In_{0.7}Ga_{0.3}As$ QW	(40 + 2x10) **x1** = 60nm
InAs QW	(40 + 2x18) **x0.92** = 70nm
InAs DQW	(40 + 2x25) **x0.87** = 78.3nm

Table 5-1. Effective perimeter of FinFET devices based on weighted average of carrier distributions.

III. Results and discussion

Long channel performance

Devices were fabricated using the novel fin definition process flow described above. Additionally, all devices demonstrated in this chapter employ a 3.5nm HfO_2 only dielectric integrated with the plasma nitride passivation layer. This leads to improved subthreshold performance for all FinFET devices consistent with the observation of improved MOSCAP CV behavior as detailed in chapter 2. Figure 5-5 shows the long channel FinFET (L_G = 1µm; W_{Fin} = 40nm) device results. Panels (b) and (c) show the results for the single and dual QW respectively. The excellent subthreshold slope of SS_{avg} = 87,92mV/dec at V_{DS} = 50mV and SS_{avg} = 90,96mV/dec at V_{DS} = 0.5V were obtained. These characteristics are attributed to the new fin definition and high-k integration scheme. The threshold voltage is measured at 0.15V and 0.1V for the single and dual QW devices respectively. Among the three devices it is seen that the dual QW FinFET provides the highest on-current at 8μA/Fin compared to 5.6μA/Fin for the single QW device at a gate overdrive of 0.6V. However both devices outperform the $In_{0.7}Ga_{0.3}As$ QW FinFET highlighted in the previous chapter. Normalizing the drive currents to the effective perimeter defined above we get drive currents of 84μA/μm and 100μA/μm for the single and dual QW devices respectively.



(a) (b) (c)

Figure 5-5. Transfer characteristics (top panel) and output characteristics (bottom panel) are shown for FinFETs on the three substrates from figure 5-1 in column (a), (b) and (c) respectively.

We benchmark the performance of the long channel devices by extracting the field effect mobility using the inverse modeling technique. Sentaurus TCAD models are calibrated to the long channel I_D-V_G characteristics (figure 5-5). After careful calibration of both subthreshold and on-state regions, charge is estimated by integrating across the device cross-section at each VG point for low drain bias ($V_{DS} = 0.05V$). At low drain bias the current is given as

$$I_D = \frac{W}{L}\mu C_{ox}(V_G - V_T)V_{DS}$$

where $C_{ox}(V_G-V_T)$ is the charge estimated from the simulation model. Mobility is then estimated from the measured drain current after dividing by the drain bias. Figure 5-6a shows the extracted mobility data for all three devices. It is observed that both the InAs QW and the dual QW structures have comparable peak mobility estimated at 3950cm^2/V.sec and 3531cm^2/V.sec respectively. However it is observed that the mobility roll-off at high carrier concentration (N_S) is steeper for the dual QW structure due to higher sidewall roughness exposure. Figure 5-6b shows the corresponding g_m extracted from long channel I_D-V_G. The DQW device shows highest gm of 373μS/μm compared to 290 μS/μm for the SQW device.

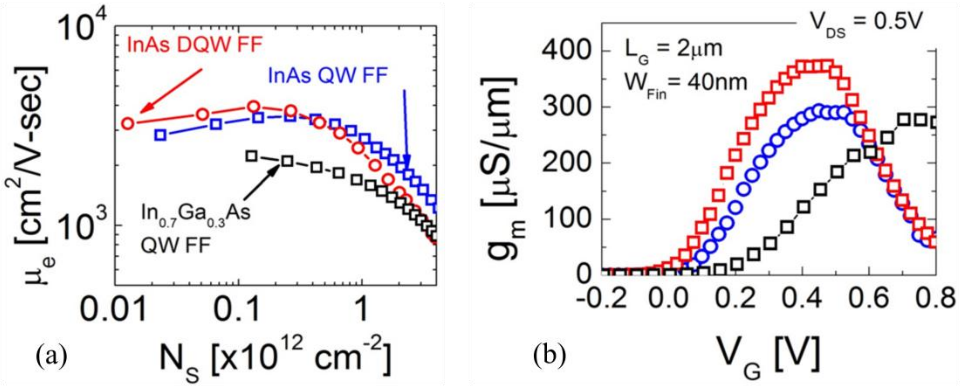


Figure 5-6. (a) Long channel FinFET mobility extracted using inverse modeling in Sentaurus TCAD after calibration to experimental long channel transfer characteristics. (b) Long channel FinFET transconductance.

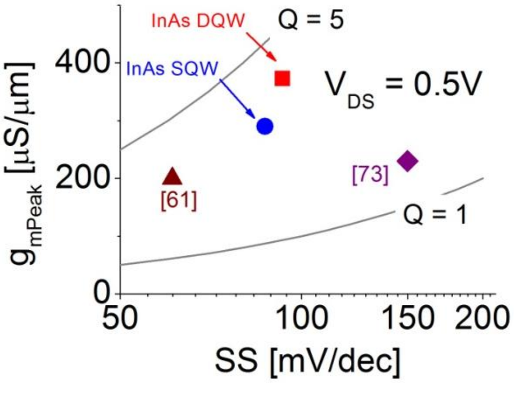


Figure 5-7. Benchmarking of experimental long channel FinFET g_m against published long channel $In_XGa_{1-X}As$/InAs devices [61,73].

From the transconductance data we see that the DQW FinFET demonstrates higher performance at comparable mobility due to the higher charge per fin supported by the two QWs. In order to benchmark the performance we use the Q-factor as described in [59]. The DQW FinFET device demonstrates the highest g_m amongst long channel III-V devices.

Intermediate channel length devices were fabricated with channel length down to 300nm. Figure 5-7 shows the transfer characteristics along with the g_m for single and dual QW channels. Both short channel FinFETs show good subthreshold behavior with SS =100mV/dec. The dual QW channel device exhibits higher g_m at 578 µS/µm compared to 513 µS/µm for the SQW device.

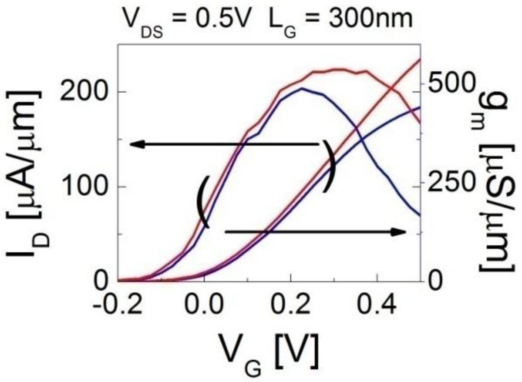


Figure 5-8. Short channel (L_G = 300nm) transfer characteristics shown for single (blue solid line) and dual (red solid line) FinFET devices. Right axis shows corresponding transconductance.

Short channel benchmarking

In order to assess the feasibility of any new technology it is necessary to benchmark the performance against state of the art Si devices. In this section we use 22nm Si FinFET technology to benchmark performance of the InAs single and dual QW FinFETs. However in order to provide a realistic benchmark the devices are first scaled down to equivalent 22nm technology dimensions using TCAD simulations. 3-D simulation models are first calibrated to experimental long channel and intermediate channel length data. The models are then scaled to 22nm technology dimensions (W_{Fin} = 8nm and L_G = 26nm) as shown in figure 5-9. Whereas the Si FinFET has a tapered sidewall with the top width of 8nm, the InAs QW devices are simulated with vertical sidewalls as shown. Also seen in figure 5-9 is the electron density profile simulated at V_G = 0.5V with I_{OFF} = 100nA/µm (high performance logic). It is observed that majority of the inversion charge is concentrated in the InAs layers (81% for SQW and 92% for DQW) indicating that the InGaAs cladding layers play a very negligible role in the conduction process. This has significant implications for the design of future InAs based devices wherein the InGaAs cladding layers may be scaled to ultrathin layers of 1nm thickness. This also allows for complete replacement of the cladding layers with other etch selective barrier layers during growth which maybe subsequently etched off during device fabrication to leave behind only the active InAs QW layers.

