

PREDICTIVE ELECTRO AND THERMAL QUANTUM TRANSPORT IN  
NANOSCALE DEVICES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yuanchen Chu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2019

Purdue University

West Lafayette, Indiana

**THE PURDUE UNIVERSITY GRADUATE SCHOOL  
STATEMENT OF DISSERTATION APPROVAL**

Dr. Gerhard Klimeck, co-chair

School of Electrical and Computer Engineering

Dr. Tillmann Kubis, co-chair

School of Electrical and Computer Engineering

Dr. Mark Lundstrom

School of Electrical and Computer Engineering

Dr. Joerg Appenzeller

School of Electrical and Computer Engineering

**Approved by:**

Dr. Dimitrios Peroulis

Head of the school of Electrical and Computer Engineering

This thesis is dedicated to my parents and my wife, for their selfless love, support and sacrifices.

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Professor Gerhard Klimeck, for accepting me as a direct PhD student in his group about 6 years ago, and for his continuous support during these years. He gave me the opportunity to study and conduct research in the field of nanoelectronics modeling, which is a valuable experience for my future career. I also want to thank him for his support in my internship at MIT, where I got the chance to work with experimentalists and was able to apply my knowledge in quantum transport to some exiting applications.

I am truly grateful to Professor Tillmann Kubis for his mentorship. He shared me with countless knowledge in quantum transport and has always been very supportive whenever I have a question. He provided me with great help in improving the quality of my research and has always been open to discuss about new ideas.

I am thankful to Professor Mark Lundstrom and Prof. Joerg Appenzeller, for their serving as my PhD committee members and for their patience in answering my questions. From their courses (ECE 656 and ECE 612), I have learned a lot about the fundamentals in carrier transport and semiconductor device physics.

I sincerely thank Dr. Mohamed Mohamed at MIT Lincoln lab for providing me a wonderful working opportunity at MIT, and for his continuous financial support in my research after I returned to Purdue. I also appreciate him for many insightful discussions and his help in improving my research works.

I would like to thank Professor Tomás Palacios at MIT for hosting me as a visiting student in his group. He has been very supportive in my research and has provided me with many helpful suggestions to guide my work during my stay at MIT. I also deeply appreciate him for introducing me to the Center for Integrated Quantum Materials (CIQM) held by Harvard University, where I got the chance to present my work and to meet many experts in this field.

I am grateful to Professor Michael Povolotskyi for many insightful discussions and for his support in the code development. He has always been very kind to me and has been very helpful in guiding me to solve various scientific and technical problems to make progress in my research.

I am thankful to Professor Xiulin Ruan and Professor Timothy S. Fisher for their helpful suggestions in my research. And I would like to thank Dr. Jingjing Shi for her efforts in our collaborative project. She has always been very patient in answering my questions. From our discussions, I have learned a lot about thermal transport and molecular dynamics simulations.

I would like to thank Dr. Jim Fonseca who helped me at the beginning of my PhD to settle down in the group. And I would like to thank Dr. Bozidar Novakovic for his support in the implementation of simulation tool on nanoHUB.

I would like to thank Dr. Prasad Sarangapani, Dr. James Charles and Daniel Lemus for their help in NEMO5 code development and valuable discussions I had with them where I learned a lot about various segments of quantum transport.

I thank Dr. Zhengping Jiang, Dr. Kai Miao, Dr. Pengyu Long, Dr. Hesam Ilatikhameneh, Dr. Tarek Ameen, Dr. Yaohua Tan, Dr. Jun Huang and Dr. Yu He for valuable discussions. I thank Dr. Yu Wang, Dr. Fan Chen, Dr. Harshad Sahasrabudhe, Xinchun Guo, Kuangchuan Wang, Dr. Namita Narendra, Dr. Jinying Wang, Dr. Daniel Mejia, Dr. Shangchun Lu, Dr. Daniel Valencia, Hanwei Hsiao, Robert Andrawis and Chinyi Chen for their help and support. I also thank all my dear friends, including Dr. Yuling Hsueh, Dr. Mehdi Salmani Jelodar, Dr. Archana Tankasala, Dr. Junzhe Geng, Evan Wilson, Dr. Gustavo Valencia, Dr. Xufeng Wang, Xin Jin, Yubo Sun, Yenpu Chen and Zhiguang Zhou for their caring and encouragement.

I would like to thank Vicki Johnson, Ashley Byrne, Leslie Schumacher, Hunter Bridge and Joseph Baylon for their help in scheduling meetings and reserving rooms.

Finally, I would like to extend my deepest gratitude to my family. Without their love and support, I would not have been able to complete my PhD.

## TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	ix
LIST OF FIGURES . . . . .	x
ABSTRACT . . . . .	xvi
1 INTRODUCTION . . . . .	1
1.1 Need for quantum transport in nanodevices . . . . .	1
1.2 Missing elements in state-of-the-art quantum transport . . . . .	4
1.2.1 An explicit screening model . . . . .	4
1.2.2 A realistic charge model for TFET . . . . .	5
1.2.3 A sophisticated scattering model for TFET . . . . .	5
1.2.4 An electro-thermal coupled model for self-heating . . . . .	6
1.3 Thermal boundary resistance . . . . .	7
1.4 GaN nanowire nFET . . . . .	8
1.5 Monolayer WTe <sub>2</sub> -ZrS <sub>2</sub> vdW TFET . . . . .	9
1.6 Negative capacitance FET . . . . .	10
1.7 Thesis outline . . . . .	10
2 FULL-BAND APPROACH FOR BAND-TO-BAND-TUNNELING DEVICES	12
2.1 Motivation . . . . .	12
2.2 The standard excess-charge approach . . . . .	13
2.3 The full-band approach . . . . .	14
2.4 Method verification . . . . .	19
2.4.1 Convergence behavior of $n_{eq}$ calculation . . . . .	19
2.4.2 Transferability of the background charge . . . . .	19
2.4.3 Transfer characteristics of silicon MOSFETs . . . . .	20
2.5 Application on TFETs . . . . .	22

