Copyright

by

Aqyan Ahmed Bhatti

2019

The Dissertation Committee for Aqyan Ahmed Bhatti Certifies that this is the approved version of the following Dissertation:

Advanced semi-classical Monte Carlo modeling of Si, Ge, InGaAs, and MoS_2 n-channel FETs for novel CMOS

Committee:

Sanjay K. Banerjee, Supervisor

Leonard F. Register, Co-Supervisor

Ananth Dodabalapur

Li Shi

Advanced semi-classical Monte Carlo modeling of Si, Ge, InGaAs, and MoS_2 n-channel FETs for novel CMOS

by

Aqyan Ahmed Bhatti

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2019

For all the teachers I ever had, who gave me the gift of knowledge

Acknowledgements

First, I would like to express thanks to my supervisors, Drs. Sanjay K. Banerjee and Leonard F. Register, for their mentorship and guidance. I was empowered to pursue my own independent research, which enabled me to grow and become a more confident and creative researcher. Dr. Banerjee deeply cares for his graduate students, whether that is securing financial support to assisting us in achieving our academic and professional goals and understands the big picture of scientific research. Meanwhile, Dr. Register understands physics at such a granular level and always seems to ask the right questions, that I was truly nourished by our discussions and humored by the concept of a universal scaling factor and the dilemma of a missing factor of two or negative sign. I would also like to thank my committee members, Drs. Li Shi and Ananth Dodabalapur for their helpful criticism and insightful comments. A kind thank you to Dr. S.V. Sreenivasan for graciously funding me for one year after my initial supervisor left and Dr. Desiderio Kovar for convincing me to join the Texas Materials Institute for my graduate studies. Moreover, Dr. Kovar also taught me to always maintain a positive outlook when performing research and to view its lows as a glass half full, and that, even if you think do not have any water, remember, at least you have glass. An apt lesson one can apply to life as well.

I also want to thank the lab members who joined me during this experience of learning, researching, and writing. In particular, Dax M. Crum for taking me under his tutelage to gently introduce me to the Monte Carlo method and training me to use the Monte Carlo software. Your patience during the long phone calls and lengthy emails was much appreciated. I want to express my gratitude to Jean Toll, Krista Seidel, Michelle

Mansolo, and the entire Microelectronics Research Center and Texas Materials Institute staff for their tireless work behind-the-scenes. Graduate school is hard enough, and their work makes life easier. Thank you to the Texas Advanced Computing Center for providing computational time and their prompt consulting staff. Also, I could have never started nor finished my Ph.D. without receiving a Cockrell School of Engineering Graduate Fellowship, which is made possible by the generous financial donations of alumni and friends.

I am forever grateful to my family. My brother, Jahshan A. Bhatti, for whose presence and wisdom has motivated me to challenge myself. Whenever I would learn a new concept, I would hurriedly run to my brother as if I found a hidden treasure, and he would try to extend what I learned further. I will never forget working on physics and mathematics problems for fun at the dining table with a pen and paper. Finally, I want to recognize the first teacher I ever had, my mother, Khaleda P. Solengi, who stressed the value of education, and hence the reason why I embarked on this journey. Her financial, emotional, and physical sacrifices and enduring love, prayers, and support are unparalleled. You have always believed in me when no else would, even myself. I love you both very dearly.

To my fellow engineers, I conclude with two pieces of advice, which proved to be very useful for me. First, if you cannot prove it, do not use it—that is, do not incessantly memorize equations, but rather understand how they are derived from fundamental equations and their associated assumptions. Second, always draw a picture, such as a circuit, free-body, band, or thermodynamic diagram, before tackling any problem.

Abstract

Advanced semi-classical Monte Carlo modeling of Si, Ge, InGaAs, and MoS_2 n-channel FETs for novel CMOS

Aqyan Ahmed Bhatti, Ph.D.

The University of Texas at Austin, 2019

Supervisor: Sanjay K. Banerjee

Co-Supervisor: Leonard F. Register

Scaling-down of silicon (Si) based complementary-metal-oxide-semiconductor (CMOS) technologies are approaching material limits. For high-performance applications, high thermal velocity channel materials, such as indium-gallium-arsenide (InGaAs) and germanium (Ge), are viable alternatives to Si to extend the limits of CMOS downscaling. The unique mechanical and electrical properties of two-dimensional atomic crystals, such as single-layer molybdenum disulfide (MoS_2), combined with soft, flexible, and curvilinear substrates, enable new device functionalities and concepts in the field of low-power flexible electronics not achievable with Si channels. While the intrinsic electron mobility of MoS_2 is rather low, strain engineering may provide a pathway for improving electron transport.

Silicon, InGaAs, Ge, and MoS_2 n-channel MOSFETs were explored via first-principles computational tools including density functional theory and particle-based ensemble semi-classical Monte Carlo methods to better understand and enable the rational design of end-of-the-roadmap CMOS and potential beyond-CMOS technologies.

The impact of contact geometry and transmissivity and gate length scaling on quasi-ballistic nanoscale Si, Ge, and InGaAs n-channel FinFETs was studied. FinFETs with end, saddle/slot, and raised source and drain contacts and the same saddle/slot contact geometry with different gate lengths, according to the projections of industry roadmaps, were simulated. Simulated Si FinFETs exhibited relatively limited degradation in performance due to non-ideal contact transmissivities, more limited sensitivity to contact geometry with non-ideal contact transmissivities, some contact-related advantage for Si $\langle 110 \rangle$ channel devices, and limited sensitivity to gate length scaling. Simulated InGaAs FinFETs were highly sensitive to modeled contact geometry, specific contact resistivity, the band structure model, and gate length scaling. Simulated Ge FinFETs showed substantial degradation due to non-ideal contact transmissivities, sensitivity to gate length scaling, and a large orientation-related advantage for Ge $\langle 110 \rangle$ channel devices. The impact of tensile strain on the intrinsic performance limits of monolayer MoS_2 n-channel MOSFETs was studied. 200 and 15 nm gate length MoS_2 MOSFETs with end contacts subject to different types and amounts of strain were simulated. Simulated MoS_2 MOSFETs displayed improved performance with strain due to lower effective mass and larger inter-valley separation, which is largely reduced due to non-ideal contact transmissivities.

Table of Contents

List of Tables xii

List of Figures xiii

Chapter 1: Introduction and Background 1

1.1 Marching towards the end-of-the-roadmap 1

1.2 Advancing CMOS beyond the Si roadmap 1

1.3 A fork in the road 4

1.4 Semiconductor device modeling and simulation 6

1.5 Dissertation overview 8

Chapter 2: University of Texas Monte Carlo Software 9

2.1 Monte Carlo history 9

2.2 Simulation methodology 10

2.3 Surface roughness 13

2.4 On contacts 16

2.5 Time evolution 18

Chapter 3: Semi-Classical Monte Carlo Study of the Impact of Contact Geometry and Transmissivity on Quasi-Ballistic Nanoscale Si and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs 22

3.1 Introduction and background 22

3.2 Simulated FinFET structures and InGaAs band structure models 27

3.2.1 FinFET structure 27

3.2.2 InGaAs band structure models 30

3.3 Common performance measures and results for unity transmissivity contacts 31

3.3.1 Transconductance, g_m 34

3.3.2 *S*, DIBL, turn-on characteristic, and on-state current 39

3.3.3 Drain current vs. drain voltage 40

3.4 Non-unity transmissivity contacts 41

3.4.1 Peak g_m, ΔV_T, and I_{on} 42

3.4.2 Drain current vs. drain voltage 44

3.5 Conclusion 47

Chapter 4: Semi-Classical Monte Carlo Study of Gate Length Scaling Impact on Quasi-Ballistic Nanoscale Si, Ge, and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs................. 48

4.1 Introduction and background 48

4.2 Simulated FinFET structure and band structure models.................................. 50

4.2.1 FinFET structure 50

4.2.2 Band structure models 52

4.3 Design rule for electrostatic integrity 54

4.4 Effect of S/D doping concentration in Si 55

4.5 Gate length scaling impact on common performance measures for unity transmissivity contacts 57

4.5.1 Transconductance, g_m 60

4.5.2 *S*, DIBL, turn-on characteristic, and on-state current 67

4.5.3 Drain current vs. drain voltage 68

4.6 Non-unity transmissivity contacts 69

4.6.1 Peak g_m, ΔV_T, and I_{on} 70

4.6.2 Drain current vs. drain voltage 72

4.7 Conclusion 75

Chapter 5: Semi-Classical Monte Carlo Study of the Impact of Tensile Strain on the Intrinsic Performance Limits of Monolayer MoS_2 n-channel MOSFETs 77

5.1 Introduction and background 77

5.2 Simulated MOSFET structure and band structure models 80

5.2.1 MOSFET structure 80

5.2.2 MoS_2 band structure models 82

5.2.3 MOSFET simulation methodology (essential elements) 85

5.3 Electronic band structure 87

5.4 Bulk drift velocity 91

5.5 Common performance measures and results for unity transmissivity contacts. 95

5.5.1 Transconductance, g_m 98

5.5.2 Drain current vs. drain voltage 99

5.6 Non-unity transmissivity contacts 102

5.6.1 Peak g_m and drain current vs. drain voltage 103

5.7 Conclusion 105

Chapter 6: Conclusion 108

6.1 Dissertation recap 108

6.2 Recommendations for future work 112

Appendix 116

References 121

Vita 135

List of Tables

Table 1.1: Basic material parameters and electrical properties of Si, Ge, and InGaAs compiled from [15]. 3

Table 3.1: Modeled FinFET dimensions. 30

Table 4.1: Modeled FinFET dimensions. 51

Table 5.1: Modeled MOSFET dimensions. 81

List of Figures

Figure 1.1: Intrinsic electron mobility versus band gap for various semiconductors, including typical III-V and transition metal dichalcogenides materials [48], [49]. 6

Figure 2.1: Comparison between UTMC electron mobility as a function of the effective electric field obtained by bulk simulations considering surface roughness scattering as well as quantum-confined phonon scattering and the experimental universal mobility curves for bulk Si-SiO_2 interface channel MOSFETs [60], [70]. 15

Figure 2.2: Flowchart of a self-consistent Monte Carlo device simulation. 19

Figure 3.1: Schematics of the simulated FinFET geometries with (a) reference end contacts, (b) saddle/slot contacts, and (c) raised source and drain (RSD) geometries. For each, a side view (lower left), a top view (top), and an end view (right) are shown. The spacer regions are not shown in order to show the underlying semiconductor fin, shaded in grey. The hatched region represents the gate metal. The gate oxide located underneath the gate metal is visible in the end views of end and saddle/slot contact FinFETs. The source and drain contact surfaces are shown in black. For the saddle/slot geometry, the source and drain contacts extend further to the side and above than shown, to the edge of the simulation region; however, only the near-source/drain-surface portions are shown for visual clarity. 29

Figure 3.2: I_{DS}-V_{GS} simulation results for L_G = 18 nm Si ⟨110⟩ (open circles), Si ⟨100⟩ (solid circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), and Γ-$In_{0.53}Ga_{0.47}As$ (open squares) FinFETs for (a) end injection, (b) saddle/slot, and (c) raised source and drain. V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method.................... 33

Figure 3.3: Dependence of (a) (centered moving average of) the peak of the transconductance g_m, (b) subthreshold swing S, (c) turn-on transition voltage ΔV_T, (d) on-current for the constant current defined threshold, I_{on}(CC), and (e) drain-induced barrier lowering, DIBL, for the end, saddle/slot, and RSD contacts to an 18 nm gate length FinFETs. 34

Figure 3.4: Injection of ballistic electrons under flat-band conditions into the source extension from perfectly injecting and absorbing saddle (left) and end (right) contacts, and equivalent contact geometries under these conditions. Note for the saddle contact geometry the direct injecting paths, illustrated by Path 1, and the reflection-mediated injection paths, illustrated by Path 2, and their counterpart paths in the equivalent geometry. There are corresponding paths into the source extension for the end contact, again illustrated by Paths 1 and 2, although those represented by Path 1 are now reflection-mediated and those represented by Path 2 are direct injecting, and their counterparts in its equivalent geometries. However, there is another category of paths for the end contact geometry, illustrated by Path 3, for which there are no counterparts for the saddle geometry. For these paths, electrons are effectively injected from an end contact at a distance $2L_c$ from the source extension, directly into the source extension. 38

Figure 3.5: (a) Alignment of the conduction channel relative to the Si conduction band energy valleys for (on the left) ⟨100⟩ and (on the right) ⟨110⟩ channel orientations on a {100} substrate. (b) UTMC-simulated carrier injection probability density per degree of the carrier injection angle with respect to the plane of the channel for channel-end-injected carriers, for Si ⟨100⟩ (solid line) and ⟨110⟩ (dashed line) channel orientations. 39

Figure 3.6: I_{DS}-V_{DS} simulation results for L_G = 18 nm Si ⟨110⟩ (open circles), Si ⟨100⟩ (solid circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), and Γ-$In_{0.53}Ga_{0.47}As$ (open squares) channel FinFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage for (a) end injection, (b) saddle/slot, and (c) raised source and drain. 41

Figure 3.7: Comparison of contacts with perfect transmissivity and imperfect transmissivity contacts on the (a) peak of the transconductance g_m, (b) turn-on transition voltage ΔV_T, and (c) the on-current I_{on} with a constant current defined threshold (I_{on}(CC)), for the end, saddle/slot, and RSD contacts to an 18 nm gate length FinFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.2 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system, including channel orientation for Si. 44

Figure 3.8: Simulated I_{DS}-V_{DS}, as for Fig. 3.6 but with non-unity transmissivity contacts: I_{DS}-V_{DS} simulation results for L_G = 18 nm Si ⟨110⟩ (open circles), Si ⟨100⟩ (solid circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), and Γ-$In_{0.53}Ga_{0.47}As$ (open squares) FinFETs. The gate overdrive voltage above the constant current threshold voltage is 0.35 V with supply V_{DD} and threshold V_T voltages of 0.6 V and 0.25 V, respectively, for (a) end injection, (b) saddle/slot (where the MV-InGaAs and Γ-InGaAs data are difficult to distinguish), and (c) raised source and drain. For InGaAs with end contacts in particular, the lack of stretch-out of this I_{DS}-V_{DS} characteristic with respect the saturation drain voltage ($V_{DS,sat}$) as compared to the data of Fig. 3.6 contrasts to what would be expected using a lumped external resistance or distributed specific contact resistivity. 46

Figure 4.1: Schematic of the simulated modeled FinFET geometry with saddle/slot contacts. A side view (lower left), a top view (top), and an end view (right) are shown. The spacers regions are not shown in order to show the underlying semiconductor fin, shaded in grey. The hatched region represents the gate metal. The gate oxide located underneath the gate metal is visible in the end views of the saddle/slot contact model devices. We note that for the saddle/slot geometry, the source and drain contacts extend further to the side and above than shown, to the edge of the simulation region; however, only the near-source/drain-surface portions are shown for visual clarity. The source and drain contact surfaces are shown in black. 51

Figure 4.2: Dependence of (a) subthreshold swing S and (b) drain-induced barrier lowering, DIBL, in Si ⟨100⟩ (solid circles), Ge ⟨100⟩ (solid squares), and MV-InGaAs (open triangles) channel FinFETs with gate length with all other device parameters have been kept unchanged. 55

Figure 4.3: Comparison of contacts with perfect transmissivity and imperfect transmissivity contacts on the (a) peak of the transconductance g_m, (b) turn-on transition voltage ΔV_T, and (c) the on-current I_{on} with a constant current defined threshold (I_{on}(CC)), for 18 nm gate length Si ⟨110⟩ FinFETs with S/D doping concentrations of 5×10^{19} cm^{-3} and 2×10^{20} cm^{-3} at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.2 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system. 57

Figure 4.4: I_{DS}-V_{GS} simulation results for (a) L_G = 18 nm and (b) L_G = 9 nm Si ⟨100⟩ (solid circles), Si ⟨110⟩ (open circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), Γ-$In_{0.53}Ga_{0.47}As$ (asterisks), Ge ⟨100⟩ (solid squares), and Ge ⟨110⟩ (open squares) FinFETs. V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method. The S/D doping concentrations were 2×10^{20} cm^{-3} for Si or Ge FinFETs and 5×10^{19} cm^{-3} for InGaAs FinFETs. 59

Figure 4.5: Dependence of (a) (centered moving average of) the peak of the transconductance g_m, (b) subthreshold swing S, (c) turn-on transition voltage ΔV_T, (d) on-current for the constant current defined threshold, I_{on}(CC), and (e) drain-induced barrier lowering, DIBL, with gate length scaling according to $L_G = 3\times W_{FIN}$. 60

Figure 4.6: Projection of constant energy surfaces onto the (a) {110} plane of Si, (b) {100} plane of Si, (c) {100} plane of Ge, (d) {110} plane of Ge, and (e) {100} plane of InGaAs. The shaded regions correspond to the sub-band of lower energy and concentric circles or ellipses are shown as dashed lines to indicate two-fold degeneracy. The direction of confinement lies into the plane of the page. For Ge, the L-valleys are located at the zone boundary, and thus, one-half of each L-valley is inside the first Brillouin zone. 66

Figure 4.7: I_D-V_{DS} simulation results for Si ⟨100⟩ (solid circles), Si ⟨110⟩ (open circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), Γ-$In_{0.53}Ga_{0.47}As$ (asterisks), Ge ⟨100⟩ (solid squares), and Ge ⟨110⟩ (open squares) FinFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage for gate lengths (fin widths) of (a) L_G = 18 nm (W_{FIN} = 6 nm) and (b) L_G = 9 nm (W_{FIN} = 3 nm). Source and drain doping concentrations were taken to be 2×10^{20} cm^{-3} and 5×10^{19} cm^{-3} for Si or Ge and InGaAs devices, respectively. 69

Figure 4.8: Comparison of contacts with perfect transmissivity and imperfect transmissivity contacts on the (a) peak of the transconductance g_m, (b) turn-on transition voltage ΔV_T, and (c) the on-current I_{on} with a constant current defined threshold (I_{on}(CC)), for the end, saddle/slot, and RSD contacts to an 18 nm and 9 nm gate length FinFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.2 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system, including channel orientation for Si and Ge. 72

Figure 4.9: As for Fig. 4.7 but with non-ideal contacts, I_D-V_{DS} simulation results for Si ⟨100⟩ (solid circles), Si ⟨110⟩ (open circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), Γ-$In_{0.53}Ga_{0.47}As$ (asterisks), Ge ⟨100⟩ (solid squares), and Ge ⟨110⟩ (open squares) FinFETs the gate overdrive voltage of 0.35 V above the constant current threshold voltage for gate lengths (fin widths) of (a) L_G = 18 nm (W_{FIN} = 6 nm) and (b) L_G = 9 nm (W_{FIN} = 3 nm), respectively. Source and drain doping concentrations were taken to be 2×10^{20} cm^{-3} and 5×10^{19} cm^{-3} for Si or Ge and InGaAs devices, respectively. 74

Figure 5.1: A side view (left) and an end view (right) of the simulated modeled MOSFET geometry. The spacers regions are not shown in order to show the underlying semiconductor fin, shaded in grey. The hatched region represents the gate metal. The gate oxide located underneath the gate metal is visible in the end views the saddle/slot contact model device. The source and drain contact surfaces are shown in black. 81

Figure 5.2: Alignment of the on-axis and off-axis K and Q conduction band valleys in the hexagonal Brillouin zone of monolayer MoS_2.................................. 82

Figure 5.3: Sketch of (a) symmetrical biaxial tensile strain ($\epsilon_x = \epsilon_y$), (b) uniaxial tensile strain only along the x-direction ($\epsilon_y = 0$), (c) uniaxial tensile strain only along the *y*-direction ($\epsilon_x = 0$), (d) uniaxial tensile strain along the *x*-direction, with associated compression along the *y*-direction ($\epsilon_y = -\nu\epsilon_x$), (e) uniaxial tensile strain along the y-direction, with associated compression along the *x*-direction ($\epsilon_x = -\nu\epsilon_y$), and (f) pure shear strain (γ), which is equivalent to (d), but with a Poisson ratio of $\nu = 0.57$ and no change in the lattice constant. 84

Figure 5.4: Dependence of (a) band gap E_g and (b) band edge valley separation of the of the K- and Q-valleys $\Delta E_{K\text{-}Q}$ under 0% to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (asterisks), uniaxial tensile strain along the *x*-direction ϵ_x (solid circles), uniaxial tensile strain along the *y*-direction ϵ_y (open circles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), and pure shear strain γ (open triangles). 89

Figure 5.5: Dependence of (a) K-valley effective mass along the x-direction, (b) K-valley effective mass along the y-direction, (c) Q-valley effective mass along the *x*-direction, and (d) Q-valley effective mass along the *y*-direction under 0% to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (asterisks), uniaxial tensile strain along the x-direction ϵ_x (solid circles), uniaxial tensile strain along the *y*-direction ϵ_y (open circles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), and pure shear strain γ (open triangles). 90

Figure 5.6: Drift velocity v_d vs. electric field F simulation results for monolayer MoS_2 at 300 K considering only phonon-limited electron transport subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). 94

Figure 5.7: Dependence of (centered moving average of) low-field phonon-limited electron mobility with strain profile at strain amounts of 1% and 3%, including unstrained MoS_2 and K-valley-only MoS_2. 95

Figure 5.8: I_{DS}-V_{GS} simulation results for L_G = 200 nm monolayer MoS_2 MOSFETs subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method. 96

Figure 5.9: I_{DS}-V_{GS} simulation results for L_G = 15 nm monolayer MoS_2 MOSFETs subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method. 97

Figure 5.10: Comparison of strain on the (centered moving average of) the peak of the transconductance g_m for 200 nm and 15 nm gate length monolayer MoS_2 MOSFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to 1% and 3% strain, respectively, are shown side by side on the same gray scale for each considered strain profile, including unstrained MoS_2 and K-valley-only MoS_2. 98

Figure 5.11: I_{DS}-V_{DS} simulation results for L_G = 200 nm monolayer MoS_2 MOSFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). 101

Figure 5.12: I_{DS}-V_{DS} simulation results for L_G = 15 nm monolayer MoS_2 MOSFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). 102

Figure 5.13: Comparison of 3% biaxial tensile strain $\epsilon_x = \epsilon_y$, uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$, and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ with perfect transmissivity and imperfect transmissivity contacts on the peak of the transconductance g_m for 200 nm and 15 nm gate length monolayer MoS_2 MOSFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.23 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system, including unstrained MoS_2 $\epsilon = 0$ and K-valley-only MoS_2........ 104

Figure 5.14: I_D-V_{DS} simulation results with perfect transmissivity and imperfect transmissivity contacts for (a) L_G = 200 nm and (b) L_G = 15 nm monolayer MoS_2 MOSFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage subject to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (dashed line) and no strain MoS_2 $\epsilon = 0$ (solid line).............. 105

Chapter 1: Introduction and Background

1.1 Marching towards the end-of-the-roadmap

According to the International Technology Roadmap for Semiconductors (ITRS), the next-generation of electronic devices need to be smaller, faster, and lower power [1]. Silicon (Si) has been the most widely used material for complementary-metal-oxide-semiconductor (CMOS) technology due to its abundance, low-cost, and favorable material properties, including a wide band gap and good thermal conductivity. However, Si-based CMOS technology is rapidly approaching limitations based on fundamental physics [2]–[4]. Performance degradation with continued scaling such as gate leakage current due to dielectric scaling, short-channel effects, relatively greater variation of threshold voltage over the die, and increasing lithography challenges and cost pose challenges to meeting roadmap specifications [4]–[6]. As a result, a consideration to novel materials, innovative device designs, or a combination of both are needed to extend the life of CMOS technology.

1.2 Advancing CMOS beyond the Si roadmap

For faster circuitry and possibly allow for a lower operating voltage, a high drain-to-source current (I_{DS}) in the ON state is desired. For quasi-ballistic transport, I_{DS} is determined by the product of the total cross-sectional charge density, $-qn_{\mathrm{b}}$ for n-channel FETs (where q is the magnitude of the fundamental unit of charge), the average source-to-drain injection velocity, v_{inj}, and the injection efficiency, γ, at the top of the source-to-channel potential barrier (which is unity in the ballistic limit) [7],

$$I_{\mathrm{DS}} = qn_{\mathrm{b}}v_{\mathrm{inj}}\gamma. \tag{1.1}$$

One way to improve carrier injection velocities, which depend on carrier mass as well as carrier energy, is by modifying the material properties, such as through the application of strain or changing the channel material entirely. On the one hand, the performance of strained Si may be reaching a plateau [8]. On the other hand, high mobility (μ) materials—indicative of some combination of light masses corresponding to high v_{inj} and/or long scattering lifetimes corresponding to high γ, if both in sublinear fashion—such as III-V and germanium (Ge) are widely regarded as potential candidates to fill the performance gap left by Si. Binary III_xV_{1-x} compound semiconductors are obtained by combining group III elements (Al, Ga, In) with group V elements (N, P, As, Sb), while further combinations are possible yielding ternary ($III_xIII_{1-x}V_y$) and quaternary ($III_xIII_{1-x}V_yV_{1-y}$) III-Vs. III-V materials boast excellent electron mobility (μ_e), for example, electrons are nearly 30 times more mobile in indium arsenide (InAs) than in Si. However, low band-gap III-V materials such as InAs also display significant band-to-band tunneling, which leads to large off-state leakage currents. An indium-gallium-arsenide alloy of $In_{0.53}Ga_{0.47}As$ (henceforth referred to as simply InGaAs) has been preferred because it is lattice-matched to InP substrates for fabrication-friendly thin film growth, has a higher mobility than GaAs, and a larger, more tunneling resistant band gap than InAs. Although excellent InGaAs n-channel FETs (nFETs) have been demonstrated, comparably performing p-channel FETs (pFETs) remain elusive due to the substantial (significantly greater than in Si) disparity between the electron and hole mobilities and thermal velocities [9]–[11]. Although both the electrons and holes mobilities of Ge are significantly higher than in Si at room temperature, early Ge-based devices suffered from significant engineering challenges precluding Ge's widespread adoption, including the poorer quality and less stable native oxide (GeO_x), higher interface state density (D_{it}) near the conduction band edge, and difficulty in developing low resistance ohmic contacts to

n-type Ge [12]. The advent of new manufacturing technologies have addressed some of these issues, and progress made on Ge pFETs have led to a reconsideration of Ge channels for future advanced devices [13], [14]. Driven by this renaissance for p-channel Ge, there is associated heightened incentive to develop high-performance Ge nFETs. Higher mobility materials, which can enable faster switching times and higher on-currents, typically have smaller band gaps, which increases standby power consumption via band-to-band tunneling leakage. Table 1.1 lists common material and electrical properties of Si, Ge, and InGaAs, including its constituent binary compounds.

	Si	**Ge**	**GaAs**	**InAs**	**$In_{0.53}Ga_{0.47}As$**
Electron mobility at 300 K, μ_e [$cm^2 V^{-1} s^{-1}$]	1,350	3,900	8,500	40,000	>8,000
Hole mobility at 300 K, μ_h [$cm^2 V^{-1} s^{-1}$]	450	1900	400	<500	350
Lattice constant, a_o [Å]	5.431	5.646	5.653	6.058	5.868
Band gap, E_g [eV]	1.12	0.66	1.42	0.35	0.75
Dielectric constant, ε_r	12	16	13	15	14
m^*_{cond}/m_e (m^*_{dos}/m_e)	0.26 (0.32)	0.16 (0.26)	0.067	0.023	0.043
Electron thermal velocity [$\times 10^7$ cm/s]	2.3	2.9	4.5	7.7	5.6
Thermal conductivity [$W cm^{-1} K^{-1}$]	1.5	0.58	0.5	0.27	0.05
Critical electric field [$\times 10^6$ V/cm]	0.25	0.1	0.004	0.002	0.2

Table 1.1: Basic material parameters and electrical properties of Si, Ge, and InGaAs compiled from [15].

Along with integrating high-mobility semiconductors, innovative designs continue to extend the limits of CMOS scaling. Alternative architectures include partially-depleted silicon on insulator (SOI), fully-depleted SOI, dual-gate SOI, and multi-gate FET [16]–[18]. In particular, the FinFET is a three-dimensional (3D) transistor

design for the 24-nm technology node and beyond that wraps the gate around the channel instead of placing it only on the top, resulting in steeper subthreshold slope and a corresponding reduced threshold voltage and higher ON-state transconductance, $g_m = dI_{\mathrm{DS}}/dV_{\mathrm{G}}$, where V_{G} is the gate voltage [16], [19]–[21]. Traditionally, CMOS circuits have been fabricated on Si $\{100\}$ substrates, substantially because this orientation resulted in low gate Si-SiO_2 interface trap densities for planar devices [22], [23]. The standard channel orientation, now as before multi-gate devices were developed, is then on that plane along a $\langle 110 \rangle$ direction. In this work we again assume $\{100\}$ substrates, but consider two channel orientations: a still standard $\langle 110 \rangle$ channel direction [24], which, as compared to it planar device predecessors, produces a non-traditional gate oxide-channel interface orientation of $\{110\}$; and a non-standard $\langle 100 \rangle$ channel direction, which, again as compared to its planar device predecessors, produces a traditional gate oxide-channel interface orientation of $\{100\}$. Production FinFETs are oriented with a $\langle 110 \rangle$ channel direction, which optimizes the hole channel mobility for both Si and Ge [25], [26].

1.3 A FORK IN THE ROAD

Two-dimensional (2-D) atomic crystals, such as the prototypical graphene, have attracted a lot of attention due to their superb electrical and mechanical properties [27]–[29]. In particular, single-layered transition metal dichalcogenides (TMDs) of the form MX_2, where M is a transition metal (M = Mo, W, Nb, Ta, Ti, Re) and X is a chalcogen (X = S, Se, or Te), have been gaining popularity. Within this family of materials, monolayer molybdenum disulfide, MoS_2, exhibits extraordinary mechanical, thermal, and electronic properties, which enables it to be used in a myriad of applications such as field-effect transistors, integrated circuits, non-volatile memory cells, solid lubricants, photodetectors, and gas sensors [30]–[35]. MoS_2 possess many desirable material

properties well-suited for applications in transistors, including low leakage current because of a substantial band gap—in contrast to gapless graphene—excellent electrostatic control due to its atomic scale thickness (the ultimate ultra-thin body), absence of dangling bonds at the surface to reduce interface traps and defects, and high mechanical flexibility. Moreover, MoS_2 integration is compatible with state-of-the-art nanofabrication processes [36] and wafer-scale device fabrication [37]. On the other hand, the challenges of fabricating MoS_2 devices include large contact resistance, interaction with surrounding environment not well-studied (environmental stability), low intrinsic charge carrier mobility, and difficult to dope. The earliest work on a single-layer MoS_2 transistor on silicon (Si) substrate was performed by B. Radisavljevic et al. in 2011, who reported an electron channel mobility and current on/off ratio of 200 $cm^2 V^{-1} s^{-1}$ and 1×10^8, respectively, at room temperature [38].

Looking beyond rigid Si-based technologies, single-layer MoS_2 devices can also be integrated with soft, flexible, and curvilinear surfaces to unlock new opportunities in the field of flexible electronics such as flexible displays, wearable electronics up to "electronic skin" and tattoos, and biosensors [39]–[41]. Flexible electronics is a disruptive technology offering devices with ultra-thin form factors and high-performance at low-cost that will be able to perform functions that conventional electronic devices cannot, including bending, rolling, folding, and stretching. Recent studies on flexible multi-layer and monolayer MoS_2 transistors on a plastic substrate of polyimide with integrated high-*k* dielectric in a back-gated device structure have shown good electrical and mechanical properties, including an on/off ratio of greater than 10^7, subthreshold slope of 82 mV per decade, and device low-field carrier mobility of 30 $cm^2 V^{-1} s^{-1}$ [42], [43]. Figure 1.1 shows a comparison of the electron mobility and band gap for several candidate semiconductors. Strain in MoS_2, which may be introduced during fabrication

due to lattice mismatch [44] or through mechanical deformation [45], has been shown to alter MoS_2's electronic and transport properties [46], [47]. Therefore, a careful study on the effects of strain on MoS_2 device performance is warranted to advance flexible electronics.

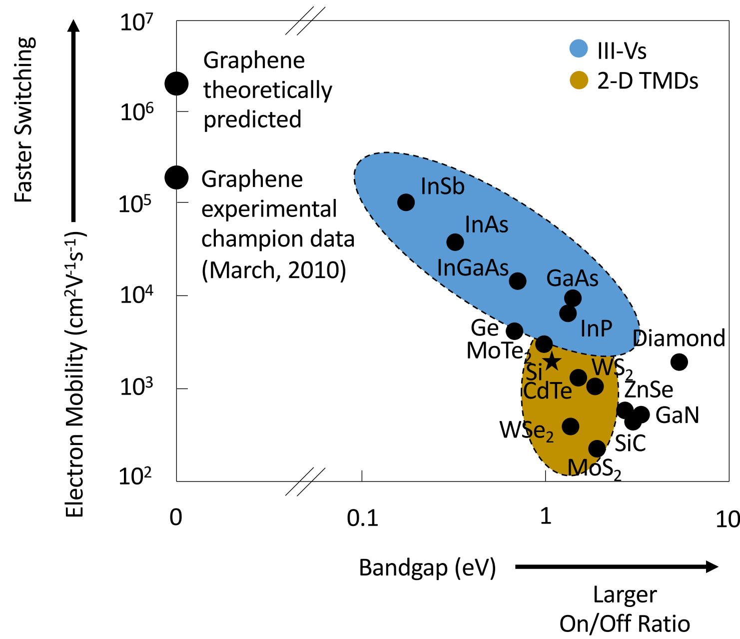

Figure 1.1: Intrinsic electron mobility versus band gap for various semiconductors, including typical III-V and transition metal dichalcogenides materials [48], [49].

1.4 SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Simulation is becoming an indispensable tool for device engineers, and together with experiments can be used to understand physical phenomena that are either difficult or impossible to measure, test hypothetical devices concepts and explore complex design spaces, and provide insights and predictions into device behavior. As feature sizes shrink into the nanometer scale regime, physical models have to be refined and extended to more accurately capture transport phenomena occurring on these scales. Degenerate

carrier concentrations that exceed the effective density of states found in modern MOSFETs, such as that in the source and drain (S/D) to reduce series resistance, invalidate classical statistics (Boltzmann statistics in the equilibrium limit). As supply voltages have not scaled accordingly, the resulting large electric fields inside devices (which rapidly change over small length scales) gives rise to hot-carrier and non-local effects. The latter include far-from-equilibrium carrier statistics, making the use of even Fermi-Dirac statistics, particularly in the channel, invalid. And an electric field in the direction perpendicular to the semiconductor channel and dielectric interface can create a narrow potential well, and the resulting quantum mechanical confinement of the free electron gas leads to quantized energy levels, valley degeneracy breaking (even without strain), and modification of the density of states.

Non-equilibrium Green's Function based (NEGF-based) quantum transport models and/or sub-band-based transport models, while offering several advantages for modeling nanoscale devices, also generally employ simplified, end-to-end source/drain carrier injection topologies. At the other end of the scale, while more realistic contact geometries can be included readily, contact orientation effects are substantially obscured in drift-diffusion or hydrodynamic simulations because even hot carriers move purely diffusively in proportion to the Fermi-level gradient as they enter the device. However, ensemble semi-classical Monte Carlo (SCMC) simulators allow for both complex contact geometries and fully- and quasi-ballistic through diffusive transport, providing an opportunity for modeling contact geometry effects in modern nanoscale devices not otherwise available. Additionally, the SCMC approach allows for, among other things, a description of carries under quantum-confinement, far-from-equilibrium transport, and the ability to include a variety of scattering mechanisms such as phonons, surface roughness, and ionized impurities.

