

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Design of scaled electronic devices based on III-V materials

### Permalink

<https://escholarship.org/uc/item/6qj9x9fs>

### Author

Wang, Lingquan

### Publication Date

2009

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Design of Scaled Electronic Devices Based on III-V Materials

A dissertation submitted in partial satisfaction of the  
requirements for the degree Doctor of Philosophy

in

Electrical Engineering (Applied Physics)

by

Lingquan Wang

Committee in charge:

Professor Peter Asbeck, Chair

Professor Prabhakar Bandaru

Professor Andrew Kummel

Professor Yuan Taur

Professor Paul Yu

2009

Copyright

Lingquan Wang, 2009

All Right Reserved

The Dissertation of Lingquan Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

Chair

University of California, San Diego

2009

DEDICATION

*To my parents*

## TABLE OF CONTENTS

Signature Page .....	iii
Dedication .....	iv
Table of Contents .....	v
List of Figures .....	xi
List of Tables .....	xxiii
Acknowledgements .....	xxiv
Vita .....	xxvi
Abstract of the Dissertation.....	xxx
Chapter 1. Introduction.....	1
1.1 Overview of CMOS Scaling .....	1
1.1.1 Mobility Enhancement by Strain Engineering .....	2
1.1.2 SCE Control .....	4
1.1.3 Difficulty in Threshold Voltage Scaling .....	4
1.2 Scaled III-V MOSFET .....	6
1.2.1 The Other Side of High Mobility .....	6
1.2.2 Quasi-ballistic Transport .....	7
1.2.3 Impact of Non-ideal Dielectric .....	8
1.2.4 Possible Device Structures for III-V MOSFET .....	8
1.2.5 Beyond Thermionic Emission Based III-V MOSFET .....	9
1.2.6 Theme of This Dissertation .....	10
1.3 Organization of the Thesis .....	12
Reference.....	15

Part I .....	17
Chapter 2. Layer Structure Design of a III-V MOSFET, Quantum Effect and Non-parabolicity .....	18
2.1 Introduction .....	18
2.2 Self Consistent 1-D Schrödinger Poisson Solver with Conduction Band Non-parabolicity .....	21
2.2.1 Objectives.....	21
2.2.2 Methodology .....	22
2.2.3 Model Verification .....	29
2.2.4 Results and Discussion .....	31
2.2.5 Model Extension .....	36
2.2.6 Further Discussion and Future Work .....	38
2.3 Capacitance Analysis of a III-V MOSFET Layer Structure Design ...	40
2.3.1 Objectives .....	40
2.3.2 A First Order Capacitance Model .....	40
2.3.3 A General Capacitance Model .....	46
2.4 Layer Design for a Representative III-V MOSFET .....	55
2.4.1 Introduction .....	55
2.4.2 Layer Structure Design for a III-V MOSFET .....	55
2.5 Summary .....	61
Reference.....	62
Chapter 3. Quasi-ballistic Transport in the Ultra-scaled III-V MOSFET.....	63
3.1 Introduction .....	63

3.2	Brief Review of Ballistic Transport Theory .....	64
3.3	Modified Ballistic Transport Theory and Ballistic Performance Comparison between III-V MOSFET and Si MOSFET .....	67
3.4	Quasi-ballistic Transport with Realistic Considerations and Performance Projections .....	73
3.5	Discussion and Summary .....	83
	Reference.....	85
Chapter 4. Source Design of a III-V MOSFET .....		86
4.1	Introduction .....	86
4.2	Simplified Model on Source Design .....	88
4.3	Simplified Model on Source Starvation .....	93
	4.3.1 Ionized Impurity Scattering .....	94
	4.3.2 Electron-Electron Scattering .....	99
	4.3.3 Source Supportable Current Computation .....	103
4.4	Summary .....	108
	Reference.....	109
Chapter 5. Dielectric Layer for III-V MOSFET .....		110
5.1	Introduction .....	110
5.2	Capacitance Voltage Characteristics with Non-ideal Dielectrics .....	113
	5.2.1 Idealized Capacitance Voltage Characteristics .....	113
	5.2.2 Impact of Interface States on Capacitance Voltage Characteristics .....	116



5.2.3	Impact of Border Traps on Capacitance Voltage Characteristics .....	125
5.3	Subthreshold Characteristics with Non-ideal Dielectrics .....	147
5.4	Summary .....	152
	Reference .....	153
Part II	.....	154
Chapter 6.	Device Physics and Ballistic Transport of III-V Nanowire Based MOSFETs .....	155
6.1	Introduction .....	155
6.2	Self Consistent Schrödinger-Poisson Solver for Radially Symmetric Nanowire Core-shell Structures .....	157
6.2.1	Introduction .....	157
6.2.2	Schrödinger and Poisson Equation Solvers .....	158
6.2.2.1	Schrödinger Solver .....	158
6.2.2.2	Poisson Solver .....	161
6.2.2.3	Self-consistent Calculation .....	163
6.2.3	Model Verification .....	166
6.2.3.1	Infinite Cylindrical Square Potential Well .....	166
6.2.3.2	Comparison with 1D Schrödinger-Poisson Solver .....	167
6.2.4	Model Applications .....	168
6.2.5	Limitation of the Model .....	175
6.2.6	Summary .....	176
6.3	Ballistic Transport in III-V Nanowire MOSFETs .....	177

6.3.1	Introduction .....	177
6.3.2	Density of States (DOS) in Nanowires .....	177
6.3.3	Ballistic Limit Current .....	183
6.3.4	Summary .....	187
6.4	Performance Comparison of Scaled III-V and Si Ballistic Nanowire MOSFETs .....	188
6.4.1	Introduction .....	188
6.4.2	Performance Comparison between Si and III-V MOSFET .....	188
6.4.3	Performance Comparison with Different EOT .....	193
6.4.4	Summary .....	196
6.5	Summary .....	197
	Reference.....	199
Part III	.....	201
Chapter 7. Tunneling FET Based on III-V Staggered Heterojunction for Ultra-low Power Applications .....		202
7.1	Introduction .....	202
7.2	Tunneling Field Effect Transistor (TFET) Based on Staggered Heterojunction for Ultra-low Power Applications .....	206
7.2.1	Introduction .....	206
7.2.2	Concept of a TFET and Material Choices.....	207
7.2.2.1	Concept of a TFET .....	207
7.2.2.2	BTB Tunneling Process .....	208
7.2.2.3	Material Consideration .....	216

7.2.3	Device Structure and Device Physics of TFETs .....	217
7.2.3.1	Device Design Considerations for Tunneling Junction of a TFET .....	217
7.2.3.2	Material Choices and Possible Technical Implementation for a TFET .....	220
7.2.3.3	Device Physics of a TFET .....	224
7.2.4	Summary .....	227
7.3	Design Considerations for Tunneling Field Effect Transistors (TFET) Based on Staggered Heterojunctions .....	229
7.3.1	Introduction .....	229
7.3.2	Baseline Device Design .....	229
7.3.3	Design Variations and Tradeoffs .....	230
7.3.3.1	Doping Strategy .....	230
7.3.3.2	Material Variation .....	235
7.3.4	Conclusion .....	237
7.4	Summary .....	238
	Reference .....	239
	Chapter 8. Summary of the Thesis.....	240

## LIST OF FIGURES

Fig. 1.1. Universal mobility curve for electrons and holes in Si. ....	2
Fig. 1.2. Illustration of the effects on conduction band of Si under stresses of different types. ....	4
Fig. 2.1. A comparison on E-k dispersions between parabolic band description and non-parabolic band description. ....	19
Fig.2.2. Schematic of the self-consistent solving procedure of Schrödinger equation and Poisson's equation including non-parabolic conduction band.....	23
Fig. 2.3. Schematic of layer structure computed.....	27
Fig. 2.4. Computation result for the layer structure shown in Fig.2: a) In-plane dispersion relationship of the 1 <sup>st</sup> and 2 <sup>nd</sup> subbands; b) Electron distribution with parabolic conduction band assumption as well as non-parabolic conduction band condition; c) Self-consistent electrostatic potential profile.....	27
Fig. 2.5. Sheet charge density as a function of the gate bias under parabolic band condition and non-parabolic conduction band condition.....	29
Fig. 2.6. Comparison between analytical and numerical result for the structure shown in the inset. ....	30
Fig. 2.7. Comparison between the in-plane dispersion relationship between analytical and numerical results.....	31
Fig. 2.8. Computed Density of States (DOS) as a function of energy. Energy 0 here reflects the position of electron Fermi level.....	32
Fig. 2.9. Evolution of the 1 <sup>st</sup> subband wavefunction as in-plane kinetic energy increases.....	33

Fig. 2.10. Computed DOS function under different gate bias (i.e. different electron occupation within the quantum well).....	34
Fig. 2.11. Derived quasi-static capacitance voltage characteristic for both parabolic and non-parabolic conduction band condition.....	35
Fig. 2.12. Electron occupation within $\Gamma$ valley and L valley under gate bias of 1.0V, with negligible L valley occupation.....	37
Fig. 2.13. Electron occupation within $\Gamma$ valley and L valley under gate bias of 1.4V, with L valley occupation accounts for ~2% the overall sheet charge density.....	37
Fig. 2.14. Schematic illustration of a III-V MOSFET layer structure and its corresponding energy band diagram. $E_{fm}$ , $E_{fs}$ are Fermi energies of the gate and the semiconductor, respectively. $E_1$ is the energy of the first quantized state.....	41
Fig. 2.15. Breakdown of contributions to the overall input capacitance from 1) oxide (high k dielectric) as gray bar, 2) quantum well capacitance due to the wave function spread as blue bar, 3) density of state capacitance as orange bar. ....	45
Fig. 2.16. Equivalent circuit for generalized gate capacitance model. ....	50
Fig. 2.17. Band alignment of selected III-V materials.....	55
Fig. 2.18. Illustration of the baseline layer structure design of a III-V MOSFET.....	56
Fig. 2.19. Carrier distribution, wave function and electrostatic potential at representative gate bias for the baseline layer structure design. ....	58
Fig. 2.20. Sheet charge density as a function of gate bias. Gate metal work function is assumed to be 4.7eV. ....	59

Fig. 2.21. Derived quasi-static CV characteristics for the baseline layer structure design. ....	59
Fig. 3.1. Schematic of the energy band diagram along the transport direction (bottom Figure) and the transverse direction (top right Figure). It is assumed that the quantization occurs within the channel. ....	64
Fig. 3.2. Sheet charge density at virtual source under one-subband ballistic condition, gate overdrive is ramped from 0.1V to 0.5V with a step of 0.1V for both InGaAs and Si.....	69
Fig. 3.3. Schematic of the energy band diagram and gate coupling to the virtual source with forward and backward going electrons under zero drain bias and high drain bias.....	70
Fig. 3.4. Computed ballistic limit I-V characteristics for an InGaAs and a Si MOSFET, with EOT thickness of 1nm assumed.....	72
Fig. 3.5. Computed ballistic limit I-V characteristics for the baseline design and its variations as presented in Chapter 2, EOT here is 0.6nm. Threshold voltages are adjusted for all designs. ....	74
Fig. 3.6. Computed ballistic limit I-V characteristics for the baseline design with series resistance at the source. ....	74
Fig. 3.7. Illustration of the Delaunay triangulation for extrinsic device characteristic computation.....	76
Fig. 3.8. Illustration of the quasi-ballistic transport with the description of transmission coefficient.....	76
Fig. 3.9. Computed quasi-ballistic output characteristics for InGaAs and Si MOSFETs, assuming various mobility based on material properties.....	78

Fig. 3.10. Computed quasi-ballistic output characteristics for InGaAs and Si MOSFETs with the series resistance being taken into account. The assumed mobilities for InGaAs and Si are 1000 cm <sup>2</sup> /V-sec and 200 cm <sup>2</sup> /V-sec, respectively. The assumed series resistances are 20 Ω-μm and 50 Ω-μm on each side, respectively. ....	79
Fig. 3.11. Model verification by comparing hardware data (35nm gate length Si NMOSFET) to model computation, with extracted mobility of 200 cm <sup>2</sup> /V-sec and series resistance of 80 Ω-μm.....	80
Fig. 3.12. Representative waveforms from succeeding stages of a ring oscillator constructed with InGaAs ballistic MOSFETs, whose characteristics are described by a Curtice Cubic model.....	82
Fig. 4.1. Schematic drawing of the carrier distribution within the channel and the source region.....	86
Fig. 4.2. Coordinates setup for the computation of source starvation.....	88
Fig. 4.3. Comparison among source designs with different doping concentrations. The geometrical dimensions of the structure simulated above are with channel thickness of 5nm and source thickness of 20nm.....	91
Fig. 4.4. Comparison for source designs with different geometrical parameters and doping levels.....	92
Fig. 4.5. Illustration of ionized impurity scattering in the k-space. $k_i$ and $k_f$ are the initial and final momentums, respectively; $\alpha$ is the angle between the two vectors.....	95
Fig. 4.6. Illustration of probability density as a function of the scattering angle. The angle may be determined by requiring the area of the shaded part beneath the curve to be equal to a chosen	

random number generated between 0 and 1. The angle at which such requirement is satisfied is chosen to be the scattering angle.....	98
Fig. 4.7. Momentum relaxation time as a function of electron energy for different doping levels.....	99
Fig. 4.8. Momentum relaxation time due to electron-electron(vacancy) scattering process mapped out on a 2-D momentum space. ....	101
Fig. 4.9. Probability density as a function of the out-scattered vacancy momentum mapped onto the 2-D momentum space.....	102
Fig. 4.10. Probability density of having a certain out-scattered electron momentum.....	103
Fig. 4.11. Overall momentum relaxation time as a function of momentum mapped onto the 2-D momentum space.....	105
Fig. 4.12. Numerical simulation result for the source limited current with ionized impurity and electron-electron scattering taken into account. ....	106
Fig. 5.1. Ideal quasi-static C-V characteristics for n-MOSCAP with 4.6nm (physical thickness) Al <sub>2</sub> O <sub>3</sub> on In <sub>0.47</sub> Ga <sub>0.53</sub> As doped at $\sim 1 \times 10^{17} \text{ cm}^{-3}$ . The dashed line represents the oxide capacitance limit.....	113
Fig. 5.2. Ideal frequency dependent C-V characteristics for n-MOSCAP with 4.6nm (physical thickness) Al <sub>2</sub> O <sub>3</sub> on In <sub>0.47</sub> Ga <sub>0.53</sub> As doped at $\sim 1 \times 10^{17} \text{ cm}^{-3}$ . The minority carrier lifetime assumed here is 1s.....	115
Fig. 5.3. Representative experimental C-V characteristics for n-MOSCAP with 4.6nm (physical thickness) Al <sub>2</sub> O <sub>3</sub> on In <sub>0.47</sub> Ga <sub>0.53</sub> As doped at $\sim 1 \times 10^{17} \text{ cm}^{-3}$ .....	117



Fig. 5.4. Time constant of interface traps as a function of the energy. ....	119
Fig. 5.5. Energy band diagram for illustration of the capacitive response of interface states....	120
Fig. 5.6. Representative simulated C-V characteristics for n-MOSCAP with 4.6nm (physical thickness) Al <sub>2</sub> O <sub>3</sub> on In <sub>0.47</sub> Ga <sub>0.53</sub> As doped at $\sim 1 \times 10^{17} \text{ cm}^{-3}$ , uniform $5 \times 10^{12} \text{ cm}^{-2} \cdot \text{eV}^{-1}$ Dit distribution.....	122
Fig. 5.7. Responsive energy range upon different small signal frequency.....	124
Fig. 5.8. Schematic drawing of border traps and coordinates setup for further modeling. The origin is set at the interface between the semiconductor and the dielectric.....	126
Fig. 5.9. Linear relationship between the maximum capacitance measured on the accumulation side and $\log(1/f)$ .....	127
Fig. 5.10. Equivalent circuit for evaluation of the border trap time constant.....	128
Fig. 5.11. Energy band diagram for illustration of the capacitive response of border traps under degenerate and non-degenerate cases.....	134
Fig. 5.12. Equivalent circuit of the gate input capacitance including border traps contribution.	137
Fig. 5.13. Experimental data on conductance as a function of frequency, showing the linear relationship as predicted by Eq. (5.27).....	139
Fig. 5.14. Representative time constant associated with border traps. Open circle symbol: low carrier concentration case ( $10^{15} \text{ cm}^{-3}$ ); Open diamond symbol: high carrier concentration case ( $3 \times 10^{19} \text{ cm}^{-3}$ ). Calculation performed in accordance to definition in Eq. (5.28).....	141
Fig. 5.15. Schematic drawing of the time domain response for the charging characteristics of the border traps. Solid line represents $E_{fs}$ , dashed line represents $E_{ft}$ . ....	142

Fig. 5.16. Representative Fermi level position as a function of location during the DC sweep.	143
Fig. 5.17. Equivalent circuit model to obtain the overall gate capacitance.....	144
Fig. 5.18. Side by side comparison between experimental data [3] and simulation result. The simulated frequencies are: 5kHz, 10kHz, 20kHz, 50kHz, 80kHz, 100kHz, 200kHz, 500kHz, 800kHz. The simulation only takes into account the electron charging of the traps.....	145
Fig. 5.19. Experimental subthreshold current at different drain biases. (Credit: Prof. Rodwell's group at UCSB).....	147
Fig. 5.20. Simulation result of the subthreshold current showing linear dependence of drain current on $V_{ds}$ bias.....	150
Fig. 5.21. Physical picture of Dit-induced DIBL effects.....	151
Fig. 6.1. Schematic of the core-shell nanowire structure and the coordinate system used.....	158
Fig. 6.2. Flow chart of the self-consistent solution of the Schrödinger equation and the Poisson equation.....	165
Fig. 6.3. Model verification for the Schrödinger solver for the case of infinite depth cylindrical potential well. Plotted in the Fig. are the first three wave functions of even parity, from both numerical calculations and analytical predictions. The inset tabulates the comparison between the eigenenergies obtained from the two methods.....	167
Fig. 6.4. Model verification by comparing the present solver with a 1D Schrödinger-Poisson slab quantum well solver for a large radius nanowire. The simulation was conducted for a 450 nm radius core-shell structure. The same thicknesses were used in the slab quantum-well solver.	

Inset shows a comparison within the entire radius range, showing good agreement between the two solutions.....168

Fig. 6.5. Model application to the self-consistent calculation of 2-DEG formation in an InAs/GaAs core shell structure. The calculated structure has an overall radius of 40 nm; the InAs core radius is 36 nm. (a) Wave functions for first few subbands. (b) Conduction-band energy for the quantum self consistent solution (solid line) and classical Poisson only solution (dashed line). Eigenenergies were also shown with thin lines, corresponding to the wave functions shown in (a). (c) Electron distribution in quantum self-consistent solution (solid line) and classical Poisson only solution (dashed line). Also provided is the solution from the slab quantum-well solver for a slab sandwich structure constituted by InAs and GaAs with thicknesses as 36 nm and 4 nm.....170

Fig. 6.6. Model application to the self-consistent calculation of 2-DEG formation in an InAs/GaAs core shell structure. The calculated structure has an overall radius of 20 nm; the InAs core radius is 18 nm. The results show a more pronounced volume electron density than in the case of Fig. 6.5.....171

Fig. 6.7. (a) Predicted Qs–V characteristic for a structure similar to that of Fig. 6.5. (b) Derived C–V characteristic for the same structure, with comparison of the classically predicted C–V curve, the horizontal line is the capacitance calculated as a co-axial cylindrical capacitor. Inset of (a) shows schematically the structure under consideration.....173

Fig. 6.8. Equivalent circuit for de-embedding quantum capacitances.  $C_{qc}$  is the collective quantum capacitance, while  $C_{qc1}$  stands for the influence due to the finite wave function spread;  $C_{qc2}$  accounts for the influence due to the reduced DOS.....174

Fig. 6.9. De-embedded quantum capacitances as functions of voltage.....175

Fig. 6.10. Computed density of states of a Si nanowire with 2.5nm radius under flat-band condition using the model presented in [15]. Solid line and dashed line represent different eigen energy ladder due to anisotropic effective mass along different direction.....	178
Fig. 6.11. Computed sheet charge density (line charge density normalized by core periphery) of the structure shown in Fig 7a inset, using a self consistent Schrödinger Poisson Solver considering cylindrical coordinates.....	179
Fig. 6.12. Equivalent capacitance network between gate and channel.....	180
Fig. 6.13. Derived C-V characteristic of the computation example. Symbols represent the overall C-V characteristic with influences from QM confinement and limited DOS.....	181
Fig. 6.14. Overall capacitance for InGaAs core and Si core with different diameters normalized to the oxide (geometrical barrier) capacitance.....	182
Fig. 6.15. Computed ballistic output characteristics for InGaAs nanowire with (a) self consistent solution and (b) one subband approximation.....	186
Fig. 6.16. Schematic drawing of nanowire transistor.....	188
Fig. 6.17. Computed ballistic line charge density as function of drain bias for (a) Si nanowire and (b) InGaAs nanowire.....	189
Fig. 6.18. Computed ballistic output characteristic as function of drain bias for (a) Si nanowire and (b) InGaAs nanowire. The drain Current is normalized to the circumference of the nanowire core.....	190
Fig. 6.19. Average injection velocities of Si and InGaAs at the virtual source.....	191

Fig. 6.20. Comparison between ballistic model and experimental result for Si Twin Nanowire FET reported in [20]. Symbols are experimental results, lines are computed results with various oxide thickness.....192

Fig. 6.21. Computed ballistic saturation current at T=0K with one subband only, with gate overdrive of 0.5V. The EOT considered here includes both oxide thickness and finite wave function depth. Solid curves are for single valley situation, dotted curves are for valley degeneracy of 6. Current is normalized to circumference of the nanowire core.  $m_{\text{eff}}/m_0$  values appropriate to InGaAs and Si are shown with symbol (o).....194

Fig. 6.22. Numerical simulation of ballistic saturation current at room temperature with gate overdrive of 0.5V. The EOT considered here includes ONLY the oxide thickness. Solid curves are for single valley situation, dotted curves are for valley degeneracy of 6. Current is normalized to circumference of nanowire core.....195

Fig. 7.1. Compilation of power consumptions of representative Intel CPU's as a function of operating clock frequency, over a range of technology nodes and product generations.....203

Fig. 7.2. Schematic drawing of the thermionic conduction mechanism which results in the >60mV/dec turn-on characteristics in conventional MOSFETs.....204

Fig. 7.3. Schematic band diagram for (a) off state and (b) on state for the source/channel junction.....207

Fig. 7.4. Illustration of the tunneling barrier for hole/electron tunnel model.....209

Fig. 7.5. Coordinate setup for the BTB tunneling computation based on two-band model.....212

Fig. 7.6. Comparison on evanescent wave vectors using hole/electron tunneling model Eq. (7.3) and two-band model Eq. (7.11).....214

Fig. 7.7. Tunneling probability obtained from hole/electron tunneling model and 2-band model as a function of electric field.....	215
Fig. 7.8. BTB tunneling probability versus electric field for $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$ and $\text{InAs}$ materials..	216
Fig. 7.9. Schematic drawing of TFET device with planar embodiment (not to scale).....	219
Fig. 7.10. (a) Approximate band line-up for selected semiconductors, (b) Schematic energy band diagram along the channel direction at on-state.....	220
Fig. 7.11. Schematic drawing of TFET device with planar embodiment (not to scale) and matched lattice constants.....	222
Fig. 7.12. Band lineup of the materials in the TFET design shown in Fig. 7.11.....	223
Fig. 7.13. Representative energy band diagram under off-state (zero gate bias) and on-state (0.3V gate bias), both with 0.3V drain bias.....	224
Fig. 7.14. (a) Transfer characteristic and (b) output characteristic of a TFET, with $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}_{0.3}\text{Sb}_{0.7}$ p+ source and $\text{In}_{0.8}\text{Ga}_{0.23}\text{As}$ n- channel.....	227
Fig. 7.15. Transfer characteristic of the baseline design TFET, design parameters same as specified in the text.....	230
Fig. 7.16. (a) Direct gate to channel coupling with aligned gate, (b) gate to channel coupling via fringing capacitance with gate-source-underlap, (c) gate coupling to both source and channel with gate-source-overlap.....	231
Fig. 7.17. Impact of gate position with respect to the tunneling junction to turn-on characteristic of TFET.....	232
Fig. 7.18. (a) Schematic drawing of inefficient turn-on when source is degenerately doped. Valence electrons with energy above (but close to) the Fermi level will start premature tunneling,	

therefore smear the subthreshold swing. (b) Transfer characteristic of devices with  $2 \times 10^{19} \text{ cm}^{-3}$  and  $5 \times 10^{19} \text{ cm}^{-3}$  source doping.....234

Fig. 7.19. On-state (squares) and off-state current (diamonds) for different band line-up conditions (via changing the Indium content within the channel). Inset shows energy gap (between source valence band and channel conduction band) at the tunneling junction as a function of Indium content.....236

## LIST OF TABLES

Table 2.1. Material parameters for the example design

Table 2.2. Material parameter used for layer structure design of a III-V MOSFET.

Table 3.1. List of parameters for propagation delay estimation

Table 5.1. Temperature dependence of the accumulation capacitance dispersion

Table 7.1. Material parameters for  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  and InAs tunneling computation

Table 7.2. Material parameters for lattice matched TFET design



## ACKNOWLEDGEMENTS

I would like to acknowledge Prof. Peter Asbeck for being a great mentor throughout my PhD study, who always enlightened me with his wisdom, nurtured me with his knowledge, influenced me with his perseverance, and set himself as a role model for myself and many of his other students.

I would like to acknowledge Prof. Yuan Taur for his mentorship through the joint program on Non-classical CMOS funded by SRC. I immensely enjoyed the work we have accomplished together, and was grateful for all the knowledge I learnt from Prof. Taur.

I am also indebted to many of my collaborators on various projects. Among them are Bo Yu, Yu Yuan, Prof. Max Fischetti (UMass), Prof. Andrew Kummel (UCSD), Prof. Mark Rodwell (UCSB), Prof. Paul McIntyre (Stanford) on the Non-classical CMOS project; Prof. Paul Yu (UCSD), Wei Lu, David Apline on GaN projects, Prof. Edward Yu (UCSD/UT) on tunneling FET project, Prof. Bandaru for serving on my committee.

As in many of the cases, I as well, have been always strongly supported by my parents. Their selfless love has gone beyond any word is capable to describe, and I can never thank them enough for everything they have done.

Section 2.2, in full, is submitted to “Solid State Electronics”. The dissertation author was the primary author of the paper.

Section 6.2, in full, is a reprint of "A numerical Schrödinger–Poisson solver for radially symmetric nanowire core–shell structures", Lingquan Wang, Deli Wang and Peter M. Asbeck, Solid State Electronics, Nov. 2006, pp 1732-1739. The dissertation author was the primary author of the paper.

Section 6.3, is part of the paper “Scaling of Nanowire Transistors (invited paper)”, Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, Yuan Taur, IEEE Transaction of Electron Devices, Nov. 2008, pp.2846-2858. The dissertation author was a co-author of this paper.

Section 6.4, is a rearrangement of the paper “Performance Comparison of Scaled III-V and Si Nanowire MOSFET”, Lingquan Wang, Bo Yu, Peter Asbeck, Yuan Taur, Mark Rodwell, International Journal of High Speed Electronics and Systems, Mar. 2009, pp.15-22. The dissertation author was the primary author.

Section 7.3, is a rearrangement of the paper “Design Considerations for Tunneling MOSFETs Based on Staggered Heterojunctions for Ultra-Low-Power Applications”, Lingquan (Dennis) Wang, Peter Asbeck, Nanoscale Material and Device Conference, 2009. The dissertation author was the primary author.

## VITA

2002	Bachelor of Engineering, Tsinghua University, Beijing, China
2004	Master of Science, University of Florida, Gainesville, FL, US
2002-2004	University of Florida Alumni Fellowship Recipient
2004-2009	Resistance Assistant, University of California, San Diego
2009	Doctor of Philosophy, University of California, San Diego, La Jolla, CA, US

## PUBLICATIONS

“Border Traps in  $\text{Al}_2\text{O}_3/\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  (100) Gate Stacks and their Passivation by Hydrogen Anneals” Eun J. Kim, Lingquan Wang, Peter M. Asbeck, Krishna C. Saraswat, and Paul C. McIntyre, accepted by *Applied Physics Letter*

“Analysis of Resistance Asymmetry Due to p-n Junctions in Graphene FETs”, Kangmu Lee, Lingquan Wang, Peter Asbeck and Jong S.Moon, submitted to *Material Research Symposium*, 2010

“Tunneling III-V MOSFETs for End of the Roadmap Applications (invited talk)”, Peter Asbeck, Lingquan Wang, Siyuan Gu, Yuan Taur and Edward Yu, International Workshop of Physics of Semiconductor Devices, 2009

“Tunneling III-V MOSFETs for End of the Roadmap Applications (invited talk)”, Peter Asbeck, Lingquan Wang, Yuan Taur and Edward Yu, *Material Research Symposium*, 2010

“In<sub>0.53</sub>Ga<sub>0.47</sub>As channel MOSFETs with self-aligned InAs source/drain formed by MEE regrowth”, Uttam Singisetti et. al. IEEE Electronic Device Letter , Nov. 2009

“Tunneling Field Effect Transistor Based on Staggered Hetero-tunnel-junction for Ultra-low Power Application”, Lingquan (Dennis) Wang, Edward Yu, Yuan Taur, Peter Asbeck, submitted to IEEE Electronic Device Letter

“Design Considerations for Tunneling MOSFETs Based on Staggered Heterojunctions for Ultra-Low-Power Applications”, Lingquan (Dennis) Wang, Peter Asbeck, Nanoscale Material and Device Conference, 2009

“Electrical Properties of Ga<sub>2</sub>O<sub>3</sub>/GaAs Interfaces and GdGaO Dielectrics in GaAs Based MOSFETs”, M. Passlack, R. Droopad, P. Fejes, L. Wang, IEEE Electronic Device Letter, Jan. 2009, pp. 2-4

“Performance Comparison of Scaled III-V and Si Nanowire MOSFET”, Lingquan Wang, Bo Yu, Peter Asbeck, Yuan Taur, Mark Rodwell, International Journal of High Speed Electronics and Systems, Mar. 2009, pp.15-22

“Performance Comparison of Scaled III-V and Si Nanowire MOSFET”, Lingquan Wang, Bo Yu, Peter Asbeck, Yuan Taur, Mark Rodwell, Lester Eastman Conference, 2008

“Scaling of Nanowire Transistors (invited paper)”, Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, Yuan Taur, IEEE Transaction of Electron Devices, Nov. 2008, pp.2846-2858

“Technology development & design for 22 nm InGaAs/InP channel MOSFETs (invited talk)”,  
Mark Rodwell et. al, Indium Phosphide and Related Material Conference, 2008

"Simulation of Electron Transport in High-Mobility-MOSFETs: Density of States Bottleneck and Source Starvation", M.V. Fischetti, L. Wang, B. Yu, C. Sachs, P. M. Asbeck, Y. Taur, and M. Rodwell, IEDM 2007

"Self Consistent Schrödinger Poisson Solver for III-V Heterostructures Accounting for Conduction Band Non-parabolicity", Lingquan Wang, Peter Asbeck, Yuan Taur, submitted to Solid State Electronics

"Self-Consistent Poisson-Schrodinger Solver Accounting for Non-Parabolic Conduction Band for III-V MOSFET", Lingquan Wang, Peter Asbeck, Bo Yu, Yuan Taur, TechCon 2007.

"A numerical Schrödinger–Poisson solver for radially symmetric nanowire core–shell structures", Lingquan Wang, Deli Wang and Peter M. Asbeck, Solid State Electronics, Nov. 2006, pp 1732-1739

"Analysis of Photoelectronic Response in Semiconductor Nanowires", Lingquan Wang, Peter Asbeck, IEEE Conference on Nanotechnology, 2006

"Pragmatic design of nanoscale multi-gate CMOS, Fossum, J.G.; Wang, L.Q.; Yang, J.W.; Kim, S.H.; Trivedi, V.P., International Electronic Device Meeting, 2004

"Trapped charge induced oxide breakdown," Lingquan Wang, Arnost Neugroschel and Gennadi Bersuker, Journal of Applied Physics, September 2004.

"Method for determination of carrier capture cross-sections at Si/SiO<sub>2</sub> interface," Lingquan Wang and Arnost Neugroschel, Electronics Letters, vol. 40, No.2, pp. 148-149, 22 Jan. 2004.

## FIELD OF STUDY

Major Field: Electrical Engineering

Focused Field: Applied Physics

ABSTRACT OF THE DISSERTATION

Design of Scaled Electronic Devices Based on III-V Materials

by

Lingquan Wang

Doctor of Philosophy in Electrical Engineering (Applied Physics)

University of California, San Diego, 2009  
Professor Peter Asbeck, Chair

One of the challenges faced by the continuous scaling of MOSFETs, is to reconcile the demands for both high speed performance and low power dissipation. The device design space to meet these two criteria using conventional embodiment of the Si bulk MOSFET structure is gradually shrinking, and thus the conventional embodiment may not be able to support the scaling towards the end of the SIA roadmap. In this dissertation, a viable solution from device design perspective to extend the semiconductor device scaling is envisioned, by exploiting the superior material properties and material versatility of the III-V semiconductor

family. On a short term basis, by replacing the Si material by high mobility III-V material (such as InGaAs) may help address the demand for high speed performance. The device involved here continues the theme of planar structure with fully depleted thin body. Various design issues are addressed in the dissertation associated the introduction of the new materials. On the intrinsic device side, much attention has been given to the reduced density of states (DOS) associated with these low effective mass materials. Analysis and discussions are presented here that illustrate device design strategies to circumvent the drawbacks resulting from the small DOS. On the extrinsic side, the incorporation of high-k dielectric has led to non-ideal characteristics at the semiconductor/dielectric interface, which is analyzed in the dissertation to assist in pinpointing the associated technical problem. For further scaling, more attention must be directed to the control of short channel effects (SCEs). Therefore, the natural next step from a planar III-V MOSFET may be a nanowire III-V MOSFET that takes advantage of gate-all-around configuration. However, due to the one dimensional nature of the nanowire, the device performance must be reevaluated with 1-D transport physics. To further proceed, especially to curb the power consumption, it is necessary to scale down the supply voltage, which however undermines performance traditionally. To reconcile these requirements, a tunneling FET based on III-V staggered heterojunctions is introduced. Simulation study has shown that significant current ( $\sim 0.4\text{mA}/\mu\text{m}$ ) may be achieved over a supply voltage of 0.3V with  $10^4$  on-off ratio. In summary, in this dissertation, a variety of feasible solutions towards the end of roadmap from device design perspective are presented.



# Chapter 1. Introduction

## 1.1 Overview of CMOS Scaling

Continuous scaling of CMOS technology over the past three decades, with a great economical success, has driven the semiconductor industry to the current fine feature size of 45nm in mass production [1]. However, as the CMOS technology becomes scaled into the deep submicron regime, significant problems have arisen due to the conflicting demands on higher performance and less power dissipation, namely short channel effects (SCE) control, mobility degradation under high electric field and difficulty of threshold voltage scaling [2]. Addressing these requirements within the current device framework, i.e. bulk silicon technology, while maintaining the pace of further scaling, is extremely challenging. There are two simultaneous aspects of this challenge: 1) to further enhance the performance of the device, i.e. to provide more drive current; 2) to effectively control the power dissipation. To address the first item, material innovation has been incorporated into the past few technology nodes: strained Si and SiGe were introduced into the device to compensate or enhance the otherwise degraded mobility due to the higher field the device will be operated at. Along the other thread, the device design constraint from the power dissipation has become an increasingly difficult challenge to meet. Fundamentally, this problem stems from the fact that the coordinated scaling of power supply voltage along with the vertical and lateral dimensions is no longer possible. Voltage scaling is impeded in that the threshold voltage of the FETs cannot be reduced from present levels ( $V_T \sim 1.25-1.3V$ ) in very large circuits, because of the resulting large subthreshold leakage currents.

In this section, we will review the status the current CMOS scaling, based on the aforementioned perspectives.

### 1.1.1 Mobility Enhancement by Strain Engineering

With the disproportionate scaling of the supply voltage and device geometry dimensions, the electric field at which the devices are operated at have increased significantly along both the transverse and longitudinal directions. The degradation of carrier mobility under higher field, as described by the “universal mobility curve” (shown in Fig. 1.1) [3], then may offset the benefit drawn from the geometrical scaling of the device.

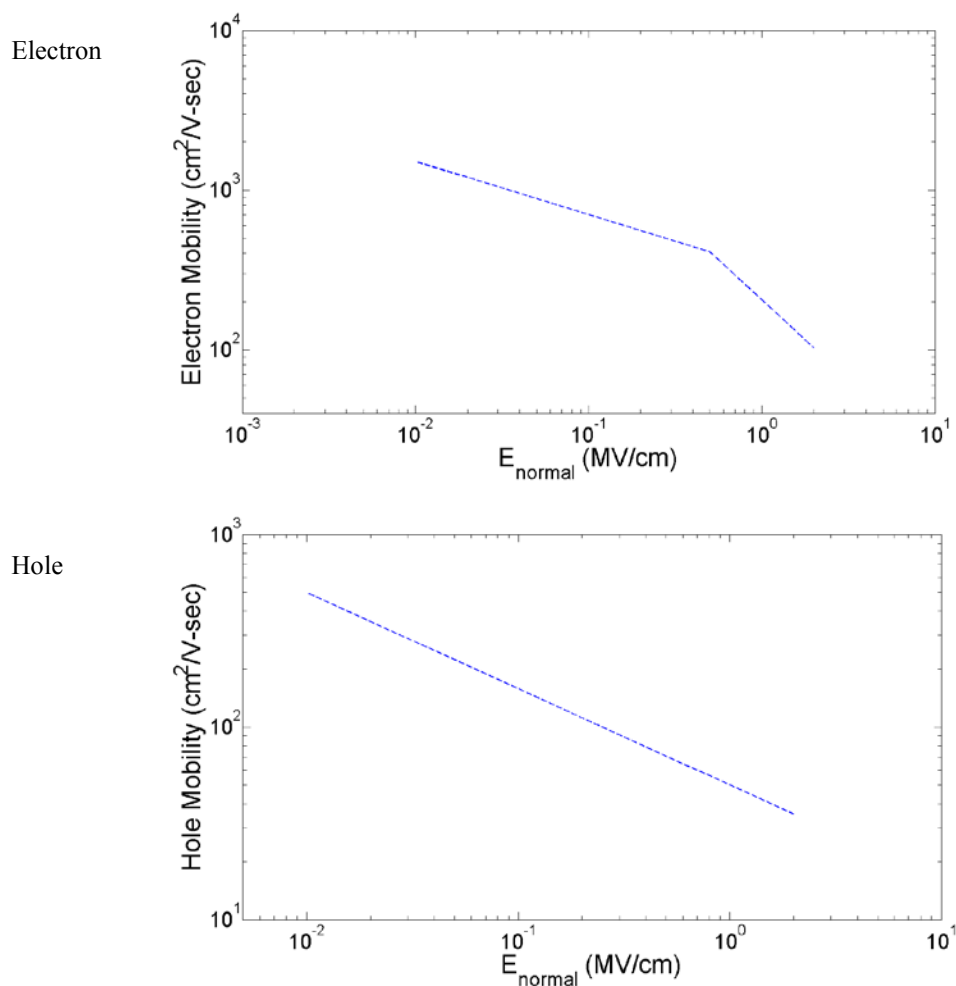


Fig. 1.1. Universal mobility curve for electrons and holes in Si.

To compensate the performance loss due to mobility degradation, strain and stress were introduced to the channel material in the forms of biaxial and uniaxial strain, which were physically implemented by using source/drain stressor and gate dielectric liners [4]. The strained Si technology was first introduced in the 90nm technology node and has been utilized through the current 45nm technology node. Much research, both theoretical and experimental [4-6], has been dedicated to this subject and constitute the current understanding of the stress effects. Presence of the strain field results in shift and split of the conduction band and the valence band. As shown in Fig. 1.2, for Si, application stress splits the 6-fold degeneracy of conduction band into two groups: 2-fold degenerated conduction band minima that are characterized by effective mass  $m_t=1.19m_0$ ; 4-fold degenerated conduction band minima that are characterized by effective mass  $m_l=1.92m_l$ . If tensile stress is applied, the 2-fold band minima assume lower energy position than the 4-fold band minima, while the opposite holds for the case of compressive stress. Therefore, in order to enhance electron mobility, it is preferable to have tensile stress so that the 2-fold band minima are populated with electrons. This further leads to overall smaller effective mass, which favorably provides higher mobility. In practice, the tensile stress was introduced via gate dielectric liner. For valence band, the strain effect is more complicated in that it not only splits the originally degenerated light hole and heavy hole bands, but also introduces significant non-parabolicity to the hole bands. In general, both compressive and tensile stress will result in mobility enhancement. However, only uniaxial compressive stress implemented with SiGe stressor was incorporated in practice. It was reported that with the introduced strain, the mobility of electron and hole may be enhanced by 20%~40% [4-5].

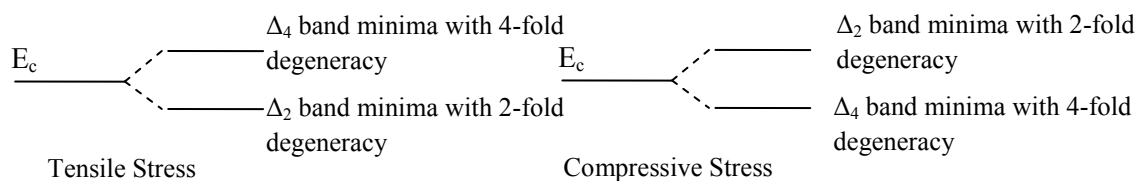


Fig. 1.2. Illustration of the effects on conduction band of Si under stresses of different types.

### ***1.1.2 SCE Control***

With the scaling of CMOS devices, many technology innovations emerged in order to control the SCE, as a mean to contain the off-state leakage. Among the many techniques, the most commonly used solution is via doping profile engineering. For sub-100nm technology nodes, Super-Steep-Retrograde (SSR) junction is commonly incorporated when defining the channel doping profile, through halo implantation [4]. However, as the channel volume becomes smaller due to continuous scaling, discrete dopant fluctuation [7] has become a severe problem that introduces undesirable process variability. To circumvent these problems, novel device configurations such as FinFET, tri-gate, gate-all-around structures have been suggested [8-10]. These devices belong to a more general device family of fully-depleted thin body SOI with typically undoped or lightly doped body. The SCE control within these devices is not through doping profile, but through geometrical confinement instead. Excellent immunity against SCE has been demonstrated experimentally within these structures [8-10] and general scaling limit on these device structures has been predicted [11-12], showing better scalability than bulk counterparts.

### ***1.1.3 Difficulty in Threshold Voltage Scaling***

The difficulty in threshold voltage scaling originates three factors: 1) given supply voltage, the lower the threshold voltage, the higher the on-state current and hence the higher the performance; 2) given subthreshold slope, the lower the threshold voltage, the exponentially higher the drain leakage; 3) the subthreshold slope is limited by  $kT$ , it may not be lower than 60mV/decade at the best under current device conduction mechanism. The joint constraints from these three factors lead to a difficult balance between performance and power consumption in terms of device design. The situation might have been different if the turn-off characteristic of a MOSFET were not specifically tied to ambient temperature, in which case the threshold voltage might be reduced provided the device turned off more rapidly.

Recently reported tunneling transistors [13-15] may provide a viable solution to this dilemma, in that the conduction mechanism of a tunneling transistor has little temperature dependence and therefore in principle may provide much more rapid turn-on and turn-off than the  $kT$ -related process (thermionic emission). However, the material choice for a tunneling transistor merits further consideration, in order to achieve high on-state current and low leakage simultaneously.

## 1.2. Scaled III-V MOSFETs

Previously, the attempt to utilize III-V semiconductor as channel material for MOSFETs, was unsuccessful due to the lack of appropriate gate dielectric with unpinned interface (negligible interfacial trap density). III-V FET devices, in its most commonly available forms (i.e. MESFET and HEMT), operate in depleted mode. The normally-on behavior of such devices has become a show-stopper for usage in digital circuit, due to the excessive power consumption. Furthermore, the gate leakage current of Schottky gate HFETs, when compared to a typical Si MOSFET, is unfortunately much higher. However, significant research interest has been rekindled on III-V material for MOSFET application, after the successful application of the high-k dielectric onto III-V semiconductor surface [16]. III-V materials are subsequently sought after for their known high mobility and celebrated high frequency capability [17-18], in hope of providing the MOSFET device further performance enhancement. On the other hand, similar device structure that are proven to have good SCE control capability in the Si technology may also be applied to the III-V MOSFETs, and thus provides no worse subthreshold leakage for a III-V MOSFET.

The idea of introducing III-V material as channel material is undoubtedly appealing, however, the device design considerations for a III-V MOSFET differs from that of a Si MOSFET significantly. In the following, we describe the background of the challenges faced by III-V MOSFET device design.

### ***1.2.1 The Other Side of the High Mobility***

The high mobility of the III-V material, as a result of their small effective masses, has led to many successful demonstrations of high speed devices. However, on the other side of the glory, small effective masses of these materials also lead to issues that must be addressed for III-V MOSFETs device designs:

1) Small density of states (DOS).

Small DOS manifests its effects in several ways: it limits the sheet charge density at given gate bias; it reduces scattering rate through having less final states for carriers to scatter to. The first effect always counterbalances the advantages provided by the high mobility, as the overall current is related to the product of the sheet charge density and the mobility. The second effect, however, has its positive and negative sides: less scattering in the channel region is preferable in that it results in higher current drive; whereas less scattering in the source is found to limit the current capability of the device, in that the source relies on the scattering process to replenish the electrons that have been injected into the channel. Inadequate scattering within the source slows down the replenish process and leads to a rather unique phenomenon only seen in the III-V MOSFET – known as source starvation.

2) Strong quantum effect.

Small effective mass, when subject to confinement, is always accompanied by more distinct quantization. As a result, in the context of MOSFET, channel that consists of material with small effective mass may expect to have larger wave function spread, which implies less effective gate coupling. This, in addition to the small DOS, limits the sheet charge density at a given gate bias.

### ***1.2.2 Quasi-ballistic Transport***

Continuous scaling of the gate length in MOSFETs gradually closes up the difference between the transport length and the mean free path of the carrier. Experimental results 65nm technology node Si nMOSFET [19] has indicated that the on-state current reached >60% of the corresponding ballistic limit current, which clearly implies the progressively migration from

conventional drift and diffusion to quasi-ballistic for the carrier transport mechanisms. It was further pointed out, through careful Monte Carlo simulation, that MOSFET based on III-V materials are more likely to approach its ballistic limit behavior [20]. Therefore, study III-V MOSFETs in their ballistic and quasi-ballistic mode, and further compare to the Si results become appropriate, for the purpose of benchmarking device performances. It is interesting to note that the advantage of high mobility appears less obvious in the context of (quasi-)ballistic transport, which needs further clarification.

### ***1.2.3 Impact of Non-ideal Dielectrics***

There are many technical factors that forge the great success of silicon CMOS devices; one of them is the near ideal interface between Si and SiO<sub>2</sub>. In contrast, one of the technical difficulties that impeded III-V FET being configured as MOSFET has been gate dielectric with negligible interfacial states. With the recent advance of incorporating high-K dielectric on III-V materials through Atomic Layer Deposition (ALD) [21] or Molecular Beam Epitaxy (MBE) [16], it is shown that it is possible to obtain unpinned interface between dielectric and III-V material. Furthermore, the need of high-K dielectric for a scaled Si MOSFET so as to control aggravated Short Channel Effect (SCE) [22], makes Si MOSFET face the same interface control problem as III-V material had long experienced. Therefore, understanding the interface quality and evaluating its impact to device performance is an important part for the success of III-V MOSFET.

### ***1.2.4 Possible Device Structures of III-V MOSFET***

To control the SCE in scaled MOSFETs, it is necessary to employ geometrical confinement to avoid excessive fluctuation introduced by discrete dopant effect. Given the same



thickness of oxide and body, the most effective geometrical confinement is provided by gate-all-around structure, which allows for the most aggressive scaling. The same concept may also be applied to III-V MOSFET. Experimental results on III-V nanowire core-shell gate-all-around have demonstrated superior control over SCE [23], which further indicated a great scaling potential for these device prototype. However, it is necessary to understand where the device stands in terms of performance, especially in this case of two-dimensional confinement and one-dimensional transport.

### ***1.2.5 Beyond Thermionic Emission Based III-V MOSFET***

As discussed previously, with the further scaling of the devices, the transistor integrated per unit area is expected to increase rapidly, which is however accompanied by the undesired hike in power density. To prevent excess power consumption, it is necessary to scale the power supply accordingly, unfortunately at a compromised performance. The conflict between the power dissipation and performance demand calls for a novel type of device that can reconcile the difficulty and open up space for device design. One of the key issues that hinder the conventional MOSFET from coordinated voltage scaling is the non-scalable  $kT$  related turn-off characteristic. To free the device design from the known 60mV/dec subthreshold swing which is associated with thermionic emission at the source/channel junction, it is perceptible to employ a different conduction mechanism that is independent (or weakly dependent) on the ambient temperature. Tunneling FET, which exploits band to band (BTB) tunneling as the primary conduction mechanism, is likely to suit for such need. However, it is still necessary to examine 1) if such device options also provide advantages on performance, 2) if a specific device embodiment incurs other off-state leakage that masks the advantage of rapid turn on and off brought by the BTB tunneling process. Along the same line, it is to be answered whether a specific type of device design may circumvent the drawbacks that may have arisen from the previous two concerns.

### ***1.2.6 Theme of This Dissertation***

This dissertation attempts to provide, from a device design perspective, analysis on III-V MOSFETs of different configurations that are suitable for corresponding stages of device scaling. On a short term basis, we envision that the planar configuration of III-V MOSFETs may further enhance the performance due to the superior mobility (or ballistic efficiency). Further scaling that requires better SCE control may be satisfied by gate-all-around embodiment of III-V MOSFETs, which is a natural form of nanowire core-shell structure. On a longer term basis, when the tradeoff between power dissipation and performance leads to narrow design space, tunneling FET based III-V staggered heterojunction may set in to reconcile the conflict and continue the saga of scaling.

The following defines the challenges to be addressed in this dissertation in detail:

For III-V planar MOSFETs:

- 1) Impacts of the small effective mass are studied extensively, in terms of resultant sheet charge density and the unique problem of source starvation. These analysis serves to provide device design guidelines to circumvent the adverse effects while maximize the benefits of superior material properties of III-Vs.
- 2) An appropriate ballistic transport model is provided that is capable to take into account the small DOS for III-V semiconductors explicitly and clarify the role of mobility in these circumstances. Performance comparison between representative III-V and Si MOSFET are provided and analyzed in ballistic or quasi-ballistic limit.
- 3) Physical models are developed to understand experimental data on realistic high-k dielectrics on III-V material obtained through experiments. Different aspects of the

non-ideal dielectrics, such as impact of interfacial states, bulk dielectric traps, are analyzed, in the context of III-V semiconductor being the channel material.

For III-V nanowire MOSFETs:

- 1) An analysis is developed for the nanowire based III-V MOSFET in parallel to the planar counterpart, again emphasizing on the impact of small(er) DOS and near ballistic transport properties.
- 2) Optimal material choices for a nanowire MOSFET in order to achieve best ballistic limit performance are discussed.

For III-V tunneling FETs:

- 1) In-depth understanding on BTB tunneling process, in terms of various material choice and band lineup situation is developed.
- 2) Explore the implementation of tunneling FET with staggered heterojunction is explored, and its performance is compared against homojunction embodiments with different materials.

### 1.3 Organization of the Thesis

This dissertation is partitioned into three parts, corresponding to the different stages of scaling as described in the last section.

Part I discusses the planar III-V MOSFET, and further divided into four chapters. Chapter 2 to 4 focus on the intrinsic device design, i.e. assuming ideal dielectric, while Chapter 5 analyzes the impact from the non-ideal dielectric. Detailed arrangement of these four chapters is described as following:

Chapter 2 lays the groundwork for the layer design, provides understanding on how sheet charge density may respond with III-V material as substrate. As mentioned previously, in order to obtain an accurate account for sheet charge density response upon gate bias while addressing the issues of strong quantum effect and conduction band non-parabolicity, self consistent numerical Schrödinger Poisson solver with non-parabolic conduction band taken into consideration was developed first. We subsequently developed analytical formalism of gate capacitance so as to gain physical insight on the impact of the small DOS. Based on these considerations, a viable layer design for a representative III-V MOSFET is presented at the end of the chapter.

Chapter 3 benchmarks the device performance of a III-V MOSFET in its (quasi-) ballistic limit against a Si MOSFET. An improved ballistic transport model that allows for small DOS associated with the III-V material taken into account is first presented. More practical considerations, such as series resistance and carrier reflection, are added to the transport model, to allow for more realistic comparison between the two embodiments. We also discussed the role of mobility in the context of quasi-ballistic transport in this chapter.

Chapter 4 discusses another effect of small DOS in III-V MOSFET, namely the source starvation. The discussion in this chapter is arranged in a progressive way: the concept of the source starvation is introduced first, followed with a simplified phenomenological model to describe the problem in an approximate fashion. A more detailed, physically-based model is further developed to allow for evaluation of the source starvation with more accuracy. Discussion on why source starvation being a unique problem for III-V MOSFET is also included.

Chapter 5 provides correlation and analysis to experimental results. The chapter focuses primarily on the effects from non-ideal gate dielectric. Impact of interfacial states and bulk dielectric traps (which are called border traps in the thesis) is evaluated on both Capacitance Voltage (CV) characteristics and Current Voltage (IV) characteristics. Physically based models were developed and further compared against experimental results (with good agreement).

Part II consists of Chapter 6, discussing the device performance of III-V Nanowire MOSFET.

Chapter 6 is arranged in its discussion in a parallel fashion to Chapter 2 and 3. As most nanowire is cylindrical as-grown, the chapter opens up with the numerical tool that allows for self consistent calculation of Schrödinger and Poisson's equations for cylindrical geometry. Device performance comparison between Si and III-V nanowire MOSFETs in their ballistic limits is then described. An interesting result arisen from the comparison is the role of EOT in these structures: the optimal material choice (in terms of the effective mass, band degeneracy) is highly dependent on the given EOT.

Part III consists of Chapter 7, discussing the device performance of III-V tunneling MOSFET with staggered heterojunction.

Chapter 7 deals with a novel type of FET device, whose primary conduction mechanism is BTB tunneling. The tunneling process itself was discussed first to illustrate the impact of material choice (narrow band gap v.s. high band gap) and band lineup choice (homojunction v.s. heterojunction) on the tunneling probability. Built upon these understandings, a baseline design of TFET with staggered heterojunction was presented and simulated with 2-D Sentaurus simulator. Design tradeoffs in terms of material choices, device doping strategy and device geometry were subsequently discussed.

Chapter 8 provides an overall summary for the entire thesis.

## Reference:

1. K. Mistry, et. al., "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging", IEDM, 2007.
2. D. A. Antoniadis, et. al., "Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations", IBM Journal of Research and Development, Vol. 50, No. 4/5, 2006, pp. 362-376.
3. Y. Taur, T. Ning, "Fundamentals of Modern VLSI Devices", Cambridge University Press, 2009
4. S. Thompson et. al., "A 90-nm Logic Technology Featuring Strained-Silicon", IEEE Transaction on Electron Devices, vol. 51, Nov. 2004, pp. 1790-1797
5. Kern (Ken) Rim, Judy L. Hoyt and James F. Gibbons, "Fabrication and Analysis of Deep Submicron Strained-Si N-MOSFET's", IEEE Transaction on Electron Devices, Vol. 47, No. 7, 2000, pp. 1406-1415.
6. Shin-ichi Takagi, Judy L. Hoyt, Jeffrey J. Welser, and James F. Gibbons, "Comparative study of phonon-limited mobility of two-dimensional electrons in strained and unstrained Si metal-oxide-semiconductor field-effect transistors", J. Appl. Phys., Aug. 1996, pp. 1567-1577
7. Hon-Sum Philip Wong, Yuan Taur, David J. Frank, "Discrete random dopant distribution effects in nanometer-scale MOSFETs", Microelectronics and Reliability, vol. 38, Sep. 1998, pp. 1447-1456
8. D. Hisamoto et. al., "FinFET-A Self-Aligned Double-Gate MOSFET Scalable to 20 nm", IEEE Transaction on Electron Devices, Vol. 47, No. 12, 2000, pp. 2320-2325.
9. B. S. Doyle, S. Datta, M. Dockzy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully depleted tri-gate CMOS transistors," IEEE Electron Device Lett., vol. 24, no. 4, pp. 263-265, Apr. 2003.
10. J. P. Colinge, "Multiple-gate SOI MOSFETs", Solid State Electronics, Vol 48, No. 6, 2004, pp. 897-905
11. Kunihiro Suzuki, Tetsu Tanaka, Yoshiharu Tosaka, Hiroshi Horie and Yoshihiro Arimoto, "Scaling Theory for Double-Gate SOI MOSFETs", IEEE Transactions on Electron Devices, vol. 40, pp.2326-2329, Dec. 1993
12. Bo Yu, Lingquan Wang, Yu Yuan, Asbeck, P.M., Yuan Taur,"Scaling of Nanowire Transistors", , IEEE Transactions on Electron Devices, vol. 55, pp. 2846-2858, Nov. 2008

13. J. Appenzeller, Y-M Lin, J. Knoch, Z. Chen, and P Avouris, "Comparing Carbon Nanotube Transistors---The Ideal Choice: A Novel Tunneling Device Design," IEEE Trans. Electr. Dev., vol. 52, pp. 2568-2576, Nov. 2005
14. O.M. Nayfeh, C. Ni Chleirigh, J.L. Hoyt and D.A. Antoniadis, "Measurement of Enhanced Gate-Controlled Band-To-Band Tunneling in Highly Strained Silicon-Germanium Diodes," IEEE Electron Device Letters, vol. 29, pp. 468-470, May 2008
15. Woo Young Choi, Byung-Gook Park, Jong Duk Lee, Tsu-Jae King Liu, "Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec," IEEE Electron Device Letters, vol. 28, pp. 743-745, August 2007
16. Karthik Rajagopalan, Jonathan Abrokwah, Ravi Droopad and Matthias Passlack, "Enhancement-Mode GaAs n-Channel MOSFET", IEEE Electron Device Letters, vol. 27, pp. 959-962, Dec 2006
17. Otfried Madelung Ed., "Semiconductors Basic Data", 2<sup>nd</sup> Edition, Springer, 1996
18. R. Lai, et. al., "Sub 50 nm InP HEMT Device with Fmax Greater than 1 THz", IEDM, 2007
19. P. Bai, et al., "A 65nm Logic Technology Featuring 35nm Gate Lengths, Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-k ILD and 1.57  $\mu$ m<sup>2</sup> SRAM Cell," International Electron Devices Meeting, pp. 657, 2004.
20. P. M. Solomon and S. E. Laux, "The Ballistic FET: Design, Capacitance and Speed Limit," IEDM Tech. Digest, pp. 511-514, 2001
21. P. McIntyre et. al., "Interface studies of ALD-grown metal oxide insulators on Ge and III-V semiconductors", Microelectronics Engineering, Vol. 86, No.7-9, 2009, pp. 1536-1539
22. SIA Roadmap
23. Bryllert T, Samuelson L, Jensen L, Wernersson L. "Vertical high mobility wrap-gated InAs nanowire transistor". DRC proceedings, 2005, pp. 157-158.



# **PART I: Thin Body Planar III-V MOSFET**

# Chapter 2. Layer Structure Design of a III-V MOSFET, Quantum Effects and Non-parabolicity

## 2.1 Introduction

A significant task of developing a suitable III-V MOSFET for VLSI application involves providing an appropriate high-K dielectric on III-V semiconductor surface. However, in addition, different device design strategies and focus need to be adopted for a III-V MOSFET due to the basic material property differences from Si. This chapter contains the very first considerations when designing a III-V MOSFET – appropriate layer structure design to deliver a respectably high current drive.