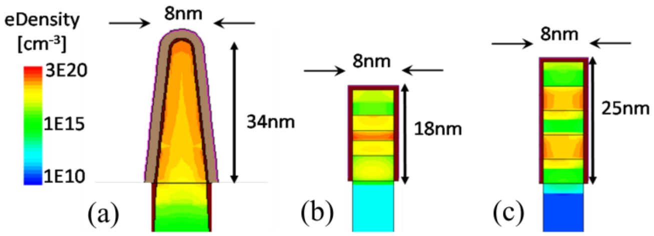


Figure 5-9. Fin cross-sections showing electron density profiles simulated at V_G = 0.5V (I_{OFF} = 100nA/µm) for scaled 8nm devices, (a) Si FinFET (b) InAs single QW FinFET and (c) InAs DQW FinFET.

A key figure of merit for short channel devices is the injection velocity (V_{inj}) at the virtual source as outlined in chapter 1. Figure 5-10 shows the extracted velocity profile along the channel length for the InAs QW devices along with Si FinFET. The single and dual QW devices show V_{inj} of 1.6×10^7cm/sec and 1.4×10^7cm/sec respectively compared to 5.3×10^6 cm/sec for 22nm Si FinFET. Thus both the single and dual QW devices have comparable injection velocity which is expected due to the fact that both systems have similar low field mobility as seen in figure 5-6. However we would expect higher current per fin from the dual QW device as it has higher InAs

QW thickness which will allow it to support higher inversion charge and thus higher $I_{ON} = Q_{inv}*V_{inj}$.

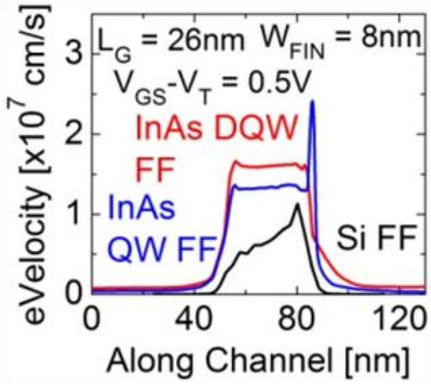


Figure 5-10. Electron velocity profiles for FinFETs projected at 22nm node using Sentaurus TCAD. Transport parameters are calibrated to experimental short channel devices.

This expectation of higher (>2x) on current is verified through the TCAD simulation models. Figure 5-11 shows the I_D-V_G transfer characteristics for the InAs single and dual QW devices in comparison to 22nm Si FinFET. The doubling of current for the InAs dual QW device is verified with demonstration of per fin I_{ON} = 27.2μA/fin compared to 13.56μA/fin for the single QW device. Additionally we note that the dual QW device supports on current comparable to the Si FinFET device with significantly lower fin perimeter (only InAs QWs conduct).

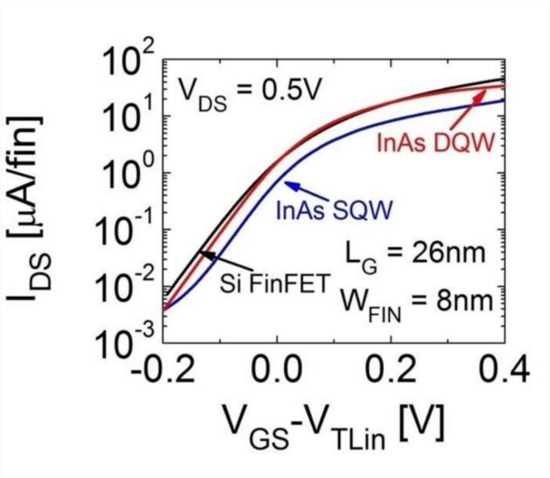


Figure 5-11. Transfer characteristics (current per fin) of InAs SQW and DQW devices benchmarked against 22nm Si FinFET technology.

Key short channel high performance logic metrics are summarized in table 2 below along with corresponding values for 22nm Si FinFETs. The above benchmarking simulations indicate that significant performance enhancement may be obtained from stacking larger number of InAs QWs in the layer structure. Due the higher V_{inj} in InAs, the current per fin may be enhanced by increasing the amount of inversion charge by stacking larger number of QWs. One of the major impediments curtailing the performance however is the raised source/drain architecture of these devices which results in high series resistance. This issue is explored in further detail in the final chapter along with possible solutions for its mitigation.

V_D = 0.5V; I_{OFF} = 100nA/µm					
Substrate	I_{DLIN} µA/µm	I_{DSAT} µA/µm	DIBL mV/V	SS_{SAT} mV/dec	g_{m_peak} µA/µm
Si FF [7]	131.5	417.5	30	71	1768
InAs QW	145.2	309.2	39	87	1125
InAs DQW	238.9	478.3	32	76	1734

Table 5- 1 Benchmarking of projected short channel performance (at 22nm node) with state of the art 22nm silicon technology. InAs dual QW FinFET devices show promising on-current.

IV. Conclusions

This chapter presented data on experimental demonstration of long channel InAs single and dual QW FinFETs. Results indicate that InAs QW heterostructures provide higher mobility even compared to high Indium percentage $In_xGa_{1-x}As$. This is also confirmed from high transconductance measured in the long channel FinFET devices. Additionally short channel projections are made using TCAD models calibrated to the measured experimental results. Both InAs single and dual QW devices demonstrate higher injection velocity compared to state of the art silicon FinFET devices. Further, both experiment and simulations establish that current per fin increases by stacking higher number of QWs. The higher amount of inversion charge generated on the sidewalls of the fin combined with the higher injection velocity results in higher current per fin. These observation indicate that III-V heterostructure FinFETs incorporating InAs can potentially outperform Silicon devices at low supply voltages (<0.5V). Finally it is observed that the InGaAs cladding layers, used to support the InAs QW growth, contribute much less to current conduction than anticipated. This is attributed primarily to the conduction band offset which favors inversion charge localization in the InAs QWs and secondly to the lower mobility in the cladding layers. This result significantly impacts future design of stacked InAs QW heterostructure FinFETs wherein the cladding layers can be made arbitrarily thin or ultimately removed altogether by using selective etch chemistry.

Chapter 6

Conclusions and future work

I. Summary and conclusions

This dissertation demonstrates important milestones towards the realization of III-V FinFETs for low power logic applications. The most important requirement for any new material system to be adopted as an alternative to Silicon is the availability of a high quality gate stack. This is critical to the demonstration of Chapter 2 describes a dilute N_2 plasma based surface clean and passivation with HfO_2 gate dielectric resulting in a high quality high-k metal-gate (HKMG) stack with maximum capacitance density of 3μF/cm^2 and CET of 1.2nm (EOT =0.8nm). This gate stack is shown to outperform more simplistic thermal ALD based high-k growth schemes with significantly lower D_{it}.

Chapter 3 addressed the next critical question about the demonstration of mobility advantage for III-V materials over silicon in 3-D confined nanoscale structures. A novel technique for accurate quantitative measurements of mobility in multi-gate nanowire geometries is demonstrated for which a new test structure, viz. the multi-nanowire gated hall bridge was developed and realized on $In_{0.7}Ga_{0.3}As$ QW substrates. A new equivalent circuit methodology for characterization was developed and verified through simulations and experiment. The primary mechanism of mobility degradation with reducing nanowire width was found to be sidewall roughness scattering. Measured Hall mobility was shown to have a peak value of 4200cm^2/V.sec at carrier density N_S = 2x10^{12}cm^{-2} at a width of 40nm. Projections for scaled nanowires (5nm width), based on experimentally calibrated models, show greater than 10x mobility advantage over Silicon nanowires. Additionally, room temperature ballistic transport was demonstrated in a

100nm hall cross structure. These observations of high mobility and ballistic transport in confined III-V structures motivates the demonstration of high performance FinFET devices.

This question is addressed in Chapter 4 where we explore channel design with varying indium percentage and quantum confinement. Thick (40nm) $In_{0.53}Ga_{0.47}As$ QW, thin (10nm) $In_{0.53}Ga_{0.47}As$ QW and thin (10nm) $In_{0.7}Ga_{0.3}As$ QW channel FinFETs were demonstrated along with a tight fin pitch process is demonstrated allowing 10 fins per μm layout width and fin width of 20nm. Long channel FinFETs exhibit peak mobility of 1040, 2085 and 3480 cm^2/Vs respectively while short channel FinFETs exhibit peak g_m of 0.445, 0.92 and 1.62 mS/μm respectively revealing that higher Indium content channels display superior transport and device performance. This was also found to be among the highest gm reported for multi-gate III-V devices.