1.5 Dissertation overview

The focus of this work is to understand and model the essential underlying physics in the operation of Si, Ge, InGaAs, and MoS_2 n-channel field-effect transistors (FETs) to identify potential performance bottlenecks and provide guidance to device designers. The organization of this dissertation is as follows. Chapter 1 reviews the challenges of scaling conventional Si-based CMOS technology and the proposed use of high mobility and thermal velocity channel materials for high-performance logic transistors and strained MoS_2 channel materials for low standby and operating power flexible electronic devices. Chapter 2 presents an advanced quantum-corrected SCMC tool for modeling the end-of-the-roadmap Si, Ge, InGaAs, and MoS_2 n-channel FETs, outlining the essential elements of our simulation methodology, including the main building blocks of the Monte Carlo algorithm, surface roughness scattering, and contact transmissivity. Chapter 3 addresses the impact of contact geometry and transmissivity on quasi-ballistic nanoscale Si $\langle 110 \rangle$ and $\langle 100 \rangle$ and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs. Chapter 4 addresses gate length scaling, and associated fin width scaling, impact on quasi-ballistic nanoscale Si $\langle 110 \rangle$ and $\langle 100 \rangle$, Ge $\langle 110 \rangle$ and $\langle 100 \rangle$, and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs. Chapter 5 addresses the impact of tensile strain on the intrinsic performance limits of monolayer MoS_2 n-channel MOSFETs. Chapter 6 concludes with a dissertation recap and recommendations for future work.

Chapter 2: University of Texas Monte Carlo Software

2.1 MONTE CARLO HISTORY

Since the late 1970s, the Monte Carlo method has been used for studying carrier transport in semiconductors and detailed reviews can be found in [50]–[53]. The Monte Carlo method is a numerical technique for solving the Boltzmann transport equation (BTE) by following the motion of carriers in both real space and momentum space, subject to stochastic scattering events determined by sequences of random numbers with specified probability distributions. Without the need for any additional physical approximations, the Monte Carlo method allows for the incorporation of carrier transport effects in a rather complete and comprehensive manner

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_{\mathbf{r}} f + \frac{q\boldsymbol{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{col}}. \quad (2.1)$$

Where q is the fundamental charge, $\hbar$ is the reduced Plank's constant, $\boldsymbol{r}$ is the carrier position in real space, $\boldsymbol{k}$ is the carrier wave vector in momentum space, $\boldsymbol{v}$ is the group velocity, $\boldsymbol{F}$ is the electric field at position $\boldsymbol{r}$, t is the time, and the distribution function $f(\boldsymbol{r},\boldsymbol{k},t)$ represents the probability for a carrier to occupy position $\boldsymbol{r}$ with momentum $\boldsymbol{k}$ at time t. The collision term depends on the microscopic scattering mechanisms present in the material system

$$\left(\frac{\partial f}{\partial t}\right)_{\mathrm{col}} = \sum_{\boldsymbol{k}'}\{S(\boldsymbol{k}',\boldsymbol{k})f(\boldsymbol{r},\boldsymbol{k}',t)[1-f(\boldsymbol{r},\boldsymbol{k},t)] - S(\boldsymbol{k},\boldsymbol{k}')f(\boldsymbol{r},\boldsymbol{k},t)[1-f(\boldsymbol{r},\boldsymbol{k}',t)]\}. \quad (2.2)$$

Where $S(\mathbf{k}',\mathbf{k})$ is the transition probability between states $\boldsymbol{k}$ and $\boldsymbol{k}'$ and $[1-f(\mathbf{r},\mathbf{k}',t)]$ term is the probability that the state $\boldsymbol{k}'$ is not occupied. To calculate these scattering rates, the Fermi Golden Rule is used

$$S(\boldsymbol{k},\boldsymbol{k}') = \frac{2\pi}{\hbar}|M(\boldsymbol{k},\boldsymbol{k}')|^2\delta\left(E_{\boldsymbol{k}} - E_{\boldsymbol{k}'} \pm \hbar\omega_{\boldsymbol{q}}\right). \tag{2.3}$$

Where $\hbar$ is the reduced Plank's constant, M is the matrix element, $E_{\boldsymbol{k}}$ and $E_{\boldsymbol{k}'}$ are the energy of the states before and after scattering, respectively, and $\hbar\omega$ is the energy of the absorbed or emitted phonon with wave vector $\boldsymbol{q}$.

2.2 Simulation Methodology

We employed our in-house quantum-corrected three-dimensional (3-D) SCMC methodology, University of Texas Monte Carlo (UTMC) [54], to study contact geometry and crystal orientation effects on carrier injection in Si, Ge, and InGaAs n-channel FinFETs and strain effects on carrier transport in MoS_2 n-channel MOSFETs, while also modeling far-from-equilibrium degenerate statistics, non-ideal contact resistivities, and quantum-confinement effects on carrier distributions in real-space and among energy valleys, and on phonon, impurity, and surface roughness scattering. The strength of our method is that no a priori assumption of an equilibrium or any specific carrier distribution is made and, no adjustable parameters are needed, unlike the effective potential approximation, to calculate the quantum-correction potentials [55]. Here we summarize some of the basic features of our simulator that impact our simulation results.

UTMC models carrier transport within 3-D device geometries considering intra- and inter-valley phonon (acoustic, optical, and polar optical), surface roughness, alloy, and (Brooks-Herring [56]) ionized impurity scattering. Intra-valley acoustic phonon scattering, alloy scattering, ionized impurity scattering, and surface roughness scattering are treated as elastic scattering processes; intra-valley optical phonon scattering, and inter-valley acoustic and optical phonon scattering are treated as inelastic scattering processes. Following the approach of Jacoboni and Fischetti, the electron energy bands

are modeled analytically with non-parabolicity corrections [50], [52], which is reasonable for the limited carrier energies encountered here. Within this approximation, the relationship between the carrier energy E and the wave vectors k_i (d =1, 2 or 3, for the dimensionality of the system) in the reference frame of the principal axes of the valley is

$$E(\boldsymbol{k})\big(1+\alpha E(\boldsymbol{k})\big)=\sum_{i=1}^{\mathrm{d}}\frac{\hbar^2 k_i^2}{2m_i}, \tag{2.4}$$

Where $\hbar$ is the reduced Plank's constant, α is the non-parabolicity correction, m_i is the component of the mass tensor along the k_i direction in the principal axes coordinate system.

Except as otherwise noted below, simulation parameters for Si and InGaAs are provided in [54]. Simulations of bulk velocity-field curves during UTMC development [54] produced excellent agreement to experimental data [50]–[52], [57], [58]. Accurate modeling of sidewall surface roughness scattering in FinFETs is more challenging, with scattering being likely dependent on channel and dielectric material and any strain thereof, materials growth and etching methods, and even detailed device geometry [59]. In this work, surface roughness parameters for ⟨100⟩ Si simply were adjusted to reproduce available channel mobility data for planar MOSFETs with high-quality Si-SiO_2 interfaces [60], as in [54]. These same surface roughness parameters then are used for Si ⟨100⟩ and Si ⟨110⟩ channel FinFETs, as well as for InGaAs FinFETs, which also leads to much the same channel mobility for simulated ⟨100⟩ and ⟨110⟩ planar Si MOSFETs. While the latter result is not consistent with mobility measurements in planar Si devices [26], it is consistent with mobility measurements in (100) and (110) sidewall Si FinFETs [59], [61]. We also have observed a relatively modest effect of surface roughness scattering in the simulated drive current of these deeply scaled FinFETs, consistent with [62] and an overall reduction in relative effect of changes in scattering on drive current as compared to mobility as the ballistic limit is approached. Others,

however, have observed greater and important effects of surface roughness scattering in simulation of deeply scaled gate-all-around FETs [63]. Thus, this modeling of sidewall surface roughness of FinFETs introduces additional uncertainty in simulated absolute and relative performance of the considered technologies, and our approach to modeling sidewall surface roughness of FinFETs may be optimistic for all devices of this work to varying degrees. However, it also provides a control and perhaps somewhat compensates for immature InGaAs MOSFET gate dielectric technology [64].

Source and drain (S/D) doping densities, such as simulated in this work, are approaching solid-solubility limits that far exceed the effective density of states of the conduction band. Because of high doping concentrations, degenerate statistics must be addressed. However, because of the far-from-equilibrium conditions encountered in nanoscale FinFETs, carrier statistics cannot be described accurately using Fermi-Dirac distributions. Instead, UTMC directly models Pauli-Blocking (PB) of scattering to obtain the far-from-equilibrium local electron occupation probabilities $f(\boldsymbol{r}, g, E, \pm)$ from the local electron populations, $N(r, g, E, \pm)$, as a function of position (*r*), energy valley (*g*) and energy (*E*), and propagation direction, forward toward the drain end (+) or backward toward the source end (−)

$$f(\boldsymbol{r}, g, E, \pm) = N(\boldsymbol{r}, g, E, \pm)/D(g, E)/2, \tag{2.5}$$

Where $D(g, E)/2$ is the position independent density of states per energy valley reduced by a factor of two for forward-going and backward-going carrier contributions.

Energy valley and position dependent quantum-corrected potentials (QCPs) are calculated to match the calculated quantum-corrected (as an approximation, for computational efficiency, for this purpose only) equilibrium semi-classical carrier distributions to the quantum mechanical distributions. The latter distributions are obtained via self-consistent coupling of Schrödinger's time-independent equation with

the Poisson's equation, while allowing for barrier penetration effects, which can moderate the effects of confinement significantly. For practicality, the QCPs are computed within two-dimensional cross sections normal to the channel direction within an effective mass approximation with a non-parabolicity correction. To approximate three-dimensional effects, the quantum corrections are ramped on starting at the onset of confinement at the source and drain extension boundaries, over a distance equal to the actual channel width. The quantum corrections then serve to increase thresholds and alter relative valley occupancy, redistribute the carriers in real-space away from potential barriers, generally increase even intra-valley phonon scattering rates, particularly for randomizing processes, and determine the surface roughness scattering rate. In this latter way, although the employed surface roughness parameters for all FinFETs here are taken as the same, the resulting surface roughness scattering rates are not.

2.3 SURFACE ROUGHNESS

As the oxide thickness decreases with each technology generation, the effective vertical electric field increases which degrades the effective surface mobility due to increased surface roughness scattering. Surface roughness of an interface, characterized by the statistical parameters of root mean square height Δ_{rms} and correlation length L_c, typically causes fluctuations in the width of the quantum well, which leads to fluctuations in the electron energy levels, and that adds to the scattering [65]. The effect of surface roughness on bound electrons was studied by Prange and Nee [65], and on carrier transport in silicon inversion layers in MOSFETs later by Fischetti *et al.* [66] in more detail. Two common models to describe the potential fluctuations at the surface are either a Gaussian or exponential distribution. The rate of surface roughness scattering is distribution independent if the product of the correlation length and carrier momentum is

much smaller than unity, an approximation that is valid to a correlation length of a couple of nanometers [67], [68], and, in this limit, the scattering process becomes randomizing. This assumption is justified based on the parameters found in the work of Goodnick *et al.* of a correlation length of 1.3 nm and a roughness height of 0.4 nm for a (100) Si-SiO_2 interface [69]. We lump all the parameters and constants, including Δ_{rms} and L_c that characterize the Si-SiO_2 interface, into a single adjustable parameter, C, determined by fitting our model to available experimental data, and the surface roughness scattering rates becomes

$$\frac{1}{\tau_{SR}} = C m_c V_{QC}^3 D(E). \quad (2.6)$$

Where m_c is the confinement mass, V_{QC}^3 is the quantum-corrected potential, and $D(E)$ is the 3-D density of states. The confinement mass is calculated by rotating the effective mass tensor in the direction of the vector containing the rate of change of the electric field components in the x-, y-, and z-directions, respectively, computed using finite differences

A mixed approach of a device simulation followed by a bulk simulation is employed to calibrate the quantum-confinement-dependent surface roughness scattering model. The quantum-corrected potentials for each valley at the middle of the channel are obtained from a self-consistent planar Si $\langle 100 \rangle$ MOSFET simulation without surface roughness scattering and the drain-to-source voltage set to 0 V as a function the average transverse electric field, i.e. the electric field normal to the interface. Next, these quantum potentials are fixed into a bulk simulation, modifying the bulk phonon scattering rates, to calculate the velocity versus field curve with the surface roughness scattering now included, and in effect, we a simulating an extremely long channel length device. Then the low-field mobility is extracted and plotted as a function of transverse electric field and the surface roughness scattering C coefficient is adjusted to match with measured

data [60], [70]. As shown in Figure 2.1, with a single adjustable parameter, reasonable agreement with experimental curves is achieved. For "bulk" channel mobility of Si (100) and Si (110), the same surface roughness leads to roughly the same surface roughness scattering.

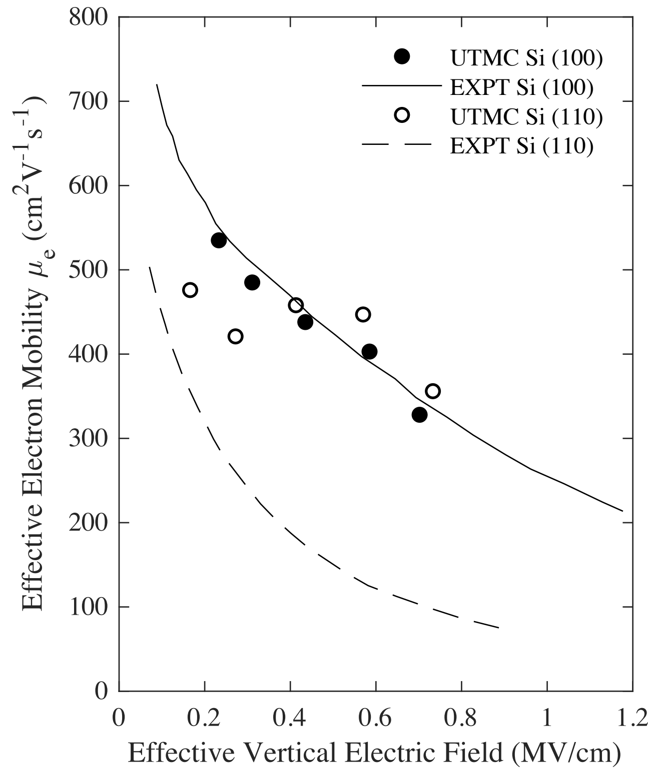

Figure 2.1: Comparison between UTMC electron mobility as a function of the effective electric field obtained by bulk simulations considering surface roughness scattering as well as quantum-confined phonon scattering and the experimental universal mobility curves for bulk Si-SiO_2 interface channel MOSFETs [60], [70].

The importance of surface roughness in deeply scaled FinFETs is still a matter of debate. Actual surface roughness may be different for FinFETs versus planar MOSFETs due to the quality of the sidewall surfaces [71] as well as different gate stack materials or material combinations [59]. Additionally, in a combination of first-principles and

experimental work [61], [62], C. D. Young et al. found that electron mobility is not significantly degraded between (100) vs. (110) fin sidewall orientations, despite substantial differences in planar devices [26]. We found surface roughness scattering to be fairly modest in these very small devices in our simulations.

2.4 On contacts

Parasitic source and drain (S/D) series resistance R_{series} can be divided into the four components: (1) extension-to-gate overlap resistance (R_{OV}), (2) S/D extension resistance (R_{EXT}), (3) deep S/D resistance ($R_{\text{S/D}}$), and (4) contact resistance between the semi-metallic silicide and the heavily doped semiconducting S/D interface (R_{C}). However, decreases in channel resistance increase the importance of series resistance. R_{series} plays an increasingly limiting role in the performance of MOSFETs near the end of the International Technology Roadmap for Semiconductors (ITRS) [1]. For nodes since 2008, R_{series} has been approximately 25% of R_{on} for Si technologies. Moreover, continually decreasing device sizes have increased the contribution to R_{series} of the contact resistance R_{C} between the semi-metallic silicide and the heavily doped semiconducting S/D interface, already about 40% of R_{series} at 50 nm gate lengths in planar Si MOSFETs [72].

The contacts are modeled as in equilibrium. At the beginning of each timestep, carriers are injected into the simulation region. The valley the carrier is injected into is determined by the ratio of the transverse density of available states, i.e. the projection of the energy contour onto the transverse *k*-plane, to the total density of available states to inject into. The current is computed by counting the net number of carriers that enter or exit a particular contact. Under overall equilibrium conditions, the net current through the

contacts vanishes on average even while electrons continue to be injected and absorbed. Details of the models and methodology can be found in [54].

To obtain the Landauer-Büttiker limit [73], [74] of specific contact resistivity, ρ_{LB}, electrons are injected from the contacts into the S/D from a surface-normal-velocity-weighted half-space Fermi-Dirac distribution, while electrons reaching the contact surface from the S/D region are perfectly absorbed. To then account for larger realistic specific contact resistivities, one method would be to add a distributed specific contact resistivity by which there is a corresponding localized voltage drop at the contact surface in proportion to the local current density [75], [76]. For this work, however, realistic contact resistivities are obtained by equally scaling down the electron injection and absorption probabilities—the electron transmissivity T—across the contact surface. Specular reflection then is used to model carriers reaching, but not being transmitted across the contact interface from the inside. Both energy and momentum parallel to the interface is conserved by reflecting into the mirror-image energy valley across the Brillouin zone, producing equal angle reflection in both $\boldsymbol{r}$ and $\boldsymbol{k}$. One way to obtain the specific contact resistivity ρ_{sp} is as one-half of $\rho_{\mathrm{net,bal}}$, the extrapolation of net resistivity ρ_{net} for current flow between two identical contacts to an inter-contact distance of zero to eliminate the contributions of scattering between the contacts to the resistivity, which can thus be related to T as follows. Consider a ballistic conduction channel (although neglecting coherence) characterized by Landauer-Büttiker resistivity, ρ_{LB} (reciprocal of conductivity σ_{LB}) between imperfect contacts modeled by transmission (T) and reflection (R) probabilities, $T_1 = 1 - R_1$, and, $T_2 = 1 - R_2$. Considering the (power) series of all possible transmission trajectories with or without internal reflections between the contacts, the net inter-contact conductivity is,

$$\sigma_{\mathrm{net,bal}} = \sigma_{\mathrm{LB}} T_1 T_2 \sum_{n=0}^{\infty} (R_1 R_2)^n = \sigma_{\mathrm{LB}} T_1 T_2 / (1 - R_1 R_2). \tag{2.7}$$

For $T_1 = T_2$ as in this work, this result reduces to,

$$\sigma_{\text{net,bal}} = \sigma_{\text{LB}} T/(2 - T). \tag{2.8}$$

The corresponding net resistivity is then,

$$\rho_{\text{net,bal}} = \rho_{\text{LB}}(2 - T)/T. \tag{2.9}$$

The resulting apparent specific contact resistivity corresponding to the transmission modeled probability T is, thus, [73], [74],

$$\rho_{\text{sp}} = \rho_{\text{LB}}(T^{-1} - 1/2). \tag{2.10}$$

Note that, intrinsically, $T \leq 1$ and $\rho_{\text{sp}}/\rho_{\text{LB}} \geq 1/2$. In this way, half of the Landaur-Büttiker resistivity, which is fundamentally a non-local quantity, nevertheless is associated with each contact by this measure of specific contact resistivity. In this way we both preserve contact geometry and surface orientation effects and avoid fully localizing the voltage drop to the contact surface in significantly ballistic devices.

2.5 Time evolution

Before device simulation, users of UTMC define the device geometry to be simulated, the physical models to be used, and the bias conditions for which electrical characteristics are to be simulated. The device structure is defined according to a rectangular coordinate system with its x-axis along the transport direction, the y-axis along the width, and z-axis along the height. The simulation ends when the total time allotted for the simulation ends, typically, on the order of tens of picoseconds. The basic building block of our SCMC algorithm is summarized with the flowchart in Fig. 2.2.

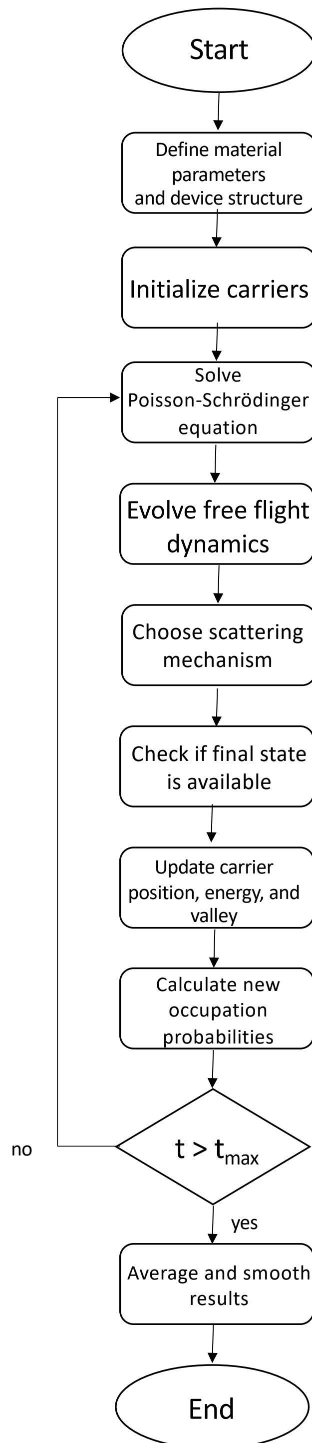

Figure 2.2: Flowchart of a self-consistent Monte Carlo device simulation.

UTMC simulates the motion of several thousand carriers, including subcarriers, through the semiconductor. This number is limited by memory constraints, but good statistics can be obtained if the simulation time is long enough. Subcarriers are used to eliminate classical artifacts of carrier-carrier scattering not subject to Pauli-Blocking to stop hot carriers, eliminate fictious self-images forces, and provide better statistics overall. The carriers are initialized with a Fermi-Dirac distribution, although that is not necessary as the correct distribution will eventually emerge, and randomly oriented momenta. During a single simulation timestep or Monte Carlo iteration, carriers undergo free flight motion and then scatter. Free flight times τ are generated according the probability distribution

$$\tau = -\frac{1}{\Gamma_o}\ln(r). \tag{2.11}$$

Where the constant Γ_o is the sum of all scattering rates at the maximum carrier energy with negligible probability of being achieved by the carrier during simulation and r is a uniform random number between 0 and 1. During free flight, the carriers drift under the influence of the electric fields according to Newton's laws of motion,

$$\begin{gathered}\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \frac{1}{\hbar}\nabla_{\boldsymbol{k}}[E(\boldsymbol{k})], \\ \frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t} = -\frac{q\boldsymbol{F}(\boldsymbol{r})}{\hbar}.\end{gathered} \tag{2.12}$$

Where q is the fundamental charge, $\hbar$ is the reduced Plank's constant, $\boldsymbol{r}$ is the carrier position in real space, $\boldsymbol{k}$ is the carrier wave vector in momentum space, $\boldsymbol{F}$ is the electric field at position $\boldsymbol{r}$, and t is the time. The timestep set by the program is divided by the free flight time. If the free flight time is longer than the timestep, then the carrier drifts according to the timestep. If the free flight time is less than the timestep, then the electron will drift and then scatter. In the worst case, an electron with thermal velocity 10^8 cm/s and timestep of 0.24 fs could move a distance of 0.24 nm, which is less than the grid spacing of 1 nm set by the program. The carrier free flight time is further sub-

divided by the time to the nearest grid site to reduce self-image forces, as discussed later. If a scattering event occurs, the carrier's state after scattering is selected based on the comparison of a random number with the scattering probability and taking into account energy conservation and probability that the state is occupied. Then, another random free flight time is generated. This process repeats for all carriers until the end of the timestep.

The Poisson equation must be solved to self-consistently to update the electrostatic potential at each grid site as the carriers move inside the device. The cloud-in-cell method is most often employed for assigning the carrier charge to the grid sites because it gives a better description of charge density but is more susceptible to self-image forces. If we move a carrier to a new position and attempted the evaluate the force on the that carrier using the forces from grid locations at the previous timestep, the carrier will feel a repulsive self-force from itself. Spurious self-image forces are reduced by solving the Poisson equation at each timestep, using the nearest grid site charge assignment to assign carrier charge to the mesh, introducing subcarriers to reduce the charge of each carrier, and decreasing the simulation timestep.

Chapter 3: Semi-Classical Monte Carlo Study of the Impact of Contact Geometry and Transmissivity on Quasi-Ballistic Nanoscale Si and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs

3.1 Introduction and Background

New materials and new device designs continue to emerge as candidates for extending CMOS scaling, including the possible use of high electron mobility and thermal velocity channel materials [6]. In direct gap III-V materials, including $In_{0.53}Ga_{0.47}As$ (InGaAs), conduction band Γ-valley electron mobilities and thermal velocities can be much greater than in silicon (Si). Substantial ballistic transport can occur on scales greater than 100 nm [50], versus on the scale of 10s of nm for Si (based on average velocity magnitude and scattering rate for thermal electrons). MOSFETs have moved to multi-gate geometries such as FinFETs for improved short channel effects [16], [19]–[21].

However, decreases in channel resistance increase the importance of series resistance. Parasitic source and drain (S/D) series resistance R_{series} plays an increasingly limiting role in the performance of MOSFETs near the end of the International Technology Roadmap for Semiconductors (ITRS) [1]. For nodes since 2008, R_{series} has been approximately 25% of R_{on} for Si technologies. Moreover, continually decreasing device sizes have increased the contribution to R_{series} of the contact resistance R_{C} between the semi-metallic silicide and the heavily doped semiconducting S/D interface, already to about 40% of R_{series} at 50 nm gate lengths in planar Si MOSFETs [72].

Common options for making contacts to multi-gate MOSFET/FinFET geometries includes dumbbell-shaped source and drain contacts, saddle or slot contacts, and raised source and drain contacts (RSD) contacts. The dumbbell S/D contact layout is like that of planar MOSFET S/D contacts in that contact holes (vias) are etched using a contact

window mask down to the surface to be contacted. However, dumbbell layouts are not area efficient, and FinFETs are moving toward pad-less fin structures, such as saddle contacts or contacts to epitaxially-thickened S/D regions [77], [78]. Saddle contacts are attractive because of a significantly smaller device footprint than the dumbbell layout, and because the saddle metal contact couples to the fin top and sidewall surfaces through a thin metal silicide interface, potentially giving rise to a larger contact area to reduce contact resistance. If making a simple saddle contact to individual fins is not possible due to tight alignment tolerances, slot contacts, a variant on the theme, can be used instead, where a thicker layer of metal silicide is deposited across the S/D of all the fins, followed by metal contact across the top of the silicide as a whole. However, the extra contact metal between the fins in slots contacts increases the parasitic gate-to-contact capacitance, which can limit circuit performance. An attractive option is to increase the fin width in the S/D semiconductor regions by epitaxial growth, even to the point of merging adjacent fins (although not modeled as such here), in the RSD structure to eliminate the contact-to-fin pitch matching requirements and increase the surface area of the contact. In addition, the RSD structure has been shown to reduce the parasitic S/D resistance and capacitance, but not at the expense of fin pitch [79]–[81]. One drawback of the RSD approach is that the conformation of the source and drain surface depends on the source and drain epitaxial faceting. For (110) sidewalls, the final fin cross section is hexagonal or diamond-shaped, and hence, the contact will land on a non-planar surface. For (100) sidewalls, the cross section of the epitaxially grown semiconductor is rectangular and contacts will land on a flat surface. In any case, a common trait of these saddle/slot and RSD contacts relevant to this work is that each may be considered as a "side" or "wrapped" contact with respect to the channel orientation.

Although not the focus of this work, a contributing device performance and modeling consideration is that direct band gap III-V materials provide high electron mobilities and thermal velocities only so long as carriers remain in the Γ-valley. Within FinFET channels, the inter-valley separations will be reduced under quantum confinement relative to the bulk due to the greatly differing effective masses, while limited quantum (density of states) capacitance will move the Fermi level up rapidly in the Γ-valley with increasing channel carrier concentration. Moreover, in near-ballistic MOSFETs channels in saturation, with primarily drain-directed electrons, that quantum capacitance is effectively halved, relative to equilibrium. This consideration of quantum confinement and quantum capacitance impacts not only potential device performance, but the types of modeling tools which can be used to assess it. Moreover, the inter-valley separations between the light-mass Γ-valley and heavy-mass peripheral L-valleys ($\Delta E_{\Gamma\text{-L}}$) and X-valleys ($\Delta E_{\Gamma\text{-X}}$) are not reliably known [82]. For $In_{0.53}Ga_{0.47}As$, a commonly cited tight-binding calculation places the bulk values of $\Delta E_{\Gamma\text{-L}}$ at 460 meV [83], while the only experimental determination places $\Delta E_{\Gamma\text{-L}}$ at 550 meV [83], [84]. Recent density-functional calculations have even estimated $\Delta E_{\Gamma\text{-L}}$ to be as large as 1.31 eV [85].

The performance of indium-gallium-arsenide (InGaAs) MOSFETs have been explored through simulation for years [86]–[92], [75], [93]. However, with varying device geometries and scales, band structure models, and simulation methods, a consistent picture has not emerged. Multi-sub-band Monte Carlo and quantum transport simulations have been performed, which intrinsically address quantum confinement effects including, but not limited to, effects on inter-valley separation [90]–[94]. These methods generally have predicted that high-mobility channel materials will lead to substantially better MOSFET performance. However, complex contact geometries are difficult to address with these tools, so the focus often is on transport through the

channel, with the S/D contacts often modeled as essentially perfect electron reservoirs via simple perfectly injecting and absorbing end contacts. Semi-classical Monte Carlo (SCMC) simulations also have been performed, allowing consideration of more realistic contact geometries, although with varying band structure models and, if any, quantum confinement models. One theme emerging from these latter simulations for high-mobility channel materials MOSFETs is source starvation associated with contact surface orientations running parallel to the channel [75], [86]–[88], along with a less clear picture of the degree of advantage, if any, of high-mobility channel materials for nanoscale devices. A recent full-band Monte Carlo simulation study of nanoscale FinFETs [75] compared the performance of a side contact geometry and of an end contact geometry of the same area (although of qualitatively different character from the one considered here and for different purposes) for Si and InGaAs FinFETs, although without modeling quantum confinement in the channel. The authors also found source starvation effects associated with side contacts, as well as with non-ideal specific contact resistivities, for all devices but more so for InGaAs devices. However, InGaAs FinFETs continued to perform better than their Si counterparts.

In this work, the effects of contact geometry and specific contact resistivity on $In_{0.53}Ga_{0.47}As$ and silicon (Si) nanoscale (18 nm channel length) n-channel FinFET performance, and the effects of models thereof, are studied, using our in-house quantum-corrected semi-classical Monte Carlo tool, UTMC. Saddle/slot contacts, RSD contacts, and a reference end contact were modeled, each with both perfectly injecting and absorbing contacts, and with contacts of more realistic specific contact resistivities, modeled here via sub-unity electron transmission probabilities (transmissivities) across the contact surface. Far-from-equilibrium degenerate statistics, and quantum-confinement effects on carrier distributions in real-space and among energy valleys and on scattering

rates are addressed. We consider Si ⟨110⟩ and Si ⟨100⟩ channel orientations, and multi-valley InGaAs (MV-InGaAs) and Γ-valley-only InGaAs (Γ-InGaAs) channel devices. The idealized Γ-InGaAs channel represents the possibility of substantially larger valley offsets than otherwise modeled here, or perhaps weaker quantum confinement in channels, as well as simulation limitations such as not modeling quantum confinement within the channel and the associated reduction in inter-valley separation, or fully ballistic simulations, whereby electrons injected into the Γ-valley in the source are unable to scatter to peripheral valleys even when energetically available.

Among our findings, echoing those of [75], Γ-InGaAs FinFETs were highly sensitive to contact geometry and specific contact resistivity, while Si FinFETs showed still significant but much less sensitivity to contact models. For idealized unity transmissivity contacts, Γ-InGaAs channel FinFETs performed best for all contact geometries, at least in terms of transconductance, and end contacts provided the best performance for all considered channel materials. For realistic contact resistivities, however, results are essentially reversed. Silicon channel FinFETs performed best for all contact geometries, and saddle/slot and RSD contacts outperformed end contacts. We also find that results for InGaAs FinFETs are sensitive to the peripheral valley energy offsets and their modification by quantum confinement within the channel. These simulation results challenge the potential of InGaAs FinFETs, but also suggest that the relative insensitivity of Si FinFET performance to contact design, and perhaps other device features, have allowed design choices that must be reconsidered to optimize InGaAs FinFET performance.

3.2 SIMULATED FINFET STRUCTURES AND INGAAS BAND STRUCTURE MODELS

3.2.1 FinFET structure

We model 18 nm gate length (L_G) and 6 nm fin width (W_{FIN}) InGaAs and Si-channel FinFETs, with reference end, saddle/slot [95], and RSD contact geometries, as shown in Fig. 3.1(a)-(c), respectively, with device geometry parameters listed in Table 3.1. The surface area of these source and drain contacts are 228 nm^2, 574 nm^2, and 752 nm^2, respectively.

A $\langle 100 \rangle$ substrate orientation is considered for all devices. For Si FinFETs, we considered both $\langle 110 \rangle$ and $\langle 100 \rangle$ channel orientations with corresponding $\{110\}$ and $\{100\}$ fin sidewall orientations, respectively. With elliptical energy valleys in Si, these different channel orientations produce different degrees of quantum confinement within the channel between channel orientations, and between otherwise-equivalent Δ-valleys for the same channel orientation. However, for the RSD contact geometry, for both Si channel orientations, we use the same rectangular geometry characteristic of $\langle 100 \rangle$ channel orientations. For InGaAs FinFETs, with the Γ-valley being spherical, we consider only the $\langle 100 \rangle$ channel orientation. We assume a 3.0 nm thick HfO_2 layer ($\varepsilon_r = 22.3$) gate oxide for an effective oxide thickness of 0.52 nm for all FinFETs for electrostatic calculations. To address near-surface barrier penetration of the wave-function for the InGaAs FinFETs, we model the oxide effective mass as that of HfO_2, $0.15 m_e$ [96]. However, for Si channel FinFETs, because there is a commonly-occurring thin SiO_2 gate-oxide interfacial layer even with high-k gate dielectrics, we model the oxide effective mass as that of SiO_2, $0.55 m_e$ [97]. The fin height (H_{FIN}) and oxide substrate thickness (H_{BOX}) of all FinFETs are 35 nm and 10 nm, respectively. The source and drain regions, located 5 nm away from the edge of the gate region, are uniformly

doped to 2×10^{20} cm^{-3} for silicon, and 5×10^{19} cm^{-3} for InGaAs, the maximum experimentally observed electrically active dopant concentrations of arsenic in silicon, and of silicon in $In_{0.53}Ga_{0.47}As$, respectively [98]–[100]. The FinFETs have a decade/nm doping profile in the 5 nm source and drain extensions (L_{EXT}).

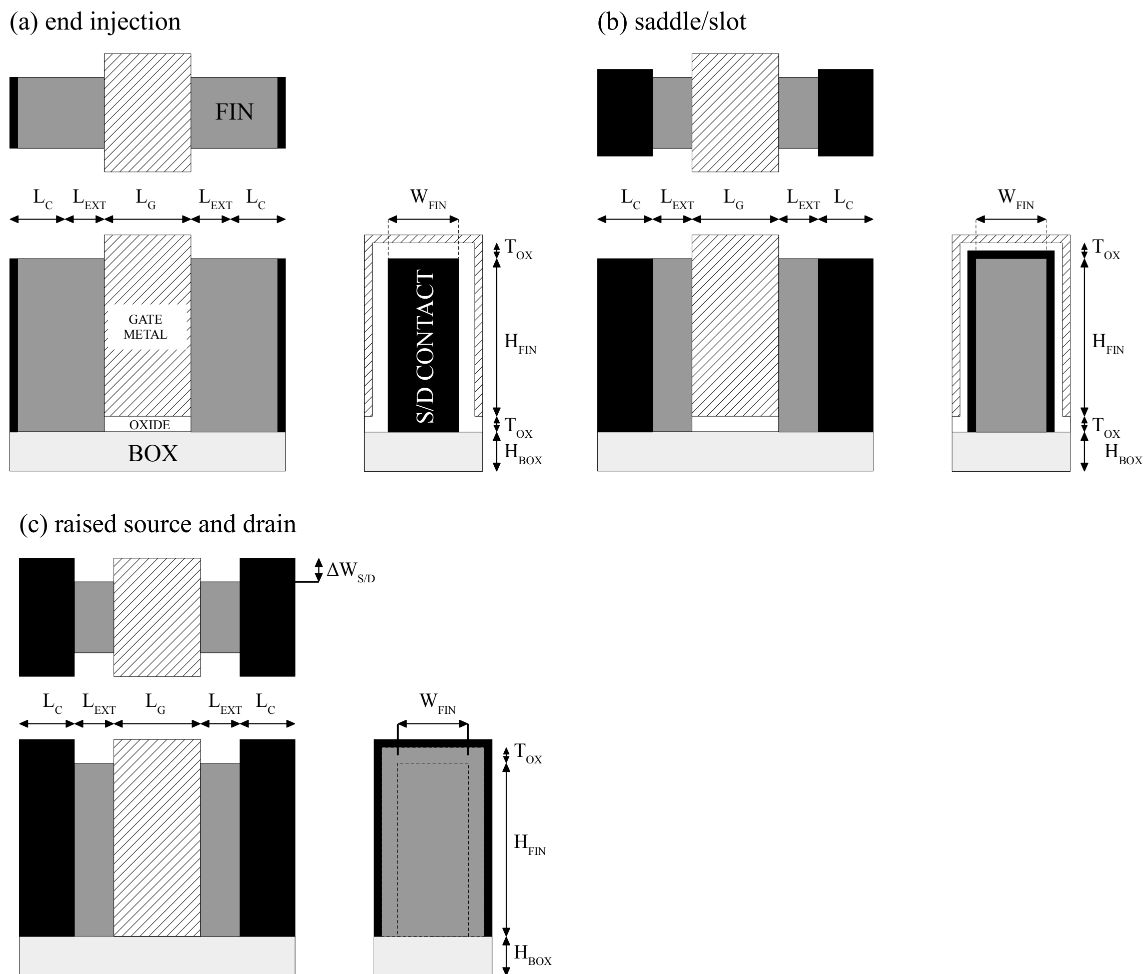

Figure 3.1: Schematics of the simulated FinFET geometries with (a) reference end contacts, (b) saddle/slot contacts, and (c) raised source and drain (RSD) geometries. For each, a side view (lower left), a top view (top), and an end view (right) are shown. The spacer regions are not shown in order to show the underlying semiconductor fin, shaded in grey. The hatched region represents the gate metal. The gate oxide located underneath the gate metal is visible in the end views of end and saddle/slot contact FinFETs. The source and drain contact surfaces are shown in black. For the saddle/slot geometry, the source and drain contacts extend further to the side and above than shown, to the edge of the simulation region; however, only the near-source/drain-surface portions are shown for visual clarity.