The notion of developing III-V MOSFETs is based on superior III-V semiconductor material properties, namely the high mobility that many III-V semiconductors possess. It is to be realized that the high mobility of III-V material, which originates from the small effective mass, does not come at no penalty. In fact, the small mass (as well as smaller degeneracy), leads to lower Density of States (DOS), (which is proportional to various powers of the effective mass under different dimensionalities<sup>1</sup>). Since the current density is the product of average carrier velocity and carrier concentration, III-V material does not automatically become superior in performance when compared to Si, in that it lags in the carrier concentration although excels in the velocity. Therefore, to carry out a feasible design of III-V MOSFET, one must evaluate the dependences of velocity and carrier concentration on the material properties, i.e. the electronic structure. Conventional treatment of semiconductor utilizes a single parameter – effective mass and assumes that the value of it prevails over the energy range of interest. This has been a good

---

<sup>1</sup> This argument has assumed parabolic Conduction Band structure such that a single effective mass prevails at all energies. A refined argument is presented in the later part of this chapter.

approximation for Si<sup>2</sup>, however it is not an accurate description for III-V materials. As a result of the smaller effective mass (which is assumed to be a function of energy here) for III-V semiconductors, the magnitude of DOS is limited as a function of energy (even more so if “parabolic effective mass” is used). In order to provide enough carriers so that the current drive is maximized, electrons must be present in a larger range of energy, as compared to Si. As electrons start to populate higher energy states, the electronic structure at these energies is no longer adequately described by the “parabolic effective mass”. The discrepancy may be better visualized by comparing electron population with E-k relationships assuming parabolic band or non-parabolic band (Fig. 2.1). If we vary the Fermi energy position, it is apparent that the higher the Fermi energy, the larger the discrepancy between parabolic band assumption (dashed line) and the more realistic non-parabolic band description (solid line). In this regard, a better description that captures the non-parabolicity of the Conduction Band must be included for accurate assessment of given layer designs.

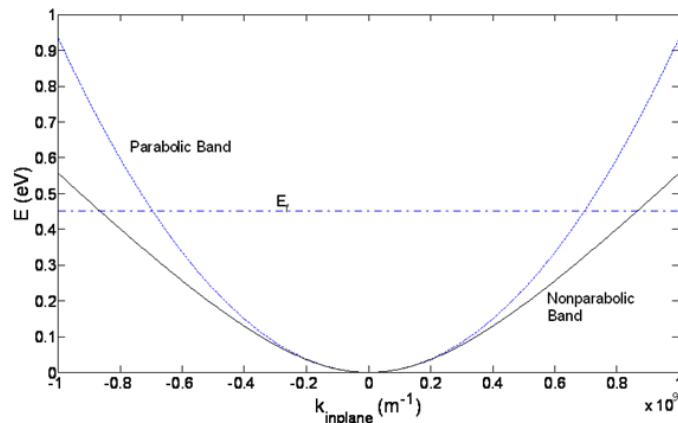


Fig. 2.1. A comparison on E-k dispersions between parabolic band description and non-parabolic band description.

An additional design concern is related to quantum mechanical effects on electron confinement. SCE controls for the advanced technology nodes require electron confinement

<sup>2</sup> Under normal gate bias conditions, the Fermi level of Si may not be swung deeply into the Conduction Band.

(either by electric field or by structural confinement) in the direction perpendicular to the current transport. Given the smaller effective masses of III-V materials, the quantum correction involved is perceivably more evident than that for Si. Again, since we have set off to assess a particular layer structure design, the quantum effect may not be overlooked.

Given the above arguments, it is quite clear that a self-consistent solver, which accounts for both non-parabolic Conduction Band and quantum confinement, must be in place to allow for accurate evaluation once a layer structure design is specified. Relevant calculations include C-V characteristics of the structure, since one of the objectives for an appropriate design is to obtain as much sheet charge density as possible. Therefore, developing a detailed understanding of the CV characteristics became a logically step as one designs the layer structure for a III-V MOSFET.

The content of this chapter is arranged in accordance to the considerations shown above. In the first part, a self-consistent 1-D Schrödinger Poisson solver which accounts for non-parabolic Conduction Band will be developed. A capacitance model specific to the situation where quantum confinement is present is introduced in the second part of this chapter.

## 2.2 Self-consistent 1-D Schrödinger Poisson Solver with Conduction Band Non-parabolicity

### *2.2.1 Objectives*

The layer structure design of III-V MOSFET requires an understanding of the band structure and corresponding quantum states for different epitaxial structures. The 1-D Poisson-Schrödinger solver reported in Ref. [1] has been very useful to evaluate such prototype device structures. However, for many III-V materials, especially the widely used high mobility Indium-containing ternaries (e.g. InGaAs), appreciable  $\Gamma$  valley non-parabolicities may cause the calculation based on a parabolic band assumption to be inaccurate. Such inaccuracy becomes pronounced when the non-parabolic parameter itself is large, and additionally, when the bulk effective mass is small. The objectives of a self-consistent Schrödinger Poisson solver that accounts for non-parabolic Conduction Band include: 1) Energies and corresponding wave functions of the eigen-states; 2) DOS function for individual subband as a function of energy derived from computed E-k dispersion relationship. It will be later noted that both of these mentioned results differ considerably from the values inferred from a parabolic band assumption.

Effects of non-parabolic band structure have been previously analyzed [2] and quantified using different methods. The perturbation method reported in [3] may only be suitable for relatively low kinetic energy electrons, while it loses its accuracy in higher energy regime. The plane wave expansion method reported in [4] is able to yield an accurate answer at a cost of high computational complexity, since a sizable number ( $>100$ ) of basis functions are typically needed. The solver we developed here intends to provide a direct method to solve the Schrödinger equation and Poisson's equation simultaneously, at a lower computational cost, while taking into

account the non-parabolicity, to provide appropriate DOS function for each subband and thus ensure correct electron distribution calculations.

This section is arranged as follows: the methodology of incorporating non-parabolicity into the self consistent calculation procedure is presented in 2.2.2, followed by the model verification in 2.2.3. In 2.2.4, the results are discussed from a physically intuitive prospective. 2.2.5 shows an extension to the current methodology for simulations involving multi-valley, which becomes an increasingly important aspect for III-V MOSFET design.

### ***2.2.2 Methodology***

The method proposed here uses successive approximations to solve the Schrödinger equation with non-parabolic band structure. The solution procedure is illustrated with the block diagram shown in Fig. 2.2. As it has become a common practice to discretize the 1D structure and solve the Schrödinger and Poisson equations with parabolic band structure, we will not repeat this procedure in detail. Readers are referred to previous works, e.g. [1].

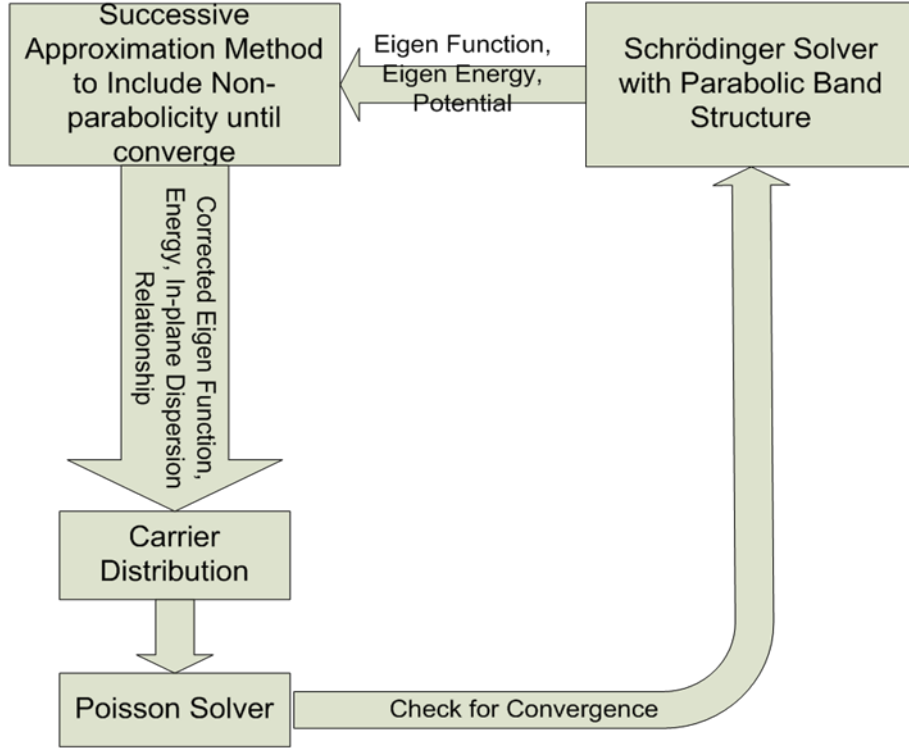


Fig. 2.2. Schematic of the self-consistent solving procedure of Schrödinger equation and Poisson's equation including non-parabolic conduction band.

We focus here how the non-parabolic band structure can be accounted for when solving the Schrödinger equation. We assume the relation between in-plane kinetic energy and in-plane wave vector can be adequately captured by the simple relationship [5]:

$$\hat{T}(1 + \alpha\hat{T}) = \frac{\hbar^2}{2m^*} \left( -\frac{\partial^2}{\partial z^2} \right) + \frac{\hbar^2 K_{xy}^2}{2m^*} \quad (2.1)$$

where  $\hat{T}$  is the kinetic energy operator,  $K_{xy}$  is the in-plane wave vector (we assume that the quantum confinement is aligned with z axis),  $m^*$  is the effective mass (bulk material parameter) and  $\alpha$  is the non-parabolicity parameter which varies with material. As becoming evident later, it

is important that we include explicitly the in-plane kinetic energy term  $\frac{\hbar^2 K_{xy}^2}{2m^*}$  here due to the

following two reasons: 1) the corresponding wave functions are influenced by the in-plane wave

vectors; 2) the corresponding eigen energies are also impacted, which results in a different E-k relationship from which the DOS is derived.

By  $\hat{T} = \hat{H} - \hat{V}$ , where  $\hat{H}$  and  $\hat{V}$  denote the full Hamiltonian and potential operator, we can further write:

$$(\hat{H} - \hat{V})(1 + \alpha\hat{H} - \alpha\hat{V}) = \hat{H} - \hat{V} + \alpha\hat{H}^2 - \alpha\hat{H}\hat{V} - \alpha\hat{V}\hat{H} + \alpha\hat{V}^2 = \frac{\hbar^2}{2m^*} \left(-\frac{\partial^2}{\partial z^2}\right) + \frac{\hbar^2 K_{xy}^2}{2m^*} \quad (2.2)$$

Let  $|\varphi_s\rangle$  be a solution to the Schrödinger equation  $\hat{H}|\varphi_s\rangle = E_s|\varphi_s\rangle$  with corresponding eigenenergy  $E_s$ . One can apply the operators on both sides of (2.2) to this wave function, which yields:

$$(1 + \bar{\alpha}\hat{H})\hat{H}|\varphi_s\rangle = \left[\frac{\hbar^2}{2m^*} \left(-\frac{\partial^2}{\partial z^2}\right) + \frac{\hbar^2 K_{xy}^2}{2m^*} + \hat{V}(1 + 2\bar{\alpha}E_s) - \alpha\hat{V}^2\right]|\varphi_s\rangle \quad (2.3)$$

where  $\bar{\alpha} = \langle\varphi_s|\alpha(z)|\varphi_s\rangle$  with position dependent non-parabolicity parameter that corresponds to different material layers.

(2.3) can be further written as:

$$(1 + \bar{\alpha}E_s)E_s|\varphi_s\rangle = \left[\frac{\hbar^2}{2m^*} \left(-\frac{\partial^2}{\partial z^2}\right) + \frac{\hbar^2 K_{xy}^2}{2m^*} + \hat{V}(1 + 2\bar{\alpha}E_s) - \langle\varphi_s|\alpha\hat{V}^2|\varphi_s\rangle\right]|\varphi_s\rangle \quad (2.4)$$

(2.4) can be converted into an iterative calculation process, as shown below:

$$(1 + \bar{\alpha}E_s^{(n)})E_s^{(n+1)}|\varphi_s^{(n+1)}\rangle = \left[\frac{\hbar^2}{2m^*} \left(-\frac{\partial^2}{\partial z^2}\right) + \frac{\hbar^2 K_{xy}^2}{2m^*} + \hat{V}(1 + 2\bar{\alpha}E_s^{(n)}) - \langle\varphi_s^{(n)}|\alpha\hat{V}^2|\varphi_s^{(n)}\rangle\right]|\varphi_s^{(n+1)}\rangle \quad (2.5)$$

where the superscripts denote the iteration indices.



If an initial guess is made of the eigenenergy and eigen wave function, and a recursive calculation prescribed by (2.5) is carried out, then if the process converges, it should yield a correct answer to the original Schrödinger equation. A reasonable initial guess may be obtained from the solution of self consistent solution to Schrödinger and Poisson equation where only parabolic band structures are considered. It typically takes less than 10 iterations for the successive approximation to converge.

For each in-plane wave vector, a corresponding eigen energy is obtained via the aforementioned iteration. Thus, an in-plane E-k relationship can be established for each subband. The DOS can be subsequently obtained with the non-parabolic dispersion relationship:

$$g(E) = \frac{1}{\pi} K_{xy} \left( \frac{dE}{dK_{xy}} \right)^{-1} \quad (2.6)$$

The total electron density can readily be derived as:

$$n = \sum_{\text{subband}} |\varphi_s|^2 \int_{E_{\text{sub}}}^{\infty} f(E) g(E) dE \quad (2.7)$$

where  $f(E)$  is the Fermi distribution in equilibrium,  $E_{\text{sub}}$  is the bottom energy of the subband in consideration.

The Poisson equation can be solved accordingly with the information of mobile electron distribution, and an iterative calculation loop is formed. Typical iterations number is around 15.

It is noteworthy that during the derivation presented above, we have ignored the fact that the operators  $\hat{H}$  and  $\hat{V}$  do not commute. To assess the error introduced by this approximation, we also solved self-consistently the Schrödinger equation and Poisson's equation using a similar but expanded method as presented in [4], taking into account the position/material dependent

non-parabolic parameter. For the representative calculation presented below, the difference between the two methods is  $\sim 3\%$  for the sheet charge density v.s. voltage characteristics, suggesting that the approximation adopted here gives reasonable accurate results, at a much lower computational cost.

Shown in Figure 2.4 is a representative result produced from the solver, with simulated structure schematically presented in Figure 2.3. The corresponding material parameters are compiled in Table 2. The structure features InGaAs channel and AlGaSb top and back barriers. The large conduction band offset ( $\sim 2.2\text{eV}$ ) between AlGaSb and InGaAs, compared to AlGaAs/InGaAs barriers, allows for larger sheet charge density to be achieved within the channel before the spillover takes place. The solver is capable of calculating self consistently the in-plane dispersion relationship (Figure 2.4a), electron distribution (Figure 2.4b), electrostatic potential (Figure 2.4c), all with conduction band non-parabolicity included. With conduction band non-parabolicity taken into account, electron density is considerably increased with respect to the prediction derived from the parabolic band assumption. Such increase implies that larger sheet charge density may be expected under the same bias condition. Plotted in Fig. 2.5 is a comparison of sheet charge density versus bias with and without non-parabolicity effect.

Table 2.1. Parameter values taken from [6, 7]

	In <sub>0.53</sub> Ga <sub>0.47</sub> As	InP	Al <sub>0.7</sub> Ga <sub>0.3</sub> Sb
$m^*/m_0$	0.041	0.079	0.18
$\alpha$ (eV <sup>-1</sup> )	2.22	0.63	N/A
$X$ (eV)	4.66	4.4	3.41
$\epsilon_r$	14.0	12.4	12.95
$E_g$ (eV)	0.78	2.42	2.75



Figure 2.3. Schematic of layer structure computed.

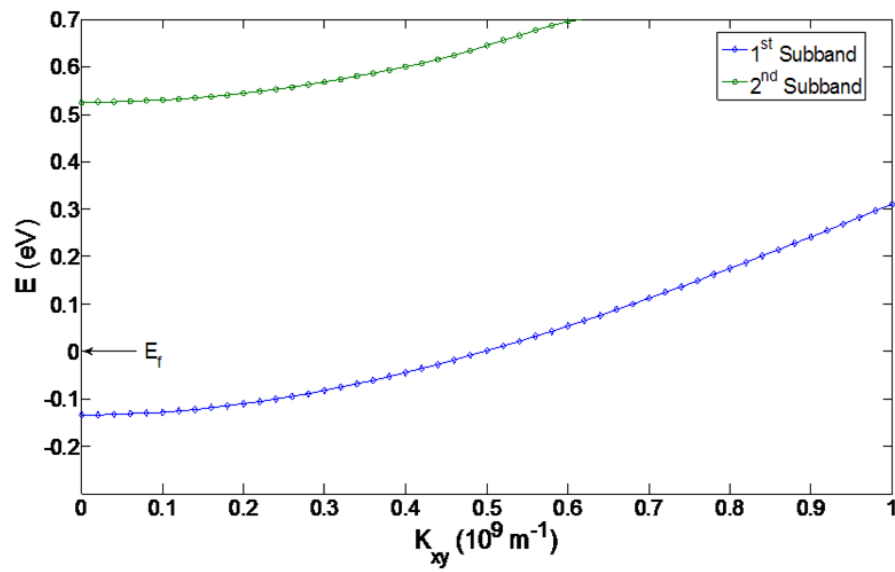


Fig. 2.4. (a)

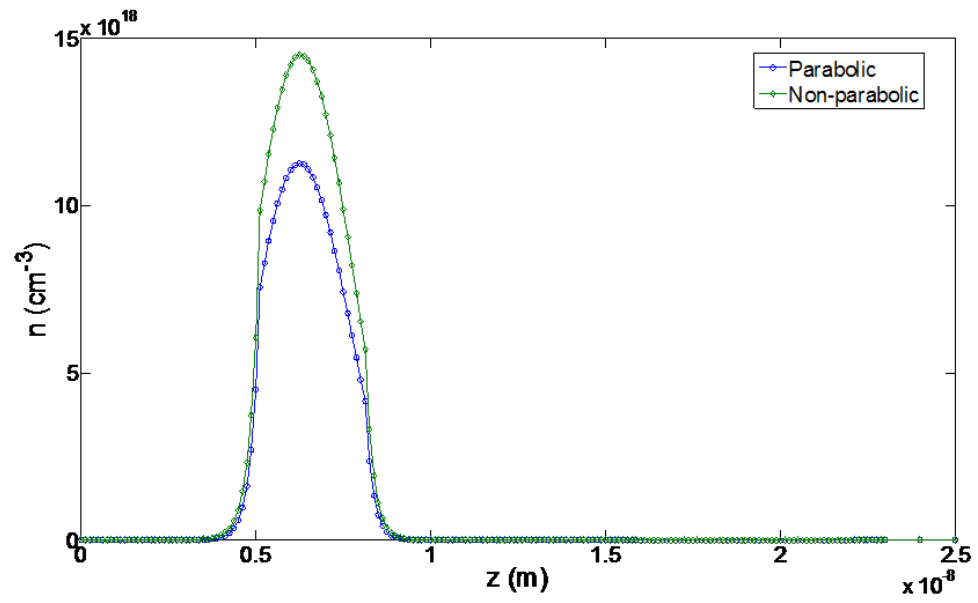


Fig. 2.4 (b)

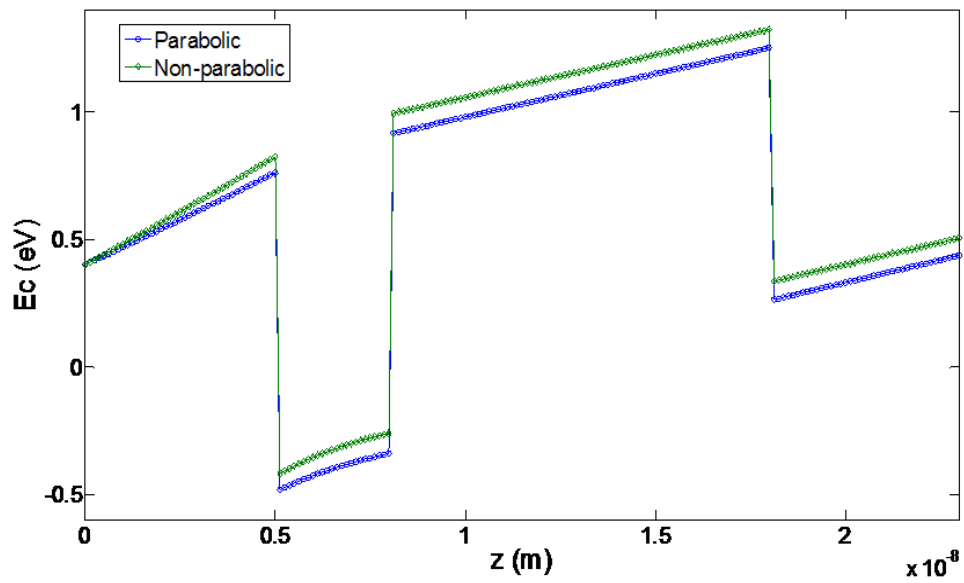


Fig. 2.4 (c)

Fig. 2.4. Computation result for the layer structure shown in Fig.2: a) In-plane dispersion relationship of the 1<sup>st</sup> and 2<sup>nd</sup> subbands; b) Electron distribution with parabolic conduction band assumption as well as non-parabolic conduction band condition; c) Self-consistent electrostatic potential profile.

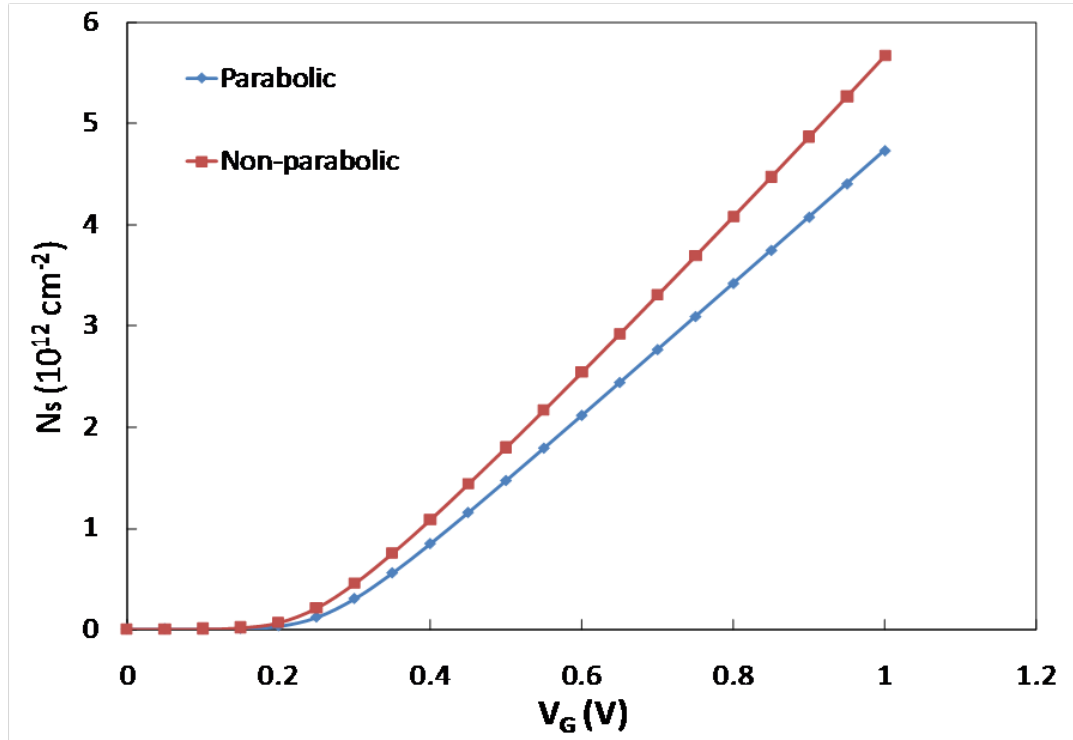


Fig. 2.5. Sheet charge density as a function of the gate bias under parabolic band condition and non-parabolic conduction band condition

### 2.2.3 Model Verification

To verify the model, we check our calculation against results obtained from the analytical formulation in [8], which is valid for low kinetic energy. The simulated structure of a GaAs quantum well sandwiched between AlGaAs layers is recapitulated in the inset of Figure 2.6. Comparison is made between the DOS's as a function of energy, as plotted in Figure 2.6. We notice that the numerical result agrees closely with the analytical formula for the first subband, whereas it deviates appreciably for the higher subband. This discrepancy can be attributed to the differences in the corresponding in-plane dispersion relationships, which are plotted in Figure 2.7 (where lines represent the numerical results while symbols correspond to the analytical formula). The deviation found in the elevated energy range stems from the fact that the analytical

formulation approximates the 2-D in-plane dispersion relationship with the bulk non-parabolic dispersion relationship. This, however, presumes that the z-direction eigen wave functions for electrons within same subband but with different kinetic energies behave exactly the same, which is not true for quantum well structures (see next section for details). In confined structures, the in-plane wave vector dependent wave function has its own dependence of eigen energies on  $K_{xy}$ 's; moreover, our computation shows that the E-k dispersion for the 2<sup>nd</sup> (or higher) subband is not a simple translational shift of that of the 1<sup>st</sup> subband, as assumed in the analytical model. Our numerical result suggests the effective mass for the second subband is 4.8% higher than predicted in [8] (where the 2<sup>nd</sup> subband E-k relationship is assumed to be a simple translational shift of that of the 1<sup>st</sup> subband).

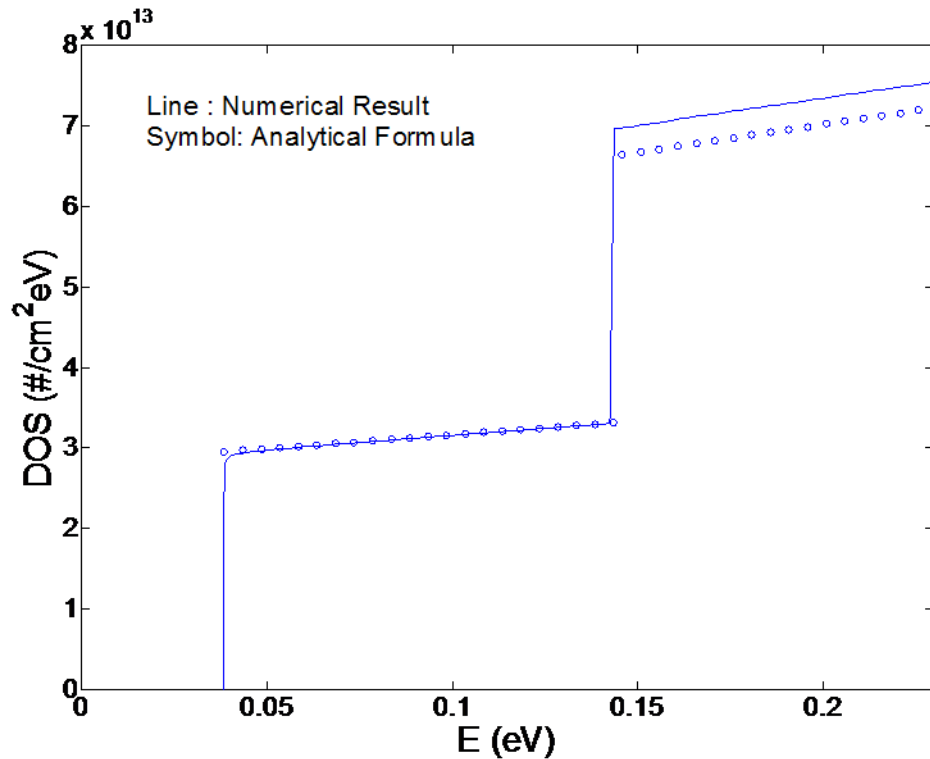


Fig. 2.6. Comparison between analytical and numerical result for the structure shown in the inset.

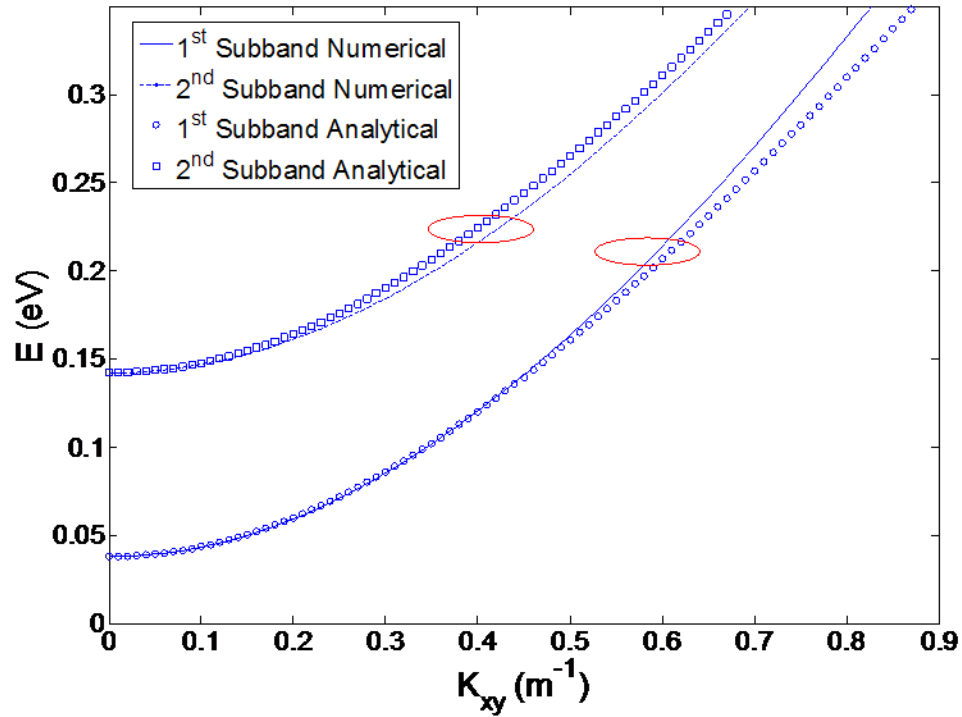


Fig. 2.7. Comparison between the in-plane dispersion relationship between analytical and numerical results.

#### 2.2.4. Results and Discussion:

As shown previously, the inclusion of non-parabolicity, leads to a larger sheet electron density at a given applied gate voltage,  $V_g$ . We attribute this increase to the increased DOS as a function of energy compared to the classical staircase scenario predicted from the parabolic band structure. Differences between the DOS functions for parabolic and non-parabolic band structures are shown in Figure 2.8.

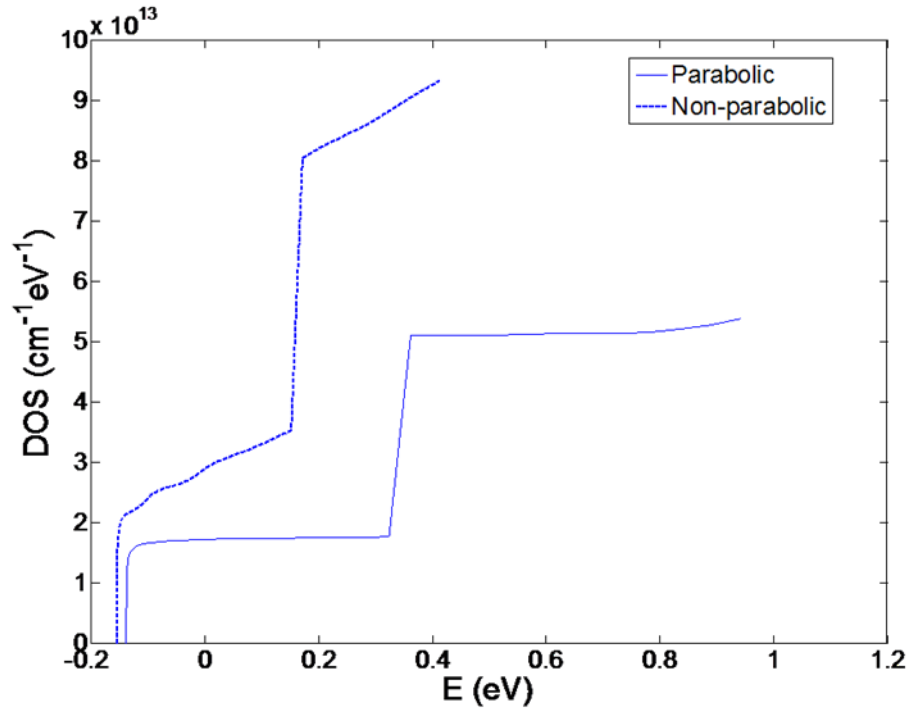


Fig. 2.8. Computed Density of States (DOS) as a function of energy. Energy 0 here reflects the position of electron Fermi level.

For the parabolic band case, it can be seen that at low energy, the DOS function follows the classically predicted staircase shape, whereas at elevated energy, a slight deviation is observed. The deviation can be attributed to the varying eigen wave function as a function of energy level. Shown in Figure 2.9, is a series of representative eigen wave functions with different in-plane wave vectors (therefore different eigen energies). For small in-plane vectors, wave functions overlap one another, and are mostly confined within the channel region (i.e. InGaAs region); whereas with increased in-plane wave vector, the wave functions start to extend in both directions into the regions where effective masses are larger (i.e. AlGaSb regions). The gradual change in wave functions can be understood as follows: with different in-plane wave vectors (eigen energies), the confinement (i.e. energy band offset and effective mass change) “seen” by the electrons at the regional boundaries are different. As a result, the associated reflection coefficients at these boundaries vary as a function of in-plane wave vector, and thereby



individual wave functions are shaped differently. It is also intuitive that the average effective mass will be influenced by how the wave function is distributed (even within the same subband), because of the different wave function distribution, and therefore the DOS function experiences a gradually changing average effective mass that deviates from the classical predicted staircase.

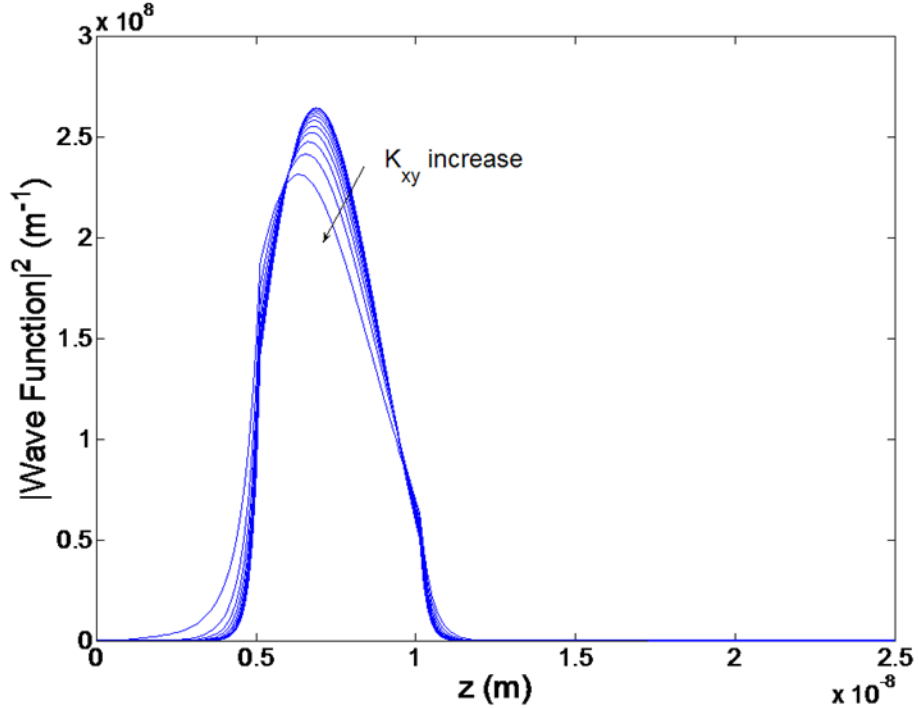


Fig. 2.9. Evolution of the 1<sup>st</sup> subband wavefunction as in-plane kinetic energy increases.

For non-parabolic band structure, the presence of additional deviations of the E-k dispersion relationship from the parabolic case leads to the more pronounced increase in the DOS function v.s. energy [8]. It is also noted that with non-parabolic band structure, the eigen-energy shifts downward by an appreciable amount. A direct way to view this is by rewriting (2.4) under the condition of zero in-plane kinetic energy (correspond to the subband bottom):

$$E_s|\varphi_s\rangle = \left[ \frac{\hbar^2}{2m^*(1+\bar{\alpha}E_s)} \left( -\frac{\partial^2}{\partial z^2} \right) + \hat{V} \frac{(1+2\bar{\alpha}E_s)}{(1+\bar{\alpha}E_s)} - \frac{1}{(1+\bar{\alpha}E_s)} \langle \varphi_s | \alpha \hat{V}^2 | \varphi_s \rangle \right] |\varphi_s\rangle \quad (2.8)$$

One notices two effects here with the rewritten Schrödinger equation with non-parabolicity: 1) the effective potential changes from  $\hat{V}$  to  $\hat{V}(1 + \frac{\bar{\alpha}E_s}{1 + \bar{\alpha}E_s}) - \frac{\langle \varphi_s | \alpha \hat{V}^2 | \varphi_s \rangle}{1 + \bar{\alpha}E_s}$ ; 2) the “effective” effective mass changes from  $m^*$  to  $m^*(1 + \bar{\alpha}E_s)$ . The first effect has minor effect in regard to the subband energy shift, as the additional  $\hat{V}(\frac{\bar{\alpha}E_s}{1 + \bar{\alpha}E_s})$  term partially cancels with  $\frac{\langle \varphi_s | \alpha \hat{V}^2 | \varphi_s \rangle}{1 + \bar{\alpha}E_s}$ . The second effect can cause downshifted eigen energies, as the effective mass is “effectively” increased by a factor of  $(1 + \bar{\alpha}E_s)$ . It is worthy of mentioning that such effect will be more pronounced as one increases the In content within the InGaAs, which leads to larger  $\bar{\alpha}$  value.

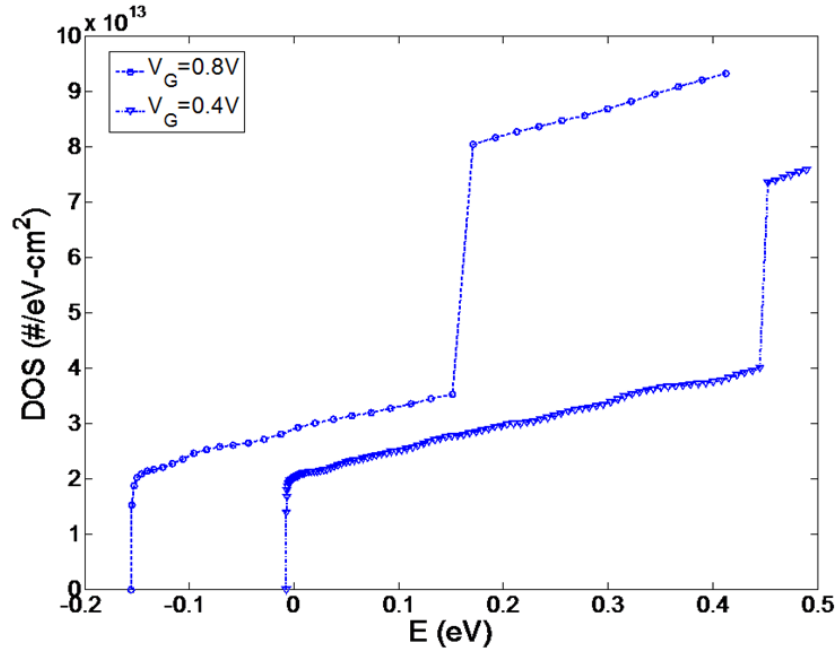


Fig. 2.10. Computed DOS function under different gate bias (i.e. different electron occupation within the quantum well).

As a result of self-consistently calculated quantum states, the density of states varies with gate bias due to different electron occupations. The presence of different electron concentration results in altered potential distribution, consequently different eigen-energies and in-plane dispersion as well. As shown in Figure 2.10, different density of states behavior is exhibited under different gate bias.

The downshifted subband energies as well as increased DOS values compared with the parabolic case clearly point to an encouraging effect for III-V FET performance: increased sheet charge density under the same  $V_g$  bias. As shown in Figure 2.5, the sheet charge density is increased by  $\sim 20\%$  under the “on condition” ( $V_G=2.0V$ ) for the non-parabolic case compared to the parabolic case. Consequently, the on state DC current drive can be expected to increase, along with an increased input capacitance. Fig. 2.11 shows the impact of larger density of states results to overall input capacitance. For comparison, both parabolic and non-parabolic cases are plotted; the geometrical barrier capacitance is also provided as a reference.

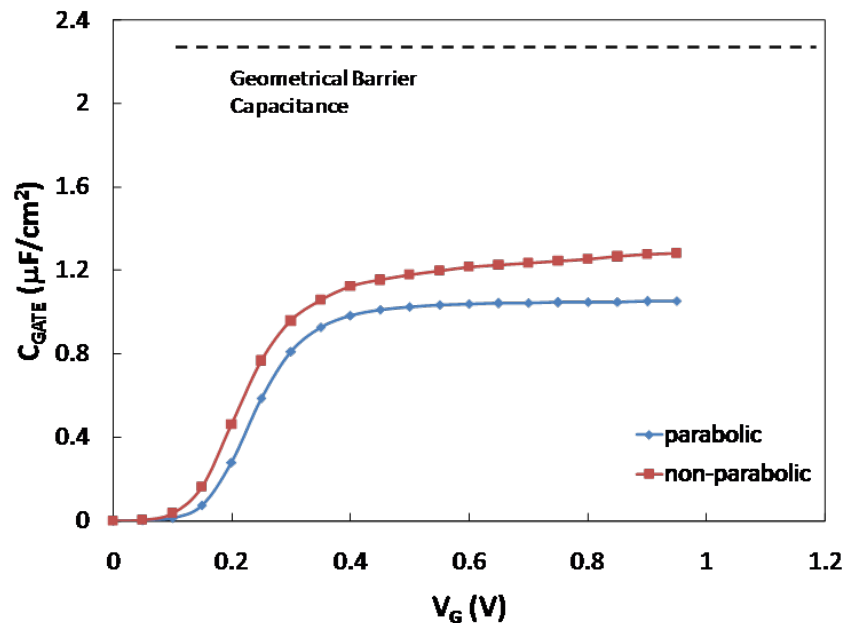


Fig.2.11. Derived quasi-static capacitance voltage characteristic for both parabolic and non-parabolic conduction band condition.

### 2.2.5. Model Extension

For III-V compound semiconductor transistors, a general concern is whether there is appreciable occupation of the L-valleys of the conduction band, which is detrimental in many ways (e.g. low mobility). The population of L-valleys may result under two circumstances: 1) electrostatically, i.e. excessively high  $V_G$  can cause occupation of L-valleys; 2) kinetically, i.e. electrons may gain enough energy from the longitudinal field and be scattered into L-valleys. Our model can evaluate the first case. The  $\Gamma$ -L valley separation is taken to be 0.63eV for InGaAs[9]. It is worthy of mentioning that the effective mass parameters for the L valley are more complicated than the  $\Gamma$  valley, in that anisotropic bulk parameters have to be incorporated into the calculation. We assume here a (100) layer growth direction, which leads to the different effective masses used for solving the Schrödinger equation (effective mass projection to  $z$  direction) and for the DOS function (effective mass projection to  $x$ - $y$  plane). Four equivalent valleys are considered when calculating the electron density.

Plotted in Fig. 2.12 and Fig. 2.13 are the self-consistent results when taking into account the L-valley electron occupation. It is noticed that at  $V_G=2.0V$ , negligible L-valley occupation is observed, whereas once the gate voltage exceeds  $V_G=2.4V$ , appreciable amount (2%) L-valley occupation can result, which is undesirable. This sets a limit on the supply voltage in certain device structures. Calculations here are based on parabolic band structure for the L valleys. As a result of the large effective masses of the L-valleys, we consider the electrons will mostly distributed close to the bottom of the bands, which will mitigate the calculation inaccuracy of not taking into account the non-parabolicity associated with the L-valleys.

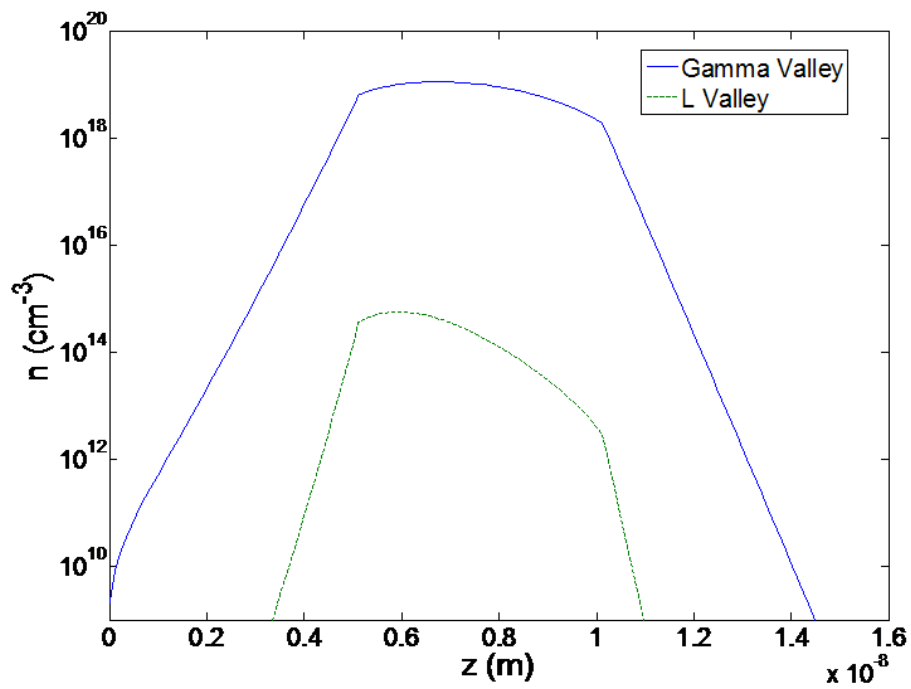


Fig. 2.12. Electron occupation within  $\Gamma$  valley and L valley under gate bias of 2.0V, with negligible L valley occupation. Computations include  $\Gamma$  valley non-parabolicity.

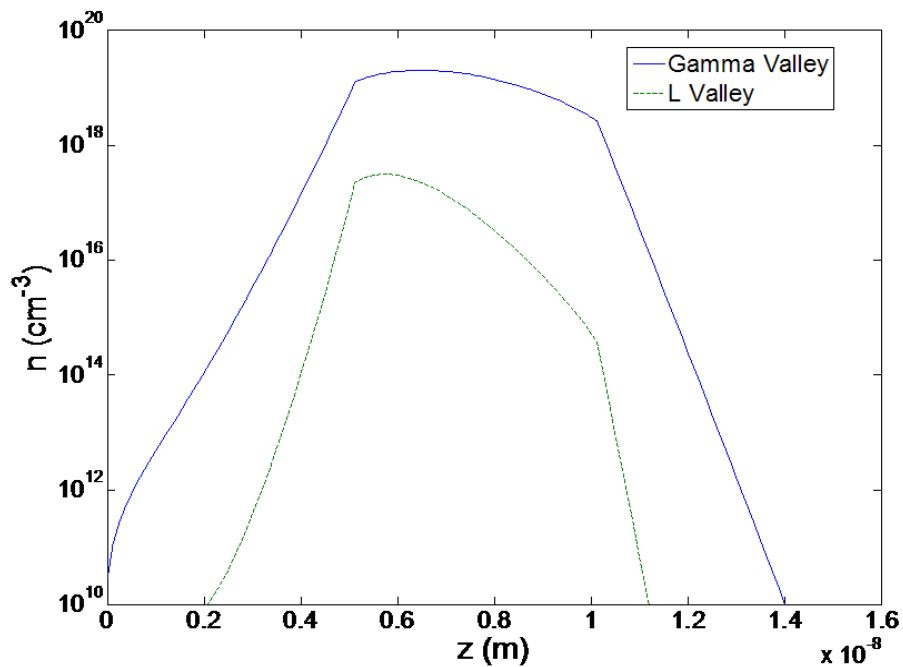


Fig. 2.13. Electron occupation within  $\Gamma$  valley and L valley under gate bias of 2.4V, with L valley occupation accounts for  $\sim 2\%$  the overall sheet charge density. Computations include  $\Gamma$  valley non-parabolicity.

### ***2.2.6. Further Discussion and Future Work***

In this section, we would like to provide further discussions of the scenarios where conduction band non-parabolicity becomes important and must be included to avoid gross error in design estimations. From the material perspective, it is obvious that if the magnitude of the material non-parabolicity parameter is significant (which is the case for most Indium contained high mobility III-V materials, e.g. InGaAs, InP, InAs), the non-parabolic Conduction Band should be considered. However, other factors also weigh in on this issue, namely channel layer thickness and (oxide) barrier height. As shown previously in Fig. 2.1, the higher the energy electrons populate, the more significant the non-parabolic effect. Under the scenario where channel layer thickness is designed to be thin enough (to address the SCE control), significant energy quantization occurs and electron population starts off at an elevated energy with respect to the bottom of the conduction band. As the layer thickness scales down (to enable ultra-short gate length), this quantization effect directly leads to significant non-parabolic deviation at the energy-wise starting point where electrons start to populate. However, it must be noticed that this argument has assumed that at least one bound state exists within the quantum well. This implies that there must be high enough barriers on both sides of the quantum well (also for sake of obtaining a large sheet charge density within the channel quantum well). In case of the previous computation example, we deliberately chose a band lineup that gives more than 1eV Conduction Band offset to ensure no spill-over occurs under the gate bias range we use. In an ideal III-V MOSFET, where oxide (high-k dielectric) is to be deposited on top of the channel quantum well, the barrier criterion is typically satisfied. In summary, we conclude that non-parabolicity becomes critical to be taken into account when 1) channel material is characterized by a large non-parabolic parameter; 2) channel quantum well is thin (the definition of “thin” depends on

individual material,  $\sim 5\text{nm}$  for high mobility III-V material); 3) sufficient barriers are present to confine the quantum well.

In an effort to develop a successful design for III-V MOSFET, we strive to obtain large sheet charge density so that the superior transport properties of III-V materials may have a suitable impact on current. Transport considerations will be the subject of following chapter; we would like to focus our discussion here on the understanding of sheet charge density. With the consideration of non-parabolicity, we have seen that the sheet charge density can be expected to increase appreciably ( $\sim 20\%$  at  $V_{gs}$  values corresponding to the “on” condition of the FETs), which implies better DC on current and transconductance. The cause of such increase is the modified DOS functions due to non-parabolicity. We noticed that although the DOS is increased by a factor of almost 2.5~2X at the energy range of interest, this increase is diluted by the presence of the layer stack. This effect will be further discussed in the next section where an equivalent capacitance model is derived.

In the last part of this section, multi-valley simulation was described. In a realistic III-V MOSFET, the previous results must be taken with suitable qualification. Given different available substrate, the strain and stress effects may lead to very different band offset as well as Gamma-L, Gamma-X separations. It is also worthwhile to consider designing a III-V MOSFET deliberately with the inclusion of L valley and X valley, to provide greater sheet charge density, while retaining reasonable transport properties along the current conduction direction by proper orientation selection [13].

The work described in this section is primarily derived from the paper “Self Consistent 1D Schrödinger Poisson Solver for III-V Heterostructures Accounting for Conduction Band Non-parabolicity”, submitted to “Solid State Electronics”, with the dissertation author being the primary author.

## 2.3 Capacitance Analysis of a III-V MOSFET Layer Structure Design

### ***2.3.1 Objective***

As we mentioned in the previous section, it is necessary to obtain reasonably large sheet charge density in a III-V MOSFET design for III-V materials to reveal their potent on superior transport properties. Although an assessment of the sheet charge density as a function of applied gate bias may be obtained with precision using the method described in Sec. 2.2, it is still worthwhile to develop more insights into the overall capacitance (esp. when the device is “on”) with different layer structure design. In this section, we will focus on the influences that are specifically due to material properties of III-V (small masses and stronger quantum effects) on the overall capacitance voltage characteristics. The section will be arranged as following: in 2.3.2, a first order model will be derived under several approximations, to provide intuitive insights on various concepts (e.g. DOS capacitance, quantum well capacitance); in 2.3.3, we will develop a more detailed model that entails more rigorous derivation; 2.3.4 will provide a brief summary.

### ***2.3.2 A First Order Capacitance Model***

In this section, we will develop a first order capacitance model, aiming to introduce concepts and lay groundwork for further introduction on more rigorous derivation that does not rely on some of the phenomenological descriptions we stipulate in this section.

To illustrate physical concepts, we consider a III-V bulk-like FET whose structure and corresponding energy band diagram are schematically shown in Fig. 2.14. In the following derivation, we assume:

1. Strong quantum confinement, so that only one subband contributes to the sheet charge density.



2. Front barrier (oxide) provides sufficient energy barrier so that the wave function associated with the first subband is well contained within the semiconductor.
3. The channel material is undoped, i.e. charge on the gate only corresponds to the channel charge.

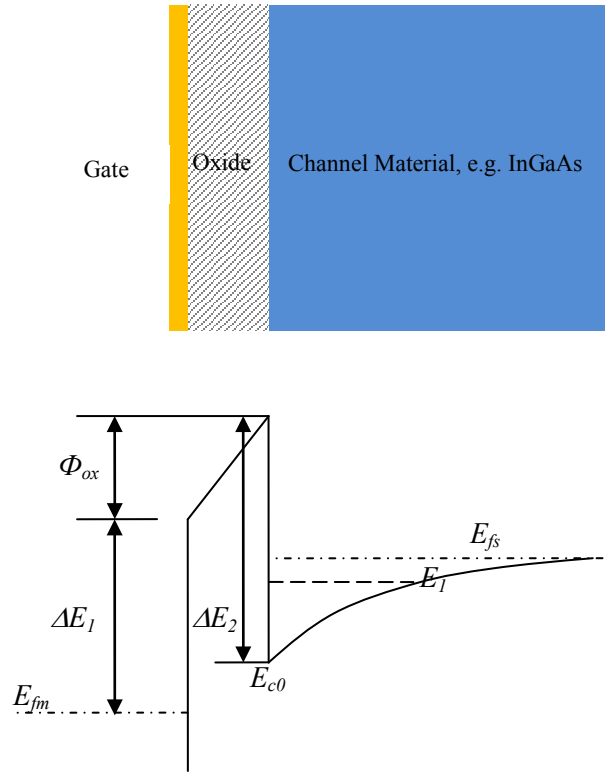


Fig. 2.14. Schematic illustration of a III-V MOSFET layer structure and its corresponding energy band diagram.  $E_{fm}$ ,  $E_{fs}$  are Fermi energies of the gate and the semiconductor, respectively.  $E_1$  is the energy of the first quantized state.

We start the derivation by writing down the relationships between energies, as illustrated above:

$$E_{fm} + \Delta E_1 + q\Phi_{ox} - \Delta E_2 + (E_1 - E_{c0}) + (E_{fs} - E_1) = E_{fs} \quad (2.9)$$

where  $E_{c0}$  is the Conduction Band bottom at the oxide/semiconductor interface,  $\Delta E_1$  and  $\Delta E_2$  are constant energy offsets. As the applied gate bias is related to the difference of the Fermi level between the gate and the semiconductor, Eq. (2.9) may be further written as:

$$qV_{gs} = q\Phi_{ox} + \Delta E_2 - \Delta E_1 + (E_1 - E_{c0}) + (E_{fs} - E_1) \quad (2.10)$$

By taking derivatives on both sides of Eq. (2.10), while noticing that the potential drop across the oxide is related to the sheet charge density by  $\Phi_{ox} = qN_s/C_{ox}$ , we have,

$$\begin{aligned} q\delta V_{gs} &= q^2\delta N_s / C_{ox} + \delta(E_1 - E_{c0}) + \delta(E_{fs} - E_1) \\ \Rightarrow \frac{1}{C_g} &= \frac{\delta V_{gs}}{q\delta N_s} = \frac{1}{C_{ox}} + \frac{\delta(E_1 - E_{c0})}{q^2\delta N_s} + \frac{\delta(E_{fs} - E_1)}{q^2\delta N_s} \end{aligned} \quad (2.11)$$

As seen from Eq. (2.11), the overall gate capacitance is composed of three parts, for a bulk like structure. We will subsequently analyze each term on the right hand side of Eq. (2.11). The first term is clearly the oxide capacitance (also known as the geometrical barrier capacitance). The understanding of the second and third term relies on better understanding on how the sheet charge density is related to various quantities. For the simplified case where only one subband is occupied, the volume charge density can be written as:

$$n(x) = |\psi(x)|^2 \cdot \int_{E_1}^{+\infty} f(E, E_{fs}) \cdot g(E) \cdot dE \quad (2.12)$$

where  $\psi(x)$  is the eigen wave function that is associated with the state. The total sheet charge density, with assumption 3, is only related to the volume electron density by the following:

$$\begin{aligned}
N_s &= \int_{x=0}^{+\infty} [|\psi(x)|^2 \bullet \int_{E_1}^{+\infty} f(E, E_{fs}) \bullet g(E) \bullet dE] dx = (\int_{x=0}^{+\infty} |\psi(x)|^2 dx) \bullet \int_{E_1}^{+\infty} f(E, E_{fs}) \bullet g(E) \bullet dE \\
&= \int_{E_1}^{+\infty} f(E, E_{fs}) \bullet g(E) \bullet dE
\end{aligned} \tag{2.13}$$

The last step of the above derivation utilizes assumption 2, such that the integration of wave function within the semiconductor is unity. To obtain physical insight without involved math, we further assume zero temperature condition.

If derivative is taken with respect to  $(E_{fs}-E_1)$  on both sides of Eq. (2.13), under zero temperature assumption, the derivative of Fermi distribution becomes a delta-function (assuming significant overdrive, i.e. Fermi level over the bottom of the Conduction Band):

$$\frac{dN_s}{d(E_{fs} - E_1)} = \int_0^{+\infty} \delta(E - E_{fs}) \bullet g(E) \bullet dE = g(E_{fs}) \tag{2.14}$$

Therefore, the third term of (2.11) is clearly related to the DOS function value at Fermi energy of the semiconductor. For this reason, we define DOS capacitance at  $T=0K$  as

$$C_{DOS} \equiv \frac{q^2 \delta N_s}{\delta(E_{fs} - E_1)} = q^2 g(E_{fs}) \tag{2.15}$$

If we further assume parabolic band structure, the DOS function is a constant for 2-DEG occupying one subband, with a value of  $\frac{m^*}{\pi \hbar^2}$ . Comparing this term between Si and III-V yields a sharp contrast: due to the larger effective mass and 6-valley degeneracy, the DOS function assumes a much larger value than that of III-V. Therefore, DOS capacitance in a Si MOSFET typically does not significantly limit the overall gate capacitance, whereas DOS capacitance in III-V is more important due to its small magnitude. As seen in Section 2.2, due to non-

parabolicity in the Conduction Band, the  $C_{DOS}$  is expected to be larger. Nevertheless, it is still a limiter of the overall achievable capacitance. A comparison on the relative importance of individual components will be provided at the end of this section with the approximated model and Section 2.3.3.