Finally these results motivate chapter 5 where the performance of binary InAs channel FinFETs is explored. InAs channels are introduced in the heterostructure as Single and dual InAs quantum wells. Stacking of QWs allows for a viable approach to increase the total InAs channel thickness per fin. Long channel mobility of $3950cm^2/V.sec$ and $3531cm^2/V.sec$ was measured for the dual and single QW channel FinFETs respectively. Experimentally calibrated device simulations scaled to 22nm node show υ_{inj} of $1.6x10^7$cm/sec, $1.4x10^7$cm/sec compared to $5.5x10^6$ cm/sec for Silicon which is about 3x lower. The higher InAs thickness and υ_{inj} result in I_{ON} of 478μA/μm for the dual QW FinFET compared to 415μA/μm for Si FinFET.

In chapters 4 and 5, projected performance for scaled devices is shown to be comparable to state of the art Silicon technology at 0.5V supply voltage. However, several improvements are needed in order to significantly enhance device performance and justify the migration to III-V materials. One of the key impediments to high drive currents is external resistance (R_{Ext}). All device demonstrations in this dissertation employ a raised source/drain architecture realized using the heavily doped cap layer grown with the channel during MBE. Raised source/drain contacts increase the access resistance to the bottom of the fins. This problem is further exacerbated in high aspect ratio fins where the target is to increase drive current by designing taller fins. Figure 6-1 shows R_{Ext} extracted for stacked InAs QW FinFETs from long channel TCAD simulations as a function of the number of QWs showing clearly the resistance penalty incurred in high aspect ratio designs with raised source/drain. R_{ext} measured for FinFETs in this thesis are larger than 200Ω-μm with specific contact resistivity ρ_S = 20 Ω- μm^2 , which is significantly higher than the target value of 0.5 Ω-μm^2 required for the 10nm node and lower. However nanoscale planar contacts with ρ_S in the range of 0.5-1.3 Ω- μm^2 have been demonstrated [74] which bodes well for III-V CMOS in general.

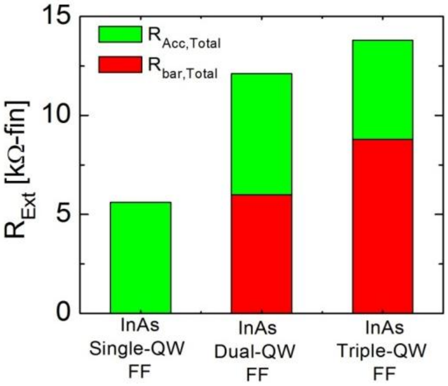


Figure 6-1. External resistance for InAs QW FinFETs with raised source/drain architectures rises rapidly with stacking of InAs QW.

These results demonstrate the necessity for embedded source/drain architectures which may be realized either through traditional implant techniques or source/drain re-growth. III-V semiconductor device structures however have traditionally employed planar architectures and thus utilized in-situ MBE growth for realizing heavily doped layers with low defect concentrations. The art of implanting n-type dopants (such as Si) in III-V materials and subsequent activation anneals is not as well established as the equivalent Silicon technology counterparts. Some of the issues that need to be carefully looked at is excessive junction leakage due to implant damage. J. J. Gu et al [32] for example observe high drain-substrate leakage and hence report source current to mask this undesirable leakage. However careful optimization of the implant process and activation anneal conditions should allow formation of high quality heavily doped source/drain junctions. Source/drain re-growth using MOCVD or MBE has been demonstrated for several recent planar InGaAs devices[60,61,73] but this is a completely unexplored frontier as far as 3-D FinFETs are concerned. Although this is a very exciting research opportunity, it is fraught with several difficult challenges. One of the most critical impediments to such work could be the starting surface stoichiometry post fin etch, which might make it impossible to nucleate subsequent layers during re-growth.

Trends demonstrated in this thesis indicate that additional refinements to the channel architecture such as migrating to the highest Indium percentage binary InAs channel will result in further performance improvement. Realizing larger InAs channel thickness will also enable high aspect ratio InAs FinFETs with higher mobility close to the bulk values of InAs. One critical question however is the capacitance penalty incurred and the resulting trade-off with transport enhancement. Due to the low effective mass in such materials the DoS bottleneck (chapter 1) could potentially negate all the benefits of enhanced transport. Since current is proportional to the product of the two quantities it is important to quantify the loss in charge in high Indium content channels relative to Silicon. Figure 6-2 shows the charge density calculated using 2-D Poisson-Schrodinger simulations for the different structures demonstrated in this work at a fin width of 8nm using nextnano. To account for the change in effective mass with quantization, 8x8 k.p bandstructure simulations are used. It can be seen that the lower DoS in III-V FinFETs results in 2x lower charge density (capacitance) which in-turn will translate to lower charge per fin. Similar trends are also observed by Majumdar in [75] for double gate $In_{0.53}Ga_{0.47}As$ and Silicon MOSFETs with 8nm channel thickness.

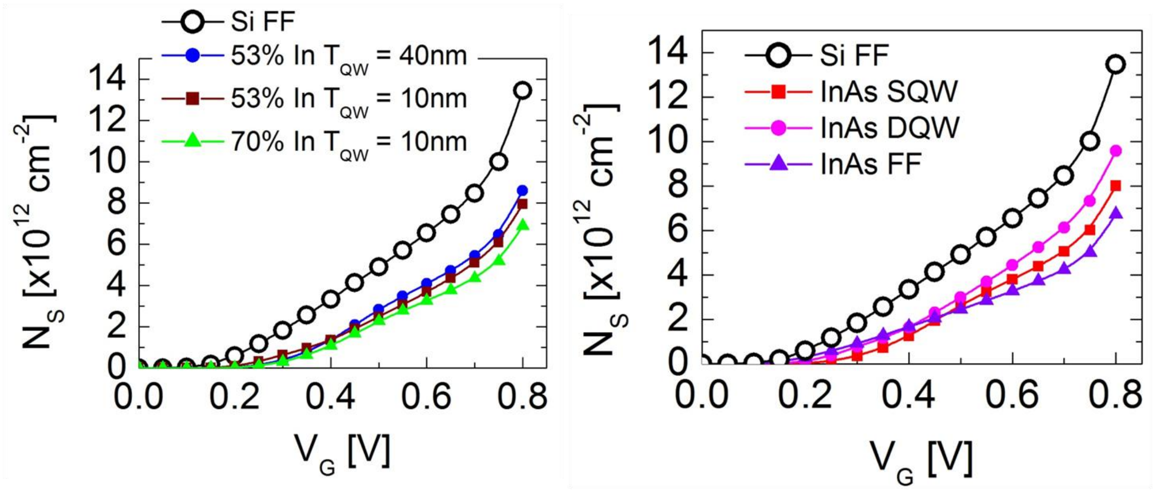


Figure 6-2. Charge density estimated using two dimensional Poisson-Schrodinger simulations for different III-V architectures explored in this work. The higher DoS in silicon results in almost 2x higher charge.

Despite the large capacitance penalty however, III-V devices still retain their competitive advantage over Silicon due to their significantly higher υ_{inj} as shown in chapters 4 and 5. Further the injection velocity is extracted at high gate and drain bias (saturation) allowing for a fair comparison to on-state performance in Silicon devices. The InGaAs QW, InAs SQW and DQW demonstrate υ_{inj} of 1.4, 1.5 and 1.6x10^7 cm/sec respectively compared to 5x10^6 cm/sec for 22nm Si FinFETs, which is more than 3x enhancement. Since I_{ON} is proportional to product of inversion charge density and injection velocity (C_G x υ_{inj}), current can be much higher for III-V FinFET devices. Figure 6-3 shows projected performance based for scaled 22nm node devices with embedded source/drain architecture along with state of the art Silicon[20]. Parameters used for scaled device simulations are derived from TCAD models calibrated to experimental results shown in this dissertation. Additionally these simulation models incorporate embedded source/drain architectures for the III-V FinFETs which allows a one-to-one comparison with Si the FinFET.