Dimension	End, Saddle/Slot	Raised Source/Drain
L_c [nm]	8	8
L_{EXT} [nm]	5	5
L_G [nm]	18	18
H_{FIN} [nm]	35	35
W_{FIN} [nm]	6	6
H_{BOX} [nm]	10	10
T_{OX} [nm]	3	3
$\Delta W_{S/D}$ [nm]	0	6

Table 3.1: Modeled FinFET dimensions.

3.2.2 InGaAs band structure models

As noted, we considered two models of the $In_{0.53}Ga_{0.47}As$ band structure, a MV-InGaAs model and, for reasons discussed in the introduction, a Γ-InGaAs model. In our MV-InGaAs model, we take the inter-valley separation between the light-mass Γ-valley and heavy-mass peripheral L-valleys and X-valleys as $\Delta E_{\Gamma\text{-L}}$ = 487 meV and $\Delta E_{\Gamma\text{-X}}$ = 610 meV, respectively, as determined by a set of bowing parameters recommended by Vurgaftman and colleagues in their comprehensive review article [101]. These values lie between the previously-noted tight-binding and experimental values of [83] and [84], respectively. With the assumed 5×10^{19} cm^{-3} doping for MV-InGaAs, the equilibrium Fermi energy is found nearly 500 meV above the conduction band edge, high enough to place approximately 40% of the equilibrium bulk carrier concentration in the L-valleys for the assumed Γ-to-L energy valley separation, as a consequence of the degenerate statistics, the much larger L-valley than Γ-valley mass, and four-fold L-valley degeneracy [54]. In contrast, in Si the Fermi energy is found only approximately 100 meV above the conduction band edge with the degenerate statistics, despite the four-fold larger assumed doping. For the Γ-InGaAs model, $\Delta E_{\Gamma\text{-L}} \to \infty$, $\Delta E_{\Gamma\text{-X}} \to \infty$. Note that this work is mute on

which model of InGaAs is physically more realistic, Γ-InGaAs or MV-InGaAs, we simply consider the consequences of both.

We would be remiss not to note that at degenerate doping levels, charge carriers are not created by ionization of donor states to the conduction band with a commensurate rise in the Fermi level, but by merging the donor states with energy valley edges and a commensurate lowering of the conduction band edge below the Fermi level [102], [103]. In this way, in particular, the effective peripheral valley separations in MV-InGaAs in the S/D would be larger than otherwise expected, and the peripheral valley occupations would be reduced or eliminated, accordingly. This physics is not addressed in the simple band-structure models of this work. However, as discussed later, the modeled ideal and non-ideal specific contact resistivities of InGaAs are only weakly dependent on the assumed energy valley separations. Within the undoped channel region, energy valley separations are not impacted by the doping, while being reduced considerably by quantum confinement. And, although there may be some advantage to reducing the fraction of carriers in the peripheral valleys in the S/D, we found previously [54] that the peripheral valleys in the channel become heavily occupied in the ON-state in modeled MV-InGaAs FinFET through inter-valley scattering even when not occupied in the modeled source and drain under lower doping, because of the previously noted quantum-confinement-reduced valley separation in the channel and limited Γ-valley quantum capacitance, particularly in saturation.

3.3 Common Performance Measures and Results for Unity Transmissivity Contacts

To analyze device performance, we initially compare transconductance (g_{m}) and the peak thereof, on-current (I_{on}), subthreshold swing (S), and drain-induced barrier

lowering (DIBL) in the off-state, as well as the turn-on abruptness, as measured by the difference (ΔV_{T}) between two different estimates of threshold voltage (V_{T}). $g_{\mathrm{m}} = \partial I_{\mathrm{DS}}/\partial V_{\mathrm{GS}}$ where I_{DS} and V_{GS} are the drain-to-source current and gate-to-source voltage, respectively, is obtained from a centered moving average over an interval of ten V_{GS} samples to reduce noise in the data. The drain-to-source voltage V_{DS} is set to the supply voltage $V_{\mathrm{DD}} = 0.6$ V in accordance with ITRS predictions [1]. V_{GS} then was swept from OFF to ON in steps of 25 mV (and somewhat beyond the 0 to 0.6 V range in practice to allow for initially unknown thresholds and exhibition of some behavior beyond the normal operating regime). I_{DS} was divided by the fin perimeter ($2H_{\mathrm{FIN}} + W_{\mathrm{FIN}}$ in Fig. 1) for the purpose of calculating current density. The turn-on abruptness measure is $\Delta V_{\mathrm{T}} = V_{\mathrm{T}}^{\mathrm{ELR}} - V_{\mathrm{T}}^{\mathrm{CC}}$. Here, $V_{\mathrm{T}}^{\mathrm{CC}}$ is the threshold voltage, as obtained by the constant current (CC) method, which is widely used in industry and serves as a reference for our I_{on} calculations, where the threshold is defined by a fixed I_{DS} target. In this work, we take $I_{\mathrm{DS}}\left(V_{\mathrm{GS}} = V_{\mathrm{T}}^{\mathrm{CC}}\right) = 0.01$ mA/µm at $V_{\mathrm{DS}} = V_{\mathrm{DD}}$. $V_{\mathrm{T}}^{\mathrm{ELR}}$ is the threshold voltage as obtained by extrapolation in the linear region (ELR) [104], i.e., by linear extrapolation from the point of maximum slope (peak g_{m}) of the I_{DS} vs. V_{GS} curve in the ON-state, back to the intercept with the V_{GS} axis. The ON-state current, I_{on}, is then calculated at the gate overdrive voltage above threshold of $V_{\mathrm{GS}} - V_{\mathrm{T}}^{\mathrm{CC}} = 0.35$ V with, again, $V_{\mathrm{DS}} = 0.6$ V and $V_{\mathrm{T}} = 0.25$ V. Thus, the reported ON-state currents are dependent on the values of both peak g_{m} and ΔV_{T}. (We note that with a $V_{\mathrm{T}}^{\mathrm{CC}}$ of 0.25 V and a constant S of 65 mV below threshold, which is roughly consistent with our results to follow, the off-state current would be on the scale of 1 nA/µm, which lies between ITRS specifications for high-performance and low-power MOSFETs [1]). $\mathrm{DIBL} = -d\Phi_{\mathrm{b}}/dV_{\mathrm{DS}}$, where Φ_{b} is the channel potential barrier, is calculated well below threshold with $V_{\mathrm{DS}} = 0.6$ V. Subthreshold swing, $S = (\ln 10) dV_{\mathrm{GS}}/d(\ln I_{\mathrm{DS}})$, is calculated well below threshold in

terms of Φ_b within a simple thermionic emission model due to the lack of sufficient statistics for direct calculation with the small currents well-below threshold, and under zero V_{DS} representing the linear regime of operation. Simulation results are provided in Figs. 3.2 and 3.3 and discussed in detail below.

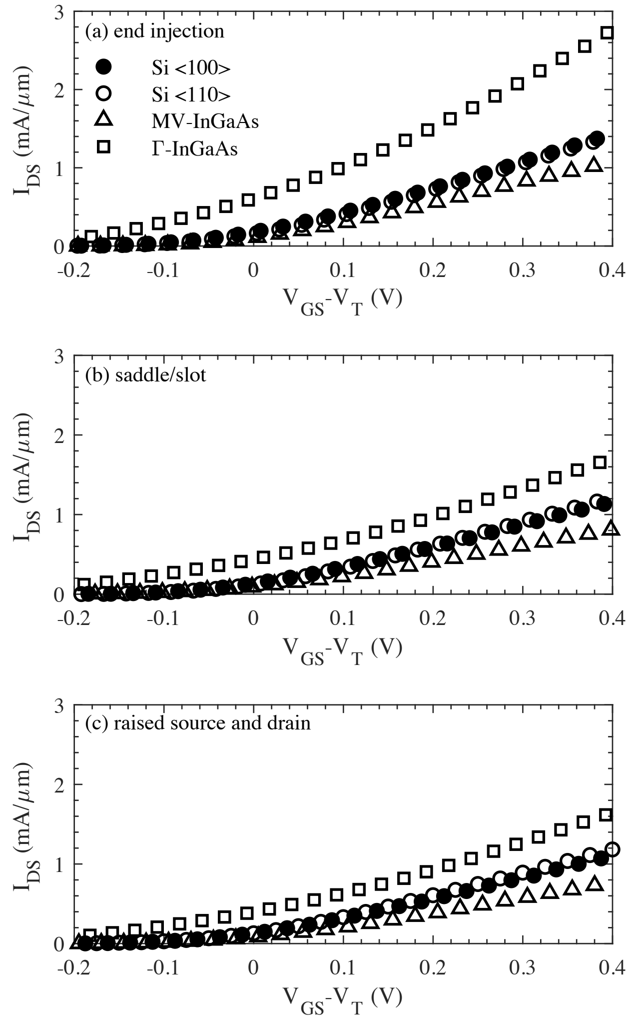

Figure 3.2: I_{DS}-V_{GS} simulation results for L_G = 18 nm Si ⟨110⟩ (open circles), Si ⟨100⟩ (solid circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), and Γ-$In_{0.53}Ga_{0.47}As$ (open squares) FinFETs for (a) end injection, (b) saddle/slot, and (c) raised source and drain. V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method.

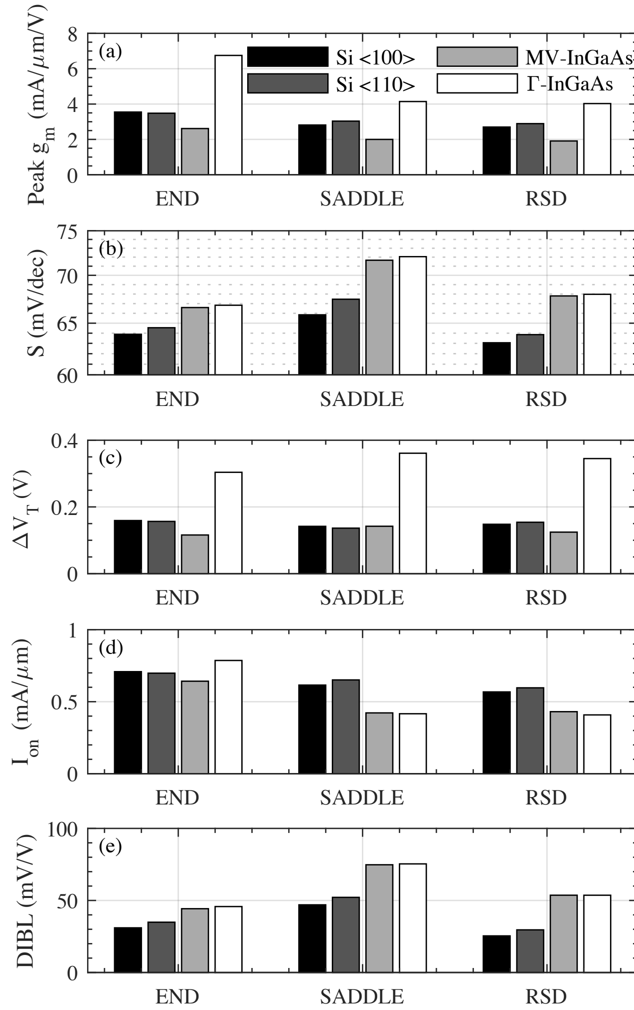

Figure 3.3: Dependence of (a) (centered moving average of) the peak of the transconductance g_{m}, (b) subthreshold swing S, (c) turn-on transition voltage ΔV_{T}, (d) on-current for the constant current defined threshold, I_{on}(CC), and (e) drain-induced barrier lowering, DIBL, for the end, saddle/slot, and RSD contacts to an 18 nm gate length FinFETs.

3.3.1 Transconductance, g_{m}

As shown in Fig. 3.2(a) and 3.3(a), Γ-InGaAs had by far the greatest peak g_{m} for end injection. The small transport mass in the Γ-valley of Γ-InGaAs produces a high channel injection velocity, which, along with limited backscattering in the channel, more than overcomes any reduced carrier concentration in the channel due to the lower

quantum capacitance. In contrast, for MV-InGaAs, the limited density of states in the Γ-valley pushes the carriers high into that valley, while quantum mechanical confinement substantially reduces the band offsets between the low density-of-states Γ-valley and high density-of-states L-valleys. Now, more readily than in the bulk considered previously, electrons transfer to L-valleys, with an accompanying decrease in group velocity and increase in scattering rate. As a result, peak g_{m} is reduced not only as compared to that of Γ-InGaAs-channel FinFETs, but also as compared to Si channel FinFETs in these simulations (analogous to reduction of the high-field electron velocity in bulk GaAs below that in bulk Si).

For the modeled saddle/slot and RSD contact geometries, the advantage of Γ-InGaAs over the other systems in peak g_{m} decreases substantially, as shown in Figs. 2(b) and (c), and 3(a). Moreover, both the RSD and saddle contacts somewhat favor a $\langle 110 \rangle$ channel orientation for Si, while, if anything, end contacts slightly favor a $\langle 100 \rangle$ channel orientation, which suggests that the $\langle 110 \rangle$ channel orientation advantage for RSD and saddle contacts is associated with contact geometry and not transport through the quantum-confined channel.

Informative of the transport physics is that, from Fig. 3.3(a), FinFET performance in terms of peak g_{m} degrades for each material system when going from the end contact geometry to the saddle/slot and RSD contact geometries, most so for Γ-InGaAs FinFETs and least so for the Si $\langle 110 \rangle$ FinFETs. For diffusive transport, however, g_{m} should be greatest for the saddle geometry, and worst for the end-contact geometry, because of the proximity of the S/D contacts to the channel and contact surface area. The juxtaposition of these two results suggests that transport in all of these simulated FinFETs leans toward ballistic, strongly so for Γ-InGaAs FinFETs and to a lesser degree for the remaining devices.

To better understand the effects of the contact geometries and source starvation for FinFETs as the ballistic limit is approached, we conceptually consider perfectly injecting and absorbing contacts and specularly-reflecting hard wall closed boundaries in the source and drain regions, as used in simulations throughout this work. To those assumptions, we add a few more chosen for illustrative value in the immediate discussion here (only): drain voltages sufficiently large that electron injection from the drain to source can be neglected; a unity transmission probability for electrons reaching the source extension with sufficient kinetic energy along the channel to overcome the channel potential barrier, and a zero transmission probability otherwise, which makes the former electrons the only ones of interest here and the source extension a perfectly absorbing boundary for these electrons of interest; and a uniform (i.e., a perfectly-screened) potential (flat-band conditions) within the source region. As illustrated in Fig. 3.4, ballistic ray tracing (as well as simple symmetry across the reflecting end contact) in this system shows that the saddle/slot contact FinFET with a reflecting boundary at the end of the source region located at L_C from the edge of the source extension boundary (Fig. 1), may be replaced by two mirror image FinFETs with a saddle/slot contact around a common source region of length $2L_C$ connected to the source extensions of both FinFETs, without affecting injection of the electrons of interest into the source extension and channel beyond. Similarly, by ballistic ray tracing, the end contact FinFET with source length L_C may be replaced by one with an end contact and source length $2L_C$ (or of any other length), which, in turn, may be replaced by one with a source length of $2L_C$ with both an end contact and a saddle/slot contact. Thus, under these assumptions, the difference between the here-considered end contact FinFET and saddle/slot contact FinFET corresponds to the difference between having a both injecting and absorbing contact, or just an absorbing contact, respectively, at the end of a source region of length

$2L_C$, with an injecting and absorbing saddle/slot contact about the source sides and top in either case. This difference is enhanced by electron injection probabilities that are peaked naturally about the surface normal direction, which selects for surfaces that are aligned perpendicular to the drain extension entry, as illustrated in Fig. 3.5 for injection about the end-contact-normal plane running along the vertical plane of Si semiconductor fins. We also note that a ray-tracing analysis under these conditions for the raised source and drain geometry gives the same results as for the saddle geometry, consistent with the similar results observed in the full simulations here. Also, consistent with this discussion, full simulation results (not shown) of the strongly ballistic Γ-InGaAs FinFETs with end-plus-saddle contacts were very similar to those provided here for end contacts.

Moreover, Si $\langle 100 \rangle$ ($\langle 110 \rangle$) channel FinFETs have $\{100\}$ ($\{110\}$) contact surfaces that promote still greater (somewhat diminish) peaking of the electron injection about the surface normal, as also shown in Fig. 3.5. As a result, the loss of injection from the end contact for the hypothetical effective $2L_C$ source length saddle/slot and RSD FinFETs relative to the reference end contact FinFETs should be at least somewhat less of a loss for Si $\langle 110 \rangle$ channel FinFETs than for Si $\langle 100 \rangle$ channel FinFETs. This expectation also is borne out in the simulation results of Fig. 3(a), where a small disadvantage in g_m for the Si $\langle 110 \rangle$ channel FinFETs vs. $\langle 100 \rangle$ channel FinFETs with the reference end contacts becomes a small advantage with wrapped contacts, despite the simulations not being in the flat-band ballistic limit. This advantage in our simulations also exists despite any contribution in the source from the small high-field advantage in bulk Si for $\langle 110 \rangle$ transport over $\langle 100 \rangle$ transport [105], also captured by UTMC, which would provide a greater loss for wrapped contacts relative to end contacts for Si $\langle 110 \rangle$ channel FinFETs. The simulated contact-related performance difference with channel orientation is more modest than the contact-shape effects and perhaps smaller than the

previously discussed uncertainty in the channel-related performance differences with orientation. Even if so, it further illustrates the common concept by which we qualitatively explain both this smaller effect and the larger contact shape effects observed in simulation.

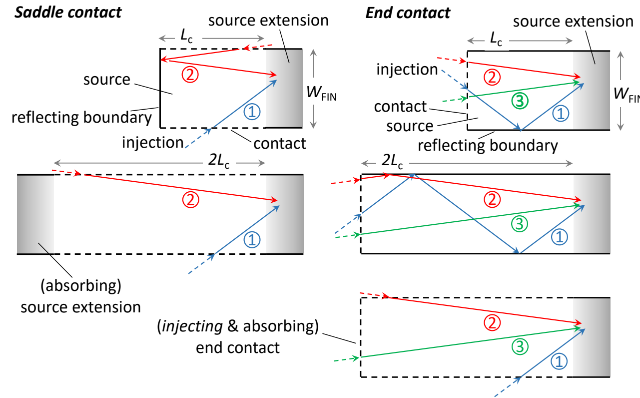

Figure 3.4: Injection of ballistic electrons under flat-band conditions into the source extension from perfectly injecting and absorbing saddle (left) and end (right) contacts, and equivalent contact geometries under these conditions. Note for the saddle contact geometry the direct injecting paths, illustrated by Path 1, and the reflection-mediated injection paths, illustrated by Path 2, and their counterpart paths in the equivalent geometry. There are corresponding paths into the source extension for the end contact, again illustrated by Paths 1 and 2, although those represented by Path 1 are now reflection-mediated and those represented by Path 2 are direct injecting, and their counterparts in its equivalent geometries. However, there is another category of paths for the end contact geometry, illustrated by Path 3, for which there are no counterparts for the saddle geometry. For these paths, electrons are effectively injected from an end contact at a distance $2L_c$ from the source extension, directly into the source extension.

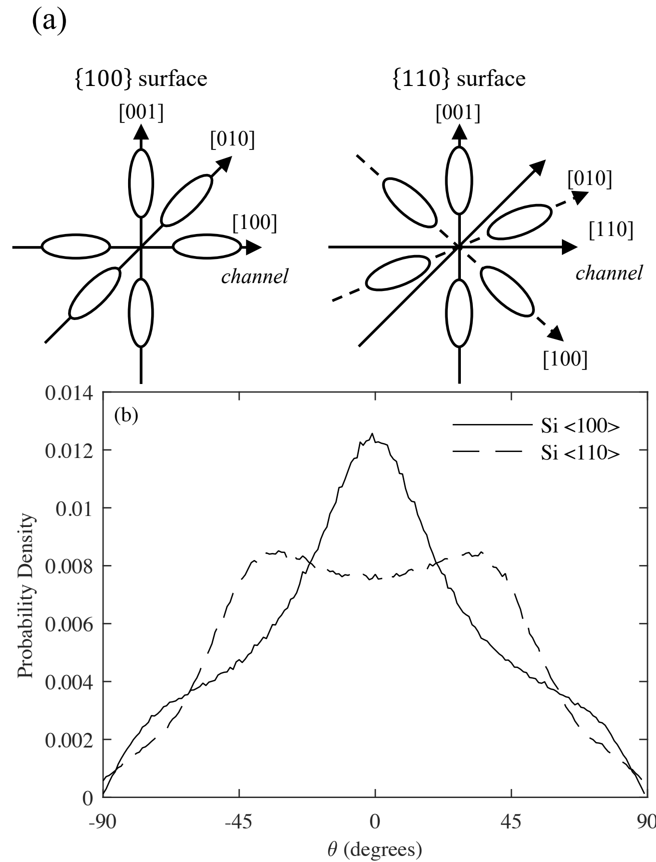

Figure 3.5: (a) Alignment of the conduction channel relative to the Si conduction band energy valleys for (on the left) ⟨100⟩ and (on the right) ⟨110⟩ channel orientations on a {100} substrate. (b) UTMC-simulated carrier injection probability density per degree of the carrier injection angle with respect to the plane of the channel for channel-end-injected carriers, for Si ⟨100⟩ (solid line) and ⟨110⟩ (dashed line) channel orientations.

3.3.2 *S*, DIBL, turn-on characteristic, and on-state current

In terms of electrostatic control, most notably, the Si channel FinFETs also have better (smaller) off-state subthreshold swing *S* and, more so, DIBL—where DIBL is measured with a narrower potential barrier along the channel—than InGaAs channel FinFETs (Figs. 3(b) and (e), respectively). The reason for this difference between Si and InGaAs channel FinFETs is likely largely the difference in dielectric constants, leading to

stronger electrostatic coupling for InGaAs to the source and drain regions. Moreover, this difference is most significant for the saddle/slot geometry with the potential pinned at the outer edges of the separate confinement regions, and the least significant for end contacts where coupling to the source and drain is weaker as the band bending extends a few screening lengths into the source and drain.

Turn-on behavior, as measured by ΔV_T, (Fig. 3.3(c)) is the slowest for Γ-InGaAs and, unlike for S and DIBL, also differs substantially from that of MV-InGaAs. These differences suggest that it may be related to the smaller quantum capacitance of the Γ-InGaAs as compared to the Si and even to MV-InGaAs channel FinFETs. As a result of this slow turn-on characteristic, the Γ-InGaAs channel FinFET has a lower I_{on} with a constant-current-defined threshold for the saddle/slot and RSD contact geometries compared to Si, despite better peak g_m. The Γ-InGaAs channel, still provides an I_{on} advantage for end contacts, but considerably less than for g_m. Moreover, we note that these relative difference in ΔV_T and the corresponding impact on I_{on} are conservative given our use of a large constant current threshold.

3.3.3 Drain current vs. drain voltage

The drain current also was calculated at the overdrive gate voltage of $V_{GS} - V_T^{CC} = 0.35$ V as a function of drain voltage V_{DS} swept from 0 V to 0.6 V in steps of 25 mV, consistent with the transistor in the on-state with $V_T^{CC} = 0.25$ V and a $V_{DD} = 0.6$ V, as shown in Fig. 3.6. All FinFETs with all contact configurations showed onset of current saturation between approximately $V_{DS} = V_{DS,sat} = 0.20$ V and 0.25 V, except for the Γ-InGaAs with end contacts, where the onset was delayed by approximately 0.05 V. However, the current saturation was the best, i.e., had the least dependence on V_{DS} above $V_{DS,sat}$, for the Γ-InGaAs FinFETs for each contact geometry.

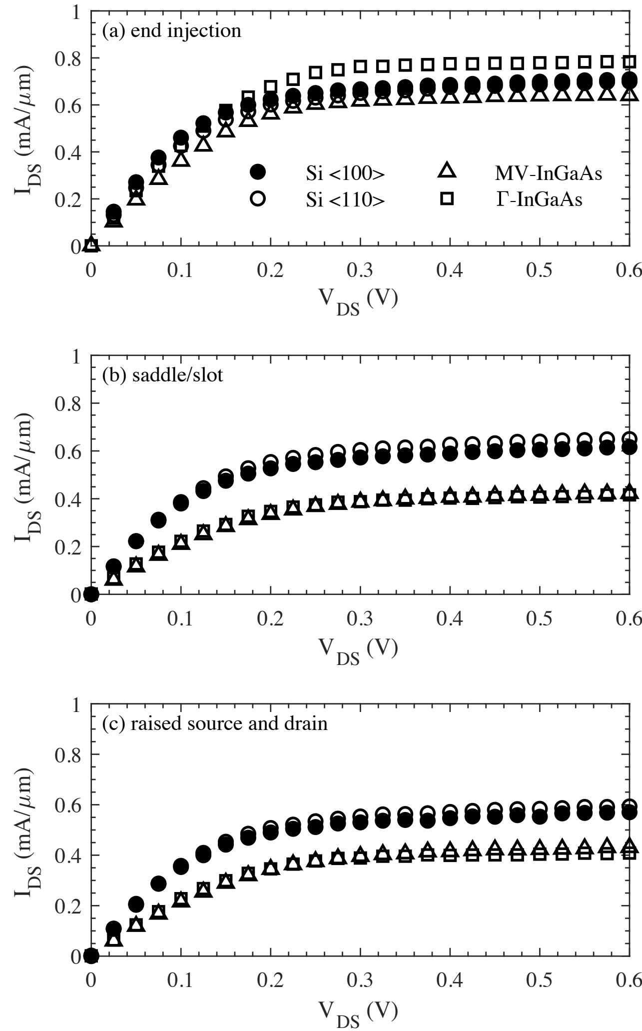

Figure 3.6: I_{DS}-V_{DS} simulation results for L_G = 18 nm Si ⟨110⟩ (open circles), Si ⟨100⟩ (solid circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), and Γ-$In_{0.53}Ga_{0.47}As$ (open squares) channel FinFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage for (a) end injection, (b) saddle/slot, and (c) raised source and drain.

3.4 Non-unity transmissivity contacts

In this section, the impact of parasitic (less than-ideal) S/D specific contact resistivity on the performance of the considered FinFETs is examined. We re-consider only the peak g_m, turn-on characteristic, ΔV_T, on state current, I_{on}, and the I_{DS} vs. V_{DS}

characteristic. S and DIBL are calculated in the absence of any significant current flow, so the specific contact resistivity is irrelevant.

For all considered FinFETs, we employ an illustrative fixed value of transmissivity of T = 0.2 as a control, which produces $\rho_{sp} = 4.5\rho_{LB}$ from Eq. (1). For silicon, ρ_{LB} = 3.0×10^{-10} Ω-cm^2 at the considered 2.0×10^{20} cm^{-3} doping concentration. Therefore, the corresponding specific contact resistivity is, ρ_{sp} = 1.35×10^{-9} Ω-cm^2, which is reasonably near a state-of-the-art reported value of 1.2×10^{-9} Ω-cm^2 [106]. For the Γ-InGaAs and MV-InGaAs channel FinFETs with ρ_{LB} values of 1.3×10^{-9} Ω-cm^2 and 1.4×10^{-9}, respectively, at the considered 5.0×10^{19} cm^{-3} doping concentration, this control value of T results in substantially larger ρ_{sp} values, of 5.9×10^{-9} Ω-cm^2 and 6.3×10^{-9} Ω-cm^2, respectively, which is still somewhat better than reported values of 7×10^{-9} Ω-cm^2 [107] for InGaAs.

3.4.1 Peak g_m, ΔV_T, and I_{on}

Overall, Fig. 3.7 shows that, as expected, non-ideal transmissivity contacts decrease the peak transconductances and on-currents, and, with the RSD and the model saddle/slot contact geometries having approximately 3.3 and 2.5 times the contact surface area as the end contact geometry, the relative reduction is the greatest for the end-contacts. The RSD geometry to some degree has greatest peak g_m for all materials systems. However, the saddle/slot geometry produces an I_{on} comparable to that of the RSD geometry for silicon ⟨110⟩ channel FinFETs, and greater than that of the RSD geometry for silicon ⟨100⟩ channel FinFETs. Moreover, also as expected, detrimental effects are the greatest on the Γ-InGaAs channel FinFETs, followed by the MV-InGaAs channel FinFETs. All Si channel FinFETs with all contact geometries now outperform all of their InGaAs channel counterparts in terms of peak g_m and, more so due to the slower

turn-on characteristic for InGaAs channel FinFETs, on-current with respect to a constant current threshold $I_{\mathrm{on}}^{(\mathrm{CC})}$. Despite substantially reduced contact transmissivity, there remains a contact-related advantage for the silicon $\langle 110 \rangle$ channel saddle/slot and RSD contact FinFETs over their silicon $\langle 100 \rangle$ channel counterparts in this work. For the ballistic flat-band approximation to the source (and, indeed, for any potential with a vanishing source-end normal field), either wrapped contact FinFET of source length L_{C} still can be replaced by two mirror images devices with a common wrapped contact of length $2L_{\mathrm{C}}$ by symmetry, despite only partially transmitting boundary conditions. And the resulting loss of injection from the end contact for these hypothetical effective $2L_{\mathrm{C}}$ source length wrapped contact FinFETs still should be less significant for the Si $\langle 110 \rangle$ channel FinFETs than for Si $\langle 100 \rangle$ channel FinFETs, providing an advantage for the former. However, the detailed explanation varies somewhat. With the partially reflecting boundary conditions, the number of electrons reaching the source extension from an end contact would be much more comparable among device orientations. However, off-normally injected electrons would do so with a smaller component of energy along the channel direction on average, and so be less likely to make it over the channel barrier to be among the electrons of interest. Therefore, the contact-related orientation advantage for the Si $\langle 110 \rangle$ channel FinFET on average per (not) injected electron, while weakened, remains. Moreover, with the reduced transmissivity, the contacts and any dependence on their geometry becomes more important to overall device performance as already seen.

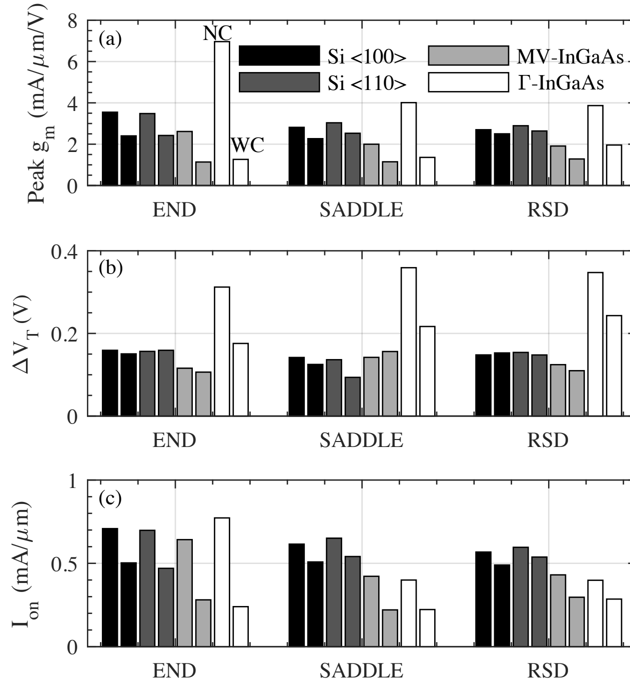

Figure 3.7: Comparison of contacts with perfect transmissivity and imperfect transmissivity contacts on the (a) peak of the transconductance g_m, (b) turn-on transition voltage ΔV_T, and (c) the on-current I_{on} with a constant current defined threshold (I_{on}(CC)), for the end, saddle/slot, and RSD contacts to an 18 nm gate length FinFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.2 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system, including channel orientation for Si.

3.4.2 Drain current vs. drain voltage

Fig. 3.8 shows drain current I_{DS} vs. drain voltage V_{DS} for the considered Si and InGaAs channel FinFETs. Performance degradation consistent with the I_{on} of Fig. 3.7 is evident. However, for InGaAs with end contacts in particular, the lack of stretch-out of this I_{DS}-V_{DS} characteristic with respect the saturation drain voltage when compared to the data of Fig. 3.6 contrasts to what would be expected using a lumped external resistance or distributed specific contact resistivity localized to the contact surface. With the latter

models, a significant portion of the energy of an injected electron gained from V_{DS} is be dropped before electron enters the source contact surface. However, in the ballistic limit, that energy is not dropped until after the electron is absorbed (if not back reflected) by the drain reservoir, just as for perfectly injecting and absorbing contacts.

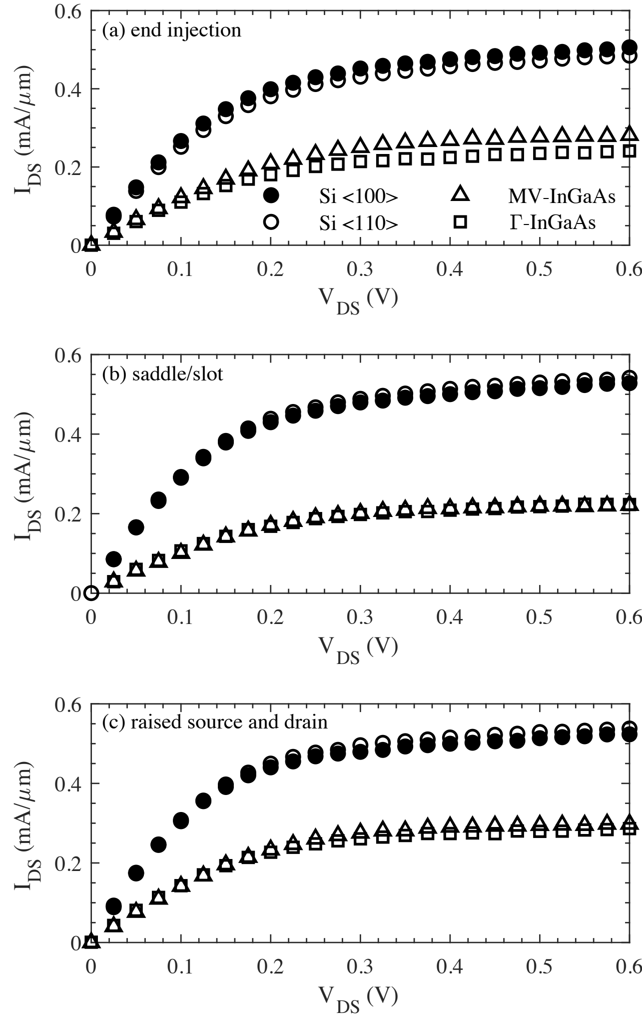

Figure 3.8: Simulated I_{DS}-V_{DS}, as for Fig. 3.6 but with non-unity transmissivity contacts: I_{DS}-V_{DS} simulation results for L_G = 18 nm Si ⟨110⟩ (open circles), Si ⟨100⟩ (solid circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), and Γ-$In_{0.53}Ga_{0.47}As$ (open squares) FinFETs. The gate overdrive voltage above the constant current threshold voltage is 0.35 V with supply V_{DD} and threshold V_T voltages of 0.6 V and 0.25 V, respectively, for (a) end injection, (b) saddle/slot (where the MV-InGaAs and Γ-InGaAs data are difficult to distinguish), and (c) raised source and drain. For InGaAs with end contacts in particular, the lack of stretch-out of this I_{DS}-V_{DS} characteristic with respect the saturation drain voltage ($V_{DS,sat}$) as compared to the data of Fig. 3.6 contrasts to what would be expected using a lumped external resistance or distributed specific contact resistivity.