The second term of Eq. (2.11) characterizes the change of the first eigen-state energy with respect to the bottom of the Conduction Band at the interface, in response to the change of overall charge density. The link between the two quantities may be established as following: upon the change An approximate treatment of this term may be done by assuming the potential profile close to the semiconductor/oxide interface to be of triangular shape. The electric field strength is given by  $\frac{qN_s}{\epsilon}$ . The eigenenergies of a triangular well is known to be related to the zeroes of Airy function

[10]:

$$(E_1 - E_{c0}) = -a_1 \left( \frac{q^2 F^2 \hbar^2}{2m^*} \right)^{1/3} \quad (2.16)$$

where  $a_1$  is the first zero of an Airy function,  $F$  is the electric filed strength. By relating the electric field strength to the sheet charge density, one can write down an explicit expression of the second term in Eq. (2.11) as following:

$$C_{qw} \equiv \frac{\delta(E_1 - E_{c0})}{q^2 \delta N_s} = -\frac{2}{3} a_1 \left( \frac{\hbar^2}{q^2 \epsilon^2} \bullet \frac{1}{2m^* N_s} \right)^{1/3} \quad (2.17)$$

We name this portion of the capacitance “quantum well” capacitance. It is clear from Eq. (2.11) that the overall capacitance for the simplified bulk MOSFET structure consists serial combination of  $C_{ox}$ ,  $C_{DOS}$  and  $C_{qw}$ . With (2.15) and (2.17), we can now compare the relative importance of the various components of the overall gate capacitance. For better visualization, the comparison will

be conducted in the following means: each capacitance component will be converted to an “EOT” thickness via:

$$C_i = \frac{\epsilon_{SiO_2}}{EOT} \quad (2.18)$$

where  $C_i$  represents one of the various components for the overall capacitance. For purpose of illustration, in the following comparison, we assume that the gate oxide has an effective oxide thickness of 1nm, and the channel material is lattice relaxed  $In_{0.53}Ga_{0.47}As$ . We examine the situation when  $6 \times 10^{12} \text{ cm}^{-2}$  sheet charge density is achieved. The comparison is provided in Fig. 2.15, where both parabolic band situation as well as (approximate) non-parabolic band situation is illustrated. The overall bar length represents the EOT associated with the overall gate input capacitance, with each contributor (e.g. oxide, quantum well capacitance, DOS capacitance) labeled with different color. The length is drawn to scale. It is seen for III-V material with small

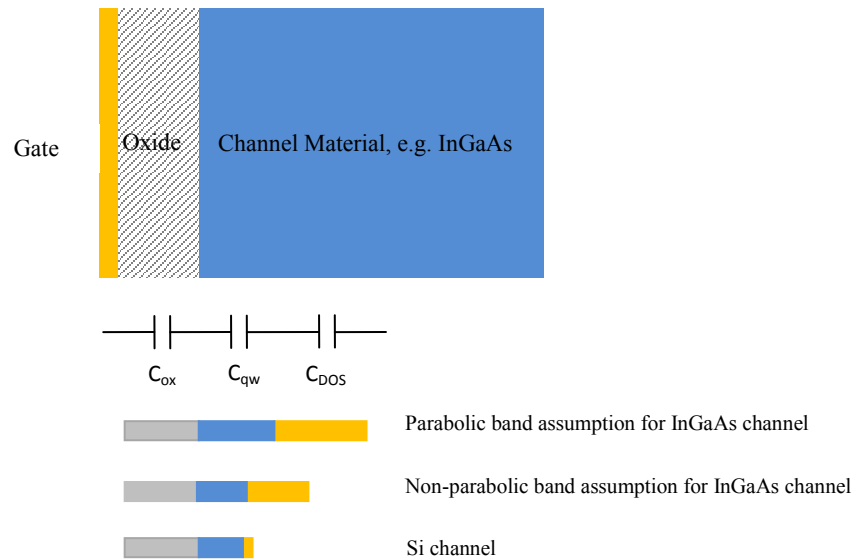


Fig. 2.15. Breakdown of contributions to the overall input capacitance from 1) oxide (high k dielectric) as gray bar, 2) quantum well capacitance due to the wave function spread as blue bar, 3) density of state capacitance as orange bar.

effective mass, that the overall input capacitance is very much dependent on the quantum effect

and the capacitance arisen from the small density of state. The later factor is almost negligible for Si MOSFET, due to the 2 valley degeneracy as well as the lager effective mass ( $0.19m_0$ ).

### ***2.3.3 A General Capacitance Model***

In this section, a more elaborate capacitance model will be presented to detailed breakdown each component of the overall capacitance. We will not limit ourselves to the assumption (3) as stated in the previous section, however still consider the case where only one eigen energy state (degenerated states) is (are) populated. The basic concepts are the same as described in Section 2.3.2, except that the more elaborated model dives into more rigorously defined quantities. Towards the end, we will provide some semi-quantitative argument for cases where multiple states are occupied.

This model is jointly developed by Bo Yu, Yu Yuan, Prof. Peter Asbeck, Prof. Yuan Taur and myself.

In the first part of this section, we will derive a constraint among different charge components within a MOSFET configuration. In the second part, we will deep dive into the individual capacitance component in relation to the physical quantities such as wave functions, DOS.

We shall start with the Poisson's equation along the gate stack direction, as depicted in Fig. 2.14 (except we do not assume substrate doping to be zero, moreover, back gate scenario is also treated here). To simplify the mathematical treatment, we assume all the materials we encounter here possess the same dielectric constant, denoted by  $\epsilon_s$ .

$$\frac{d^2\psi}{dx^2} = \frac{q}{\epsilon_s} [n(x) - p(x) + N'(x)] \quad (2.19)$$

where  $N'(x)$  is the net doping ( $N_a - N_d$ ) as a function of depth. The origin of the depth is defined at the interface between semiconductor and the front oxide. For the last two terms in bracket on the right hand side of Eq. (2.19), it has a physical meaning of depletion charge (for an nMOSFET) within the depletion region (the boundary of the depletion region is labeled as  $W_m$ ). We will, from now on, denote the depletion charge as  $N(x)$ . Eq. (2.19) is rewritten as:

$$\frac{d^2\psi}{dx^2} = \frac{q}{\epsilon_s} [n(x) + N(x)] \quad 0 < x < W_m \quad (2.20)$$

Integrating Eq. (2.20) yields:

$$\int_x^{W_m} \left( \frac{d^2\psi}{dx^2} \right) dx' = \frac{d\psi}{dx} \Big|_{x=W_m} - \frac{d\psi(x)}{dx} = \int_x^{W_m} \frac{q}{\epsilon_s} [n(x') + N(x')] dx' \quad (2.21)$$

$$\frac{d\psi}{dx} = \frac{d\psi}{dx} \Big|_{x=W_m} - \int_x^{W_m} \frac{q}{\epsilon_s} [n(x') + N(x')] dx'$$

The term  $\frac{d\psi}{dx} \Big|_{x=W_m}$  corresponds to, according to Gauss's Law, the back boundary electric field,

which can be expressed as  $\frac{Q_m}{\epsilon_s}$ , where  $Q_m$  is the sheet charge at the substrate metal plate. By this

treatment, we generalize the derivation to situations including bulk structure as well as thin body/double gate structure. The potential drop over the range from  $x=0$  to  $x=W_m$  may be found by integrate Eq. (2.21):

$$\int_0^{W_m} \frac{d\psi}{dx} dx' = \psi(W_m) - \psi(0) = \int_0^{W_m} \left( \frac{Q_m}{\epsilon_s} - \int_x^{W_m} \frac{q}{\epsilon_s} [n(x') + N(x')] dx' \right) dx'' \quad (2.22)$$

By integration by part, Eq. (2.22) can be further written as:

$$\psi(W_m) - \psi_s = \frac{Q_m W_m}{\epsilon_s} - \frac{q}{\epsilon_s} \int_0^{W_m} x \bullet n(x) dx - \frac{q}{\epsilon_s} \int_0^{W_m} x \bullet N(x) dx \quad (2.23)$$

where  $\psi_s$  denotes the electrostatic potential at the surface (not referred to bulk potential). To obtain the capacitive component, we differentiate Eq. (2.23) with respect to  $\psi_s$  on both sides:

$$-1 = \frac{W_m}{\epsilon_s} \frac{\partial Q_m}{\partial \psi_s} - \frac{x_{av}}{\epsilon_s} \frac{\partial}{\partial \psi_s} \int_0^{W_m} q \bullet n(x) dx - \frac{W_{av}}{\epsilon_s} \frac{\partial}{\partial \psi_s} \int_0^{W_m} q \bullet N(x) dx \quad (2.24)$$

where  $x_{av}$  and  $W_{av}$  are defined through the following relationships:

$$\begin{aligned} x_{av} &= \int_0^{W_m} x \bullet \frac{\partial n(x)}{\partial \psi_s} dx / \int_0^{W_m} \frac{\partial n(x)}{\partial \psi_s} dx \\ W_{av} &= \int_0^{W_m} x \bullet \frac{\partial N(x)}{\partial \psi_s} dx / \int_0^{W_m} \frac{\partial N(x)}{\partial \psi_s} dx \end{aligned} \quad (2.25)$$

From Eq. (2.25), it is clear that the quantities  $x_{av}$  and  $W_{av}$  have the physical meaning of centroid of differential inversion charge and differential depletion charge, respectively.

If we further define the following quantities:

$$\text{Total inversion charge: } Q_i = - \int_0^{W_m} q \bullet n(x) dx$$

$$\text{Total depletion charge: } Q_d = - \int_0^{W_m} q \bullet N(x) dx$$

We can rewrite Eq. (2.24) as:

$$-1 = \frac{W_m}{\epsilon_s} \frac{\partial Q_m}{\partial \psi_s} + \frac{x_{av}}{\epsilon_s} \frac{\partial Q_i}{\partial \psi_s} + \frac{W_{av}}{\epsilon_s} \frac{\partial Q_d}{\partial \psi_s} \quad (2.26)$$



Eq. (2.26) reveals the relationship among surface potential and different charge components (back contact charge, inversion charge, depletion charge), with each of them correspond to a capacitive component. We define them as:

$$C_m = -\frac{\partial Q_m}{\partial \psi_s} \quad C_i = -\frac{\partial Q_i}{\partial \psi_s} \quad C_d = -\frac{\partial Q_d}{\partial \psi_s} \quad (2.27)$$

Therefore, we obtain a constraint among various charge (capacitance) components:

$$1 = \frac{W_m}{\epsilon_s} C_m + \frac{x_{av}}{\epsilon_s} C_i + \frac{W_{av}}{\epsilon_s} C_d \quad (2.28)$$

It is noted that Eq. (2.28) is generalized expression, which is applicable to bulk structure, thin film SOI and double gate structure.

So far we have obtained a general relationship among various capacitances, we will now inspect individual capacitance in regard to physical quantities.

Similar to what has been established in the previous section, with the generalized model, the total gate input capacitance is:

$$C_g = \frac{\partial Q_g}{\partial V_g} = \frac{\partial Q_g}{\partial \psi_{ox} + \partial \psi_s} = \frac{1}{\frac{1}{\partial Q_g / \partial \psi_{ox}} + \frac{1}{\partial Q_g / \partial \psi_s}} = \frac{1}{1/C_{ox} + 1/C_s} \quad (2.29)$$

where  $C_s$  stands for the capacitance contribution from semiconductor. Given the charge neutrality condition,  $Q_g + Q_d + Q_i + Q_m = 0$ , the semiconductor capacitance can be further decomposed into the summation of  $(C_i + C_d + C_m)$ .

In the previous section, we have treated  $C_i$  as it was assumed that inversion charge is the only charge that exists inside the semiconductor. It is also shown that  $C_i$  is composed of two capacitances that are defined as DOS capacitance and quantum well capacitance, respectively. Therefore, according to the generalized derivation, the overall capacitance configuration is given in Fig. 2.16.

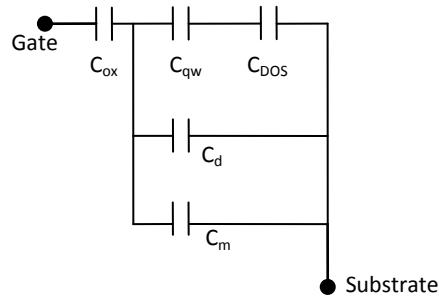


Fig. 2.16. Equivalent circuit for generalized gate capacitance model.

To obtain a more general physical insight, we will not assume 1) significant gate overdrive and 2) triangular potential well in the following derivation. However, to obtain a closed form solution, we will have to assume parabolic dispersion throughout the layer stack.

As defined in Section 2.3.2, DOS capacitance is defined as:

$$C_{DOS} \equiv -\frac{q\partial Q_i}{\partial(E_{fs} - E_1)} \quad (2.30)$$

In case of 2DEG, the sheet charge density may be analytically obtained assuming parabolic band dispersion relationship.

$$Q_i = q \frac{m^*}{\pi \hbar^2} kT \ln[1 + \exp(\frac{E_f - E_1}{kT})] \quad (2.31)$$

Substitute Eq. (2.31) into Eq. (2.30), we obtain

$$C_{DOS} \equiv q^2 \frac{m^*}{\pi \hbar^2} \frac{1}{1 + \exp(\frac{E_{fs} - E_1}{kT})} \quad (2.32)$$

Eq. (2.32) is valid for bias range from accumulation to inversion. Under the condition of strong gate overdrive, Eq. (2.32) reduces to Eq. (2.15) under T=0K.

We will now probe further into the quantum well capacitance. As defined previously, the quantum well capacitance is given as:

$$C_{qw} \equiv - \frac{\partial Q_i}{\partial(\frac{E_1 - E_{c0}}{q})} = - \frac{\partial Q_i / \partial \psi_s}{\partial(\frac{E_1 - E_{c0}}{q}) / \partial \psi_s} = \frac{C_i}{\partial(\frac{E_1 - E_{c0}}{q}) / \partial \psi_s} \quad (2.33)$$

We denote the wave function associated with the first eigen energy state as  $\Phi_1$ . The term

$\partial(\frac{E_1 - E_{c0}}{q}) / \partial \psi_s$  can be further derived, according to first order perturbation theory:

$$\partial(\frac{E_1 - E_{c0}}{q}) / \partial \psi_s = \langle \phi_1 | \frac{\partial(\frac{\hat{H} - E_c}{q})}{\partial \psi_s} | \phi_1 \rangle = \langle \phi_1 | \frac{\partial(\psi_s - \psi(x))}{\partial \psi_s} | \phi_1 \rangle \quad (2.34)$$

If we further assume that  $Q_m=0$ , we have

$$\psi_s - \psi(x) = \frac{q}{\epsilon_s} [\int_0^x x' n(x') dx' + \int_0^x x' N(x') dx' + x \int_x^{W_m} n(x') dx' + x \int_x^{W_m} N(x') dx'] \quad (2.35)$$

We shall treat term by term the derivative with respect to  $\psi_s$  in the following.

$$\begin{aligned}
\langle \phi_1 | \frac{\partial \int_0^x x' n(x') dx'}{\partial \psi_s} | \phi_1 \rangle &= \int_0^{W_m} |\phi_1(x)|^2 \left( \int_0^x x' \frac{\partial n(x')}{\partial \psi_s} dx' \right) dx \\
&= \int_0^{W_m} \frac{n(x)}{\int_0^{W_m} n(x) dx} \left( \int_0^x x' \frac{\partial n(x')}{\partial \psi_s} dx' \right) dx \\
&= \int_0^{W_m} x \frac{\partial n(x)}{\partial \psi_s} dx - \int_0^{W_m} x \frac{\partial n(x)}{\partial \psi_s} \left( \int_0^x \frac{n(x')}{\int_0^{W_m} n(x) dx} dx' \right) dx
\end{aligned}$$

$$\langle \phi_1 | \frac{\partial \int_0^x x' N(x') dx'}{\partial \psi_s} | \phi_1 \rangle = \int_0^{W_m} \frac{n(x)}{\int_0^{W_m} n(x) dx} \left( \int_0^x x' \frac{\partial N(x')}{\partial \psi_s} dx' \right) dx = 0$$

In the above equation, we used the approximation that the inversion charge is physically separated from the depletion charge.

$$\langle \phi_1 | \frac{\partial (x \int_x^{W_m} n(x') dx')}{\partial \psi_s} | \phi_1 \rangle = \int_0^{W_m} \frac{\partial n(x)}{\partial \psi_s} \left( \int_0^x \frac{x' n(x')}{\int_0^{W_m} n(x) dx} dx' \right) dx$$

$$\langle \phi_1 | \frac{\partial (x \int_x^{W_m} N(x') dx')}{\partial \psi_s} | \phi_1 \rangle = \int_0^{W_m} \frac{\partial N(x)}{\partial \psi_s} \left( \int_0^x \frac{x' n(x')}{\int_0^{W_m} n(x) dx} dx' \right) dx$$

To obtain a neat form for Eq. (2.33), we define the following quantities:

$$x_c = \frac{\int_0^{W_m} x \bullet n(x) dx}{\int_0^{W_m} n(x) dx} \tag{2.36}$$

$$x_{sh} = \frac{\int_0^{W_m} \frac{\partial n(x)}{\partial \psi_s} \left( \int_0^x \frac{(x-x')n(x')}{\int_0^{W_m} n(x)dx} dx' \right) dx}{\int_0^{W_m} \frac{\partial n(x)}{\partial \psi_s} dx} \quad (2.37)$$

The quantity defined in Eq. (2.36) has a clear physical meaning of centroid of the inversion charge distribution, while the quantity defined in Eq. (2.37) is typically small, which is the second moment of the inversion charge distribution, conceptually linked to the spreading of inversion charge (wave function). With the facilitation of the quantities defined in Eq. (2.25), (2.36) and (2.37), Eq. (2.34) can be simplified as:

$$\partial \left( \frac{E_1 - E_{c0}}{q} \right) / \partial \psi_s = \frac{q}{\epsilon_s} (x_{av} - x_{sh}) \int_0^{W_m} \frac{\partial n(x)}{\partial \psi_s} dx + \frac{q}{\epsilon_s} x_c \int_0^{W_m} \frac{\partial N(x)}{\partial \psi_s} dx = \frac{x_{av} - x_{sh}}{\epsilon_s} C_i + \frac{x_c}{\epsilon_s} C_d \quad (2.38)$$

In Eq. (2.38), we made use of the definition of  $C_i$  and  $C_d$  as defined in Eq. (2.28).

From Eq. (2.38), it is possible to obtain an expression with more physical meaning for the quantum well capacitance.

$$C_{qw} = \frac{C_i}{\frac{x_{av} - x_{sh}}{\epsilon_s} C_i + \frac{x_c}{\epsilon_s} C_d} = \frac{\epsilon_s}{x_{av} - x_{sh} + x_c} \frac{C_d}{C_i} \quad (2.39)$$

It is instructive to understand the physical meaning of Eq. (2.39). For an undoped FET design (i.e.  $C_d=0$ ), the quantum well capacitance is essentially related to the length of the centroid of the differential inversion charge (note that this quantity is different from the centroid of the inversion charge), since  $x_{sh}$  is typically much smaller.

With all the derivation so far, we will be able to revisit the definition of an equivalent oxide thickness that is associated with the overall input capacitance, as we did in the previous

section. However, with the inclusion of the depletion charge here, we will have to use the capacitance constraint given Eq. (2.28). Assume zero back substrate charge, we have

$$\frac{x_{av}}{\epsilon_s} C_i + \frac{W_{av}}{\epsilon_s} C_d = 1$$

It has been shown that  $C_i$  is the serial combination of  $C_{DOS}$  and  $C_{qw}$ , the equivalent oxide thickness associated with  $C_i$  is therefore given by (with Eq. (2.30) and Eq. (2.39)):

$$C_i = \frac{\epsilon_s \left(1 - \frac{x_c}{W_{av}}\right)}{x_{DOS} + x_{av} - x_{sh} - x_c \frac{x_{av}}{W_{av}}} \equiv \frac{\epsilon_s}{x_i} \quad (2.40)$$

$$x_i = x_{av} + \frac{W_{av}}{W_{av} - x_c} (x_{DOS} - x_{sh})$$

Given the equivalent circuit configuration in Fig. 2.16, we can further work out the equivalent oxide thickness associated with the overall input capacitance, denoted by  $x_{inv}$ .

$$x_{inv} = x_{ox} + x_{av} + \frac{W_{av}}{W_{av} - x_c} (x_{DOS} - x_{sh}) + \frac{x_{ox}}{W_{av} - x_c} (x_{DOS} - x_{sh}) \quad (2.41)$$

Compared to the expression provided in the previous section, the inclusion of the depletion charge accentuates the importance of the  $C_{DOS}$ .

## 2.4 Layer Design for a Representative III-V MOSFET

### 2.4.1 Introduction

Given the methodologies that were described in Section 2.2 and 2.3, we will set out to design the layer structure for a representative III-V MOSFET. In this section, we will first present a baseline layer design, with explanation on the role of each layer. Numerical simulation result is then supplemented to verify the design concepts. For completeness, a brief discussion on the fabrication challenges of such a device will be provided with references on recent progress of this subject.

### 2.4.2 Layer Structure Design for a III-V MOSFET

It is quite clear that an appropriate layer design must be capable to support sufficient sheet charge density over a limited gate voltage swing. To address this criteria, adequate gate input capacitance must be achieved. As seen from Sec. 2.3, thinning down physical thickness of the gate dielectric is necessary, but not sufficient. In addition to the thin gate dielectric, we need to engineer the wave function to be as close to the gate as possible, in order to reduce the quantum well capacitance.

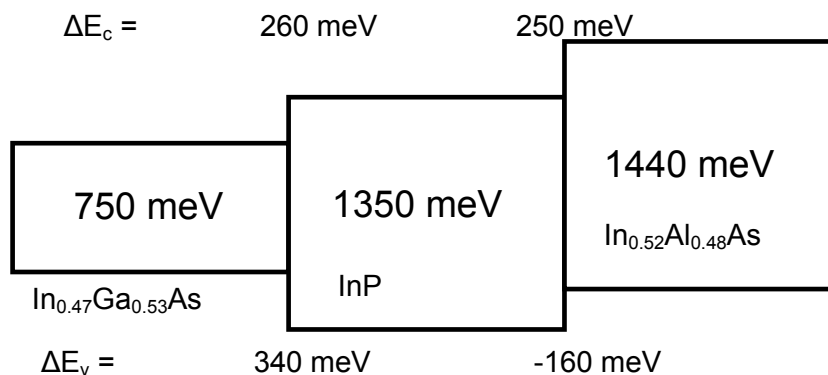


Fig. 2.17. Band alignment of selected III-V materials, data taken from [6,7].

A cross section of the baseline layer structure design is depicted in Fig. 2.18. The gate dielectric is chosen to give an EOT of 0.6nm. The channel layer is designed to consist of two materials –  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  as the top and primary layer while  $\text{InP}$  as the bottom and auxiliary layer of the channel region. This arrangement is based on: 1) due to the band alignment configuration (see Fig. 2.17) between  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  and  $\text{InP}$ , the energetic confinement force the wave function towards the surface, 2) at higher gate bias, the  $\text{InP}$  layer can be turned on, providing additional density of states (due to its higher effective mass), and thus give higher sheet charge density. There is a back barrier layer that is made of  $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$  to provide further confinement, followed with a  $\text{p}^+$   $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  ground plane. The vertical stack of the intrinsic device is made of materials that are lattice matched to an  $\text{InP}$  substrate, therefore minimal strain/stress is introduced from the material arrangement. The electrical configuration of the layer design resembles that of an asymmetric double gate device, to provide good electrostatic integrity when the lateral dimension of the device get scaled down. The detailed thickness of each layer is given below: high-k dielectric: 0.6nm EOT;  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  channel layer: 2.5nm;  $\text{InP}$  channel layer: 2.5nm;  $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$  back barrier layer: 3.5nm.

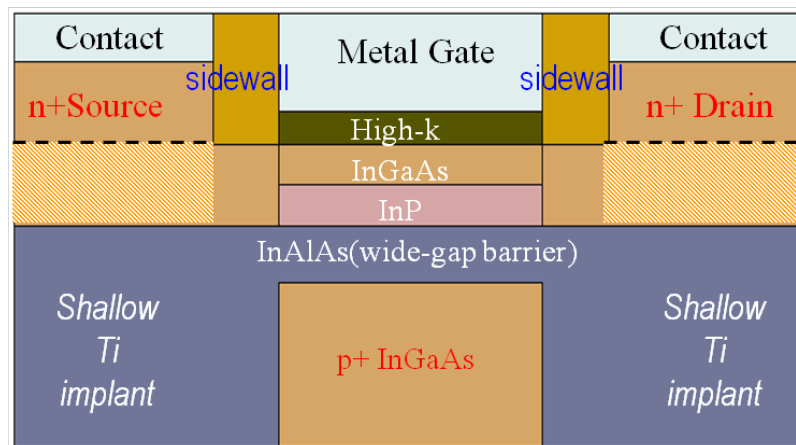


Fig. 2.18. Illustration of the baseline layer structure design of a III-V MOSFET.



To verify the feasibility of the layer structure design, numerical simulation using the method described in Section 2.2 is performed to account for the non-parabolic conduction band profile. The relevant material parameters for each layer are given in Table 2.2.

Numerical simulation results are summarized in Fig. 2.19, where wave function, electrostatic potential and carrier distribution are shown under representative gate bias.

Table 2.2. Material parameter used for layer structure design of a III-V MOSFET.

	In <sub>0.53</sub> Ga <sub>0.47</sub> As	InP	InAlAs	Fictitious High K
$m^*/m_0$	0.041	0.079	0.08	0.5
$\alpha$ (eV <sup>-1</sup> )	1.22	0.63	N/A	N/A
$X$ (eV)	4.66	4.4	4.16	2.0
$\epsilon_r$	14.0	12.4	12.45	13.0
$E_g$ (eV)	0.78	1.42	1.46	6.0

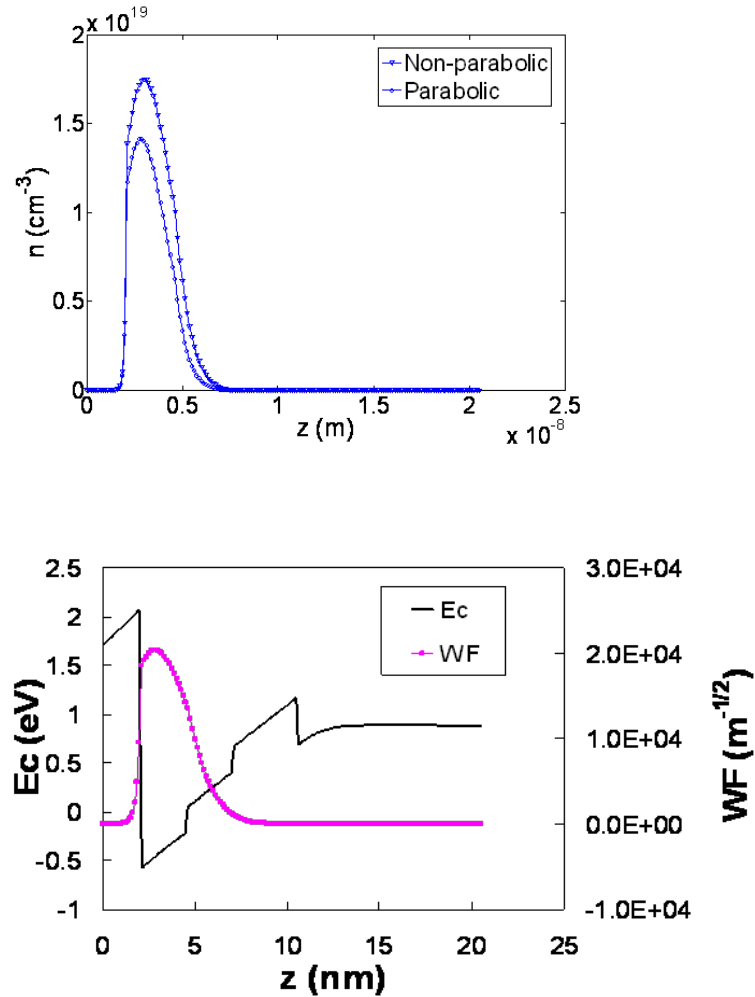


Fig. 2.19. Carrier distribution, wave function and electrostatic potential at representative gate bias for the baseline layer structure design.

The sheet charge density as a function of gate bias is plotted in Fig. 2. 20 and the derived quasi-static C-V characteristics are shown in Fig. 2.21. It is seen here again, the importance of accounting for the non-parabolic conduction band. It must be mentioned here that the gate work function is assumed to be chosen on an as-needed basis. For this specific design, the gate work function used is 4.7eV, resulting  $\sim 0.3\text{V}$  threshold voltage. It is noticed that the composite channel design is capable to deliver  $\sim 8 \times 10^{12}$   $\text{cm}^{-2}$  sheet charge density at gate overdrive of 0.7V.

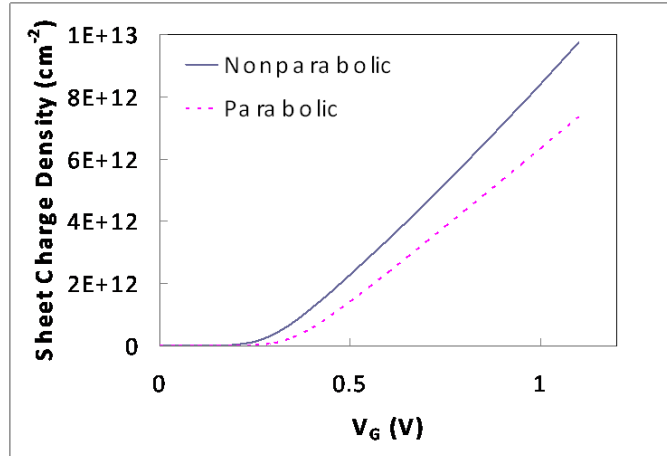


Fig. 2.20. Sheet charge density as a function of gate bias. Gate metal work function is assumed to be 4.7eV.

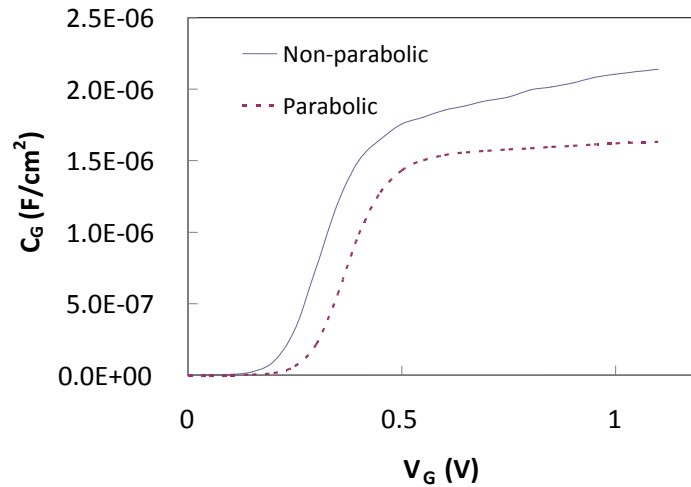


Fig. 2.21. Derived quasi-static CV characteristics for the baseline layer structure design.

The current drive capability and some benchmark of this baseline layer structure design are the subjects of following chapters, along with a few variations based upon this design. The fabrication of such device is difficult, however feasible. The technology has been developed by a group at UCSB and a first device of this type has been experimentally realized and published in [11]. There are many practical challenges, including the selective etching strategy, ultra-thin sidewall and selective regrowth of the source/drain region, on top of the integration of a high

quality high-k dielectric layer. We did not adopt the ion implantation technique for source/drain formation, which is used elsewhere [12] to fabricate the III-V MOSFET. The design considerations behind are: 1) doping/activation may not achieve the level that a proper design desires (see Chpt 3 for detail); 2) implantation damage may hurt the mobility and introduce high series resistance at source and drain. The regrowth methodology adopted in the device fabrication employs InAs material to form the source and drain, due to the good surface mobility associated with InAs. The readers are directed to the paper [11] for recent progress on the technology development of such III-V MOSFET device.

## 2.5 Summary

In this chapter, we addressed the various physical concerns while designing the layer structure. To carry out an appropriate layer structure design, it is necessary to understand the impact of non-parabolic conduction band. The methodology of treating the problem with such consideration is described in Section 2.2. Furthermore, to enable a respectful sheet charge density to be achieved within a certain layer structure design, the various components that contribute to the overall gate capacitance is further analyzed, in Section 2.3. Section 2.4 brought out the baseline layer structure design of a representative III-V MOSFET, whose transport properties are to be analyzed in the following chapter.

Section 2.2, in full, is submitted to “Solid State Electronics”. The dissertation author was the primary author of the paper.

## Reference:

1. I-H. Tan, G.L. Snider, L.D. Chang and E.L. Hu, “A self-consistent solution of Schrödinger–Poisson equations using a nonuniform mesh”, *J Appl Phys* 68 (1990) (8), pp. 4071–4076
2. C. Wetzel, R. Winkler, M. Drechsler, B. K. Meyer, U. Rössler, J. Scriba, J. P. Kotthaus, V. Härle and F. Scholz, “Electron effective mass and nonparabolicity in  $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$  quantum wells”, *Physics Review B*, vol. 53, pp. 1038-1041
3. U. Ekenberg, “Nonparabolicity effects in a quantum well: Sublevel shift, parallel mass, and Landau levels”, *Physics Review B*, vol. 40, pp. 7714-7726
4. C. Troger, H. Kosina, and S. Selberherr, “Modeling nonparabolicity effects in silicon inversion layers”, *SISPAD '97*, pp. 323-326
5. M. Lundstrom, “Fundamentals of Carrier Transport” 2nd Ed. Cambridge University Press
6. M. Shur, “Handbook series on semiconductor parameters” vol.2
7. O. Madelung edited, “Semiconductors – Basic Data”, 2<sup>nd</sup> edition, Springer, 1996
8. V. Ariel Altschul, A. Fraenkel, and E. Finkman, “Effects of band nonparabolicity on two-dimensional electron gas”, *Journal of Applied Physics*, vol. 71, pp. 4382-4384
9. S. Picozzi, A. Continenza, R. Asahi, W. Mannstadt, A. J. Freeman, W. Wolf, E. Wimmer and C. B. Geller, “Volume and composition dependence of direct and indirect band gaps in ordered ternary III-V semiconductor compounds: A screened-exchange LDA study”, *Physical Review B*, vol. 61, pp. 4677-4684
10. Airy functions, see for example, [://en.wikipedia.org/wiki/Airy\\_function](http://en.wikipedia.org/wiki/Airy_function)
11. Uttam Singiseti et. al., “ $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  channel MOSFETs with self-aligned InAs source/drain formed by MEE regrowth”, accepted by *IEEE Electronic Device Letter* , 2009
12. Xuan, Y.Q. Wu, and P.D. Ye, “High-Performance Inversion-Type Enhancement-Mode InGaAs MOSFET With Maximum Drain Current Exceeding 1 A/mm” , *IEEE Electron Devices Letters* Vol. 29, No. 4, 2008, pp. 294
13. Private communication with Prof. Mark Rodwell (UCSB)

## Chapter 3. Quasi-ballistic Transport in Ultra-scaled III-V

### MOSFETs

#### 3.1 Introduction

In this chapter, we deal with the transport properties of ultra-scaled III-V MOSFETs. Given the small gate length the device design is targeted at (22nm technology node and beyond), the transport distance becomes comparable to the mean free path (MFP) of a carrier, which implies that the carrier transport becomes ballistic or quasi-ballistic. In fact, it has been shown previously that close-to-ballistic transport is more likely to happen for III-V materials in comparison to Si at the same transport length [1]. Therefore, it is appropriate to formulate the transport study for an ultra-scaled III-V MOSFET based on ballistic transport theories.

We will cover the following topics within this chapter: ballistic transport theory, (which was originally developed by Natori [2] and later on enriched by Lundstrom et. al. [3]), will be reviewed in brief in Section 3.3. A modified version of the ballistic transport mechanism taking into account the small DOS of III-V's is then presented in Section 3.3, together with a calculation of the ideal ballistic characteristics for a representative III-V MOSFET and its Si counterpart. In Section 3.4, we consider non-ideal factors such as series resistance and reflections due to finite mobility considerations. Section 3.5 discusses limitation of the quasi-ballistic model and provides an overall summary for this chapter.

### 3.2 Brief Review of the Ballistic Transport Theory

The ballistic transport model for a MOSFET was first presented by Natori in [2]. The theory is based on the following assumptions:

- 1) As shown in Fig. 3.1, along the transport direction in the channel, there is a point (typically called “virtual source”) where the lateral electric field is zero. It is further assumed that at this point, the carrier distribution is well described by a Fermi-Dirac distribution with the Fermi energy in the source.

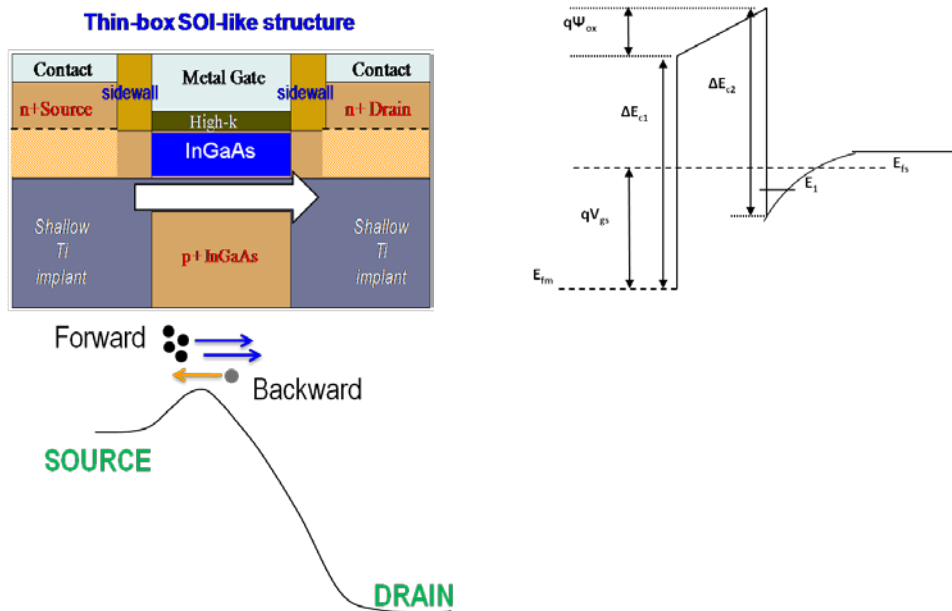


Fig. 3.1. Schematic of the energy band diagram along the transport direction (bottom figure) and the transverse direction (top right figure). It is assumed that the quantization occurs within the channel.

- 2) From the point virtual source onward, all the forward going carriers are originated from source (because of no scattering); all the backward going carriers are originated from the drain.



- 3) Due to the conservation of energy as a result of no scattering, all carriers originated from the source have energies corresponding to their Fermi-Dirac distribution at the virtual source, which is described by the source Fermi level; all carriers originated from the drain are described by Fermi-Dirac distribution with the drain Fermi level.
- 4) The sheet charge density at the virtual source is assumed to be proportional to  $V_{GS}-V_t$ .

We will briefly recapitulate the key results of the original theory put forward by Natori here, and provide more discussion within the next section of a formulation to account for the small density of states issue that prevails within the III-V material. For a full account of Natori's model, readers are referred to the original paper [2] or the excellent recapitulation in [4].

Assume only one subband is occupied for the MOSFET, (assuming an n-channel device here) the forward and backward going charge densities at the virtual source may be obtained via the following integration:

$$\begin{aligned}
 N_s^+ &= \frac{m^*}{2\pi\hbar^2} \int_{E_1}^{+\infty} \frac{1}{1 + \exp\left(\frac{E - E_{fs}}{kT}\right)} dE = \frac{kTm^*}{2\pi\hbar^2} \ln\left[1 + \exp\left(\frac{E_1 - E_{fs}}{kT}\right)\right] \\
 N_s^- &= \frac{m^*}{2\pi\hbar^2} \int_{E_1}^{+\infty} \frac{1}{1 + \exp\left(\frac{E - E_{fs} + qV_{DS}}{kT}\right)} dE = \frac{kTm^*}{2\pi\hbar^2} \ln\left[1 + \exp\left(\frac{E_1 - E_{fs} + qV_{DS}}{kT}\right)\right]
 \end{aligned} \tag{3.1}$$

where  $N_s^+$  and  $N_s^-$  are the forward and backward going sheet charge density at the virtual source;  $E_1$  is the ground state energy.

The computation of the forward and backward going currents may be carried out by integrating in the 2-D k-space. Assume the current transport direction is aligned with y axis, the forward and backward going currents can be expressed as:

$$I^+ = \frac{q}{4\pi^2} \int_{-\infty}^{+\infty} dk_x \int_0^{+\infty} \frac{\hbar k_y / m^*}{1 + \exp\left\{\frac{E_1 + [\hbar^2(k_y^2 + k_x^2) / 2m^*] - E_{fs}}{kT}\right\}} dk_y$$

$$I^- = \frac{q}{4\pi^2} \int_{-\infty}^{+\infty} dk_x \int_{-\infty}^0 \frac{\hbar k_y / m^*}{1 + \exp\left\{\frac{E_1 + [\hbar^2(k_y^2 + k_x^2) / 2m^*] - E_{fs} + qV_{DS}}{kT}\right\}} dk_y \quad (3.2)$$

In the above derivations, we made use of the following equalities:  $E_{fs} - E_{fd} = qV_{DS}$ , and the y direction velocity is  $\hbar k_y / m^*$ . The overall charge density at the virtual source point is given by the sum of the forward and backward going charge densities, whereas the current is given by the difference between the forward and backward going current.

$$Q = q(N_s^+ + N_s^-) = \frac{kTm^*}{2\pi\hbar^2} \left\{ \ln\left[1 + \exp\left(\frac{E_1 - E_{fs}}{kT}\right)\right] + \ln\left[1 + \exp\left(\frac{E_1 - E_{fs} + qV_{DS}}{kT}\right)\right] \right\} \quad (3.3)$$

To uniquely determine the charge density and current at certain bias condition, it is necessary to compute the corresponding  $E_1 - E_{fs}$  value. This is done, in Natori's model, via equating the inversion charge expressed in Eq. (3.3) to  $C_{ox}'(V_{GS} - V_t)$ , where  $C_{ox}'$  is the equivalent oxide capacitance. Given the value of  $E_1 - E_{fs}$ , the integration in Eq. (3.2) may be obtained readily and the overall current is  $(I^+ - I^-)$ .

In the following section, we will modify this formulation slightly to better account for the DOS capacitance for III-V material.

### 3.3 Modified Ballistic Transport Theory and Ballistic Performance Comparison between III-V MOSFET and Si MOSFET

In this section, we will present a formulation that solves the Poisson equation rigorously along the transverse direction (x-direction), under the assumption that only one-subband is occupied and the band structure is described with parabolic dispersion (to obtain a close form solution). By referring to the energy band diagram in Fig. 3.1, assuming 1-D electrostatic is an adequate description for this problem, we have the following relationships among energies:

$$E_{fs} - (E_{fs} - E_I) - (E_I - E_c) + \Delta E_{c2} - q\Psi_{ox} - \Delta E_{c1} = E_{fm} \quad (3.4)$$

Using that  $E_{fs} - E_{fm} = qV_{gs}$ , and by referencing other energy levels to  $E_{fs}$  ( $E_{fs} = 0$ ), one can obtain an equation as:

$$V_{gs} = (\Delta E_{c1} - \Delta E_{c2})/q - E_I/q + (E_I - E_c)/q + \Psi_{ox} \quad (3.5)$$

Assuming that  $E_I - E_c$  is relatively a constant above threshold, the above equation can be simplified as

$$V_{gs} = V_t - E_I/q + \Psi_{ox} \quad (3.6)$$

In Eq. (3.6), we attributed the term  $(\Delta E_{c1} - \Delta E_{c2})/q + (E_I - E_c)/q$  to the threshold voltage. Given that  $\Psi_{ox}$  is determined by  $q(N_s^+ + N_s^-)/C_{ox}'$ , one can solve the following equation to obtain  $E_I$ :

$$V_{gs} - V_t + \frac{E_I}{q} = \frac{1}{C_{ox}'} \frac{2\pi q k T g m_t}{h^2} \left[ \ln \left( 1 + e^{-\frac{E_I}{kT}} \right) + \ln \left( 1 + e^{-\frac{E_I + qV_{ds}}{kT}} \right) \right] \quad (3.7)$$

where  $C_{ox}'$  contains oxide capacitance and quantum well capacitance;  $g$  is valley degeneracy;  $m_t$  is the in-plane effective mass.

Once the  $E_I$  is solved, the overall current may be calculated with Eq. (3.2).

The aforementioned formalism takes explicit account of the small DOS for the III-V material, by means of solving the gate direction electrostatic explicitly. This differs from the previous formulation in [2], where the DOS capacitance was treated as a voltage independent quantity and was included in  $C_{ox}$ . As a result, in [2], the corresponding equation for (3.7) did not have the  $E_l/q$  term.

In the following, we will compare the ballistic limit result for a Silicon MOSFET and a III-V MOSFET (using  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  as the channel material). We will first examine the sheet charge density as function of drain bias at the virtual source for both cases (shown in Fig. 3.2). It is noticed that at all gate biases, the sheet charge density of the InGaAs MOSFET changes drastically as drain bias is increased, whereas for the Si case, the change is relatively small.

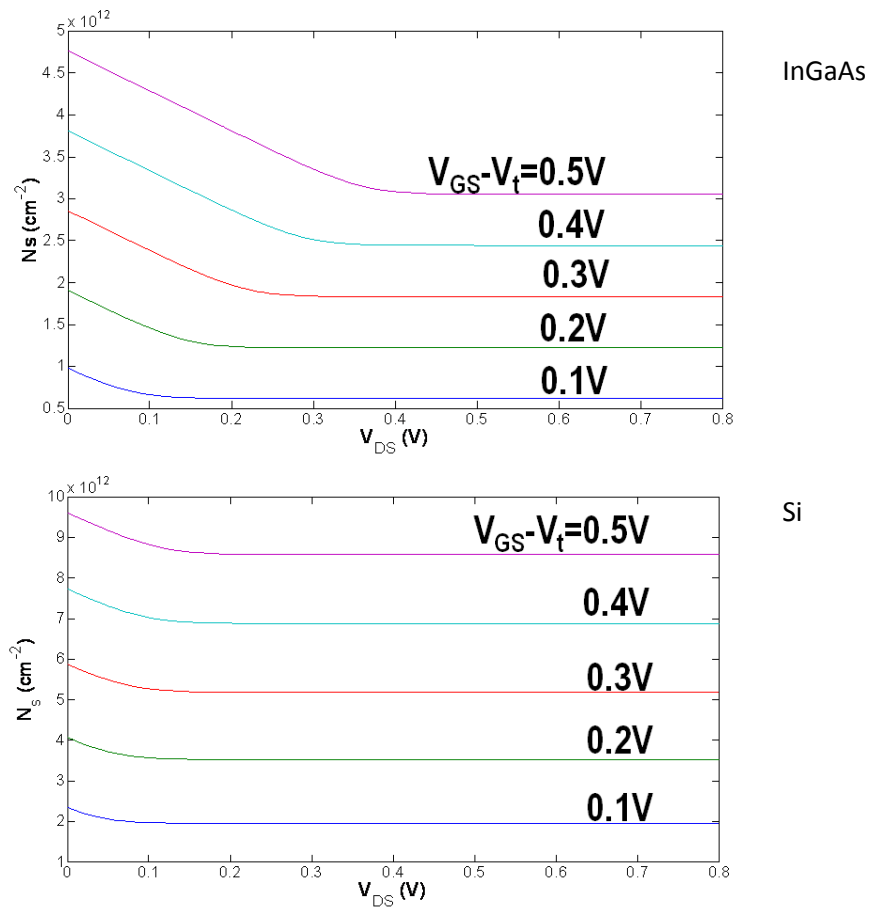


Fig. 3.2. Sheet charge density at virtual source under one-subband ballistic condition, gate overdrive is ramped from 0.1V to 0.5V with a step of 0.1V for both InGaAs and Si.

There are two aspects that we would like to understand for the drain bias dependence of the virtual source sheet charge density: 1) why does the sheet charge density decrease with increased drain bias; 2) why does the decrease in the InGaAs case much more significant than the Si MOSFET case. The first question may be readily understood by inspecting the band diagram sketched in Fig. 3.3 under different drain bias conditions. As mentioned previously, the sheet charge density at the virtual source is composed of the forward and backward going electrons, which are in thermal equilibrium with the Fermi level in source and drain, respectively. Under zero drain bias, equal contributions are made to the overall sheet charge density at the virtual

source from the forward and backward going electrons. Whereas, under high drain bias, due to the energy barrier present between the drain and virtual source, virtually no backward going electron contribute to the sheet charge density at the virtual source. It is however to be clarified that the sheet charge density does not therefore reduce to half – the transverse direction Poisson equation must be satisfied at all times and therefore requires more forward going electrons be present.

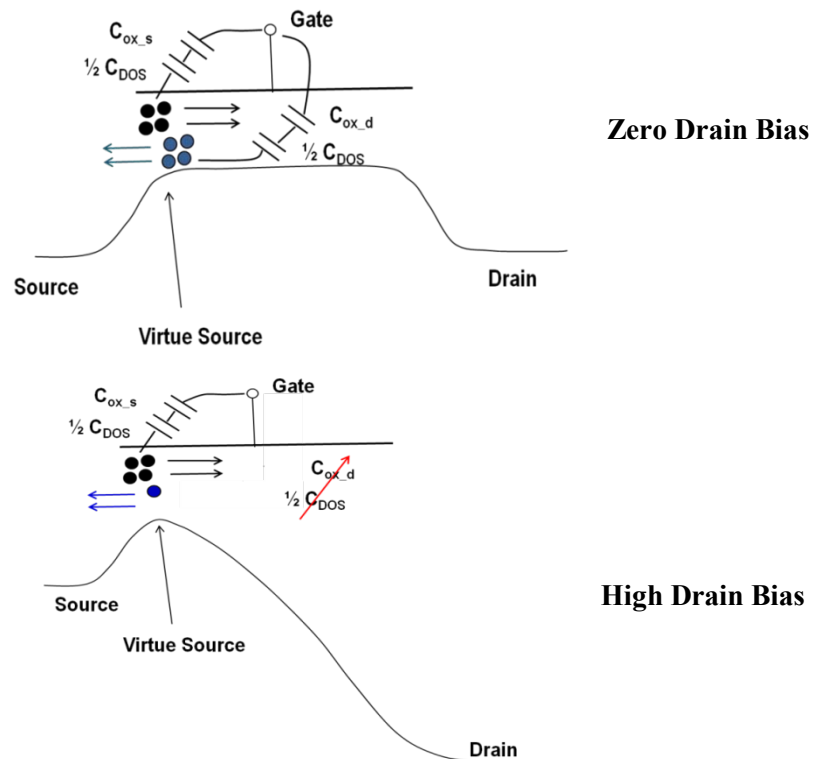


Fig. 3.3. Schematic of the energy band diagram and gate coupling to the virtual source with forward and backward going electrons under zero drain bias and high drain bias.

To understand the reason why InGaAs MOSFETs lose more sheet charge density with increased drain bias, we continue to use the capacitance model that we developed in Chapter 1. As depicted in Fig. 3.3, the capacitance coupling for both forward and backward going electrons approximately consist of the equivalent oxide capacitance (with quantum well capacitance

included) and the DOS capacitance. For both cases (Si and InGaAs), the coupling capacitance between the gate and forward/backward going electrons is determined not only by  $C_{ox}$ , but also by  $C_{DOS}$ , half of which comes from forward going electrons while the other half from backward going electrons. Under zero drain bias condition, equal contributions are made by the forward and backward going electrons, therefore the gate is coupling to the channel at the virtual source by  $C_{ox}/(1/2C_{DOS}+1/2C_{DOS})$ ; under high drain bias, as the coupling to backward going electrons vanishes, the gate is coupling to the channel at the virtual source by  $C_{ox}/1/2C_{DOS}$ . For the InGaAs case, due to the small  $C_{DOS}$ , the overall coupling capacitance is largely influenced by losing  $1/2C_{DOS}$  as the drain bias is raised. However, for the Si case, due to the multi-valley degeneracy ( $g=2$ ) and larger effective mass, the  $C_{DOS}$  is much larger, and thus poses a milder effect to the overall gate-virtual source coupling. We therefore see a much smaller decrease in sheet charge density in the Si case.

We may now examine the ballistic current limit for both Si and InGaAs MOSFETs. The calculated ballistic I-V characteristics are plotted in Fig. 3.4. The effective gate oxide thickness is assumed to be 1 nm for both cases. We notice that FETs made from both materials perform comparably in terms of current drive, with InGaAs MOSFET slightly better. It is further shown that such comparison is highly dependent on the effective oxide thickness, as the overall gate capacitance that determines the charge density is a combination of the oxide capacitance and DOS capacitance [5]. In [5], the authors pointed out that with currently achievable EOT, InGaAs provides advantages in delivering higher current drive; however, if the EOT can be made much smaller (e.g.  $<0.4\text{nm}$ ), Si would be a better choice in terms of ballistic current limit.

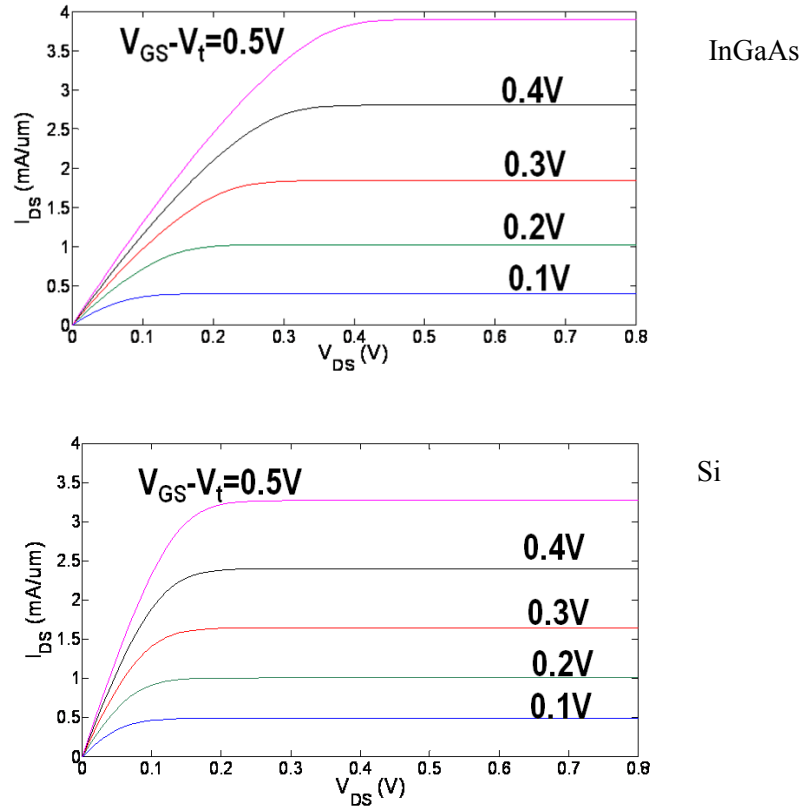


Fig. 3.4. Computed ballistic limit I-V characteristics for an InGaAs and a Si MOSFET, with EOT thickness of 1 nm assumed.

Up to now, we have discussed the I-V characteristics under the ballistic transport assumption. In reality, there are many non-ideal factors that will affect the overall performance, namely series resistances, rare scattering events that reflect part of the forward/backward carrier flows. These considerations will be treated in the next section.



### 3.4 Quasi-ballistic Transport with Realistic Considerations and Performance Projections

In this section, we will develop understanding of the (quasi-)ballistic transport in a MOSFET device with more realistic considerations such as non-parabolic conduction band, series resistance at the source and drain and reflection due to rare scattering events. We will further discuss the role of the mobility and its implication in how closely a MOSFET may behave to its ballistic limit.

We start our discussion with the conduction band non-parabolicity and its impact on the current drive capability of a given structure. As mentioned in the first chapter, significant underestimation of the sheet charge density may occur if non-parabolicity is not taken into account for the III-V material. This non-parabolic conduction band effect was not included for ballistic current computation in prior literature.

In the following computation, we use the structure that was designed as our baseline structure, with EOT thickness of 0.6nm. Several structure variations are also included, via changing the relative composition of the composite channel (i.e. changing the thickness of InP/InGaAs while keep the entire stack thickness constant at 5nm). It must be noted that in this computation, we consider only high drain bias situation, such that the approximation can be made at the virtual source that no backward going electrons are present. It is found, as shown in Fig. 3.5, for all variation considered, the current drive capability is comparable, exceeding  $6\text{mA}/\mu\text{m}$  (note that the EOT is lower in this case than the computation presented in the previous section).

The computation result should be considered as a limit that the specific channel designs can support, without any extrinsic components being taken into account. An immediate impact to the drain current level may result from the presence of series resistance. The impact of this may

be readily understood through the source degeneration configuration of an MOSFET, where the current is degraded by  $1/(1 + g_m R_s)$ , where  $g_m$  is the intrinsic transconductance. The computed impact of source series resistance is shown in Fig. 3.6. It is noted that  $>50 \Omega\text{-}\mu\text{m}$  source resistance may significantly degrade the device performance. The series resistance is defined through  $R_s = \sqrt{R_{sh} \cdot \rho_c}$ , where  $R_{sh}$  is the sheet resistivity and  $\rho_c$  is the contact resistance.

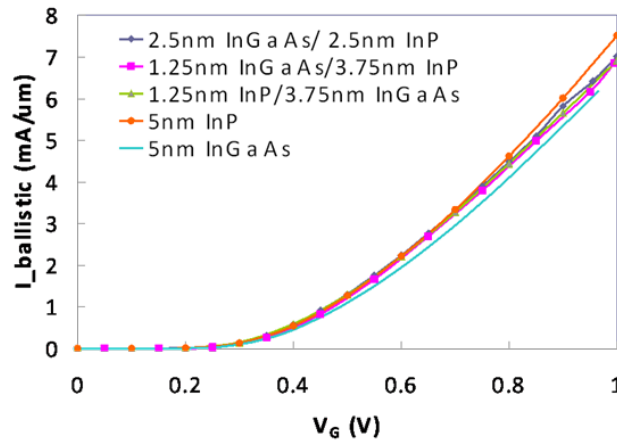


Fig. 3.5. Computed ballistic limit I-V characteristics for the baseline design and its variations as presented in Chapter 1, EOT here is 0.6nm. Threshold voltages are adjusted for all designs.

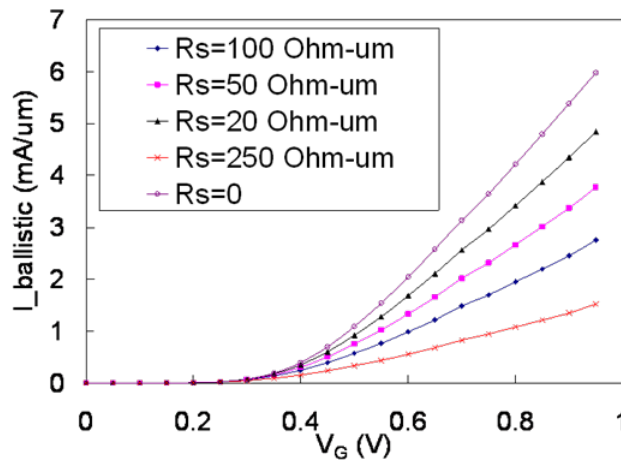


Fig. 3.6. Computed ballistic limit I-V characteristics for the baseline design with series resistance at the source.

From the computation, it is necessary to keep the source series resistance low (preferably no greater than  $20 \text{ } \Omega\text{-}\mu\text{m}$ ) to maintain a reasonably high current drive. It is therefore required to maintain low sheet resistance as well as contact resistance at the same time, a simple estimation indicates that a combination of contact resistance  $<2 \times 10^{-8} \text{ } \Omega\text{-cm}^2$ , and source doping  $>3 \times 10^{19} \text{ cm}^{-3}$  with a mobility  $>600 \text{ cm}^2/\text{V}\text{-sec}$  may deliver the targeted source resistance level.

In addition to the transfer characteristics ( $I_{ds}$  v.s.  $V_{gs}$ ), we are also interested in the output characteristics ( $I_{ds}$  v.s.  $V_{ds}$ ) of a (quasi-)ballistic FET with the non-ideal factors. We include the following two considerations: 1) in presence of series resistances at source and drain; 2) in presence of rare scattering events. For the first non-ideality factor, the intrinsic and extrinsic device behaviors are linked via the following equations:

$$\begin{aligned} V_{gs}^{ext} &= V_{gs}^{int} + I_{ds} \cdot R_s \\ V_{ds}^{ext} &= V_{ds}^{int} + I_{ds} \cdot (R_s + R_d) \end{aligned} \tag{3.8}$$

Conceptually, the computation of the extrinsic characteristic may be carried out given Eq. 3.8. However, from a standpoint of implementation, it is easier to start with the intrinsic device characteristic so that a table of  $(V_{gs}^{int}, V_{ds}^{int}, I_{ds})$  may be obtained. A corresponding table consisting of  $(V_{gs}^{ext}, V_{ds}^{ext}, I_{ds})$  may be subsequently derived. It is worthwhile to mention the algorithmic realization of the interpolation for the desired bias points. Matlab provides a very convenient library function to perform Delaunay triangulation (function “delaunay”), so that a triangulated surface is constructed on existed extrinsic  $(V_{gs}^{ext}, V_{ds}^{ext}, I_{ds})$  points (an example is provided in Fig. 3.7). The interpolation is subsequently done on the constructed surface. Another approach that may be taken is using the Newton-Ralphson iteration, starting directly from the extrinsic bias condition (not used in the following computation).

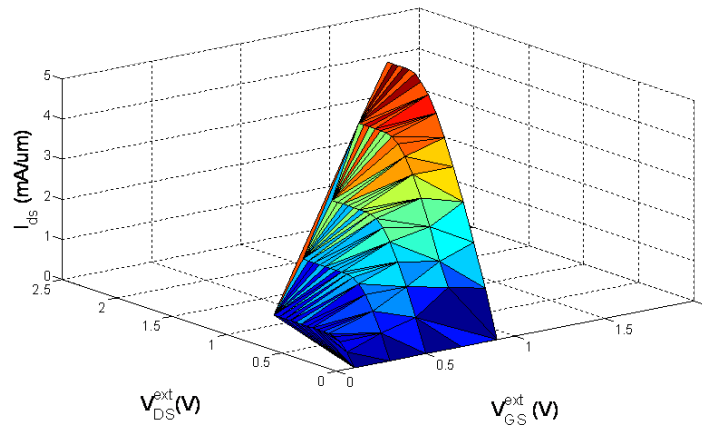


Fig. 3.7. Illustration of the Delaunay triangulation for extrinsic device characteristic computation.