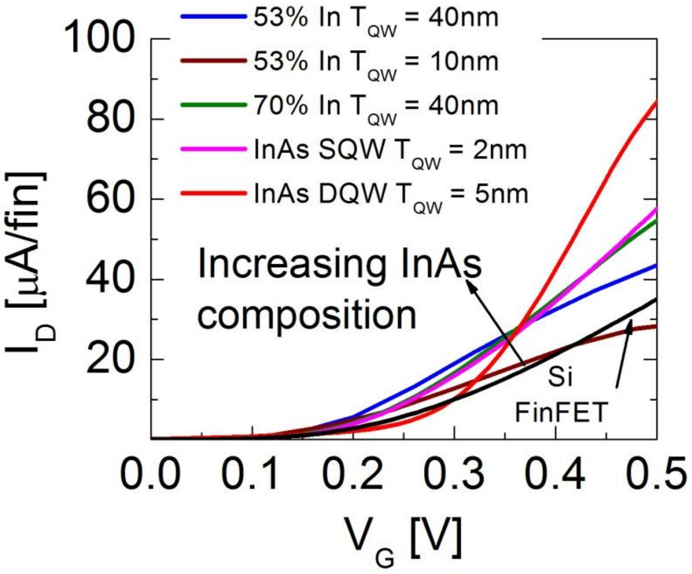


Figure 6-3. TCAD simulation of scaled devices comparing projected performance with 22nm Silicon FinFETs [20].

Impact of confinement on transport : confined phonon modes

Although υ_{inj} has been shown to be higher in the simulation results discussed so far, there are several new scattering mechanisms not considered in the simulations which could result in deviations from the projected performance. One of the most important mechanisms is the emergence of confined phonon dispersion modes in extremely scaled nanostructures[45,76,77]. For Silicon nanowires Ramayya *et al* [45] show that at low transverse fields, where phonon scattering dominates, scattering from confined acoustic phonons results in about a 10% decrease in the mobility with respect to the bulk phonon approximation. Further as the wire cross section is reduced electron mobility drops because of detrimental increase in both electron-acoustic phonon and electron-surface roughness scattering rates, eroding the benefits of volume inversion and subband modulation. Pokatilov *et al* [76] show that acoustically mismatched barriers dramatically influence the quantized phonon spectrum of GaN/AlN wurtzite nanowire structures. The barriers with lower sound velocity compress the phonon energy spectrum and reduce the phonon group velocities in the nanowire while barrier materials with higher sound velocity have an opposite effect. Also in the case of extremely confined quantum wells there is stronger coupling to confined phonon modes resulting in fast electron relaxation. This has important consequences for mobility in extremely scaled FinFET channels. However, it is also shown that it may be possible to tune the strength of the electron-phonon interaction in a desired way by varying the core and cladding layers thicknesses. This however might be at odds with the requirements dictated by electrostatic constraints in such devices thus making it a problem of choosing the appropriate trade-off. Thus it is important to keep in mind that new scattering mechanisms might potentially overshadow the benefits we may expect with high mobility materials and strong confinement (volume inversion).

Despite these cautions there still remains the tantalizing question of whether high aspect ratio, binary InAs channel (100% In content) FinFETs might be able to demonstrate the ultimate performance. However, growing thick relaxed InAs channels is key to realizing such tall fin devices. This requires careful re-design of the underlying buffer layer in order to minimize the strain (compressive bi-axial) arising from the large lattice constant of InAs (6.06A). The next session addresses some of these concerns and outlines some strategies towards realizing such devices.

II. Future work: High aspect ratio InAs channel FinFETs

Pseudomorphic growth of InAs on InP/$In_{0.52}Al_{0.48}As$ buffers is necessarily limited to QW thickness of less than 5nm in order to accommodate the strain. However, for InAs QW with T_{QW} < 10nm, mobility rolls-off as the sixth power of TQW due to surface roughness. Additionally as outlined in chapter 3, patterning into three dimensional structures causes further mobility degradation due to sidewall roughness (also see [78]). Thus in order to exploit the true potential of InAs channels it is necessary to obtain good starting material with high mobility which dictates the minimum T_{QW} to be larger than 10nm.

Several recent studies have addressed the issue of growing relatively thicker InAs QW channels for various applications ranging from optoelectronics to HEMT devices [79-82]. In this section we outline growth strategies for integrating such QWs on InP substrates which have historically enabled the highest performance HEMT devices. Thus one approach towards realizing thicker InAs channels is to grade the $In_XAl_{1-X}As$ buffer from the lattice matched 52% Indium to higher Indium percentage so as to approach the lattice constant of InAs, as shown in figure 6-4. One undesirable consequence however is the reduced conduction band offset (200meV) between the channel and buffer. This could potentially lead to large unwanted source/drain leakage currents in extremely scaled transistors.

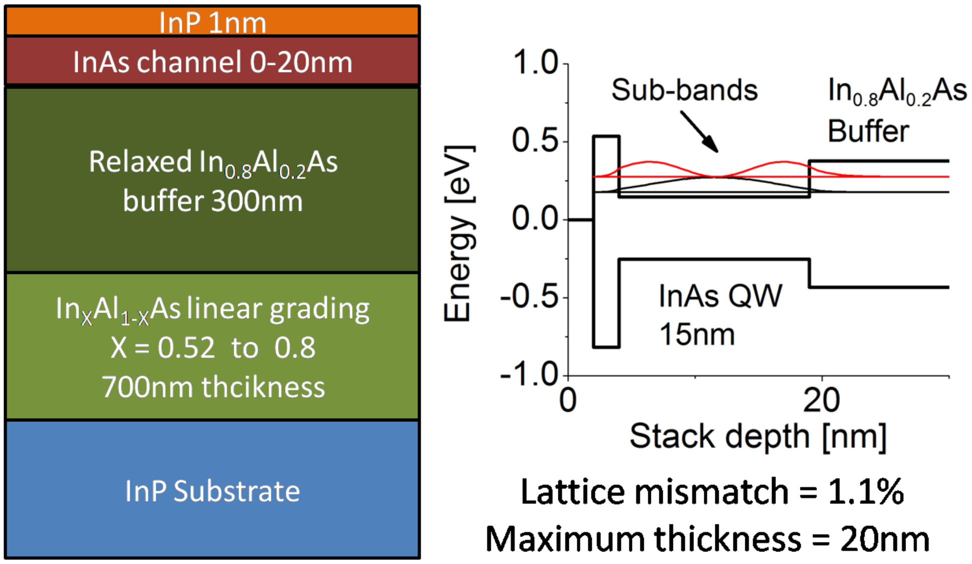

InP 1nm
InAs channel 0-20nm
Relaxed In0.8Al0.2As buffer 300nm
InXAl1-XAs linear grading
X = 0.52 to 0.8
700nm thcikness
InP Substrate
1.0
0.5
0.0
-0.5
-1.0
Energy [eV]
Sub-bands
In0.8Al0.2As Buffer
InAs QW 15nm
0
20
Stack depth [nm]
Lattice mismatch = 1.1%
Maximum thickness = 20nm

Figure 6-4. Linear grading of $In_XAl_{1-X}As$ buffer to accommodate thick pseudomorphic InAs channel (0-20nm). The lattice mismatch still limits maximum thickness in order to accommodate strain. Shown on the right is the corresponding band diagram.

As outlined in chapter 1, it is important to grow the channel material on top of a wide bandgap buffer layer that provides a large conduction band offset so as to act as a punch through stopper and suppress sub-channel leakage currents. Buffer layers that are closely lattice matched to InAs and satisfy this property are predominantly Sb-based alloys such GaSb, AlSb and finally $AlAs_{0.16}Sb_{0.84}$ which is lattice matched to InAs. Figure 6-5 shows the band alignments for these systems relative to InAs. It is evident that all of them provide large conduction band offsets to InAs as desired which implies good electron confinement within the InAs channel. Although GaSb is closely lattice matched it exhibits a broken gap alignment with InAs of the order of 150meV which is detrimental to preventing band-to-band tunneling near the source/drain regions in FET type devices. Thus it is desirable to move to a system incorporating an AlSb buffer which exhibits a nearly straddling band alignment. Figure 6-6 shows a possible layer structure using a metamorphic buffer growth. The buffer layer of AlSb is almost fully relaxed and accommodates the 8% lattice mismatch relative to the InP substrate.