3.5 Conclusion

The effects of contact geometry and specific contact resistivity on $In_{0.53}Ga_{0.47}As$ and Si nanoscale (18 nm channel length) n-channel FinFETs performance, and the effects of models thereof, were studied using a quantum-corrected semi-classical Monte Carlo method. Saddle/slot, raised source and drain (RSD), and reference end contacts were modeled. Both ideal perfectly injecting and absorbing contacts and those with more realistic specific contact resistivities were considered. Far-from-equilibrium degenerate statistics, quantum-confinement effects on carrier distributions in real-space and among energy valleys and on scattering, and quasi-ballistic transport were modeled. Silicon $\langle 110 \rangle$ channel and Si $\langle 100 \rangle$ channel FinFETs, multi-valley InGaAs channel FinFETs with conventionally-reported InGaAs energy valley offsets (MV-InGaAs), and a reference idealized Γ-valley-only InGaAs (Γ-InGaAs) channel FinFETs were simulated. Among our findings, echoing those of [75], InGaAs FinFETs were highly sensitive to contact geometry and specific contact resistivity and to the band structure model, while Si FinFETs showed still significant but much less sensitivity to contact models. For example, for idealized unity transmissivity contacts, Γ-InGaAs channel FinFETs performed best for all contact geometries, at least in terms of transconductance, and end contacts provided the best performance for all considered channel materials. For realistic contact resistivities, however, results of this work are essentially reversed. Silicon channel FinFETs performed best for all contact geometries, and saddle/slot and RSD contacts outperformed end contacts. These simulation results challenge the potential of InGaAs channel FinFETs, but they also suggest that the relative insensitivity of Si channel FinFET performance to contact design, and perhaps other device features, have allowed design choices that must be reconsidered to optimize InGaAs channel FinFET performance.

Chapter 4: Semi-Classical Monte Carlo Study of Gate Length Scaling Impact on Quasi-Ballistic Nanoscale Si, Ge, and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs

4.1 Introduction and Background

To assess the viability of silicon (Si), germanium (Ge), and indium-gallium-arsenide (InGaAs) channel materials in future CMOS technology nodes, realistic S/D contact geometries and the effects of quantum mechanical confinement must be considered. In quasi-ballistic devices, "wrapped" contact geometries can limit device performance because it becomes harder to get charge into the channel, which otherwise would not be expected with end contacts as typically employed in full quantum [93], [94] or multi-sub-band Boltzmann transport (Monte Carlo or deterministic) simulations [63], [90]–[92]. These types of simulations have predicted InGaAs MOSFETs to perform better than their Si counterparts. However, it has also been shown that relative performance expectations can vary significantly with contact geometry [75]. Additionally, quantum-confinement can remove band degeneracy even without the application of strain, and thus, Si and Ge MOSFET performance would depend on the sidewall surface orientation [94], [108], [109].

Our work explores the limitations and challenges of gate length scaling, and associated fin width scaling, on n-channel Si, Ge, and $In_{0.53}Ga_{0.47}As$ (InGaAs) FinFET performance, while considering more realistic contact geometries and specific contact resistivities, surface sidewall orientation effects in Si and Ge, and peripheral valleys in InGaAs. Si $\langle 110 \rangle$, Si $\langle 100 \rangle$, multi-valley $In_{0.53}Ga_{0.47}As$ with conventionally-reported energy valley offsets (MV-InGaAs), idealized Γ-valley-only $In_{0.53}Ga_{0.47}As$ (Γ-InGaAs), Ge $\langle 110 \rangle$, and Ge $\langle 100 \rangle$ channel devices were modeled.

To this end, we employed our in-house quantum-corrected 3D semi-classical Monte Carlo (SCMC) tool, University of Texas Monte Carlo (UTMC), which allows for, among other things, modeling of far-from-equilibrium degenerate statistics, non-ideal contacts, quantum-confinement effects on carrier distributions in real-space and among energy valleys and confinement-dependent scattering rates, quasi-ballistic transport, short and long-range inelastic scattering, and complex contact geometries [54].

Gate length scaling and associated fin width scaling is demonstrated to have deleterious effects on device performance, most so for InGaAs and Ge ⟨100⟩ channel FinFETs and least so for Si ⟨110⟩ and Ge ⟨110⟩ channel FinFETs. We identified source starvation and quantum-confinement as performance issues for the scalability of FinFETs. Use of realistic contact geometries and transmissivities exacerbates source starvation, the inability of the source and drain (S/D) regions to replenish carriers in the channel [75], [86]–[88]. Fin width scaling further exacerbated source starvation effects and increase quantum-confinement effects in the channel for better (Si ⟨100⟩ channels and Ge ⟨110⟩ channels) or worse (MV-InGaAs channels), in the channel. The relative effect of source starvation is greater for longer mean-free path particles, such as Γ-valley electrons in InGaAs. The performance of Si and Ge FinFETs also depends on channel orientation via injection through the contacts and transport within the channel. Major results include ⟨110⟩ channel Ge scaled the best, and Γ-InGaAs scaled the worst with the most realistic contact model. Also include ⟨110⟩ channel Ge performed the best and MV-InGaAs performed worst under all simulation conditions. However, particularly with regard to comparing absolute device performance among channel materials, we acknowledge an uncertainty associated with using the same surface roughness (but not surface roughness scattering) for all materials as a control given uncertainties in the actual surface roughness with being dependent on channel and dielectric material and any

strain thereof, materials growth and etching methods, and even detailed device geometry [59], [110]. Our results are substantially influenced by quantum confinement in the channel and wrapped saddle/slot contact geometries and sub-unity transmissivities.

4.2 Simulated FinFET structure and band structure models

4.2.1 FinFET structure

The simulated device structure, illustrated in Fig. 4.1, is a saddle-contacted FinFET with either Si $\langle 110 \rangle$, Si $\langle 100 \rangle$, MV-InGaAs, Γ-InGaAs, Ge $\langle 110 \rangle$, and Ge $\langle 100 \rangle$ as the channel material with device geometry parameters listed in Table 4.2. Model devices included a 3 nm thick HfO_2 ($\varepsilon_r = 22.3$) gate oxide for an effective oxide thickness of 0.52 nm for all FinFETs for electrostatic calculations. To address near-surface barrier penetration of the wave-function for the InGaAs and Ge FinFETs, we model the oxide effective mass as that of HfO_2, $0.15 m_e$ [96]. However, for Si channel FinFETs, because there is a commonly-occurring thin SiO_2 gate-oxide interfacial layer even with high-k gate dielectrics, we model the oxide effective mass as that of SiO_2, $0.55 m_e$ [97]. The doping gradient within the 5 nm source and drain extensions (L_{EXT}) is 1 nm/decade and the channel is intrinsic with uniform doping, corresponding to no gate overlap or underlap.

We performed two gate length (L_G) scaling studies. In the first study, we scaled the gate length via L_G = 27, 21, 18, and 15 nm, while all other parameters were fixed, to obtain a design rule between the gate length and fin width (W_{FIN}) with respect to electrostatic control. In the second study, using the gate length to fin width design rule for Si as a control, we further scaled the FinFET gate length to L_G = 18, 15, 12, and 9 nm with all other parameters fixed [1].

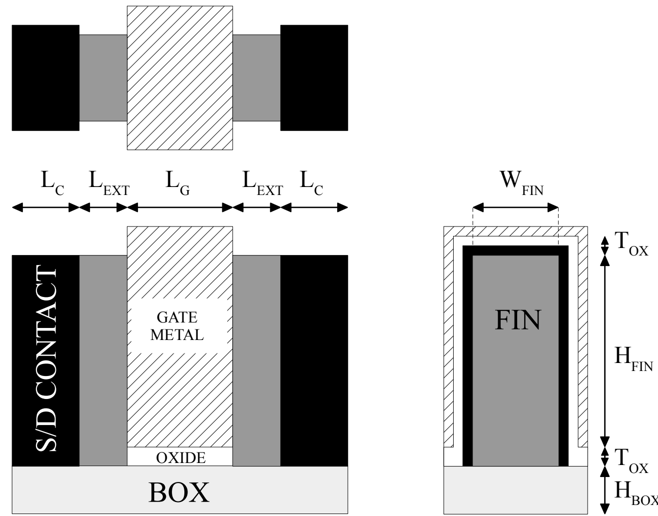

Figure 4.1: Schematic of the simulated modeled FinFET geometry with saddle/slot contacts. A side view (lower left), a top view (top), and an end view (right) are shown. The spacers regions are not shown in order to show the underlying semiconductor fin, shaded in grey. The hatched region represents the gate metal. The gate oxide located underneath the gate metal is visible in the end views of the saddle/slot contact model devices. We note that for the saddle/slot geometry, the source and drain contacts extend further to the side and above than shown, to the edge of the simulation region; however, only the near-source/drain-surface portions are shown for visual clarity. The source and drain contact surfaces are shown in black.

Dimension	**Saddle/Slot**
L_c [nm]	8
L_{EXT} [nm]	5
L_G [nm]	18, 15, 12, 9
H_{FIN} [nm]	35
W_{FIN} [nm]	6, 5, 4, 3
H_{BOX} [nm]	10
T_{OX} [nm]	3

Table 4.1: Modeled FinFET dimensions.

4.2.2 Band structure models

In our Si model, we considered 6 ellipsoidal Δ-valleys for the conduction band. Solid solubility data of Si suggests that arsenic might be active up to concentrations of 2×10^{21} cm^{-3}, but in practice it is difficult to actually achieve electrically active arsenic concentrations above 2×10^{20} cm^{-3}, where the corresponding equilibrium Fermi level is 100 meV above the conduction band edge considering degenerate statistics [99], [100]. We considered Si ⟨110⟩ channel devices and Si ⟨100⟩ channel devices due to differences in the conduction band-edge quantization mass can alter the effects quantum-confinement.

In our Si model, we considered 6 ellipsoidal Δ-valleys for the conduction band. Solid solubility data of Si suggests that arsenic might be active up to concentrations of 2×10^{21} cm^{-3}, but in practice it is difficult to actually achieve electrically active arsenic concentrations above 2×10^{20} cm^{-3}, which we use, where the corresponding equilibrium Fermi level is 100 meV above the conduction band edge considering degenerate statistics [99], [100].

In our conventional MV-InGaAs model, we considered 1 Γ-, 4 L-, and 3 X-valleys for the conduction band. The Γ-valley is modeled as spherical; the L- and X-valleys are modeled as ellipsoidal. In-situ Si doping during the epitaxial growth of InGaAs have yielded carrier concentrations of up to 5×10^{19} cm^{-3} [111], [112], which we use here. Because of the uncertainty in valley band-edge separations between the light-mass Γ-valley and heavy-mass peripheral L-valleys ($\Delta E_{\Gamma\text{-L}}$) and X-valleys ($\Delta E_{\Gamma\text{-X}}$) [82], [85], we also decided to study transport in the limiting case of a single Γ-valley-only InGaAs model with no satellite valleys. Γ-InGaAs also represents transport behavior where carriers injected into the Γ-valley from the contacts are not allowed to scatter into the peripheral valleys in the channel, such as for fully ballistic models of transport, for

larger valley offsets than those considered here, or for neglecting quantum-confinement in the channel, which reduces peripheral valley. For such highly degenerate doping, the equilibrium Fermi levels for MV-InGaAs and Γ-InGaAs are nearly 500 meV and 650 meV above the conduction band edge, respectively.

In our Ge model, we considered 1 Γ-, 4 L-, and 6 Δ-valleys for the conduction band, as in [52]. The Γ-valley is modeled as spherical; the L- and X-valleys are modeled as ellipsoidal. Maximum activated carrier concentrations of 2×10^{20} cm^{-3} have been obtained in heavily doped phosphorous n-type Ge thin films [113], which we use here, and which places the equilibrium Fermi level 160 meV above the conduction band edge.

While an attempt is made at simulating devices that are realistic, despite the very different technological maturities between Si, Ge, and InGaAs FinFETs, and obey physical limits, we note many assumptions inherent in the employed models (e.g. band structure for InGaAs, conservative estimates of surface roughness scattering, control value of contact transmissivity) invite improvement and led to optimistic, at least for now, but not necessarily unphysical, scenarios. Yet, the results presented in this work provide a good compromise between the necessary levels of sophistication to capture enough of the essential physics so as to be qualitatively accurate and view relevant trends and the computational burden.

Our simulation approach and its physical models that are used have been validated by comparing program output against several bulk and interface mobility experimental data sets, achieving excellent agreement, and requiring only small tuning of various phonon coupling constants [50]–[52], [57], [58], [60]. All our final simulation parameters such as valley-specific effective masses, non-parabolicity constants, and deformation potentials for Si and InGaAs are assembled in [54] and for Ge are tabulated in the Appendix.

4.3 DESIGN RULE FOR ELECTROSTATIC INTEGRITY

We determined a design rule relationship between L_G and W_{FIN} with respect to electrostatic control by calculating both DIBL and S well-below threshold. Holding all parameters constant including W_{FIN} = 6 nm, the device in Fig. 4.1 is scaled via L_G = 27, 21, 18, and 15 nm. Setting a benchmark for electrostatic control of $S < 70$ mV/decade and DIBL < 70 mV/V, a design rule of $L_G \geq 3 \times W_{FIN}$ is more than sufficient for Si and Ge FinFETs, whereas MV-InGaAs FinFETs require somewhat longer channels, approximately $L_G \geq 3.5 \times W_{FIN}$, as shown in Fig. 4.2, to meet these metrics. Electrostatic control is poorer in InGaAs devices most likely due to its higher dielectric permittivity, which leads to stronger coupling of the channel with the source and drain. Ge devices behave more like Si devices, despite having a larger dielectric constant than InGaAs, pointing to the combined effects of channel quantum capacitance and dielectric permittivity as a more important consideration than channel dielectric alone, which under these simulated conditions, favors Si devices. As L_G increases, S appears to converge to the thermodynamic limit of 60 mV/decade at room temperature. Thereafter, we simulated all devices using the gate length to fin width design rule for Si of $L_G = 3 \times W_{FIN}$ as a control, acknowledging that the subthreshold behavior would be worse for InGaAs devices.

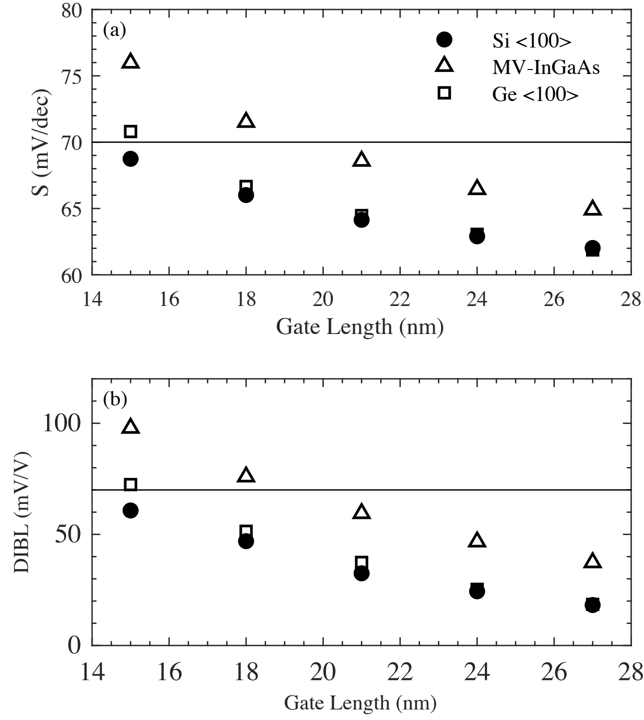

Figure 4.2: Dependence of (a) subthreshold swing S and (b) drain-induced barrier lowering, DIBL, in Si ⟨100⟩ (solid circles), Ge ⟨100⟩ (solid squares), and MV-InGaAs (open triangles) channel FinFETs with gate length with all other device parameters have been kept unchanged.

4.4 Effect of S/D doping concentration in Si

In Fig. 4.3, the performance of L_G = 18 nm Si ⟨110⟩ FinFETs are compared at S/D doping concentrations (Fermi-level) of 5×10^{19} cm^{-3} ($E_f = 24$ meV) and 2×10^{20} cm^{-3} ($E_f = 102$ meV) with perfect and reduced contact transmissivity. A fourfold increase in S/D doping had a relatively modest impact on device performance with ideal contact transmissivity. Peak g_m and I_{on} are improved approximately by a factor of 1.3 and 1.1, respectively, due to increased electron conductivity in the S/D, and the rate of turn-on is almost 1.3 times (33%) longer because the source and drain become more metallic. Similarly, DIBL and S, not shown here, are increased with larger S/D doping to the S/D

regions coupling more strongly to the channel. However, upon reducing the contact transmissivity to T = 0.2, Si devices with the S/D doped to 2×10^{20} cm^{-3} show about a factor of two enhancement in peak g_{m} and more similar ΔV_{T} compared to Si devices with S/D doped to 5×10^{19} cm^{-3}. Higher S/D doping not only improves the S/D injection efficiency by increasing the number of available electrons, but also increases the ionized impurity scattering rate due to increased doping and become more momentum randomizing due to decreased screening length, which can help to redirect side-injected carriers in the S/D into the channel to ameliorate source starvation. But increasing the S/D doping in InGaAs has diminishing returns once the Fermi level moves into the satellite L-valleys.

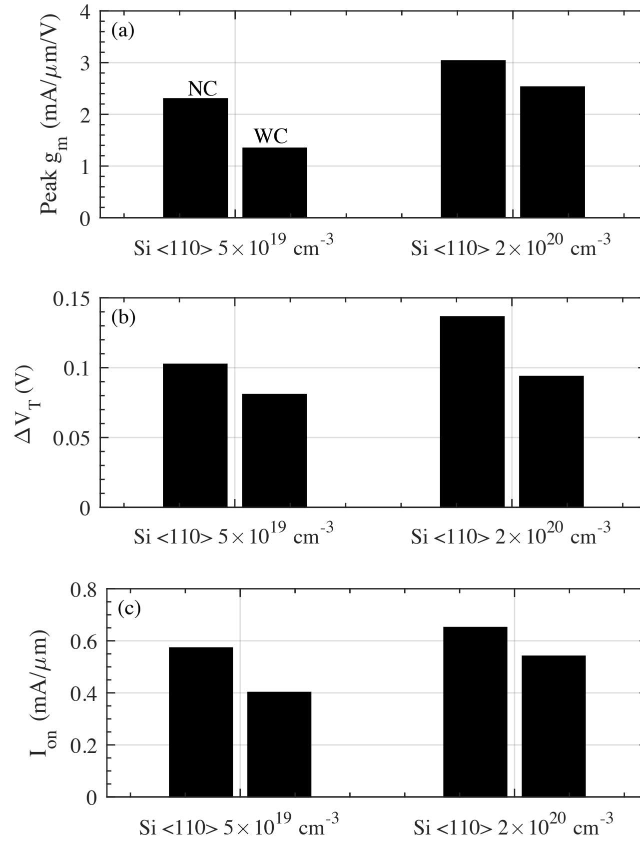

Figure 4.3: Comparison of contacts with perfect transmissivity and imperfect transmissivity contacts on the (a) peak of the transconductance g_m, (b) turn-on transition voltage ΔV_T, and (c) the on-current I_{on} with a constant current defined threshold (I_{on}(CC)), for 18 nm gate length Si ⟨110⟩ FinFETs with S/D doping concentrations of 5×10^{19} cm^{-3} and 2×10^{20} cm^{-3} at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.2 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system.

4.5 Gate length scaling impact on common performance measures for unity transmissivity contacts

Next, we studied gate length and fin width scaling on device performance. The gate length of the device from Fig. 4.1 is then scaled via $L_G = 3 \times W_{FIN}$ according to possible device nodes at L_G = 18, 15, 12, and 9 nm with all other parameters held fixed,

including the oxide thickness. As before, we simulated InGaAs devices using the same gate length to fin width ratio even though in practice they would require somewhat longer channel lengths to achieve the same level of electrostatic control as Si devices. At these gate lengths, the fin aspect ratio, H_{FIN}/W_{FIN}, ranges from 6 to 12, increasing with decreasing fin width, which is comparable to 7.57 (53/7~7.57), currently used in Intel's 10 nm FinFET technology [114]. Figs. 4.4(a) and 4.4(b) shows the I_{DS}-V_{GS} curves with L_G = 18 nm and L_G = 9 nm, respectively. Fig. 4.5 reports common performance measures and results at all the gate lengths considered in this study.

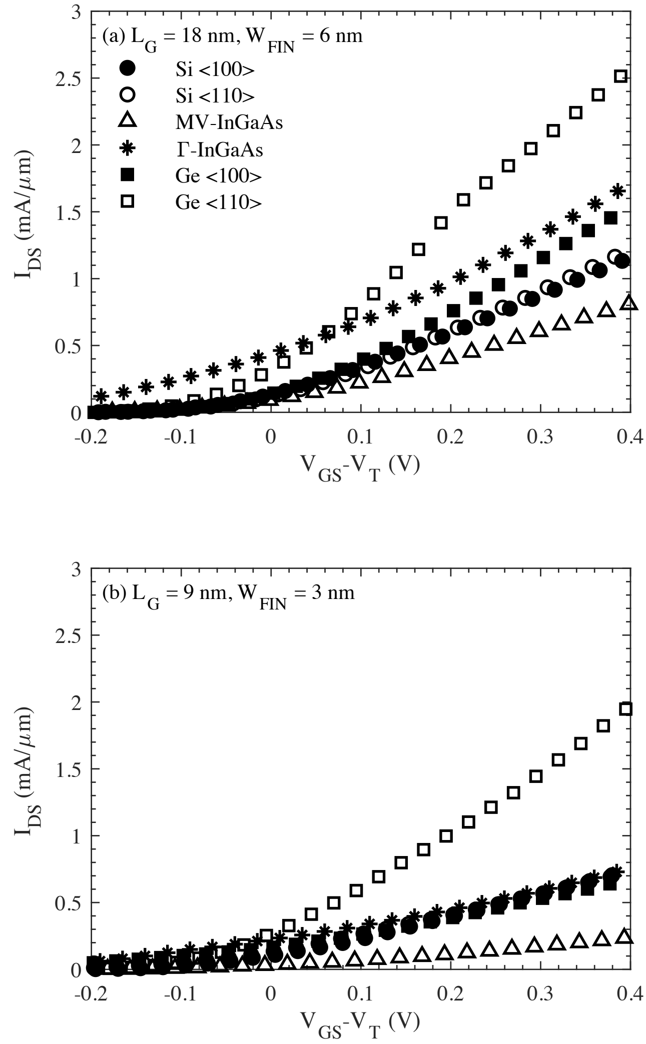

Figure 4.4: I_{DS}-V_{GS} simulation results for (a) L_G = 18 nm and (b) L_G = 9 nm Si ⟨100⟩ (solid circles), Si ⟨110⟩ (open circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), Γ-$In_{0.53}Ga_{0.47}As$ (asterisks), Ge ⟨100⟩ (solid squares), and Ge ⟨110⟩ (open squares) FinFETs. V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method. The S/D doping concentrations were 2×10^{20} cm^{-3} for Si or Ge FinFETs and 5×10^{19} cm^{-3} for InGaAs FinFETs.

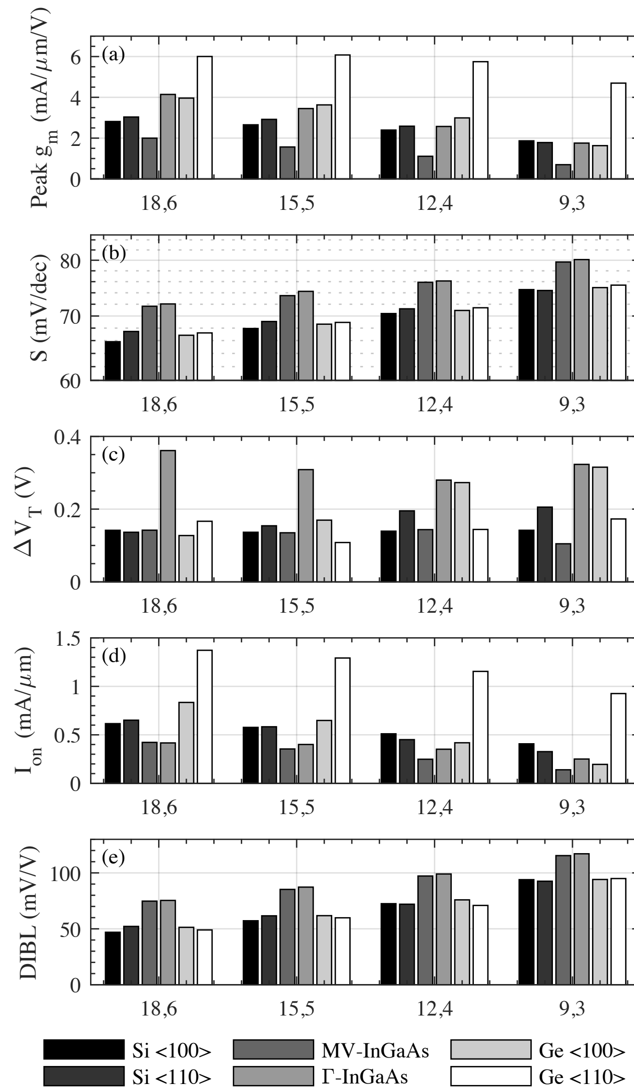

Figure 4.5: Dependence of (a) (centered moving average of) the peak of the transconductance g_m, (b) subthreshold swing S, (c) turn-on transition voltage ΔV_T, (d) on-current for the constant current defined threshold, I_{on}(CC), and (e) drain-induced barrier lowering, DIBL, with gate length scaling according to $L_G = 3 \times W_{FIN}$.

4.5.1 Transconductance, g_m

According to Figs. 4.4 and 4.5(a), gate length scaling and associated fin width scaling degrades g_m. One apparent drawback of shrinking of the fin width is that it

gradually reduces the top gate device area, effectively leaving only the fin sidewalls to carry the majority of the current. At L_G = 18 nm, Ge ⟨110⟩ channel devices show a factor of two enhancement in peak g_m, and to a lesser extent by Ge ⟨100⟩ channel devices, over Si devices due to higher injection velocities and comparable channel quantum (density of states) capacitance provided by Ge L-valleys, which is consistent with Ge n-channel nanowire simulations [108]. A limiting InGaAs device modeled with a Γ-valley-only shows a factor of 1.4 enhancement in peak g_m over Si devices as the large injection velocity is able to compensate for reduced carrier concentration in the channel due to the limited quantum capacitance. On the other hand, MV-InGaAs devices perform the worst as peripheral valleys in the channel become heavily occupied in the ON-state through inter-valley scattering, and even when not occupied in the unconfined S/D regions under conditions of lower doping, carriers will inevitably transfer to the peripheral valleys in the channel [54]. Silicon ⟨110⟩ channel devices perform somewhat better than Si ⟨100⟩ channel devices as a result of favorable alignment of the conduction band energy valleys relative to the channel direction. In all material systems, gate length scaling reduces g_m. Source starvation removes the peak g_m advantage by Γ-InGaAs devices over Si devices at L_G = 12 nm (W_{FIN} = 4 nm), which is consistent with the performance reduction found in high-aspect ratio InGaAs FinFETs [115] and ultra-thin-body InAs MOSFETs [116]. Stronger quantum confinement causes Ge ⟨100⟩ channel devices to underperform Si devices at L_G = 9 nm, whereas Ge ⟨110⟩ channel devices, for all gate lengths, outperform Si devices by factor of two in peak g_m owing to the smallest conductivity mass of the L-valley and moderated quantum confinement. At L_G = 9 nm, Si devices slightly favor ⟨100⟩ channel orientations over ⟨110⟩ channel orientations in terms of peak g_m and even more so in terms of I_{on}. Scaling performance in FinFETs can be divided into two regimes separated by a critical fin width of about 4 nm in our simulations, above which, device

performance is source-limited via source starvation and below which, device performance is, or at least also is substantially, channel-limited via quantum-confinement.

Gate length scaling impact on device performance can be decomposed into relative contributions from carrier transport in the source region and the channel region. The inability to redirect side-injected carriers into the channel causes source starvation in saddle/slot and raised source and drain contact geometries, and the relative effect is greater for high thermal velocity materials as transport becomes increasingly quasi-ballistic transport at device scales. It becomes even harder to inject carriers into the channel for all material systems as the fin width becomes narrower, i.e. carriers are unable to pass through the narrow slit of the channel (fin) cross-section because of their injection orientation. InGaAs FinFETs are highly sensitive to source starvation due to carriers occupying the light-mass Γ-valley as seen by the rapid reduction in peak g_m with scaling in Fig. 4.5(a). Source starvation due to contact orientation and narrow-slit injection in Si FinFETs is somewhat counterbalanced as carriers are able to enter into the channel via momentum scattering and due to higher S/D doping. Additionally, the alignment of the L- and Δ-valleys in Ge $\langle 100 \rangle$ and Si $\langle 110 \rangle$ channel devices, respectively, results in injected carriers to be peaked naturally away from the contact normal, which means carriers have a greater probability to enter into the channel when considering ray tracing. Although, the loss of the perfectly reflecting boundary at the end of a source region with narrower fins may diminish this contact-related advantage. Yet as the fin width decreases, the transconductance in Ge $\langle 100 \rangle$ channel devices decrease further due to increased occupancy of Δ-valleys in the channel via stronger quantum-confinement, as detailed subsequently, relative to Si devices. We did not shrink the contact length, which would reduce the impact of narrower fin widths by reducing the

transfer time carriers potentially spend in the S/D regions, where they can be absorbed at the contacts, before they are injected into the channel. However, any reduction in contact area would otherwise increase contact resistance.

Device performance is also limited by transport in the channel as a result of stronger quantum-mechanical confinement. The decrease in carriers in the channel is consistent with the reduction in normal field at fixed gate voltage as narrower fins are unable to pull down the symmetric triangular quantum wells formed at the channel sidewalls. We do not scale the oxide, which would mitigate this loss through increased gate capacitance; however, it should be noted that the effective oxide thickness modeled in this work has already been extended (scaled) to end-of-the-roadmap for FinFETs [1]. As a consequence, quantum-confinement effects are exacerbated, the triangular quantum well begins to resemble a square well, the latter having a greater degree of confinement, and carriers are redistributed among the energy valleys through scattering via reduced inter-valley separations, which results in worse device performance. For MV-InGaAs, the limited density of states in the Γ-valley, $m_{\Gamma}^{*} = 0.042m_{\mathrm{e}}$, pushes carriers high into that valley, while quantum mechanical confinement substantially reduces the band offsets between the low density-of-states Γ-valley and high density-of-states L-valleys, and electrons readily transfer into the L-valleys in the channel, which have much slower carriers and much higher scattering rates. Likewise in Ge devices, quantum-confinement reduces the bulk inter-valley separation between the L- and Δ-valleys, nominally $\Delta E_{\text{L-}\Delta}$ = 173 meV, enabling electrons to readily transfer from the L-valleys, $m_{\mathrm{L}}^{*} = 0.26m_{\mathrm{e}}$, into the Δ-valleys, $m_{\Delta}^{*} = 0.48m_{\mathrm{e}}$, where transport in these valleys is expected to behave similar to that of a heavier-mass version of Si ⟨100⟩ devices. A competing effect emerges, as the ostensible benefit of greater occupation of larger density of states and heavier-mass subsidiary valleys would be increased quantum capacitance and moderated

quantum-confinement, respectively; however, the larger amounts of scattering in these valleys reduces the carrier's average injection velocity and injection efficiency (backscattering) at the barrier-top to reduce the drive current of these transistors further. Quantum-confinement also increases the average energy of carriers, which can overall reduce carrier velocities via the combination of substantial non-parabolicity and higher energy, most notably in Γ-InGaAs devices. Furthermore, the gate oxide-semiconductor interface can also modify the degree of confinement of the wavefunction that follows from their spreading into the oxide barriers, which in this work, penalizes InGaAs more so than Si [95].

There are very little differences in electrostatics among the two orientations of Si and Ge considered, and the orientation-dependent performance with scaling, beyond that of source starvation, can be attributed to quantum-confinement, or at least these two sources become convolved. Channel orientation leads to different degrees of quantum-confinement within the channel and source and drain extensions for electrons within the various otherwise-equivalent band-edge Δ-valleys or L-valleys in Si and Ge, respectively, as shown in Fig. 4.6, including the Γ-valley in InGaAs. In Si $\langle 100 \rangle$ channel devices, quantum-confinement breaks the six-fold degeneracy of the Δ-valleys, resulting in two lower-lying valleys, with circular constant-energy contours, and four higher-lying valleys with elliptical constant-energy contours once projected onto the $\{100\}$ surface due to the differences in the confinement mass of the valleys. The confinement mass (m_{c}) and channel (conductivity) effective mass (m_{ch}), i.e. mass along the direction of transport, of the lower two valleys is $m_{\mathrm{c}} = 0.98m_{\mathrm{e}}$ and $m_{\mathrm{ch}} = 0.19m_{\mathrm{e}}$, respectively. Meanwhile, in Si $\langle 110 \rangle$ channel devices with $\{110\}$ sidewall surfaces, quantum-confinement lifts a pair of Δ-valleys above four Δ-valleys. The lower four valleys have a $m_{\mathrm{c}} = 0.32m_{\mathrm{e}}$ and $m_{\mathrm{ch}} = 0.585m_{\mathrm{e}}$, respectively, and the latter channel effective mass is much heavier than

Si $\langle 100 \rangle$ channel devices, which leads to lower velocities. As a result of having the largest quantization mass and lightest channel effective mass, Si $\langle 100 \rangle$ devices begin to somewhat outperform Si $\langle 110 \rangle$ devices for fin widths less than 4 nm, which is consistent with the limit of ideal end injection. Compared to Ge $\langle 100 \rangle$ channel devices with $m_{\mathrm{c}} = 0.16 m_{\mathrm{e}}$ and $m_{\mathrm{ch}} = 0.20 m_{\mathrm{e}}$, Ge $\langle 110 \rangle$ channel devices results in a larger confinement mass and lighter channel mass of $m_{\mathrm{c}} = 0.29 m_{\mathrm{e}}$ and $m_{\mathrm{ch}} = 0.112 m_{\mathrm{e}}$, respectively, and the latter mass is about halfway between the channel mass of InGaAs and Si $\langle 100 \rangle$.