Besides the series resistances, we would like to extend the consideration to include the situation where scattering is present but scarce (the quasi-ballistic situation). The underlying theory is presented in [3], where a “ $kT$ ” length is used to characterize the reflection coefficient. The overall current, after considering the rare scattering, is the original ballistic current times the transmission coefficient (shown schematically in Fig. 3.8).



Fig. 3.8. Illustration of the quasi-ballistic transport with the description of transmission coefficient.

The transmission coefficient is determined via:

$$T = \frac{L_0}{L_{scatter} + L_0} \quad (3.9)$$

where  $L_0$  is the mean free path, this quantity is determined by the following considerations:

For long channel device, the device characteristics are well described by drift diffusion model. Under very small  $V_{ds}$  bias condition, the quasi-ballistic transport model prediction should agree with the drift diffusion model prediction, as the lateral field effect is negligible for this situation and the scattering length approaches gate length. We may equate the current for the two conditions and obtain the mean free path given as following

$$L_0 = \frac{\mu k T \pi \ln(1 + E_{fs} / kT)}{q \langle v_{inj} \rangle F_{1/2}'(E_{fs} / kT)} \quad (3.10)$$

The scatter length is determined by:

$$L_{scatter} = \min(L_{kT}, L_G) \quad (3.11)$$

Via the average time interval between two consecutive scattering events, the overall current is linked to the low field mobility, which works in favor for the channel with high mobility to start with. In Fig. 3.9, the effects for transmission coefficients assuming different mobilities are shown for both InGaAs and Si quasi-ballistic FET, with an assumed gate length of 25nm.

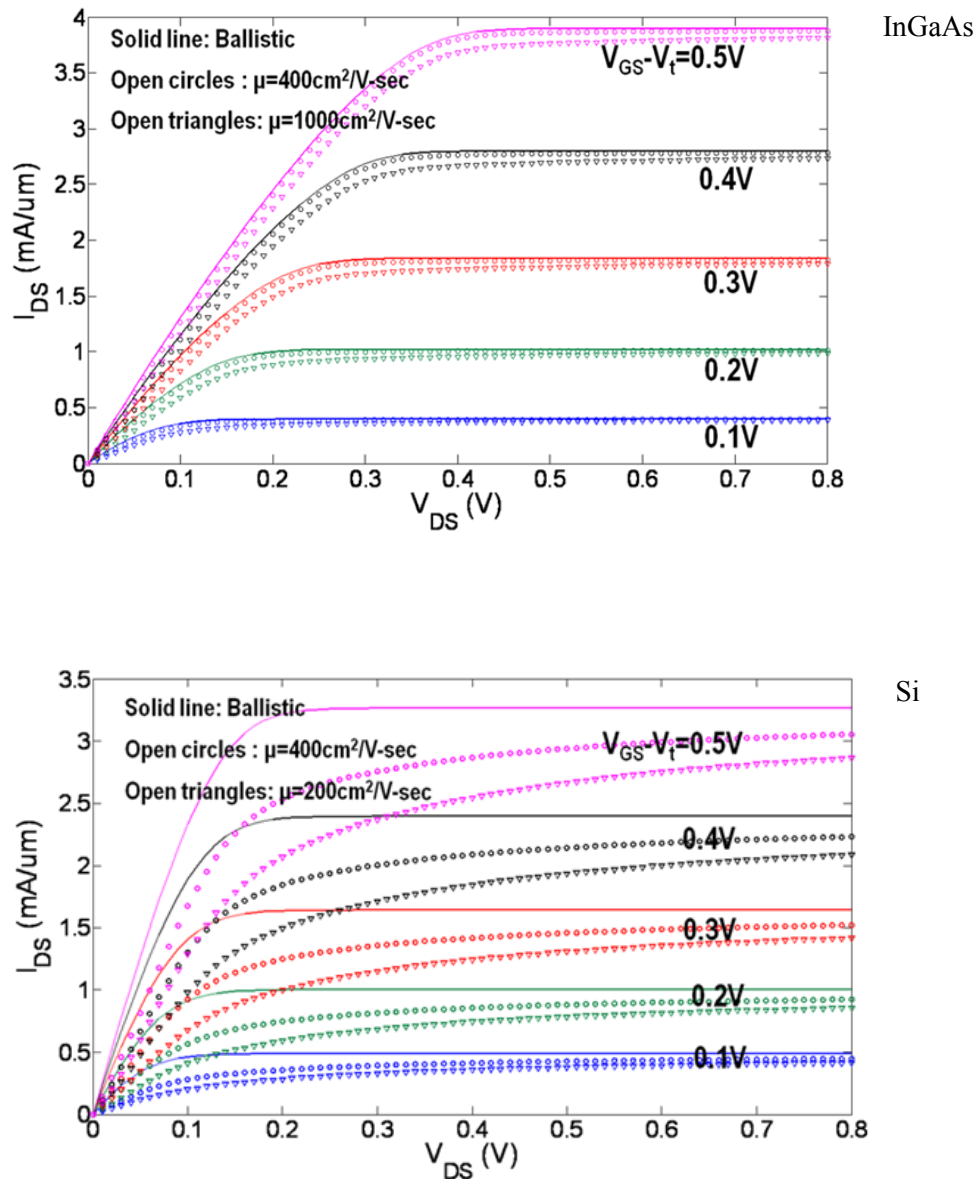


Fig. 3.9. Computed quasi-ballistic output characteristics for InGaAs and Si MOSFETs, assuming various mobility based on material properties.

It is shown that the “more realistic” output characteristics deviates from ideal ballistic limit characteristics more rapidly as mobility drops below  $1000 \text{ cm}^2/\text{V-sec}$  (which is the Si case).

To obtain a complete picture including both effects, we compare the output characteristics for InGaAs and Si MOSFET as shown in Fig. 3.10. We assumed here a smaller series

resistance for InGaAs, again due to its high mobility (it is the low field mobility that determines the series resistance associated with the source since near equilibrium condition holds within the source).

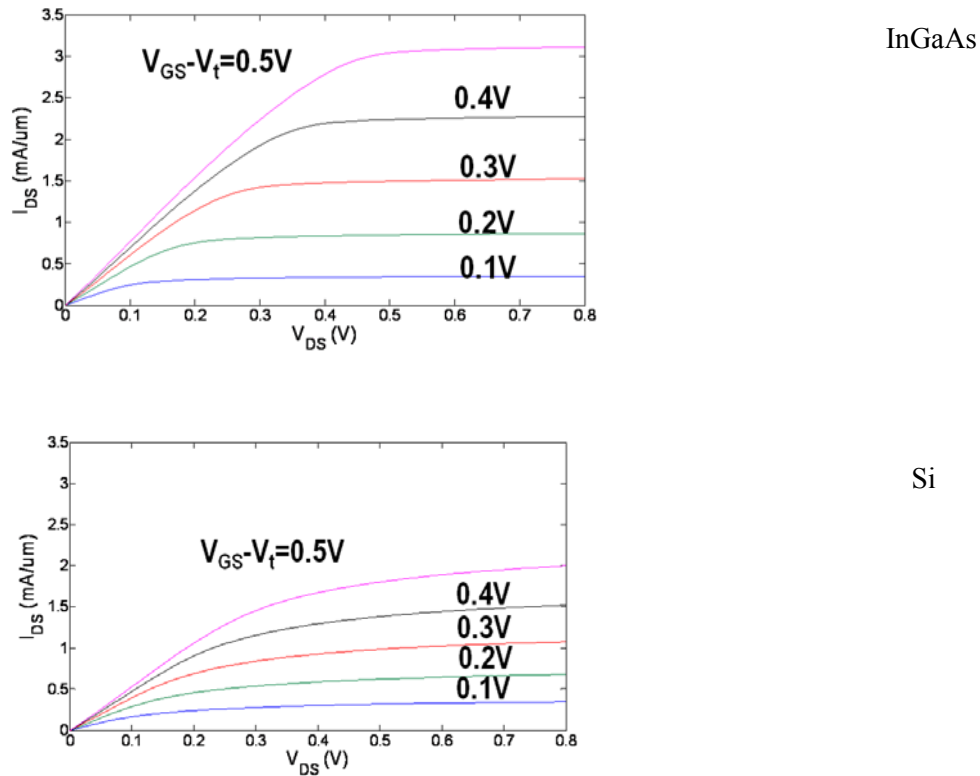


Fig. 3.10. Computed quasi-ballistic output characteristics for InGaAs and Si MOSFETs with the series resistance being taken into account. The assumed mobilities for InGaAs and Si are  $1000 \text{ cm}^2/\text{V}\cdot\text{sec}$  and  $200 \text{ cm}^2/\text{V}\cdot\text{sec}$ , respectively. The assumed series resistances are  $20 \text{ } \Omega\text{-}\mu\text{m}$  and  $50 \text{ } \Omega\text{-}\mu\text{m}$  on each side, respectively.

It is clear that the mobility, although it plays no role in the pure ballistic transport picture, influences the quasi-ballistic transport properties by determining the transmission coefficient and the series resistance. Therefore, introducing high mobility material to the channel does provide an advantage in this regard.

To provide some verification to the forgoing simulation result, we compared our simple model to existing hardware data taken from [6]. A Si NMOSFET designed for 65nm technology

node with 35nm physical gate length may be fitted reasonably well with mobility of  $200 \text{ cm}^2/\text{V}\cdot\text{sec}$ , series resistance of  $80 \text{ }\Omega\cdot\mu\text{m}$  and no further parameter tuning.

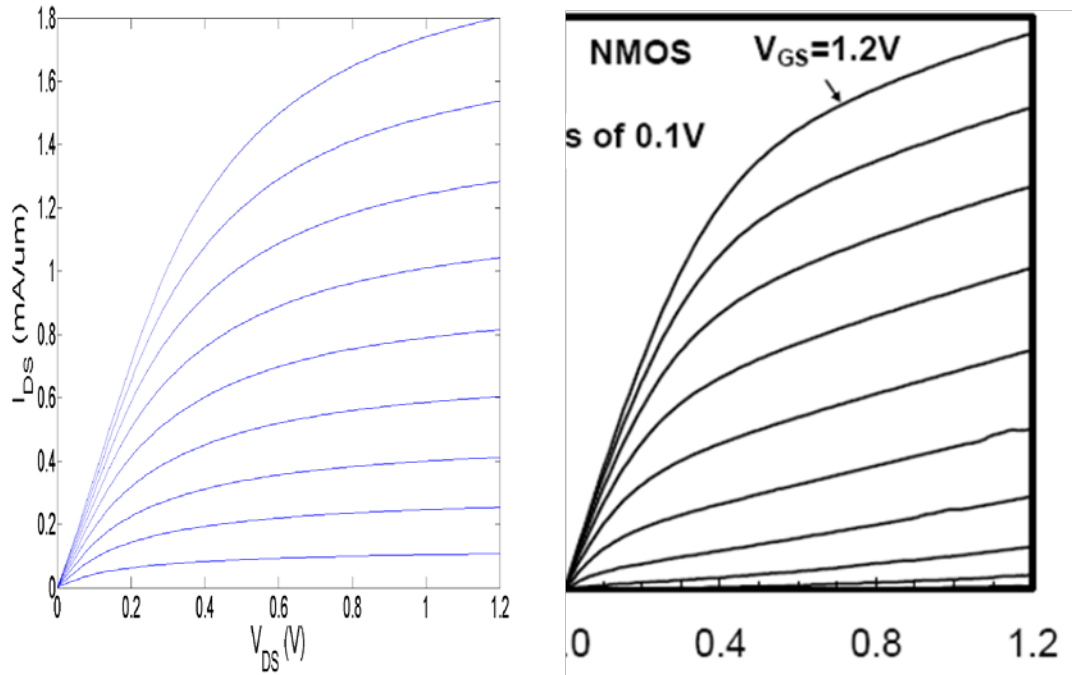


Fig. 3.11. Model verification by comparing hardware data (35nm gate length Si nMOSFET) to model computation, with extracted mobility of  $200 \text{ cm}^2/\text{V}\cdot\text{sec}$  and series resistance of  $80 \text{ }\Omega\cdot\mu\text{m}$ .  $V_{GS}$  step of the curves is 0.1V.

From the obtained I-V characteristics and Q-V characteristics under the quasi-ballistic condition, we are able to further project the performance if such devices were to be used in a CMOS inverter. We assume comparable output characteristics may be achieved by a pMOSFET via proper device sizing (even though this will be very difficult for a purely III-V based technology). Two methods are implemented here to obtain the propagation delay, through analytical model as reported in [7] and through numerical simulation by fitting the device characteristics to a Curtice Cubic Model [8].



To estimate the propagation delay analytically, it is necessary to calculate effective current drive and the capacitive loads. The effective current drive is defined as [7]:

$$I_{eff} = \frac{1}{2}[I_{DS}(V_{GS} = V_{DD}/2, V_{DS} = V_{DD}) + I_{DS}(V_{GS} = V_{DD}, V_{DS} = V_{DD}/2)] \quad (3.12)$$

The capacitive loads contain intrinsic components (gate to channel coupling) and parasitic components (fringing capacitances, which is estimated from parameters of current technology node, since it does not change significantly from one technology node to another). The intrinsic components are obtained from the Q-V characteristics similar to what was depicted in Fig. 3.2, but with quasi-ballistic factors considered. The estimated individual component values are listed in Table 3.1.

Table 3.1. List of parameters for propagation delay estimation

	$I_{eff}$ (mA/ $\mu$ m)	$C_{intrinsic}$ (fF/ $\mu$ m)	$C_{fringe}$ (fF/ $\mu$ m/side)	$C_{total}$ (fF/ $\mu$ m)
InGaAs	3.5	0.29	0.3	4.3
Si	1.7	0.58	0.3	5.7

The propagation delay is calculated as [7]

$$\tau = \frac{C_{eff}V_{DD}}{2I_{eff}} \quad (3.13)$$

Resultant propagation delays are 0.87ps and 1.43ps for the InGaAs device and the Si device, respectively. To verify the correctness of the analytical model, we further fit the device characteristics to a Curtice Cubic model, and performed a simulation on a ring oscillator constructed with such devices. Representative waveforms taken from succeeding stages of the ring oscillator are shown in Fig. 3.12 for an InGaAs quasi-ballistic FET. The extracted propagation is 0.83ps, in excellent agreement with the analytical result.

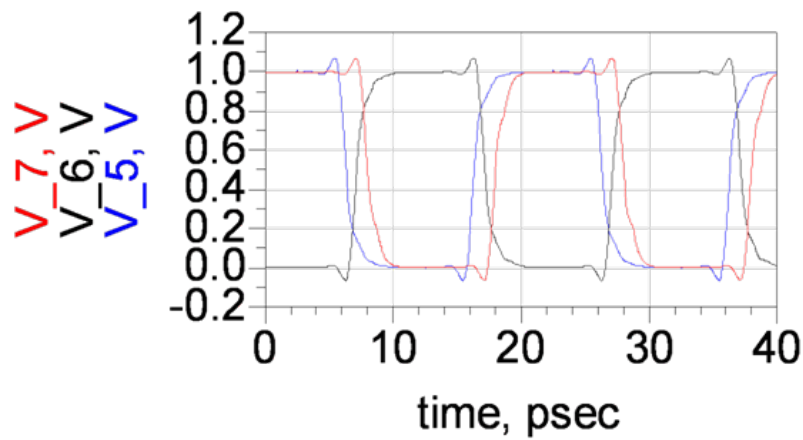


Fig. 3.12. Representative waveforms from succeeding stages of a ring oscillator constructed with InGaAs ballistic MOSFETs, whose characteristics are described by a Curtice Cubic model.

### 3.5 Discussions and Summary

In this section, we will discuss some limitations of the (quasi-)ballistic model presented previously and summarize the proceeding chapter.

The quasi-ballistic model presented here is aimed to provide guidelines and first order estimations for actual device designs. Its accuracy degrades when extremely short devices (or devices with bad SCE's) are considered. Via contrast and comparison, we found that high mobility, although does not provide a direct benefit, does result in better ballistic efficiency and low series resistance. These improvements may help InGaAs to behave closer to its ballistic limit and outperform its Si counterpart. It is also shown that at pure ballistic limit, InGaAs is only slightly better than Si in terms of on state current drive.

There are still potential limitations of the (quasi-)ballistic formulation presented to date. Two basic assumptions may not hold true in practical device operation. The first assumption that virtual source is always in thermal equilibrium with the source may fail for extremely short channel devices. It has been shown in [9] through full band Monte Carlo simulations that the virtual source may be physically far away from the source, which limits the rate of carrier exchanges between the source and virtual source. Therefore, thermal equilibrium may not be readily established. The continuous streaming of carriers with longitudinal momentum via the channel also undermines the thermal equilibrium assumption (this will be discussed further in next chapter). The second consideration comes from the electrostatic perspective. Under the situation where large current density occurs, the carrier density close to the drain will have a significant effect on the electrostatic potential close to the source region. Such modulation (which can be understood as DIBL-like effect) changes the carrier density at the virtual source, adding

additional 2-D effects to the current model. Such a 2-D effect is addressed in [10] via adding a coupling capacitance between drain and virtual source.

The formalism may be further expanded to accommodate more than one subband to be significantly occupied. The computation may make use of a slightly modified self-consistent Schrödinger Poisson solver. The only modification that needs to be done is to take into account the  $V_{DS}$  dependent carrier computation for each subband, similar to that has been described by Eq. (3.1) for the first subband. i.e. modify the Fermi Dirac function in this case.

## Reference:

1. P. M. Solomon and S. E. Laux, "The Ballistic FET: Design, Capacitance and Speed Limit," IEDM Tech. Digest, pp. 511-514 (2001)
2. K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor", J. Appl. Phys. vol. 76, pp. 4879-4890
3. M. S. Lundstrom and Z. Ren, "Essential physics of nanoscale MOSFETs," IEEE Trans. Electron Devices, vol. 49, pp. 133-141, Jan. 2002
4. Yuan Taur, Tak Ning, "Fundamentals of Modern VLSI Devices", 2<sup>nd</sup> Edition, Cambridge University Press, 2009
5. Mark Rodwell et. al, "Technology development & design for 22 nm InGaAs/InP channel MOSFETs (invited talk)", Indium Phosphide and Related Material Conference, 2008
6. P. Bai, et al., "A 65nm Logic Technology Featuring 35nm Gate Lengths, Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-k ILD and 0.57  $\mu\text{m}^2$  SRAM Cell," International Electron Devices Meeting, pp. 657, 2004.
7. M. H. Na, E. L. Nowak, W. Haensch, J. Cai, "The effective drive current in CMOS inverters," pp. 121, IEDM 2002
8. ADS manual, Agilent Inc.  
[://edocs.soco.agilent.com/display/ads2008U2/Curtice3+Model+\(Curtice-Cubic+GaAsFET+Model\)](http://edocs.soco.agilent.com/display/ads2008U2/Curtice3+Model+(Curtice-Cubic+GaAsFET+Model))
9. M. V. Fischetti, S. Jin, T.-W. Tang, P. Asbeck, Y. Taur, S. E. Laux, M. Rodwell and N. Sano, "Scaling MOSFETs to 10 nm: Coulomb effects, source starvation, and virtual source model", Journal of Computational Electronics, Jul. 2009.
10. M. Lundstrom, J. Guo, "Nanoscale Transistors: Device Physics, Modeling And Simulation", Springer, 2006

## Chapter 4. Source Design of III-V MOSFETs

### 4.1 Introduction

It is shown in previous chapters that a III-V MOSFET, if designed appropriately, may provide significant output current drive for high performance applications. However, it must be noticed that in the previous analysis, we have assumed that the carriers that are needed to sustain the high current are always readily supplied by the source. This requires that the source replenishes the vacancies left by carriers exiting the source in a timely fashion. Such conditions have mostly been met in the bulk Si CMOS device designs so far, however this may not be the case in the III-V MOSFET that we propose here.

In this section, we intend to introduce a qualitative picture to describe the nature of this problem, while following up in the succeeding sections with more quantitative models. Sketched in Fig. 4.1 is a basic picture of the carrier distribution in the channel and the source region. We would like to point out several aspects of this problem qualitatively in this section.

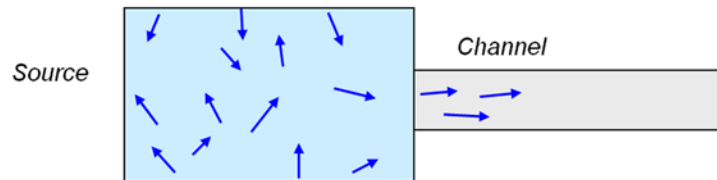


Fig. 4.1. Schematic drawing of the carrier distribution within the channel and the source region.

Before the discussion, it is necessary to present some of the definitions and caveats. In the following, the current transport direction is referred to longitudinal direction, while the epitaxial growth direction is referred to transverse direction. It is also necessary to consider the schematic drawing above within the framework of the entire device, which was shown previously

in Sec. 2.4. In order to suppress the Short Channel Effects (SCE) and enable the device design to work down to 20nm gate length, the channel is designed to be a thin quantum well with a p<sup>+</sup> ground plane underneath (resembles an asymmetrical double gate device). In this regard, the most efficient transport of electrons occurs for electrons with a primarily longitudinal momentum component, so that few reflections from the quantum well top and bottom boundaries happen during the course of transport. Such a situation closely resemble to the photon transmission within an optical fiber. In fact, if we consider the channel as a waveguide for electrons, we may borrow some well defined terms and established wisdom from the optical fiber research to illustrate the problem here. Given the thickness of the channel (2.5nm~5nm), the “numerical aperture” of such a waveguide is relatively small. Moreover, as the majority of the electrons that travel within the channel possess a large longitudinal momentum component, the vacancies left behind in the source are specifically associated with these longitudinal electrons. On the other hand, the electrons that dwell in the source initially have a random distribution of momentum. As large quantities of longitudinal electrons are pulled out from the source, it is necessary for the source to replenish these vacancies via scattering processes. When the replenishing process is not fast enough, the overall output current suffers, and we name this phenomenon “source starvation”. It will become clear later that this problem is unique to III-V due to its small DOS, while the Si MOSFET exhibit insignificant source starvation (the “numerical aperture” issue, however, will still impact the current for a Si MOSFET). Dominant scattering processes within the source include ionized impurity scattering and electron-electron scattering, both of which are related to the doping concentration of the source.

In this section, we will present discussions in the following sequence: a back-to-the-envelop model is provided in section 4.2, followed by a more refined model that entails various scattering mechanisms in section 4.4.

## 4.2 Simplified Model on Source Design

In this section, we present a simplified model that considers the electrons within the source as classical gas. We will defer the detailed discussion on the scattering process to the next section; instead we simply assume a phenomenological lifetime here for all electrons regardless of their energies and momentums.

We start to consider the problem with following geometry as shown Fig. 4.2.

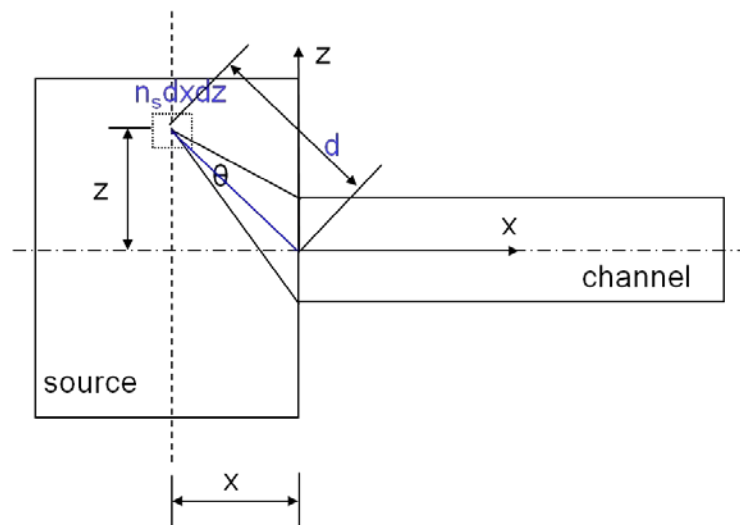


Fig. 4.2. Coordinates setup for the computation of source starvation.

We define the following quantities:

$t_{ch}$  – thickness of the channel

$t_s$  – thickness of the source

$x_s$  – width of the source (extent on the x-direction)

$n_s$  – volume carrier density, assumed to be the same as the doping concentration (doping level higher than Mott transition point [1])



$\theta$  – angle between the k-vector and the x-axis for a specific electron

$\tau$  – average lifetime of the electron in the source

For an infinitesimal area  $dx dz$  centered at  $(x, z)$ , the total number of carriers that reside within the area is given by  $n_s dx dz$  (normalized by width). We further assume that these electrons distribute uniformly in the momentum space (k-space). However, it must be realized that only electrons whose k-vectors are directed towards the channel may be collected and subsequently contribute to the overall flow. In Fig. 4.2, such directions are bounded by the two line segments that connect the point of  $(x, z)$  and the upper/lower boundary of the channel region. Mathematically, this criterion translates into the allowed range for  $\theta$ , whose boundaries are given by the following equations:

$$\begin{aligned}\theta_1 &= \arctan\left(\frac{z - t_{ch}/2}{x}\right) \\ \theta_2 &= \arctan\left(\frac{z + t_{ch}/2}{x}\right)\end{aligned}\tag{4.1}$$

With the assumed uniform carrier distribution in the momentum space, the number of carriers that reside with a unit angular range is given by  $\frac{n_s dx dz}{2\pi}$ .

Not all electrons at point  $(x, z)$  within the appropriate k-vector direction may reach the channel, due to the scattering along the way. We approximate the channel collection probability of an electron that resides initially at point  $(x, z)$  by  $\exp(-d/\lambda)$ , where  $d$  is the distance between point  $(x, z)$  and the channel center, given by  $d = \sqrt{x^2 + z^2}$ ; and  $\lambda$  is the mean free path which is taken here to be the product of the average lifetime of an electron and its average thermal velocity.

With these considerations, we may calculate the forward going current contribution (normalized to width) of the electrons from the infinitesimal area  $dx dz$ .

$$dI = \frac{1}{2\pi} \cdot \frac{q}{\tau} \int_{\theta_1}^{\theta_2} d\theta \cdot (n_s dx dz) \cdot \cos\theta \cdot \exp\left(-\frac{\sqrt{x^2 + z^2}}{v_{th} \cdot \tau}\right) \quad (4.2)$$

The  $\cos\theta$  term in Eq. (4.2) accounts for the x direction only current transport.

Eq. (4.2) may be further integrated throughout the entire source region, and the total per unit width current may be found, as expressed below:

$$I = \frac{n_s}{\pi} \cdot \frac{q}{\tau} \int_0^{-x_s} dx \int_0^{t_s/2} dz \int_{\theta_1}^{\theta_2} d\theta \cdot \cos\theta \cdot \exp\left(-\frac{\sqrt{x^2 + z^2}}{v_{th} \cdot \tau}\right) \quad (4.3)$$

We utilized the symmetrical property with respect to the x axis while deriving Eq. (4.3) and only carried out integration along the z direction up to  $t_s/2$ .

Using Eq. (4.3), we may examine several impact to the drain current with various source design parameters. As we showed previously in the mathematics section, various device parameters may impact the outcome of overall supportable current:

1. Electron lifetime  $\tau$ , influenced by doping level (and electron concentration) within the source.
2. Channel width, influences  $\theta_1$  and  $\theta_2$ , or the collection angle range.
3. Channel thickness, only expecting influences to the overall current when the thickness is small. When the thickness exceeds a certain limit, little benefit results due to the exponential factor for the collection probability.

The results of the computation are shown in the Fig. 4.3 and Fig. 4.4. Fig. 4.3 shows the effect of source doping with a source thickness of 20nm (raise S/D configuration) and channel thickness of 5nm. It is seen that the maximum sustainable current increases with increased source doping. Given a source doping level, the output current initially increases with average lifetime, as the MFP increases proportionally with the lifetime and henceforth enhances the effective collection range of the channel. As average lifetime further increases, the decrease in the replenish rate becomes dominant, and causes a slump in the overall output current. Fig. 4.4 shows the effects of other parameters such as channel thickness and source thickness. It is necessary to clarify the design options: 1) Nominal design corresponds to  $2 \times 10^{19} \text{ cm}^{-3}$  source doping, channel thickness of 2.5nm and source thickness of 10nm; 2) Raised S/D design corresponds to 20nm source thickness; 3) Higher doping refers to  $5 \times 10^{19} \text{ cm}^{-3}$ ; 4) Thicker channel refers to a channel thickness of 5nm. It is seen from the figure that for optimal current sustainability, raise S/D (which is equivalent to thicker source configuration), higher source doping and thicker channel are necessary.

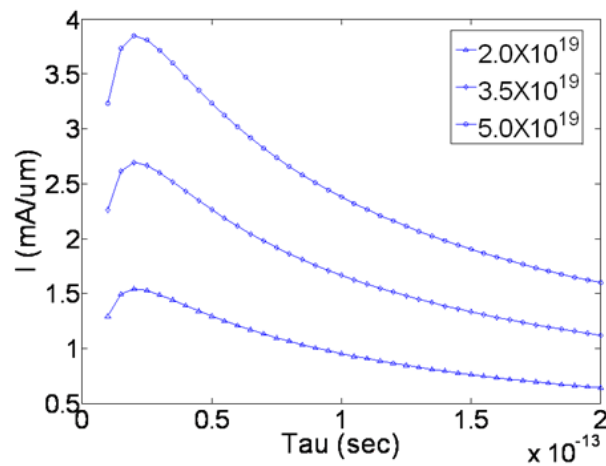


Fig. 4.3. Comparison among source designs with different doping concentrations. The geometrical dimensions of the structure simulated above are: channel thickness of 5nm and source thickness of 20nm.

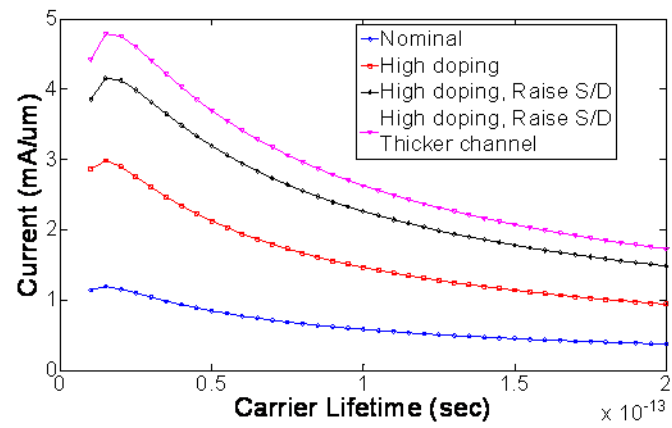


Fig. 4.4. Comparison of source designs with different geometrical parameters and doping levels.

### 4.3 Simplified Model on Source Starvation

In this section, we will present a physical model that describes the “source starvation” phenomenon. The model is based upon the geometrical coordinates as illustrated in Fig. 4.2. The primary purpose of the model is to calculate the average lifetime a function carrier momentum and therefore give a better account on how scattering mechanisms affect the source current supply capability.

At a given position  $(x, z)$  at the junction, let us consider an electron which is described by  $(k_x, k_z)$  in momentum space, being removed from the source side to contribute to overall current flow. At the same time that the electron is removed, a vacancy is created at the junction. Electrons that dwell at nearby locations (however not necessarily with the same momentum) will attempt to fill up the vacancy so that the overall current may be sustained. There are many possible electrons that may fulfill such a task, which makes consideration from an electron standpoint more involved. Therefore, in the following treatment, we instead trace the vacancy itself (analogous to the concept of holes in the valence band except that the vacancy resides in the conduction band). The process of an electron filling the vacancy may be considered as an inverse procedure to that of a vacancy occupying a state of an electron  $(k_x', k_z')$  in the momentum space. The momentum of the state may not necessarily match the original momentum  $(k_x, k_z)$ , and the process that bridges the momentum transition is scattering. In this section, we intend to develop a method along the same line to assess the replenishing rate and further compute the sustainable current level given a specific design. To achieve this goal, it is necessary to obtain the average lifetime of a specific vacancy that is left behind by an electron with  $(k_x, k_z)$  momentum, subject to the scattering processes that take place within the source.

As mentioned previously, there are two dominant scattering processes within the source – impurity scattering and electron-electron scattering. Since we are treating the vacancy as if it is an electron, the same scattering considerations apply to the vacancy. The fill-up of a vacancy may be completed with an electron in a different state with the momentum space through either process. The considerations of the individual processes will be detailed in the next two subsections, followed by a subsection discussing the computation of current level with the knowledge of average lifetime.

#### ***4.4.1 Ionized Impurity Scattering***

The dominant scattering mechanisms within the source region are ionized impurity scattering and electron-electron scattering. In this sub-section, we focus on ionized impurity scattering.

Similar to the normal ionized impurity scattering process involving electrons, the scattering of the vacancy also possesses an elastic property, i.e. the kinetic energy (magnitude of the momentum vector) of the vacancy after the scattering remains the same as before the scattering. The momentum, however, does not necessarily remain the same. The transition in the k-space is schematically illustrated in Fig. 4.5, where the final momentum lies on the same equi-energy contour (circle in this case if a parabolic band is assumed).

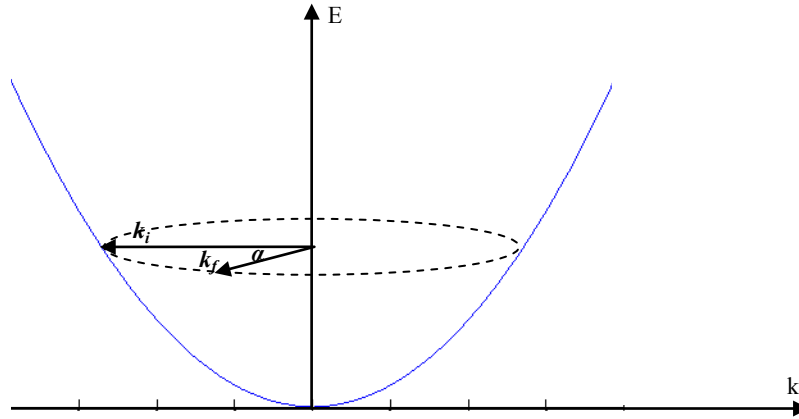


Fig. 4.5. Illustration of ionized impurity scattering in the k-space.  $k_i$  and  $k_f$  are the initial and final momentums, respectively;  $\alpha$  is the angle between the two vectors.

Once the scattering happens, the resultant momentum (in terms of its direction) may be determined via a random choosing procedure weighed by the scattering rate as a function of scattering angle. It will be shown shortly that the average lifetime for the ionized impurity scattering is merely a function of the momentum magnitude, which leaves the determination of the exact angle after scattering unnecessary for current computation. However, for completeness, we include the method that may be followed to determine the direction of the final momentum below as well.

Given the above arguments, there are two tasks to be fulfilled here:

- 1) Determine the scattering rate (or lifetime as the inverse of the scattering rate) associated with the ionized impurity scattering. It must be noted the computation is aimed to obtain momentum relaxation time instead of energy relaxation time, since the scattering must randomize the distribution in the momentum space such that specific states with large longitudinal momentum that have been streamed out by the channel get replenished within the source.

- 2) Once the scattering rate is determined, the ionized impurity scattering event is determined via the procedure in the previous sub-section. We determine the momentum of final states with the consideration of the large angle discrimination mentioned above.

The implementation of step 1) may be achieved readily by using the established ionized impurity scattering results [2], as given by:

$$\tau(\vec{k}) = \frac{16\sqrt{2m^*}\pi\epsilon_s^2}{N_I q^4} \left[ \ln(1+\gamma^2) - \frac{\gamma^2}{1+\gamma^2} \right]^{-1} E^{3/2}(k) \quad (4.4)$$

where  $N_I$  is the volume density of the ionized impurity,  $\epsilon_s$  is the dielectric constant of the material.

Additionally, the parameter  $\gamma$  is given by:

$$\gamma = 8m^* E(k) L_D^2 / \hbar^2 \quad (4.5)$$

where  $L_D$  is the Debye length to take into account the screening effect, and  $h$  is the Planck's constant. It must be noted that the source is heavily doped such that the Debye length for a highly

degenerated semiconductor should be used instead ( $L_D = \sqrt{\frac{\epsilon_s (kT)^{3/2}}{\sqrt{E_f - E_c}} \cdot \frac{1}{q^2 N_c}}$ ).

The implementation of step 2) makes use of two random numbers  $r_{i1}$  and  $r_{i2}$ , both of have values between 0 and 1. The first random number  $r_{i1}$  is used to determine the angle between the initial and final momentum directions, which span over the range of  $0$  to  $\pi$ . From literature [2], the angular dependence of the scattering rate  $S(\mathbf{k}, \mathbf{k}')$  is given by:

$$S(\vec{k}, \vec{k}') \propto \left[ \frac{1}{(4|k|^2 \sin^2(\alpha/2) + (1/L_D)^2)} \right]^2 \quad (4.6)$$



where  $\alpha$  is the angle between the initial and final momentum. We define the following probability density function:

$$p(\alpha) = \frac{\left[ \frac{1}{(4|k|^2 \sin^2(\alpha/2) + (1/L_D)^2)} \right]^2}{\int_0^\pi \left[ \frac{1}{(4|k|^2 \sin^2(\alpha/2) + (1/L_D)^2)} \right]^2 d\alpha} \quad (4.7)$$

To determine the angle  $\theta$  between final and initial state momentums, we require that  $\theta$  satisfies the following equation:

$$r_{il} = \int_0^\theta p(\alpha) d\alpha \quad (4.8)$$

where  $r_{il}$  is a random number generated between 0 and 1. By using the probability density function defined in Eq. (4.9) when computing  $\theta$ , we ensured that the scattering angle reflects the correct large angle discrimination for the ionized impurity scattering process. A representative probability density function as defined in Eq. (4.9) is plotted in Fig. 4.6, and the procedure of determining  $\theta$  is illustrated in the figure as well.

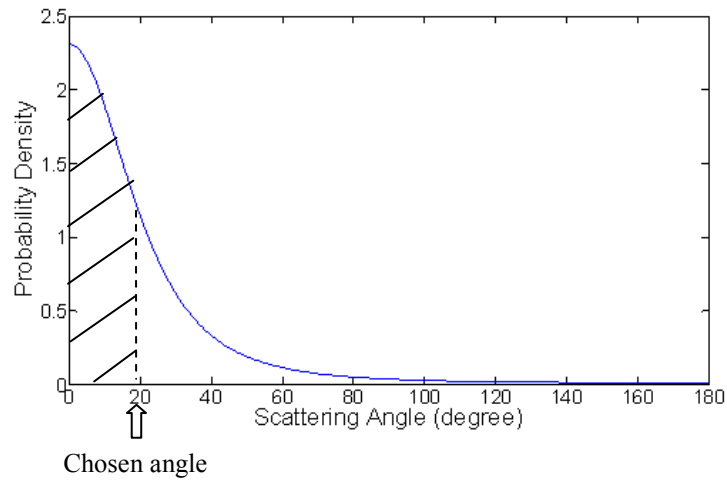


Fig. 4.6. Illustration of probability density as a function of the scattering angle. The angle may be determined by requiring that the area of the shaded region beneath the curve to be equal to a random number generated between 0 and 1. The angle at which such a requirement is satisfied is chosen to be the scattering angle.

It is realized that the angle between two momentum vectors is defined in terms of the absolute value, i.e. there are two possible final states that are separated from the initial momentum by the same angle. Therefore, we need a second random number  $r_{i2}$  to determine whether the final state momentum is parted from the initial state momentum by  $\theta$  or  $-\theta$  (counterclockwise direction is defined as positive). If  $r_{i2}$  is between 0 and 0.5, we assert that the angle is  $\theta$ , otherwise  $-\theta$ .

Representative momentum relaxation time is plotted as a function of energy (momentum magnitude) in Fig. 4.7, where doping level is used as a stepping parameter.

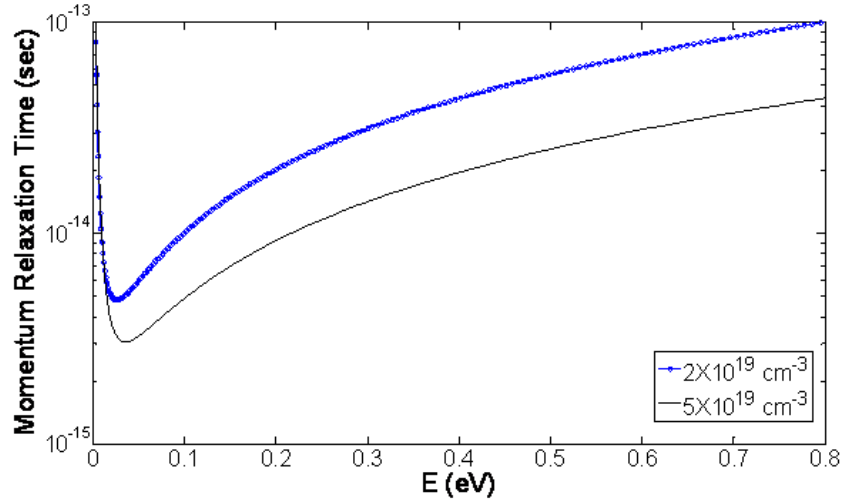


Fig. 4.7. Momentum relaxation time as a function of electron energy for different doping levels.

#### 4.4.2 Electron-Electron Scattering

In this sub-section, we discuss another possible scattering mechanism that may take place within the source region – electron-electron scattering. During the course of scattering, the following must be satisfied:

- 1) Before and after the collision, the momentum must be conserved:

$$\vec{k}_v' + \vec{k}_e' = \vec{k}_v + \vec{k}_e \quad (4.9)$$

where the primed quantity denotes the final state, footnote “v” denotes the vacancy related quantity and footnote “e” denotes the electron related quantity.

- 2) The vacancy may collide with electrons with any momentum, and the collision may be considered as elastic [2]. Via momentum conservation and kinetic energy conservation, the momentums for both the vacancy and the electron may be found.

Similar to the ionized impurity scattering described in the last sub-section, we need to identify the following quantities to implement the simulation:

- 1) The momentum relaxation time to facilitate the determination of scattering type.
- 2) The possible positions in the momentum space of electrons that may collide with the vacancy.

We will see in the following discussion that 1) and 2) are closely related and it is simpler to start with considering 2) from a physically intuitive way. If we denote the transition rate as  $S(\mathbf{k}_e, \mathbf{k}_v; \mathbf{k}_e', \mathbf{k}_v')$  between the initial state pair  $(\mathbf{k}_e, \mathbf{k}_v)$  to the final state pair  $(\mathbf{k}_e', \mathbf{k}_v')$ , we realize that in order for such a transition to happen, an initial state  $\mathbf{k}_e$  must be occupied by an electron while the final state  $\mathbf{k}_e'$  must be empty. Mathematically, such a transition is subject to the factor of  $f(\vec{k}_e) \bullet [1 - f(\vec{k}_e')]$ .

The scattering rate may be derived from the Fermi Golden rule, and is given by:

$$S(\vec{k}_v, \vec{k}_e; \vec{k}_v', \vec{k}_e') = \frac{2\pi q^4}{\hbar \varepsilon_s^2} \bullet \frac{1}{(|\vec{k}_v - \vec{k}_v'|^2 + \frac{1}{L_D^2})^2} \quad (4.10)$$

Eq. (4.10) is a slightly modified result from what was given in [2], without the assumption that the magnitude of  $\mathbf{k}_v$  and  $\mathbf{k}_v'$  are the same. The momentum relaxation time can be further computed with:

$$\frac{1}{\tau(\vec{k}_v)} = \sum_{\vec{k}_e} \sum_{\vec{k}_v'} S(\vec{k}_v, \vec{k}_e; \vec{k}_v', \vec{k}_e') \left(1 - \frac{|\vec{k}_v'|}{|\vec{k}_v|} \cos \alpha\right) f(\vec{k}_e) \bullet [1 - f(\vec{k}_e')] \quad (4.11)$$

Plotted in Fig. 4.8 is the momentum relaxation time as a function of the 2-D momentum of the vacancy. In the plot, it is assumed that the doping level is N-type  $2 \times 10^{19} \text{ cm}^{-4}$ .

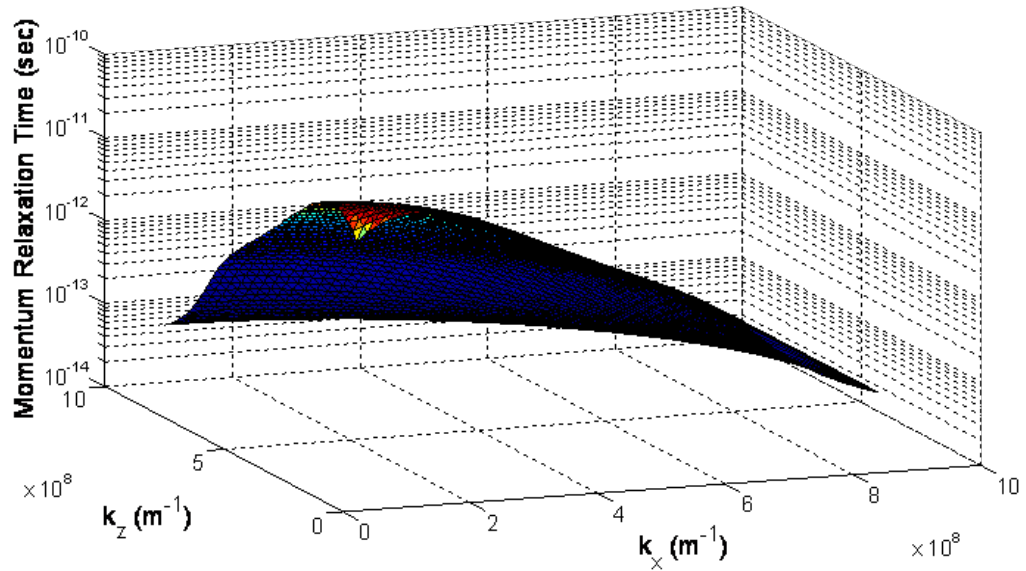


Fig. 4.8. Momentum relaxation time due to the electron-electron(vacancy) scattering process mapped out on a 2-D momentum space.

Once it is determined that the electron-vacancy collision has happened, we can further identify two quantities (although not necessary for the current computation purpose): final state momentum for the vacancy and initial state momentum for the electron. Selection of the final state momentum for the vacancy makes use of the following probability density function with a random number  $r_{el}$  generated between 0 and 1:

$$p(\vec{k}_v') = \frac{\sum_{\vec{k}_e} S(\vec{k}_v, \vec{k}_e; \vec{k}_v', \vec{k}_e') \left(1 - \frac{|\vec{k}_v'|}{|\vec{k}_v|} \cos \alpha\right) f(\vec{k}_e) \cdot [1 - f(\vec{k}_e')]}{\sum_{\vec{k}_v'} \sum_{\vec{k}_e} S(\vec{k}_v, \vec{k}_e; \vec{k}_v', \vec{k}_e') \left(1 - \frac{|\vec{k}_v'|}{|\vec{k}_v|} \cos \alpha\right) f(\vec{k}_e) \cdot [1 - f(\vec{k}_e')]} \quad (4.12)$$

Plotted in Fig. 4.9 is a representative situation described by Eq. (4.12) with a chosen  $\vec{k}_v$ . The droop of the probability density function beyond  $k_x = 6 \times 10^8 \text{ m}^{-1}$  originates from the modulation of

the Fermi-Dirac distribution (the resultant electron momentums reside in states that are not occupied in the momentum space).

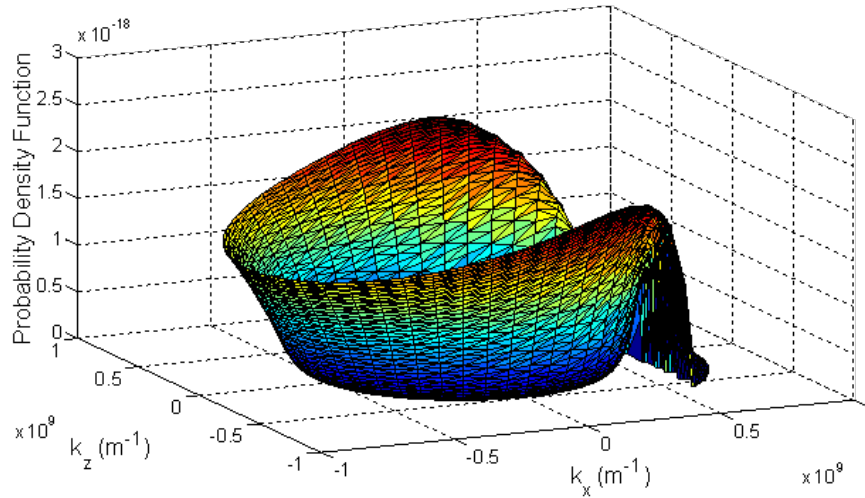


Fig. 4.9. Probability density as a function of the out-scattered vacancy momentum mapped onto the 2-D momentum space.

The selection of the electron must link with the likelihood of such a collision happening, i.e. determined by the factor of  $f(\vec{k}_e) \cdot [1 - f(\vec{k}_e')]$ . Every  $\mathbf{k}_e, \mathbf{k}_e'$  may be uniquely determined through momentum conservation and energy conservation. One can therefore define a probability density function similar to the one expressed by Eq. (4.9), using  $f(\vec{k}_e) \cdot [1 - f(\vec{k}_e')]$  with proper normalization. Attention needs to be paid to the fact that  $\mathbf{k}_e$  is a vector; therefore a double integral is utilized along with the random number  $r_{e2}$  generated between 0 and 1 to determine the electron momentum. A representative probability density function is shown in Fig. 4.10 for electron-vacancy scattering as a function electron momentum given certain initial and final vacancy momentums.

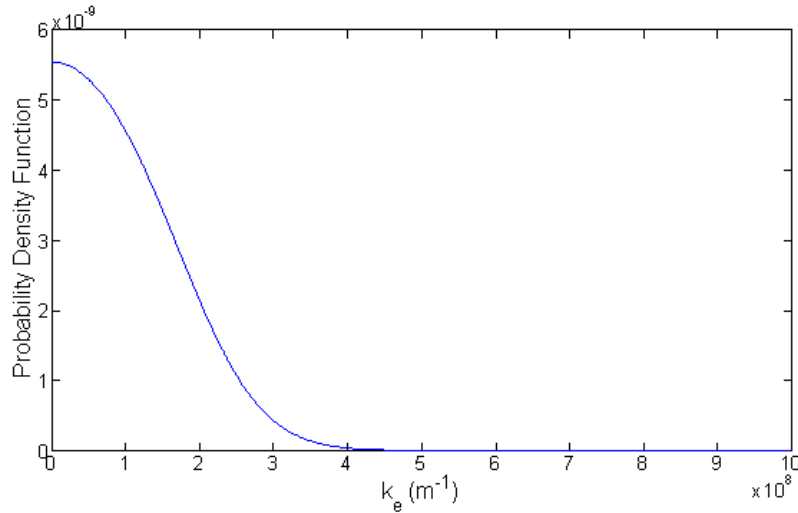


Fig. 4.10. Probability density of having a certain out-scattered electron momentum.

To this point, all the quantities have been determined for the electron-vacancy scattering.

### 4.3.3 *Source Supportable Current Computation*

Before evaluating the sustainable current level that is limited by a certain source design, it is instructive to compare the two scattering mechanisms that have been discussed in the previous two sections. Comparing Fig. 4.7 and Fig. 4.8, we notice that the momentum relaxation time for the two scattering mechanisms differs considerably, although they all fall into the category of the Coulomb interaction. The ionized impurity scattering exhibits a shorter momentum relaxation time in general, as the scattering process only involves one constraint – the energy of the initial and final states must be the same. Whereas, in addition to the energy conservation, there are more constraints for the electron-electron (vacancy) scattering: the states involved in this process are subject to the Fermi-Dirac distribution, so that only states within the appropriate energy range may participate in the scattering. This is evident from Fig. 4.10, where only electrons with momentums falling within a small range contribute to the scattering given the initial and final vacancy states. Whereas, for ionized impurity scattering, the scattering center

(which is the ionized impurity) does not have such a constraint. Mathematically, this constraint is represented by the Fermi-Dirac distribution that weighs into the relaxation time evaluation.

Since the vacancy state replenishing time is highly dependent on the source doping level, in the following discussion, we focus on the evaluation of the supportable current from the source with various doping levels. The procedure is outlined below:

- 1) Given a doping level, the momentum relaxation times may be evaluated for both the ionized impurity scattering process and the electron-electron scattering process as functions of the vacancy momentums (2D vector).
- 2) The overall momentum relaxation time, as a function of 2D momentum, may be obtained via:

$$\frac{1}{\tau_{overall}} = \frac{1}{\tau_{impurity}} + \frac{1}{\tau_{electron}} \quad (4.13)$$

- 3) The overall current may be evaluated by a sum over the momentum space and real space similar to what we have done in Sec. 4.2, except that the momentum relaxation time is now connected to physical scattering processes.

$$I = \frac{q}{\pi} \cdot \sum_{\vec{k}_v} \frac{f(\vec{k}_v)}{\tau(\vec{k}_v)} \int_0^{-x_s} dx \int_0^{t_s/2} dz \int_{\theta_1}^{\theta_2} d\theta \cdot \cos \theta \cdot \exp\left(-\frac{m^* \cdot \sqrt{x^2 + z^2}}{\hbar |\vec{k}_v| \cdot \tau(\vec{k}_v)}\right) \quad (4.14)$$

where  $f(\mathbf{k}_v)$  is the Fermi-Dirac distribution function,  $\tau(\mathbf{k}_v)$  is the momentum relaxation time that includes both the ionized impurity scattering mechanism and the electron-electron scattering mechanism. It is noted that Eq. (4.14) prescribes a triple integral within the k-space and double integral within the real space. The computation cost of such a procedure is quite high, with one data point taking ~10 hours on a 2.0GHz dual-core CPU, 2GB RAM Thinkpad laptop .



A representative overall momentum relaxation time is plotted over the 2-D momentum space shown in Fig. 4.11, with a source doping level of  $2 \times 10^{19} \text{ cm}^{-4}$ . The source supportable current is plotted as a function of source doping level in Fig. 4.12.

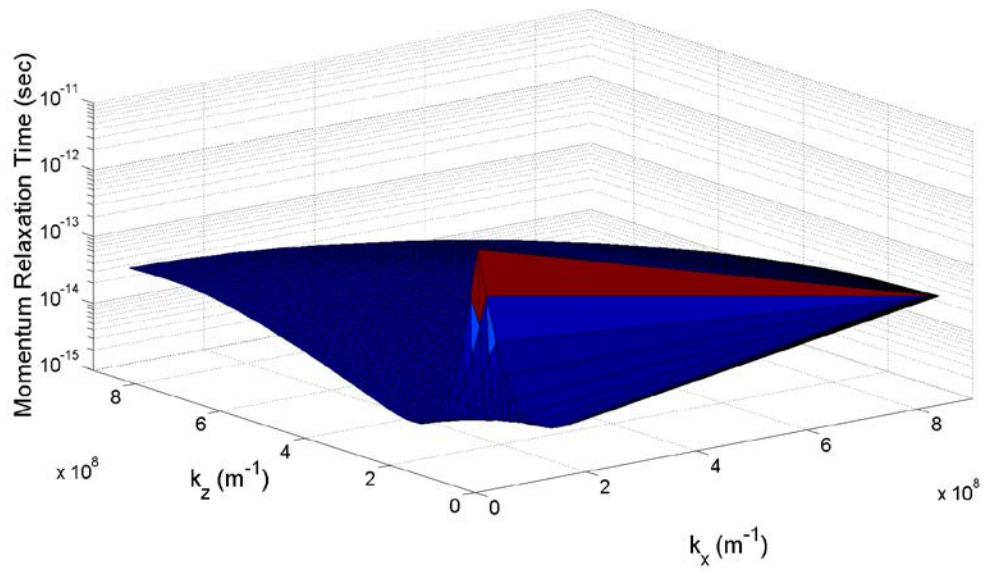


Fig. 4.11. Overall momentum relaxation time as a function of momentum mapped onto the 2-D momentum space.

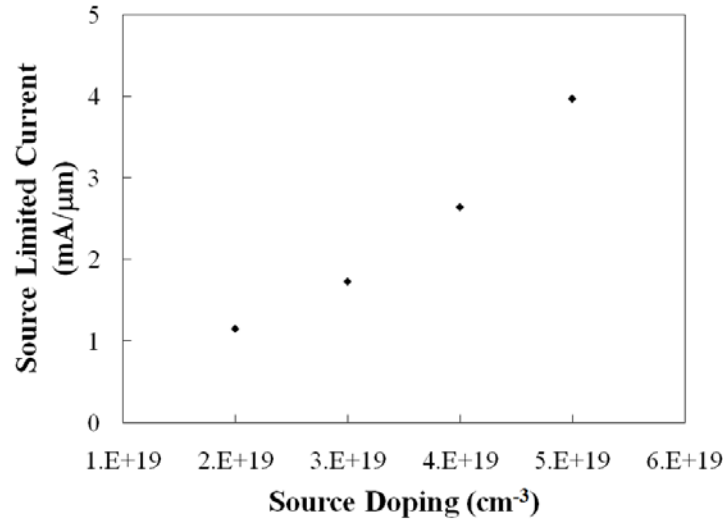


Fig. 4.12. Numerical simulation results for the source limited current with ionized impurity and electron-electron scattering taken into account.

It must be noted that the results in Fig. 4.12 does not include the effect from non-parabolic conduction band in the source. Qualitatively, the non-parabolicity increases the DOS and therefore increases the possible means for scattering. As a result, it is expected that the effective lifetime will decrease which may lead to a larger source limited current. It is also noticed that the source limited current is a super-linear function of the doping level. It may be understood from the following physically intuitive perspectives: 1) For ionized impurity scattering, higher doping not only increases the scattering center (see Eq. (4.4), where  $\tau$  is inversely proportional to  $N_I$ ), but also elevates the Fermi level within the source which increases the possible states that the vacancy may scatter to. Therefore, the replenishing process is super-linearly dependent on doping level for ionized impurity scattering. 2) A similar argument may hold for electron-electron scattering, and additionally due to the binary nature of this scattering process, its rate is approximately proportional to  $n^2$ . Given these considerations, it is expected that the resultant source limited current will be a super-linear function of doping.

It is also worthwhile to point out that the current limit from the source is only one aspect of the limits. Among others, the device current may also be limited by the channel design (e.g. ballistic limit). The goal for source design is to unveil the full potential of the channel, i.e. to provide a larger source limited current than the channel limited current.

The source starvation problem has not been identified previously in Si MOSFET design, and is unique to the III-V MOSFET design. It is worthwhile to understand this point from a physical perspective. There are two major differences in the source design for a Si MOSFET and a III-V MOSFET. 1) The source of a Si MOSFET is typically doped very high, in the range of low to mid  $10^{20} \text{ cm}^{-4}$ . 2) Silicon has a much larger density of states due to higher mass and multi-valley degeneracy. The first factor directly results in higher scattering rate for the ionized impurity scattering (directly proportional to the volume density of the ionized impurity, as seen from Eq. (4.4)). The second factor may be understood as following: when an electron is scattered from an initial state to a final state, the number of possible scattering is proportional to the number of the final states; having higher value of DOS means more states within the same energy interval, therefore also increases the scattering rate and replenish the vacant states more readily.

In this section, we presented a physical model to describe the source starvation problem that is unique to the III-V MOSFET designs. Dominant scattering mechanisms such as ionized impurity scattering and electron-electron scattering are described and implemented in computation. It is found that the ionized impurity scattering accounts for more replenishing the vacant longitudinal states given its shorter momentum relaxation time in comparison to the electron-electron scattering.

## 4.4 Summary

In this chapter, we discussed another limiting factor to the overall current capability for a specific III-V MOSFET design – source starvation. It is necessary to realize that for a III-V MOSFET, the current capability may be limited not only by the channel design (e.g. the channel mobility), but also by the source configuration. Through the simplified model that treats the electrons within the source as ideal gas (Sec. 4.2), we learned various factors may affect the final performance of the device: geometry of the source (thickness, width), geometry of the channel and the doping level of the source. We further verified the findings by modeling with actual scattering mechanisms that may be dominant in the source, as presented in Sec. 4.4. It must be realized that although the model in Sec. 4.3 takes into account physical scattering processes, is only a first order description of the actual situation. We have assumed in the model that the source does not deviate significantly from its thermal equilibrium state, which is only true with a reasonable source design. For a design with significantly lower doping, this assumption may not be valid. From a more elaborate simulation [3] performed in Damocles [4], which is a full band Monte Carlo simulator, it is understood that significant deviation from quasi-equilibrium state is observed for lower source doping concentrations. This inappropriate source design further limits the overall device performance.