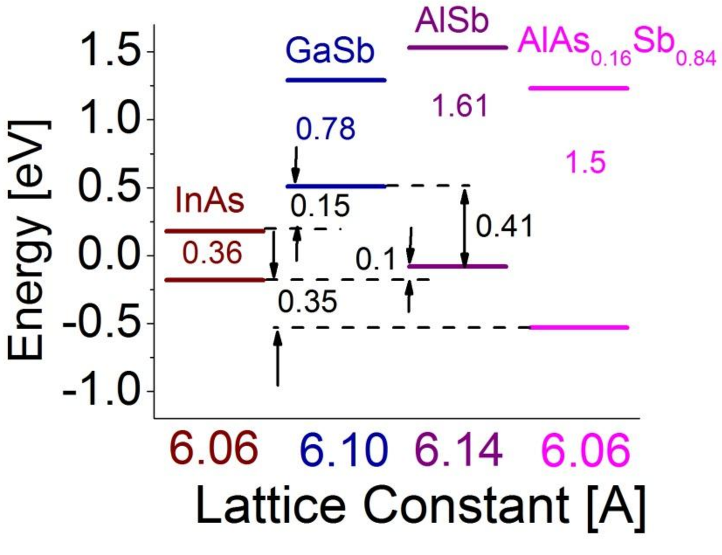


Figure 6-5. Band gaps and alignment of various Sb based buffer layers relative to InAs.

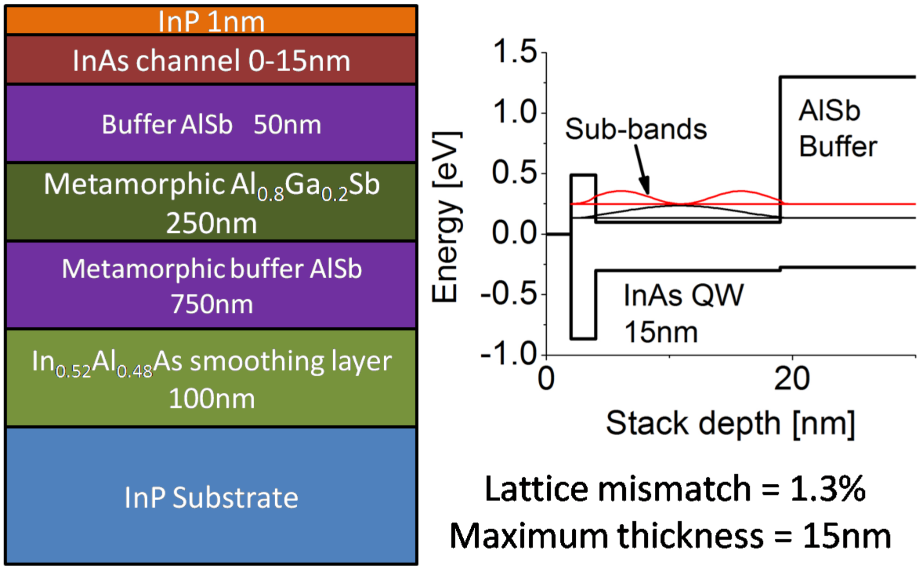

InP 1nm
InAs channel 0-15nm
Buffer AlSb 50nm
Metamorphic $Al_{0.8}Ga_{0.2}Sb$ 250nm
Metamorphic buffer AlSb 750nm
$In_{0.52}Al_{0.48}As$ smoothing layer 100nm
InP Substrate
1.5
1.0
0.5
0.0
-0.5
-1.0
Energy [eV]
Sub-bands
AlSb Buffer
InAs QW 15nm
0
20
Stack depth [nm]
Lattice mismatch = 1.3%
Maximum thickness = 15nm

Figure 6-6. A thick AlSb metamorphic buffer growth on InP/InAlAs can provide a closely lattice matched system to InAs along with large conduction band offsets required for electron confinement. Shown on the right is the corresponding band diagram.

The 250 nm layer of $Al_{0.8}Ga_{0.2}Sb$ allows for a mesa isolation etch that stops in the AlGaSb and hence prevents exposure of AlSb outside the active region to air. This is necessary as moisture causes deterioration in the AlSb layers which is in turn detrimental to device operation. Similar heterostructures have been demonstrated to have a mobility of 21,300 cm^2/V.sec at sheet charge density of $1.6x10^{12}/cm^2$ [79,80]. However, both structures discussed thus far still suffer a lattice mismatch of 1.1% compressive and 1.3% tensile strains respectively with respect to the underlying buffer which limits the maximum epitaxial thickness. In order to realize perfectly lattice matched buffer layers it is necessary to migrate from AlSb towards AlAs. Figure 6-7 shows the design of a lattice matched buffer incorporating $AlAs_{0.16}Sb_{0.84}$. Also as observed from figure 6-5 this provides excellent conduction band offset while simultaneously increasing the valence band offset favorably as well.

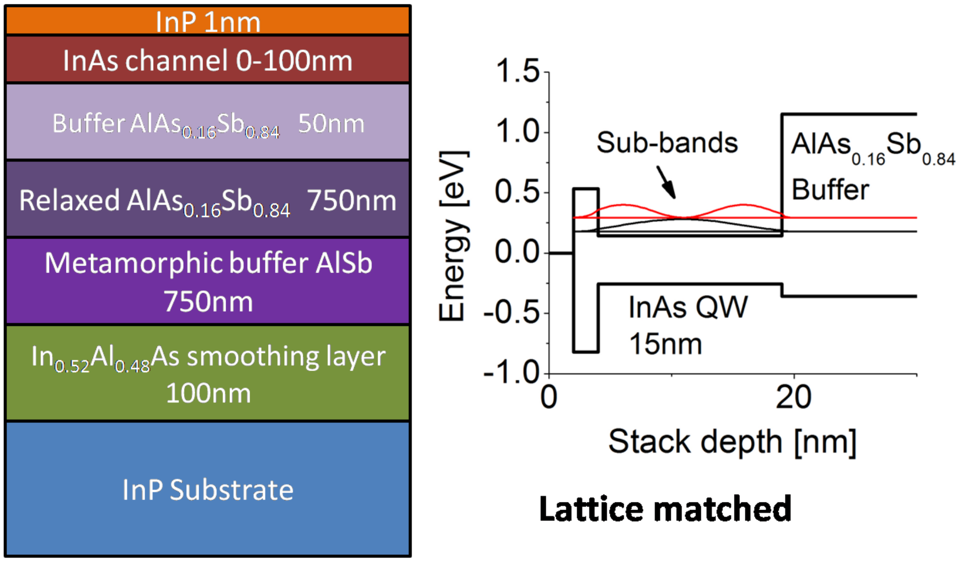

InP 1nm
InAs channel 0-100nm
Buffer AlAs$_{0.16}$Sb$_{0.84}$ 50nm
Relaxed AlAs$_{0.16}$Sb$_{0.84}$ 750nm
Metamorphic buffer AlSb 750nm
In$_{0.52}$Al$_{0.48}$As smoothing layer 100nm
InP Substrate
1.5
1.0
0.5
0.0
-0.5
-1.0
Energy [eV]
Sub-bands
AlAs$_{0.16}$Sb$_{0.84}$ Buffer
InAs QW 15nm
0
20
Stack depth [nm]
Lattice matched

Figure 6-7. Thick metamorphic AlSb growth followed by the growth of thick relaxed AlAsSb can provide a perfectly lattice matched buffer layer which also satisfies the criteria of large conduction band offset (good electron confinement). Shown on the right is the corresponding band diagram.

One of the difficulties however with the growth of AlAsSb is the large miscibility gap at the optimal growth temperatures [82]. This makes it is difficult to obtain reproducible lattice matched layers. To find the lattice matching conditions for AlAsSb on InAs was grown at a temperature of 525°C by A. Wilk *et al* in [82]. This relatively high substrate temperature was used to achieve high crystalline quality while avoiding thermal degradation of the InAs substrate.

The three layer structures discussed above provide different possible pathways for growing thick InAs channels required in order to realize high aspect ratio InAs fins. This is necessary for demonstrating high performance FinFETs capable of providing substantially superior performance compared to state of the art Si FinFET devices. Such demonstrations will firmly establish the potential for III-V materials to replace Silicon at extremely scaled technology nodes of 7nm or lower.

Bibliography

[1] M. Lundstrom, "Moore's Law forever?," *Science*, vol. 299, no. 5604, pp. 210-211, January 2003.

[2] Robert Chau, Brian Doyle, Suman Datta, Jack Kavalieros and Kevin Zhang, "Integrated nanoelectronics for the future," *Nature Materials*, vol. 6, pp. 810-812, 2007.

[3] Ravi Pillarisetty, "Academic and industry research progress in germanium nanodevices," *Nature*, vol. 479, pp. 324-328, November 2011.