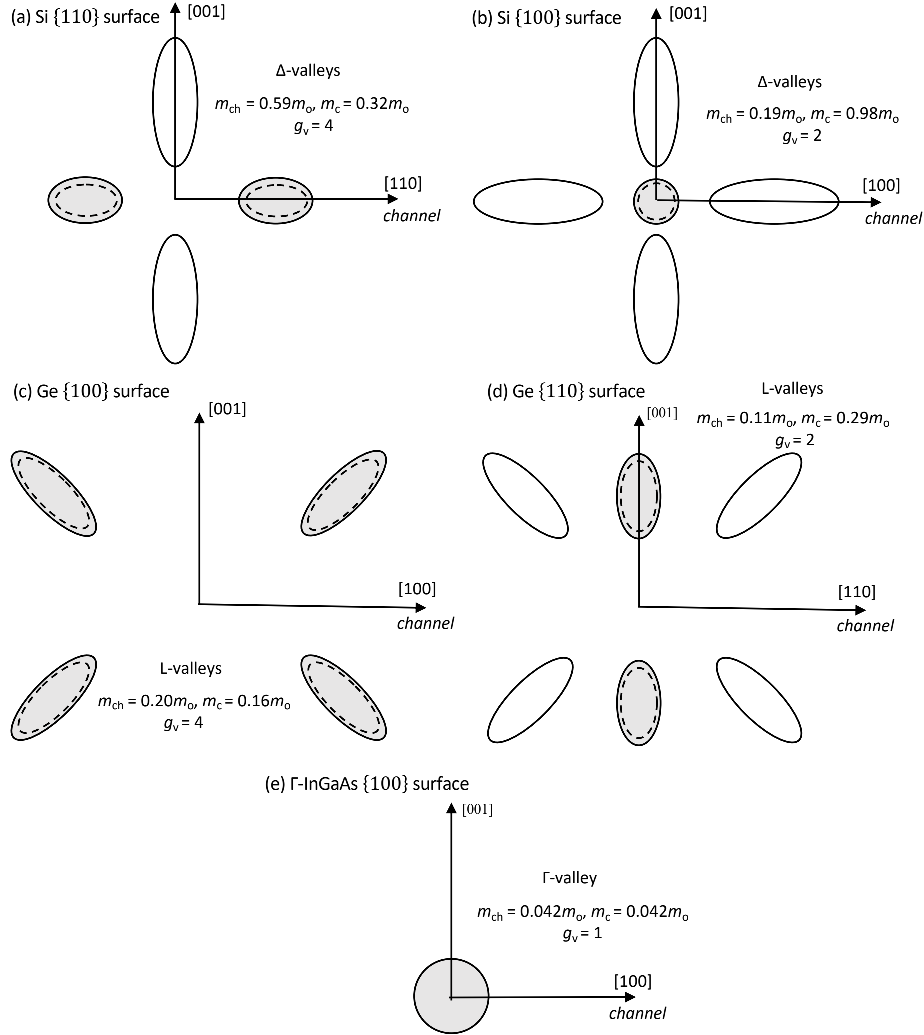

Figure 4.6: Projection of constant energy surfaces onto the (a) {110} plane of Si, (b) {100} plane of Si, (c) {100} plane of Ge, (d) {110} plane of Ge, and (e) {100} plane of InGaAs. The shaded regions correspond to the sub-band of lower energy and concentric circles or ellipses are shown as dashed lines to indicate two-fold degeneracy. The direction of confinement lies into the plane of the page. For Ge, the L-valleys are located at the zone boundary, and thus, one-half of each L-valley is inside the first Brillouin zone.

4.5.2 *S*, DIBL, turn-on characteristic, and on-state current

According to Figs. 4.5(b), (e), gate length scaling worsens (increases) S and DIBL. Silicon and Ge devices exhibit better gate control than both models of InGaAs due to the high dielectric constant in InGaAs, resulting in the channel to couple strongly to the S/D regions. DIBL is severely aggravated as gate length scaling brings the source and drain regions into closer proximity of the channel and to each other. Shortening of the gate length not only increases the parasitic capacitance, but also decreases the gate contact area along the fin sidewalls leading to reduced gate capacitance.

In all cases, Figs. 4.5(c), (d) show that gate length scaling and fin width scaling cause the on-current to decrease due to reduced charge density and/or reduced injection velocity as a result of source starvation and quantum-confinement, whereas the effect of scaling on the turn-on transition voltage is mixed. We also note that V_{on} increases for all studied devices with scaling as stronger confinement increases the sub-band energy splitting and therefore larger carrier concentrations are required to achieve the same performance defined at a fixed gate voltage above constant-current threshold, which is consistent with reduced DOS and increased threshold voltage experimentally observed in thin body SOI FinFETs [117]. Although Γ-InGaAs devices show promising g_{m}, the turn-on transition voltage ΔV_{T} is the longest, which differs from otherwise equivalent MV-InGaAs devices with the exception of included satellite valleys, suggesting limited quantum capacitance as to the reason for the slower turn-on. At L_{G} = 9 nm, Ge $\langle 110 \rangle$ channel devices deliver the largest I_{on} and ΔV_{T}, whereas Ge $\langle 100 \rangle$ channels have progressively worse turn-on behavior due to stronger threshold voltage shift. Meanwhile, ΔV_{T} for both MV-InGaAs and Si $\langle 100 \rangle$ devices remained relatively constant with scaling, whereas Si $\langle 110 \rangle$ devices showed an increase, which is consistent with stronger

quantum-confinement. Si ⟨100⟩ devices show a greater on-current than Si ⟨110⟩ devices after W_{FIN} = 4 nm due to the poorer turn-on characteristic in the latter orientation.

4.5.3 Drain current vs. drain voltage

The drain current also was calculated at the overdrive gate voltage of $V_{\mathrm{GS}} - V_{\mathrm{T}}^{\mathrm{CC}} = 0.35$ V as a function of drain voltage V_{DS} swept from 0 V to 0.6 V in steps of 25 mV, consistent with the transistor in the on-state with $V_{\mathrm{T}}^{\mathrm{CC}} = 0.25$ V and a $V_{\mathrm{DD}} = 0.6$ V, as shown in Fig. 4.7. At L_{G} = 18 nm, all devices showed onset of current saturation at $V_{\mathrm{DS}} = V_{\mathrm{DS,sat}}$ = 0.20 V and 0.30 V, except for Ge ⟨100⟩ FinFET, where the onset was delayed by approximately 0.1 V, and Ge ⟨110⟩ FinFET struggles to saturate. The greater depth of the Fermi-level in InGaAs devices compared to Si devices causes $\mathrm{V}_{\mathrm{DS,sat}}$ to be stretched-out. MV-InGaAs FinFETs performed the worst attributed to confinement-reduced inter-valley energy separations leading to transfer of electrons to heavier-mass satellite valleys, consistent with our g_{m} results. At L_{G} = 9 nm, the $\mathrm{V}_{\mathrm{DS,sat}}$ for Si ⟨110⟩ devices increased by 0.05 V and decreased by 0.05 V for Ge devices. Meanwhile, $\mathrm{V}_{\mathrm{DS,sat}}$ for both InGaAs materials systems was weakly dependent on gate length and showed better current saturation.

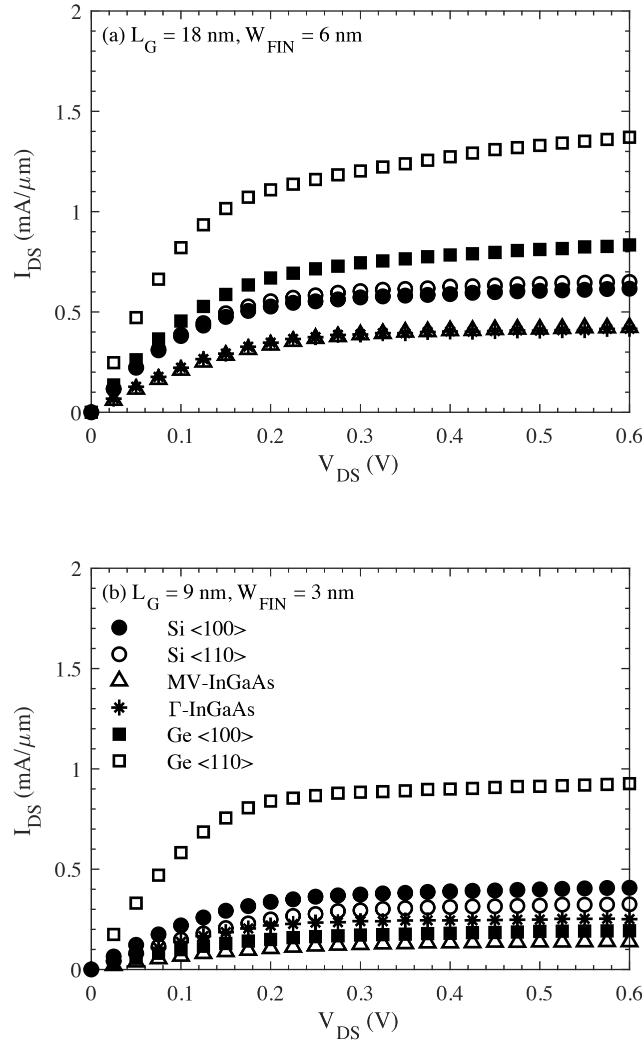

Figure 4.7: I_D-V_{DS} simulation results for Si ⟨100⟩ (solid circles), Si ⟨110⟩ (open circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), Γ-$In_{0.53}Ga_{0.47}As$ (asterisks), Ge ⟨100⟩ (solid squares), and Ge ⟨110⟩ (open squares) FinFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage for gate lengths (fin widths) of (a) L_G = 18 nm (W_{FIN} = 6 nm) and (b) L_G = 9 nm (W_{FIN} = 3 nm). Source and drain doping concentrations were taken to be 2×10^{20} cm^{-3} and 5×10^{19} cm^{-3} for Si or Ge and InGaAs devices, respectively.

4.6 Non-unity transmissivity contacts

As device dimensions are scaled down, electrical contact resistance comprises a significant fraction of R_{on}. In this section, the impact of reduced contact transmissivity on

peak g_m, turn-on characteristic ΔV_T, on-state current I_{on}, and the I_{DS} vs. V_{DS} characteristic of L_G = 18 nm and L_G = 9 nm FinFETs is re-examined. An illustrative control value of T = 0.20 was chosen, which, for Si, with ρ_{LB} = 3.0×10^{-10} Ω-cm^2 at the considered 2.0×10^{20} cm^{-3} doping concentration, corresponds to a specific contact resistivity of 1.35×10^{-9} Ω-cm^2, near a state-of-the-art reported value of 1.2×10^{-9} Ω-cm^2 [106]. For the Γ-InGaAs and MV-InGaAs devices with ρ_{LB} values of 1.3×10^{-9} Ω-cm^2 and 1.4×10^{-9}, respectively, at the considered 5.0×10^{19} cm^{-3} doping concentration, this control value of T results in substantially larger ρ_{sp} values, of 5.9×10^{-9} Ω-cm^2 and 6.3×10^{-9} Ω-cm^2, respectively. For the Ge devices with a ρ_{LB} value of 3.5×10^{-10} Ω-cm^2 at the considered 2.0×10^{20} cm^{-3} doping concentration, a control value of T = 0.20 corresponds to a specific contact resistivity of 1.6×10^{-9} Ω-cm^2. Nevertheless, these values of ρ_{sp} may be optimistic for InGaAs systems with reported specific resistivities of 7×10^{-9} Ω-cm^2 [107] and even more so for Ge systems with lowest specific resistivities reported of 6.8×10^{-8} Ω-cm^2 [118].

4.6.1 Peak g_{m}, ΔV_{T}, and I_{on}

Overall, Fig. 4.8 shows contact resistance decreased peak g_{m} and I_{on}, as expected. The relative effect of fixed contact resistivity is greatest for Γ-InGaAs and Ge ⟨110⟩ FinFETs, even considered optimistically so, and least so for Si ⟨110⟩ FinFETs. In fact, at L_G = 18 nm, the peak g_m advantage over Si devices by Γ-InGaAs devices assuming ideal contacts dissipates upon considering non-ideal contacts, owing to worse source starvation which now starts earlier. MV-InGaAs FinFETs perform the worst overall. Ge FinFETs continue to have the largest peak g_{m}, but only a slight advantage for Ge ⟨110⟩ channel devices over Ge ⟨100⟩ devices. At L_G = 9 nm, both orientations of Si devices now outperform all of their InGaAs and Ge ⟨100⟩ counterparts in terms of peak

g_{m} and even more so in terms of I_{on} because of the poorer turn-on characteristic in the latter material systems. However, there remains a distinct advantage for Ge $\langle 110 \rangle$ channel devices over all other materials systems because the quantum-confined band structure results in large channel capacitance (multiple valley degeneracy) and high injection velocity (light channel effective mass).

Initially (L_{G} = 18 nm) with ideal contacts, device performance is primarily source-limited via source starvation due to contact and surface orientation effects, but scaling of the fin width introduces narrow-slit injection issues, including the loss of the perfectly reflecting boundaries at the end of the source region, and the effect of source starvation saturates, whereas the effect of quantum confinement would continue to increase such that device performance is, or at least also is substantially, channel-limited via quantum-confinement. As the contact transmissivity is reduced, source injected carriers can reflect off the contact surfaces in the source and enter the channel. However, for electron injection probabilities that are peaked naturally about the surface normal direction, such as in Si $\langle 100 \rangle$ and Ge $\langle 110 \rangle$ channel devices, carriers possibly spend more time reflecting of the contacts and the likelihood of being absorbed at a contact surface increase. For these reasons, Si $\langle 110 \rangle$ channel devices help against source starvation and outperform Si $\langle 100 \rangle$ devices in terms of peak g_m at L_{G} = 9 nm. A similar contact-related advantage also exists for Ge $\langle 100 \rangle$ channel devices, but the differences in quantum confinement between $\langle 100 \rangle$ and $\langle 110 \rangle$ channel devices are more important.

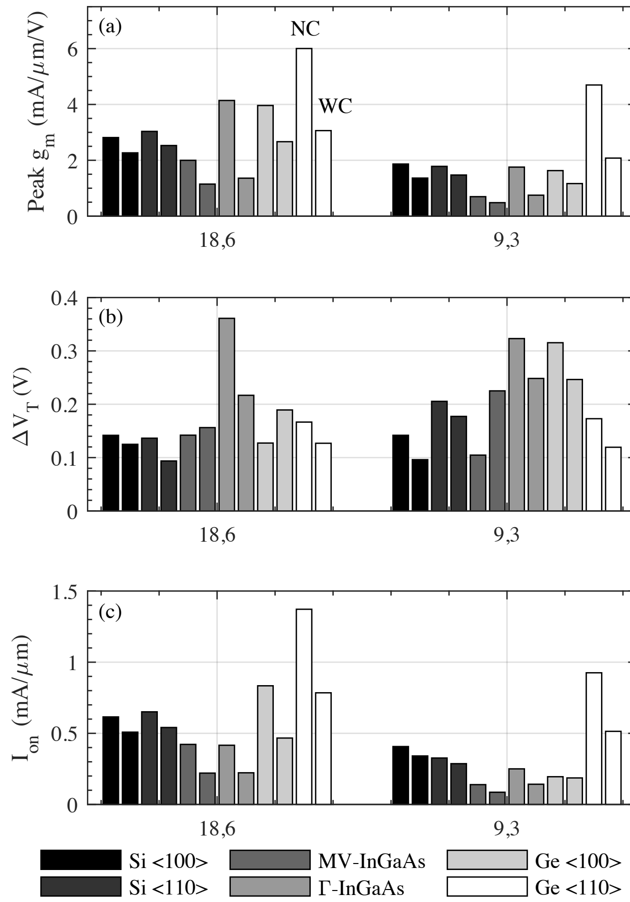

Figure 4.8: Comparison of contacts with perfect transmissivity and imperfect transmissivity contacts on the (a) peak of the transconductance g_m, (b) turn-on transition voltage ΔV_T, and (c) the on-current I_{on} with a constant current defined threshold (I_{on}(CC)), for the end, saddle/slot, and RSD contacts to an 18 nm and 9 nm gate length FinFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.2 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system, including channel orientation for Si and Ge.

4.6.2 Drain current vs. drain voltage

Fig. 4.9 shows drain current I_{DS} vs. drain voltage V_{DS} for the considered devices. For all materials systems, we found that the drain saturation voltage to achieve current

saturation was insensitive to contact transmissivity at the longest and shortest gate length devices studied because the voltage drop due to specific contact resistivity is not localized to the contact surface. Instead, modeling of specific contact resistivity in this limit corresponds to reducing the S/D injection efficiency and the primary limitation of InGaAs devices is the limited S/D doping leads to source starvation with realistic contacts.

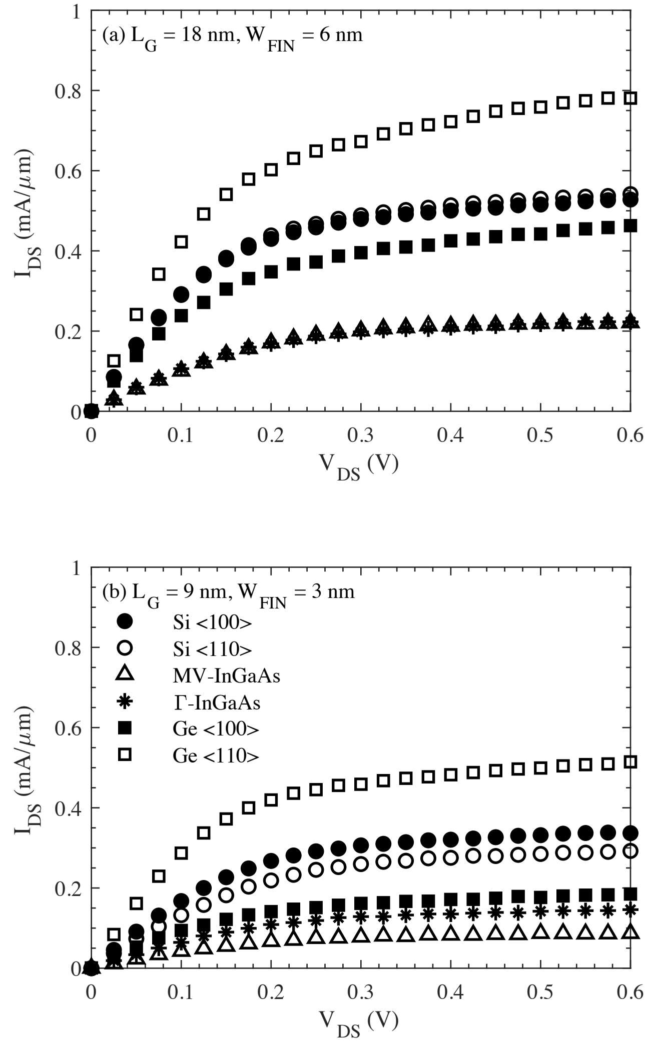

Figure 4.9: As for Fig. 4.7 but with non-ideal contacts, I_D-V_{DS} simulation results for Si ⟨100⟩ (solid circles), Si ⟨110⟩ (open circles), MV-$In_{0.53}Ga_{0.47}As$ (open triangles), Γ-$In_{0.53}Ga_{0.47}As$ (asterisks), Ge ⟨100⟩ (solid squares), and Ge ⟨110⟩ (open squares) FinFETs the gate overdrive voltage of 0.35 V above the constant current threshold voltage for gate lengths (fin widths) of (a) L_G = 18 nm (W_{FIN} = 6 nm) and (b) L_G = 9 nm (W_{FIN} = 3 nm), respectively. Source and drain doping concentrations were taken to be 2×10^{20} cm^{-3} and 5×10^{19} cm^{-3} for Si or Ge and InGaAs devices, respectively.

4.7 CONCLUSION

Silicon CMOS scaling is approaching limitations of fundamental physics, and thus, new materials and new structures have been proposed to reach the end-of-the-roadmap. For instance, high-mobility and high thermal velocity channel materials such as InGaAs and Ge present an alternative to Si channels. The FinFET design offers higher drive current per unit area per input voltage and tighter confinement for better electrostatic control than conventional planar MOSFETs. Unlike drift-diffusion and hydrodynamic simulations or quantum simulations, Monte Carlo methods provide a more detailed microscopic picture of carrier transport. The limitation and challenges of gate length scaling, channel orientation, and more reasonable contact transmissivities on Si, Ge, and InGaAs nanoscale n-channel FinFET performance with saddle/slot contacts are explored using a quantum-corrected semi-classical Monte Carlo method. Our simulation framework allows for quantum corrections to address quantum mechanical confinement, direct calculation of the occupation probabilities without assuming Fermi statistics to account for far-from-equilibrium degenerate carrier populations and to model the Pauli-blocking of scattering, and reduced contact transmissivity to model experimental contact resistivities. Silicon $\langle 110 \rangle$, Si $\langle 100 \rangle$, MV-InGaAs with conventionally reported energy valley offsets, idealized Γ-valley only InGaAs, Ge $\langle 110 \rangle$, and Ge $\langle 100 \rangle$ channel devices were modeled. A design rule for the gate length to fin width ratio of $L_{\mathrm{G}} \geq 3 \times W_{\mathrm{FIN}}$ was obtained for Si and Ge FinFETs such that S and DIBL lie below 70 mV/decade and 70 mV/V, respectively, compared to InGaAs FinFETs, which required longer channels to achieve these benchmarks.

We found that source starvation and quantum-confinement pose a challenge to gate length and fin width scaling in FinFETs. Although conduction in the light mass Γ-valley in InGaAs devices leads to high injection velocities, such carriers experience

greater quasi-ballistic transport, which leads to the earlier onset of source starvation, the choking of source injected carriers into the channel, compared to Si devices. In MV-InGaAs devices, quantum-confinement effects enable electrons to populate and conduct in heavier-mass peripheral valleys in the channel, and thus reducing both the injection velocity and efficiency and carrier concentration in the channel. Adding the effect of reduced contact transmissivity significantly degrades the performance of InGaAs devices even further due to reduced S/D injection efficiency. In Ge $\langle 100 \rangle$ devices, increased confinement pushes L-valley electrons into relatively heavier-mass Δ-valleys, and their advantage over Si devices vanishes at L_G = 9 nm even with ideal contacts. Ge $\langle 110 \rangle$ channel devices outperform all other devices in terms of g_m and I_{on} with and without reduced contact transmissivity due to high density of states, including multiple valley degeneracy, heavy confinement mass, and high injection velocity. Device performance in Si devices possess greater robustness against gate length scaling by mitigating undesirable quantum-confinement and source starvation side-effects, even using the former effect to remove valley degeneracy, and reduced contact transmissivity that typically hinder device performance in Ge or InGaAs devices. Both channel orientations in Si and Ge devices are expected to converge in the ray tracing limit from decreasing the fin width as devices become equally source starved and the loss of the perfectly reflecting boundary at the end of a source region but pick up an additional effect of quantum-confinement, which is exacerbated for $\langle 110 \rangle$ channels in Si and $\langle 100 \rangle$ channels in Ge. Our results provide deeper insights to the material options for scaled FinFETs, which had not been discussed in the literature previously. Our findings show the relative importance of source and drain contacts than any intrinsic channel advantage at these devices scales, and current drivability depends on the S/D injection efficiency, which can improved through better contact design.

Chapter 5: Semi-Classical Monte Carlo Study of the Impact of Tensile Strain on the Intrinsic Performance Limits of Monolayer MoS_2 n-channel MOSFETs

5.1 INTRODUCTION AND BACKGROUND

In order to overcome the bulk nature of silicon (Si), two-dimensional (2-D) channel materials offer greater immunity to short-channel effects [38], [119], [120]. Despite the extraordinary mobility of electrons in graphene, the gapless band structure restricts its employment in field effect transistor (FET) applications [121], [122]. Transition metal dichalcogenides (MX_2) are layered materials composed of a transition metal (M) layer sandwiched between two chalcogen (X) atomic layers, such as molybdenum disulfide, MoS_2, that possess outstanding electrical, optical, and mechanical properties. MoS_2 offers several attractive properties as a channel material in FETs, including a sizable band gap, lack of surface dangling bonds, ultra-thin body, and a high degree of mechanical flexibility. Molybdenum disulfide is not a direct replacement for silicon in high-speed, high-performance applications but present a departure from traditional roadmaps, with the potential to forge a new path of low-cost and high volume, ultra-low power, transparent, ultra-thin and ultra-light, flexible electronics. Potential applications include flexible displays, wearable electronics, smart fabrics, and sensing devices that can be rolled, stretched, folded, and bent without losing functionality [39]–[41]. Additionally, MoS_2 can be easily isolated and stacked with other layered materials, e.g. graphene and hexagonal boron nitride, to form flexible contacts and dielectrics [123]. Meanwhile, some of the challenges facing fabrication of MoS_2 devices include large scale and defect-free film growth with controllable thickness, environmental stability, mitigating substrate and dielectric interface effects, reducing specific contact resistivity, and difficulty doping [124].

For this work, the intrinsic, i.e., phonon-limited, performance limits, of monolayer MoS_2 n-channel metal-oxide-semiconductor field effect transistors (MOSFETs) as a function of peripheral valley energy, contact transmissivity, gate length, and type and amount of tensile strain has been studied using a semi-classical Monte Carlo (SCMC) method. Electron effective masses, non-parabolicity constants, and conduction band-edge energy offsets are extracted from density functional theory (DFT) calculations of the electronic band structures of strained MoS_2. Multi-valley MoS_2 with DFT-calculated energy valley offsets, and reference idealized K-valley-only MoS_2 (K-MoS_2) are considered. The idealized K-MoS_2 material system represents the possibility of substantially larger valley offsets than otherwise modeled here and provides a reference point for the effects of the higher-lying energy valleys. The bulk drift velocity versus electric field characteristics of MoS_2 are simulated. We found tensile strain enhances the bulk low-field electron mobility, primarily due to reduced K-valley effective mass, increases peak and saturation velocities, and leads to negative differential resistance (NDR) at high fields. 200 and 15 nm gate length MoS_2 channel MOSFETs are modeled, the former representative of long-channel experimental devices and the latter of ultimately desired nanodevices. Both perfect and imperfect transmissivity contacts are simulated. These MoS_2 channel MOSFETs were highly sensitive to non-ideal contact transmissivities, most so with strain, and to the band structure model, relatively insensitive to the amount of strain, and some orientation-related advantage for biaxial tensile strain. And while our results suggest more limited improvement in device performance, it may bear out the motivation for the use of tailored strain profiles to tune device characteristics.

Strain effects, whether intentional or unintentional, potentially can alter the electronic structure of MoS_2, leading to lowering of the electron effective mass and improvement in mobility [125], [126]. Various methods to strain MoS_2 has been demonstrated in a number of studies [45], [123], [127]–[131], including biaxial strains of up to 6% using pressure differences across a suspended MoS_2 membrane. A common method to apply biaxial strain to MoS_2 is via thermal coefficient of expansion mismatch [127], [129], with a maximum reported strain of 1% [127]. During fabrication, the lattice mismatch between a deposited gate oxide layer of HfO_2 and monolayer MoS_2 channel could introduce 0.3% to 0.6% tensile train [44], which results in a substantial enhancement in the carrier mobilities from 0.5 and 3 $cm^2\ V^{-1}\ s^{-1}$ [132], typically observed in bulk films, to 200 $cm^2\ V^{-1}\ s^{-1}$ [38]. Local biaxial straining has also been recently proposed to generate a funnel for excitons, useful for photovoltaics and photodetectors [133]. Uniaxial strain via bending is ubiquitous in flexible electronics applications where the two most common failure mechanisms are cracks in the gate dielectric and buckling delamination, limiting the applied strain to 0.8% [45]. Another aspect of the weakly bonded 2-D TMDs atomic layers is that they can be isolated and stacked with other layered semiconductors to construct a wide range of Van der Waals heterostructures without the limitation of lattice matching. For example, MoS_2 channel transistors with hexagonal boron nitride gate dielectrics can support larger amounts of strain than with conventional dielectrics with little performance degradation [123]. Nevertheless, these reported strain limits are far less than highly crystalline and defect-free monolayer MoS_2, which can sustain up to 11% strain [134], because the efficiency of transfer of strain from the substrate to the MoS_2 layer depends on the Young's modulus of the substrate.

Non-equilibrium Green's functions (NEGF-based) (full-quantum) ballistic simulations [119], [135], analytical ballistic models (top-of-the-barrier model) [136], or

compact models fit to experimental data [137] are commonly used to study the performance of MoS_2 transistors. However, electron-phonon scattering can be expected to continue to play an important role in carrier transport at feature sizes in targeted applications. The effects of uniaxial strain on polar optical phonon scattering [138] and full-quantum simulations with electron-phonon scattering [139] in MoS_2 transistors has been studied, but both of these studies neglected the effects of strain on the electronic structure. For this work, we simulated the phonon-limited device performance of strained monolayer MoS_2 using a SCMC method that allows for far-from-equilibrium degenerate statistics, non-ideal contacts, quasi-ballistic transport inaccessible through drift-diffusion and hydrodynamic simulations, and scattering mechanisms not readily accessible through NEGF simulations, among other things. Our results provide a greater understanding of the electronic structure of MoS_2 under strain, and associated MOSFET device physics.

5.2 SIMULATED MOSFET STRUCTURE AND BAND STRUCTURE MODELS

5.2.1 MOSFET structure

The 200 nm and 15 nm gate length (L_G) monolayer MoS_2 channel MOSFETs with end-injecting contacts and a 50 nm channel width (W_{CH}), sufficiently wide that edge effects are negligible, are shown in Fig. 5.1. The device geometry parameters are listed in Table 5.1. An electron mean free path of 15 to 22 nm for MoS_2 has been suggested [119], [140], such that phonon scattering is expected to occur at these device scales. In-plane end contacts have been reported to achieve a small contact resistance [141]. The width of these end contacts are equal to the width of the channel. For electrostatic modeling, we assumed a 3 nm thick HfO_2 ($\varepsilon_r = 22.3$) gate oxide, corresponding to an effective oxide thickness of 0.52 nm, a channel thickness (H_{CH}) of 6.5 Å, corresponding to the single layer thickness of a mechanically exfoliated MoS_2 flake using the scotch-tape method

[132], and an SiO_2 substrate thickness (H_{BOX}) of 10 nm. The source and drain (S/D) regions are uniformly doped to a sheet donor doping of 1×10^{13} cm^{-2}, which results in degenerate electron statistics with the Fermi level 11.6 meV above the conduction band edge under equilibrium conditions. Such degenerate doping concentrations has been reported to be possible for MoS_2 using potassium as an adatom dopant [142] and nearly achievable using chloride molecular doping [143]. Devices have a decade/nm doping profile in the 5 nm source and drain extensions.

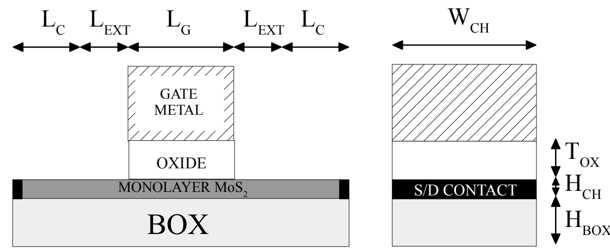

Figure 5.1: A side view (left) and an end view (right) of the simulated modeled MOSFET geometry. The spacers regions are not shown in order to show the underlying semiconductor fin, shaded in grey. The hatched region represents the gate metal. The gate oxide located underneath the gate metal is visible in the end views the saddle/slot contact model device. The source and drain contact surfaces are shown in black.

Dimension	**MoS_2 MOSFET**
L_c [nm]	8
L_{EXT} [nm]	5
L_G [nm]	200, 15
H_{CH} [nm]	0.65
W_{CH} [nm]	50
H_{BOX} [nm]	10
T_{OX} [nm]	3

Table 5.1: Modeled MOSFET dimensions.

5.2.2 MoS_2 band structure models

For our conventional multivalley MoS_2 model, we included two K-valleys, located at the corners of the hexagonal Brillouin zone, and six Q-valleys, located within the Brillouin zone along the same directions as the K-valleys, corresponding to the lowest and second lowest sets of band minima in the conduction band, respectively, as shown in Fig. 5.2. The energy separation between the K-valleys and Q-valleys, $\Delta E_{K\text{-}Q}$, is unsettled in the literature. Theoretical estimates range from 80 meV to 300 meV [144]–[148]. The only experimental evidence to date does not provide a separation but only places it at ≈ 60 meV [149]. However, these experimental results were performed on potassium intercalated MoS_2 to induce electron doping into the conduction band, which causes structural changes in MoS_2, which, in turn, is expected to alter the MoS_2 electronic properties [150]. Given such ambiguity, and possibly within our own results, we also considered a limiting K-valley-only MoS_2 model (K-MoS_2) with no satellite valleys ($\Delta E_{K\text{-}Q} \rightarrow \infty$).

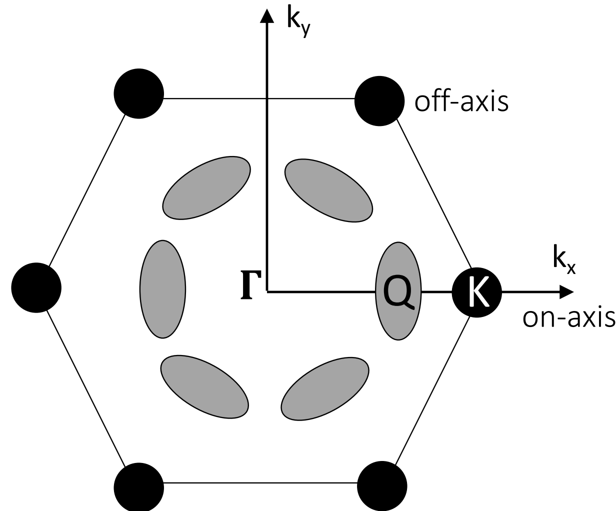

Figure 5.2: Alignment of the on-axis and off-axis K and Q conduction band valleys in the hexagonal Brillouin zone of monolayer MoS_2.

DFT simulations for this work were performed with the Vienna Ab-initio Simulation Package (VASP) [151], [152], using the Projector augmented-wave (PAW) pseudopotential formalism and Perdew-Burke-Ernzerhof (PBE) modification of the generalized gradient approximation (GGA) for approximating the exchange-correlation potential for Mo and S atoms [153], [154]. A 10 Å vacuum spacer along the z-axis was created to eliminate spurious interlayer interactions due to periodic boundary conditions. Illustrative (i) asymmetrical uniaxial tensile strain $\epsilon_{\mathrm{x}} \neq \epsilon_{\mathrm{y}}$, (ii) symmetrical biaxial tensile strain $\epsilon_{\mathrm{x}} = \epsilon_{\mathrm{y}}$, and (iii) positive symmetrical pure shear strain $\gamma_1 = -\gamma_2$ were modeled by deforming the unit cell along the *x*- and/or *y*-direction. Uniaxial tensile strain along the *x*-direction (*y*-direction), with associated compression along the *y*-direction (*x*-direction) according to Poisson's ratio, i.e. $\varepsilon_{\mathrm{y(x)}} = -\nu\varepsilon_{\mathrm{x(y)}}$, was considered. Here, the Poisson ratio was taken to be 0.27 [134]. Also, uniaxial tensile strain only along the *x*-direction (*y*-direction) is considered $\varepsilon_{\mathrm{y(x)}} = 0$, which models the situation in which a MoS_2 layer is strongly adhered to a mismatched substrate that stretches it only along one direction, but not along the other. For symmetrical biaxial tensile strain, the unit cell is stretched uniformly along both the *x*-and *y*-directions, $\varepsilon_{\mathrm{x}} = \varepsilon_{\mathrm{y}}$, which preserves the hexagonal symmetry of the lattice. Pure shear strain of $\gamma > 0$ corresponds to a rotation of the lattice vectors and modeled by decreasing the angle between in-pane lattice vectors. The atomic positions were allowed to relax while keeping the lattice vectors fixed until an energy and force convergence of 10^{-6} eV and 10^{-3} eV/Å, respectively, was reached for each of the strained structures. A Monkhorst-Pack Brillouin-zone grid of 6×6×1 k-points and a cut off energy of 500 eV was adopted for obtaining the relaxed structure. Subsequently, the optimized structures were used to carry out band structure calculations with an identical set of simulation parameters to extract valley energy separations, effective masses, and non-parabolicity constants.