## Reference:

1. N. Mott, "Metal-Insulator Transition", Rev. Mod. Phys., vol. 40, Oct. 1968, pp.677-683
2. M. Lundstrom, "Fundamentals of Carrier Transport" 2nd Ed. Cambridge University Press
3. M.V. Fischetti, L. Wang, B. Yu, C. Sachs, P. M. Asbeck, Y. Taur, and M. Rodwell, "Simulation of Electron Transport in High-Mobility-MOSFETs: Density of States Bottleneck and Source Starvation", IEDM 2007
4. M. V. Fischetti, Steve Laux, "Damocles Theoretical Manual", IBM, 1996

## Chapter 5. Dielectric Layers for III-V MOSFETs

### 5.1 Introduction

As mentioned at the very beginning of this part, although III-V materials have been known to have great material properties for a long time, a MOSFET that utilize these superior properties has only been demonstrated recently [1]. Despite the fact that III-V materials thrived in different forms of FET (e.g. HFETs) which exhibit the highest speed thus far for all the FET devices, the same device configuration is not suitable for highly integrated digital applications. It is worthwhile to point out some of the reasons here: 1) HFETs are typically depletion mode devices (to minimize series resistance at source and drain since self-aligned source and drain is unavailable in the past III-V technology); 2) high gate bias may induce excessive gate leakage as gate to barrier layer junction is Schottky. Therefore, in order to exploit the superior properties of III-V material, it is necessary to find out a viable solution to configure a III-V MOSFET with self-aligned source and drain (so that enhancement mode operation is useful), and to have an appropriate oxide to effectively reduce the gate leakage current. A III-V MOSFET configuration may provide solutions to the previous problems; however such development has been hindered by the poor oxide/III-V semiconductor interface property. It has been long noticed that the oxide/III-V interface is plagued with large density of interface states, which sometimes results in pinned surface, rendering inefficient gate modulation. The first successful oxide (GaGdO) on III-V material (GaAs) was demonstrated by Passlack et.al. [1], which exhibits unpinned surface, low density of interfacial trap density and high transconductance measured from the experimental devices with these oxide.

It is instructive to review in brief the progress made for the GaGdO/GaAs interface and difficulty it has overcome, to provide references in further studies on other material systems. As

pointed out by Passlack [2], one of the major difficulties in depositing oxide of any kind of III-V surface such as GaAs is to avoid the undesired native oxide formation (in this sense, Si is extremely lucky to have high quality native oxide). One of the native oxides of GaAs formed by Gallium and oxygen may take two forms,  $\text{Ga}_2\text{O}$  and  $\text{Ga}_2\text{O}_3$ . The mixture of these two oxides further results in unsaturated bonds and induce large amount of interface trap density that eventually cause frequency dispersion of the device characteristics and in the worst case, surface pinning. The breakthrough made on this material system is done via adding Gd into the oxide, which induces the Ga to take only one form of ionic state (+3) and results in a stable compound of GaGdO. The result on the oxide has been reported in [1] and regarded as a successful demonstration of oxide implementation within III-V material system. It is desirable to work out a comparable scheme for the InGaAs channel material since it offers even better material properties than GaAs. As learned from the previous example, the chosen oxide must not encourage (preferably suppress) the formation of the native oxide (as the native oxide typically implies inferior interface property). Furthermore, for the purpose of scaled device design (to enable high current drive and control short channel effects), it is also necessary to have scalable oxide down to  $\text{EOT} < 1\text{nm}$ . Given these considerations, it is encouraging that the augmented  $\text{Al}_2\text{O}_3$  deposited on InGaAs surface via Atomic Layer Deposition (ALD) may be capable of addressing these needs. We will review briefly here some properties of the  $\text{Al}_2\text{O}_3$  by ALD, since in the following part of the chapter device characteristics with this material will be analyzed in more details.

To obtain high quality oxide (and interface), it is very important to start with pristine surface. It has been reported that one of the precursors, TMA (Trimethylaluminum), cleans the InGaAs surface by removing the native oxide. To further improve the surface quality, in-situ arsenic cap was deposited on top of the InGaAs sample to be discussed in this chapter, and was subsequently desorbed thermally under high vacuum condition immediately prior to the  $\text{Al}_2\text{O}_3$

deposition. It was shown via XPS study that absorption peaks associated with  $\text{As}_2\text{O}_3$ ,  $\text{Ga}_2\text{O}_3$ ,  $\text{Ga}_2\text{O}_3$  and  $\text{In}_2\text{O}_3$  are not evident. Only with these conditions, it is possible to focus the study on the “intrinsic” properties of the dielectric film on top of the InGaAs. In this chapter, we will provide analysis on both Capacitance Voltage characteristics and the Current Voltage characteristics based on experimental samples. It may be seen, although with the cautions taken, much improvement is still needed before the dielectric quality becomes acceptable for a practical device. As most of the indications on the interface quality are inferred from the MOSCAP study, a significant part of this chapter will be focused on the analysis and interpretation of the measured CV characteristics. On the other hand, as we would also like to understand the impact of non-ideal dielectric to the IV characteristics, the last section will be dedicated to the analysis on the IV characteristics with the presence of various traps.



## 5.2 Capacitance Voltage Characteristics with Non-ideal Dielectrics

### 5.2.1 Idealized Capacitance Voltage Characteristics

In order to understand all the non-ideal effects that are added to the practical MOSCAP structure, in this subsection, we would like to establish the ideal characteristics of the III-V MOSCAP.

In the following computational example, we consider a sample that has 5.6nm physical thickness  $\text{Al}_2\text{O}_3$  on top of the n-type  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  material. The C-V characteristics may be computed by solving the Poisson's equation along gate stack direction. The calculation of the electrostatic potential is similar to the Poisson solver used in Section. 1.2 where the Newton-Ralphson method is used. Once the electrostatic potential is obtained, carrier concentration may be computed via  $n(x) = N_c F_{1/2} \left[ \frac{E_f - E_c(x)}{kT} \right]$ ,  $p(x) = N_c F_{1/2} \left[ \frac{E_v(x) - E_f}{kT} \right]$ , where  $F_{1/2}(x)$  is the Fermi integral of the order  $\frac{1}{2}$ . The total sheet charge may be subsequently assessed on a quasi-static basis. Once we have the Q-V characteristics, C-V characteristics may be computed accordingly as the derivative of the Q-V characteristics. This computation was carried out by numerically solving Poisson's equation, assuming the parabolic band structure.

There are several points that are worthwhile to point out here:

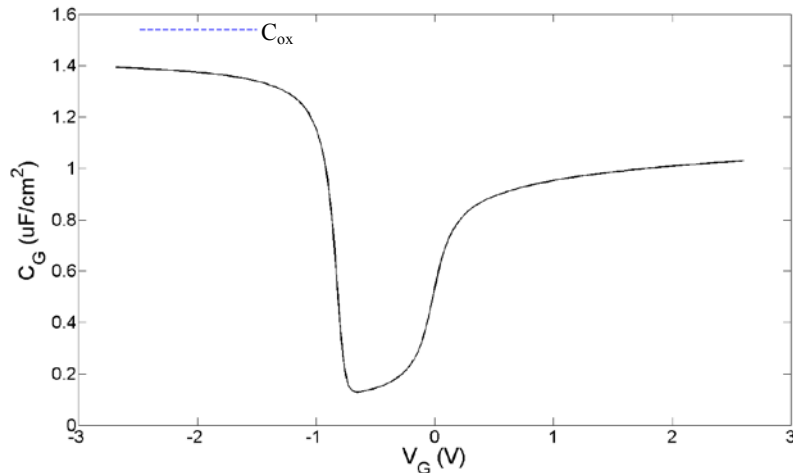


Fig. 5.1. Ideal quasi-static C-V characteristics for n-MOSCAP with 5.6nm (physical thickness)  $\text{Al}_2\text{O}_3$  on  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  doped at  $\sim 1 \times 10^{17} \text{ cm}^{-3}$ . The dashed line represents the oxide capacitance limit.

- 1) The asymmetry between the inversion side (negative voltage side) and the accumulation side (positive voltage side) is due to the difference between the hole effective mass and electron effective mass. As explained in Chapter 1, the small electron effective mass results in a much smaller  $C_{\text{DOS}}$ , which connects in series with the oxide capacitance and henceforth decreases the overall gate input capacitance.
- 2) It must be understood that this curve presented in Fig. 1 assumes quasi-static situation, i.e. adequate time is allocated to allow generation of the minority carriers between different voltages applied to the gate.

We will analyze in the following, with ideal interface/dielectric film, how the minority carrier generation depends on small signal frequency. It is helpful to define several time scales here for future convenience:

- 1) Sweeping rate: the rate at which DC bias to the gate is changed. This is typically a longer time scale that corresponds to a frequency on the order of Hz.

- 2) Small signal frequency: the frequency of the small AC voltage signal that is added to the bias on top of the DC bias. The frequencies of such signals typically range from kHz to MHz.

In the following, we would like extend the previous analysis to include the frequency dependence of the minority carrier generation on the inversion side. For simplicity, we assume the time constant associated with minority carrier generation is much smaller than the time scale of DC voltage sweep. Furthermore, we assume that the phase lag between the AC current signal and AC voltage signal is predominantly due to the minority carrier generation.

The flow of the computation may be summarized in two steps: 1) with the given DC bias voltage, a quasi-static electrostatic potential is established and the carrier distribution may be evaluated accordingly; 2) with the given AC bias, the minority carrier (hole in this case) generation is limited by the generation lifetime, the response of the small signal charge may be accounted by the admittance of the branch:

$$Y = \frac{\tau\omega^2 C}{1 + \omega^2\tau^2} + \frac{j\omega C}{1 + \omega^2\tau^2} \quad (5.1)$$

where  $\tau$  is the time constant of the minority carrier response. This is equivalent to say that the small signal charge response  $\tilde{Q}$  is scaled from the quasi-equilibrium value  $Q$  by the factor of  $1/(1 + \omega^2\tau^2)$  to account for the phase lag and magnitude change. Shown in Fig. 5.2 is a representative computational result with ideal dielectric but taking into account the frequency response of the minority carriers. The time constant assumed for minority carrier (holes) here is 1  $\mu$ s.

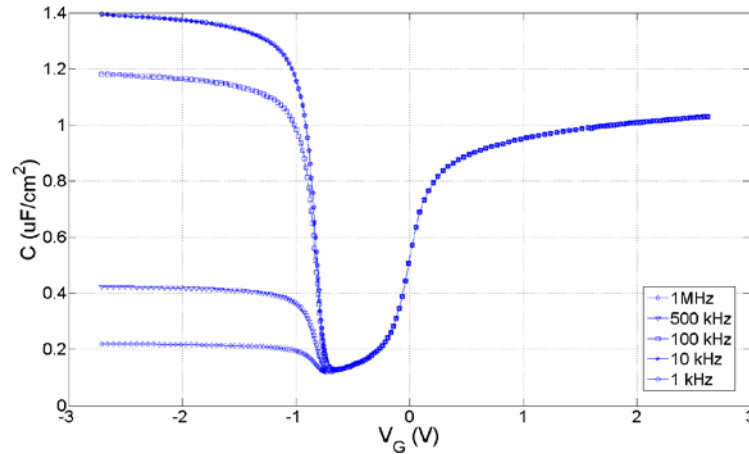


Fig. 5.2. Ideal frequency dependent C-V characteristics for n-MOSCAP with 5.6nm (physical thickness)  $\text{Al}_2\text{O}_3$  on  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  doped at  $\sim 1 \times 10^{17} \text{ cm}^{-3}$ . The minority carrier time constant assumed here is  $1 \mu\text{s}$ .

It is worthwhile to point out that there is a minute difference between the minority carrier lifetime and the time constant for minority carrier to respond to the CV measurement. The time constant that matters for the CV measurement is related to the AC current flow upon the application of the small AC signal. In other words, as the AC signal wiggles back and forth, the corresponding AC current that contributes to the capacitance must as well flow back (i.e. being depleted) and forth (i.e. being accumulated). This time constant is, however, not the same as the time constant related to the generation rate, which only corresponds to the accumulation process. In the previous calculation, we use a phenomenological time constant to represent this effect. It is noted that this time constant is much longer than that of the minority carrier lifetime (on the order of 10's nanosecond).

### 5.2.2 Impact of Interface States on Capacitance Voltage Characteristics

Interface states may arise from the unsaturated bonds at the interface between the dielectric and the semiconductor. In this section, we will discuss how the presence of interface

states affects the CV characteristics, in the context of III-V material being the substrate, and show correlations with experimental data.

As established in [3], interface states are charged/discharged accordingly to the gate bias and contribute to the overall capacitance. However, given the different lifetime which is dependent on the interface states' energy position, it may or may not respond to the higher frequency small signal. A typical signature of the interface state impacted CV characteristics is the dispersion across various frequencies. Nevertheless, care must be taken before attribute frequency dispersions to solely the presence of interface states. We will show in the following, in case of III-V substrate, only certain dispersion may be attributed to interface states while others may be due to border traps, which will be the topic of next subsection.

Representative experimental CV characteristics are shown below in Fig. 5.3, in which n-type  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  is used as the substrate and 5.6nm  $\text{Al}_2\text{O}_3$  as the gate dielectric. The sample is fabricated and measured by Eunji Kim at Stanford University. The result below has reflected the improvement on the CV characteristics by a critical Hydrogen anneal step that has removed significant amount of frequency dispersion. Details of Hydrogen passivation step that drastically improves the dielectric film quality may be found in [4]. For the as-grown sample result (inset of Fig. 5.3), evident dispersion appears at depletion/weak accumulation region ( $\sim$  between 0 V and 1V gate bias) as well as accumulation region ( $\sim$  between 2V and 3V). It is however less clear whether the dispersion on the inversion side is due to frequency response of minority carrier and/or interface traps.

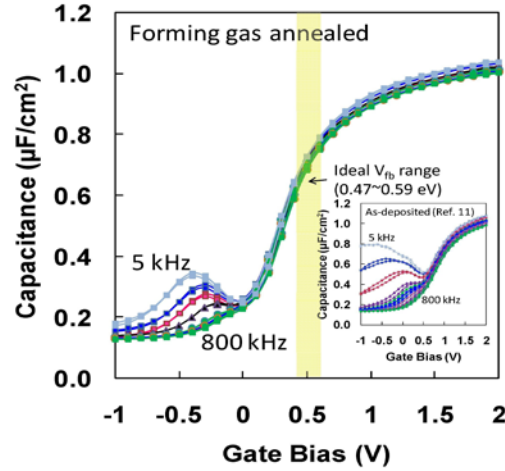


Fig. 5.3. Representative experimental C-V characteristics [3] for n-MOSCAP with 4.6nm (physical thickness)  $\text{Al}_2\text{O}_3$  on  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  doped at  $\sim 1 \times 10^{17} \text{ cm}^{-3}$ .

In the following discussions, we will focus on the dispersion in the depletion/weak accumulation and strong accumulation regions. To understand the frequency dispersion, especially in the context of the InGaAs substrate, it is necessary to understand the time constant associated with the interface traps at different energies. The dependence on energy may be described by Eq. (5.2).

$$\tau(E) = \tau_0 \cdot \exp[\Delta E / kT] \quad (5.2)$$

where  $\Delta E$  is the energy difference between the interface state and the band edge.

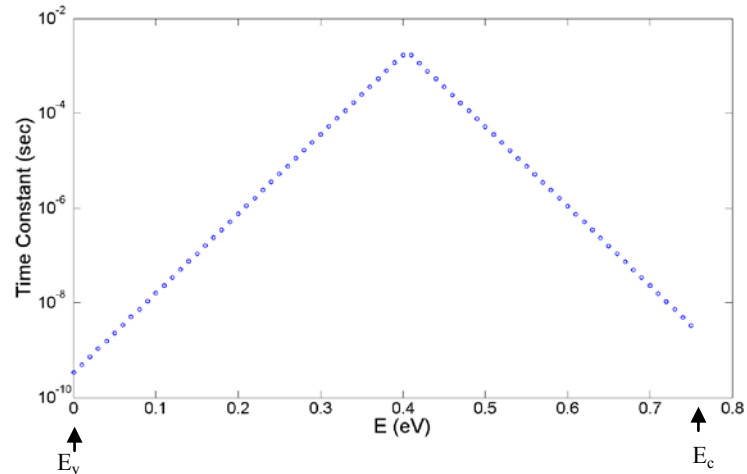


Fig. 5.4. Time constant of interface traps as a function of the energy.

As reported in [2], for narrow band material as  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$ , the factor  $\tau_0$  is on the order of  $10^{-9}$  sec (it may be computed by  $(\sigma v_i N_c)^{-1}$ , where  $\sigma$  is the capture cross section on the order of  $10^{-14} \sim 10^{-15}$   $\text{cm}^2$ ,  $v_i$  is the thermal velocity of the electrons and  $N_c$  is the effective density of states at the conduction band edge). Therefore, the time constant for the interface traps distributed across the bandgap is within the range of nano-second to mili-second. It is worth of note that for the interface states that are close to valence band, it is likely that the expression for  $\tau_0$  needs to be changed accordingly to  $(\sigma v_i N_v)^{-1}$ . Due to the asymmetry between the electron effective mass and hole effective mass in InGaAs material, the factor is slightly different, giving smaller time constant for hole trapping into the interface states. The computed result using Eq. (5.2) is shown in Fig. 5.5. It is also noticed that given the longest time constant across the bandgap falls into the mili-second range, the time interval (on the order of second) between two consecutive DC bias point is long enough to establish quasi-equilibrium between the semiconductor and interface states.

The other physical parameter that greatly influences the overall interfacial charging situation is the charge neutrality level (CNL) of the interface states. The CNL is defined as the

Fermi level position at which the net charge given by all the interface states is zero. In the following computation, we will assume that the charge neutrality level is at midgap position within the bandgap.

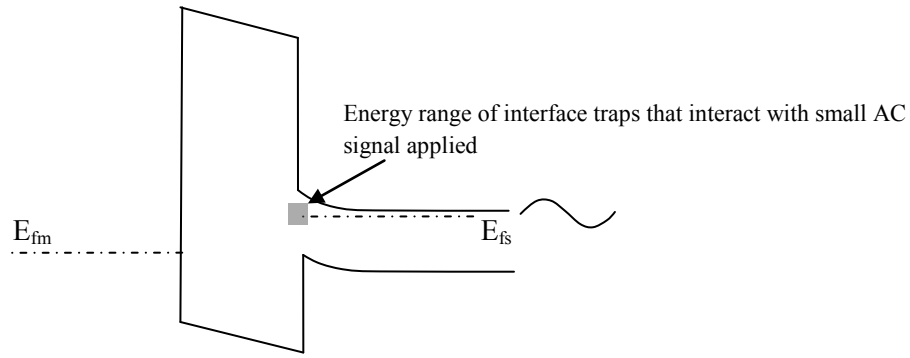


Fig. 5.5. Energy band diagram for illustration of the capacitive response of interface states.

Fig. 5.5 illustrates how the interface traps get charged up as the Fermi level of semiconductor sweeps in accordance to the sweeping of gate voltage. During a CV sweep, DC bias is changed on a slow basis, and establishes a Fermi level position throughout the semiconductor and the interface states. With the addition of small signal, the Fermi level oscillates on a faster basis on the energy scale within the vicinity of the Fermi level position established by the DC bias. Depending on the (relative) position of Fermi level established by the DC bias, the Fermi level will “see” interface states with different time constants at different stages during the DC sweep, as shown in Fig. 5.5. We may again use the similar argument as in section 5.2.1 for the reduction of the capacitance response due to the small signal frequency, i.e., the interface state charge will respond to the oscillation with reduced magnitude by a factor of  $1/1 + \omega^2 \tau^2$  (same as that was described in Eq. 5.1), where  $\tau$  depends on DC bias established Fermi level.



In the following, we will describe the computation procedure for the CV characteristics with the consideration of the interface states.

- 1) Compute the DC condition. The only complication is to include the interface state charge. Denoting  $E_{fs}$  as the Fermi level in the semiconductor,  $E_{CNL}$  as the charge neutrality level for the interface states,  $D_{it}$  as the interface state density in unit of  $\#/eV\text{-cm}^2$ , we can write down the interface charge as the following integral:

$$Q_{it} = q \int_{E_v}^{E_c} D_{it}(E) \cdot [f(E, E_{fs}) - f(E, E_{CNL})] dE \quad (5.3)$$

where  $f(E, E_f)$  is the Fermi distribution function. When  $E_f$  is above  $E_{CNL}$ , net negative charge results for the interface states, which leads to positive charge of the same magnitude on the gate. If we assume the distribution of the interface states is uniform on the energy scale, Eq. (5.3) may be further reduced to:

$$Q_{it} = q D_{it} kT \left\{ \ln \frac{[1 + \exp(\frac{E_{fs} - E_v}{kT})] \cdot [1 + \exp(\frac{E_{CNL} - E_c}{kT})]}{[1 + \exp(\frac{E_{fs} - E_c}{kT})] \cdot [1 + \exp(\frac{E_{CNL} - E_v}{kT})]} \right\} \quad (5.4)$$

- 2) Given the established DC condition, we treat the AC signal as a perturbation. We may first evaluate the gate charge response assuming the AC signal has zero frequency. The difference between the total charge with and without perturbation stems from change in the majority carrier charge  $\Delta Q_{MA}$ , minority carrier charge  $\Delta Q_{MI}$ , depletion charge  $\Delta Q_D$  and interface state charge  $\Delta Q_{IT}$ , among which  $\Delta Q_{MI}$  and  $\Delta Q_{IT}$  are frequency dependent and is scaled by  $1/1 + \omega^2 \tau^2$  at finite frequency.

With these considerations, a frequency dependent CV curve may be computed including the response from interface states. A representative computation result is given in Fig. 5.6, with the assumption of the interface states are uniformly distributed across the bandgap with the value of  $5 \times 10^{12} \text{ cm}^{-2}\text{-eV}^{-1}$ . It is noted that only classical solution is pursued here, i.e. no quantum mechanical effects have been taken into account at this point. It is also noted here that the above methodology may be used to calculate CV characteristics for any arbitrary distribution of interface states.

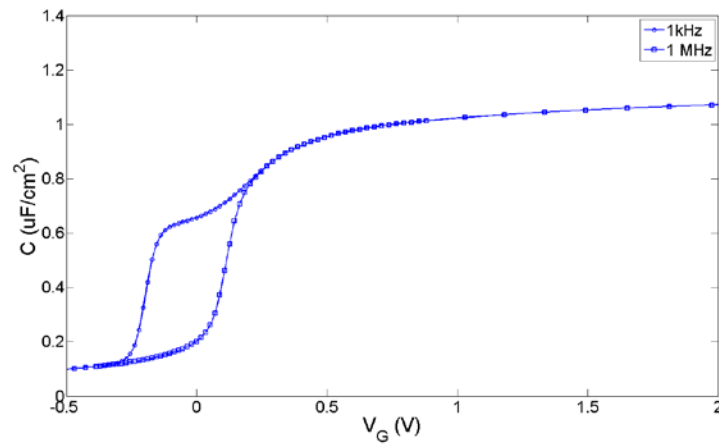


Fig. 5.6. Representative simulated C-V characteristics for n-MOSCAP with 5.6nm (physical thickness)  $\text{Al}_2\text{O}_3$  on  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  doped at  $\sim 1 \times 10^{17} \text{ cm}^{-3}$ , uniform  $5 \times 10^{12} \text{ cm}^{-2}\text{-eV}^{-1}$  Dit distribution.

There are several features here that are worth of further discussion:

- 1) Frequency dispersion due to the presence of  $D_{it}$  is only evident in the weak inversion, depletion and weak accumulation regime. The reason lies in that only interface traps with long enough time constant may show frequency dependent response (in that traps with shorter time constant are always capable to respond to small signals). As given in Eq. 5.2, traps with longer time constant are located around the mid gap. Therefore, only when quick Fermi level oscillation (oscillation produced by AC

signal around the quiescent point set by DC sweep) sweeps within these energy ranges, the charge response from the interface states will show frequency dependence. On the other hand, interface states that are located close to the band edge will not contribute to the frequency dispersion since the time constants involved are too small to be sensed by 1MHz. This is also why no dispersion due to the interface traps is shown in the strong accumulation region (when the DC established Fermi level swings into the conduction band).

- 2) For the low frequency (1kHz) CV curve, we observe shoulders around -0.25V. The presence of shoulder is originated from the capacitance response from the interface states. In this bias range (where semiconductor is in depletion and only contributes a small depletion capacitance), the additional charge response is provided by the interface state charge. The early “turn-on” observed in 1kHz CV curve around -0.25V is a result of this kind. When the DC bias further increases, capacitance response from the semiconductor starts to take over and results a second “turn-on” for the CV result. Overall, this behavior produces the shoulder shape in the CV characteristics seen. As the time constant is a strong function interface state energies, at different sweeping frequencies, the small signal essentially picks up response from interface states of different energy ranges. This effect is schematically shown in Fig. 5.7, where we mark the responsive energy range under different frequency utilizing the previous time constant result shown in Fig. 5.5. As seen, the lower the sweeping frequency, the wider energy range from which the interface states will respond and contribute to the capacitive response. Therefore, the shoulder position/height may vary with different band edge time constants. The height of the shoulder is influenced by the  $D_{it}$  level and distribution.

- 3) It is to be noted that computation for minority carriers (hole for this case) is more complicated. In addition to the energy dependent time constant as shown in Fig. 5.7, one must also take into account the time constant associated with minority carrier generation time. This effect is not included in the computation presented above.

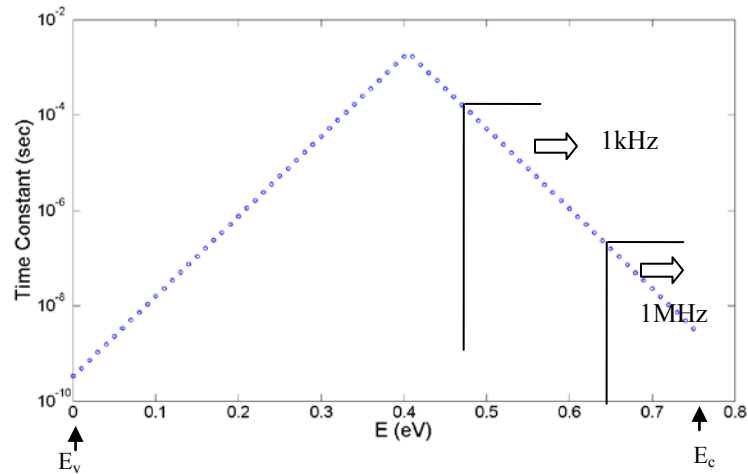


Fig. 5.7. Responsive energy range upon different small signal frequency.

Comparing Fig. 5.6 to the experimental results, we notice that the frequency dispersion on the strong accumulation end may not be explained by the presence of interface states itself. Additional experimental evidence suggests that the accumulation end dispersion is almost temperature independent, which clearly sets it apart from primarily relating to the interface states. To explain the dispersion on the accumulation side, we introduce another type of trap – border trap, which is the primary topic of the following subsection.

### ***5.2.3 Impact of Border Traps on Capacitance Voltage Characteristics***

In this section, we will introduce another type of trap that also contributes to the dispersion but less noticed in previous literature. We had seen in the previous subsection that the frequency dispersion caused by interface traps itself may not be adequate to explain what has been experimentally observed, i.e. the dispersion in the strong accumulation region, unless we assert the existence of slow interfacial traps at the band edges or inside the conduction/valence band. However the conventional interpretation disagrees with the following considerations: 1) the mechanisms of having very long trapping constants at band edges that are comparable to those at the midgap are not explained in the conventional interpretation, given the carrier concentrations that would have reached when the Fermi level approaches the band edges; 2) the trapping/detrapping at the interface in the conventional picture is known to have strong temperature dependence, which originates from the temperature dependence of its time constant, which was however found to be opposite to the experimental results. In light of these observations, we consider another trapping/detrapping mechanism that involves the possible defects/traps located within the deposited high-k dielectrics – referred to as border traps in the following text.

This subsection will be arranged in the following sequence: the physical picture will be introduced first to convey the concept of the border traps, followed by a model that depicts the behavior of border traps. A novel method of estimating the border trap level will be then presented followed by further discussion on the limitation of current border trap model. A summary will be provided at the end.

The border traps are traps that are distributed across the entire high-k dielectric layer in both real space and in energy, as schematically shown in Fig. 5.8. The carrier exchange between the semiconductor and the high-k dielectric is through a tunneling process (which is little

dependent on temperature – the fine point of temperature dependence will be discussed later in this subsection). Since these traps are distributed across the bulk of the dielectric, they are sometimes notated as bulk traps in literature [4]. However, as the tunneling probability sharply decreased with the penetration depth, the most effective traps to exchange carriers with the semiconductor are the traps located close to the semiconductor/dielectric interface (border), they are thus referred here as border traps.

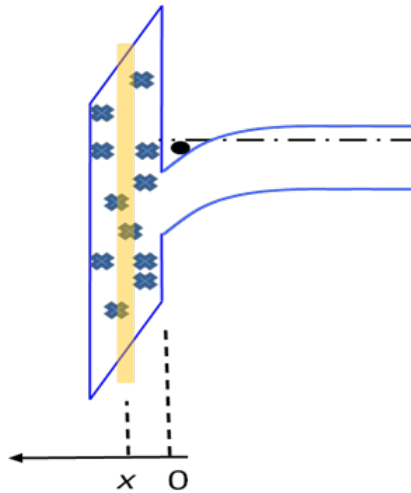


Fig. 5.8. Schematic drawing of border traps and coordinates setup for further modeling. The origin is set at the interface between the semiconductor and the dielectric.

Before we start to model the impact on the CV characteristic due to border traps, it is worthwhile to understand the essential features of the dispersions at the accumulation side. In addition to the experimental results shown in Fig. 5.3, if we plot the maximum capacitance value obtained at the highest gate bias applied during the measurement, as a function of  $\log(1/f)$ , we further observe an almost linear relationship within the frequency range of 100Hz~800kHz between the two quantities, as shown in Fig. 5.9. The plot has been made available by Dr. Eunji Kim of Stanford, and may be found in [3].

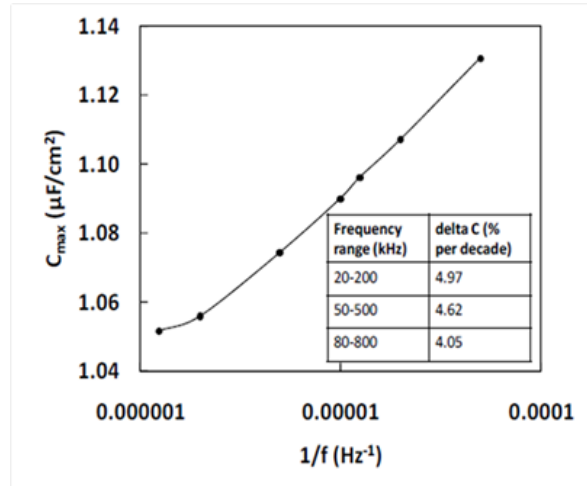


Fig. 5.9. Linear relationship between the maximum capacitance measured on the accumulation side and  $\log(1/f)$ . Plot is taken from a collaborative work documented in [3].

Additional experiments have also been conducted to explore the temperature dependence of such dispersion. The dispersion amount is listed in Table 5.1, as a recast of the data originally presented in [3]. As indicated by the data below, little temperature dependence may be observed on the frequency dispersion.

Table 5.1. Temperature dependence of the accumulation capacitance dispersion

Frequency Range	Frequency Dispersion (% per decade)		
	-40 °C	Room Temperature	110 °C
20Hz- 200kHz	5.65	5.98	5.97
50Hz- 500kHz	5.50	5.90	5.62
80Hz- 800kHz	5.46	5.64	5.05

In the model presented below, the proposed physical mechanism must be capable of explaining the two features observed above: 1) linear dependence for frequency dispersion on  $\log(1/f)$ , 2) little temperature dependence.

To understand the frequency response of trapping/detrapping associated with the border traps, it is appropriate to clarify the time constant associated with the process. We will start this analysis by describing the tunneling process. When the Fermi levels differ between the semiconductor and the border traps with the application of small AC signal, electron exchange happens via tunneling process, involving charging and discharging the states in the border traps. The time constant for traps that are distributed between  $E - \delta E/2$  and  $E + \delta E/2$  and physically located at  $x$  may be computed by using an RC equivalent circuit, as shown below:

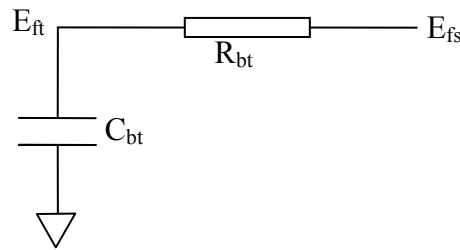


Fig. 5.10. Equivalent circuit for evaluation of the border trap time constant.

We will compute the two individual components to obtain the RC time constant. It is worthwhile to point out that the current derivation is only concerned with capacitance and resistance contribution from the border traps that are distributed within the  $\delta E$  range around energy  $E$ , while the total contribution throughout the energy range will be carried out via integration later on.

The border trap capacitance is defined as:



$$C_{bt} = -\frac{qdn_t}{dE_{ft}/q} = -q^2 N_t \delta E \frac{df_t}{dE_{ft}} \quad (5.5)$$

where  $qdn_t$  is the trapped charge density with the  $\delta E$  range around energy  $E_t$  at position  $x$ , and is expressed by the product of available states ( $N_t \delta E$ ) and Fermi distribution function, where  $N_t$  is the number of bulk trap per unit volume per unit energy.

$$C_{bt} = \frac{q^2 N_t \delta E}{kT} \bullet f_t \bullet (1 - f_t) \quad (5.6)$$

It is to be noted that if Eq. (5.6) is integrated over the entire energy range, it will give the DOS capacitance of border traps, i.e.  $q^2 N_t$ .

The computation of  $R_{bt}$  may be implemented by relating the tunneling current to the difference in Fermi level (voltage).

The tunneling current received by border traps within  $\delta E$  range may be computed by the current difference between forward going (semiconductor  $\rightarrow$  border traps) and backward going (border traps  $\rightarrow$  semiconductor).

The forward going current is given by

$$J_f = q \bullet (N_t \delta E) \bullet (1 - f_t) \bullet \sigma \bullet [g_s(E_s) \bullet \delta E_s] \bullet f_s(E_s) \bullet v_x(E_s) \bullet e^{-\gamma x}$$

$$\gamma = \frac{2}{\hbar} \sqrt{2m^* (\Delta E_C - E)} \quad (5.7)$$

where  $\delta E_s$  is the energy interval for the electrons from semiconductor. Similarly the backward going current is given by

$$J_b = q \cdot (N_t \delta E) \cdot f_t \cdot \sigma \cdot [g_s(E_s) \cdot \delta E_s] \cdot [1 - f_s(E_s)] \cdot v_x(E_s) \cdot e^{-\gamma x} \quad (5.8)$$

The net current is then given by:

$$J_{net} = q \cdot (N_t \delta E) \cdot \sigma \cdot [g_s(E_s) \cdot \delta E_s] \cdot [f_s - f_t] \cdot v_x(E_s) \cdot e^{-\gamma x} \quad (5.9)$$

In the above equation,  $f_s$  is the Fermi distribution function on the semiconductor side,

$$f_s = [1 + \exp(\frac{E_s - E_{fs}}{kT})]^{-1} ; \text{ while } f_t \text{ is the Fermi distribution function of border traps,}$$

$$f_t = [1 + \exp(\frac{E_t - E_{ft}}{kT})]^{-1}. \text{ As the most effective traps are located close to the border, the border}$$

trap Fermi level is almost the same as the semiconductor Fermi level ( $E_{fs} \approx E_{ft}$ ). Furthermore,

as the energy selection posed by tunneling process (energy conservation),  $E_s$  is equal to  $E_t$ .

Therefore, the term  $f_s f_t$  may be further written as:

$$f_s - f_t = \frac{E_{fs} - E_{ft}}{kT} \cdot f_s \cdot (1 - f_s) \quad (5.10)$$

The current received by the border traps that lie in the energy range of  $\delta E$  may be obtained by integrating through all possible  $E_s$ . The integration may be written as:

$$J_{net} = \frac{q^2 \sigma N_t}{kT} \frac{(E_{fs} - E_{ft})}{q} \cdot \int_0^{+\infty} g_s(E_s) \cdot (1 - f_s) \cdot f_s \cdot v_x(E_s) \cdot e^{-\gamma x} \cdot \delta E \cdot dE_s \quad (5.11)$$

It is worthwhile to consider different situations. In the following, we will discuss two scenarios where non-degenerate or degenerate occupation is assumed for the semiconductor.

For non-degenerate case, the Fermi level is considered to be well below the conduction band edge. Therefore, the  $1 - f_s$  factor in Eq. (5.11) may be approximated to be 1. The integration in Eq. (5.11) may therefore be written as:

$$J_{net} = \frac{q^2 \sigma N_t (E_{fs} - E_{ft})}{kT} \cdot n_s \cdot e^{-\gamma x} \cdot \delta E \cdot \bar{v}_x(E_s) \quad (5.12)$$

$$\text{where } n_s = \int_0^{+\infty} g_s(E) \cdot f_s(E) \cdot dE$$

In deriving Eq. (5.12), we have used an average carrier velocity and taken it out from the integrand. We also assume that the  $\gamma$  factor is a slow varying function of energy, and is taken out from the integrand as well. With these mathematical manipulations, we may further evaluate the resistance of the branch as shown in Fig. 5.10.

$$R_{bt} = \frac{E_{fs} - E_{ft}}{qJ_{net}} = \left[ \frac{q^2 \sigma N_t}{kT} \cdot n_s \cdot e^{-\gamma x} \cdot \delta E \cdot \bar{v}_x(E_s) \right]^{-1} \quad (5.13)$$

The RC time constant may therefore be written as:

$$\tau_{bt} = \frac{1}{n_s \sigma \bar{v}_x} e^{\gamma x} = \tau_0 e^{\gamma x} \quad (5.14)$$

It is to be noted that this quantity  $\tau_{bt}$ , with the assumptions presented previously, is the same for all energies, once the Fermi level position is established. The result shown in Eq. (5.14) is also in accordance to the conclusion on the time constant drawn in [Heiman and Warfield]. The overall capacitance contributed by border traps may be assessed by integrating the imaginary part of the admittance over the entire energy range. Similar to the expression in previous sections, the admittance associated with the border traps may be written as:

$$Y = \frac{j\omega C_{bt}}{1 + \omega^2 \tau_{bt}^2} + \frac{\omega^2 \tau_{bt} C_{bt}}{1 + \omega^2 \tau_{bt}^2} \quad (5.15)$$

The overall contribution from the border trap is:

$$C_{bt\_overall} = \int_0^{t_{ox}} dx \int_0^{+\infty} \frac{1}{1 + \omega^2 \tau_{bt}^2} \frac{q^2 N_t}{kT} f_t (1 - f_t) dE \quad (5.16)$$

In the above derivation, we used the result of Eq. (5.6). Given that  $\tau_{bt}$  is not energy dependent, Eq. (5.16) becomes:

$$\begin{aligned} C_{bt\_overall} &= \frac{q^2 N_t}{kT} \int_0^{t_{ox}} \frac{1}{1 + \omega^2 \tau_0^2 e^{-2\gamma x}} dx \int_0^{+\infty} f_t (1 - f_t) dE \\ &= \frac{q^2 N_t}{2\gamma} [\ln(1 + \omega^2 \tau_0^2) - \ln(e^{-2\gamma t_{ox}} + \omega^2 \tau_0^2)] \end{aligned} \quad (5.17)$$

To arrive at Eq. (5.17), we have assumed that the quantity  $\gamma$  has little energy dependence, which is appropriate when  $\Delta E_c$  is large.

For the degenerate case, where  $E_f$  is inside the conduction band, we may no longer approximate the quantity  $1 - f_s$  to be 1 for all energies in the conduction band. We notice that in Eq. (5.11),  $f_s \cdot (1 - f_s)$  is a strongly peaked function around  $E_f$ , whose integration over the entire energy range is  $kT$ . For mathematical simplicity, we may approximate this function as a square function with width of  $kT$  around  $E_f$  and value of 1 within the  $kT$  width. (This is equivalent to approximate this function as square function with width of  $\Delta^* kT$  around  $E_f$  and value of  $1/\Delta^*$  within the  $\Delta^* kT$  width).

With the above considerations, using Eq. (5.11), the resistance associated with the tunneling may be computed as:

$$R_{bt} = \frac{E_{fs} - E_{ft}}{qJ_{net}} = \begin{cases} \infty & (E - \frac{\delta E}{2}, E + \frac{\delta E}{2}) \not\subset (E_f - \frac{kT}{2}, E_f + \frac{kT}{2}) \\ \frac{1}{q^2 \sigma N_t} \frac{e^{\gamma x}}{g(E)v(E)\delta E} & (E - \frac{\delta E}{2}, E + \frac{\delta E}{2}) \subset (E_f - \frac{kT}{2}, E_f + \frac{kT}{2}) \end{cases} \quad (5.18)$$

Similarly Eq. (5.6) may be written in this fashion as well

$$C_{bt} = \begin{cases} 0 & (E - \frac{\delta E}{2}, E + \frac{\delta E}{2}) \not\subset (E_f - \frac{kT}{2}, E_f + \frac{kT}{2}) \\ \frac{q^2 N_t \delta E}{kT} & (E - \frac{\delta E}{2}, E + \frac{\delta E}{2}) \subset (E_f - \frac{kT}{2}, E_f + \frac{kT}{2}) \end{cases} \quad (5.19)$$

The time constant is therefore

$$\tau_{bt}(x, E) = \begin{cases} \infty & (E - \frac{\delta E}{2}, E + \frac{\delta E}{2}) \not\subset (E_f - \frac{kT}{2}, E_f + \frac{kT}{2}) \\ \frac{e^{\gamma x}}{\sigma g(E)v(E)kT} & (E - \frac{\delta E}{2}, E + \frac{\delta E}{2}) \subset (E_f - \frac{kT}{2}, E_f + \frac{kT}{2}) \end{cases} \quad (5.20)$$

When  $\delta E$  does not belong to the  $kT$  width, there is essentially no charge exchange, therefore, from a physical perspective, the time constant is infinite.

We would like to point out here that for the degenerate case, the quantity  $\tau_{bt}$  is both a function of energy and position; in comparison to the energy independent  $\tau_{bt}$  for the non-degenerate case. This point is further illustrated in Fig. 5.11, where the band diagram is shown for non-degenerate and degenerate cases. As seen, for the non-degenerate case (Fermi level within the band gap), as the carrier exchange must happen within the conduction band, only the tail of

$f_s \bullet (1 - f_s)$  contributes to the current (which further determines the time constant). Whereas, for the degenerate case, the primary contributor to the current is located around  $E_f$ , due to the sharply peaked shape of  $f_s \bullet (1 - f_s)$ . In other words, electrons with energies that are beneath the Fermi level by more than a few  $kT$  may not contribute to the overall current, since no empty states are available within the oxide. The same argument holds for electrons with energies that are above the Fermi level by a few  $kT$ . This situation compares drastically with the non-degenerate case, where electrons within the conduction band always “see” empty states within the oxide.

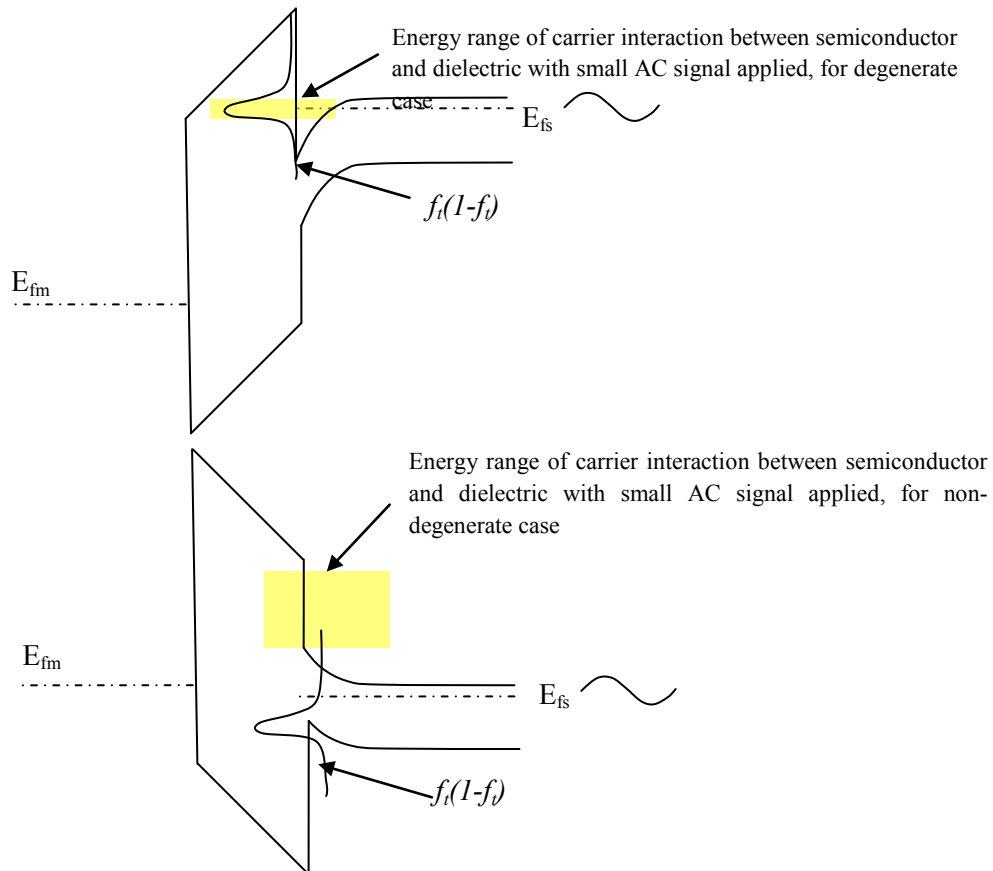


Fig. 5.11. Energy band diagram for illustration of the capacitive response of border traps under degenerate and non-degenerate cases.

It is therefore worthwhile to compare the time constant we derived so far for the two cases:

$$\tau_{bt}(x, E) = \begin{cases} \frac{e^{\gamma x}}{n_s \overline{\sigma v}_x} & \text{non-deg.} \\ \frac{e^{\gamma x}}{\sigma g(E_f) v(E_f) kT} & \text{deg.} \end{cases} \quad (5.21)$$

We may again evaluate, for the degenerate case, the overall contribution to the capacitance due to the border traps, by integrating the imaginary part of the admittance. For mathematical simplicity, we define the following quantity:

$$\tau_0 = \frac{1}{\sigma g(E_f) v(E_f) kT} \quad (5.22)$$

$$\begin{aligned} C_{bt\_all} &= \int_0^{t_{ox}} dx \int_0^{\Delta E_c} \frac{1}{1 + \omega^2 \tau^2} \frac{q^2 N_t}{kT} f_t(1 - f_t) = \int_0^{t_{ox}} dx \frac{(q^2 / kT) \bullet N_t \bullet kT}{1 + \omega^2 \tau_0^2 e^{2\gamma x}} \\ &= \frac{q^2 N_t}{2\gamma} [\ln(1 + \omega^2 \tau_0^2) - \ln(e^{-2\gamma t_{ox}} + \omega^2 \tau_0^2)] \end{aligned} \quad (5.23)$$

It is to be realized that the quantity  $\tau_0$  also assumes different forms for degenerate and non-degenerate cases. It is interesting that Eq. (5.23) is identical to Eq. (5.17) in the mathematical form, despite the  $\tau_0$  has different values. In cases when the dielectric thickness is not too thin, (it may be determined numerically, to be  $>3\text{nm}$ , assuming 2eV band offset for  $\text{Al}_2\text{O}_3$  and  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  substrate), the contribution from the term  $e^{-2\gamma t_{ox}}$  may be neglected. Eq. (5.17) and Eq. (5.23) may be further reduced to:

$$C_{bt\_ox} = \frac{q^2 N_{bt}}{2\gamma} \ln\left(1 + \frac{1}{\omega^2 \tau_0^2}\right) \quad (5.24)$$

At low frequencies (i.e.  $\frac{1}{\omega^2 \tau_0^2} \gg 1$ ), Eq. (5.12) becomes  $\frac{q^2 N_{bt}}{\gamma} \ln\left(\frac{1}{\omega \tau_0}\right)$ , proportional to  $\ln\left(\frac{1}{\omega}\right)$ ,

which agrees with the first feature that we extracted from the experimental result.

From Eq. (5.24), the volume density of the border traps may also be estimated, with the information on the capacitance dispersion at different frequencies. Again, under low frequency limit:

$$C_{bt\_all} \approx \frac{q^2 N_{bt} \hbar}{\sqrt{2m^* \Delta E_C}} \ln\left(\frac{1}{\omega}\right) + const \quad (5.25)$$

The constant term in Eq. (5.25) is the integration result of  $\ln(1/\tau_0)$  term. It is observed from Eq. (5.25) that the slope with regard to  $\ln(1/f)$  is directly related to the volume border trap density, hence the density may be extracted from the dispersion v.s.  $\ln(1/f)$  slope.

To implement such extraction, we make use of the equivalent circuits shown in Fig. 5.12, where  $C_{acc}$  denotes the accumulation capacitance from the semiconductor (which includes the contributions from quantum well capacitance as well as the DOS capacitance);  $C_{ox}$  is the oxide capacitance. Given the dielectric information of the sample (4.6nm physical thickness  $\text{Al}_2\text{O}_3$  with relative dielectric constant  $\sim 8$ ), the capacitive response from semiconductor and border trap may be computed accordingly. The frequency dispersion is considered to be caused by the border traps. From Fig. 5.7, the slopes of the capacitance versus  $\ln(1/f)$  may be read as  $s=3.3 \times 10^{-4} \text{ F/m}^2$ . However, it must be realized that this slope is a diluted result due to the presence of  $C_{ox}$  and  $C_{acc}$  in the equivalent circuit shown in Fig. 5.12.

Denoting the slope of  $C_{bt}$  v.s.  $\ln(1/f)$  as  $S$  (which is equal to  $\frac{q^2 N_{bt} \hbar}{\sqrt{2m^* \Delta E_C}}$ ), the overall

capacitance may be written as:



$$C_{overall} = \frac{[S \cdot \ln(1/f) + C_{acc}] + C_{ox}}{C_{ox} + C_{acc} + S \cdot \ln(1/f)} \approx \frac{C_{ox}}{C_{ox} + C_{acc}} \cdot S \cdot \ln(1/f) + \frac{C_{ox} \cdot C_{acc}}{C_{ox} + C_{acc}} \quad (5.26)$$

In Eq. (5.26), we have assumed that the capacitance contribution from the border traps is a small perturbation for the overall capacitance. It is seen that the  $\ln(1/f)$  dependence of the  $C_{bt}$  is diluted

by a factor of  $\frac{C_{ox}}{C_{ox} + C_{acc}}$ , which is estimated to be  $\frac{1}{4}$  for this case. By equating the two

quantities:  $s = \frac{q^2 N_{bt} \hbar}{\sqrt{2m^* \Delta E_c}} \cdot \frac{C_{ox}}{C_{ox} + C_{acc}}$ , the density of the border trap is estimated to be

$7.8 \times 10^{19} \text{ cm}^{-3} \text{ eV}^{-1}$ . The effective mass used here for the oxide is  $0.5m_0$ , and the  $\Delta E_c$  here is taken to be  $1.7 \text{ eV}$ .

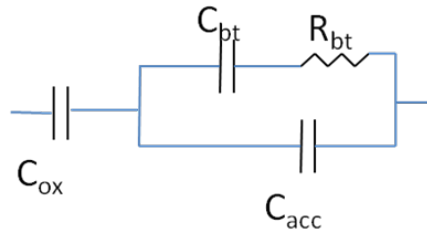


Fig. 5.12. Equivalent circuit of the gate input capacitance including border traps contribution.

From the equivalent circuits shown in Fig. 5.12, we would like to note again the difference that the substrate material will make to the measurement. If border traps with similar densities exist in the dielectric of same thickness, however on substrates of different materials, frequency dispersion will be more visible for the sample that contains smaller effective mass substrate. The reason lies in a much smaller DOS capacitance that limits the overall accumulation capacitance, which makes the frequency dispersion more observable in the experiment. On the other hand, the dispersion becomes less evident for the substrate with larger effective mass and thus large DOS capacitance, which in series combination with the oxide capacitance will mask

the frequency dispersion. (In this regard, the frequency dispersion on the accumulation end is less visible for high-k dielectrics on Si).

The temperature dependence (independence) of the frequency dispersion is another aspect that makes border trap response distinct. The (almost) independence observed from experiments may as well be explained by the previous model. The data compiled in Table 5.1 essentially suggests the slope of dispersion as a function of  $\ln(1/f)$  does not change with temperature. This is confirmed by the slope as shown in Eq. (5.25), which is temperature independent.

We may also utilize the derivation above towards the conductance obtained from CV measurement. In the non-degenerate case, the conductance is given by:

$$\begin{aligned}
 G_{overall} &= \frac{q^2 \omega^2 N_t}{kT} \int_0^{t_{ox}} \frac{\tau_0 e^{x\gamma}}{1 + \omega^2 \tau_0^2 e^{2x\gamma}} dx \int_0^{\Delta E_c} f_t (1 - f_t) dE \\
 &= \frac{q^2 N_t \omega}{\gamma} [\arctan(\omega \tau_0 e^{\gamma t_{ox}}) - \arctan(\omega \tau_0)] \\
 &\approx \frac{q^2 N_t \omega}{\gamma} \bullet \frac{\pi}{2}
 \end{aligned} \tag{5.27}$$

In the above derivation, the quantity  $\tau_0$  is defined in Eq. (5.14).

For the degenerate case, the derivation assumes the same form as presented in Eq. (5.27), except that the quantity  $\tau_0$  is defined in Eq. (5.22). From Eq. (5.27), it is evident that the conductance has a linear dependence on frequency. This is also verified by experiment results (a different set of samples, credit: Dr. Shin, Stanford, Yu Yuan, UCSD), as shown in Fig. 5.13. We may again extract the value of the border trap density from the slope of the graph. Using the same set of material parameters as before, we obtain  $8.3 \times 10^{19} \text{ cm}^{-3} \text{ eV}^{-1}$  border trap density.

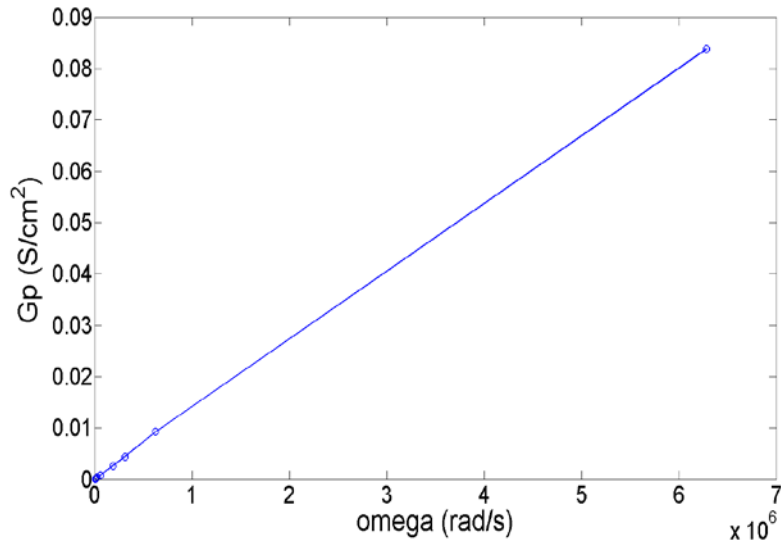


Fig. 5.13. Experimental data on conductance as a function of frequency, showing the linear relationship as predicted by Eq. (5.27).

In the following, we outline the computational approach for the CV characteristics including border trap response. We will make use of the physical understanding that we developed previously on the time constants to guide our computation. The overall calculation of CV characteristics may be divided into two segments: “DC” computation and “AC” computation.

“DC” computation starts with establishing the initial bias condition. It is assumed that the initial bias is applied long enough so that a true “DC” condition may be reached. Followed by this step, we would like to examine if a true “DC” status has been reached within the time interval between applications of two consecutive DC biases. To answer this question, we will use the previous knowledge developed for the border trap time constant. In the following computational result showing the position dependent time constant, instead of evaluating the two extreme cases (i.e., non-degenerate and degenerate cases), we calculate numerically the integration given in Eq. (5.11) to evaluate the overall current. Corresponding resistance is then computed that enables the

further evaluation for the time constant at each energy point. We define the position dependent time constant as follows:

$$\tau_{bt}(x) = \int_0^{\Delta E_c} \frac{qJ_{net}}{E_{fs} - E_{ft}} \frac{q^2 N_t}{kT} \cdot f_t \cdot (1 - f_t) dE \quad (5.28)$$

The corresponding result of Eq. (5.28) is shown in Fig. 5.15. For this computation (and later computations), we use the following parameters:

Oxide thickness:  $t_{ox}=4.6nm$

Relative dielectric constant of oxide:  $\epsilon_r=8$

Border trap volume density:  $N_t=8 \times 10^{19} \text{ cm}^{-3} \text{ eV}^{-1}$

Capture cross section:  $\sigma=10^{-19} \text{ cm}^2$

The  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  material parameters may be found in Table 2.2.

It is seen from the figure, with different semiconductor Fermi level, the resultant time constants are very different. For low carrier concentration case (Fermi level low), the time constant is much longer. For traps that are located far from the semiconductor/dielectric interface, the time constant is even longer than that of the DC bias dwell time. In this situation, one may not assume that the application of the DC bias has established a quasi-equilibrium Fermi level throughout the semiconductor/dielectric stack.

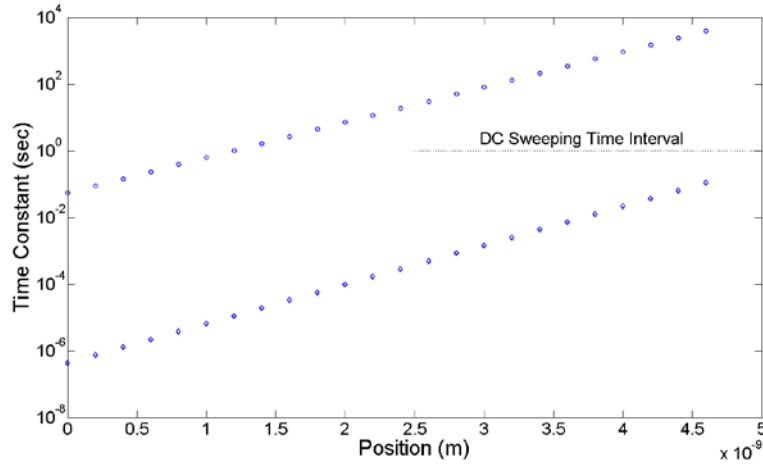


Fig. 5.14. Representative time constant associated with border traps. Open circle symbol: low carrier concentration case ( $10^{15} \text{ cm}^{-3}$ ); Open diamond symbol: high carrier concentration case ( $3 \times 10^{19} \text{ cm}^{-3}$ ). Calculation performed in accordance to definition in Eq. (5.28).

On the other hand, for high carrier concentration case, the time constant of the entire stack becomes much smaller, and DC bias may establish a quasi-equilibrium Fermi level throughout the film (but the time constants are still large enough to create frequency dispersion for small signals). In light of these thoughts, we formulate our “DC” computation with the following steps:

1. We start our calculation by specifying the surface potential. The Poisson’s equation may be readily solved to obtain the semiconductor charge distribution. For a thin slice of dielectric located at  $x$ , upon a change of semiconductor Fermi level position, the incremental charge that gets into border traps is given by:

$$\Delta Q_{BT\_DC} dx = C_{bt} dx \bullet \Delta E_{fs} \{1 - \exp[-T / \tau_{bt}(x)]\} \quad (5.29)$$

where  $T$  is the time interval between two consecutive “DC” bias points, and  $\tau_{bt}(x)$  is the position dependent time constant of border traps. This procedure is essentially the charging process of a capacitor, which is schematically shown in Fig. 5.15. It

must be noticed that each time when the DC bias steps up, one must reevaluate the time constant for a given position, in that the occupations of the traps have changed, which further affects the current flow. Therefore, the filling situation of the border traps is not only a function of the current DC bias, but also the DC bias history. This effect is more evident for the depletion/weak accumulation region, where time constant is comparable to the DC sweeping time. Whereas, when the Fermi level swings into strong accumulation, due to the much smaller time constant (compare to DC sweeping time), the Fermi level for the border traps ( $E_{ft}$ ) catches up the  $E_{fs}$ , during the time scale of DC sweeping time. We should also notice that the time constant is strongly affected by the physical location of the trap (due to the tunneling process involved).