[4] J. A. del Alamo, "Nanometre-scale electronics with III−V compound semiconductors," *Nature*, vol. 479, pp. 317-323, November 2011.

[5] ITRS. (2013) International Technology Roadmap for Semiconductors. [Online]. http://www.itrs.net/

[6] (2014) extremetech. [Online]. http://www.extremetech.com/

[7] Anisur Rahman, Jing Guo, and Supriyo Datta and Mark Lundstrom, "Theory of ballistic nanotransistors," *IEEE Transactions on Electron Devices*, vol. 50, no. 9, pp. 1853-1864, September 2003.

[8] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor," *Jornal of Applied Physics*, vol. 76, pp. 4879-4890, 1994.

[9] S. Luryi, "Qauntum Capacitance Devices," *Applied Physics Letters*, vol. 52, no. 6, pp. 501-503, February 1988.

[10] T. Ando, A. B. Fowler, and F. Stern, "Properties of 2-dimensional electron systems," *Reviews of Modern Physics*, vol. 54, no. 2, pp. 437-672, 1982.

[11] F. Stern and W. E. Howard, ""Properties of semiconductor surface inversion layers in the

electric quantum limit," *Phys. Rev.*, vol. 163, pp. 816-835, 1967.

[12] K. Hikosaka, S. Sasa, and N. Harada and S. Kuroda, "Current-Gain Cutoff Frequency Comparison of InGaAs HEMTs," *IEEE Electron Device Letters*, vol. 9, no. 5, pp. 241-243, May 1988.

[13] K. L. Tan et al., "60-GHz Pseudomorphic Al0.25 Ga 0.75 As /In O.2sGa 0.72 As Low-Noise HEMT's," *IEEE Electron Device Letters*, vol. 12, no. 1, pp. 23-26, January 1991.

[14] Yoshimi Yamashita et al., "Pseudomorphic In0:52Al0:48As/In0:7Ga0:3As HEMTs With an Ultrahigh fT of 562 GHz," *IEEE Electron Device Letters*, vol. 23, no. 10, pp. 573-575, October 2002.

[15] Keisuke Shinohara, Yoshimi Yamashita, Akira Endoh, Kohki Hikosaka, and Toshiaki Matsui, "Extremely High-Speed Lattice-Matched InGaAs/InAlAs High Electron Mobility Transistors," *Japanese Journal of Applied Physics*, vol. 41, no. 4B, pp. 437-439, April 2002.

[16] I. Watanabe et al., "High Transconductance of 2.25 S/mm Observed at 16K for 195-nm-Gate In0:75Ga0:25As=In0:52Al0:48As HEMT Fabricated on (411)A-Oriented InP Substrate," *IEEE Electron Device Letters*, vol. 26, no. 7, pp. 425-428, July 2005.

[17] K. Shinohara et al., "547-GHz ft In0:7Ga0:3As–In0:52Al0:48As HEMTs With Reduced Source and Drain Resistance," *IEEE Electron Device Letters*, vol. 25, no. 5, pp. 241-243, May 2004.

[18] M. W. Dvorak, and D. H. Chow C. R. Bolognesi, "High-Transconductance Delta-Doped InAs/AlSb HFET's with Ultrathin Silicon-Doped InAs Quantum Well Donor Layer," *IEEE Electron Device Letters*, vol. 19, no. 3, pp. 83-86, March 1998.

[19] Member, IEEE, Walter Kruppa, Member, IEEE, Brian R. Bennett, Doewon Park,Steven W. Kirchoefer, Member, IEEE, Robert Bass, and Harry B. Dietrich, Member, IEEE J. Brad

Boos, "AlSb/InAs HEMT’s for Low-Voltage, High-Speed Applications," *IEEE Transactions on Electron Devices*, vol. 45, no. 9, pp. 1869-1875, September 1998.

[20] C. Auth et al., "A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *IEEE Symposium on VLSI Technology*, 2012, pp. 131-132.

[21] Sun Dae Suk et al., "High Performance 5nm radius Twin Silicon Nanowire MOSFET(TSNWFET) :Fabrication on Bulk Si Wafer, Characteristics, and Reliability," in *IEEE International Electron Devices Meeting*, 2005, pp. 717-720.

[22] H. S. Wong, D. J. Frank, and Y. and Stork, J. M. C. Taur, "“Design and performance considerations for sub-0.1 um double-gate SOI MOSFET’s," in *IEEE International Electron Devices Meeting*, 1994, pp. 747-750.

[23] G. D. Wilk and R. M. and Anthony, J. M. Wallace, "High-κ gate dielectrics: Current status and materials properties considerations," *Journal of Applied Physics*, vol. 89, no. 10, pp. 5243-5275, May 2001.

[24] L. Lin and J. Robertson, "Defect states at III-V semiconductor oxide interfaces," *Applied Physics Letters*, vol. 98, p. 082903, 2011.

[25] Evgueni A. and Kummel, Andrew C. Chagarov, "Density functional theory simulations of amorphous high-κ oxides on a compound semiconductor alloy: a-Al2O3/InGaAs(100)-(4×2), a-HfO2/InGaAs(100)-(4×2), and a-ZrO2/InGaAs(100)-(4×2)," *The Journal of Chemical Physics*, vol. 135, pp. 244705 1-17, December 2011.

[26] P. D. Ye et al., "GaAs metal–oxide–semiconductor field-effect transistor with nanometer-thin dielectric grown by atomic layer deposition," *Applied Physics Letters*, vol. 83, no. 1, pp. 180-182, July 2003.

[27] Martin M. Frank et al., "HfO2 and Al2O3 gate dielectrics on GaAs grown by atomic layer deposition," *Applied Physics Letters*, vol. 86, p. 152904, 2005.

[28] J. and Falabretti, B. Robertson, "Band offsets of high K gate oxides on III-V semiconductors," *Journal of Applied Physics*, vol. 100, no. 014111, pp. 1-8, July 2006.

[29] Varistha Chobpattana et al., "Nitrogen-passivated dielectric/InGaAs interfaces with sub-nm equivalent oxide thickness and low interface trap densities," *Applied Physics Letters*, vol. 102, p. 022907, 2013.

[30] A. Ali et al., "Small-Signal Response of Inversion Layers in High-Mobility In0.53Ga0.47As MOSFETs Made With Thin High-K Dielectrics," *IEEE Transactions on Electron Devices*, vol. 57, no. 4, pp. 742-748, April 2010.

[31] M. Radosavljevic et al., "Electrostatics Improvement in 3-D Tri-gate Over Ultra-Thin Body Planar InGaAs Quantum Well Field Effect Transistors with High-K Gate Dielectric and Scaled Gate-to-Drain/Gate-to-Source Separation.," in *International Electron Device Meeting*, 2011, pp. 33.1.1-33.1.4.

[32] J. J. Gu et al., "20−80nm Channel Length InGaAs Gate all-around Nanowire MOSFETs with EOT=1.2nm and Lowest SS=63mV/dec.," in *International Electron Devices Meeting*, 2012, p. 27.6.1−27.6.4.

[33] R. Kotlyar, B. Obradovic, P. Matagne, M. Stettler, and M. D. Giles, "Assessment of room-temperature phonon-limited mobility in gated silicon nanowires.," *Applied Physics Letters*, vol. 84, no. 25, p. 5270−5272., June 2004.

[34] A. C. Ford et al., "Diameter-Dependent Electron Mobility of InAs Nanowires.," *Nano Letters*, vol. 9, no. 1, pp. 360-365, 2009.

[35] k. Tomioka, M. Yoshimura, and T. Fukui, "A III−V nanowire channel on silicon for high-

performance vertical transistors.," *Nature*, vol. 488, pp. 189-192, May 2012.

[36] A. W. Dey, J. Svensson, B. M. Borg, M. Ek, and L.-E. Wernersson, "Single InAs/GaSb Nanowire Low-Power CMOS Inverter.," *Nano Letters*, vol. 12, pp. 5593-5597, 2012.

[37] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Sub-30 nm InAs Quantum-Well MOSFETs with Self-aligned Metal Contacts and Sub-1 nm EOT HfO2 Insulator," in *Internatinal Electron Devices Meeting*, 2012, p. 32.1.1−32.1.14.

[38] O. Gunawan et al., "Measurement of Carrier Mobility in Silicon Nanowires," *Nano Letters*, vol. 8, no. 6, pp. 1566-1571, 2008.