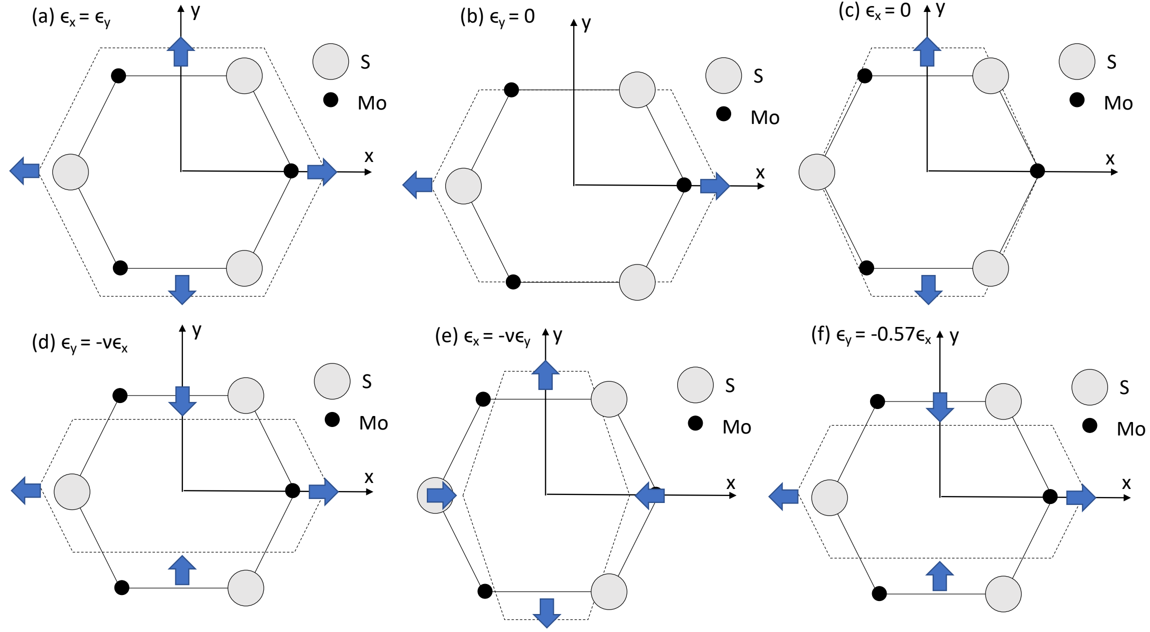

Figure 5.3: Sketch of (a) symmetrical biaxial tensile strain ($\epsilon_x = \epsilon_y$), (b) uniaxial tensile strain only along the x-direction ($\epsilon_y = 0$), (c) uniaxial tensile strain only along the *y*-direction ($\epsilon_x = 0$), (d) uniaxial tensile strain along the *x*-direction, with associated compression along the *y*-direction ($\epsilon_y = -\nu\epsilon_x$), (e) uniaxial tensile strain along the y-direction, with associated compression along the *x*-direction ($\epsilon_x = -\nu\epsilon_y$), and (f) pure shear strain (γ), which is equivalent to (d), but with a Poisson ratio of $\nu = 0.57$ and no change in the lattice constant.

Assuming a non-parabolic dispersion relation [50], the effective masses along the *x*-direction (m_x) and *y*-direction (m_y), and the non-parabolicity constants (α) for the K- and Q-valleys are extracted from the DFT-calculated electronic band structure. The allowable scattering mechanisms and corresponding phonon energies have been determined from first-principles calculations by Li et al. [147], and we use these in our simulations. However, the coupling constants for the various phonon modes have been adjusted to reproduce available experimental velocity-field data [155]. All simulation parameters for unstrained MoS_2, such as valley-specific effective masses, non-parabolicity constants, and deformation potentials are assembled in the Appendix.

A combination of theoretical studies and experimental data have identified Coulomb impurities arising from fixed ionized impurity charges at the bottom substrate, remote interfacial phonons from the oxide dielectric, traps, and defects as key performance bottlenecks to MoS_2 device performance. Coulomb scattering dominates at low temperatures and can be suppressed with high-*k* dielectrics [156], [157]. However, the high dielectric constant and soft polar phonon vibration mode in HfO_2 may lead to exacerbated surface polar optical phonon scattering. A thin buffer layer of parylene between the MoS2 channel and HfO_2 gate oxide can be used to reduce the surface polar phonon scattering from HfO_2 [158]. Atomically flat TMDs exhibit negligible surface scattering, in contrast to the severe surface scattering exhibited by Si. Thus, the focus of this work is the performance limits of unstrained and strained MoS_2 MOSFETs subject to intrinsic scattering, i.e., electron-phonon scattering. We adopted the deformation potentials and phonon energies from [147], while the former have been adjusted to reproduce available experimental data [155]. While strain induces shifts in the phonon modes [138], [159], we have assumed fixed deformation potentials and phonon energies as the strains considered here are small and even certain phonon mode energies are relatively insensitive to strain [160]. We have considered acoustic and optical intravalley and intervalley phonon scattering; contributions from other phonon modes have been found to be relatively modest due to weak coupling [161].

5.2.3 MOSFET simulation methodology (essential elements)

Our in-house University of Texas Semi-Classical Monte Carlo (UTMC) software [54] models carrier transport within complex device geometries and materials considering intra- and inter-valley phonon (acoustic, optical, and polar optical), surface roughness (SR), alloy, and ionized impurity scattering. The electron energy bands are modeled

analytically with non-parabolicity corrections, which is appropriate at these energy scales [50], [52]. For this work we have created a version of UTMC with 2D transport, while retaining the 3D electrostatics.

Because of high doping concentrations, we must consider degenerate statistics. However, because of the far-from-equilibrium conditions encountered in these devices, we cannot approximate the carrier statistics accurately using Fermi-Dirac distributions. Instead, we directly model Pauli-Blocking (PB) of scattering to obtain the far-from-equilibrium local electron occupation probabilities from the local electron populations, $N(r, E, g, \pm)$, as a function of position (r), energy valley (g) and energy (E), and propagation direction, forward toward the drain end (+) or backward toward the source end (−).

Contact non-ideality is simulated directly within our SCMC framework via a reduction below unity in the probability for an electron to be transmitted across the contact interface in either direction, T. Equal angle reflection is used to model carriers reaching, but not being transmitted across the contact interface from the inside. The resulting apparent specific contact resistivity is, $\rho_{\mathrm{sp}} = \rho_{\mathrm{LB}}(T^{-1} - 1/2)$, where ρ_{LB} is the Landauer-Büttiker ballistic resistivity [73], [74]. In this the electron energy loss associate the added contact resistivity is dropped gradually within the device on the scale of the carrier mean-free path, rather than being dropped outside of the device as within a lumped contact resistance model, which also avoids computationally burdensome post-processing of contact resistance effects associated with the latter model [162]. (Moreover, although the contact geometry is simple, here, effects of more complicated contact geometries also would be preserved in this way). In this work, we use a position and energy independent transmission probability for simplicity, but not as a limit of the method.

5.3 ELECTRONIC BAND STRUCTURE

We extracted pertinent band structure parameters of MoS_2 with 0% to 3% strain in steps of 1% strain using the GGA functional in DFT framework as detailed above. Effective masses were computed using central finite differences at the local minima of the K-valley and Q-valley. Non-parabolicity constants were obtained by fitting the valleys to analytic bands [50], [52]. Fig. 5.4 shows the change in the band gap E_g and the K-valley to Q-valley energy separation, $\Delta E_{K\text{-}Q}$. Fig. 5.5. shows the change in the electron effective mass along the x-direction (m_x) and y-direction (m_y) of the on-x-axis K and Q-valleys of MoS_2 subject to the here-considered strain profiles.

The optimized lattice constant and calculated direct band gap, with electron and valence band edges at the K-points, of unstrained monolayer MoS_2 was 3.14 Å and 1.8 eV, respectively, which is consistent with previous theoretical studies [147], [148] and close to the experimentally determined values [163] [164]. The inter-valley separation between the light-mass K-valleys and heavy-mass peripheral Q-valleys is $\Delta E_{K\text{-}Q} = 139$ meV. The extracted masses along the x- and y-directions of the K-valley are nearly identical, $m_x^{\mathrm{K}} = m_y^{\mathrm{K}} = 0.47 m_e$, where m_e denotes the electron rest mass, in agreement with previous first-principles calculations [165], [166]. These estimates are below the only experimental estimates reported to date of $m_{\mathrm{K}} = 0.67 m_{\mathrm{e}}$ [149], but the experimental results employed potassium intercalation, which may cause structural transformations. On the other hand, the Q-valleys are highly anisotropic with masses of $m_{\mathrm{x}}^{\mathrm{Q}} = 0.58 m_{\mathrm{e}}$ and $m_{\mathrm{y}}^{\mathrm{Q}} = 1.1 m_{\mathrm{e}}$.

After strain is applied, direct-to-indirect band gap and even semiconductor-to-metal transitions are induced, which is consistent with previous first-principles calculations [128], [167] and with photoluminescence measurements of the optical band gap [160]. The E_g is linearly reduced by 212 meV/%, 107 meV/%, 104 meV/%, and 70

meV/% for symmetrical biaxial tensile strain ($\epsilon_{\mathrm{x}} = \epsilon_{\mathrm{y}}$), pure shear strain ($\gamma$), uniaxial tensile strain only along the *x*-direction ($\epsilon_{\mathrm{y}} = 0$) or uniaxial tensile strain only along the *y*-direction ($\epsilon_{\mathrm{x}} = 0$), and uniaxial tensile strain along the *x*-direction (ϵ_{x}) or uniaxial tensile strain along the *y*-direction (ϵ_{y}), respectively. Our results are significantly larger than the reported shrinkage of the optical band gap at a rate of 100 meV/% [128] and 50 meV/% [168] with biaxial and uniaxial tensile strain, respectively. $\Delta E_{\mathrm{K\text{-}Q}}$ increases by 139 meV/%, 120 meV/%, 90 meV/%, 75 meV/%, 74 meV/%, and 54 meV/%, for biaxial tensile strain, pure shear strain, uniaxial tensile strain only along the *x*-direction, uniaxial tensile strain only along the *y*-direction, uniaxial tensile strain along the *x*-direction, and uniaxial tensile strain along the *y*-direction, respectively.

Non-parabolicity constants were found to be relatively insensitive to strain. Except for symmetrical biaxial tensile strain, strain breaks the hexagonal symmetry of the lattice, and consequently can remove the isotropy of the K-valley and warp the shape of the off-axis Q-valleys. The effective masses of the K-valley decrease with all forms of strain. For illustrative purposes, the effect of strain on the on-axis Q-valley effective masses were calculated. We found the change in $m_{\mathrm{x}}^{\mathrm{Q}}$ to be more mixed than the K-valley mass change, and all forms of strain raise $m_{\mathrm{y}}^{\mathrm{Q}}$. These trends are consistent with previous theoretical studies [169], [170]. Overall, the relative effect of strain on the effective masses of the K- and Q-valleys, $\Delta E_{\mathrm{K\text{-}Q}}$, and E_{g} depend on the type and amount of strain, within the ranges of strain considered. Strain along the y-direction (face-to-face of the hexagonal lattice) has a greater relative effect on distorting the lattice, e.g. increasing the lattice constant, increasing the Mo-S bond distance, and decreasing the S-Mo-S bond angle, decreasing $m_{\mathrm{x}}^{\mathrm{K}}$, and increasing $m_{\mathrm{y}}^{\mathrm{Q}}$, whereas strain along the x-direction (point-to-point of the hexagonal lattice) and shear strain has a greater relative effect on decreasing $m_{\mathrm{y}}^{\mathrm{K}}$, increasing $m_{\mathrm{x}}^{\mathrm{Q}}$, and increasing $\Delta E_{\mathrm{K\text{-}Q}}$. Based on these initial results, strain can

potentially be employed to improve electron transport in MoS_2 by reducing inter-valley transfer to heavier-mass Q-valleys via increased $\Delta E_{K\text{-}Q}$ and increasing carrier velocities and reducing scattering in the K-valleys via reduced K-valley effective mass.

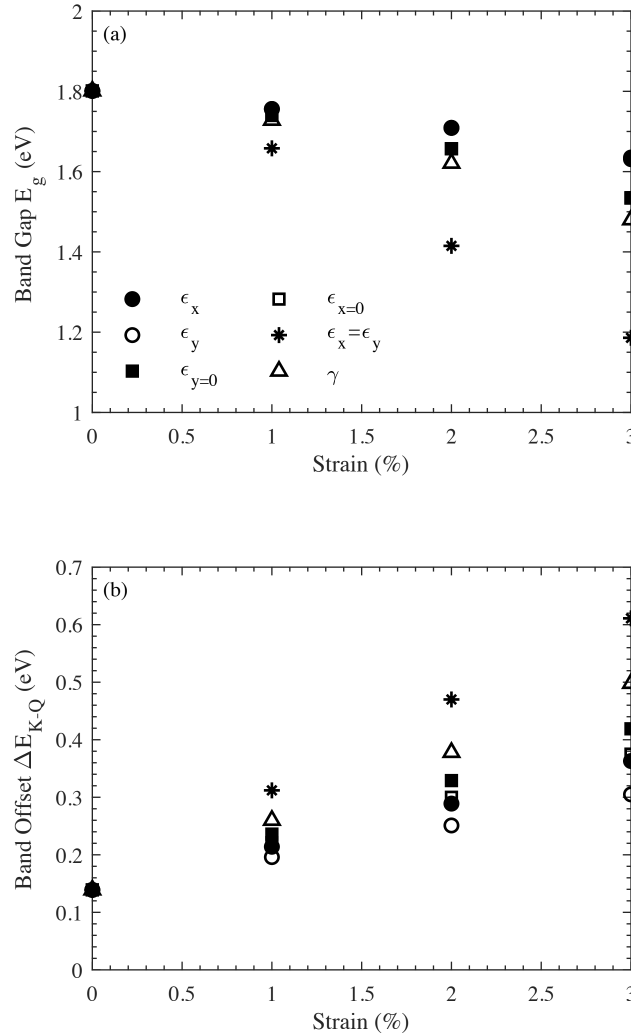

Figure 5.4: Dependence of (a) band gap E_g and (b) band edge valley separation of the of the K- and Q-valleys $\Delta E_{K\text{-}Q}$ under 0% to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (asterisks), uniaxial tensile strain along the *x*-direction ϵ_x (solid circles), uniaxial tensile strain along the *y*-direction ϵ_y (open circles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), and pure shear strain γ (open triangles).

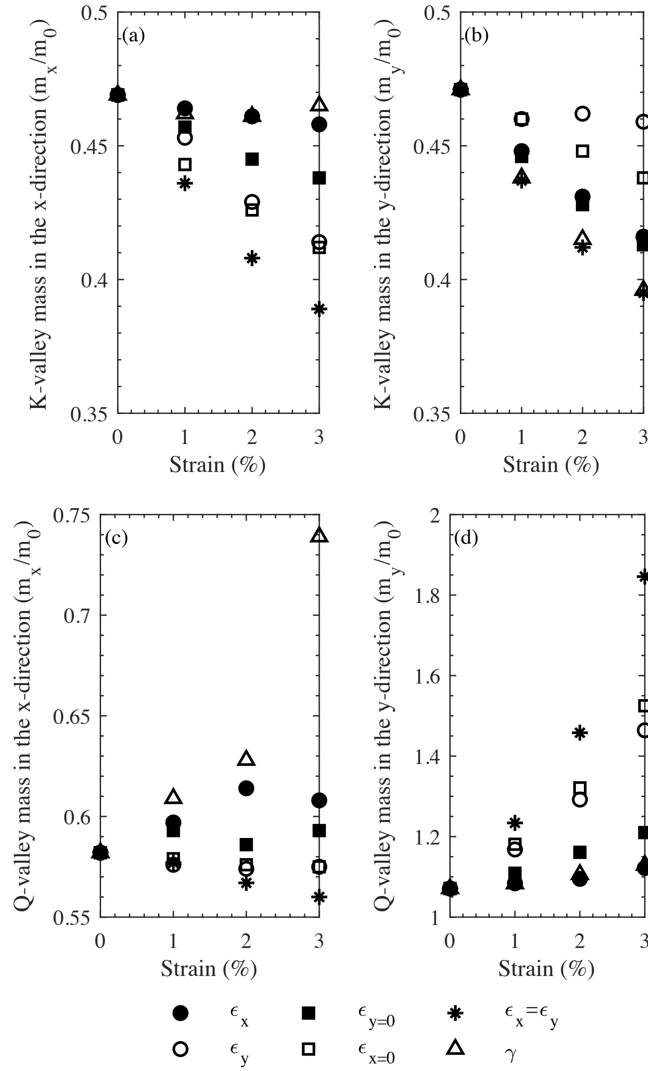

Figure 5.5: Dependence of (a) K-valley effective mass along the x-direction, (b) K-valley effective mass along the y-direction, (c) Q-valley effective mass along the *x*-direction, and (d) Q-valley effective mass along the *y*-direction under 0% to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (asterisks), uniaxial tensile strain along the x-direction ϵ_x (solid circles), uniaxial tensile strain along the *y*-direction ϵ_y (open circles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), and pure shear strain γ (open triangles).

5.4 Bulk drift velocity

While other types of strain were considered in DFT, biaxial tensile strain and uniaxial tensile strain only along the *x*- and *y*-directions not only had the largest effect on the inter-valley energy separation and effective mass of the K-valley, but also are most likely to occur physically, and are studied in further detail in these subsequent sections. For biaxial strain, the hexagonal symmetry of the lattice is preserved, and the off-axis K-valley and Q-valley effective masses, i.e. longitudinal and transverse effective mass, are changed according to DFT calculations of the on-axis values. On the other hand, for uniaxial strain, the off-axis K-valley masses are similarly changed according to on-axis DFT results, whereas both the on-axis and off-axis Q-valley effective masses are fixed to their unstrained values as a good approximation because strain has a much larger effect on increasing inter-valley band-edge separation and lowering the K-valley effective mass than changing the Q-valley mass. Moreover, the change in mass of Q-valley mass is small compared to the change in mass of the K-valley and the decrease in Q-valley occupancy due to larger inter-valley offset is expected to diminish their relative importance. Also, consistent with this discussion, mobility results of MoS_2 with biaxial strain at 1% and no mass change of the Q-valley were very similar to those provided here for MoS_2 with biaxial strain at 1% and Q-valley mass change.

UTMC is used to compute bulk charge carrier characteristics of MoS_2, including drift velocity (v_d) versus electric field (F), phonon-limited low-field electron mobility (μ_e), saturation velocity ($v_{d,sat}$), and peak velocity ($v_{d,peak}$). $\mu_e = \partial v_d / \partial F$ is calculated by centered moving average of the central finite differences of the drift velocity at low electric fields. $v_{d,sat}$ is evaluated at 100 kV/cm and $v_{d,peak}$ is the maximum drift velocity before velocity saturation. Figs. 5.6 and 5.7 show the v_d versus F curves and μ_e for 1% and 3% strain obtained from UTMC simulations at 300 K, respectively. The main

features of these results are that with increasing strain (i) μ_e increases, (ii) $v_{d,sat}$ increases, (iii) $v_{d,peak}$ increases, and (iv) negative differential resistance (NDR) behavior at high electric fields increases. Having been calibrated to reproduce experimental data, unstrained MoS_2 μ_e and $v_{d,sat}$ are 126 cm^2 V^{-1} s^{-1} and 3.6×10^6 cm/s, respectively, which is in agreement with theoretical estimates of μ_e and $v_{d,sat}$ of 130 cm^2 V^{-1} s^{-1} [147] and 3.4×10^6 to 4.8×10^6 cm/s [146], respectively. Considering only intra- and inter-valley scattering in the K-valleys, μ_e and $v_{d,sat}$ are 134 cm^2 V^{-1} s^{-1} and 5.0×10^6 cm/s, respectively, and the former mobility is only slightly larger than unstrained MoS_2, which illustrates the small effect of increasing $\Delta E_{K\text{-}Q}$ has on transport. Our calculated mobility is significantly lower than theoretically predicted phonon-limited mobility in K-MoS_2 of 320 cm^2 V^{-1} s^{-1} [147] and 410 cm^2 V^{-1} s^{-1} [166]. Predicted mobilities as large as 690 cm^2 V^{-1} s^{-1} have been reported elsewhere when inter-valley transfer to the Q-valleys is not considered [146]. This large difference between mobility predictions stems from our simulation approach to calibrate our unstrained model to reproduce experimental data. In realistic devices, defects such as charged-impurity scattering from sulfur vacancies, substrate screening, and effects of the dielectric environment such as remote phonon or surface optical phonon scattering operate at low-fields to further reduce mobility, and to compensate for the extra scattering mechanisms not accounted for in our model, the intrinsic electron-phonon coupling constants in both the K- and Q-valleys are increased.

With 1% strain, μ_e and $v_{d,sat}$ ranges from 142 to 168 cm^2 V^{-1} s^{-1} and 3.8×10^6 to 4.1×10^6 cm/s, respectively, which increases due to lighter K-valley effective mass and reduced inter-valley scattering via increased inter-valley separation, respectively. Uniaxial tensile strain along the *y*-direction show a greater enhancement in mobility than uniaxial tensile strain along the *x*-direction because the latter type of strain had a smaller change in m_x^K as compared to the former type of strain. As expected, μ_e is further

enhanced with 3% strain, ranging from 156 to 191 $cm^2 V^{-1} s^{-1}$, whereas significant NDR behavior with 3% strain causes the increase in $v_{d,sat}$ (4.2×10^6 to 4.4×10^6 cm/s) to be smaller relative to μ_e. NDR is caused by hot-electron transfer from the light-mass K-valleys into the heavier mass Q-valleys, which have much slower carriers and much higher scattering rates. Larger amounts of strain increase the relative energy separation between these two valleys and more electrons occupy the K-valleys; however, with increasing electric field, the population of the K-valleys decreases as carriers scatter into the Q-valley, which results in a more pronounced NDR. The electric field at which $v_{d,peak}$ occurs of 5×10^4 V/cm is somewhat unchanged with the type and amount of strain. From a bulk perspective, strained monolayer MoS_2 for n-channel MOSFETs should be beneficial.

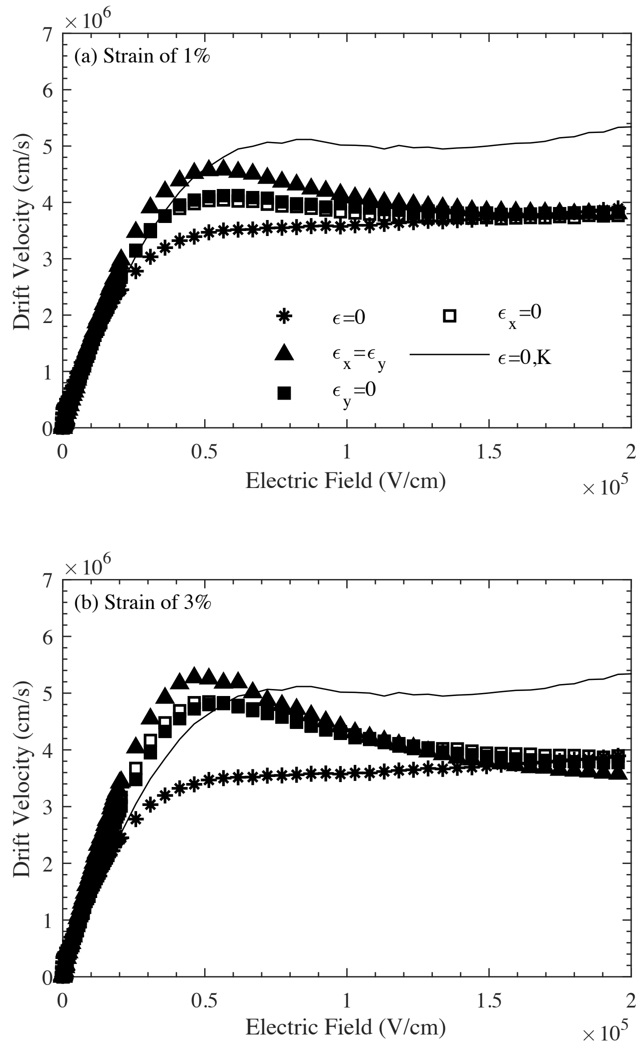

Figure 5.6: Drift velocity v_d vs. electric field F simulation results for monolayer MoS_2 at 300 K considering only phonon-limited electron transport subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line).

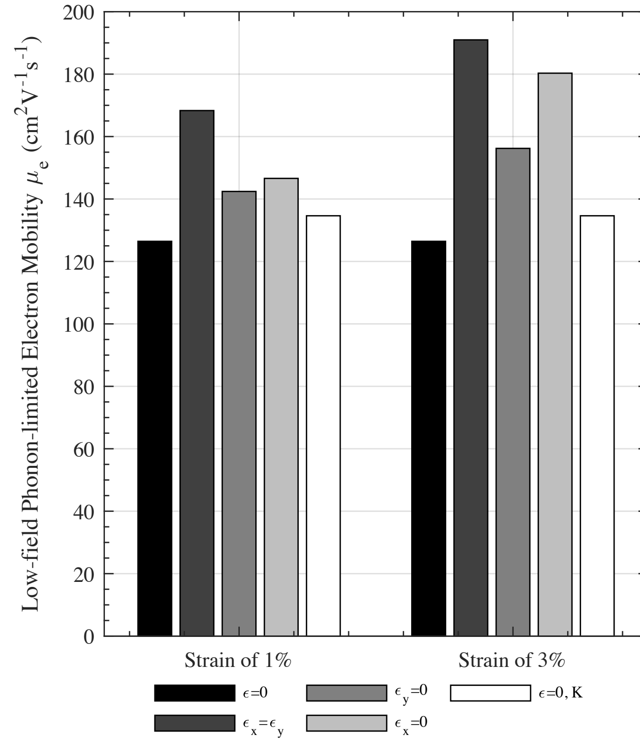

Figure 5.7: Dependence of (centered moving average of) low-field phonon-limited electron mobility with strain profile at strain amounts of 1% and 3%, including unstrained MoS_2 and K-valley-only MoS_2.

5.5 COMMON PERFORMANCE MEASURES AND RESULTS FOR UNITY TRANSMISSIVITY CONTACTS

UTMC simulations of L_G = 200 nm and 15 nm MoS_2 n-channel MOSFETs with 1% and 3% strain were performed. Simulation results are provided in Figs. 5.8, 5.9, and 5.10 and discussed in detail below.

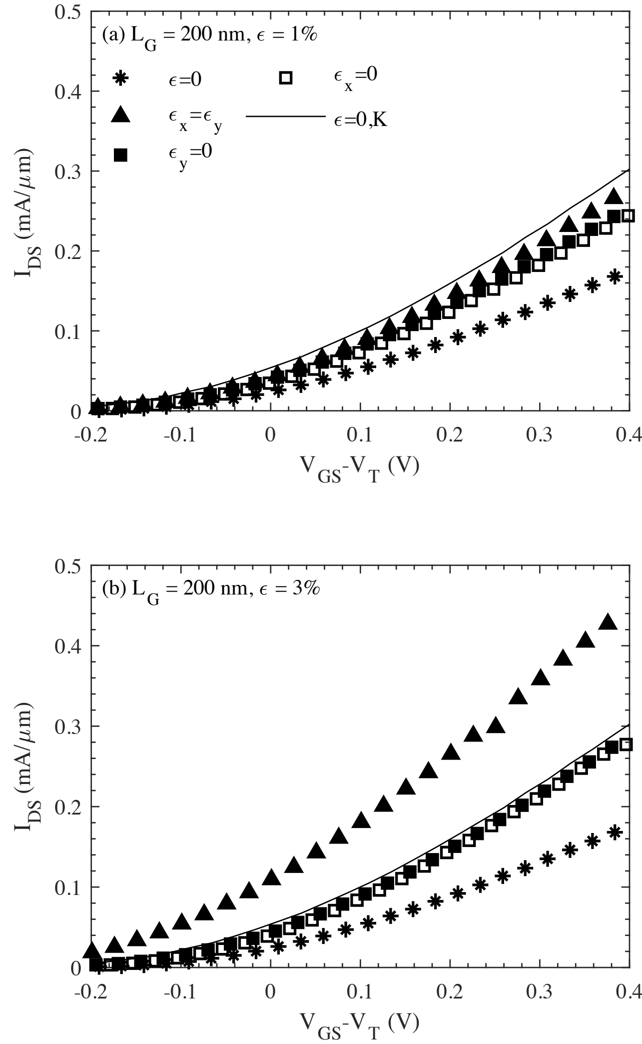

Figure 5.8: I_{DS}-V_{GS} simulation results for L_G = 200 nm monolayer MoS_2 MOSFETs subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method.

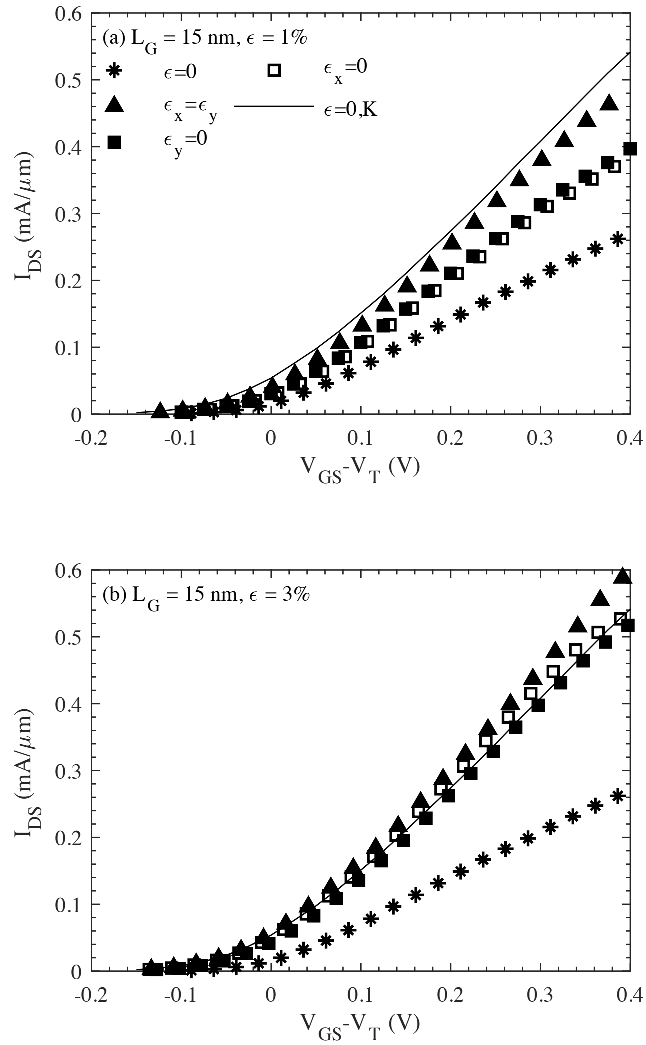

Figure 5.9: I_{DS}-V_{GS} simulation results for L_G = 15 nm monolayer MoS_2 MOSFETs subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line). V_{DS} = 0.6 V. For visual clarity with respect to transconductance, the threshold voltage is that obtained using the extrapolation in the linear regime method.

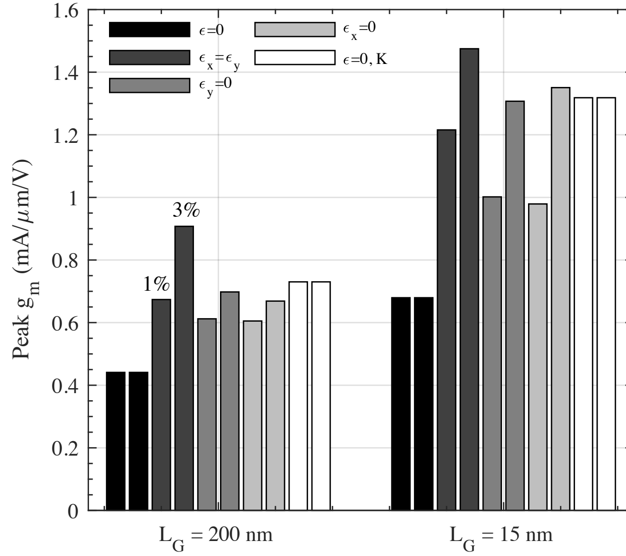

Figure 5.10: Comparison of strain on the (centered moving average of) the peak of the transconductance g_m for 200 nm and 15 nm gate length monolayer MoS_2 MOSFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to 1% and 3% strain, respectively, are shown side by side on the same gray scale for each considered strain profile, including unstrained MoS_2 and K-valley-only MoS_2.

5.5.1 Transconductance, g_m

At L_G = 200 nm, MoS_2 with 1% strain for all types of strain provides an enhancement over unstrained MoS_2 in terms of peak g_m; however, strained MoS_2 MOSFETs underperform unstrained K-valley-only MoS_2 MOSFETs owing to electrons transferring into the Q-valleys in the channel via inter-valley scattering even at $\Delta E_{K\text{-}Q}$ = 300 meV. Transconductance enhancement with biaxial strain is similar to bulk mobility simulations where initially carriers occupy the Q-valleys, but occupation of these valleys goes away beyond a certain amount of strain via reduced inter-valley transfer and what remains is the effect in the mass change of the K-valley. Similarly, increasing the amount of uniaxial tensile strain has a relatively modest effect on peak g_m at these device scales due the already large energy separation between the K- and Q-valleys and relatively weak

lowering of the K-valley mass, and continue to perform worse than unstrained K-MoS_2 devices in terms of transconductance.

At L_G = 15 nm, electron transport at these device scales is expected to be quasi-ballistic, and as a result, the increase in peak g_{m} with strain relative to unstrained devices is greater compared to 200 nm channel length devices due to the stronger effect of lowering the K-valley effective mass and less backscattering. For example**,** MoS_2 MOSFETs show about a 20% enhancement in peak g_{m} from 1% to 3% strain as we pick up a greater effect of lowering the effective mass of the K-valley, but, uniaxial tensile strain continues to underperform K-MoS_2 MOSFETs. Based on our results, the relative effect of strain primarily depends on the change of the K-valley effective mass and, to a lesser extent, on $\Delta E_{\text{K-Q}}$. The inconsistency of the latter value in the DFT calculations adds difficulty to accurately evaluating the role of strain, but if DFT is more accurate than experiment, than strain would have less impact than would otherwise be expected.

5.5.2 Drain current vs. drain voltage

The drain current also was calculated at the overdrive gate voltage of $V_{\mathrm{GS}} - V_{\mathrm{T}}^{\mathrm{CC}} = 0.35$ V as a function of drain voltage V_{DS} swept from 0 V to 0.6 V in steps of 25 mV, consistent with the transistor in the on-state with $V_{\mathrm{T}}^{\mathrm{CC}} = 0.25$ V and a $V_{\mathrm{DD}} = 0.6$ V, as shown in Figs. 5.11 and 5.12. At L_G = 200 nm, unstrained and strained MoS_2 MOSFET showed onset of current saturation at $V_{\mathrm{DS}} = V_{\mathrm{DS,sat}} = 0.35$ V, which is consistent with long-channel device behavior, in that, drain current saturation occurs when $V_{\mathrm{DS}} \geq V_{\mathrm{GS}} - V_{\mathrm{T}}$. For all devices, the current had little dependence on V_{DS} above $V_{\mathrm{DS,sat}}$. The onset of current saturation for L_G = 15 nm unstrained MoS_2 MOSFET decreases to $V_{\mathrm{DS,sat}}$ = 0.3 V, which points towards more quasi-ballistic transport than diffusive transport. For MoS_2 MOSFETs with 1% biaxial tensile strain, uniaxial tensile

strain only in the x-direction, and uniaxial tensile strain only in the *y*-direction, $V_{\mathrm{DS,sat}}$ = 0.4 V, 0.32 V, and 0.3 V, respectively, which also decreases compared to the results at L_G = 200 nm. Additionally, the current saturation above $V_{\mathrm{DS,sat}}$ is worse for the strained MoS_2 MOSFETs, especially with biaxial strain. With 3% strain, increasing V_{DS} above about 0.48 V and 0.42 V for biaxial and uniaxial tensile strain, results in NDR from inter-valley transfer between the lighter-mass K-valleys and heavier-mass Q-valleys.

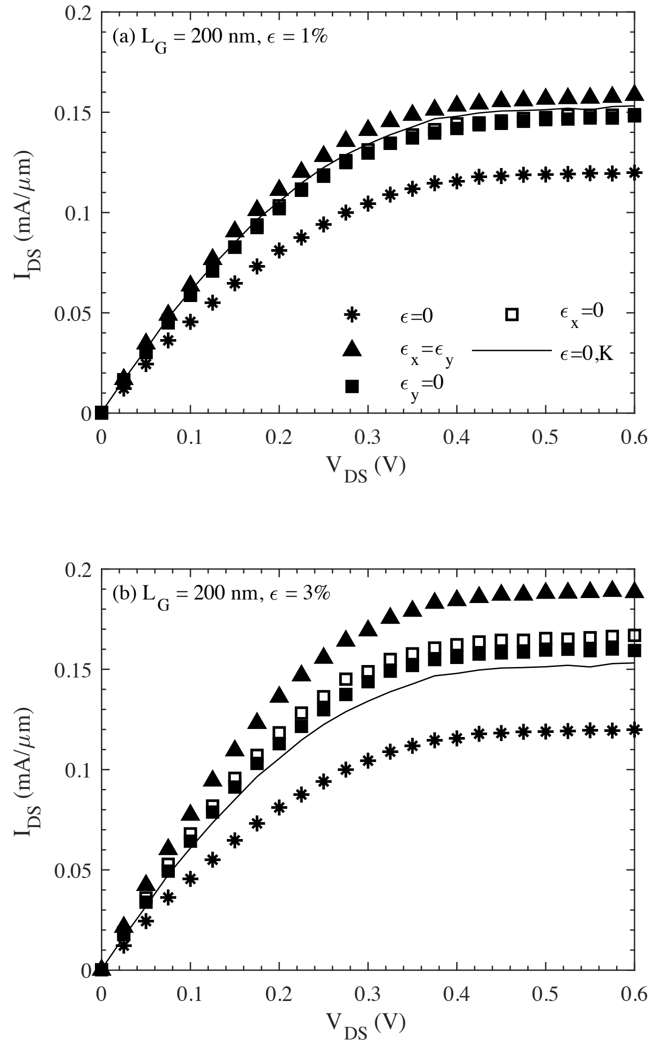

Figure 5.11: I_{DS}-V_{DS} simulation results for L_G = 200 nm monolayer MoS_2 MOSFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line).