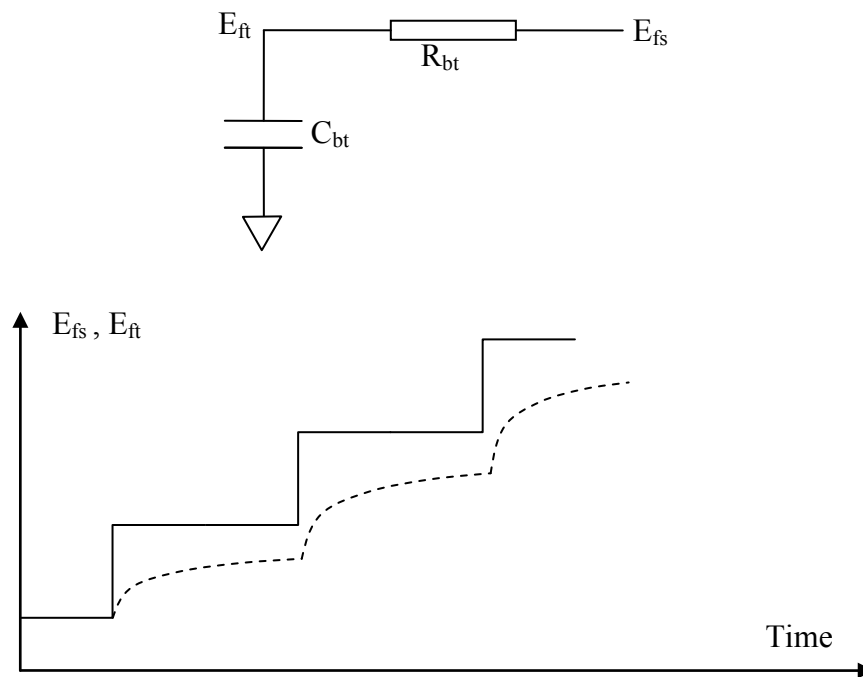


Fig. 5.15. Schematic drawing of the time domain response for the charging characteristics of the border traps. Solid line represents  $E_{fs}$ , dashed line represents  $E_{ft}$ .

Fig. 5.16 is a computed result showing how the Fermi level of the traps varies as a function of position at a DC bias in the weak accumulation region ( $V_G=0.5V$ ). It may be seen that at the interface, the exchange of carriers between border traps and semiconductor is more rapid, which leads a close-following border trap Fermi level close to the interface in respect to the semiconductor Fermi level. In comparison, for positions that are far from the interface, Fermi level lags the semiconductor in a significant way, which is another representation of the non-quasi-static property across the entire stack during DC sweep.

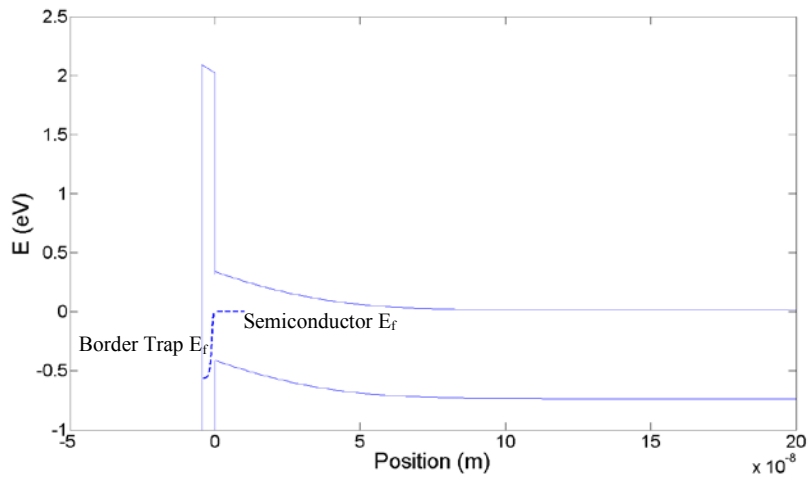


Fig. 5.16. Representative Fermi level position as a function of location during the DC sweep.

2. Upon a change of the surface potential, the incremental gate bias may be calculated accordingly given the increments of the semiconductor charge and the border trap charge.

$$\Delta V_G = \frac{\Delta Q_{semi}}{C_{ox}} + \int_0^{t_{ox}} (t_{ox} - x) \frac{\Delta Q_{BT-DC}(x)}{\epsilon} dx \quad (5.30)$$

On top of the “DC” result, “AC” response may be further computed. The computation procedure is similar to the one outlined for the interface traps, except that the time constants for the border traps are not only energy dependent, but also position dependent. Therefore, integration over the thickness of the dielectric layer is needed.

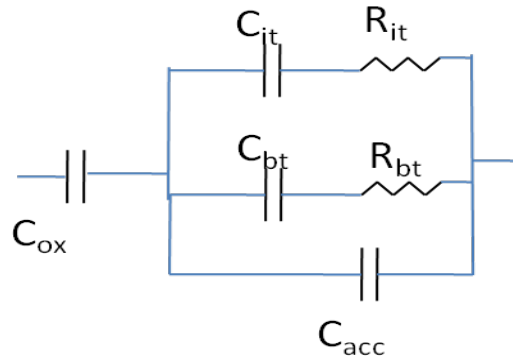


Fig. 5.17. Equivalent circuit model to obtain the overall gate capacitance.

With the computed AC response from both the interface traps and the border traps, we make use of the following equivalent circuit model to obtain the overall gate capacitance, as shown in Fig. 5.17. The admittance for each branch may be expressed as:

$$\begin{aligned}
 Y_{bt} &= \int_0^{t_{ox}} dx \left[ \frac{j\omega C_{bt}}{1 + \omega^2 \tau_{bt}(x)^2} + \frac{\tau_{bt}(x)\omega^2 C_{bt}}{1 + \omega^2 \tau_{bt}(x)^2} \right] \\
 Y_{it} &= \frac{j\omega C_{it}}{1 + \omega^2 \tau_{it}^2} + \frac{\tau_{it}\omega^2 C_{it}}{1 + \omega^2 \tau_{it}^2} \\
 \frac{1}{Y_{total}} &= \frac{1}{Y_{ox}} + \frac{1}{Y_{bt} + Y_{it} + Y_{semi}} \\
 C_{total} &= \frac{\text{Im}(Y_{total})}{\omega}
 \end{aligned} \tag{5.31}$$

where  $Y_{ox}$  is the admittance associated with oxide capacitance,  $Y_{semi}$  is the admittance associated with semiconductor capacitance.



Shown in Fig. 5.18 is a representative computation result in comparison with the experimental data, including only the electron trapping effect of the border traps and interface traps with different small signal frequencies. DC bias interval is assumed to be 1sec between consecutive bias points. As seen, the addition of the border trap effect successfully recreates the frequency dispersion feature that was not explainable by only including the effect from interface traps. With trapping effect, CV curves exhibit stretched out in comparison to the quasi-static solution shown in Fig. 5.1. The dispersion in the weak accumulation region is also shown by the simulation. The dispersion in the weak accumulation region is also shown by the simulation. The computed accumulation region dispersion also reflects the diminishing feature as small signal frequency increases, which is also exhibited in experimental results. There are several issues that must be noted: 1) border trap may also interact with holes, which is however not included in this computation, 2) the distribution of the border trap used in the simulation is a uniform distribution, while in reality the distribution may take other shapes. The detailed distribution may influence the shape of the turn-on characteristics of the CV curve.

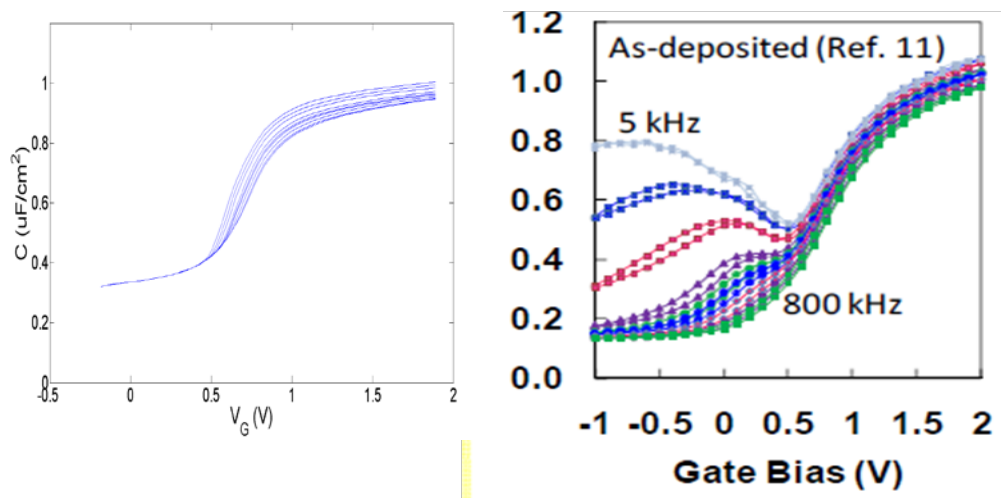


Fig. 5.18. Side by side comparison between experimental data [3] and simulation result. The simulated frequencies are: 5kHz, 10kHz, 20kHz, 50kHz, 80kHz, 100kHz, 200kHz, 500kHz, 800kHz. The simulation only takes into account the electron charging of the traps.

We would like to summarize the understanding on border traps characteristics as the following:

- 1) Traps are distributed throughout the high-k layer in real space, and across the entire energy band gap range of the dielectric. However, the most effective traps (that respond to small signals) are located close to the interface.
- 2) Different small signal frequencies may penetrate different depth, thus pick up response from different portion of the dielectric.
- 3) Quasi-static situation may not be established readily during the DC sweep in a typical LCR meter measurement. Different sweep rate settings will lead different set of CV curves.
- 4) Due to the limited DOS capacitance, capacitive response of border traps is more visible.

It is also necessary to point out that the preceding analysis may not be directly applicable to very thin high-k dielectric, as significant direct tunneling current may take place from semiconductor to the gate. The trapping within the dielectric may therefore be greatly reduced (in that even carriers may be caught by a trap, but they may escape via tunneling process).

The dissertation author would like to acknowledge Eunji Kim, Dr. Shin and Prof. McIntyre at Stanford for providing experimental data. The dissertation author would also like to acknowledge Yu Yuan at UCSD for enlightening discussion on the conductance analysis for border traps.

### 5.3 Subthreshold Characteristics with Non-ideal Dielectrics

In this section, we discuss a new phenomenon in subthreshold I-V characteristics and its physical origin, which is related to the presence of large interface trap density. In the early round of the experiment (device fabrication and measurement result by Prof. Rodwell's group at UCSB), we noticed a strong dependence of the drain current on  $V_{ds}$  in the subthreshold regime, as shown in Fig. 5.19. The exhibited large shift between transfer characteristics indicates a serious DIBL-like effects, even though the device has long channel. This contrasts sharply to the behavior of a "normal" MOSFET behavior, where the dependence of the drain current on  $V_{ds}$  is expected to be  $I_{ds} \propto [1 - \exp(-qV_{ds}/kT)]$ , as shown in the Section 3.3.2 of [5].

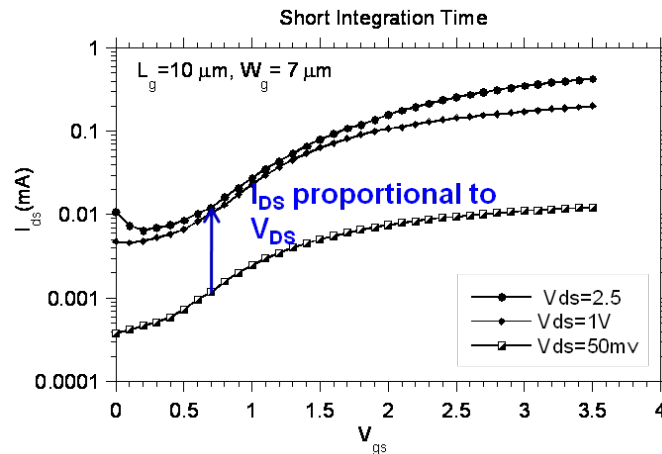


Fig. 5.19. Experimental subthreshold current at different drain biases. (Credit: Prof. Rodwell's group at UCSB)

The almost linear proportionality of the drain current, together with the large subthreshold swing may indicate the presence of excessive interface trap density. To quantify this effect, we adopt gradual channel approximation (GCA) and include the impact of the interface state charge. In the following derivation, we also assume uniform distribution of

interfacial trap density on the energy scale. We define x direction as the direction that is perpendicular to the current transport.

With the presence of charged interfacial traps in subthreshold region, the x direction electrostatic requires:

$$V_G = V_{fb} + \psi_s + \frac{Q_s}{C_{ox}} + \frac{Q_{int}}{C_{ox}} \quad (5.32)$$

where  $Q_{int}$  is the charge trapped at the interface,  $Q_s$  the total sheet charge of the semiconductor. Denote the neutral Fermi level of the interface traps as  $E_n$ , the total trapped (negative) charge can be written as:

$$Q_{int} = qD_{it} \int_{E_v}^{E_c} dE \left( \frac{1}{1 + e^{(E-E_f)/kT}} - \frac{1}{1 + e^{(E-E_n)/kT}} \right) = qkTD_{it} \ln \left( \frac{1 + e^{(E_f-E_v)/kT}}{1 + e^{(E_f-E_c)/kT}} \cdot \frac{1 + e^{(E_n-E_c)/kT}}{1 + e^{(E_n-E_v)/kT}} \right) \quad (5.33)$$

To simplify Eq. (5.33), we define several quantities:

$$\begin{aligned} c_1 &\equiv \frac{1 + e^{(E_n-E_c)/kT}}{1 + e^{(E_n-E_v)/kT}} \\ c_2 &\equiv e^{(E_f-E_v)/kT} \\ c_3 &\equiv e^{(E_f-E_c)/kT} \\ V &= E_f / q \end{aligned} \quad (5.34)$$

On the other hand, the semiconductor charge includes depletion charge as well as charge due to accumulated carrier at the interface. The associated total sheet charge density may be related to the electrostatic potential as:

$$Q_s = Q_D + Q_i = -\frac{\sqrt{2q\psi_s\epsilon_s N_a}}{C_{ox}} - \sqrt{\frac{\epsilon_s q N_a}{2\psi_s}} (kT/q) \left(\frac{n_i}{N_a}\right)^2 e^{q(\psi_s - V)/kT} \quad (5.35)$$

Therefore, Eq. (5.35) maybe written as:

$$V_g = V_{fb} + \psi_s - \frac{\sqrt{2q\psi_s\epsilon_s N_a}}{C_{ox}} - \sqrt{\frac{\epsilon_s q N_a}{2\psi_s}} (kT/q) \left(\frac{n_i}{N_a}\right)^2 e^{q(\psi_s - V)/kT} - \frac{qkTD_{it}}{C_{ox}} \ln\left(c_1 \cdot \frac{1 + c_2 e^{q(\psi_s - V)/kT}}{1 + c_3 e^{q(\psi_s - V)/kT}}\right) \quad (5.36)$$

Eq. (5.36) may be solved numerically to obtain the surface potential  $\psi_s$ , which may be further utilized to compute the subthreshold current, given by:

$$I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\epsilon_s q N_a}{2}} (kT/q) \left(\frac{n_i}{N_a}\right)^2 \int_0^{V_{ds}} e^{q(\psi_s - V)/kT} \frac{1}{\sqrt{\psi_s}} dV \quad (5.37)$$

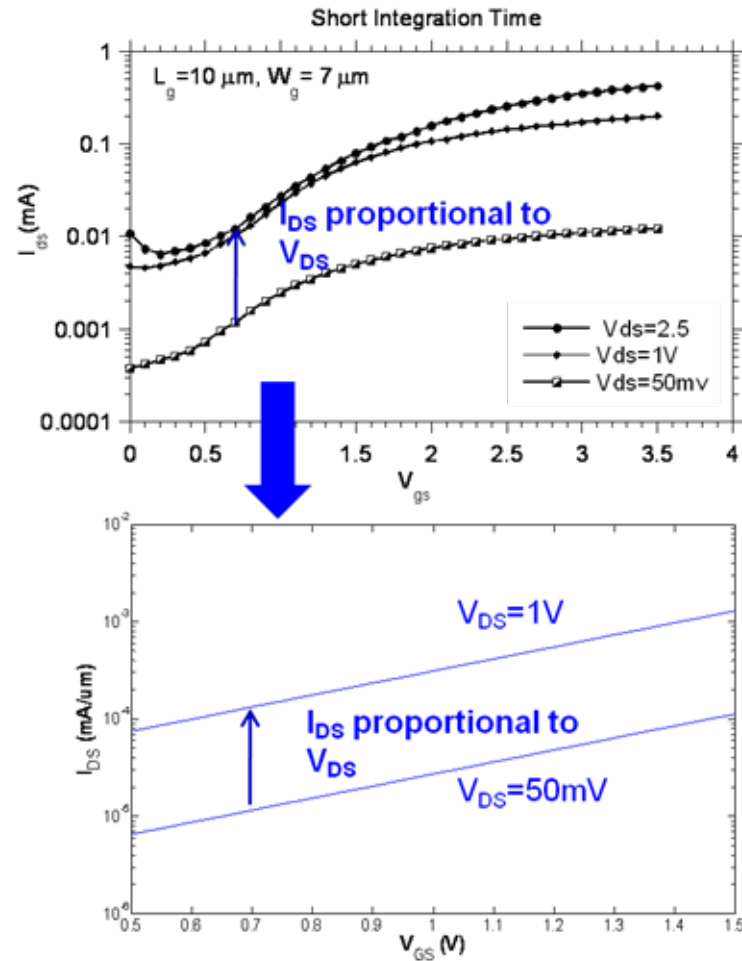


Fig. 5.20. Simulation result of the subthreshold current showing linear dependence of drain current on  $V_{ds}$  bias.

The simulation result, shown in Fig. 5.20 agrees well with the experiment result in terms of the  $V_{ds}$  proportionality of drain current within subthreshold regime. It may be further concluded that this DIBL-like effect is due to large interface state density. We therefore name it as Dit-induced-DIBL effect.

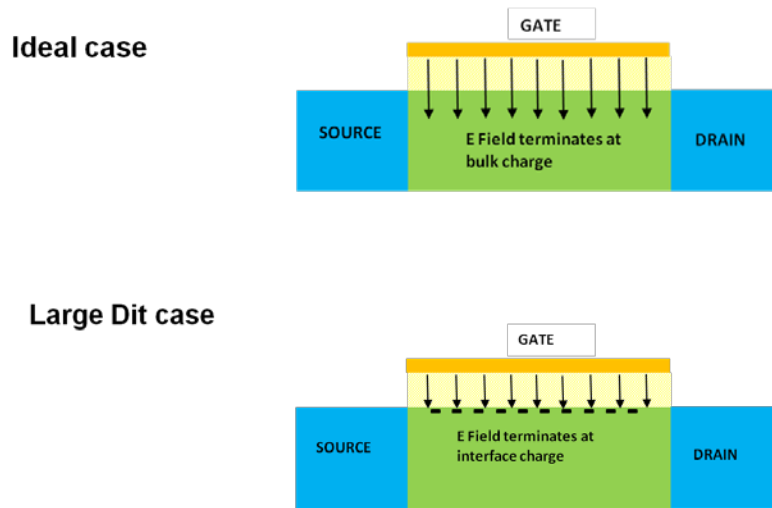


Fig. 5.21. Physical picture of Dit-induced DIBL effects.

Such behavior may also be understood from a physically intuitive perspective, as sketched in Fig. 5.21. For a normal long channel MOSFET, the electric field originated from the gate terminates at the semiconductor charge. In this case, in the vicinity of the semiconductor dielectric interface, vertical electric field is dominant and results in negligible DIBL effect. However, when large amount of interface states is present, the vertical electric field terminates primarily at the interfacial charge, leaving much smaller electric field penetrate into the semiconductor. Therefore, the lateral electric field may become comparable to (in extreme cases, even larger than) the vertical field within the semiconductor close to the interface. Significant charge sharing may therefore result, and lead to pronounced DIBL effect.

In this section, we introduce and analyze a novel phenomenon – Dit-induced-DIBL effect. In addition to the impact on CV characteristics as presented in the previous section, excessive Dit may also lead to large dependence of subthreshold drain current on drain bias.

## 5.4 Summary

In this chapter, we focused on analysis on effects that originate from non-ideal high-k dielectric. Both interface traps and border traps are discussed and analyzed, within the context of III-V (InGaAs) material as the substrate.

For CV characteristics, it is found that:

- 1) Interface states may cause frequency dispersions at weak accumulation/depletion regime, but may not be held responsible for the dispersions at strong accumulation – by that time, the interface state lifetime around  $E_f$  will be much shorter and respond to the small signal on an as-needed basis.
- 2) Border traps may be responsible for the dispersion seen in the strong accumulation. The carrier exchange between semiconductor and border traps is via tunneling, leading to position dependence time constant for traps inside the dielectric. The most effective traps are located close to the border of the semiconductor and dielectric (also where its name comes from). Small signals with different frequencies may pick up response from traps located up to different depth. In a typical LCR measurement, the applied DC bias may not establish a quasi-static situation between semiconductor and dielectric.

For transfer characteristics, it is found that excessive interface traps may induce strong DIBL like effect, even for long channel devices.

Work presented in this chapter has been greatly benefited from collaboration with Prof. Rodwell's group at UCSB and Prof. McIntyre's group at Stanford.



## Reference:

1. Karthik Rajagopalan, Jonathan Abrokwah, Ravi Droopad and Matthias Passlack, "Enhancement-Mode GaAs n-Channel MOSFET", IEEE Electron Device Letters, vol. 27, pp. 959-962, Dec 2006
2. Matthias Passlack, "III-V metal-oxide-semiconductor technology", Indium Phosphide and Related Material Conference, 2008
3. E. Nicollian, J. Brews, "MOS (Metal Oxide Semiconductor) Physics and Technology", Wiley Interscience, 1982
4. Eunji Kim, Lingquan Wang, Peter M. Asbeck, Krishna C. Saraswat, and Paul C. McIntyre, "Border Traps in  $\text{Al}_2\text{O}_3/\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  (100) Gate Stacks and their Passivation by Hydrogen Anneals", submitted to Appl. Phys. Lett.
5. Y. Taur, T. Ning, "Fundamentals of Modern VLSI Devices", 2<sup>nd</sup> Ed., Cambridge University Press, 2009

## **PART II. III-V Nanowire MOSFET**

## Chapter 6. Device Physics and Ballistic Transport of III-V

### Nanowire Based MOSFET

#### 6.1 Introduction

Further scaling of MOSFETs is facing diminishing design space which is bound by performance requirement and leakage control. The latter entails a number of issues, including, but not limited to: direct tunneling leakage through the oxide and SCE related drain leakage. We have seen in the previous part, by using III-V semiconductor as channel material, the device performance may be improved. However, to address the other part of the challenge, i.e. effective control of the leakage current, especially for device designs of extremely short gate length, structure confinement needs to be introduced for the device geometry. Along the evolution line of single gate (planar MOSFET), double gate (e.g. FINFET [1]), tri-gate [2], which was outlined by SIA roadmap [3], it is clear that the ultimate SCE control may be achieved by using gate-all-around structure (also known as wrapped gate structure).

Meantime, the recently advancement in nanowire growth and nanowire device fabrication [4-7] have clearly made nanowire based MOSFET a contender for the possible ultra-scaled device structures. The particular core-shell configuration that may be implemented through radial growth of the nanowire has provided a natural template for a gate-all-around FET. Therefore, combining the structural advantage of core-shell nanowire with the material excellence of III-V semiconductor seems to be a reasonable choice for the next generation short channel MOSFET.

Following a similar discussion sequence as in Part I, we arrange this chapter in the following sequence: in section 6.2, we will develop a necessary tool, i.e. self-consistent

Schrödinger Poisson solver for device analysis on core-shell nanowire MOSFET with radial symmetry; in section 6.3, we will develop a parallel formulation of ballistic transport in this nanowire 1-D transport system; in section 6.4, a comparison will be made on a Silicon nanowire MOSFET and a III-V nanowire MOSFET under ballistic limit.

Work completed in this chapter/part is supported by NSF NIRT program.

## 6.2 Self Consistent Schrödinger-Poisson Solver for Radially Symmetric Nanowire Core-shell Structures

### 6.2.1 Introduction

The rapid advancement in semiconductor nanowire growth technology has motivated numerous research efforts aimed at different applications of nanowires in electronics, optics and biology [4–7]. Among various structures under investigation, the nanowire core-shell structure (with a radial variation in material characteristics, such as semiconductor composition) has been very popular, since it provides great versatility for use in devices such as field-effect transistors [8, 9], photoemitters and photodetectors. To exploit the unique traits stemming from the 1-D quantum structure of nanowires, systematic understanding of the electrical and optical properties is important. In general, the determination of the electronic energy levels and potential distribution requires the self-consistent solution of the Schrödinger equation and the Poisson equation under cylindrical coordinates. Previous work has shown that a general-purpose self-consistent Schrödinger –Poisson solver can be formulated for Cartesian coordinates (as needed for representative planar epitaxial structures) [10]. This paper reports on a numerical solver and its applications for nanowires, exploiting cylindrical symmetry.

The numerical Schrödinger Poisson self-consistent solver accounts for quantum confinement as well as the 1-D nature of the density of states in nanowires. The structure considered and its associated coordinate system are illustrated in Fig. 6.1. The solver is generally applicable to nanowires structures with arbitrary material and doping dependence in the radial direction (the core-shell structure pictured in Fig. 6.1 has one change in material composition along the radius). The solver uses the conventional finite-difference method to solve the

Schrödinger equation. This implies trade-offs of solution accuracy and computation complexity; values chosen in this work provide reasonable computational times for nanowires of physical interest ( $R \sim 5\text{--}500\text{ nm}$ ).

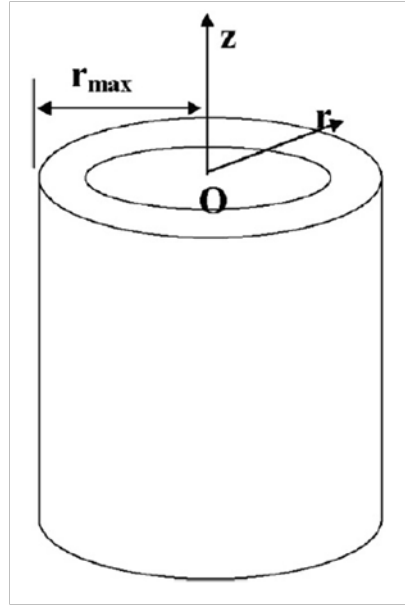


Fig. 6.1. Schematic of the core-shell nanowire structure and the coordinate system used.

From the self-consistent solution, the potential profile, wave functions, electron density and further information can be derived. As an example, the paper illustrates the analysis of capacitance–voltage characteristics for a nanowire core–shell structure configured as an FET, with a radially-deposited metallic gate. For the example case shown in the paper, the gate capacitance is reduced to  $1/3$  of the geometrical barrier limited value, and is only 70% of the classically predicted value (for which only the Poisson equation is solved).

This section is organized in the following manner. Governing equations are discussed in Section 6.2.2, with emphasis on comparing solutions with cylindrical coordinates and Cartesian coordinates. In Section 6.2.3, the solver is tested for a cylindrical constant potential well (where

the exact wave functions are known in analytical form) and a core–shell structure with large dimensions (which asymptotically approaches the slab quantum-well solution). The latter result was compared with the solution given by a well-established one-dimensional solver [10] for validation. Section 6.2.4 gives an example application of the solver in which a core–shell nanowire structure with metallic gate is characterized in terms of electron distribution and quasi-static C–V characteristics; and mechanisms that cause the gate capacitance reduction are analyzed and de-embedded. Model limitations are briefly discussed in Section 6.2.5, while a summary of the paper is provided in Section 6.2.6.

## 6.2.2 *Schrödinger and Poisson Equation Solvers*

### 6.2.2.1 *Schrödinger Solver*

In the presence of varying material composition, the corresponding spatially-varying effective masses have to be taken into account in the Schrödinger equation. The following form of the Schrödinger equation results:

$$-\frac{\hbar^2}{4} \left[ \nabla^2 \left( \frac{\Psi}{m^*} \right) + \frac{1}{m^*} \nabla^2 \Psi + V \Psi \right] = E \Psi \quad (6.1)$$

The effective masses are taken to be energy-independent (thus non-parabolicity effects are neglected). We apply the separation of variables method by assuming the overall wave function can be written as

$$\Psi(r, \theta) = R(r) \Theta(\theta) \quad (6.2)$$

The angular part of the wave function has rotational symmetry, which leads to a solution as:

$$\Theta(\theta) = e^{i\nu\theta} \text{ with boundary condition: } \Theta(\theta) = \Theta(\theta + 2n\pi) \Rightarrow \nu \text{ is an integer} \quad (6.3)$$

The radial part of the wave function satisfies:

$$-\left[ \frac{\hbar^2}{2m^*(r)} \left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{\nu^2}{r^2} \right) + \frac{\hbar^2}{2} \frac{d(1/m^*)}{dr} \frac{d}{dr} + \frac{\hbar^2}{4} \frac{d^2(1/m^*)}{dr^2} + \frac{\hbar^2}{4} \frac{1}{r} \frac{d(1/m^*)}{dr} \right] R + V(r)R = ER \quad (6.4)$$

It may be noted that in cylindrical coordinates, unlike the Cartesian coordinate counterparts, there are additional terms associated with  $r^{-n}$  which bring in a singularity at  $r=0$ . It is therefore necessary to give particular attention to the behavior of solutions in the neighborhood of the origin. Typical nanowire potentials vary slowly around the  $r=0$  point, and can therefore be treated as constants when determining the asymptotic solution. As a result, the Schrödinger equation assumes the following form around  $r=0$ :

$$-\left[ \frac{\hbar^2}{2m^*(r)} \left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{\nu^2}{r^2} \right) \right] R + V(r)R = ER \quad (6.5)$$

Solutions to the above equation are J-type Bessel functions with integer orders [8], which can be expressed as an infinite polynomial series summation, with  $\nu$  as the order of the Bessel function:

$$J_\nu = \sum_{l=0}^{\infty} \frac{(-1)^l}{2^{2l+\nu} l!(\nu+l)!} x^{2l+\nu} \quad (6.6)$$

from which the effect of the Hamiltonian operator can be determined at  $r=0$ :



$$\hat{H}J_v|_{r \rightarrow 0} = \begin{cases} -\frac{\hbar^2}{2m^*} \frac{d^2}{dr^2} J_0 & v=0 \\ -J_v & \text{otherwise} \end{cases} \quad (6.7)$$

With the knowledge of the Hamiltonian's behavior near the origin, discretization of the Hamiltonian elsewhere is straightforward. The boundary condition implemented in the solver enforces that the wave function goes to zero at the outer shell of the cylinder (i.e. an infinite potential is assumed confining the electrons to the nanowire). Solving for eigenvalues and eigenfunctions of the Hamiltonian matrix yields subband energies and corresponding wave functions. The electron concentration is further calculated with the following:

$$n = \sum_m |\Psi_m(r)|^2 \int_{E_m}^{\infty} f(E) g(E) dE \quad (6.8)$$

$$g(E) = \frac{1}{\pi\hbar} \sqrt{\frac{2m^*}{E - E_m}}$$

where  $g(E)$  is the 1-D density of states, and  $f(E)$  is the Fermi–Dirac distribution. The summation over subband indices accounts for contributions from all possible subbands (labeled by footnote  $m$ ), while the integration accounts for contributions belonging to a specific subband (and varying wave vector along the nanowire).

#### 6.2.2.2 Poisson Solver

Two Poisson solvers are implemented in the overall self-consistent scheme. The first gives a Poisson solution for classical electron distributions only, in conjunction with the Fermi–Dirac distribution. The result is then fed into the Schrödinger solver as an initial guess of the potential profile.

The second Poisson solver calculates the potential distribution with the knowledge of electron concentration as a function of position (where the electron concentration is derived from the Schrödinger solver). Detailed implementation of the individual Poisson solvers are described below.

For the initial guess Poisson solver, the following set of equations is solved self-consistently with Newton–Raphson iteration:

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr}\right)\varphi = \frac{q(N_d^+ - n)}{\varepsilon} \quad (6.9)$$

$$n = N_c F_{1/2}\left(\frac{E_f - E_c}{kT}\right) \quad \text{while} \quad E_c = -q\varphi + \Delta E_c(r)$$

Here  $F_{1/2}(E)$  is the Fermi integral of order 1/2.  $N_d^+$  is the radial distribution of ionized donors (or acceptors);  $E_f$  is the Fermi level, taken to be 0 (it serves as the reference level for energy). The surface boundary condition, i.e. the surface potential, is determined by the externally applied voltage. Interfacial charging at the core–shell boundary could be readily accounted for by adding the potential contribution due to the interfacial charge. If cylindrical symmetry is assumed, and also the energy-wise distribution of surface states is known, an extra potential with the form

$$\varphi' = -\frac{\sigma}{\varepsilon} r_{core} \ln \frac{r}{r_{core}}$$

should be included for  $r > r_{core}$  (within the shell region), here is the interfacial charge density per unit area.

The second Poisson solver calculates the potential distribution with a known electron distribution. In order to improve the numerical stability, the differential equation is converted into an integral equation. We denote by  $\delta\varphi$  the difference between the potential profile of two consecutive iterations, and by  $\delta n$  the difference between the electron concentrations as a function of position in successive iterations. The equivalent Poisson equation is written as

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr}\right) \delta\varphi = \frac{q \delta n}{\varepsilon} \quad (6.10)$$

Integrating the above equation once, results in the following:

$$\frac{d(\delta\varphi)}{dr} = \frac{1}{r} \int_0^r \frac{r' q}{\varepsilon} (\delta n) dr' \quad (6.11)$$

In the above equation, we have made use of the identity  $\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr}\right) \delta\varphi = \frac{1}{r} \frac{d}{dr} \left(r \frac{d(\delta\varphi)}{dr}\right)$ .

As a result of symmetry, the electric field at the center is zero, i.e.  $(\delta\varphi)'|_{r=0}=0$ . With a second integration, the expression for  $\delta\varphi$  becomes

$$\delta\varphi = \int_0^r dr' \left[ \frac{1}{r'} \int_0^{r'} \frac{r'' q}{\varepsilon} (\delta n) dr'' \right] + C \quad (6.12)$$

If the surface potential is known and fixed (by means of applying a fixed gate bias or Fermi pinning), then  $\delta\varphi|_{r=r_{max}}=0$ , by which the constant C can be specified.

### 6.2.2.3 Self-consistent calculation

The self-consistent calculation is carried out as described in the flow chart shown in Fig. 6.2. An initial solution of the classical Poisson equation in conjunction with the Fermi–Dirac distribution is fed into the Schrödinger solver.

Subsequently the calculated wave functions and electron concentration as a function of position are provided to the Poisson solver, from which a new electro-static potential is calculated. The new electro-static potential is compared with the previous one; if it satisfies the convergence

criteria, then the calculation is terminated. If not, the new potential is again fed into the Schrödinger solver, and the process is iterated until convergence is reached.

In representative cases, a radial grid with 100–200 equally spaced grid points was established. Convergence could be established with 10–30 iterations. Calculations were carried out on a desktop computer.

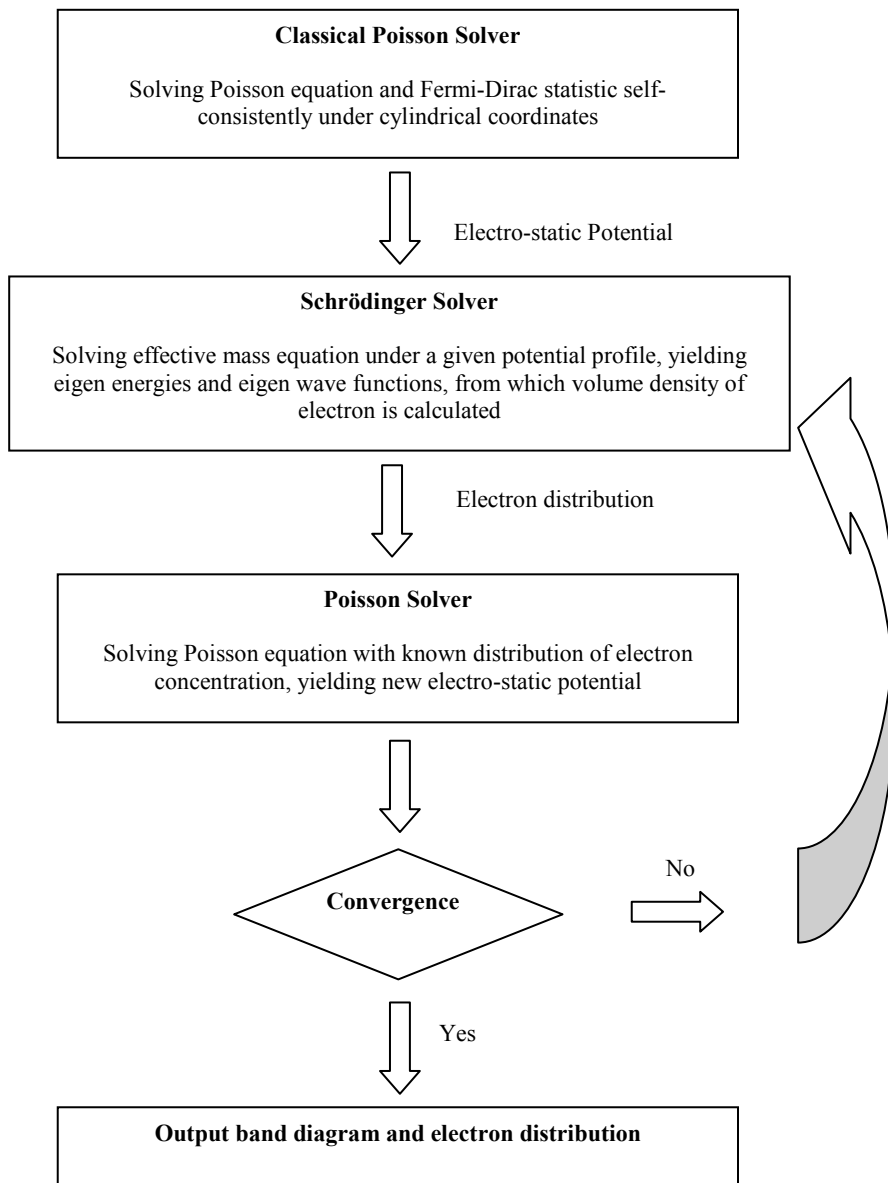


Figure 6.2. Flow chart of the self-consistent solution of the Schrödinger equation and the Poisson equation.

### 6.2.3 Model verification

#### 6.2.3.1 Infinite cylindrical square potential well

The Schrödinger solver was tested under a specific model potential where exact solutions were known as J-type Bessel functions of various orders [11]. The potential was set to be zero inside the core, and infinite in the shell region, which is referred to here as the infinite cylindrical constant potential well. Contributions to the potential from the electron concentration are ignored. The corresponding eigen wave vector values are determined analytically via

$$k_n \cdot r = z_n \quad (6.13)$$

where  $z_n$  represents the  $n^{\text{th}}$  root of the J-type Bessel function. Eigenenergies were subsequently determined with the knowledge of eigen wave vectors by

$$E_n = \frac{\hbar^2 k_n^2}{2m^*} \quad (6.14)$$

Shown in Fig. 6.3 is the comparison between numerical and analytical even parity wave functions, while the inset tabulates the eigenenergies obtained from both methods. Excellent agreement was observed in this case for the numerical solution and the analytical solution.

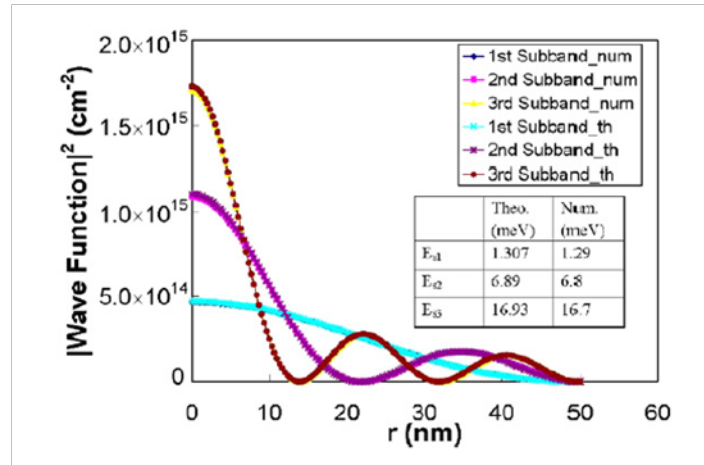


Fig. 6.3. Model verification for the Schrödinger solver for the case of infinite depth cylindrical potential well. Plotted in the figure are the first three wave functions of even parity, from both numerical calculations and analytical predictions. The inset tabulates the comparison between the eigenenergies obtained from the two methods.

### 6.2.3.2 Comparison with 1D Schrödinger–Poisson solver

To verify the entire self-consistent scheme, we further examined our model in comparison with a well-known 1D Schrödinger–Poisson solver [10]. Although this solver was designed for slab structures, when the core–shell nanowire structure has sufficiently large radius, it should approach the slab case limit. The device dimensions were chosen to be  $0.405 \mu\text{m}$  as the core radius and  $0.045 \mu\text{m}$  as the shell thickness. The simulated temperature was set to be 25 K, to avoid excessive influence from the thermal voltage yet still be able to capture the physical essence of the problem. The comparison between our solution and the 1D solution (with slab thickness set to be equal to the radius) is shown in Fig. 6.4. Also shown in a dashed line is the classical solution of the Poisson equation only. Good agreement was demonstrated, with a less than 3% mismatch between the two solutions. The difference we believe is due to the equation form itself, i.e. in cylindrical coordinates, the Laplacian operator gives an additional  $\frac{1}{r} \frac{d}{dr}$  term,

which is a slowly decaying term and still contributes to the wave functions at the radius considered in the given example.

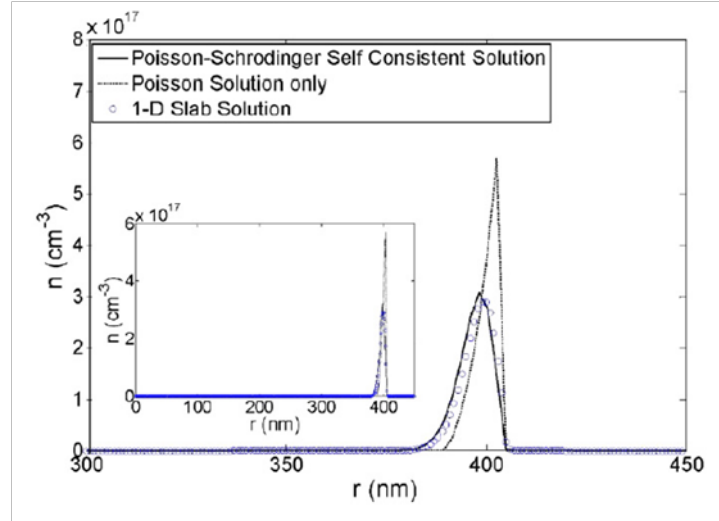


Fig. 6.4. Model verification by comparing the present solver with a 1D Schrödinger-Poisson slab quantum well solver for a large radius nanowire. The simulation was conducted for a 450 nm radius core-shell structure. The same thicknesses were used in the slab quantum-well solver. Inset shows a comparison within the entire radius range, showing good agreement between the two solutions.

#### 6.2.4 Model Applications

Core-shell nanowire structures are of substantial interest for electrical and optoelectronic applications. Core-shell configurations can utilize a great diversity of semiconductor combinations, since the nanowires are relatively tolerant to lattice mismatch between materials, and the associated stress and strain fields. Quantitative understanding of electron concentrations is important for further work, particularly for the two-dimensional electron gas established at the interface between semiconducting materials in appropriately designed nanowires. We therefore applied our model to a representative device of this type. Shown in Fig. 6.5 are the computed results for an InAs/ GaAs core shell structure, where the inner core is formed of InAs (with radius



36 nm), and the outer shell is of GaAs (so that the overall radius is 40 nm). Bandgap energies and electron masses were taken from [12]. It was assumed that there were no interface states at the boundaries of the materials, and that the band offset in the conduction band corresponds to 0.62 eV (which reflects a 58–42% split in conduction band offset and valence band offset [12]). Fig. 6.5a presents wave functions for the first few subbands for the core–shell structure. It can be noted that there is a set of subbands corresponding to radial variation of the wave function, in a manner that resembles the subband structure of slab quantum well structures. Each of these subbands has a further splitting associated with the angular variation of the wave functions. The radial dependence of the wave functions varies slightly with the angular quantum number, as shown in Fig. 6.5(a). Inasmuch as the externally applied potential has circular symmetry, we can label each subband with s, p, d ... following the well-established convention. Differences between the radial wave functions due to the angular variation are more pronounced for the higher energy subbands. To illustrate, first few eigenenergies are shown with thinner lines in Fig. 6.5(b).

The conduction band profile and the electron concentration are shown in Fig. 6.5(b) and (c), respectively. It is observed that although the conduction band for the quantum wire is below the classically predicted value (which is based on 3-D bulk statistics), the electron density is much smaller near the surface, due to the form of the low energy wave functions and smaller density of states in the 1-D system. Also shown in Fig. 6.5(b) is the result obtained from the 1D Schrödinger–Poisson slab solver, in which the slab thickness was set to be equal to the radius. Compared to its slab counterpart, the cylindrical solution exhibits more quantum-like corrections, i.e. the peak of the electron concentration is pushed more towards the center, and the concentration peaks at a lower value. Characteristics of quantum-wires with different radius can differ significantly. Shown in Fig. 6.6 are the conduction band profile and electron concentration as functions of position in a smaller device featuring 20 nm radius (but otherwise identical

characteristics to the device in Fig.6.5). In the smaller size device, with the same surface potential, the electron density throughout the volume is significantly larger than the 40 nm radius case (which compares well with the known results of increasing volume inversion with thinner slabs in the slab-like FinFET devices [13]).

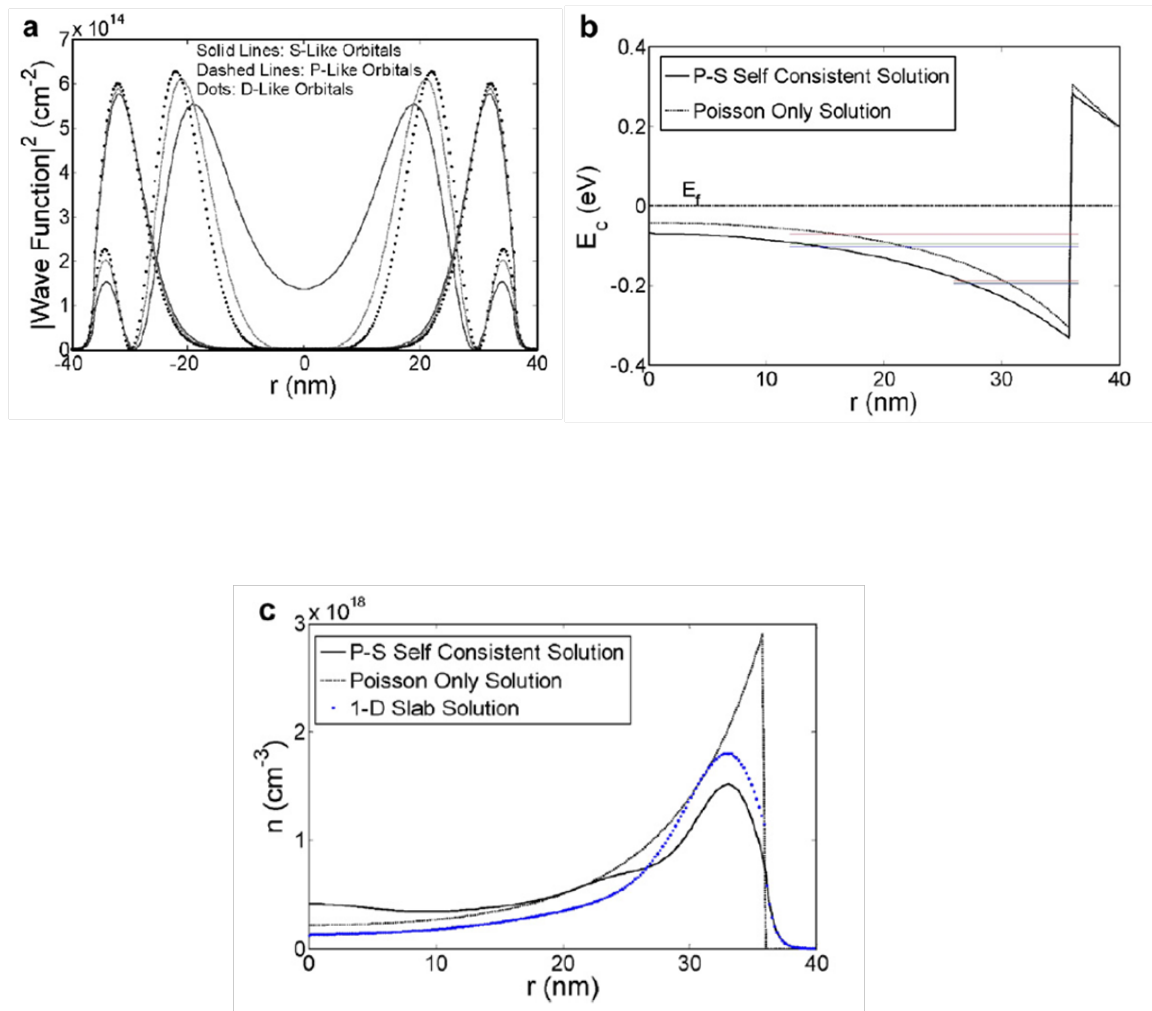


Fig. 6.5. Model application to the self-consistent calculation of 2-DEG formation in an InAs/GaAs core shell structure. The calculated structure has an overall radius of 40 nm; the InAs core radius is 36 nm. (a) Wave functions for first few subbands. (b) Conduction-band energy for the quantum self consistent solution (solid line) and classical Poisson only solution (dashed line). Eigenenergies were also shown with thin lines, corresponding to the wave functions shown in (a). (c) Electron distribution in quantum self-consistent solution (solid line) and classical Poisson only solution (dashed line). Also provided is the solution from the slab quantum-well solver for a slab sandwich structure constituted by InAs and GaAs with thicknesses as 36 nm and 4 nm.

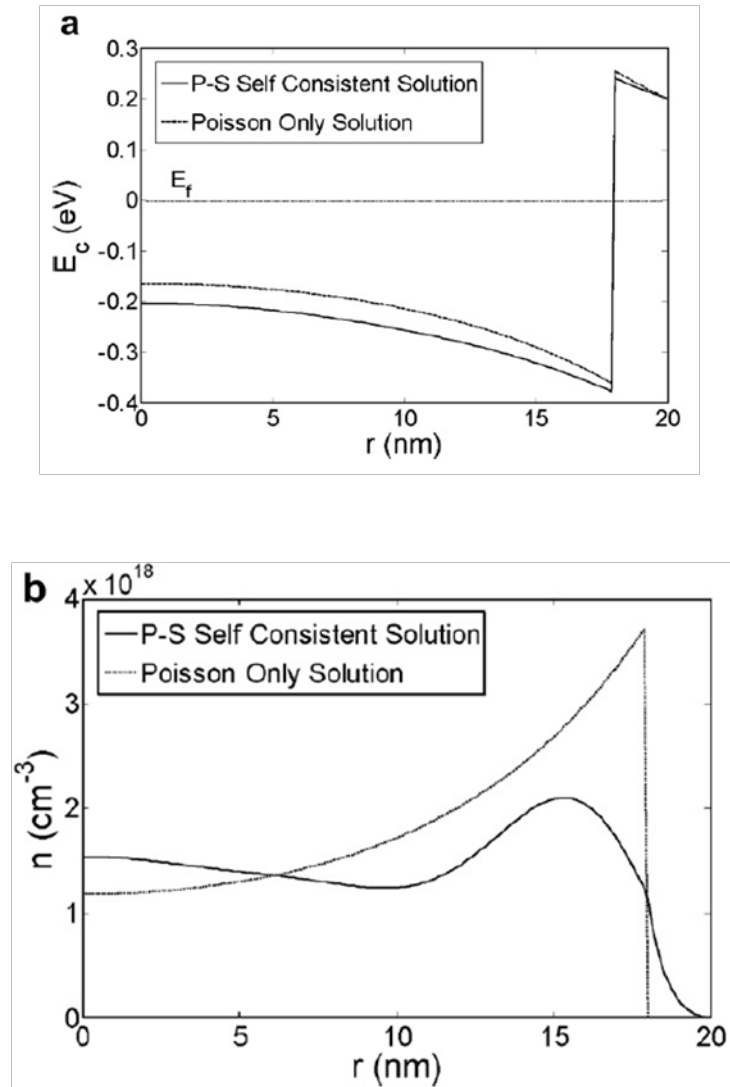


Fig. 6.6. Model application to the self-consistent calculation of 2-DEG formation in an InAs/GaAs core shell structure. The calculated structure has an overall radius of 20 nm; the InAs core radius is 18 nm. The results show a more pronounced volume electron density than in the case of Fig. 6.

The general-purpose Schrödinger–Poisson solver described here can be used in various nanowire device contexts. The numerical solver provides the capability to find discrete eigenenergies, identify corresponding wave functions, specify the electron concentration as a function of position, and calculate electrostatic potential and radial a representative nanowire FET,

with core-shell structure electric fields. From these basic functions additional capabilities can be derived, such as calculating the capacitance-voltage characteristics of FET structures. The geometry of a representative nanowire FET, with core-shell structure and metal gate, is shown in Fig. 6.7a. For simplicity yet without losing any physical essence, we ignored here the contact potential between a metal gate and the shell material. For a given voltage difference between the metal Fermi level and the Fermi level of electrons in the nanowire, the overall charge  $Q$  can be calculated, and the  $C$ - $V$  curve can be further evaluated through  $C=dQ/dV$ . Fig. 6.7 is a demonstration of this capability for a 40 nm radius InAs/GaAs core-shell structure with 4 nm thick GaAs barrier layer. Interestingly, the  $Q$ - $V$  curves exhibit ideal 60 mV/dec “subthreshold” characteristic, which is expected for the ultra-thin-body fully depleted FET devices. It is noticed that the resulting gate capacitance is considerably lower than its geometrical barrier limit (only  $\sim 1/3$  the value calculated as a co-axial cylinder capacitor between the gate metal and the InAs/GaAs interface); and is  $\sim 70\%$  of the classically predicted capacitance.

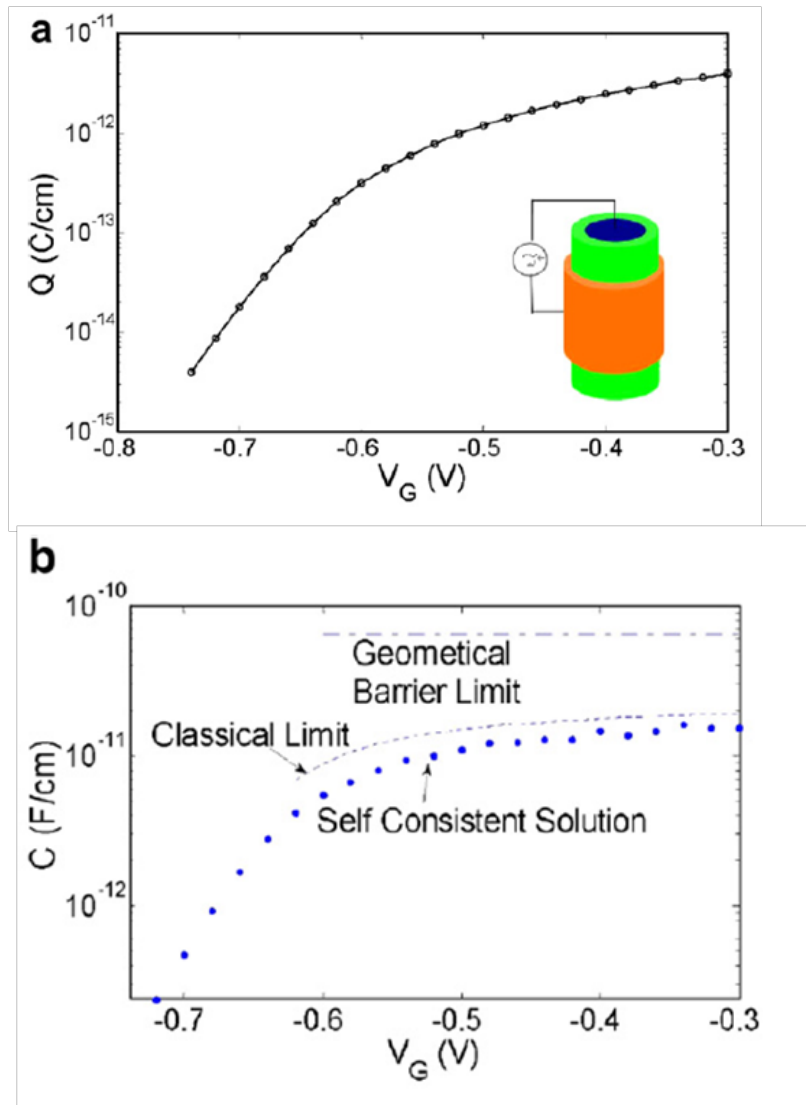


Fig. 6.7. (a) Predicted  $Q_s$ - $V$  characteristic for a structure similar to that of Fig. 6. (b) Derived  $C$ - $V$  characteristic for the same structure, with comparison of the classically predicted  $C$ - $V$  curve, the horizontal line is the capacitance calculated as a co-axial cylindrical capacitor. Inset of (a) shows schematically the structure under consideration.

The reduction of gate capacitance can be understood via the carrier distribution: self-consistent calculation predicts that the charge centroid is pushed away from the InAs/ GaAs interface both classically and, additionally from the quantum effects known as the energy quantization and reduced 1-D density of state (DOS) (see Figs. 5(c) and 6(b)) [14]. The capacitance reduction is considered detrimental in terms of reduced transistor transconductance.

To characterize effects from different mechanisms, we employ a phenomenological equivalent circuit depicted in Fig. 6.8, where the gate capacitance  $C_g$  is considered to be a series combination of the classically predicted gate capacitance and a quantum correction capacitance  $C_{qc}$ .  $C_{qc}$  originates from the finite spread of the wave function associated with each energy level and the reduced DOS from the low dimensionality of the structure. To further distinguish the relative importance of these mechanisms, we introduce  $C_{qc1}$  and  $C_{qc2}$  to account for their individual contributions; the total quantum capacitance  $C_{qc}$  is a series combination of  $C_{qc1}$  and  $C_{qc2}$  (see Fig. 6.8).

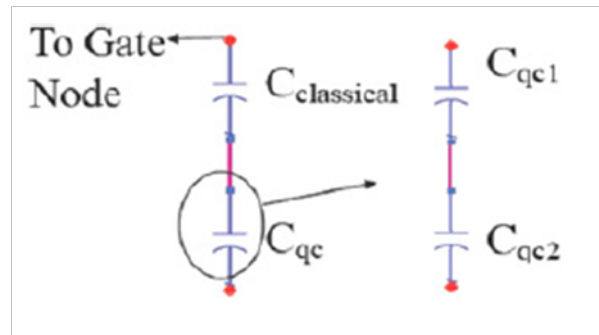


Fig. 6.8. Equivalent circuit for de-embedding quantum capacitances.  $C_{qc}$  is the collective quantum capacitance, while  $C_{qc1}$  stands for the influence due to the finite wave function spread;  $C_{qc2}$  accounts for the influence due to the reduced DOS.

The de-embedding proceeds as following (using results presented in Fig. 6.7): first,  $C_{qc}$  is de-embedded given the classically predicted value and the self-consistent quantum mechanical solution (shown in Fig. 6.9); second,  $C_{qc1}$ , which represents the influence from finite wave function spread, is obtained by substituting the 3-D DOS function into Eq. (6.8) instead of the 1-D DOS function used previously, so that the 1-D DOS influence is excluded; third, with the knowledge of  $C_{qc}$  and  $C_{qc1}$ ,  $C_{qc2}$  was calculated via the equivalent circuit presented in Fig. 6.8. For this particular example, it is shown that the low dimensionality is the dominant mechanism of the reduced gate capacitance.