[39] D. Shahrjerdi et al., "Hall mobility measurements in enhancement-mode GaAs field-effect transistors with Al2O3 gate dielectric.," *Applied Physics Letters*, vol. 97, p. 213506, 2010.

[40] D. Veksler et al., "Quantification of interfacial state density (Dit) at the high-k/III-V interface based on Hall effect measurements," *Journal of Applied Physics*, vol. 112, p. 054504, 2012.

[41] K. Storm et al., "Spatially resolved Hall effect measurement in a single semiconductor nanowire," *Nature Nanotechnology*, vol. 7, pp. 718-722, 2012.

[42] Ch. Blomers et al., "Hall effect measurements on InAs nanowires.," *Applied Physics Letters*, vol. 101, p. 152106, 2012.

[43] J. P. DeGrave, D. Liang, and S. Jin, "A General Method to Measure the Hall Effect in Nanowires: Examples of FeS2 and MnSi.," *Nano Letters*, vol. 13, pp. 2704-2709, 2013.

[44] Synopsys Inc., "Sentaurus Device User Guide," User Guide version f-2011.09 ed.;, 2011.

[45] E. B. Ramayya, D. Vasileska, S. M. Goodnick, and I. J. Knezevic, "Electron Transport in Silicon Nanowires: The Role of Acoustic Phonon Confinement and Surface Roughness Scattering," *Journal of Applied Physics*, vol. 104, p. 063711, 2008.

[46] S. Poli, M. G. Pala, T. Poiroux, S. Deleonibus, and G. Baccarani, "Size Dependence of Surface-Roughness-Limited Mobility in Silicon-Nanowire FETs.," *IEEE Transactions on Electron Devices*, vol. 55, no. 11, pp. 2968-2975, 2008.

[47] S. Chuang et al., "Ballistic InAs Nanowire Transistors.," *Nano Letters*, vol. 13, pp. 555-558, 2013.

[48] J. J. Hou et al., "Diameter dependence of electron mobility in InGaAs nanowires.," *Applied Physics Letters*, vol. 102, p. 093112, 2013.

[49] 3nextnano ©, http://www.nextnano.de.

[50] J. Wang and M. Lundstrom, "Ballistic Transport in High Electron Mobility Transistors," *IEEE Transactions on Electron Devices*, vol. 50, no. 7, pp. 1604-1609, 2003.

[51] K. von-klitzing, G. Dorda, and M. Pepper, "New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance.," *Physical Review Letters*, vol. 45, pp. 494-497, 1980.

[52] D. C. Tsui, H. L. Stormer, and A. C. Gossard, "Two-Dimensional Magnetotransport in the Extreme Quantum Limit.," *Physical Review Letters*, vol. 48, pp. 1559-1562, 1980.

[53] R. de Picciotto, H. L. Stormer, L. N. Pfeiffer, K. W. Baldwin, and K. W. West, "Four-terminal resistance of a ballistic quantum wire.," *Nature*, vol. 411, pp. 51-53, 2001.

[54] M. Z. Hasan and C. L. kane, "Topological insulators.," *Reviews of Modern Physics*, vol. 82, pp. 3045-3067, 2010.

[55] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, "Ballistic carbon nanotube transistors," *Nature*, vol. 424, pp. 654-657, 2003.

[56] A. M. Gilbertson et al., "Room temperature ballistic transport in InSb quantum well nanodevices," *Physical Review B*, vol. 83, p. 075304, 2011.

[57] A. S. Mayorov et al., "Micrometer-scale ballistic transport in encapsulated graphene at room temperature," *Nano Letters*, vol. 11, pp. 2396-2399, 2011.

[58] V. T. Arun et al., "Investigation of InxGa1−xAs FinFET architecture with varying indium (x) concentration and quantum confinement," in *IEEE Symposium on VLSI Technology*, 2014, pp. 978 1-2.

[59] G. Doornbos and M. Passlack, "Benchmarking of III–V n-MOSFET maturity and feasibility for future CMOS," *IEEE Electron Device Letters*, vol. 31, no. 10, pp. 1110-1112, October 2010.

[60] S.W. Chang et al., "InAs N-MOSFETs with record performance of Ion = 600 μA/μm at Ioff = 100 nA/μm (Vd = 0.5 V)," in *International Electron Devices Meeting*, 2013, pp. 16.1.1-16.1.4.

[61] S. Lee et al., "Record Ion (0.50 mA/μm at VDD = 0.5 V and Ioff = 100 nA/μm) 25 nm-gate-length ZrO2/InAs/InAlAs MOSFETs," in *IEEE Symposium on VLSI Technology*, 2014, pp. 1-2.

[62] T.-W. Kim et al., "Sub-100 nm InGaAs quantum-well (QW) tri-gate MOSFETs with Al2O3/HfO2 (EOT < 1 nm) for low-power logic applications," in *International Electron Devices Meeting*, 2013, pp. 16.3.1-16.3.4.

[63] Sang-Hyeon Kim et al., "High Performance Tri-Gate Extremely Thin-Body InAs-On-Insulator MOSFETs With High Short Channel Effect Immunity and Vth Tunability," *IEEE Transaction on Electron Devices*, vol. 61, no. 5, pp. 1354-1360, May 2014.

[64] Jianqiang Lin, Xin Zhao, Tao Yu, Dimitri A. Antoniadis, and Jesús A. del Alamo, "A New Self-aligned Quantum-Well MOSFET Architecture Fabricated by a Scalable Tight-Pitch Process," in *International Electron Devices Meeting*, 2013, pp. 421-424.

[65] Claes Thelander, Linus E. Fröberg, Carl Rehnsted, and Lars Samuelson and Lars-Erik Wernersson, "Vertical Enhancement-Mode InAs Nanowire Field-Effect Transistor With 50-nm Wrap Gate," *IEEE Transactions on Electron Devices*, vol. 29, no. 3, pp. 206-208, March 2008.

[66] X. Zhou et al., "In0.53Ga0.47As quantum-well MOSFET with source/drain regrowth for low power logic applications," in *IEEE Symposium on VLSI Technology*, 2014, pp. 1-2.

[67] Ioffe physico-technical Institute. Electronic archive: New Semiconductor Materials. Characteristics and Properties. http://www.ioffe.ru/SVA/.

[68] Yasuyuki Kobayashi and Naoki Kobayashi, "In Situ Interface Control of Pseudomorphic InAs/InP Quantum Well Structure Growth by Surface Photo-Absorption," *Japanese Journal of Applied Physics*, vol. 31, no. 12A, pp. 3988-3994, September 1992.

[69] Cristopher C. Eugster, Tom P. E. Broekaert, and Jesh A. del Alamo and Clifton G. Fonstad, "An InAlAs/InAs MODFET," *IEEE Electron Device Letters*, vol. 12, no. 12, pp. 707-709, December 1991.

[70] Dae-Hyun Kim and Jesus A. del Alamo, "Logic Performance of 40 nm InAs HEMTs," in *International Electron Devices Meeting*, 2007, pp. 629-632.

[71] Dae-Hyun Kim and Jesús A. del Alamo, "30-nm InAs Pseudomorphic HEMTs on an InP Substrate With a Current-Gain Cutoff Frequency of 628 GHz," *IEEE Electron Device Letters*, vol. 29, no. 8, pp. 830-833, August 2008.

[72] A. V. Thathachary et al., "Impact of Sidewall Passivation and Channel Composition on InxGa1−xAs FinFET Performance," *IEEE Electron Device Letters*, vol. 36, no. 2, pp. 117-119, February 2015.

[73] Y. Sun et al., "High-Performance CMOS-Compatible Self-Aligned In0.53Ga0.47As

MOSFETs with GMSAT over 2200 μS/μm at VDD = 0.5 V," in *IEEE International Electron Devices Meeting*, 2014, pp. 582-585.

[74] U. Singisetti et al., "Ultralow resistance in situ Ohmic contacts to InGaAs/InP," *Applied Physics Letters*, vol. 93, pp. 183502-4, November 2008.

[75] Amlan Majumdar, "Semiconductor Capacitance Penalty per Gate in Single- and Double-Gate FETs," *IEEE Electron Device Letters,* vol. 35, no. 6, pp. 609-611, June 2014.