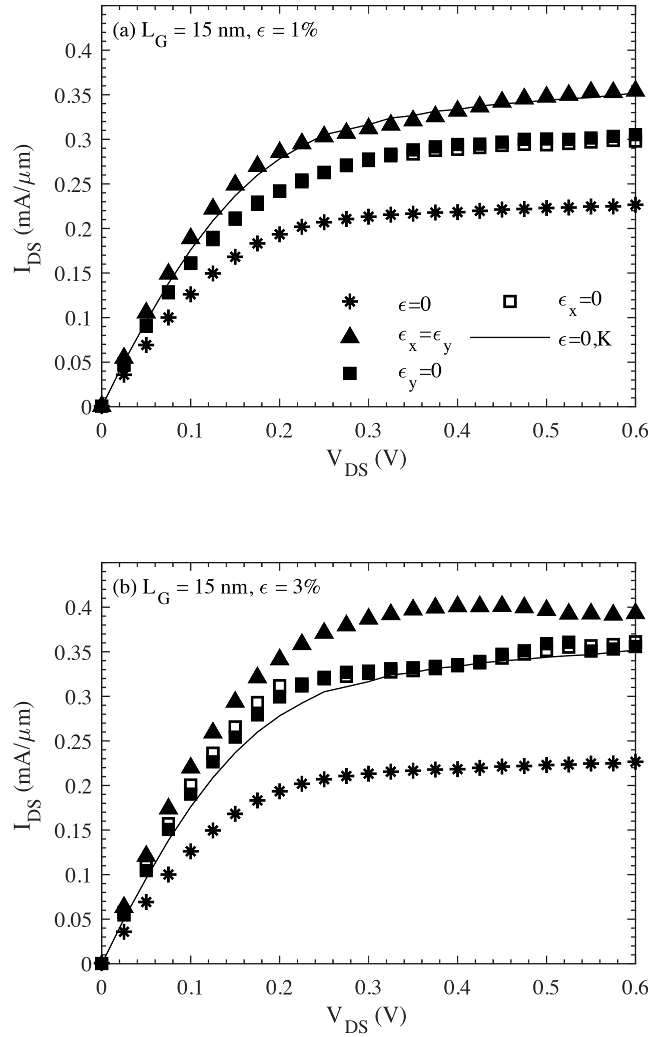

Figure 5.12: I_{DS}-V_{DS} simulation results for L_G = 15 nm monolayer MoS_2 MOSFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage subject to (a) 1% and (b) 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (solid triangles), uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$ (solid squares), and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ (open squares), including unstrained MoS_2 $\epsilon = 0$ (asterisks) and K-MoS_2 (solid line).

5.6 Non-unity transmissivity contacts

Parasitic specific contact resistivity between the semiconductor and contact interface remains a challenge in realizing high-performance MoS_2 MOSFETs. In this

section, the impact S/D electrical contact resistance on the performance of unstrained and 3% biaxially and uniaxially tensile strained MoS_2 MOSFETs is examined via sub-unity electron transmission probabilities across the contact surface. An illustrative control value of $T = 0.23$ was chosen, which, for MoS_2, with ρ_{LB} = 52.85 Ω-μm at the considered 1.5×10^{20} cm^{-3} doping concentration, corresponds to a specific contact resistivity of 200 Ω-μm, consistent with a reported state-of-the-art specific contact resistivity value using phase engineered contacts [171].

5.6.1 Peak g_m and drain current vs. drain voltage

Overall, Fig. 5.13 shows contact resistance decreased peak g_m as expected. The relative effect of specific contact resistivity is greatest for strained MoS_2 devices and least so for unstrained MoS_2 devices. The peak transconductance and on-current advantage over unstrained MV-MoS_2 devices by MoS_2 devices with 3% biaxial or uniaxial tensile strain assuming ideal contacts is largely reduced or entirely vanishes, respectively, upon considering non-ideal contacts at both 200 nm and 15 nm channel length devices. At L_G = 200 nm, carriers experience more bulk channel resistance and the S/D contacts have less of an effect, and biaxially strained MoS_2 devices continues to provide an improvement in peak g_m over its unstrained counterparts, but poorer turn-on behavior. As the channel length is scaled down to 15 nm, the relative contribution of contact resistance to the total device resistance grows, and strained MoS_2 devices are more sensitive to specific contact resistivity due to the smaller on-channel resistance than unstrained devices. As a result, the performance advantage in peak g_m by biaxially strained MoS_2 MOSFETs is largely reduced or somewhat vanishes over unstrained MV-MoS_2 and K-MoS_2 MOSFETs, respectively. While our results challenge the potential of strain in MoS_2 MOSFETs, it

does motivate the use of biaxial strain over uniaxial strain for boosting device performance.

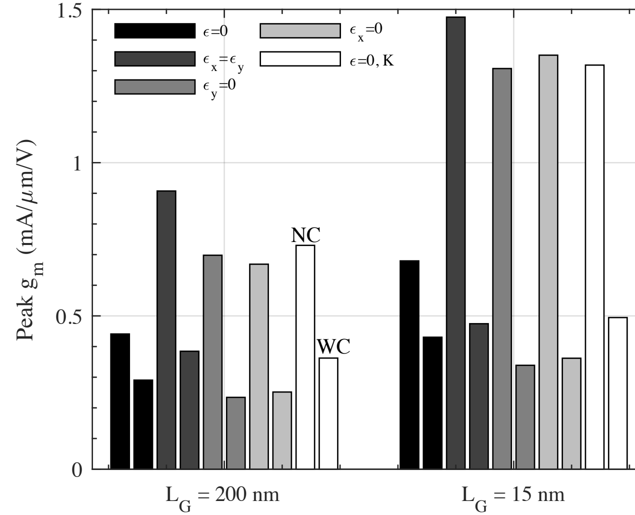

Figure 5.13: Comparison of 3% biaxial tensile strain $\epsilon_x = \epsilon_y$, uniaxial tensile strain only along the *x*-direction $\epsilon_y = 0$, and uniaxial tensile strain only along the *y*-direction $\epsilon_x = 0$ with perfect transmissivity and imperfect transmissivity contacts on the peak of the transconductance g_m for 200 nm and 15 nm gate length monolayer MoS_2 MOSFETs at V_{DS} of 0.6 V. Here, bar pairs corresponding to unity transmissivity (with no added specific contact resistivity) "NC" and to 0.23 transmissivity (with added specific contact resistivity) "WC", respectively, are shown side by side on the same gray scale for each considered material system, including unstrained MoS_2 $\epsilon = 0$ and K-valley-only MoS_2.

Fig. 5.14 shows drain current I_{DS} vs. drain voltage V_{DS} for L_G = 200 nm and L_G = 15 nm monolayer MoS_2 MOSFETs subject to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (dashed line) and unstrained MoS_2 $\epsilon = 0$ (solid line). We found that the drain saturation voltage to achieve current saturation was insensitive to contact transmissivity both for long channel and short channel length devices for these material systems.

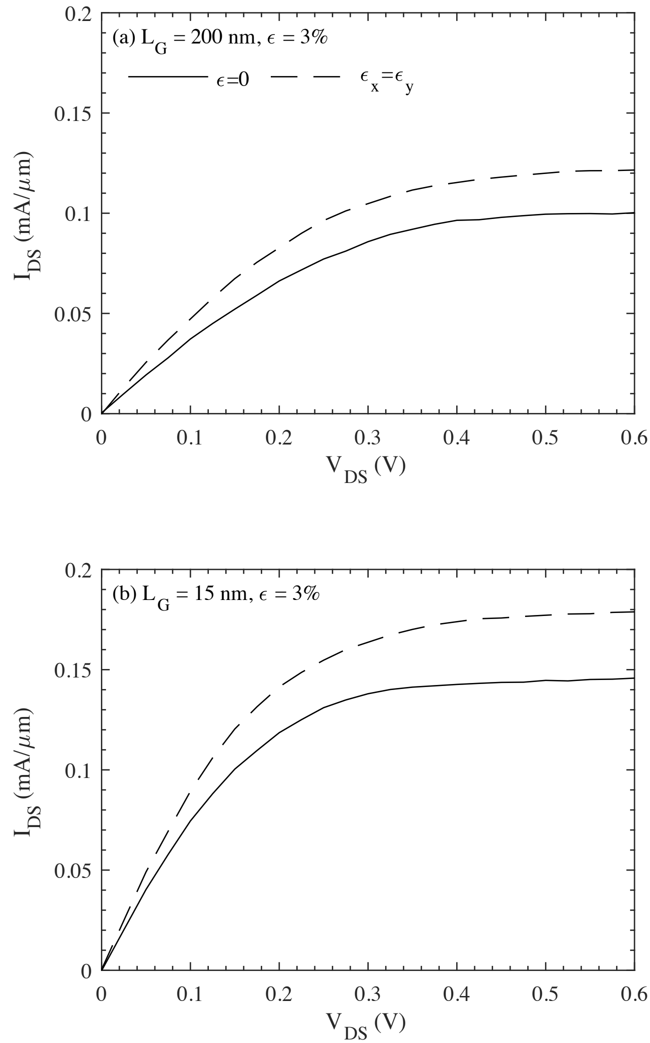

Figure 5.14: I_D-V_{DS} simulation results with perfect transmissivity and imperfect transmissivity contacts for (a) L_G = 200 nm and (b) L_G = 15 nm monolayer MoS_2 MOSFETs at the gate overdrive voltage of 0.35 V above the constant current threshold voltage subject to 3% biaxial tensile strain $\epsilon_x = \epsilon_y$ (dashed line) and no strain MoS_2 $\epsilon = 0$ (solid line).

5.7 Conclusion

Two-dimensional materials such as MoS_2 are undergoing rapid development for flexible, transparent, lightweight, low-cost, and ultra-low power applications due to their atomically thin body with sizeable band gaps, absence of dangling bonds, and mechanical

robustness. Unlike silicon, which typically breaks at strain levels of 1.5%, MoS_2 can sustain much larger strain levels and even opens up the possibility of time-dependent and local straining. Strains may arise in MoS_2 during fabrication due to differences between the thermal expansion coefficient of the film and substrate, or introduced during mechanical deformations due to folding, stretching, and bending. In this work, we study the intrinsic performance limits of MoS_2 n-channel MOSFETs using a semi-classical Monte Carlo method as a function of tensile strain, ideality of peripheral valley energy, and reduced contact transmissivity. Uniaxial tensile strain the *x*- or *y*-directions, symmetrical biaxial tensile strain in both *x*- and *y*-directions, and symmetrical pure shear strain are modeled. Band structure parameters, including valley effective masses, non-parabolicity constants, and band offsets are extracted from density functional theory calculations. The considered phonons and related coupling constants for unstrained MoS_2 have been calibrated to reproduce experimental data and are also used for strained MoS_2, which is reasonable given the applied strain is within linear limits. Among our findings, the electronic structure is highly sensitive to the type of strain and amount of strain applied due to changes in bond lengths and bond angles to modulate the coupling strengths of the Mo and S orbitals. Tensile strain can decrease the size of the band gap, including cause an indirect-to-direct gap transition, increase inter-valley offsets, and decrease K-valley effective masses. As a result, low-field mobility is substantially enhanced due to lighter K-valley effective mass and reduction in inter-valley scattering to heavier-mass Q-valleys. Simulated devices included 200 nm and 15 nm gate length MoS_2 n-channel MOSFETs with end contacts. Strain in MoS_2 channel devices can enhance the on-state transconductance and current; however, the relatively weak effect of strain on the K-valley effective mass causes the enhancement in performance to saturate. For instance, K-valley only MOSFETs, representative of the inconsistencies of the inter-

valley energy separation between the K- and Q-valleys reported in recent publications, outperformed otherwise identical strained devices within the entire range of strain values simulated, expect for MoS_2 with biaxial strain, which had strongest effect on the K-valley effective mass. With reduced contact transmissivity, the performance advantage of strained MoS_2 MOSFETs over unstrained MoS_2 MOSFETs is more limited, especially for short channel length devices.

Evaluating the practical role of strain is difficult to ascertain as details and magnitude of the predicted effect of strain depend on the estimation of $\Delta E_{K\text{-}Q}$. The choice of pseudopotential and exchange-correlation functional, [146], [172] within density functional theory calculations give different theoretical estimates of $\Delta E_{K\text{-}Q}$; however, we use the generalized gradient approximation to the exchange-correlation energy because it more closely reproduces experimental parameters such as the lattice constant and band gap. Ultimately, strain can alter the electronic structure of monolayer MoS_2 devices and presents both challenge and opportunities for device performance. Our theoretical simulations will help to interpret experiments and guide strain engineering in monolayer MoS_2 MOSFETs.

Chapter 6: Conclusion

6.1 Dissertation recap

Novel materials and device designs for end-of-the-roadmap CMOS and potential beyond CMOS applications have been considered, including the use high electron mobility and thermal velocity channel materials, FinFET device geometries, and emerging two-dimensional channel materials. I have shown the need for simulation, and, in particular, particle-based Monte Carlo methods, to understand the essential physics underlying the operation of Si, Ge, InGaAs, and MoS_2 n-channel field-effect transistors (FETs). The goal was not merely to reproduce experimental results, but also to yield physical insight on device operation by decomposing relevant device metrics into the contribution of different fundamental transport mechanisms. The results of this work will provide guidance to device designers, assist researchers in directing future research and development, and benchmarking of compact models. Additionally, this project will provide a basis for future modeling efforts made by our research group. This dissertation is organized in six chapters and one appendix, as follows.

Chapter 1 describes how scaling of Si FETs is reaching the limits of performance, which has spurred the development of novel channel materials and device designs for end-of-the-roadmap CMOS technology. High mobility and thermal velocity channel materials are being considered for high-performance complementary logic applications, for increased switching speeds, reduction in power dissipation density, and better gate control of the channel. In addition, two-dimensional materials have potential beyond-CMOS applications and are forging their own path, going where Si cannot follow, in the field of low-standby and operating power flexible electronics. Chapter 1 also compares various simulation approaches for modeling transport in semiconductors. For providing a

detailed picture of transport in deeply scaled FETs, the Monte Carlo method is a flexible, general, and powerful numerical approaching to solving the Boltzmann Transport Equation with room for quantum corrections for non-classical effects.

Chapter 2 summarizes some of the essential elements of our advanced ensemble semi-classical Monte Carlo transport simulator, UTMC, which address the effects of quantum-confinement, degenerate carrier populations, non-ideal contacts, and allows exploration of both conventional and un-conventional CMOS device geometries. Chapter 2 also addresses other critical details of UTMC implementation, such as the main Monte Carlo algorithm, our semi-empirical approach to modeling surface roughness, and modeling of the source and drain contacts.

Chapter 3 addresses the impact of contact geometry and transmissivity on quasi-ballistic nanoscale Si ⟨110⟩ and ⟨100⟩ and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs. The effects of contact geometry and specific contact resistivity on $In_{0.53}Ga_{0.47}As$ (InGaAs) and silicon (Si) nanoscale (18 nm channel length) n-channel FinFETs performance, and the effects of models thereof, are studied using a quantum-corrected semi-classical Monte Carlo method. Saddle/slot, raised source and drain (RSD), and reference end contacts are modeled. Both ideal perfectly injecting and absorbing contacts and those with more realistic specific contact resistivities are considered. Far-from-equilibrium degenerate statistics, quantum-confinement effects on carrier distributions in real-space and among energy valleys and on scattering, and quasi-ballistic transport are modeled. Silicon ⟨110⟩ channel and Si ⟨100⟩ channel FinFETs, multi-valley InGaAs channel FinFETs with conventionally-reported InGaAs energy valley offsets (MV-InGaAs), and reference idealized Γ-valley-only InGaAs (Γ-InGaAs) channel FinFETs are simulated. Among our findings, InGaAs channel FinFETs are highly sensitive to modeled contact geometry and specific contact resistivity and to the band structure model, while Si channel FinFETs

showed still significant but much less sensitivity to contact models. For example, for idealized unity transmissivity contacts, Γ-InGaAs channel FinFETs performed best for all contact geometries, at least in terms of transconductance, and end contacts provided the best performance for all considered channel materials. For realistic contact resistivities, however, results are essentially reversed. Silicon channel FinFETs performed best for all contact geometries, and saddle/slot and RSD contacts outperformed end contacts.

Chapter 4 addresses gate length scaling impact on quasi-ballistic nanoscale Si $\langle 110 \rangle$ and $\langle 100 \rangle$, Ge $\langle 110 \rangle$ and $\langle 100 \rangle$, and $In_{0.53}Ga_{0.47}As$ n-channel FinFETs. The effects of gate length scaling and specific contact resistivity on silicon (Si), $In_{0.53}Ga_{0.47}As$ (InGaAs), and germanium (Ge) nanoscale n-channel FinFETs performance are explored again using our quantum-corrected semi-classical Monte Carlo method. The saddle/slot contact geometry is assumed. We found InGaAs channel FinFETs performance to be most sensitive to gate length scaling, reduced contact transmissivity, and sensitive to the assumed peripheral energy valley offset. Quantum-confinement can eliminate otherwise expected benefits of light-mass Γ-valley electrons in MV-InGaAs devices over otherwise identical Si devices, despite higher injection velocities, due to increased occupation of heavier-mass satellite valleys in the channel and performed the poorest under all simulation scenarios. Without consideration of the peripheral valleys, illustrative of the uncertainties about peripheral valley energy offsets and degree of quantum-confinement, source starvation of Γ-valley electrons arises to depress transconductance. Ge offers greater channel quantum capacitance than InGaAs and a lighter conductivity effective mass than Si. However, the transconductance advantage over Si devices is limited for Ge $\langle 110 \rangle$ channel devices due to substantial degradation with reduced contact transmissivity, and the advantage vanishes for Ge $\langle 100 \rangle$ channel devices due to increased occupancy of heavier-mass Ge Δ-valleys via quantum-confinement with scaling. In contrast, simulated

Si devices exhibited relatively limited sensitivity to gate length scaling and more limited degradation in performance due to non-ideal contact transmissivities. FinFET performance can be divided into two regimes separated by a critical fin width of about 4 nm in our simulations, above which, device performance is source-limited via source starvation and below which, device performance is, or at least also is substantially, channel-limited via quantum-confinement.

Chapter 5 discusses the impact of tensile strain on the intrinsic performance limits of monolayer MoS_2 n-channel MOSFETs. The effects of tensile strain, peripheral valleys, and contact transmissivity on the intrinsic performance limits of monolayer molybdenum disulfide (MoS_2) nanoscale n-channel MOSFETs are studied using a semi-classical Monte Carlo method, that of the preceding chapters adapted for the 2D geometry considered here. Density functional theory calculations were performed to parametrize the electronic band structure of MoS_2 subject to tensile and shear strain. Tensile strain decreases the band gap, increases the inter-valley band-edge energy separation between the light-mass K-valleys and heavier-mass Q-valleys, and decreases the K-valley effective mass. These changes strongly depend on the direction and the amount of the applied strain. We found symmetrical biaxial tensile strain and uniaxial tensile strain only along the x- or y-directions to have the largest effect. Bulk drift velocity versus electric field simulation showed the low-field phonon-limited electron mobility is enhanced, peak and saturation drift velocities are increased, and high-field negative differential resistance is more pronounced with increasing strain. Both 200 and 15 nm gate length MoS_2 MOSFETs with end-contacts with perfect and reduced contact interface transmissivity were simulated. Simulated MoS_2 devices exhibited large sensitivity to specific contact resistivity, most so with strain, and to the band structure model, limited sensitivity to the amount of strain, and some direction-related advantage for biaxial tensile strain. Our

results elucidate the interplay between strain and electron transport in MoS_2 transistors and suggests that strain engineering may provide a pathway to improve electron mobility and boost device performance in MoS_2 MOSFETs.

The appendix contains the simulation and materials parameters for Ge and MoS_2 developed during the course of this dissertation work.

6.2 Recommendations for Future Work

Although the amount of experimental work on novel electronic devices has increased, there still exists a chasm between interpreting experimental results and microscopic mechanistic details of transport. Our work attempts to bridge the two together as the ability to predict a range of important physics at an unprecedented level of detail is making atomic-scale computational methods an invaluable research tool to gain unique insight into device physics in combination with experimental studies. Possible pathways to continue the work established by this dissertation include using the MC simulator to investigate emerging channel materials and device geometries and adding more physics to the existing MC simulator.

Going forward, it is suggested that similar studies be carried out for other material systems and device geometries to compare the merits and shortcomings of competing end-of-the-roadmap technologies. The role of electrical contact resistance also may deserve further study as this work has shown how contact resistance can comprise a significant fraction of the on-state resistance in ultra-scaled devices. Accurate simulation of off-state subthreshold leakage current is critical to technology projection and device design. Addition of rare-event enhancements would be beneficial because in the subthreshold thermionic emission of electrons from the tail of the quasi-equilibrium electron distribution in the source can overcome the energy barrier in the channel

producing leakage current, but, necessarily for good device performance, not enough to otherwise overcome sampling error substantially below threshold.

For continued studies of end-of-the-roadmap CMOS applications, MC simulation of strained silicon or silicon-germanium (Si_xGe_{1-x}) n-channel FinFETs would help further reveal the extent, if any, of the performance advantage Ge devices have over other competing material systems. Epitaxial growth of silicon on SiGe substrates is well-known to improve carrier mobilities by removing band degeneracy via strain. Silicon-germanium channels with their Si-like DOS and III-V-like conductivity effective mass could lead to optimized FinFET sidewall orientations that moderate quantum-confinement effects and provide large along-channel thermal velocities. A challenge of CMOS scaling is that it must be done for both n-channel and p-channel devices and understanding both electron and hole transport is important to the assessment alternative channel materials. Hole transport for the channel materials studied in this work and those proposed here can be implemented in MC software, as detailed subsequently. Even FinFETs are expected to hit scaling limits and therefore, device design is expected to become increasingly more important. Highly geometrically confined device structures such as gate-all-around nanosheets and nanowires may be studied to understand their performance limits for possibly extending the limits of CMOS technology.

Novel transistor concepts based on low-dimensional materials are currently being explored as potential replacements or extensions to CMOS technology. Through the work of this dissertation, the MC simulation software is capable of simulating transport in 2-D materials. 2-D crystals such as the graphene family (e.g. graphene, hexagonal boron nitride, boron and nitrogen co-doped graphene, fluorographene, graphene oxide), transition metal dichalcogenides (e.g. WS_2, $MoSe_2$, WSe_2), Xenes (e.g. silicene, germanene, phosphorene), MXenes (i.e. 2-D carbides and nitrides), and exotic

topological insulators (spin-dependent DOS and transport required) present rich physics to be studied. Moreover, Van der Waals heterostructures, which combine disparate materials together with novel hybrid properties, offer stackable platforms to build devices and can also be investigated. For this pursuit, band structures (shape of the energy bands, values of the effective masses, valley energy positions), phonon dispersion relations, and electron-phonon interaction potentials for these materials may be obtained using density functional theory calculations. Calibration of the scattering models are obtained by matching MC simulation of drift velocity versus electric field curves with measured data. While there has been an increasing number of theoretical studies of 2-D semiconducting materials, experimental studies have been sparse thus far, and calibration may be difficult. However, once the band structure is known and scattering coupling constants are verified, the Monte Carlo simulator can be calibrated to bulk transport properties, to then model device characteristics. In addition to 2-D material systems, the simulator could be further amended to handle 1-D or 0-D (where carriers merely change states by scattering) materials by updating appropriate energy, velocity, and density of states relationships.

Further improvements to UTMC include more comprehensive physical models and more efficient computer algorithms. To be able to better estimate the device performance of modern scaled MOSFETs, material models that incorporate all relevant physical mechanisms are required to capture the full microscopic picture of carrier transport. New physics such as hole transport, calculation of contact transmissivity, and exchange-correlation effects are some of the opportunities to be pursued. Monte Carlo transport of holes is identical to that of electrons; however, accurate modeling of holes using analytical bands is not as straightforward as that for electrons due to their warped, quartic ellipsoidal valleys. The complex band structure in some case perhaps can be

approximated by parabolic and spherical energy valleys for some purposes, but analytic models based on k-dot-p methods could provide a more accurate and general approach. Subsequent calculations of velocity field curves can be used to calibrate scattering strengths with available experimental data. In this work, we employed a fixed transmission probability to model the effects of specific contact resistivity. Further refinement of this approach could include modeling the potential barrier that is formed at the metal-semiconductor interface as a 1-D triangular barrier. One possibility, would be directly solving the 1-D Schrödinger equation normal to the interface coupled with the Poisson solution. Alternatively, a modified Wentzel-Kramers-Brillouin (WKB) approximation could be employed to more efficiently estimate the interface transmissivity as a function of energy [173], [174].

While advances in computer power have enabled femtosecond simulations, simulation timescales still remain a challenge for collecting good statistics with Monte Carlo. For instance, the ITRS high-performance target for I_{off} of 100 nA/μm is an order of magnitude smaller than what we can practically achieve with direct simulation, i.e., without rare event statistical enhancement. Statistical enhancement in Monte Carlo simulations are especially useful when the device behavior is governed by rare events such as device operation in the subthreshold regime. Two approaches to enhance statistics are population control and event-biasing. Population control techniques are based on the heuristic idea of splitting of the carriers entering a given phase space region of interest. Event-biasing techniques enrich the statistics by biasing the probabilities associated with the transport of classical carriers and apply a weight to the carriers to correct for the bias. To accurately simulate off-state behavior and compute I_{on}/I_{off} ratios, rare event statistical enhancement could be implemented.

Appendix

Ge Monte Carlo Simulation Parameters

Listed are the simulated band structure and scattering parameters for Ge including the lattice constant (a_0), mass density (ρ), speed of sound (v_s), relative dielectric permittivity (ε_r^0), electron affinity ($q\chi$), non-parabolicity constant (α), valley effective mass (m), acoustic deformation potential (Δ_{ac}), deformation field (DK), phonon energy ($\hbar\omega$), and inter-valley separation (E). Simulation parameters are consistent with previous Monte Carlo studies except for adjusting the values of the deformation potential.

Parameter	Ge	Units
a_0	5.658	Å
ρ	5.32	g/cm^3
v_s^l	5.4	$\times 10^5$ cm/s
v_s^t	3.2	$\times 10^5$ cm/s
ε_r^0	16.2	-
$q\chi$	4.00	eV
$E_{\Gamma L}$	0.135	eV
E_{LX}	0.173	eV
α_Γ	0.85	eV^{-1}
m_Γ	0.062	-
Δ_{ac}^Γ	6.25	eV
$DK_{(\Gamma \to L)}$	4.88	$\times 10^8$ eV/cm
$\hbar\omega_{(\Gamma \to L)}$	269	K
$DK_{(\Gamma \to X)}$	4.78	$\times 10^8$ eV/cm
$\hbar\omega_{(\Gamma \to X)}$	267	K
α_L	0.33	eV^{-1}
m_t^L	0.112	-
m_l^L	1.454	-
Δ_{ac}^L	6.25	eV
$DK_{(L \to X)}$	4.65	$\times 10^8$ eV/cm
$\hbar\omega_{(L \to X)}$	265	K
$DK_{(L \to L)}$	5.26	$\times 10^8$ eV/cm
$\hbar\omega_{(L \to L)}$	278	K
$DK_{(L \to L)}^{op}$	3.5	$\times 10^8$ eV/cm
$\hbar\omega_{(L \to L)}^{op}$	430	K
α_X	0.14	eV^{-1}
m_t^X	0.288	-

m_l^X	1.353	-
Δ_{ac}^X	6.25	eV
$DK_{(X\rightarrow X)}$	3.78	$\times 10^8$ eV/cm
$\hbar\omega_{(X\rightarrow X)}$	356	K

MoS_2 Monte Carlo Simulation Parameters

Listed are the simulated band structure and scattering parameters for unstrained single-layer MoS_2 including the lattice constant (a_0), mass density (ρ), speed of sound (v_s), relative dielectric permittivity (e_r^0), electron affinity ($q\chi$), non-parabolicity constant (α), valley effective mass (m), acoustic deformation potential (Δ_{ac}), deformation field (DK), phonon energy ($\hbar\omega$) and the corresponding phonon wave-vector is shown in parenthesis, and inter-valley separation (E). Simulation parameters are consistent with previous Monte Carlo studies except for adjusting the values of the deformation potential.

Parameter	MoS_2	Units
a_0	3.14	Å
ρ	3.1×10^{-7}	g/cm^2
v_s^l	6.6	$\times10^5$ cm/s
e_r^0	6.4	-
$q\chi$	4.3	eV
ΔE_{K-Q}	0.139	eV
α_K	0.5	eV^{-1}
m_K	0.47	-
$D_{ac}^{(K\to K)}$ (Γ)	8.1	eV
$D_{op}^{(K\to K)}$ (Γ)	5.8	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(K\to K)}$ (Γ)	49.5	meV
$D_{ac}^{(K\to K')}$ (K)	1.4	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(K\to K')}$ (K)	26.1	meV
$D_{op}^{(K\to K')}$ (K)	2	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(K\to K')}$ (K)	46.8	meV
$D_{ac}^{(K\to Q)}$ (Q)	1.40	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(K\to Q)}$ (Q)	20.7	meV
$D_{op}^{(K\to Q)}$ (Q)	2.85	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(K\to Q)}$ (Q)	48.1	meV
$D_{ac}^{(K\to Q)}$ (M)	6.6	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(K\to Q)}$ (M)	24.2	meV
$D_{op}^{(K\to Q)}$ (M)	8.4	$\times10^8$ eV/cm

$\hbar\omega_{op}^{(\mathrm{K}\to\mathrm{Q})}$ (M)	47.5	meV
α_{Q}	0.5	eV^{-1}
m_t^{Q}	1.07	-
m_l^{Q}	0.582	-
$D_{\mathrm{ac}}^{(\mathrm{Q}\to\mathrm{Q})}$ (Γ)	2.8	eV
$D_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (Γ)	7.1	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (Γ)	49.5	meV
$D_{\mathrm{ac}}^{(\mathrm{Q}\to\mathrm{Q})}$ (Q)	2.1	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(\mathrm{Q}\to\mathrm{Q})}$ (Q)	20.7	meV
$D_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (Q)	4.8	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (Q)	48.1	meV
$D_{\mathrm{ac}}^{(\mathrm{Q}\to\mathrm{Q})}$ (M)	2.0	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(\mathrm{Q}\to\mathrm{Q})}$ (M)	24.2	meV
$D_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (M)	4.0	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (M)	47.5	meV
$D_{\mathrm{ac}}^{(\mathrm{Q}\to\mathrm{Q})}$ (K)	4.8	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(\mathrm{Q}\to\mathrm{Q})}$ (K)	26.1	meV
$D_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (K)	6.5	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(\mathrm{Q}\to\mathrm{Q})}$ (K)	46.8	meV
$D_{\mathrm{ac}}^{(\mathrm{Q}\to\mathrm{K})}$ (Q)	2.25	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(\mathrm{Q}\to\mathrm{K})}$ (Q)	20.7	meV
$D_{op}^{(\mathrm{Q}\to\mathrm{K})}$ (Q)	3.6	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(\mathrm{Q}\to\mathrm{K})}$ (Q)	48.1	meV
$D_{\mathrm{ac}}^{(\mathrm{Q}\to K')}$ (M)	6.6	$\times10^8$ eV/cm
$\hbar\omega_{ac}^{(\mathrm{Q}\to K')}$ (M)	24.2	meV
$D_{op}^{(\mathrm{Q}\to K')}$ (M)	9.9	$\times10^8$ eV/cm
$\hbar\omega_{op}^{(\mathrm{Q}\to K')}$ (M)	47.5	meV

References

[1] "ITRS 2.0 Home Page," *International Technology Roadmap for Semiconductors*. [Online]. Available: http://www.itrs2.net/. [Accessed: 26-Feb-2018].

[2] S. Thompson, P. Packan, and M. Bohr, "MOS Scaling: Transistor Challenges for the 21st Century," p. 19.

[3] J. D. Meindl, "Limits on Silicon Nanoelectronics for Terascale Integration," *Science*, vol. 293, no. 5537, pp. 2044–2049, Sep. 2001.

[4] J. D. Plummer and P. B. Griffin, "Material and process limits in silicon VLSI technology," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 240–258, 2001.

[5] K. J. Kuhn, M. Y. Liu, and H. Kennel, "Technology options for 22nm and beyond," in *Junction Technology (IWJT), 2010 International Workshop on*, 2010, pp. 1–6.

[6] K. J. Kuhn, "Considerations for Ultimate CMOS Scaling," *IEEE Transactions on Electron Devices*, vol. 59, no. 7, pp. 1813–1828, Jul. 2012.

[7] M. S. Lundstrom, "On the mobility versus drain current relation for a nanoscale MOSFET," *IEEE Electron Device Lett.*, vol. 22, no. 6, pp. 293–295, Jun. 2001.

[8] K. Rim, "Scaling of strain-induced mobility enhancements in advanced CMOS technology," in *Solid-State and Integrated-Circuit Technology, 2008. ICSICT 2008. 9th International Conference on*, 2008, pp. 105–108.

[9] Y. Q. Wu *et al.*, "Atomic-layer-deposited Al2O3/GaAs metal-oxide-semiconductor field-effect transistor on Si substrate using aspect ratio trapping technique," *Applied Physics Letters*, vol. 93, no. 24, p. 242106, Dec. 2008.

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

[11] N. Mukherjee *et al.*, "MOVPE III–V material growth on silicon substrates and its comparison to MBE for future high performance and low power logic applications," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2011, pp. 35–1.

[12] R. Xie, T. H. Phung, W. He, M. Yu, and C. Zhu, "Interface-Engineered High-Mobility High-k/Ge pMOSFETs With 1-nm Equivalent Oxide Thickness," *IEEE Transactions on Electron Devices*, 2009.

[13] S. Takagi and M. Takenaka, "Advanced CMOS technologies using III-V/Ge channels," in *VLSI Technology, Systems and Applications (VLSI-TSA), 2011 International Symposium on*, 2011, pp. 1–2.

[14] F. Bellenger *et al.*, “High FET Performance for a Future CMOS GeO_2-Based Technology,” *IEEE Electron Device Letters*, vol. 31, no. 5, pp. 402–404, May 2010.

[15] M. E. Levinshtein and S. L. Rumyantsev, “SILICON (Si),” in *Handbook Series on Semiconductor Parameters*, 2 vols., WORLD SCIENTIFIC, 1996, pp. 1–32.

[16] D. Hisamoto *et al.*, “FinFET-a self-aligned double-gate MOSFET scalable to 20 nm,” *IEEE Transactions on Electron Devices*, vol. 47, no. 12, pp. 2320–2325, 2000.

[17] H.-S. Wong, D. J. Frank, and P. M. Solomon, “Device design considerations for double-gate, ground-plane, and single-gated ultra-thin SOI MOSFET’s at the 25 nm channel length generation,” in *Electron Devices Meeting, 1998. IEDM’98. Technical Digest., International*, 1998, pp. 407–410.

[18] B. S. Doyle *et al.*, “High performance fully-depleted tri-gate CMOS transistors,” *IEEE Electron Device Letters*, vol. 24, no. 4, pp. 263–265, Apr. 2003.

[19] Shiying Xiong and J. Bokor, “Sensitivity of double-gate and finfet devices to process variations,” *IEEE Transactions on Electron Devices*, vol. 50, no. 11, pp. 2255–2261, Nov. 2003.

[20] X. Huang *et al.*, “Sub 50-nm finfet: Pmos,” in *Electron Devices Meeting, 1999. IEDM’99. Technical Digest. International*, 1999, pp. 67–70.