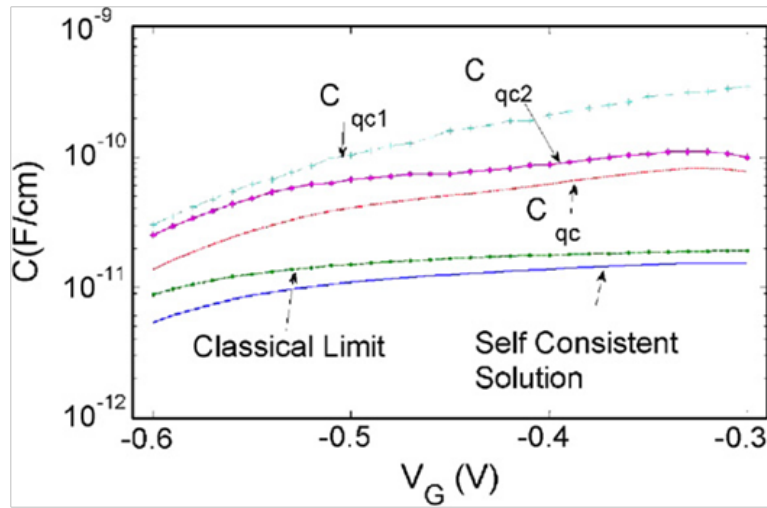


Fig. 6.9. De-embedded quantum capacitances as functions of voltage

### 6.2.5 Limitation of the Model

Various simplifications have been introduced into the solver which can limit the accuracy in a number of situations. As noted above, the effective mass is assumed to have a constant value. Thus non-parabolicity effects are not taken into account. The implied isotropy of the effective mass omits a number of details present in the conduction band of Si and other multi-valley semiconductors. Our present treatment of holes is also associated with a simple effective mass; the full complexity of light and heavy hole bands and their interactions are ignored.

### 6.2.6 Summary

A general-purpose Schrödinger–Poisson solver was developed for symmetric structures in cylindrical coordinates. This solver is useful for a wide range of semiconductor nanowire problems. The solver allows evaluation of discrete eigenenergies, and wave functions for a given

geometry, based on which the electro-static potential and the electron concentration can be calculated, all in a self-consistent manner. Self-consistent C–V characteristics of a given structure can be readily deduced from the model, and compared to their classical counterparts, which allows for assessment of the impact of quantum capacitance. A de-embedding method is further introduced to distinguish individual contributions from finite wave function spread and low dimensionality to the quantum capacitance.

Work completed in this section has been published in the paper “A numerical Schrödinger–Poisson Solver for Radially Symmetric Nanowire Core–shell Structures”, *Solid State Electronics*, Nov. 2006.



## 6.3 Ballistic Transport in III-V Nanowire MOSFETs

### 6.3.1 Introduction

Nanowire structures exhibit different I-V characteristics than their planar counterparts. One source of difference is reduced density of states (DOS) in the case of the nanowire, and as a result, reduced electron density and reduced capacitance between channel and gate. Another source of difference is that for nanowires with short gate length, transport is expected to be close to ballistic limits. In this section, the impact of the density of states is first reviewed. Subsequently, ballistic current limit in nanowires is presented. Values of current obtained from this ballistic analysis must be regarded as upper limits to the current obtained in real devices, which will be unavoidably influenced by scattering.

### 6.3.1 Density of States (DOS) in Nanowires

For very small nanowire diameter, the density of electron states for the nanowire Density of States (DOS) in Nanowires approaches the result expected for one dimensional semiconductors, which differs from the value in two and three dimensions. In conventional planar devices as well as double gate structures, carriers reside in the channel where confinement is only present in one direction, leaving the carriers as a two-dimensional electron gas. In this case, if parabolic band structure is assumed, the DOS function is simply a staircase function, with each step height equal to  $\frac{m^*}{\pi\hbar^2}$ . By comparison, in a nanowire FET structure, confinement is present along two directions (radial and azimuthal); accordingly, the DOS function takes a different form which can be written as:

$$g(E) = \sum_{\mu, \nu} \frac{1}{\pi \hbar} \sqrt{\frac{2m^*}{E - E_{\mu, \nu}}} u(E - E_{\mu, \nu}) \quad (6.15)$$

where  $m^*$  is the effective mass along the unconfined direction;  $E_{\mu, \nu}$  represents energy minima of individual subbands with band indices  $\mu$  and  $\nu$ ;  $u(x)$  is the unit step function which takes value of 1 when  $x \geq 0$  and is 0 otherwise.

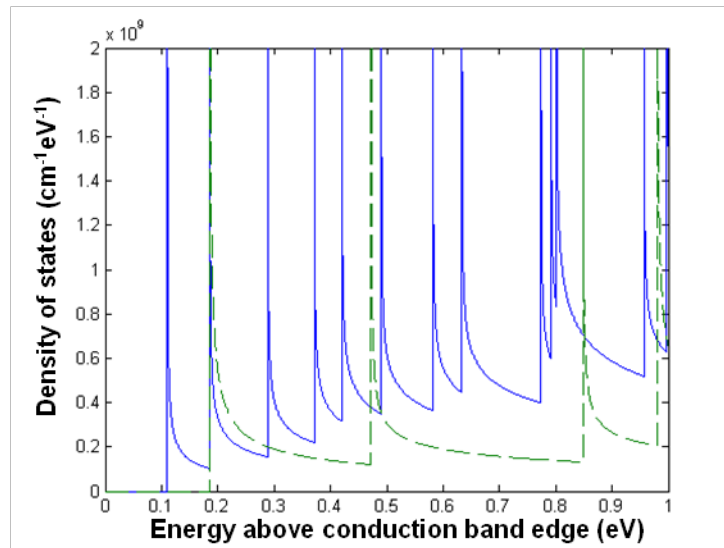


Fig. 6.10. Computed density of states of a Si nanowire with 2.5nm radius under flat-band condition using the model presented in [15]. Solid line and dashed line represent different eigen energy ladder due to anisotropic effective mass along different direction.

Figure 6.10 illustrates, for example, the density of states for a silicon nanowire of radius of 2.5 nm, computed below threshold, according to the model described in [15]. DOS contributions associated with different valleys of the conduction band are different, according to whether or not the ellipsoid is aligned along the direction of the nanowire. For gate biases above threshold, the electrostatic potential introduced by the channel electron density further modifies the density of states. The eigen energy of each subband,  $E_{\mu, \nu}$ , in general, may be obtained from a self-consistent Schrödinger Poisson solver that takes into account the self-consistent electrostatic

potential of the electron density, as well as the cylindrical geometry [16]. The spatial distribution of the carrier density can be written as:

$$n(r) = \sum_{\mu,\nu} \frac{1}{\pi\hbar} |\psi_{\mu,\nu}(r)|^2 \int_{E_{\mu,\nu}}^{\infty} \sqrt{\frac{2m^*}{E - E_{\mu,\nu}}} \cdot \frac{1}{1 + e^{\frac{E - E_f}{kT}}} \cdot dE \quad (6.16)$$

where  $\psi_{\mu,\nu}(r)$  is the wave function associated with the subband with band index  $(\mu, \nu)$ . After integration of the wave function in Eq. (6.16) over the cross section, the line charge density can generally be simplified to

$$N \approx \sum_{\mu,\nu} \frac{1}{\pi\hbar} \left( \int_{E_{\mu,\nu}}^{\infty} \sqrt{\frac{2m^*}{E - E_{\mu,\nu}}} \cdot \frac{1}{1 + e^{\frac{E - E_f}{kT}}} \cdot dE \right) \quad (6.17)$$

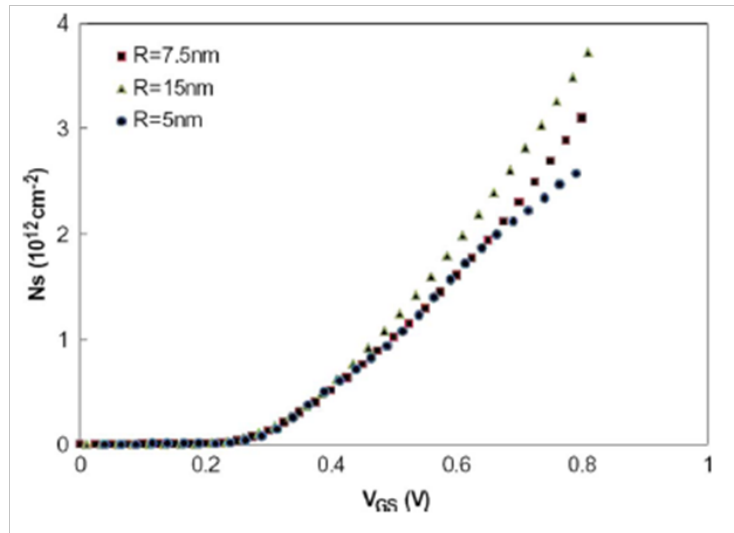


Fig. 6.11. Computed sheet charge density (line charge density normalized by core periphery) of the structure shown in Fig 7a inset, using a self consistent Schrödinger Poisson Solver considering cylindrical coordinates.

Figure 6.11 shows the charge density computed according to Eq. (6.17) for a core-shell III-V nanowire of the structure illustrated in the inset of Fig. 6.7(a). The core material is  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ , while the shell (barrier) corresponds to  $\text{Al}_{0.35}\text{Ga}_{0.65}\text{Sb}$  (which functions approximately as the dielectric layer of a MOSFET, with an equivalent oxide thickness of 0.55nm). The curves shown are for core radius of 5nm, 7.5nm and 15nm. The charge densities were obtained by, normalizing the corresponding line charge densities to the core periphery. For the large radius case the carrier density value is comparable to the expected value for planar FETs, but there are reductions for nanowires with smaller dimensions.

With the charge voltage relationship calculated above, quasi-static C-V characteristics can be derived. An overall equivalent circuit for the gate-channel capacitance of the nanowire is shown in Fig. 6.12. Here  $C_{ox}$  corresponds to the barrier or insulator geometric capacitance;  $C_{qw}$  represents the modification to the capacitance associated with the finite depth of the wave function below the core/shell interface; and  $C_{DOS}$  is the DOS contribution. The overall C-V relation obtained for the nanowires with core radius of 5nm is shown in Fig 6.13.

As shown in the figure, influences on the overall capacitance characteristics from quantum confinement and limited DOS are significant. The overall capacitance is only ~15% of the geometrical barrier limit, where the whole structure is considered to be a concentric cylinder capacitance (and assumes a charge sheet present at the InGaAs/AlGaSb interface). The overall capacitance is still considerably smaller (~35%) than the capacitance value with quantum confinement, if 1.5 nm average wave function depth is assumed. This result suggests that the DOS will be a significant capacitance (and current) limiter in scaled core/shell FET structures. Also suggested is the fact that scaling the dielectric thickness to very small dimensions is not

necessary in order to maintain high current. The figure also shows carrier population in multiple subbands with increased gate bias, which is marked by distinct “humps” in the C-V curve. Furthermore, a slight decrease in capacitance results as gate voltage increases between the “humps”; this is due to decreasing DOS as a function of energy exhibited in Eq. (6.15).

To gain further physical insight in terms of the impact of 1-D DOS on overall capacitance, one may perform the derivative with respect to  $V_G$  upon line charge density.

$$\begin{aligned}
 C_{total} &= \frac{qdN}{dV_G} = \frac{qdE_f}{dV_G} \frac{d}{dE_f} \sum_{\mu,\nu} \frac{1}{\pi\hbar} \left( \int_{E_{\mu,\nu}}^{\infty} \sqrt{\frac{2m^*}{E - E_{\mu,\nu}}} \cdot \frac{1}{1 + e^{\frac{E - E_f}{kT}}} \cdot dE \right) \\
 &= \frac{qdE_f}{dV_G} \sum_{\mu,\nu} \frac{1}{\pi\hbar} \int_{E_{\mu,\nu}}^{\infty} dE \cdot \sqrt{\frac{2m^*}{E - E_{\mu,\nu}}} \cdot \frac{d}{dE_f} \left( \frac{1}{1 + e^{\frac{E - E_f}{kT}}} \right)
 \end{aligned} \tag{6.18}$$

The  $\frac{d}{dE_f} \left( \frac{1}{1 + e^{\frac{E - E_f}{kT}}} \right)$  term can generally be approximated as a delta function  $\sim \delta(E - E_f)$ , and

thus the overall capacitance is approximately  $\frac{dN}{dV_G} \propto \frac{dE_f}{dV_G} \sum_{\mu,\nu} \frac{1}{\pi\hbar} \sqrt{\frac{2m^*}{E_f - E_{\mu,\nu}}}$ . This expression

illustrates that small effective mass will limit the attainable capacitance (and charge density).

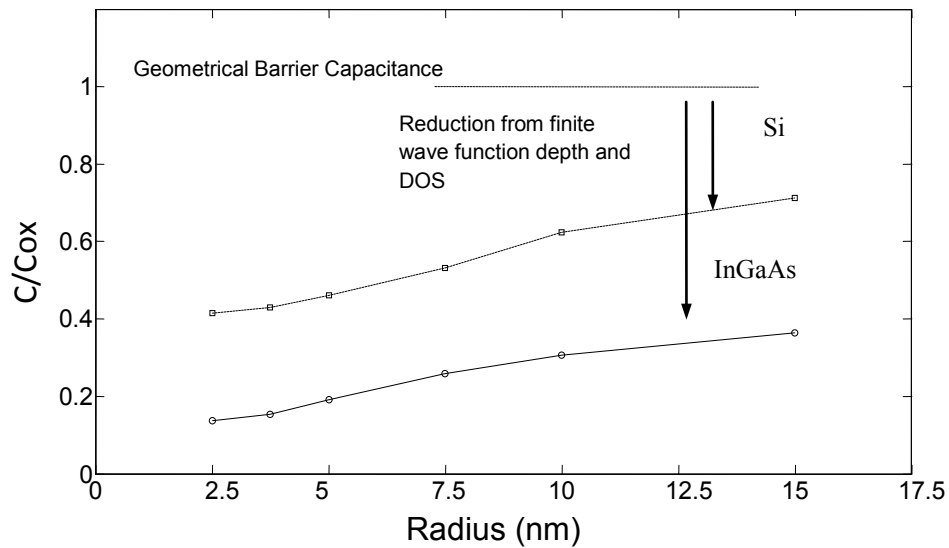


Fig. 6.14. Overall capacitance for InGaAs core and Si core with different diameters normalized to the oxide (geometrical barrier) capacitance.

To this point, the discussion has assumed a single valley conduction band, such as occurs in most III-V materials. In case of Si, one must consider degeneracy of the conduction band (which is six-fold in the case of 3D unstressed materials, but which splits into two-fold plus four-fold sets as explained above). Figure 6.14 illustrates the behavior of capacitance v.s. nanowire radius computed for representative Si nanowire MOSFETs and for III-V structures, normalized to the geometric capacitance. The results indicate that the DOS reduction is much more significant for III-V materials than for Silicon.

### 6.3.3 Ballistic Limit Current

With scaling, the gate length of nanostructure FETs is approaching the carrier mean free path. For even shorter dimensions, transport can be considered to be ballistic or quasi-ballistic. In planar devices, ballistic efficiency of ~60% has already been demonstrated experimentally [17]. Study of ballistic transport behavior provides an upper limit to the current drive that can be

attained. The ballistic transport model was originally proposed by Natori [18] and elaborated by Lundstrom and coworkers [19], using simplified assumptions for the charge voltage relationship. In the following, the ballistic transport will be recapped, followed by specific computed examples.

Along the transport direction, there is typically a point within the channel where the longitudinal electric field is zero, which is often referred to as the “virtual source”. It is assumed quasi-equilibrium conditions apply here, with forward going electrons filling up the band according to the Fermi level in the source, and backward-going electrons according to the Fermi level in the drain.

It is worthwhile noticing that due to the two dimensional confinement, a group of closely spaced eigen-states is present. By contrast, in the planar case, subbands are generally found to be separated more widely. Thus, to account the charge and current correctly in the nanowire case, multiple subbands need to be considered.

The forward going current can be written as:

$$J_+ = q \sum_{\mu, \nu} \int_{E_{\mu, \nu}}^{\infty} n_+(E) v_+(E) dE = q \sum_{\mu, \nu} \int_{E_{\mu, \nu}}^{\infty} \frac{1}{2} g(E) f(E, E_{fs}) v_+(E) dE \quad (6.19)$$

where  $g(E)$  is the DOS function and  $f(E, E_f)$  is the Fermi distribution function;  $v(E)$  is the carrier velocity. Note that as only forward going electrons are counted here, half of the DOS function is used as the other half is associated with the backward going electrons.

Similarly, backward going current can be written as:

$$J_- = q \sum_{\mu,\nu} \int_{E_{\mu,\nu}}^{\infty} n_-(E) v_-(E) dE = q \sum_{\mu,\nu} \int_{E_{\mu,\nu}}^{\infty} \frac{1}{2} g(E) f(E, E_{f_s} - qV_{DS}) v_-(E) dE \quad (6.20)$$

The expression for the forward going and backward going current can be further simplified, yielding the total current as expressed below:

$$J_{total} = J_+ - J_- = \frac{q}{\pi\hbar} \sum_{\mu,\nu} \int_{E_{\mu,\nu}}^{\infty} dE \left( \frac{1}{1 + e^{\frac{E - E_{f_s}}{kT}}} - \frac{1}{1 + e^{\frac{E - E_{f_s} + qV_{DS}}{kT}}} \right) \quad (6.21)$$

$$J_{total} = \frac{kTq}{\pi\hbar} \sum_{\mu,\nu} [\ln(1 + e^{(E_{f_s} - E_{\mu,\nu})/kT}) - \ln(1 + e^{(E_{f_s} + qV_{DS} - E_{\mu,\nu})/kT})]$$

The total charge in the nanowire at the position of the virtual source may be similarly computed. The resulting value is  $V_{DS}$  dependent, since  $V_{DS}$  determines the Fermi level controlling the backward going electrons. The resultant value of charge is:

$$N_{total} = \frac{\sqrt{2m^*}}{\pi\hbar} \sum_{\mu,\nu} \int_{E_{\mu,\nu}}^{\infty} \frac{1}{\sqrt{E - E_{\mu,\nu}}} \left( \frac{1}{1 + e^{(E - E_{f_s})/kT}} + \frac{1}{1 + e^{(E - E_{f_s} + qV_{DS})/kT}} \right) dE \quad (6.22)$$

In the planar device case, a useful simplification of this formalism is the so-called ‘‘one subband’’ approximation, which assumes only one subband is occupied above threshold. Such consideration is reasonable when the channel is sufficiently narrow (such as in the case of carbon nanotubes), and thus subbands are widely separated relative to  $kT$  and to  $V_{GS} - V_t$ , implying minimal population in the higher subbands. With this consideration, an exact solution of  $E_{f_s} - E_t$  (energy separation between source Fermi level and 1<sup>st</sup> subband bottom) can be obtained via the following equation (without solving the Schrödinger equation self consistently).



$$V_{GS} - V_t = \frac{1}{q}(E_{fS} - E_1) + \frac{q\sqrt{2m^*}}{\pi\hbar C_{ox}'} \int_{E_1}^{\infty} \frac{1}{\sqrt{E - E_1}} \left( \frac{1}{1 + e^{(E - E_{fS})/kT}} + \frac{1}{1 + e^{(E - E_{fS} + qV_{DS})/kT}} \right) dE \quad (6.23)$$

where  $C_{ox}'$  is the effective oxide capacitance including the finite wave function depth. The second term represents the potential drop over the oxide. In the nanowire case, the single subband approximation is often insufficient, however, because of the tight spacing of energy levels. With the low  $C_{DOS}$ , often the Fermi level in the nanowire follows closely the value of  $V_{GS} - V_t$  and strongly degenerate electron populations are obtained. Even subbands that lie several  $kT$  in energy above the lowest subband can thus be occupied. A more accurate approach is to solve self-consistently Schrödinger Poisson equations, considering band filling explicitly with Fermi levels  $E_{fS}$  and  $E_{fS} - qV_{DS}$  to account for forward and backward going electrons. Multiple subbands are included when necessary. Detailed computation considerations may be found in Ref [16]. The importance of this procedure is illustrated in Figure 6.15(a) and (b) below, in which the ballistic current limit is shown for the same InGaAs/AlGaSb core/shell structure as shown before. Results computed using only the lowest energy state and using the full self-consistent Schrodinger Poisson solver are shown. The equivalent oxide thickness used here is 0.55nm.

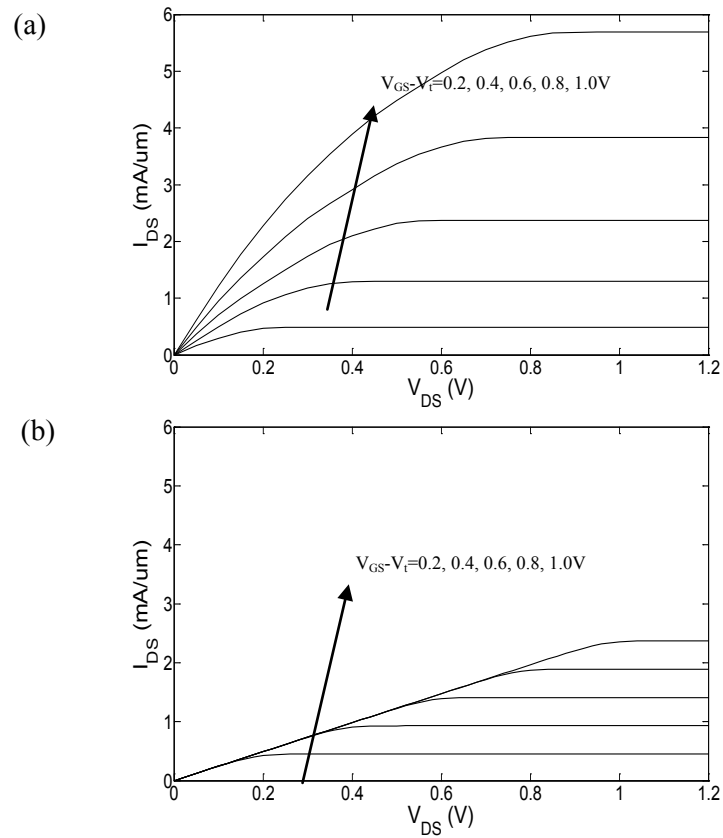


Fig. 6.16. Computed ballistic output characteristics for InGaAs nanowire with (a) self consistent solution and (b) one subband approximation.

For the one subband case, the on-resistance  $dI_{DS}/dV_{DS}$  at low  $V_{DS}$  is fixed at the value of  $h/2e^2$ , as expected for the resistance of a single quantum channel. The computation shows that this is a significant underestimation, however, for the range of gate voltages applied. The discrepancy between the two computation methods mainly stems from the crowded subband distribution in the nanowire scenario, i.e. subbands are no longer separated widely enough that higher order subbands occupation can be ignored. The closely packed subband profile is mainly due to the azimuthal confinement which lifts the degeneracy of eigen-energy resulting from radial alone confinement [16].

### **6.3.4 Summary**

In this section, we formulated the ballistic model, with 1-D transport consideration, for the nanowire MOSFET. We emphasized the importance of 1-D DOS and the necessity to include multiple subbands for the computation of ballistic current, due to the azimuthal confinement in a core-shell nanowire. We will further compare the performance difference between a Si nanowire MOSFET and a III-V nanowire MOSFET under the ballistic limit.

The work reported in this section is part of the paper titled “Scaling of Nanowire Transistor”, published in IEEE Transaction of Electron Device, Nov., 2008.

## 6.4 Performance Comparison of Scaled III-V and Si Ballistic Nanowire MOSFETs

### 6.4.1 Introduction

In this section, we will use the theoretical formulation that we established in the last section and compare the ballistic limit performance of III-V and Si nanowire MOSFETs (NWMOSFETs). The simulation reveals interesting tradeoffs between the two types of material. High mass and multi-valley degeneracy may have a positive or negative influence on the saturation current, depending on the effective oxide thickness (EOT) of the device. In all cases, high mass and multi-valley degeneracy results in smaller on-resistance in the ballistic limit at low drain bias than low mass, low valley degeneracy.

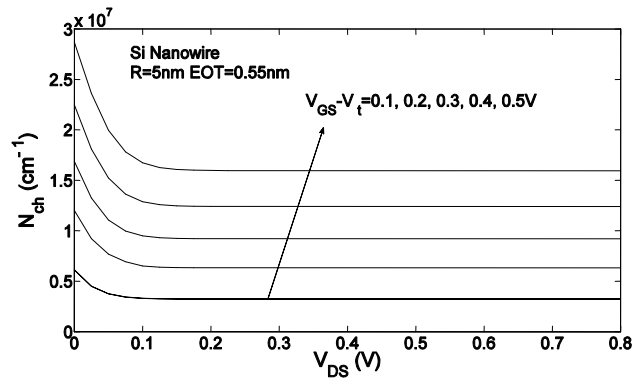
### 6.4.2 Performance Comparison between Si and III-V Nanowire MOSFET

The structure simulated numerically is a gate-all-around (GAA) MOSFET, as shown schematically in Fig. 6.16. Two types of MOSFETs under study differ only in the core (channel) material: one with III-V core material taken to be  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  with effective mass of 0.041, and the other with Si as the core material. The shell insulators are the same for comparison purposes. In our simulation, Si was approximated as an isotropic material with effective mass of 0.258 ( $3 \cdot m_i \cdot m_l / (2m_l + m_l)$ ) and degeneracy factor of 6.

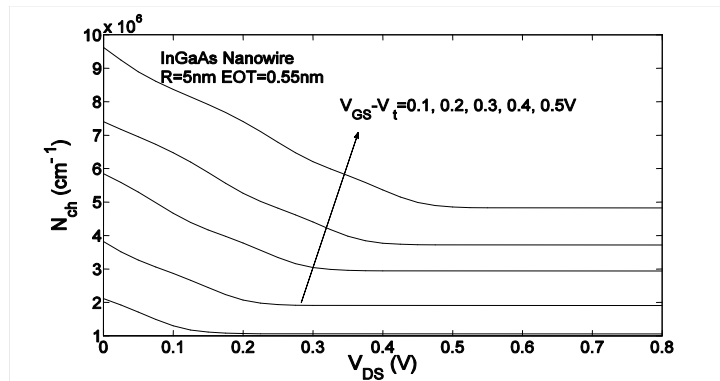


Fig. 6.16. Schematic drawing of nanowire transistor.

Figure 6.17 and 6.18 show a comparison of charge density and ballistic output I-V characteristics of a Si nanowire FET and an InGaAs nanowire FET of comparable dimensions, respectively. Two features are observed here: InGaAs MOSFET delivers higher saturation current, whereas the Si MOSFET exhibits smaller turn-on resistance at low  $V_{DS}$ .

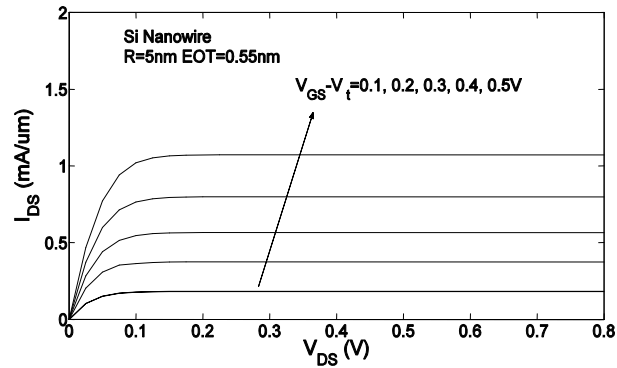


(a)

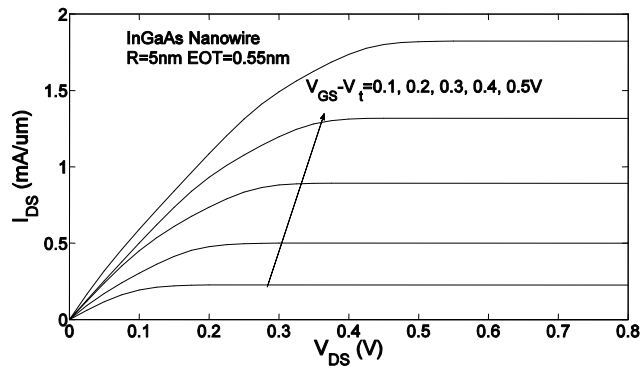


(b)

Fig. 6.17. Computed ballistic line charge density as function of drain bias for (a) Si nanowire and (b) InGaAs nanowire.



(a)



(b)

Fig. 6.18. Computed ballistic output characteristic as function of drain bias for (a) Si nanowire and (b) InGaAs nanowire. The drain Current is normalized to the circumference of the nanowire core.

To understand the higher current in the III-V NWMOSFET, the average carrier velocity is shown in Fig. 6.19, respectively. It is intuitive that with low density of states (low mass and degeneracy), less charge exists in III-V NWMOSFET. For the structures considered here, at representative bias conditions of  $V_{DS}=0.8\text{ V}$  and  $V_{GS}=V_t+0.5\text{ V}$ , the charge density in the Si device is  $\sim 3\text{X}$  that of the InGaAs device. The detailed ratio is dependent on the oxide capacitance. However, due to the same reason, the Fermi level within the III-V nanowire follows closely the

gate voltage, driving the core material into heavy degeneracy, which induces a sharp increase on average carrier velocity. As shown in Fig. 4, although the ratio of average velocity between InGaAs and Si starts at a factor of  $\sim 2.5$  ( $\sim \sqrt{m_{Si}^*/m_{InGaAs}^*}$ ) under non-degenerate conditions, this factor increases to  $\sim 8X$  as the device is driven into heavy degeneracy. For this last case, the increment in carrier velocity outweighs the disadvantage of less charge, thus yielding a higher saturation current for III-V NWMOSEFET. A steplike behavior can also be noted in the InGaAs velocity curve, which is due to the discretized eigen energy ladder.

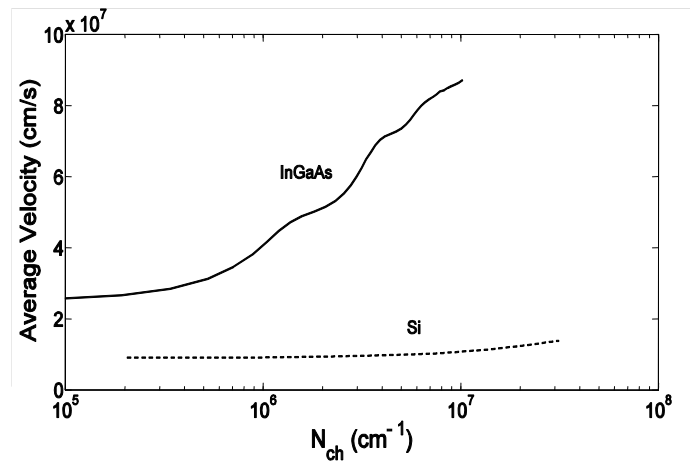


Fig. 6.19. Average injection velocities of Si and InGaAs at the virtual source.

It must be noted that the treatments for both Si and InGaAs devices are based on an assumed parabolic band structure. For III-V materials with strong non-parabolicity (such as InGaAs), this approximation is only strictly valid in a relatively low bias range ( $V_{GS} - V_t \leq 0.5V$ ). With non-parabolicity, the computation task of solving Schrödinger's equation quickly becomes very complex due to the anisotropic band nonparabolicity in the nanowire geometry.

The lower turn on resistance exhibited by the computed characteristics of the above Si device is a result of higher DOS. This result has been calculated for the ballistic limit. For

practical devices, when scattering is present, the turn on resistance will also be influenced by the low field mobility of the material. The advantage in turn-on resistance of the Si device may diminish as a result of its lower mobility than in InGaAs.

To compare our simplified model to experimental values, we computed the ballistic current limits for the Si Twin Nanowire FET structure reported in [20]. As shown in Fig. 17, with  $t_{\text{ox}}=3\text{nm}$ , the experimental nanowire FETs reach a current level of about 75% of their estimated ballistic limit at very small dimensions ( $R=3\text{nm}$ ), while the fraction changes to  $\sim 50\%$  as the nanowire radius increases to 9nm. The figure also shows projected current levels if the oxide thickness can be further scaled down. The discrepancies between the ballistic model and experimental results may arise from parasitic resistance and finite mobility within the nanowires. These factors, which are strongly dependent on device structure, material choice and fabrication process, in general will modify the device behavior and reduce the current from its ballistic limit.

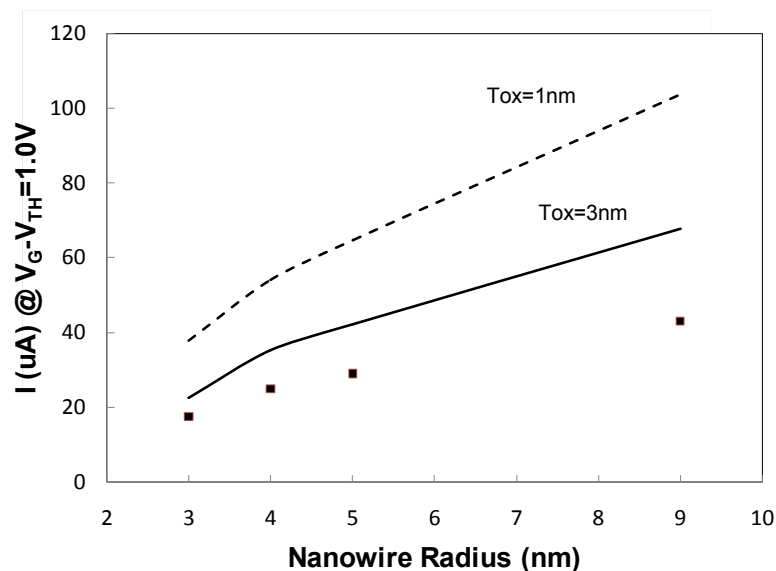


Fig. 6.20. Comparison between ballistic model and experimental result for Si Twin Nanowire FET reported in [20]. Symbols are experimental results, lines are computed results with various oxide thickness.



### 6.4.3 Performance Comparison with Different EOT

In this subsection, we explore the role of EOT on the performance. To illustrate the physics, we use single-subband approximation and low temperature limit. It can be shown that the forward going line charge density of a core/shell nanowire structure, at T=0K, can be written as:

$$N = \frac{g_d}{2\pi\hbar} \int_{E_1}^{E_{fs}} \sqrt{\frac{2m^*}{E - E_1}} dE \quad (6.24)$$

where  $g_d$  is the valley degeneracy. The  $C_{DOS}$  term can therefore be simplified into a form resembling the 1-D DOS:

$$C_{DOS} = (q^2 g_d / \pi\hbar) \sqrt{m^* / 2(E_{fs} - E_1)} \quad (6.25)$$

$C_{qw}$  is related to the width of the wave function, and is generally slowly varying with gate bias [16]. Therefore,  $C_{qw}$  and  $C_{ox}$  may be combined to form an effective oxide capacitance  $C_{ox}'$ .

At T=0K, from Eq. (6.21), the ballistic current can be written as

$$I = g_d q (E_{fs} - E_1) / \pi\hbar. \quad (6.26)$$

where  $(E_f - E_1)$  can be solved via the capacitance network through the relation

$$\frac{1}{q} d(E_{fs} - E_1) = d(V_G - V_{th}) \cdot \frac{C_{ox}'}{C_{ox}' + C_{DOS}} \quad (6.27)$$

The ballistic current can thus be expressed as:

$$I = \frac{q^2 g_d}{\pi\hbar} [(V_G - V_{th}) - \frac{m^* q^3 g_d^2}{\pi^2 \hbar^2 C_{ox}'^2} \left( \sqrt{1 + \frac{2\hbar^2 \pi^2 C_{ox}'^2 (V_G - V_{th})}{q^3 g_d^2 m^*}} - 1 \right)] \quad (6.28)$$

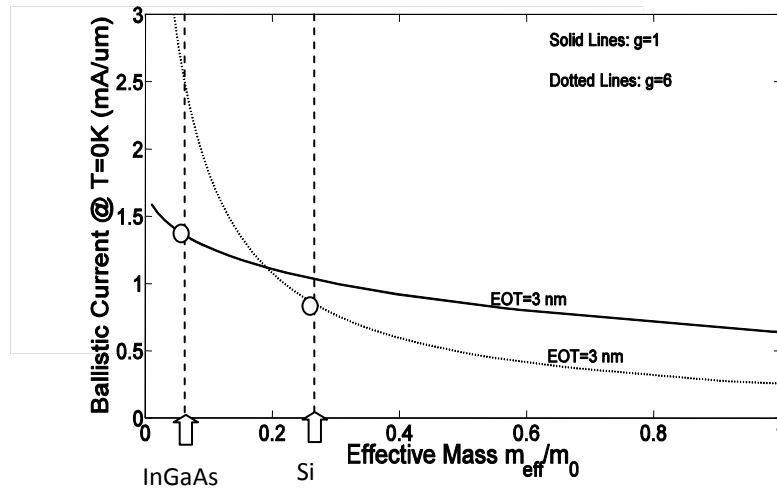


Fig. 6.21. Computed ballistic saturation current at  $T=0K$  with one subband only, with gate overdrive of  $0.5V$ . The EOT considered here includes both oxide thickness and finite wave function depth. Solid curves are for single valley situation, dotted curves are for valley degeneracy of 6. Current is normalized to circumference of the nanowire core.  $m_{\text{eff}}/m_0$  values appropriate to InGaAs and Si are shown with symbol (o).

Plotted in Fig. 6.21 is the ballistic current expressed by Eq. (6.28) with various effective mass and degeneracy combinations. In general, the ballistic current decreases with effective mass with a given EOT value. Interestingly, depending on the actual effective oxide thickness, the ballistic current may or may not benefit from high effective mass and high valley degeneracy. This crossover behavior can be understood by considering extreme conditions:

- (1)  $C_{ox} \gg C_{DOS}$ : the current approaches a constant value of  $I = (g_d q^2 / \pi \hbar) (V_G - V_{th})$ , which is proportional to the degeneracy factor. This asymptotic behavior implies that the current becomes (almost) independent of EOT at low effective mass.
- (2)  $C_{ox} \ll C_{DOS}$ : by expanding the square root term in Eq. (6.28) to the second order, we obtain another asymptotic expression,  $I = \pi \hbar C_{ox}^{-1/2} (V_G - V_{th})^2 / (2m * g_d q)$ . This indicates that the current decreases with increased effective mass and valley degeneracy. This effect shows up at the high effective mass and with high EOT value (see EOT=3nm curve,

where dotted line representing material with higher valley degeneracy fall below the solid line representing lower valley degeneracy at high effective mass).

These two opposite asymptotic trends thus predict a crossover between low and high valley degeneracy as the effective mass increases. The position where the crossover occurs is dependent on the relative magnitude of the (effective) oxide capacitance and the DOS capacitance. Therefore, high mass or high valley degeneracy may have a positive or negative effect on the device performance in the ballistic limit. The same trend is confirmed with the full numerical simulation that takes into account finite temperature and multiple subbands, as plotted in Fig. 6.22. It is worthwhile to notice that for  $EOT=0.55\text{nm}$ , InGaAs (symbol star) exhibits higher ballistic current than Si (symbol cross).

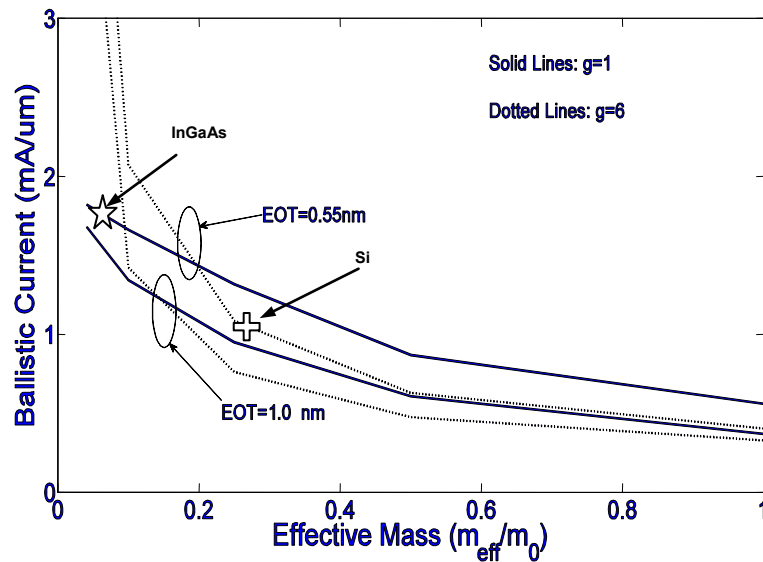


Figure 6.21 Numerical simulation of ballistic saturation current at room temperature with gate overdrive of 0.5V. The EOT considered here includes ONLY the oxide thickness. Solid curves are for single valley situation, dotted curves are for valley degeneracy of 6. Current is normalized to circumference of nanowire core.

#### **6.4.4 Summary**

In conclusion, ballistic III-V NWMOSFETs may outperform Si NWMOSFET in terms of saturation current depending on the available equivalent insulator thickness. Large mass and greater valley degeneracy always result in a smaller turn on resistance at low drain bias in the ballistic limit. Valley degeneracy may have positive or negative influence on saturation current depending on the EOT.

The work in this section is published in the paper titled: “Performance Comparison of Scaled III-V and Si Ballistic Nanowire MOSFETs”, International Journal of High Speed Electronics and Systems, Mar. 2009

## 6.5 Summary

In this chapter, we explore on the device design aspect of the III-V nanowire (core-shell) MOSFET in gate-all-around configuration. We developed a self consistent Schrödinger-Poisson solver for cylindrical nanowire core-shell structure. It was found that the density of state significantly reduces the gate input capacitance for nanowire MOSFET. The self consistent solution is further utilized extensively in latter computation in the chapter. A parallel formulation of ballistic transport model was developed for the nanowire transistor. Performance comparison between InGaAs nanowire MOSFET and Si nanowire MOSFET with the otherwise same structure is conducted. It was found that the thickness of EOT is of great importance to determine which material may give better performance. For very thin EOT, Si may outperform InGaAs nanowire MOSFET; while for practical EOT, InGaAs is better in terms of ballistic limit performance. Section 6.2, in full, is a reprint of "A numerical Schrödinger-Poisson solver for radially symmetric nanowire core-shell structures", Lingquan Wang, Deli Wang and Peter M. Asbeck, Solid State Electronics, Nov. 2006, pp 1732-1739. The dissertation author was the primary author of the paper.

Section 6.3, is part of the paper "Scaling of Nanowire Transistors (invited paper)", Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, Yuan Taur, IEEE Transaction of Electron Devices, Nov. 2008, pp.2846-2858. The dissertation author was a co-author of this paper.

Section 6.4, is a rearrangement of the paper "Performance Comparison of Scaled III-V and Si Nanowire MOSFET", Lingquan Wang, Bo Yu, Peter Asbeck, Yuan Taur, Mark Rodwell,

International Journal of High Speed Electronics and Systems, Mar. 2009, pp.15-22. The dissertation author was the primary author.

## Reference:

1. D. Hisamoto et. al., "FinFET-A Self-Aligned Double-Gate MOSFET Scalable to 20 nm", *IEEE Transaction on Electron Devices*, Vol. 47, No. 12, 2000, pp. 2320-2326.
2. B. S. Doyle, S. Datta, M. Dookzy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully depleted tri-gate CMOS transistors," *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 263–265, Apr. 2003.
3. SIA Roadmap
4. Huang Y, Lieber CM. "Integrated nanoscale electronics and optoelectronics: exploring nanoscale science and technology through semiconductor nanowires." *Pure Appl Chem*, 2004, vol. 76, pp. 2051–68.
5. Hahm J, Lieber CM. "Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors", *Nano Lett* 2004, vol 4, pp. 51–4.
6. Gu Y, Kwak E-S, Lensch JL, Allen JE, Odom TW, Lauhon LJ. "Nearfield scanning photocurrent microscopy of a nanowire photodetector", *Appl Phys Lett*, 2005, vol.87, 043111.
7. Duan X, Huang Y, Agarwal R, Lieber CM. "Single-nanowire electrically driven lasers". *Nature* 2003, vol.421, pp.241–6.
8. Cui Y, Zhong Z, Wang D, Wang WU, Lieber CM. "High performance silicon nanowire field effect transistors". *Nano Lett* 2003, vol.3, pp.149–52. Bryllert T, Samuelson L, Jensen L, Wernersson L. "Vertical high mobility wrap-gated InAs nanowire transistor". DRC proceedings, 2005, p. 157–8.
9. Tan I-H, Snider GL, Chang LD, Hu EL. "A self-consistent solution of Schro"dingere–Poisson equations using a nonuniform mesh". *J Appl Phys* 1990, vol.68(8), pp.4071–6.
10. [://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html](http://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html).
11. Sze SM. *Physics of semiconductor devices*. 2nd ed. John Wiley & Sons; 1981 [Appendix G].
12. Bruhach J, Silov AYu, Haverkort JEM, Vleuten Wvd, Wolter JH. "Coupling of ultrathin InAs layers as a tool for band-offset determination", *Phys Rev B* 1999, vol.59(15), pp.10315–26.
13. Lopez-Villanueva JA, Cartujo-Cassinello P, Gamiz F, Banqueri J, Palma AJ. "Effects of the inversion-layer centroid on the performance of double-gate MOSFETs". *IEEE Trans Electron Dev* 2000, vol.47(1), pp.141–6.
14. Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, Yuan Taur, "Scaling of Nanowire Transistor", published in *IEEE Transaction of Electron Device*, Nov., 2008.

15. Lingquan Wang, Deli Wang, Peter Asbeck, “A numerical Schrödinger-Poisson solver for radially symmetric nanowire core-shell structures” , Solid State Electronics, vol. 50, pp 1732-1739
16. S. Tyagi, et.al, “An advanced low power, high performance, strained channel 65nm technology” , IEDM 2005
17. Kenji Natori, “Ballistic metal-oxide-semiconductor field effect transistor”, Journal of Applied Physics, vol. 76, pp. 4879
18. A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, “Theory of ballistic nanotransistors,” IEEE Trans. Electron Devices, vol. 50, no. 9, pp. 1853–1864, Sep. 2003.
19. S. Suk, et. al, “Investigation on nanowire size dependency on TSNWFET”, IEDM 2007



## **PART III. III-V Tunneling MOSFET**

# Chapter 7. Tunneling FET Based on III-V Staggered Heterojunction for End of the Roadmap Application

## 7.1 Introduction

One of most fundamental problems facing by the current device scaling is the power consumption, which consists of active and stand-by power. As pointed out in [1], as device scaling proceeds into the sub-100nm feature size regime, the stand-by power becomes a significant portion of the overall power consumption, while providing no benefit to any functionality of the IC. Over the history of device scaling, the stand-by power has increased much more rapidly than the active power consumption. If this trend continues, the author of [1] further pointed out that the stand-by power will catch up with the active power when the devices are scaled down to 10~20nm gate length.

The power consumption of a single device further influences the overall power consumption of an IC chip. Although various circuit techniques (such as multi-threshold device strategy, power management module, multi-core architecture) have been implemented, the ever increasing power consumption (for example, in a CPU) has become an extremely important concern, not to mention the global notion to go “green”. It is therefore reasonable to retrospect and contemplate the possible innovations one can make to the fundamental device design to address this power challenge. Shown in Fig. 7.1 is a compilation of power consumption of representative CPU products from Intel over the past two decades, with relevant information (such as power supply, technology node) supplied with the graph. The technical specifications were taken from the product manuals from Intel website [2].

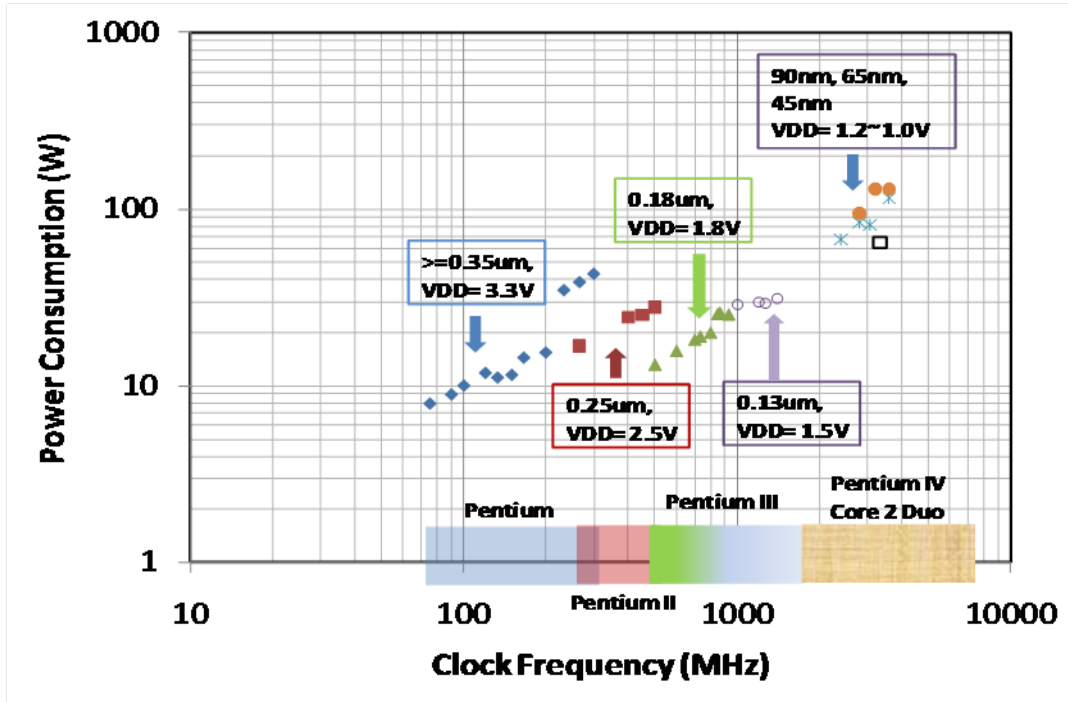


Fig. 7.1. Compilation of power consumptions of representative Intel CPU's as a function of operating clock frequency, over a range of technology nodes and product generations.

It is seen that down to  $0.18\mu\text{m}$  technology node, the supply voltage was scaled almost proportionally to the feature size of a given technology. Accompanied with the voltage scaling, we also observe a sharp decrease of the power consumption although the operation frequency has been steadily increased. However, from  $130\text{nm}$  technology node onward, the supply voltage scaling apparently lags the feature size scaling. As a consequence, the device scaling did not specifically benefit the CPU in terms of its power consumption. An increasing trend in power consumption is observed among CPU of more recent generations as operational frequency goes up.

It is clear that the demand of higher performance and the requirement to conserve power is closing up the design space of a MOSFET. From the performance standpoint, in order to effectively drive the non-scaling interconnect and parasitic, the device must be capable to provide

large enough current drive, which calls for large enough gate overdrive ( $V_{GS}-V_t$ ). On the other hand, threshold voltage must be high enough to prevent excessive leakage, since the best subthreshold roll-off under the current MOSFET structure is limited by temperature: 60mV/dec, i.e.  $\ln(10)\frac{kT}{q}/dec$ . The concurrent demand for high enough  $V_t$  and high enough gate overdrive has prevented the supply voltage being scaled down. To circumvent this dilemma, it is desired to have device that can be turned on more rapid than 60mV/dec.

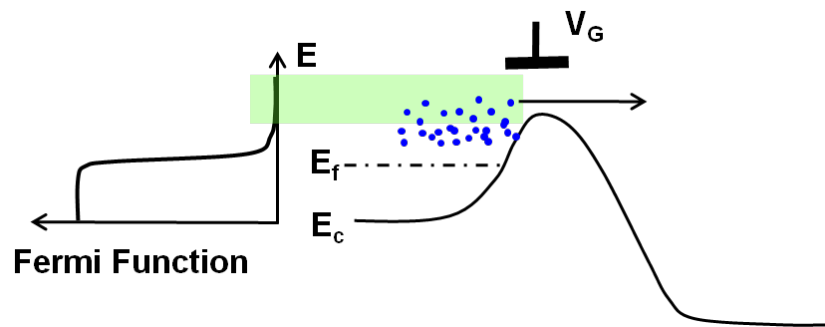


Fig. 7.2. Schematic drawing of the thermionic conduction mechanism which results in the  $>60\text{mV/dec}$  turn-on characteristics in conventional MOSFETs.

It is instructive to review why the conventional MOSFET is limited by 60mV/dec (more quantitative discussion will be shown later). As shown in Fig. 7.2, for a conventional MOSFET operated in subthreshold regime, the electrons that can surmount the source to channel barrier are distributed on the tail of the Fermi-Dirac function on the energy scale. In this case, Fermi-Dirac distribution may well be expressed in form of Maxwell-Boltzmann distribution, therefore we expect  $\exp(-E/kT)$  turn on pattern, which unfortunately ties the turn-on characteristics to the ambient temperature. In other words, it is the thermionic emission mode of the conduction that prevails in conventional MOSFET that determines the 60mV/dec under room temperature.

To break this technical barrier, it is necessary to investigate into conduction mechanisms other than thermionic emission. Tunneling is known to have minor temperature dependence, which makes using tunneling as the primary conduction mechanism from source to channel an attractive option for the novel transistor. It has been demonstrated previously, in various material systems such as Carbon Nanotube [3], Silicon [4] and SiGe [5], the sub-60mV/dec turn on characteristics. However, it must also be realized that the same device must also be capable to deliver high current when fully turned on, which is yet demonstrated by these previous experiment results. In this chapter, we will analyze the necessary criteria for a MOSFET based on tunneling to deliver high on-state current. We will propose possible device structures that may be suitable for this purpose.

This chapter will be arranged in the following sequence: section 7.2 will be focused on the basic device physics of a tunneling FET, possible physical realization, material choice, etc; section 7.3 will further explore the device design space of a TFET based on III-V staggered heterojunction; section 7.4 will summarize the chapter.

## 7.2 Tunneling Field Effect Transistors (TFET) Based on Staggered Heterojunctions for Ultra-low Power Applications

### 7.2.1 Introduction

As MOSFET devices become aggressively scaled, the increasing off-state leakage has become one of the important concerns along with the on-state current drive. The non-scaling property of the off-state current in a conventional MOSFET device stems from thermionic injection at the source channel junction. The resulting minimum 60mV/dec subthreshold slope (SS) fundamentally limits the reduction of off-state current [6]. To overcome this barrier, novel transistor structures incorporating a tunneling junction between source and channel have been proposed [7] and experimentally demonstrated [4-5] to exhibit sub-60mV/dec subthreshold slope. In this section, we discuss possible implementations for TFET and discuss the probable choice to enable both rapid turn-on and high on-state current. We will later focus on TFET the option of III-V heterojunction with staggered band lineup for the source channel junction so as to enable ultra-low voltage operation without compromising leakage. The flexibility of hetero-integration of diverse III-V materials allows a good balance to be found between power consumption and performance. For the representative device design shown in this letter, indicated by simulations, the Tunneling Field Effect Transistor (TFET) is capable of working with 0.3V voltage supply with  $\sim 5 \times 10^{-5}$  mA/ $\mu\text{m}$  off-state leakage and  $\sim 0.4$  mA/ $\mu\text{m}$  on-state current drive, as a result of extremely steep turn-on characteristics enabled by the tunneling injection mechanism. This section will be arranged in the following sequence: concept of TFET and its material choice will be first discussed, followed by discussion on possible device structures and their implementations. Simulation results from 2D device simulator Sentaurus [8] will then be presented and discussed.

## 7.2.2 Concept of a TFET and Material Choices

### 7.2.2.1 Concept of a TFET

The TFET utilizes a tunneling junction (reverse biased p-n junction) between the source and channel. A schematic band diagram is shown in Fig. 7.3 to explain the basic device operation. The band diagram in Fig. 7.3(a) is a representative situation under “off” condition, where the tunneling from the source valence band to the channel conduction band is prohibited. As gate bias increase (Fig. 7.3(b)), the conduction band position on the channel side is pulled down, making conduction band states available for electrons from the source valence band to tunnel into. The TFET thus turns on.

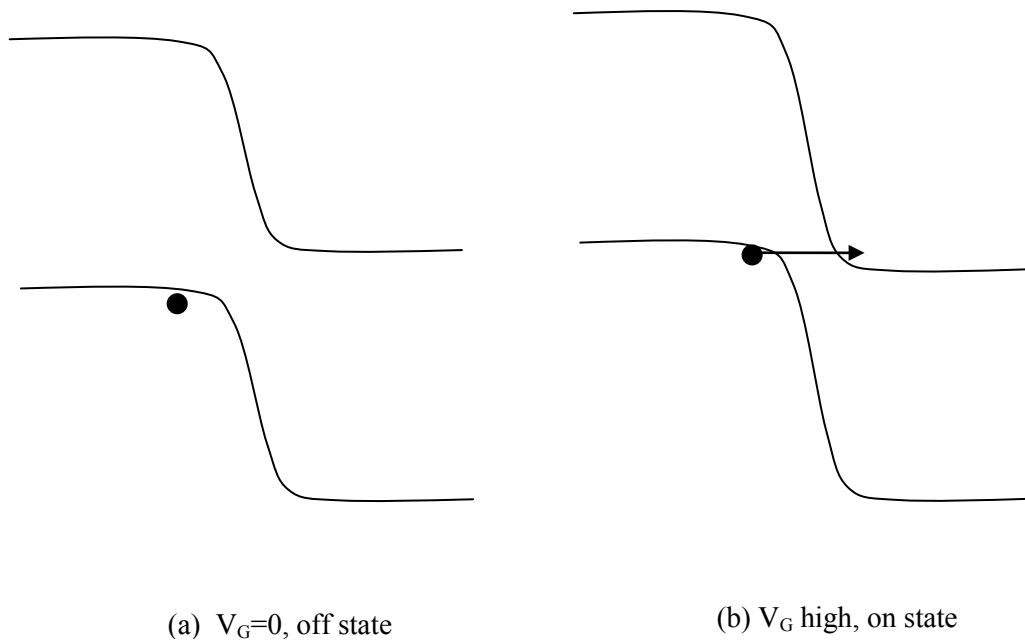


Fig. 7.3. Schematic band diagram for (a) off state and (b) on state for the source/channel junction.

The tunneling mechanism that we utilize here is the so-called Band-to-Band (BTB) tunneling, which not only provides the rapid turn-on characteristics in a TFET, but also determines how much current the transistor will be able to deliver when the transistor is turned on. Therefore, it is appropriate to analysis the tunneling process and understand how the material choice affects the tunneling probability as well as current drive.

### 7.2.2.2 BTB Tunneling Process

BTB tunneling process has been previously studied extensively in the literature [9] for homojunctions. It has been suggested that the tunneling current density for reverse biased homojunction may be expressed by the following expression:

$$J = a \cdot V_{eff} \cdot F \cdot \exp\left(-\frac{b}{F}\right) \quad (7.1)$$

where  $V_{eff}$  is the junction bias,  $F$  is the electric field, and constant  $a$  and  $b$  are defined as:

$$a = \frac{Aq^3}{4\pi^2\hbar^2} \sqrt{\frac{2m^*}{E_g}}$$

$$b = \frac{4\sqrt{m^*}E_g^{3/2}}{3q\hbar} \quad (7.2)$$

where  $m^*$  is the effective mass, and  $E_g$  is the bandgap of the material. The influences of various material parameters may be derived from these equations:

- 1) The energy bandgap of the material: tunneling current will decrease exponentially as the bandgap of the material increases, as seen from the “ $b$ ” term in Eq. (7.2). This may limit the potential for wider bandgap material such as Si to deliver high on-state current. This simple observation also explains why only  $\sim 10 \mu\text{A}/\mu\text{m}$  on-state current was observed in a



recently reported Si-based tunneling FET [5]. For a modern VLSI device, it is however desired to have an on-state current on the order of  $1\text{mA}/\mu\text{m}$ , which is  $\sim 100\text{X}$  greater than the reported current level in a Si TFET.

- 2) The effective mass: the presence of effective mass in the exponential also suggests that larger effective mass may not be preferable for high current drive purpose.

Although Eq. (7.1) gives a reasonably good description of how various material parameters affect the overall tunneling current, it is less clear on several physical perspectives: 1) the tunneling starts off from a valence band electron (which possesses the hole effective mass), however it finishes as a conduction band electron (which possesses the electron effective mass). It is therefore questionable whether an electron effective mass or a hole effective mass should be used to characterize the tunneling. 2) In Eq. (7.2), the details of the tunneling barrier that is faced by the electron during the tunneling process are not reflected. A connection between the barrier and material parameters such as energy bandgap is lacking, which makes the physical picture incomplete.