[76] E. P. Pokatilov and D. L. Nika and A. A. Balandin, "Acoustic-phonon propagation in rectangular semiconductor nanowires with elastically dissimilar barriers," *Phys. Rev. B,* vol. 72, p. 113311, 2005.

[77] Evgenii P. Pokatilov, Denis L. Nika, and and Alexander A. Balandin, "Confined electron-confined phonon scattering rates in wurtzite AlN/GaN/AlN heterostructures," *Journal of Applied Physics*, vol. 95, no. 10, pp. 5626-5632, May 2004.

[78] Herbert Kroemer, "The 6:1 A family (InAs, GaSb, AlSb) and its heterostructures: a selective review," *Phisica E,* vol. 20, pp. 196-203, 2003.

[79] Brian R. Bennett, Richard Magno, J. Brad Boos, and Walter Kruppa and Mario G. Ancona, "Antimonide-based compound semiconductors for electronic devices: A review," *Solid State Electronics*, vol. 49, pp. 1875-1895, November 2005.

[80] X. Wallart, J. Lastennet, and D. Vignaud and F. Mollot, "Performances and limitations of InAs/InAlAs metamorphic heterostructures on InP for high mobility devices," *Applied Physics Letters*, vol. 87, pp. 043504 1-3, July 2005.

[81] G. Moschetti et al., "DC characteristics of InAs/AlSb HEMTs at cryogenic temperatures," in *IEEE conference on Indium Phosphide and related materials (IPRM)*, 2009, pp. 323-325.

[82] A. Wilk et al., "MBE growth of InAs/InAsSb/AlAsSb structures for mid-infrared lasers,"

Journal of Crystal Growth, vol. 223, pp. 341-348, 2001.

Appendix

Process flow for III-V FinFET fabrication

E-beam litho markers + S/D pads

- MMA EL 11 4000rpm 45secs + 150C 3mins; PMMA A3 4000rpm 45secs + 180C 3mins;
- e-beam litho dose = 410uC/cm^2
- Develop
  - 1:3 MIBK: IPA for 3mins
  - 1:1 MIBK:IPA for 10secs
  - IPA for 1min (no DI rinse after)
  - N_2 Blow dry
- Sputter metal deposition
  - Mo (pwr1 gun1) 5mTorr; 200W; 400secs; 18C
  - Pt (pwr3 gun3) 5mTorr; 200W; 400secs; 18C
  - Total thickness = 100nm
- Lift-off in hot ACETONE (80C) for 15mins; hot PRS 3000 for 20mins + IPA rinse + DI rinse + N2 blow dry

Mesa isolation

- Surpass 3000/4000 1min soak + DI rinse + 100°C 1min bake
- ZEP 520A ***2500***rpm 60secs spin + 180°C bake for 3min; e-beam litho dose = 320uC/cm^2
- Develop in N-Amyl Acetate 3mins; IPA 1min;N_2 blow dry
- *Etch Recipe name:* Baseline Si Trench etch recipe in PM1(Versalock)

- Cl_2 30sccm; 700W Coil; 150W chuck; 2mTorr; Total etch time = 18secs; (Depth = 90nm (total thickness of Mo + InGaAs)
- Strip in PRS 3000 (80°C) + IPA rinse + DI rinse + N_2 blow dry; Etch depth = 150nm

Gate recess etch

- ALD Al_2O_3 110°C ; 300 cycles; measured thickness = 225A
- Surpass 3000 (or 4000) 1min + DI rinse + 100°C 1min bake
- ZEP 520A 2500rpm 60secs + 180°C bake 3mins; e-beam litho dose = 108x4 uC/cm^2 for small features; 350 uC/cm^2 for large features
- Develop in N-Amyl acetate 3mins ; IPA dip 1min; N_2 blow dry
- Etch recipe name: PM2 SiO_2 Baseline etch recipe for Al_2O_3
- CF_4 50sccm; 700W Coil; RF1 150W chuck (325V); 4mTorr; *30secs* etch
- $HCl:H_2O$ 1:3 (to etch n++ doped InP cap for 60secs)
- NOTE: Al_2O_3 hard mask withstands above etch for >90 secs (selectivity >100)
- Strip in PRS 3000 (80°C) + IPA rinse + DI rinse + N_2 blow dry

Fin Definition

- Acetone + IPA clean
- ZEP 1;1 3000 rpm 50 secs + 180°C bake 3mins; e-beam litho dose = 50x4 uC/cm^2 for small features; feature dose scaling included dose1 = 1; dose2 = 1.1x; dose3 = 1.2x
- Develop in N-Amyl acetate 3mins ; IPA dip 1min; N_2 blow dry gentle
- Etch recipe name: ULVAC BLE InGaAs T etch recipe

- Chuck temperature = Room Temp ; Cl_2 6sccm; N_2 24sccm; 5mTorr; 900W coil; 60W chuck; Etch time = 16secs (etch depth = 23nm @ RT)
- ZEP removal; Hot PRS 3000 (80C) for 10mins + IPA rinse + DI rinse + N2 blow dry
- ***Etch damage anneal***
- RTA tool 375°C; FGA ambient; 10 litres/min flow; 15mins

Gate Stack: High-k deposition

- 10:1 BOE dip 3mins; DI H_2O rinse; N_2 blow dry; Transfer into load-lock with minimal delay
- Cluster tool ALD chuck temperature at 332°C
- TMA + N_2 plasma recipe
- Channel A: 0.04secs TMA; 10secs purge;
- Channel C: 6 secs N_2 plasma; 10 secs purge; 116sccm Ar + 4sccm N_2; 125W power; 6 cycles;
- NOTE: remove TMA pulse for first cycle
- **Gate stack growth**
- Reduce chuck temperature to 271°C
- Run HfO_2 deposition recipe
- *Channel A:* 0.2 secs TDMAH pulse; 10secs purge; *Channel B:* 0.3secs H_2O; 20secs purge

Gate electrode definition

- MMA EL 11 3000rpm 45secs + 150C 3mins; PMMA A3 3000rpm 45secs + 180C 3mins;

- e-beam litho dose = 400uC/cm^2
- Develop in 1:3 MIBK: IPA for 3mins + 1:1 MIBK:IPA for 10secs followed by IPA rinse for 1min (no DI rinse after) +N_2 Blow dry
- **Thermal Ni evaporation + e-beam gold evaporation**
- Use Al_2O_3 coated tungsten boat; 2 Ni pellets
- Semicore evaporation: 48% power for 0.3A/sec (Lab 18 evaporation: 85% power for 0.2A/sec)
- NOTE: Lab 18 gives conformal coating with rotation; ***load only 1 small pellet***
- Total thickness = 600A Ni + 800A Gold
- **Lift-off:** dip in hot acetone (85°C) 15mins + Hot PRS 3000 15mins + IPA rinse + DI H_2O rinse + N_2 dry
- **Forming Gas Anneal**
- Chuck temp. ***375°C*** set point; 75sccm Ar + 100sccm H_2; *Anneal time = 20mins* including equilibration

VITA

Arun V. Thathachary

Arun V. Thathachary hails from Bengaluru (formerly Bangalore), renowned as the Silicon Valley of India and the present capital of the southern Indian state of Karnataka. He completed his elementary schooling and high school from Kendriya Vidhyalaya, IISc campus, Bengaluru. Arun received his B.S. degree in Electronics and Communication Engineering from the M. S. Ramaiah Institute of Technology, Bengaluru, India, in 2006 and then went to pursue Masters degree at the Indian Institute of Science (IISc), Bengaluru, India. He received his M.S. degree from the department of Electrical Communication Engineering in 2009 and continued there as a research assistant from 2009 to 2011.

He then moved to Pennsylvania State University, University Park, PA and joined the research group of Dr. Suman Datta in 2011 where he worked towards his doctorate degree in the field of III-V compound semiconductor FinFETs for low power logic applications. He obtained the Ph.D. degree in 2015 from the Department of Electrical Engineering at Penn State. After his PhD, he is headed to Intel Corp. where he will be working as Device Engineer in the NAND Flash memory group in Santa Clara, CA.

During his PhD, Arun published 8 journal and conference papers with 5 first author publications among them. He was awarded the Melvin P. Bloom Outstanding Doctoral Research Award in Electrical Engineering. The award recognizes outstanding achievement in scholarship and professional accomplishment, and he was honored to be one of the three graduate students within Electrical Engineering at Penn State to receive the award. He can be contacted at his email arunvthathachary@gmail.com.