[21] M. Guillorn *et al.*, “FinFET performance advantage at 22nm: An AC perspective,” in *2008 Symposium on VLSI Technology*, Honolulu, HI, USA, 2008, pp. 12–13.

[22] J. P. Campbell and P. M. Lenahan, “Density of states of Pb1 Si/SiO2 interface trap centers,” *Applied Physics Letters*, vol. 80, no. 11, pp. 1945–1947, Mar. 2002.

[23] L.-Å. Ragnarsson and P. Lundgren, “Electrical characterization of Pb centers in (100)Si–SiO2 structures: The influence of surface potential on passivation during post metallization anneal,” *Journal of Applied Physics*, vol. 88, no. 2, pp. 938–942, Jul. 2000.

[24] M. Yang *et al.*, “High performance CMOS fabricated on hybrid substrate with different crystal orientations,” in *Electron Devices Meeting, 2003. IEDM’03 Technical Digest. IEEE International*, 2003, pp. 18–7.

[25] Y.-J. Yang, W. S. Ho, C.-F. Huang, S. T. Chang, and C. W. Liu, “Electron mobility enhancement in strained-germanium n-channel metal-oxide-semiconductor field-effect transistors,” *Applied Physics Letters*, vol. 91, no. 10, p. 102103, Sep. 2007.

[26] Min Yang *et al.*, “Performance dependence of CMOS on silicon substrate orientation for ultrathin oxynitride and HfO $_2$ gate dielectrics,” *IEEE Electron Device Letters*, vol. 24, no. 5, pp. 339–341, May 2003.

[27] S. Z. Butler *et al.*, "Progress, Challenges, and Opportunities in Two-Dimensional Materials Beyond Graphene," *ACS Nano*, vol. 7, no. 4, pp. 2898–2926, Apr. 2013.

[28] D. Akinwande, N. Petrone, and J. Hone, "Two-dimensional flexible nanoelectronics," *Nature Communications*, vol. 5, p. 5678, Dec. 2014.

[29] A. Nathan *et al.*, "Flexible Electronics: The Next Ubiquitous Platform," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1486–1517, May 2012.

[30] M. Chhowalla and G. A. J. Amaratunga, "Thin films of fullerene-like MoS_2 nanoparticles with ultra-low friction and wear," vol. 407, p. 4, 2000.

[31] H. Wang *et al.*, "Integrated Circuits Based on Bilayer MoS_2 Transistors," *Nano Letters*, vol. 12, no. 9, pp. 4674–4680, Sep. 2012.

[32] S. Bertolazzi, D. Krasnozhon, and A. Kis, "Nonvolatile Memory Cells Based on MoS_2/Graphene Heterostructures," *ACS Nano*, vol. 7, no. 4, pp. 3246–3252, Apr. 2013.

[33] H. Li *et al.*, "Fabrication of Single- and Multilayer MoS_2 Film-Based Field-Effect Transistors for Sensing NO at Room Temperature," *Small*, vol. 8, no. 1, pp. 63–67, Jan. 2012.

[34] S. Wachter, D. K. Polyushkin, O. Bethge, and T. Mueller, "A microprocessor based on a two-dimensional semiconductor," *Nature Communications*, vol. 8, p. 14948, Apr. 2017.

[35] O. Lopez-Sanchez, D. Lembke, M. Kayci, A. Radenovic, and A. Kis, "Ultrasensitive photodetectors based on monolayer MoS_2," *Nature Nanotechnology*, vol. 8, no. 7, pp. 497–501, Jul. 2013.

[36] M. Chen *et al.*, "Nanoimprint-Assisted Shear Exfoliation (NASE) for Producing Multilayer MoS_2 Structures as Field-Effect Transistor Channel Arrays," *ACS Nano*, vol. 9, no. 9, pp. 8773–8785, Sep. 2015.

[37] A. Tarasov *et al.*, "Highly Uniform Trilayer Molybdenum Disulfide for Wafer-Scale Device Fabrication," *Advanced Functional Materials*, vol. 24, no. 40, pp. 6389–6400, Oct. 2014.

[38] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, "Single-layer MoS_2 transistors," *Nature Nanotechnology*, vol. 6, no. 3, pp. 147–150, Mar. 2011.

[39] Y. Yang, X. Yang, Y. Tan, and Q. Yuan, "Recent progress in flexible and wearable bio-electronics based on nanomaterials," *Nano Res.*, vol. 10, no. 5, pp. 1560–1583, May 2017.

[40] Y. J. Park *et al.*, "All MoS_2-Based Large Area, Skin-Attachable Active-Matrix Tactile Sensor," *ACS Nano*, vol. 13, no. 3, pp. 3023–3030, Mar. 2019.

[41] M. Park, Y. J. Park, X. Chen, Y.-K. Park, M.-S. Kim, and J.-H. Ahn, "MoS_2-Based Tactile Sensor for Electronic Skin Applications," *Adv. Mater.*, vol. 28, no. 13, pp. 2556–2562, Apr. 2016.

[42] H.-Y. Chang *et al.*, "High-Performance, Highly Bendable MoS_2 Transistors with High-K Dielectrics for Flexible Low-Power Systems," *ACS Nano*, vol. 7, no. 6, pp. 5446–5452, Jun. 2013.

[43] H.-Y. Chang *et al.*, "Large-Area Monolayer MoS_2 for Flexible Low-Power RF Nanoelectronics in the GHz Regime," *Advanced Materials*, vol. 28, no. 9, pp. 1818–1823, Mar. 2016.

[44] G. Plechinger, F.-X. Schrettenbrunner, J. Eroms, D. Weiss, C. Schüller, and T. Korn, "Low-temperature photoluminescence of oxide-covered single-layer MoS_2," *physica status solidi (RRL) - Rapid Research Letters*, vol. 6, no. 3, pp. 126–128, Mar. 2012.

[45] C. R. Zhu *et al.*, "Strain tuning of optical emission energy and polarization in monolayer and bilayer MoS_2," *Physical Review B*, vol. 88, no. 12, Sep. 2013.

[46] Y. Chai, S. Su, D. Yan, M. Ozkan, R. Lake, and C. S. Ozkan, "Strain Gated Bilayer Molybdenum Disulfide Field Effect Transistor with Edge Contacts," *Sci Rep*, vol. 7, no. 1, p. 41593, Feb. 2017.

[47] J. W. Christopher *et al.*, "Monolayer MoS_2 Strained to 1.3% With a Microelectromechanical System," *J. Microelectromech. Syst.*, vol. 28, no. 2, pp. 254–263, Apr. 2019.

[48] B. M. S, G. Y. A, and M. M. P, *Handbook Series On Semiconductor Parameters, Vol. 2: Ternary And Quaternary Iii-v Compounds*. World Scientific, 1996.

[49] T. Kanazawa *et al.*, "Few-layer HfS_2 transistors," *Scientific Reports*, vol. 6, no. 1, Apr. 2016.

[50] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys.*, vol. 55, no. 3, pp. 645–705, Jul. 1983.

[51] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, "Electron drift velocity in silicon," *Phys. Rev. B*, vol. 12, no. 6, pp. 2265–2284, Sep. 1975.

[52] M. V. Fischetti, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. I. Homogeneous transport," *IEEE Transactions on Electron Devices*, vol. 38, no. 3, pp. 634–649, 1991.

[53] C. Moglestue, "Monte Carlo particle simulation of the hole-electron plasma formed in a p-n junction," *Electron. Lett.*, vol. 22, no. 7, p. 397, 1986.

[54] D. M. Crum, A. Valsaraj, J. K. David, L. F. Register, and S. K. Banerjee, “Methods for modeling non-equilibrium degenerate statistics and quantum-confined scattering in 3D ensemble Monte Carlo transport simulations,” *Journal of Applied Physics*, vol. 120, no. 22, p. 224301, Dec. 2016.

[55] D. K. Ferry, “Effective potentials and the onset of quantization in ultrasmall MOSFETs,” *Superlattices and Microstructures*, vol. 28, no. 5–6, pp. 419–423, Nov. 2000.

[56] H. Brooks, “Scattering by ionized impurities in semiconductors,” in *Physical Review*, 1951, vol. 83, pp. 879–879.

[57] C. Jacoboni, F. Nava, C. Canali, and G. Ottaviani, “Electron drift velocity and diffusivity in germanium,” *Physical Review B*, vol. 24, no. 2, p. 1014, 1981.

[58] V. Balynas, A. Krotkus, A. Stalnionis, A. T. Gorelionok, N. M. Shmidt, and J. A. Tellefsen, “Time-resolved, hot-electron conductivity measurement using an electro-optic sampling technique,” *Applied Physics A Solids and Surfaces*, vol. 51, no. 4, pp. 357–360, Oct. 1990.

[59] K. Akarvardar *et al.*, “Impact of Fin Doping and Gate Stack on FinFET (110) and (100) Electron and Hole Mobilities,” *IEEE Electron Device Lett.*, vol. 33, no. 3, pp. 351–353, Mar. 2012.

[60] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, “On the universality of inversion layer mobility in Si MOSFET’s: Part I-effects of substrate impurity concentration,” *IEEE Transactions on Electron Devices*, vol. 41, no. 12, pp. 2357–2362, Dec. 1994.

[61] C. D. Young *et al.*, “(110) and (100) Sidewall-oriented FinFETs: A performance and reliability investigation,” *Solid-State Electronics*, vol. 78, pp. 2–10, Dec. 2012.

[62] C. D. Young *et al.*, “Critical discussion on (100) and (110) orientation dependent transport: nMOS planar and FinFET,” in *2011 Symposium on VLSI Technology - Digest of Technical Papers*, 2011, pp. 18–19.

[63] O. Badami, F. Driussi, P. Palestri, L. Selmi, and D. Esseni, “Performance comparison for FinFETs, nanowire and stacked nanowires FETs: Focus on the influence of surface roughness and thermal effects,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2017, pp. 13.2.1-13.2.4.

[64] E. P. Gusev *et al.*, “Ultrathin high-K gate stacks for advanced CMOS devices,” in *Electron Devices Meeting, 2001. IEDM’01. Technical Digest. International*, 2001, pp. 20–1.

[65] R. E. Prange and T.-W. Nee, “Quantum spectroscopy of the low-field oscillations in the surface impedance,” *Physical Review*, vol. 168, no. 3, p. 779, 1968.

[66] Seonghoon Jin, M. V. Fischetti, and Ting-Wei Tang, “Modeling of Surface-Roughness Scattering in Ultrathin-Body SOI MOSFETs,” *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2191–2203, Sep. 2007.

[67] S. Yamakawa *et al.*, “Study of interface roughness dependence of electron mobility in Si inversion layers using the Monte Carlo method,” *Journal of Applied Physics*, vol. 79, no. 2, p. 911, 1996.

[68] D. Nikonov, H. Pal, and G. Bourianoff, *Scattering in NEGF: Made simple*. Nov, 2009.

[69] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, “Surface roughness at the Si (100)-SiO 2 interface,” *Physical Review B*, vol. 32, no. 12, p. 8171, 1985.

[70] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, “On the universality of inversion layer mobility in Si MOSFET’s: Part II-effects of surface orientation,” *IEEE Transactions on Electron Devices*, vol. 41, no. 12, pp. 2363–2368, 1994.

[71] C. J. Petti, J. P. McVittie, and J. D. Plummer, “Characterization of surface mobility on the sidewalls of dry-etched trenches,” in *Technical Digest., International Electron Devices Meeting*, 1988, pp. 104–107.

[72] S.-D. Kim, C.-M. Park, and J. C. Woo, “Advanced model and analysis of series resistance for CMOS scaling into nanometer regime. II. Quantitative analysis,” *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 467–472, 2002.

[73] R. Landauer, “Spatial variation of currents and fields due to localized scatterers in metallic conduction (and comment),” *Journal of Mathematical Physics*, vol. 37, no. 10, pp. 5259–5268, Oct. 1996.

[74] M. Büttiker, “Four-Terminal Phase-Coherent Conductance,” *Physical Review Letters*, vol. 57, no. 14, pp. 1761–1764, Oct. 1986.

[75] R. Kim, U. E. Avci, and I. A. Young, “Computational Study of Geometrical Designs for Source/Drain Contacts to Reduce Parasitic Resistance in Extremely Scaled MOSFETs,” *IEEE Transactions on Electron Devices*, vol. 66, no. 3, pp. 1189–1196, Mar. 2019.

[76] *Sentaurus Device Monte Carlo User Guide*. Mountain View, CA, USA: Synopsys, Inc., 2015.

[77] J. Kavalieros *et al.*, “Tri-Gate Transistor Architecture with High-k Gate Dielectrics, Metal Gates and Strain Engineering,” in *2006 Symposium on VLSI Technology, 2006. Digest of Technical Papers.*, Honolulu, HI, USA, 2006, pp. 50–51.

[78] C.-H. Jan *et al.*, “A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high

density SoC applications," in *2012 International Electron Devices Meeting*, San Francisco, CA, USA, 2012, pp. 3.1.1-3.1.4.

[79] M. Saitoh, Y. Nakabayashi, K. Uchida, and T. Numata, "Short-Channel Performance Improvement by Raised Source/Drain Extensions With Thin Spacers in Trigate Silicon Nanowire MOSFETs," *IEEE Electron Device Letters*, vol. 32, no. 3, pp. 273–275, Mar. 2011.

[80] J.-P. Colinge, *FinFETs and Other Multi-Gate Transistors*. Springer Science & Business Media, 2008.

[81] J. Kedzierski *et al.*, "Extension and source/drain design for high-performance finFET devices," *IEEE Transactions on Electron Devices*, vol. 50, no. 4, pp. 952–958, Apr. 2003.

[82] T. P. O'Regan and P. K. Hurley, "Calculation of the capacitance-voltage characteristic of GaAs, $In_{0.53}$ $Ga_{0.47}As$, and InAs metal-oxide-semiconductor structures," *Applied Physics Letters*, vol. 99, no. 16, p. 163502, Oct. 2011.

[83] W. Porod and D. K. Ferry, "Modification of the virtual-crystal approximation for ternary III-V compounds," *Physical Review B*, vol. 27, no. 4, p. 2587, 1983.

[84] K. Y. Cheng, A. Y. Cho, S. B. Christman, T. P. Pearsall, and J. E. Rowe, "Measurement of the *Γ-L* separation in $Ga_{0.47}In_{0.53}$ As by ultraviolet photoemission," *Applied Physics Letters*, vol. 40, no. 5, pp. 423–425, Mar. 1982.

[85] G. Greene-Diniz, M. V. Fischetti, and J. C. Greer, "Energies of the *X*- and *L* - valleys in $In_{0.53}Ga_{0.47}As$ from electronic structure calculations," *Journal of Applied Physics*, vol. 119, no. 5, p. 055707, Feb. 2016.

[86] M. V. Fischetti *et al.*, "Simulation of electron transport in high-mobility MOSFETs: Density of states bottleneck and source starvation," in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 2007, pp. 109–112.

[87] M. V. Fischetti *et al.*, "Theoretical Study of Some Physical Aspects of Electronic Transport in nMOSFETs at the 10-nm Gate-Length," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2116–2136, Sep. 2007.

[88] S. E. Laux, "A Simulation Study of the Switching Times of 22- and 17-nm Gate-Length SOI nFETs on High Mobility Substrates and Si," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2304–2320, Sep. 2007.

[89] H. Tsuchiya, A. Maenaka, T. Mori, and Y. Azuma, "Role of Carrier Transport in Source and Drain Electrodes of High-Mobility MOSFETs," *IEEE Electron Device Letters*, vol. 31, no. 4, pp. 365–367, Apr. 2010.

[90] E. Caruso, D. Lizzit, P. Osgnach, D. Esseni, P. Palestri, and L. Selmi, "Simulation analysis of III-V n-MOSFETs: Channel materials, Fermi level pinning and biaxial strain," in *2014 IEEE International Electron Devices Meeting*, San Francisco, CA, USA, 2014, pp. 7.6.1-7.6.4.

[91] M. Rau *et al.*, “Performance projection of III-V ultra-thin-body, FinFET, and nanowire MOSFETs for two next-generation technology nodes,” in *2016 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2016, pp. 30.6.1-30.6.4.

[92] F. M. Bufler, G. Eneman, N. Collaert, and A. Mocuta, “Monte Carlo Benchmark of $In_{0.53}Ga_{0.47}As$ and Silicon-FinFETs,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 13.3.1-13.3.4.

[93] S. H. Park *et al.*, “Performance Comparisons of III–V and Strained-Si in Planar FETs and Nonplanar FinFETs at Ultrashort Gate Length (12 nm),” *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2107–2114, Aug. 2012.

[94] R. Kim, U. E. Avci, and I. A. Young, “CMOS performance benchmarking of Si, InAs, GaAs, and Ge nanowire n- and pMOSFETs with Lg=13 nm based on atomistic quantum transport simulation including strain effects,” in *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, USA, 2015, pp. 34.1.1-34.1.4.

[95] D. M. Crum *et al.*, “Impact of gate oxide complex band structure on n-channel III–V FinFETs,” in *Simulation of Semiconductor Processes and Devices (SISPAD), 2015 International Conference on*, 2015, pp. 250–253.

[96] Y. T. Hou, M. F. Li, H. Y. Yu, and D.-L. Kwong, “Modeling of tunneling currents through HfO_2 and $(HfO_2)_x(Al_2O_3)_{1-x}$ gate stacks,” *IEEE Electron Device Letters*, vol. 24, no. 2, pp. 96–98, Feb. 2003.

[97] N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, “Modeling study of ultrathin gate oxides using direct tunneling current and capacitance-voltage measurements in MOS devices,” *IEEE Transactions on Electron Devices*, vol. 46, no. 7, pp. 1464–1471, 1999.

[98] F. A. Trumbore, “Solid solubilities of impurity elements in germanium and silicon,” *The Bell System Technical Journal*, vol. 39, no. 1, pp. 205–233, Jan. 1960.

[99] P. M. Rousseau, P. B. Griffin, W. T. Fang, and J. D. Plummer, “Arsenic deactivation enhanced diffusion: A time, temperature, and concentration study,” *Journal of Applied Physics*, vol. 84, no. 7, pp. 3593–3601, Oct. 1998.

[100] S. Thompson *et al.*, “Source/drain extension scaling for 0.1/spl mu/m and below channel length MOSFETs,” in *VLSI Technology, 1998. Digest of Technical Papers. 1998 Symposium on*, 1998, pp. 132–133.

[101] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, “Band parameters for III–V compound semiconductors and their alloys,” *Journal of Applied Physics*, vol. 89, no. 11, pp. 5815–5875, Jun. 2001.

[102] R. R. Parsons, “On the origin of photoluminescence in heavily-doped silicon,” *Solid State Communications*, vol. 29, no. 11, pp. 763–766, Mar. 1979.

[103] M. Eswaran, B. Bergersen, J. A. Rostworowski, and R. R. Parsons, “Impurity band states in SI(P),” *Solid State Communications*, vol. 20, no. 8, pp. 811–813, Nov. 1976.

[104] A. Ortiz-Conde, F. G. Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, “A review of recent MOSFET threshold voltage extraction methods,” *Microelectronics Reliability*, vol. 42, no. 4–5, pp. 583–596, 2002.

[105] C. Canali, G. Ottaviani, and A. Alberigi Quaranta, “Drift velocity of electrons and holes and associated anisotropic effects in silicon,” *Journal of Physics and Chemistry of Solids*, vol. 32, no. 8, pp. 1707–1720, Jan. 1971.

[106] C.-N. Ni *et al.*, “Ultra-low NMOS contact resistivity using a novel plasma-based DSS implant and laser anneal for post 7 nm nodes,” in *2016 IEEE Symposium on VLSI Technology*, Honolulu, HI, USA, 2016, pp. 1–2.

[107] U. Singisetti *et al.*, “Ultra-Low Resistance Ohmic Contacts to InGaAs/InP,” in *2007 65th Annual Device Research Conference*, South Bend, IN, USA, 2007, pp. 149–150.

[108] Raseong Kim, U. E. Avci, and I. A. Young, “Ge Nanowire nMOSFET Design With Optimum Band Structure for High Ballistic Drive Current,” *IEEE Electron Device Letters*, vol. 36, no. 8, pp. 751–753, Aug. 2015.

[109] T. Low *et al.*, “Electron mobility in Ge and strained-Si channel ultrathin-body metal-oxide semi conductor field-effect transistors,” p. 4.

[110] C. H. Lee *et al.*, “Ge MOSFETs performance: Impact of Ge interface passivation,” in *2010 International Electron Devices Meeting*, San Francisco, CA, USA, 2010, pp. 18.1.1-18.1.4.

[111] A. Lidow, J. F. Gibbons, V. R. Deline, and C. A. Evans, “Solid solubility of selenium in GaAs as measured by secondary ion mass spectrometry,” *Applied Physics Letters*, vol. 32, no. 9, pp. 572–573, May 1978.

[112] T. Fujii, T. Inata, K. Ishii, and S. Hiyamizu, “Heavily Si-doped InGaAs lattice-matched to InP grown by MBE,” *Electronics Letters*, vol. 22, no. 4, pp. 191–192, 1986.

[113] S. Prucnal *et al.*, “Ultra-doped n-type germanium thin films for sensing in the mid-infrared,” *Scientific Reports*, vol. 6, no. 1, Sep. 2016.

[114] S. Natarajan *et al.*, “A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 µm 2 SRAM cell size,” in *Electron Devices Meeting (IEDM), 2014 IEEE International*, 2014, pp. 3–7.

[115] A. Vardi, J. Lin, Wenjie Lu, X. Zhao, and J. A. del Alamo, “High aspect ratio InGaAs FinFETs with sub-20 nm fin width,” in *2016 IEEE Symposium on VLSI Technology*, Honolulu, HI, USA, 2016, pp. 1–2.

[116] Sang-Hyeon Kim *et al.*, “High Performance Tri-Gate Extremely Thin-Body InAs-On-Insulator MOSFETs With High Short Channel Effect Immunity and V_{th} Tunability,” *IEEE Transactions on Electron Devices*, vol. 61, no. 5, pp. 1354–1360, May 2014.

[117] J. B. Chang *et al.*, “Scaling of SOI FinFETs down to fin width of 4 nm for the 10nm technology node,” p. 2.

[118] C. C. Hsu, C. H. Chou, S. Y. Wang, W. C. Chi, C. H. Chien, and G. L. Luo, “Fabricating a n^+-Ge contact with ultralow specific contact resistivity by introducing a PtGe alloy as a contact metal,” *Applied Physics Letters*, vol. 107, no. 11, p. 113503, Sep. 2015.

[119] Y. Yoon, K. Ganapathi, and S. Salahuddin, “How Good Can Monolayer MoS_2 Transistors Be?,” *Nano Letters*, vol. 11, no. 9, pp. 3768–3773, Sep. 2011.

[120] Y. Pan *et al.*, “Near-ideal subthreshold swing MoS_2 back-gate transistors with an optimized ultrathin HfO_2 dielectric layer,” *Nanotechnology*, vol. 30, no. 9, p. 095202, Mar. 2019.

[121] B. Chen, H. Huang, X. Ma, L. Huang, Z. Zhang, and L.-M. Peng, “How good can CVD-grown monolayer graphene be?,” *Nanoscale*, vol. 6, no. 24, pp. 15255–15261, Oct. 2014.

[122] S. V. Morozov *et al.*, “Giant Intrinsic Carrier Mobilities in Graphene and Its Bilayer,” *Phys. Rev. Lett.*, vol. 100, no. 1, p. 016602, Jan. 2008.

[123] G.-H. Lee *et al.*, “Flexible and Transparent MoS_2 Field-Effect Transistors on Hexagonal Boron Nitride-Graphene Heterostructures,” *ACS Nano*, vol. 7, no. 9, pp. 7931–7936, Sep. 2013.

[124] X. Tong, E. Ashalley, F. Lin, H. Li, and Z. M. Wang, “Advances in MoS_2-Based Field Effect Transistors (FETs),” *Nano-Micro Lett.*, vol. 7, no. 3, pp. 203–218, Jul. 2015.

[125] K. He, C. Poole, K. F. Mak, and J. Shan, “Experimental Demonstration of Continuous Electronic Structure Tuning via Strain in Atomically Thin MoS_2,” *Nano Letters*, vol. 13, no. 6, pp. 2931–2936, Jun. 2013.

[126] R. Roldán, A. Castellanos-Gomez, E. Cappelluti, and F. Guinea, “Strain engineering in semiconducting two-dimensional crystals,” *Journal of Physics: Condensed Matter*, vol. 27, no. 31, p. 313201, 2015.

[127] R. Frisenda *et al.*, “Biaxial strain tuning of the optical properties of single-layer transition metal dichalcogenides,” *npj 2D Mater Appl*, vol. 1, no. 1, p. 10, Dec. 2017.

[128] D. Lloyd *et al.*, “Band Gap Engineering with Ultralarge Biaxial Strains in Suspended Monolayer MoS_2,” *Nano Letters*, vol. 16, no. 9, pp. 5836–5841, Sep. 2016.

[129] G. Plechinger *et al.*, “Control of biaxial strain in single-layer molybdenite using local thermal expansion of the substrate,” *2D Materials*, vol. 2, no. 1, p. 015006, Mar. 2015.

[130] Y. Y. Hui *et al.*, “Exceptional Tunability of Band Energy in a Compressively Strained Trilayer MoS_2 Sheet,” *ACS Nano*, vol. 7, no. 8, pp. 7126–7131, Aug. 2013.

[131] H. Li *et al.*, “Optoelectronic crystal of artificial atoms in strain-textured molybdenum disulphide,” *Nat Commun*, vol. 6, no. 1, p. 7381, Nov. 2015.

[132] K. S. Novoselov *et al.*, “Two-dimensional atomic crystals,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 30, pp. 10451–10453, 2005.

[133] A. Castellanos-Gomez *et al.*, “Local Strain Engineering in Atomically Thin MoS_2,” *Nano Letters*, vol. 13, no. 11, pp. 5361–5366, Nov. 2013.

[134] S. Bertolazzi, J. Brivio, and A. Kis, “Stretching and Breaking of Ultrathin MoS_2,” *ACS Nano*, vol. 5, no. 12, pp. 9703–9709, Dec. 2011.

[135] V. Mishra *et al.*, “Screening in Ultrashort (5 nm) Channel MoS_2 Transistors: A Full-Band Quantum Transport Study,” *IEEE Transactions on Electron Devices*, vol. 62, no. 8, pp. 2457–2463, Aug. 2015.

[136] L. Liu, S. B. Kumar, Y. Ouyang, and J. Guo, “Performance Limits of Monolayer Transition Metal Dichalcogenide Transistors,” *IEEE Transactions on Electron Devices*, vol. 58, no. 9, pp. 3042–3047, Sep. 2011.

[137] A. Nourbakhsh *et al.*, “MoS_2 Field-Effect Transistor with Sub-10 nm Channel Length,” *Nano Letters*, vol. 16, no. 12, pp. 7798–7806, Dec. 2016.

[138] K. A. Khair and S. S. Ahmed, “Effects of uniaxial strain on polar optical phonon scattering and electron transport in monolayer MoS_2 FETs,” in *Nanotechnology (IEEE-NANO), 2017 IEEE 17th International Conference on*, 2017, pp. 246–249.

[139] A. Sengupta, A. Chanana, and S. Mahapatra, “Phonon scattering limited performance of monolayer MoS_2 and WSe_2 n-MOSFET,” *AIP Advances*, vol. 5, no. 2, p. 027101, Feb. 2015.

[140] S. B. Adler and J. W. Smith, “Effects of long-range forces on oxygen transport in yttria-doped ceria: simulation and theory,” *J. Chem. Soc., Faraday Trans.*, vol. 89, no. 16, pp. 3123–3128, 1993.

[141] Z. Cheng, K. Price, and A. D. Franklin, “Contacting and Gating 2-D Nanomaterials,” *IEEE Transactions on Electron Devices*, vol. 65, no. 10, pp. 4073–4083, Oct. 2018.

[142] H. Fang *et al.*, “Degenerate n-Doping of Few-Layer Transition Metal Dichalcogenides by Potassium,” *Nano Lett.*, vol. 13, no. 5, pp. 1991–1995, May 2013.

[143] L. Yang *et al.*, “Chloride Molecular Doping Technique on 2D Materials: WS_2 and MoS_2,” *Nano Letters*, vol. 14, no. 11, pp. 6275–6280, Nov. 2014.

[144] F. Zahid, L. Liu, Y. Zhu, J. Wang, and H. Guo, “A generic tight-binding model for monolayer, bilayer and bulk MoS_2,” *AIP Advances*, vol. 3, no. 5, p. 052111, May 2013.

[145] T. Cheiwchanchamnangij and W. R. L. Lambrecht, “Quasiparticle band structure calculation of monolayer, bilayer, and bulk MoS_2,” *Physical Review B*, vol. 85, no. 20, May 2012.

[146] Z. Jin, X. Li, J. T. Mullen, and K. W. Kim, “Intrinsic Transport Properties of Electrons and Holes in Monolayer Transition Metal Dichalcogenides,” *Physical Review B*, vol. 90, no. 4, Jul. 2014.

[147] X. Li, J. T. Mullen, Z. Jin, K. M. Borysenko, M. Buongiorno Nardelli, and K. W. Kim, “Intrinsic electrical transport properties of monolayer silicene and MoS_2 from first principles,” *Physical Review B*, vol. 87, no. 11, Mar. 2013.

[148] A. Kumar and P. K. Ahluwalia, “Electronic structure of transition metal dichalcogenides monolayers 1H-MX_2 (M = Mo, W; X = S, Se, Te) from ab-initio theory: new direct band gap semiconductors,” *The European Physical Journal B*, vol. 85, no. 6, Jun. 2012.

[149] T. Eknapakul *et al.*, “Electronic Structure of a Quasi-Freestanding MoS_2 Monolayer,” *Nano Letters*, vol. 14, no. 3, pp. 1312–1316, Mar. 2014.

[150] A. Andersen, S. M. Kathmann, M. A. Lilga, K. O. Albrecht, R. T. Hallen, and D. Mei, “First-Principles Characterization of Potassium Intercalation in Hexagonal 2H-MoS_2,” *J. Phys. Chem. C*, vol. 116, no. 2, pp. 1826–1832, Jan. 2012.

[151] G. Kresse and J. Hafner, “Ab initio molecular dynamics for liquid metals,” *Phys. Rev. B*, vol. 47, no. 1, pp. 558–561, Jan. 1993.

[152] G. Kresse and J. Hafner, “Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium,” *Physical Review B*, vol. 49, no. 20, p. 14251, 1994.

[153] G. Kresse and D. Joubert, “From ultrasoft pseudopotentials to the projector augmented-wave method,” *Physical Review B*, vol. 59, no. 3, p. 1758, 1999.

[154] P. E. Blöchl, “Projector augmented-wave method,” *Physical Review B*, vol. 50, no. 24, p. 17953, 1994.

[155] K. K. H. Smithe, C. D. English, S. V. Suryavanshi, and E. Pop, "High-Field Transport and Velocity Saturation in Synthetic Monolayer MoS_2," *Nano Letters*, vol. 18, no. 7, pp. 4516–4522, Jul. 2018.

[156] Z. Yu *et al.*, "Realization of Room-Temperature Phonon-Limited Carrier Transport in Monolayer MoS_2 by Dielectric and Carrier Screening," *Adv. Mater.*, vol. 28, no. 3, pp. 547–552, Jan. 2016.

[157] B. Radisavljevic and A. Kis, "Mobility engineering and a metal–insulator transition in monolayer MoS_2," *Nature Materials*, vol. 12, no. 9, pp. 815–820, Sep. 2013.

[158] B. Chamlagain *et al.*, "Mobility Improvement and Temperature Dependence in $MoSe_2$ Field-Effect Transistors on Parylene-C Substrate," *ACS Nano*, vol. 8, no. 5, pp. 5079–5088, May 2014.

[159] C. Rice *et al.*, "Raman-scattering measurements and first-principles calculations of strain-induced phonon shifts in monolayer MoS_2," *Physical Review B*, vol. 87, no. 8, Feb. 2013.

[160] H. J. Conley, B. Wang, J. I. Ziegler, R. F. Haglund Jr., S. T. Pantelides, and K. I. Bolotin, "Bandgap Engineering of Strained Monolayer and Bilayer MoS_2," *arXiv:1305.3880 [cond-mat]*, May 2013.

[161] N. Huo, Y. Yang, Y.-N. Wu, X.-G. Zhang, S. T. Pantelides, and G. Konstantatos, "High carrier mobility in monolayer CVD-grown MoS_2 through phonon suppression," *Nanoscale*, vol. 10, no. 31, pp. 15071–15077, 2018.

[162] N. Neophytou, T. Rakshit, and M. S. Lundstrom, "Performance Analysis of 60-nm Gate-Length III–V InGaAs HEMTs: Simulations Versus Experiments," *IEEE Transactions on Electron Devices*, vol. 56, no. 7, pp. 1377–1387, Jul. 2009.

[163] N. Wakabayashi, H. G. Smith, and R. M. Nicklow, "Lattice dynamics of hexagonal MoS_2 studied by neutron scattering," *Physical Review B*, vol. 12, no. 2, pp. 659–663, Jul. 1975.

[164] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically Thin MoS_2: A New Direct-Gap Semiconductor," *Physical Review Letters*, vol. 105, no. 13, Sep. 2010.

[165] W. S. Yun, S. W. Han, S. C. Hong, I. G. Kim, and J. D. Lee, "Thickness and strain effects on electronic structures of transition metal dichalcogenides: 2H-MX_2 semiconductors (M = Mo, W; X = S, Se, Te)," *Physical Review B*, vol. 85, no. 3, Jan. 2012.

[166] K. Kaasbjerg, K. S. Thygesen, and K. W. Jacobsen, "Phonon-limited mobility in n-type single-layer MoS_2 from first principles," *Physical Review B*, vol. 85, no. 11, Mar. 2012.

[167] E. Scalise, M. Houssa, G. Pourtois, V. Afanas'ev, and A. Stesmans, "Strain-induced semiconductor to metal transition in the two-dimensional honeycomb structure of MoS_2," *Nano Research*, vol. 5, no. 1, pp. 43–48, Jan. 2012.

[168] B. R. Carvalho, L. M. Malard, J. M. Alves, C. Fantini, and M. A. Pimenta, "Symmetry-Dependent Exciton-Phonon Coupling in 2D and Bulk MoS_2 Observed by Resonance Raman Scattering," *Physical Review Letters*, vol. 114, no. 13, Apr. 2015.

[169] H. Peelaers and C. G. Van de Walle, "Effects of strain on band structure and effective masses in MoS_2," *Physical Review B*, vol. 86, no. 24, Dec. 2012.

[170] E. S. Kadantsev and P. Hawrylak, "Electronic structure of a single MoS_2 monolayer," *Solid State Communications*, vol. 152, no. 10, pp. 909–913, May 2012.

[171] R. Kappera *et al.*, "Phase-engineered low-resistance contacts for ultrathin MoS_2 transistors," *Nature Materials*, vol. 13, no. 12, pp. 1128–1134, Dec. 2014.

[172] M. Hosseini, M. Elahi, M. Pourfath, and D. Esseni, "Strain-Induced Modulation of Electron Mobility in Single-Layer Transition Metal Dichalcogenides MX_2 (M = Mo, W; X = S, Se)," *IEEE Transactions on Electron Devices*, vol. 62, no. 10, pp. 3192–3198, Oct. 2015.

[173] L. F. Register, E. Rosenbaum, and K. Yang, "Analytic model for direct tunneling current in polycrystalline silicon-gate metal–oxide–semiconductor devices," *Appl. Phys. Lett.*, vol. 74, no. 3, pp. 457–459, Jan. 1999.

[174] Fei Li, S. P. Mudanai, Yang-Yu Fan, L. F. Register, and S. K. Banerjee, "Physically based quantum-mechanical compact model of MOS devices substrate-injected tunneling current through ultrathin (EOT /spl sim/ 1 nm) SiO/sub 2/ and high-/spl kappa/ gate stacks," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1096–1106, May 2006.

Vita

Aqyan A. Bhatti is pursuing a Ph.D. in Materials Science and Engineering at the University of Texas at Austin, where he also received his B.S. in Mechanical Engineering in 2013. His research is focused on understanding and modeling the underlying physics in the operation of nanoscale solid-state electronic devices.

Permanent address: abhatti@utexas.edu

This dissertation was typed by Aqyan Ahmed Bhatti.