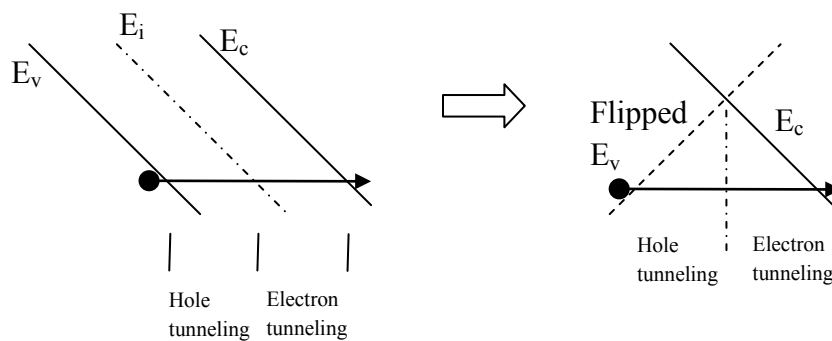


Fig. 7.4. Illustration of the tunneling barrier for hole/electron tunnel model.

An intuitive way to understand the Eq. (7.1) and (7.2) is illustrated in Fig. 7.4. In the following, we consider the situation where in-plane wave vector to be zero ( $k_y=k_z=0$ ). We assume that the first half of entire tunneling distance is accomplished by hole tunneling (backward into the valence band from the mid-point M); while the other half by electron tunneling into the conduction band. Since the positive direction of energy points downwards on the energy band diagram, the first half of the tunneling is equivalent to an electron tunneling through a barrier where the barrier is constructed by flipping the valence band profile with respect to the  $x$  axis. With the aid of the band diagram shown in Fig. 7.3, the imaginary part of the evanescent wave vector within the tunneling barrier may be written as:

$$\kappa = \begin{cases} \frac{\sqrt{2m_v^*}}{\hbar} \sqrt{qFx} & x \in (0, \frac{E_g}{2qF}) \\ \frac{\sqrt{2m_c^*}}{\hbar} \sqrt{\frac{E_g}{2} - qFx} & x \in (\frac{E_g}{2qF}, \frac{E_g}{qF}) \end{cases} \quad (7.3)$$

In spirit of the WKB method [10], the tunneling probability may be expressed as:

$$T = \exp(-2 \int_0^{E_g/(qF)} \kappa dx) = \exp[-\frac{2E_g^{3/2}}{3qF\hbar} (\sqrt{m_v^*} + \sqrt{m_c^*})] \quad (7.4)$$

As the dominant contribution from the valence band tunneling is provided by light hole band, and the light hole effective mass is close to the conduction band effective mass (as can be seen from the  $\mathbf{k}\cdot\mathbf{p}$  method [11]), we may further simplify Eq. (7.4) by assuming  $m^*=m_c^*=m_v^*$ :

$$T = \exp(-\frac{4E_g^{3/2}}{3qF\hbar} \sqrt{m^*}) \quad (7.5)$$

We note that Eq. (7.5) resembles the exponential term in Eq. (7.1), and we develop an understanding that it was derived under the physical picture of hole and electron tunneling half

way of the entire barrier length, respectively. We also made the assumption that light hole mass and electron mass are equal in order to arrive at the same mathematical form.

However, the picture of hole and electron tunneling half across of the barrier, although intuitive, still needs further verification. To develop further understanding and illustrate the essential physics while not involving excessive mathematical procedure, we base the following derivation on the two-band model [11]. By employing the  $\mathbf{k}\cdot\mathbf{p}$  method, the dispersion relationship concerning only two bands (i.e. conduction band and light hole band) may be written as:

$$\frac{\hbar^2 k^2}{2m} = \left(E - \frac{1}{2}E_g + \frac{P^2}{m}\right) - \sqrt{\left(E - \frac{1}{2}E_g + \frac{P^2}{m}\right)^2 + E(E_g - E)} \quad (7.6)$$

where  $m$  is electron mass (not the effective mass);  $E_g$  is the energy Bandgap;  $P$  is defined as:

$$P = \langle u_{0,v} | \hat{p} | u_{0,c} \rangle = \langle u_{0,c} | \hat{p} | u_{0,v} \rangle \quad (7.7)$$

where  $u_{0,v}$  and  $u_{0,c}$  are Bloch functions associated with valence band and conduction band, respectively. The value of the matrix element described in Eq. (7.7) may be extracted from the conduction band or valence band effective masses:

$$\begin{aligned} \frac{m}{m_v^*} &= \left(1 - \frac{2P^2}{mE_g}\right) \\ \frac{m}{m_c^*} &= \left(1 + \frac{2P^2}{mE_g}\right) \end{aligned} \quad (7.8)$$

where  $m_c^*$  and  $m_v^*$  are effective masses for conduction band and valence band, respectively.

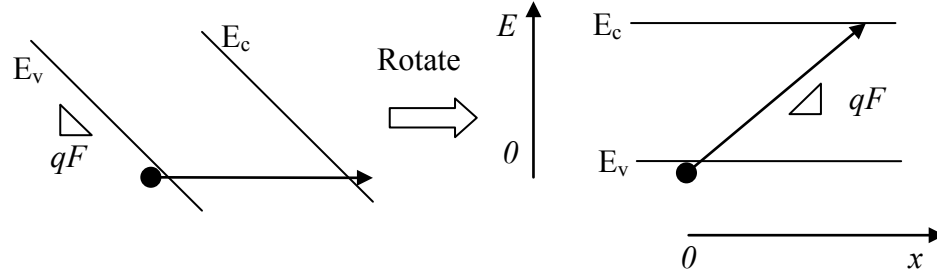


Fig. 7.5. Coordinate setup for the BTB tunneling computation based on two-band model.

In the BTB tunneling problem, we may assume the electric field  $F$  (to avoid confusion with energy notation) is constant within the tunneling distance of interest. We use the coordinate set up illustrated in Fig. 7.5 to assist our analysis. As shown, a valence electron starts off from  $x=0$ , tunneling into the forbidden gap, resulting an imaginary wave vector. The exact value of the wave vector may be calculated by Eq. (7.6). It should be noted that the energy  $E$  in Eq. (7.6) is always referred to valence band top. It is therefore more convenient to rotate the energy band diagram as shown in Fig. 7.5(b). The energy of the electron that undergoes the tunneling process, in the new coordinate system may be expressed by:

$$E = qFx \quad (7.9)$$

The total tunneling distance may be determined by:

$$d_{\text{tunnel}} = E_g / (qF) \quad (7.10)$$

Substitute Eq. (7.9) to Eq. (7.6), we have

$$\kappa = \text{Im}(k) = \frac{\sqrt{2m}}{\hbar} \sqrt{\sqrt{\left(qFx - \frac{1}{2}E_g + \frac{P^2}{m}\right)^2 + qFx(E_g - qFx)} - \left(qFx - \frac{1}{2}E_g + \frac{P^2}{m}\right)} \quad (7.11)$$

The tunneling probability may be obtained by integrating Eq. (7.11) once with respect to  $x$  from 0 to  $d_{tunnel}$ , in accordance with the WKB method.

$$T = \exp\left(-2 \int_0^{d_{tunnel}} \kappa dx\right) \quad (7.12)$$

To benchmark the validity of the intuitive physical picture, we compare the imaginary parts of evanescent wave vectors as functions of position obtained under different assumptions, as shown in Fig. 7.7. Results obtained from Eq. (7.3), (7.11) are included in the plot. The computation uses  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  as an example under the uniform field of 0.1 MV/cm. Si is not chosen as an example due to its indirect bandgap; tunneling between valence band and conduction band will involve momentum change of the carrier, which is not modeled in the preceding treatment. In the two computations, wave vectors derived from Eq. (7.6) are considered to be the most accurate. The material parameters used in the computation are listed in Table 7.1.

Table 7.1. Material parameters for  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  and InAs tunneling computation [12,13]

	$m_c^*$	$m_v^*$	$E_g$	$P$
$\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$	$0.041m_0$	$0.045m_0$	0.75eV	$1.13 \times 10^{-24} \text{ kg} \cdot \text{m}/\text{sec}$
InAs	$0.023m_0$	$0.026m_0$	0.36	$1.06 \times 10^{-24} \text{ kg} \cdot \text{m}/\text{sec}$

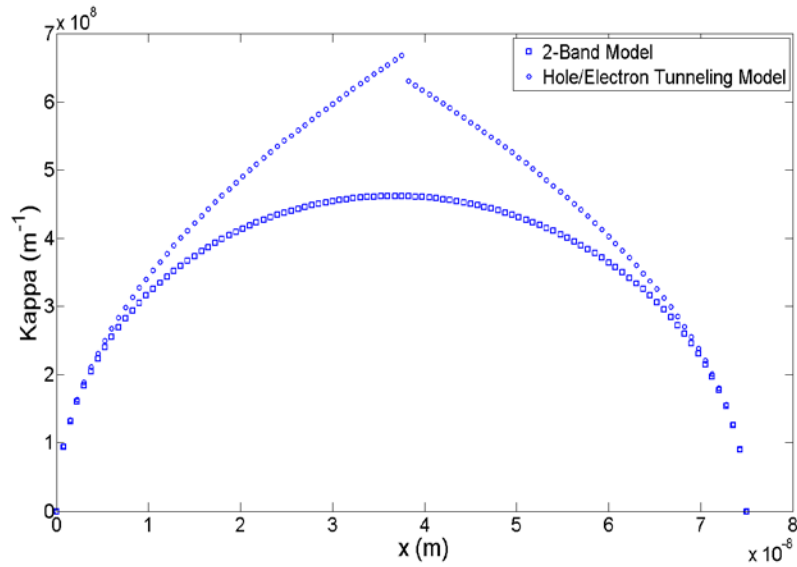


Fig. 7.6. Comparison on evanescent wave vectors using hole/electron tunneling model Eq. (7.3) and two-band model Eq. (7.11).

As shown in Fig. 7.6, significant difference exists between the hole/electron tunneling model and 2-band model, although values close the band edges agree well (in the vicinity of starting point and ending point of the tunneling process). Integrating Fig. 7.6 yields tunneling probability: the tunneling probabilities under this particular electric field (0.1MV/cm) are  $4.8 \times 10^{-29}$  and  $2.5 \times 10^{-24}$ . To further compare the difference between the two models, we plot in Fig. 7.7 the tunneling probabilities as a function of electric field  $F$ . Again, we observe significant difference between the two model predictions.

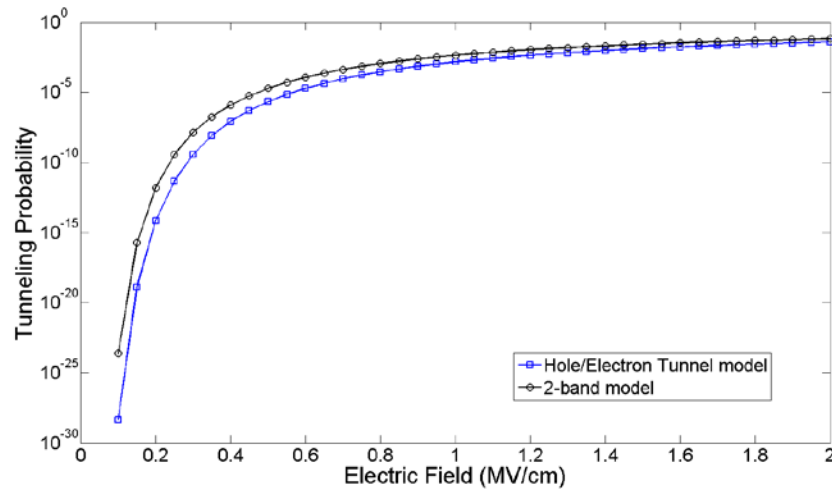


Fig. 7.7. Tunneling probability obtained from hole/electron tunneling model and 2-band model as a function of electric field.

It is also seen from Fig. 7.7 that significant tunneling does not occur until very high electric field is imposed across the tunneling junction. As the total tunneling distance may be determined as  $E_g/qF$ , the tunneling probability is therefore intuitively dependent on the energy bandgap of the material. In Fig. 7.8, we compare the tunneling probability as a function of electric field for  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  and  $\text{InAs}$ , to illustrate the impact of the energy band gap to the tunneling probability (both computations use the 2-band model). This specific computation may be used towards the appropriate choice of the material for TFET tunneling junction.

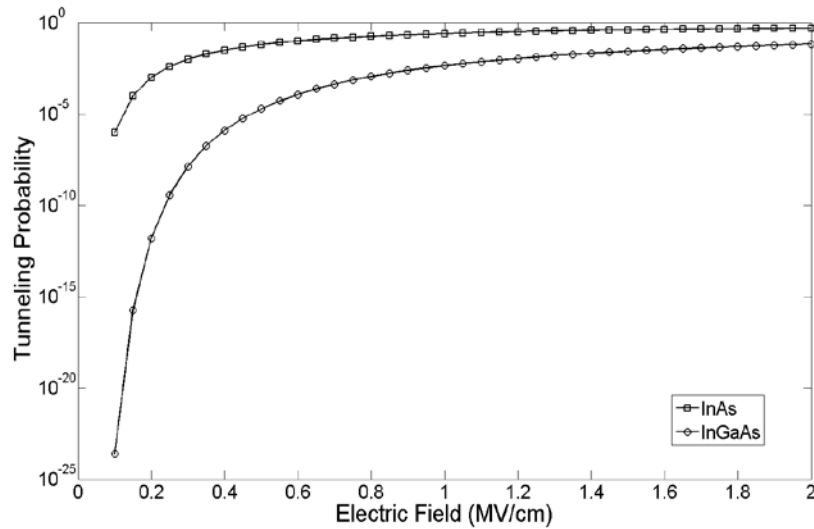


Fig. 7.8. BTB tunneling probability versus electric field for  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  and InAs materials.

### 7.2.2.3 Material Consideration

One of the TFET design objectives is to deliver high current when the device is turned on. To achieve this goal, significant BTB tunneling must occur, which requires large tunneling probability. As seen from Fig. 7.8, it is necessary to have either high electric field or lower bandgap to tunneling through. The high electric field may be achieved by having high doping concentration at the junction, while the lower bandgap may be available by choosing the appropriate material (e.g. low bandgap material like InAs suits for this purpose).

The other TFET design objective is to have low off-state leakage. It may be seen from Fig. 7.3(a) that the BTB tunneling is effectively forbidden at the off-state. However, we must take into account other possible leakage mechanisms that may contribute to the overall off-state leakage. As discussed above, low bandgap material is preferred as large tunneling probability may be achieved at low electric field, however, it is also to be noted that low bandgap material



may also be affected by larger thermal generation rate. Once an electron-hole pair is generated, the hole will be swept to the source (by the E-field at the source/channel junction) while the electron will be further collected by the drain: the generation current thus contributes to the overall leakage. For SRH generation, if we assume the defects from which the electron-hole pair is generated are located at the mid-gap, we may attribute the dependence of generation rate on the energy band gap as:

$$G \propto \exp\left(-\frac{E_g}{2kT}\right) \quad (7.13)$$

The exponential dependence on the energy bandgap implies rapidly increasing generation current as the choice of material involves smaller bandgap. This clearly poses a conflict to the notion of using narrow bandgap material to enhance the tunneling probability. More discussion will be provided in regard to this design tradeoff in the latter sections.

### ***7.2.3 Device Structure and Device Physics of TFETs***

In this subsection, we further discuss possible implementations of a TFET, as well as the rationale behind the choice of a heterojunction over a homojunction for the source/channel tunnel junction.

#### ***7.2.3.1 Device Design Considerations for Tunneling Junction of a TFET***

The source/channel junction for a TFET, as mentioned previously, is replaced by a reverse biased p-n junction (instead of a forward bias p-n junction). There are, in general, two choices in terms of material configurations for the junction – homojunction and heterojunction.

In case of a homojunction, as mentioned previously, we may not be able to achieve significant current flow (on the order of  $\sim 1\text{mA}/\mu\text{m}$ ) while complying with small supply voltage (which implies lower electric field) using a wide bandgap material. On the other hand, if narrow bandgap material were used for the source/channel, it would induce large thermal generation currents that would dominate the leakage level of a given design. The high leakage were shown in simulations by Datta et.al [7], and poses a general concern of having narrow bandgap material for the source/channel junction. Given these considerations, the design space is very small for a homojunction TFET to achieve high on-state current and low off-state leakage at the same time.

A more careful examination may be done of the tunneling probability for the tunneling junction. In the tunneling process, the barrier that the valence electron from the source sees is closely related to the separation of the channel conduction band and source valence band, in the vicinity of the junction. For future reference, we label this quantity as  $E_{sep}$ . In a homojunction implementation, this value happens to be the energy bandgap of the material  $E_{sep}=E_g$ . On the other hand, as far as the concern on the thermal generation current goes, it depends on the energy bandgap of the materials for source ( $E_{gsrc}$ ), channel ( $E_{gch}$ ) and drain ( $E_{gdrn}$ ). Again, for a homojunction implementation, these values coincides with each other ( $E_{gsrc} = E_{gch} = E_{gdrn} = E_g$ ). However, for a heterojunction embodiment, it is possible to have different values for  $E_{sep}$ ,  $E_{gsrc}$ ,  $E_{gch}$  and  $E_{gdrn}$ , and thus resolve the conflict we discussed earlier between the tunneling probability (on-state current) and thermal generation (off-state current). An effective choice of this type is a staggered heterojunction, in which  $E_{sep}$  may be selected to be relatively small (to give high tunneling current) while others may be larger to suppress the thermal generation.

A representative device implementation using a heterojunction is shown Fig. 7.9, based on InGaAs/AlGaSb material system. Band lineups of selected materials of relevance are depicted in Fig. 7.10 with a schematic energy band diagram shown in a same figure. It is to be noted that

given the complexity of layers, it is necessary to discuss effects from stress and strain, which will be dealt in the next subsection. The band diagram shown Fig. 7.10 is based on relaxed material parameters. The purpose of Fig. 7.10 is to show the flexibility of the heterojunction, which gives greater design space by allowing for a smaller  $E_{sep}$  (large tunneling current) and relatively larger  $E_{gsrc}$ ,  $E_{gch}$  and  $E_{gdrn}$  (low leakage current). We will further discuss the role of stress and strain in the context of the device design of a TFET in the following subsection.

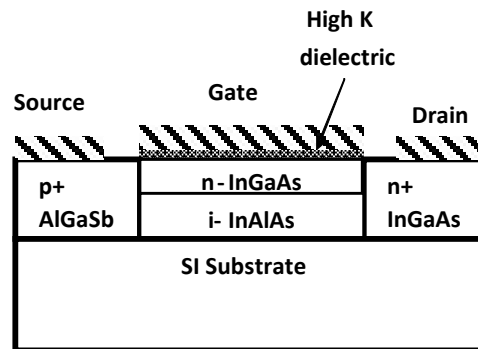


Fig. 7.9. Schematic drawing of TFET device with planar embodiment (not to scale).

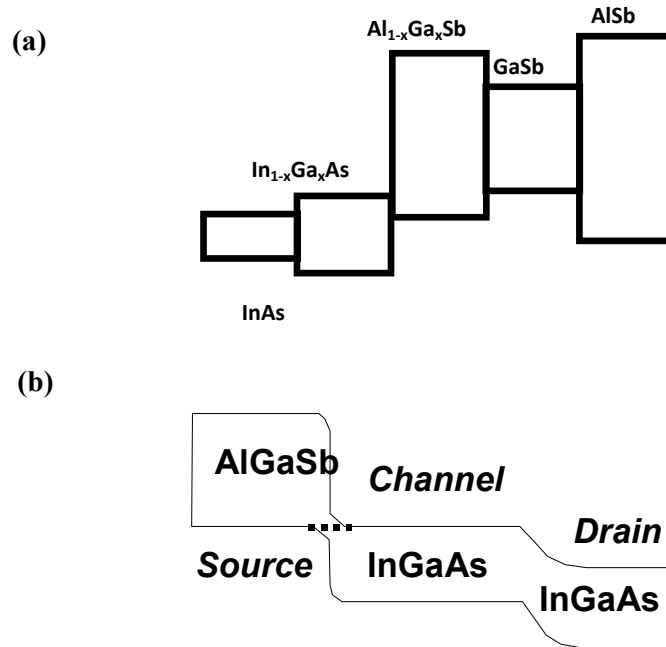


Fig. 7.10. (a) Approximate band line-up for selected semiconductors, (b) Schematic energy band diagram along the channel direction at on-state.

### 7.2.3.2 Material Choices and Possible Implementation of a TFET

In practice, due to the lattice mismatch among different materials, stress and strain will be introduced to the coherently stressed layers. The stress and strain due to material and lattice mismatch have impacts on: 1) the energy band gap 2) the relative band lineup. We would like to point out here that the material variations occur not only along the growth direction, but also along the current transport direction (referred as longitudinal direction below). Therefore, strain and stress effects take place in both directions.

Along the growth direction, the layer may be grown in a coherently strained fashion, if materials with mismatched lattice constants were to be deposited sequentially. To prevent dislocations, the layer thicknesses must be kept below the corresponding critical thicknesses. As

the channel layer is preferably thin for SCE control purposes, the criterion for critical thicknesses may be satisfied. However, along the longitudinal direction, the strain effects are more complicated:

- 1) The channel region and source region are individually coherently strained to the (virtual) substrate, along the growth direction. However, when the segments are brought into contact along the longitudinal direction, the mismatch in lattice constant at the source/channel interface creates another constraint on the source length, if coherently strained source is desired.
- 2) It must be noticed that the lattice constants for considering the source/channel interface lattice mismatch are not the relaxed lattice constants. Rather, the lattice constants along the vertical edge is related to the Poisson ratio of the material given the strain create at the interface with the substrate. As the source length is typically on the order of  $\sim 100\text{nm}$ , to avoid dislocations being generated at this length scale becomes more challenging.

Given these technical difficulties of having strained layers, it is beneficial to have lattice matched design for the planar structure (For modeling purposes, the material parameters are better known for lattice relaxed material than for strained material as well). A possible design is shown in Fig. 7.11 with the relevant material properties tabulated in Table 7.2. The relative band lineup is illustrated in Fig. 7.12. The doping profile incorporated in this baseline design is specified as: Source region: p-type uniform  $2 \times 10^{19} \text{ cm}^{-3}$ ; Channel region: undoped; Drain region:  $5 \times 10^{19} \text{ cm}^{-3}$ . The gate length simulated in this design is  $100\text{nm}$ . The thickness of the channel layer is  $10\text{nm}$ . The choice for the doping level for individual regions will be the subject of latter sections, where design tradeoffs associated with doping will be presented.

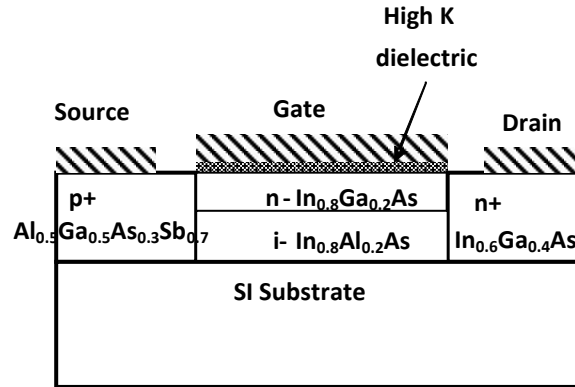


Fig. 7.11. Schematic drawing of TFET device with planar embodiment (not to scale) and matched lattice constants.

Table 7.2. Material parameters for lattice matched TFET design [12,13]

	$m_c^*$	$m_v^*$	$E_g$	Lattice Constant ( $\text{\AA}$ )	Electron Affinity (V)
$\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}_{0.3}\text{Sb}_{0.7}$	$0.09m_0$	$0.11m_0$	1.38eV	5.98	3.60
$\text{In}_{0.8}\text{Ga}_{0.2}\text{As}$	$0.034m_0$	$0.036m_0$	0.50eV	5.98	4.73
$\text{In}_{0.8}\text{Al}_{0.2}\text{As}$	$0.051m_0$	$0.057m_0$	0.77eV	5.98	4.51

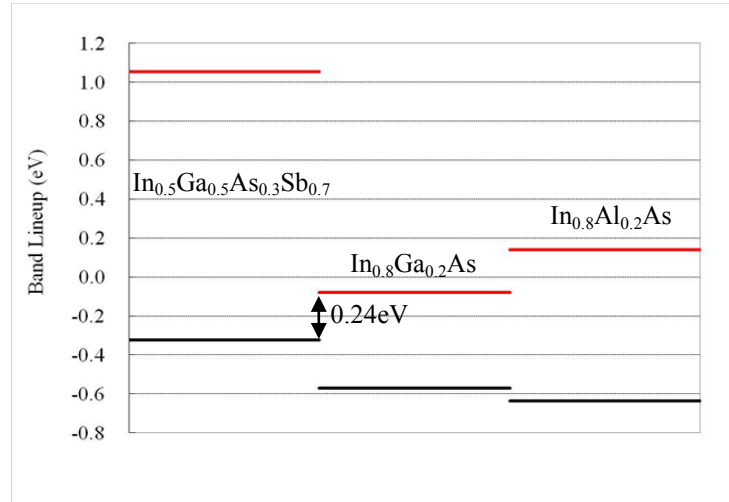


Fig. 7.12. Band lineup of the materials in the TFET design shown in Fig. 7.11.

It is also instructive to discuss the possible implementation of a TFET to make the design more practical. Given the structure provided in Fig. 7.11, the following technological challenge must be addressed:

- 1) Formation of the Source/Drain region, since they consist of materials that are presumably different from the channel material. Layer structures in the III-V arena are typically produced by MBE or MOCVD, giving the layer thickness accuracy to the order of a monolayer ( $2\sim 3\text{\AA}$ ). Using these techniques, it is conventional to have material variation along the growth direction, while on the other hand, less standard to have material variation along directions that are perpendicular to the growth direction. The recently reported selective regrowth technique [14] makes the implementation possible. In the reported work, different material was deposited as source and drain in a III-V MOSFET after the channel are patterned.
- 2) The growth of a quaternary with mixed group V ( $\text{InGaAsSb}$  source) is challenging. However, successful growth has been demonstrated previously in literature. The

same principle may be applied to the regrowth process to allow for the source formation.

- 3) Inclusion of the high-k dielectric. This issue has been discussed in Chapter 4 and will not be repeated here.

### 7.2.3.3 Device Physics of aTFET

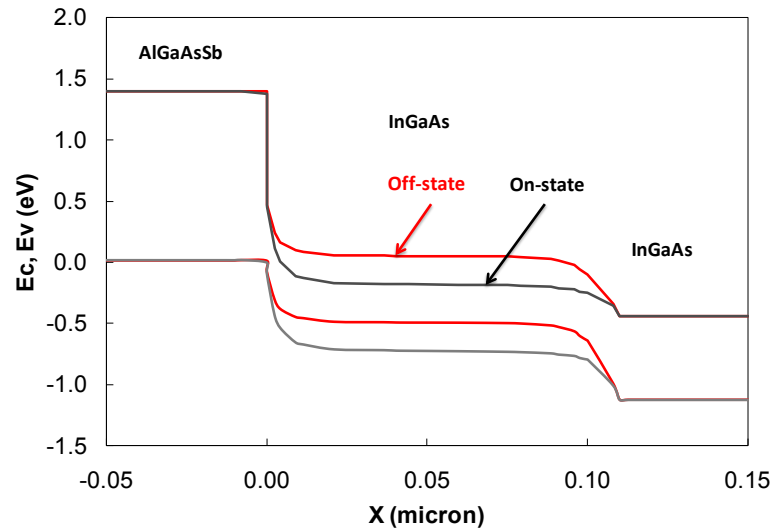


Fig. 7.13. Representative energy band diagram under off-state (zero gate bias) and on-state (0.3V gate bias), both with 0.3V drain bias.

The energy band diagrams of the TFET (design shown in Fig. 7.11) under different bias conditions are shown in Fig. 7.13, under zero gate bias (off state) and high gate bias (on state), both with high drain bias. The tunneling process has been discussed in Sec. 7.2.2.2 and will not be repeated here. It is noted that both heavy holes and light holes are involved in this process; however, the tunneling probabilities for light holes are much greater and comprise the primary tunneling current. At zero gate bias, as shown, the tunneling process is prohibited due to lack of available final states in the conduction band of the channel region, at the corresponding energy. As gate bias increases, the conduction band of the channel is pulled down which reduces the energy barrier as well as the tunneling distance. Significant BTB tunneling current thus occurs at



the source-channel junction and provides the primary current flow.

The same material system may be utilized to implement p-type tunneling FET by swapping the InGaAs and GaAlAsSb material. The device operates in a similar fashion to the n-type TFET.

To characterize the current, we employ the Landauer's formalism [15]: the tunneling current at the source channel tunneling junction can be expressed as:

$$J = q \int v_x(k_x, k_{||}) \bullet g_{sd}(k_{||}) \bullet T(k_x, k_{||}) \bullet [f_s(\vec{k}, E_{fs}) - f_d(\vec{k}, E_{fd})] \bullet d\vec{k} \quad (7.14)$$

where  $E_{fs}$ ,  $E_{fd}$  are quasi-Fermi levels at tunneling source and destination, respectively;  $f_s(\mathbf{k}, E_{fs})$ ,  $f_d(\mathbf{k}, E_{fd})$  are Fermi-Dirac distribution functions;  $g_{sd}(k_{||})$  is the joint density of state function;  $v_x(\mathbf{k})$  is the forward going velocity of the carriers;  $T(\mathbf{k})$  is the transmission coefficient (tunneling probability in this case). The integration is carried out over the energy range where source valence band and channel conduction band overlap. We may use Eq. (7.14) to further quantify how the TFET differs from a conventional MOSFET and provides steeper turn on characteristics than 60mV/dec.

In conventional MOSFETs that rely on thermionic injection,  $T(E)$  becomes almost unity for carriers with sufficient kinetic energy to overcome the barrier at the source-channel junction, whereas quickly drop to zero if the kinetic energy falls short. As discussed previously, the steepness of the turn-on characteristic is thus dominated by the availability of high energy carriers within the source, which corresponds to the tail of the Fermi-Dirac distribution with 60mV/dec slope at room temperature. However, in the TFET case (e.g. n-channel TFET), the steepness of the turn-on is controlled by the tunneling probability. As shown in Eq. (7.11) and (7.12), the tunneling probability depends on material properties such as effective mass and energy band gap, which depend weakly on temperature. Therefore the turn-on steepness is freed from the 60mV/dec thermal limit. The on-set of the current flow is controlled by two factors: 1)

Availability of the states on the channel side, which depends on the band alignment between the source valence band and channel conduction band. This factor may be modulated by the gate bias.

2) Tunneling probability when an empty final state is available. Factor 1) and 2) are typically intertwined. For instance, as depicted in Fig. 7.13, at off-state, if a valence electron were to traverse the barrier, it must tunnel very long distance until it find an empty state with appropriate energy to tunnel into. This suppression of the tunneling process in this scenario may be attributed to both factor 1 (lack of appropriate final states close by) and 2 (large tunneling distance).

Shown in Fig. 7.14 are the representative simulated transfer characteristics and output characteristics of the baseline design of device. The device is biased with a supply voltage of 0.3V, as it is designed for ultra-low operation voltage applications. A metal work function of 4.6eV is assumed which provides the appropriate threshold voltage. The simulation is carried out with commercial TCAD software Sentaurus. The Sentaurus modeling parameters were validated by analysis of the two-band BTB direct tunneling process, using the physical parameters listed in Table 7.2. Sub-60mV/dec SS is exhibited over a range of several decades of current, giving the device very sharp turn-on characteristics. For reference purposes, best possible FET turn-on characteristics based on conventional thermionic injection mechanism (60mV/dec) are also shown in the figure. With the use of a gate dielectric of 0.6nm EOT, this device provides an on-state current drive of  $\sim 0.4\text{mA}/\mu\text{m}$ , at an off-state current of  $\sim 5 \times 10^{-5}\text{mA}/\mu\text{m}$ , and therefore achieves an on-off current ratio  $\sim 10^4$ , over a small voltage swing of 0.3V. It is also noted that for the output characteristic, simulation predicts a super-linear I-V behavior in the triode regime, which may be understood as following: the initial slow turn-on is due to the cancelation of tunneling current traveling along opposite direction. As drain bias increases, the forward tunneling current becomes dominant and results in more rapid increase in the drain current.

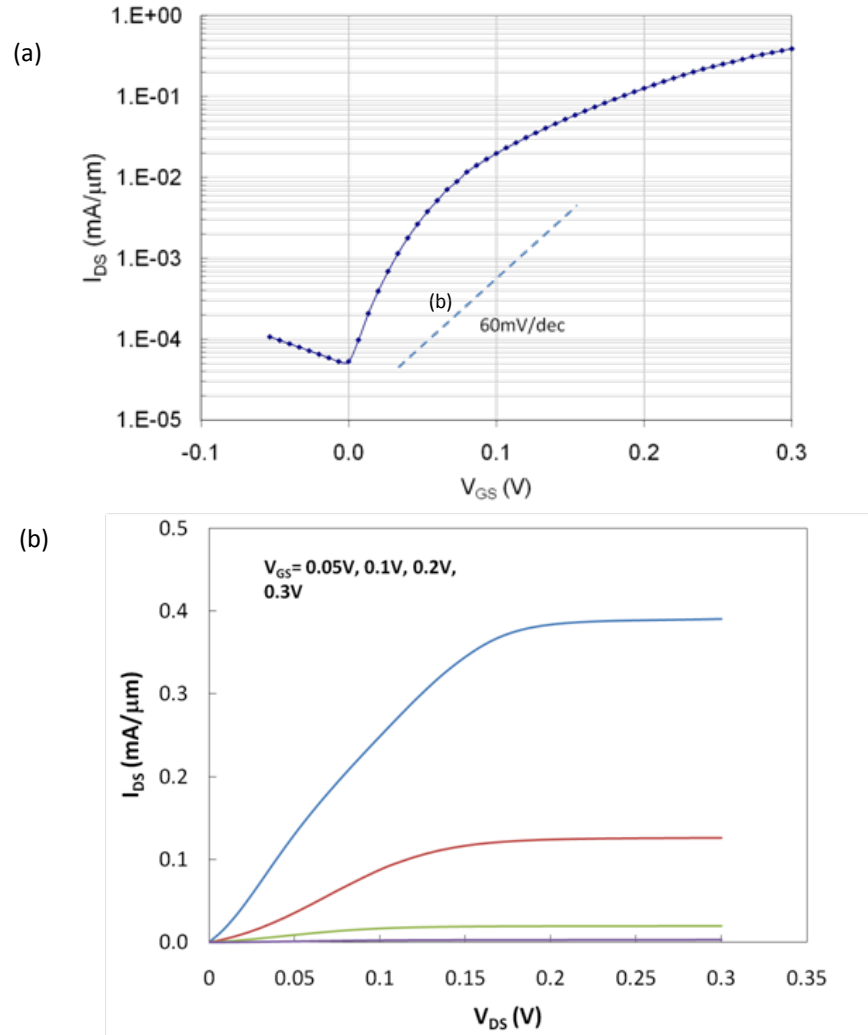


Fig. 7.14. (a) Transfer characteristic and (b) output characteristic of a TFET, with  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}_{0.3}\text{Sb}_{0.7}$  p+ source and  $\text{In}_{0.8}\text{Ga}_{0.23}\text{As}$  n- channel.

## 7.2.4 Summary

In this section, we lay down the necessary physics for designing a tunneling FET based on staggered heterojunctions. A representative design of this kind is suitable for ultra-low power supply applications. For an example provided, with 0.3V supply voltage, the device is predicted to be capable of delivering  $0.4\text{mA}/\mu\text{m}$  current at on-state with an off-state current of  $\sim 5 \times 10^{-5}$  mA/ $\mu\text{m}$ . Analysis suggests that staggered heterojunction embodiment of the TFET may offer a

better design compromise in terms of performance (on-state current) and power consumption (off-state leakage) than its homojunction TFET counterpart. We will discuss further design tradeoffs of a TFET based on staggered heterojunction in the following section.

## 7.3 Design Considerations for Tunneling Field Effect Transistors (TFET) Based on Staggered Heterojunctions

### 7.3.1 Introduction

For a TFET, a successful design must be capable of delivering high on-state current over a limited voltage swing. This not only implies achieving sub-60mV/dec subthreshold swing, but also to maintain the steep turn-on characteristics over several decade of drain current range. In this chapter, we seek to address these problems from a device design perspective, and illustrate how device design parameters may affect the overall characteristics. For this purpose, we utilize the device structure that was shown in Fig. 7.9, and neglect the effects of strain and stress between mismatched lattices. The same principle may be applied to other designs such as the design discussed in section 5.2.3.

### 7.3.2 Baseline Device Design

In this section, we will refer the baseline design as depicted in Fig. 7.9, which is different from the lattice matched design shown in the previous section. The source consists of p+  $\text{Al}_{0.15}\text{Ga}_{0.85}\text{Sb}$  doped to  $2 \times 10^{19} \text{ cm}^{-3}$ ; channel consists of n-  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  doped at  $10^{15} \text{ cm}^{-3}$ ; the drain consists of n+  $\text{In}_{0.4}\text{Ga}_{0.6}\text{As}$  doped at  $5 \times 10^{19} \text{ cm}^{-3}$ ; there is a graded layer between the channel and drain to provide smooth transition. The gate length simulated in this design is 100nm. The thickness of the channel layer is 10nm with EOT of 2nm. The structure is schematically shown in Fig. 7.10. The purpose of this section is to show the impact of device designs to the overall performance, rather than results due to the actual material properties.

Simulated transfer characteristics of this device are shown in Fig. 7.13, as a reference for further discussion and comparison.

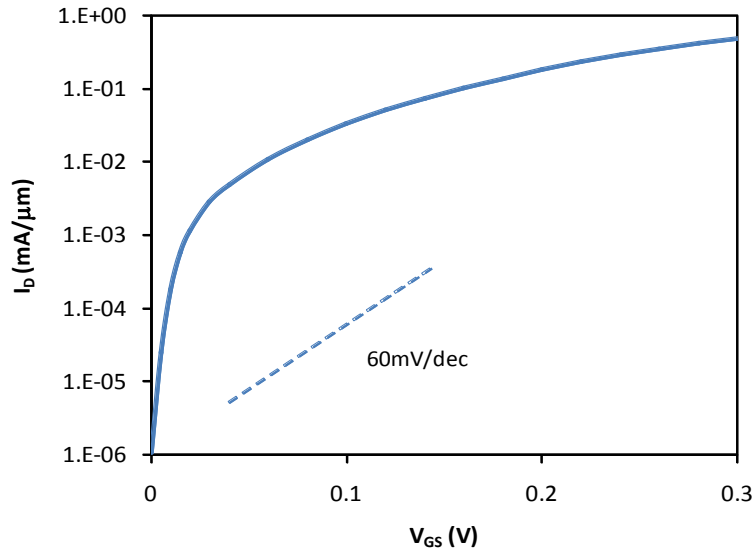


Fig. 7.15. Transfer characteristic of the baseline design TFET, design parameters same as specified in the text.

### 7.3.3 Design Variations and Tradeoffs

A useful design of a staggered heterojunction TFET must preserve a steep subthreshold swing (SS). In the following, we will discuss impact to the device characteristics with various design parameters/strategies.

#### 7.3.3.1 Doping Strategy

As discussed previously, the device turn-on is enabled by modulating the channel conduction band position via capacitive coupling between the gate and the channel. The coupling efficiency thus becomes critical in determining the steepness of the turn-on characteristics, i.e. subthreshold swing. As shown in Fig. 7.16(a) - (c), various coupling conditions may be represented by adjusted capacitance network, where capacitance due to inversion charge is neglected within the subthreshold region. When gate edge is perfectly aligned to the tunneling junction (Fig. 7.16(a)), a fraction  $C_{ox}/(C_{ox}+C_{d_{ch}})$  of the gate bias change is translated into the

change of conduction band position, where  $C_{ox}$  represents the oxide capacitance and  $C_{d_{ch}}$  the depletion capacitance of the channel. Maximization of this efficiency is needed here to minimize the subthreshold slope, which is equivalent to minimizing the depletion capacitance. Therefore a lightly doped channel is desired.

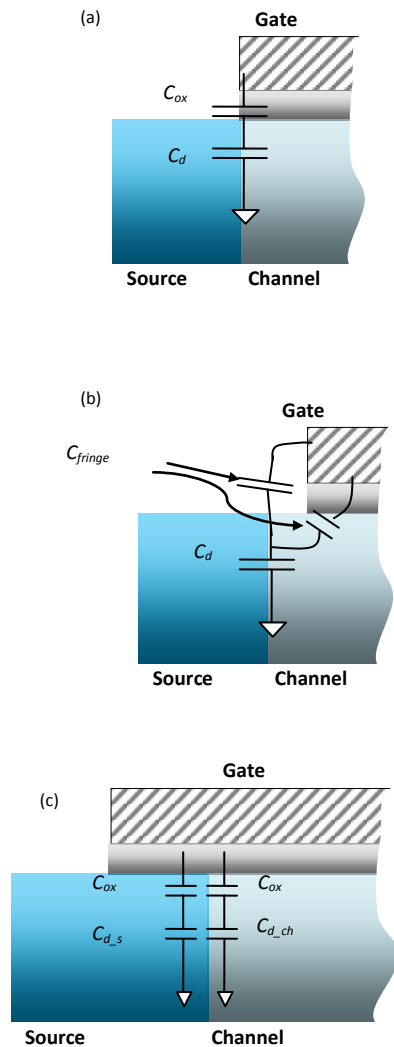


Fig. 7.16. (a) Direct gate to channel coupling with aligned gate, (b) gate to channel coupling via fringing capacitance with gate-source-underlap, (c) gate coupling to both source and channel with gate-source-overlap.

If any underlap region is present, as shown in Fig. 7.16(b), the coupling efficiency is undermined since the gate modulation is achieved through fringing field (effectively reduced  $C_{ox}$ ). Moreover, the maximum achievable current is also harmed by the presence of series resistance from the underlap region. In case of gate overlapping the source, gate modulation will affect both source and channel regions. While the channel conduction band is pulled down by  $C_{ox}/(C_{ox}+C_{d_{ch}}) \square V_G$ , the source valence band adjacent to the tunnel junction is also pulled down by  $C_{ox}/(C_{ox}+C_{d_{S}}) \square V_G$ , where  $C_{d_{S}}$  stands for the depletion capacitance in the p-type source region. The effective gate modulation that affects the tunneling process is the difference between these two expressions, since the tunneling is controlled by the relative band positions between the source valence band and channel conduction band. The overlap is therefore also detrimental to the subthreshold swing. Shown in Fig. 7.17 is a comparison of the transfer characteristics among various gate alignment configurations. It can be seen that accurate gate alignment is critical to preserve sharp turn-on characteristic and to obtain high on-state current drive.

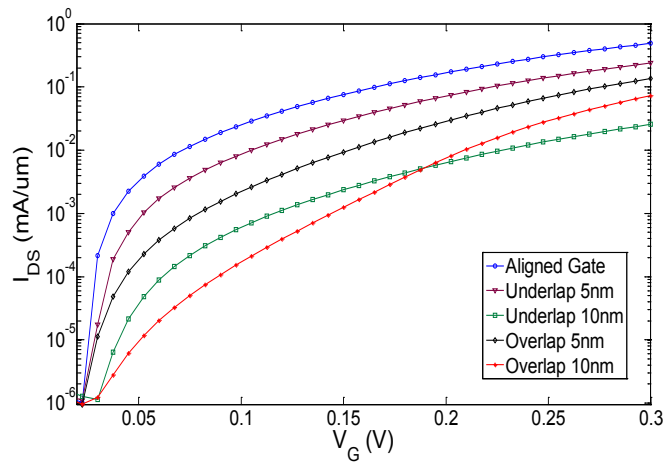


Fig. 7.17. Impact of gate position with respect to the tunneling junction to turn-on characteristic of TFET.



In a conventional MOSFET, source region are almost always doped very heavily, to adequately supply carriers for transport. This common practice is, however, not desired for TFET to achieve steepest SS. As discussed in the previous section, the turn-on is enabled by the tunneling process between occupied valence band states and unoccupied conduction band states. To achieve the best efficiency of turning on (thus the steepest SS), every paired valence band state and conduction band state should be utilized to transport electrons via tunneling. However, very high source doping will drive the device from the optimum condition. This impact may be understood by the band diagram depicted in Fig. 7.18(a). As shown, source Fermi level  $E_{fs}$  lies below the valence band edge  $E_v$ , corresponding to highly degenerate doping. Under finite temperature, valence band states between  $E_v$  and  $E_{fs}$  are partially filled by valence electrons according to the Fermi Dirac distribution. When gate voltage is ramped up, channel conduction band states will first pair up with valence band states whose energy ranges from  $E_v$  to  $E_{fs}$ . However, as these states are partially, only a (small) fraction of paired valence/conduction states are capable to contribute to the tunneling conduction and thus renders slower turn-on. In contrast, when  $E_{fs}$  lies above the valence band, majority of the source valence band states are filled, giving ample of electrons to contribute to conduction when paired up states become available, and henceforth exhibit faster turn-on behavior. A comparison of transfer characteristic with source doping level of  $2 \times 10^{19} \text{ cm}^{-3}$  and  $5 \times 10^{19} \text{ cm}^{-3}$  is shown in Fig. 7.16(b). It can be seen that higher source doping results in slower turn-on as well as higher leakage current floor. Nevertheless, it must be noticed that although lower doping level in the source region is beneficial for subthreshold swing, it is not desirable to introduce excessive series resistance from the lightly doped region. A possible compromise may introduce moderately high doping region only adjacent to the tunnel junction, while having higher doping far from the junction.

Another consideration that needs to be taken into account is the impact of source doping to the electric field at the tunneling junction. As source doping increases, the electric field

intensifies at the tunnel junction, thus provides higher tunneling probability through the tunnel junction. This effect is preferable in terms of high on-state current, however, as discussed previously not desirable for sharp turn-on characteristics. The tradeoff may be gauged with specific application requirement.

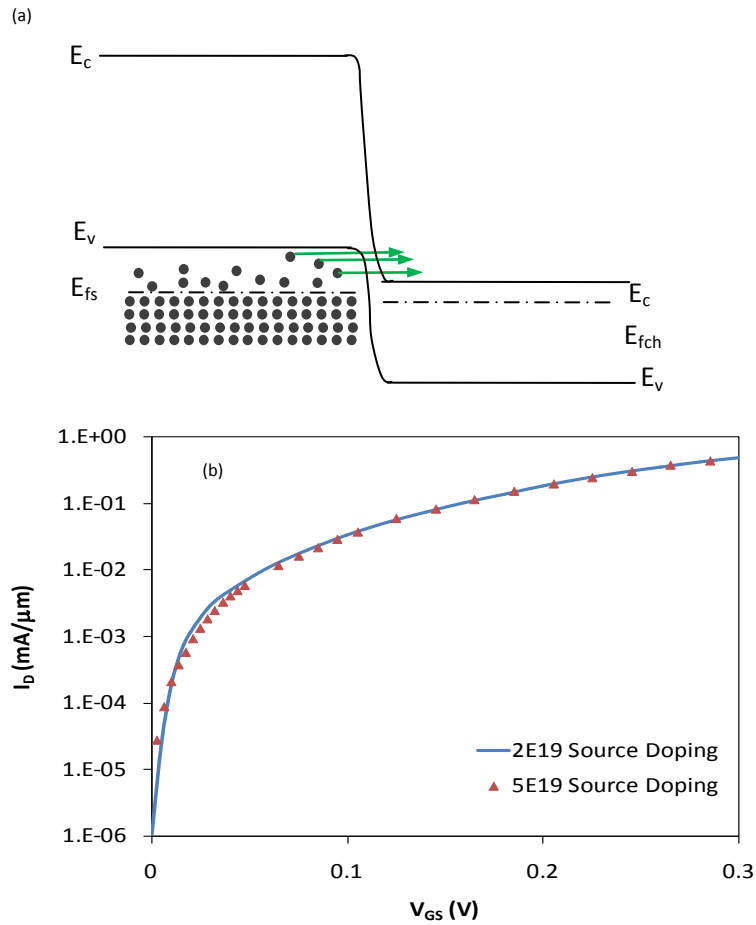


Fig. 7.18. (a) Schematic drawing of inefficient turn-on when source is degenerately doped. Valence electrons with energy above (but close to) the Fermi level will start premature tunneling, therefore smear the subthreshold swing. (b) Transfer characteristic of devices with  $2 \times 10^{19} \text{ cm}^{-3}$  and  $5 \times 10^{19} \text{ cm}^{-3}$  source doping.

### 7.3.3.2 Material Variation

To optimize the balance between off-state and on-state current via material variation, we realize the following constraints for the device design. The first constraint on channel material choice comes from the supply voltage. In order to sufficiently turn on the device, i.e. to allow sufficiently large  $T(E)$  in Eq. (7.14) for the on-state, the energy gap present at the tunneling junction may be at best only a fraction of the supply voltage. This bounds the Indium content for the InGaAs channel on the minimum. For this particular design example incorporating  $\text{Al}_{0.15}\text{Ga}_{0.85}\text{Sb}$  as the source material, the minimum Indium content in the InGaAs channel is found to be  $\sim 50\%$ , which corresponds to an energy gap of  $\sim 0.23\text{eV}$  at the tunnel junction. After the turn-on, as shown in Eq. (7.14), the current is limited the term of joint density of state,  $g_{sd}(E)$ . Similar to the radiative recombination process,  $g_{sd}(E)$  is determined by a reduced effective mass, which is given by  $m_r^{-1} = m_h^{-1} + m_e^{-1}$  assuming parabolic band description [16], where hole mass in the source material and electron mass in the channel material are involved. In this particular design example, the limiting factor of the joint density of states is the electron mass, i.e. the on-state current is limited by available receiving states on the channel side for source valence electrons to tunnel into. The number of available states is affected by (1) the channel conduction band position (the lower it is, the more states are available); and (2) the effective mass of the channel material (higher effective mass leads to higher joint density of states). It must be noted that (1) and (2) may not be independent tuned via material composition. Given the same gate bias, the channel conduction band position may appear at a lower position if the initial energy gap is small, which implies high Indium content; whereas to obtain higher joint density of states, it is desirable to have higher effective mass, which points to less Indium content. This tradeoff is shown in Fig. 7.19. Initial increase of the on-state current as Indium content increases shows the benefit to joint density of states from having a smaller initial energy gap between the source

valence band and channel conduction band. However, this benefit gradually diminishes when the effect of small electron mass of the channel becomes dominant. As a result, the on-state current turns around as the Indium content of the InGaAs channel exceeds 0.7.

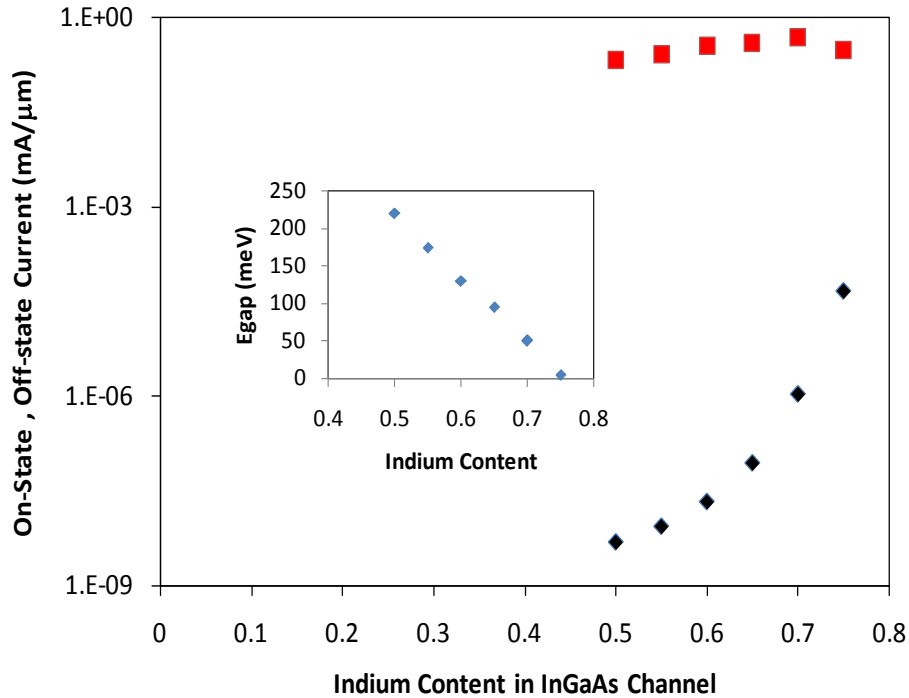


Fig.7.19. On-state (squares) and off-state current (diamonds) for different band line-up conditions (via changing the Indium content within the channel). Inset shows energy gap (between source valence band and channel conduction band) at the tunnelling junction as a function of Indium content.

Another important concern for low power application is the stand-by leakage current. Under zero gate bias, the BTB tunneling is barred by the lack of available receiving states. The most likely leakage current source will be the thermally generated electron-hole pairs at various point of the channel. As the generation rate is proportional to  $\exp(-E_{sep}/nkT)$ , where  $E_{sep}$  may be more appropriately understood as the separation between the conduction band and the valence band. It is obvious that the most vulnerable spot for the thermal generation is at the tunnel

junction, where the smallest separation between conduction band and valence band takes place. Therefore, although a small energy gap (high Indium content) at the tunnel junction may be beneficial for on state current, it may also well lead to large leakage, as shown by the diamond symbols in Fig. 7.19. Given these material considerations, with source being  $\text{Al}_{0.15}\text{Ga}_{0.85}\text{Sb}$ , having Indium content ranging from 50% ~ 70% maybe appropriate to yield reasonably large on-state current without incurring excessive leakage.

### **7.3.4 Conclusion**

In this section, key design considerations of a staggered heterojunction TFET have been discussed. Gate alignment, lightly doped channel and appropriate source doping strategy is found to be critical factors to preserve sharp turn-on characteristic. Channel material choice may introduce tradeoffs between on-state and off-state current, 50%~70% indium content may be appropriate to achieve good on-state current while keeping off-state current acceptable.

The content of this section is published in the proceedings of IEEE “Nanoscale Material and Device Conference”, 2009.

## 7.4 Summary

In this chapter, we discussed a novel type of device that may be turned on much more rapidly than the 60mV/dec thermionic limit in the conventional MOSFET. The device is therefore suitable for ultra-low power supply application and potentially capable to reconcile the conflict between power consumption and performance in a MOSFET design. We further show that an embodiment based on staggered heterojunction provides relaxed design space to balance the demand of high on-state current as well as low off-state leakage.

A representative device design is simulated to be able to provide  $\sim 0.4\text{mA}/\mu\text{m}$  over a gate swing of 0.3V, at a leakage current level of  $\sim 5 \times 10^{-5} \text{mA}/\mu\text{m}$ . We further discussed the device design issues and possible optimizations. It is found that among the key device guidelines are: (good) gate alignment, undoped channel, moderately high doped source and application dependent material composition choices.

Section 7.3, is a rearrangement of the paper “Design Considerations for Tunneling MOSFETs Based on Staggered Heterojunctions for Ultra-Low-Power Applications”, Lingquan (Dennis) Wang, Peter Asbeck, Nanoscale Material and Device Conference, 2009. The dissertation author was the primary author.

## Reference:

1. Wilfried Haensch, et. al., IBM Journal of Research and Development, 4/5, 2006
2. [://www.intel.com](http://www.intel.com)
3. J. Appenzeller, Y-M Lin, J. Knoch, Z. Chen, and P Avouris, "Comparing Carbon Nanotube Transistors---The Ideal Choice: A Novel Tunneling Device Design," IEEE Trans. Electr. Dev., vol. 52, pp. 2568-2576, Nov. 2005.
4. O.M. Nayfeh, C. Ni Chleirigh, J.L. Hoyt and D.A. Antoniadis, "Measurement of Enhanced Gate-Controlled Band-To-Band Tunneling in Highly Strained Silicon-Germanium Diodes," IEEE Electron Device Letters, vol. 29, pp. 468-470, May 2008.
5. Woo Young Choi, Byung-Gook Park, Jong Duk Lee, Tsu-Jae King Liu, "Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec," IEEE Electron Device Letters, vol. 28, pp. 743-745, August 2007.
6. Yuan Taur, Tak Ning, Fundamentals of Modern VLSI Devices. Cambridge University Press, 1998, ch. 4.
7. Suman Datta, "Sub-quarter Volt Supply Voltage III-V Tunnel Transistors for Green Nanoelectronics," as discussed in IEEE Semiconductor Interface Specialist Symposium 2008.
8. Synopsis Sentaurus Manual, Synopsis, 2008
9. S. Sze Physics of Semiconductor Devices, 2nd Edition, John Wiley & Sons, 1981
10. See for example, E. F. Schubert, "Quantum Mechanism Applied to Semiconductor Devices", [://www.ecse.rpi.edu/~schubert/Course-ECSE-6968%20Quantum%20mechanics/Ch00%20Prelims.pdf](http://www.ecse.rpi.edu/~schubert/Course-ECSE-6968%20Quantum%20mechanics/Ch00%20Prelims.pdf)
11. J. Ziman, "Principles of the Theory of Solids", Cambridge University Press; 2<sup>nd</sup> edition
12. M. Levinshtein, S. Rumyantsev, M. Shur edited, Handbook Series on Semiconductor Parameters, vol. 2. , World Scientific, 1999.
13. O. Madelung edited, Semiconductors – Basic Data. 2nd Edition, Springer, 1996
14. Uttam Singiseti et. al. , "In<sub>0.53</sub>Ga<sub>0.47</sub>As channel MOSFETs with self-aligned InAs source/drain formed by MEE regrowth", IEEE Electronic Device Letter, Nov. 2009
15. See for example, Mark Lundstrom, Fundamentals of Carrier Transport, 2nd Edition, Cambridge University Press, 2000
16. See for example, L. Coldren, S. Corzine, "Diode Lasers and Photonic Integrated Circuits", Johns Wiley & Sons, 1995

## Chapter 8. Summary of the Thesis

In this dissertation, design considerations on III-V MOSFET of various types (thermionic and tunneling) were presented, and are summarized below:

For planar III-V MOSFETs:

- 1) Small DOS due to the small effective mass of III-V semiconductors has been a great concern of the layer structure, in that it limits the overall sheet charge density that may be achieved upon given gate bias. Although significant portion of the sheet charge density may be recovered due to the non-parabolicity in the conduction band, the effect of the small DOS may still cause a degradation of the overall gate capacitance from the oxide capacitance to be as large as 30%.
- 2) Small effective mass also results in more pronounced quantization effects, which push the charge distribution centroid farther from the semiconductor/dielectric interface. The degradation of the overall gate capacitance due to this effect was also evaluated through self-consistent Schrödinger-Poisson solutions with non-parabolic conduction band taken into account.
- 3) In ballistic limit, a representative III-V (InGaAs) MOSFET may be slightly better than an otherwise same Si MOSFET, in terms of on-state current. However, the high mobility of the III-V materials enables a III-V MOSFET to behave more closely to its ballistic limit (as a result of less scattering and lower series resistance). Therefore, in quasi-ballistic limit, it is estimated that an InGaAs MOSFET may significantly outperform a Si MOSFET.
- 4) Small DOS of III-V materials also results in a unique problem for III-V MOSFET – source starvation. To circumvent this adverse effect, it is necessary to design the source geometry appropriately (wide enough channel and source) and dope the source heavily enough ( $5 \times 10^{19} \text{ cm}^{-3}$ , which is much more than common practice for III-V HFETs).



- 5) For non-ideal high-k dielectric on III-V semiconductors, we provided a comprehensive analysis on effects of the border traps within the dielectric. It is found that the typical frequency dispersion in different gate bias region may be due to different types of traps. In depletion/weak accumulation region, the dispersion is likely due to interface traps and border traps, while the dispersion in the strong accumulation may be explained by border traps. Furthermore, the dispersion amount of capacitance in the strong accumulation region is predicted to be inversely proportional to  $\ln(1/f)$  by the model, and verified by the experimental results.

For III-V nanowire MOSFET, a parallel analysis was presented in terms of the impact of the further reduced DOS (due to 2D confinement) and performance projection.

- 1) It is found that the small DOS significantly reduces the gate input capacitance for nanowire MOSFET, to ~30% of the capacitance defined by the oxide capacitance.
- 2) The thickness of EOT is of great importance to determine which material may give better performance. For very thin EOT, Si may outperform InGaAs nanowire MOSFET; while for practical EOT, InGaAs is better in terms of ballistic limit performance.

For III-V tunneling MOSFET

- 1) We analyzed the BTB tunneling process and compared the different possible design embodiments based on homojunction and heterojunction. It is found that the staggered heterojunction design provides relaxed design space to balance the demand of high on-state current as well as low off-state leakage, and is simulated to be capable of provide 0.4mA/ $\mu\text{m}$  on-state current over 0.3V gate bias, at an off-state current of  $5 \times 10^{-5}$  mA/ $\mu\text{m}$ .
- 2) We discussed the device tradeoffs of a III-V tunneling FET, it is found that the device guidelines are: (good) gate alignment, undoped channel, moderately high doped source

and application dependent material composition choices to reflect the tradeoff between on-state and off-state current requirements.