	Page
2.5.1 Silicon TFET . . . . .	22
2.5.2 p-GaSb/n-InAs HJTTFET . . . . .	24
2.6 Incompatibility of ECA with NEGF . . . . .	26
2.7 Jacobian matrix for Newton iteration . . . . .	27
2.8 Summary . . . . .	29
<b>3 THERMAL BOUNDARY RESISTANCE PREDICTION WITH NEGF AND MD SIMULATIONS . . . . .</b>	<b>31</b>
3.1 Motivation . . . . .	31
3.2 The benchmark system . . . . .	32
3.3 Simulation approach . . . . .	33
3.3.1 MD . . . . .	33
3.3.2 NEGF . . . . .	34
3.4 Thermal boundary resistance: NEGF versus MD . . . . .	37
3.5 Inelastic scattering in NEGF . . . . .	38
3.6 Role of phonon scattering in energy current across interface . . . . .	40
3.7 Summary . . . . .	40
<b>4 5 nm GATE LENGTH GaN NANOWIRE nFET . . . . .</b>	<b>42</b>
4.1 Motivation . . . . .	42
4.2 Simulation approach . . . . .	42
4.3 Results and discussions . . . . .	43
4.3.1 Device structures . . . . .	43
4.3.2 Benchmarking GAA GaN with Si NW-nFET . . . . .	43
4.3.3 Gate efficiency and current normalization . . . . .	46
4.3.4 Influence of geometrical shape . . . . .	47
4.4 Summary . . . . .	49
<b>5 MONOLAYER WTe<sub>2</sub>-ZrS<sub>2</sub> vdW TFET . . . . .</b>	<b>52</b>
5.1 Motivation . . . . .	52
5.2 Simulation approach . . . . .	54

	Page
5.3 Device design . . . . .	55
5.4 n-type TFET operating principle . . . . .	59
5.5 p-type TFET operating principle . . . . .	63
5.6 Device performance . . . . .	66
5.7 Summary . . . . .	70
REFERENCES . . . . .	72
A PLOT LOCAL DENSITY OF STATES USING TECPLOT . . . . .	87
A.1 Introduction . . . . .	87
A.2 Data preparation . . . . .	87
A.3 Plot figure in Tecplot . . . . .	87
B FINITE ELEMENT IMPLEMENTATION OF LANDAU-KHALATNIKOV EQUATION . . . . .	93
B.1 Introduction . . . . .	93
B.2 Equation discretization . . . . .	93
B.3 Solve with Newton's method . . . . .	98
B.4 Solve with iterative method I . . . . .	99
B.4.1 Computing with trapezoidal integration . . . . .	100
B.4.2 Computing with high order integration . . . . .	102
B.5 Solve with iterative method II . . . . .	102
C AN ELECTRO-THERMAL COUPLED QUANTUM TRANSPORT MODEL	105
C.1 Introduction . . . . .	105
C.2 Büttiker probe RGF algorithm . . . . .	106
VITA . . . . .	111

## LIST OF TABLES

Table	Page
2.1 The heuristic interpolation factor $\lambda(E)$ of the three heuristic ECA models in band-to-band tunneling devices. All given formulas and numbers are dimensionless. The interpolation factor $\lambda(E)$ is applied when the energy is above the valence band edge and below the conduction band edge. Below (above) the valence (conduction) band edge, $\lambda(E)$ is equal 1 (0). If transport happens exclusively in valence or conduction bands (e.g. in MOSFETs), no interpolation factor is applied in ECA. . . . .	16
4.1 Performance benchmark of square Si-, GaN- and triangular GaN-NW-nFETs of comparable cross-sectional area. . . . .	50

## LIST OF FIGURES

Figure	Page
1.1 The evolution of semiconductor transistors. From the left to right: planar, fully depleted fin and gate-all-around structure. For each new generation, transistors become smaller, faster and consume less power. This figure is taken from Samsung newsroom [1]. . . . .	2
1.2 42 years of microprocessor trend data. Original data up to the year 2010 collected and plotted by M.Hotowitz, F.Labonte, O.Shacham, K.Olukotun, L.Hammond and C.Batten. New plot and data collected for 2010 – 2017 by K.Rupp. This figure is taken from Dr. Karl Rupp’s personal website [2].	3
1.3 A schematic diagram of heat paths in planar and tri-gate bulk structures. The electron current conduction direction is perpendicular to the plane of the figure. © 2013 IEEE. Reprinted, with permission, from Ref. [6]. . . .	6
2.1 A typical integration contour used for $n_{eq}$ in Eq. (C.4). The Fermi level of the left contact $\mu_L$ is set to 0. Poles enclosed by the contour are marked by crosses and highlighted in the inset. The inset also illustrates the dense distribution of mesh points around the Fermi level that ensure a well resolved contact Fermi function. The arrow on the contour indicates the direction of the integral in the complex plane. . . . .	18
2.2 The atom resolved electron density $n_{eq}$ vs number of energy points per momentum point for the real energy integral in Eq. (C.5) (a) and for the complex contour integral of Eq. (C.6) (b). The converged number of 4 valence Si electrons is indicated with a solid line to guide the eye. The $n_{eq}$ converges with only 30 energy points in Eq. (C.6) while no convergence is observed when solved along the real energy axis even with an immense number of energy points. . . . .	20
2.3 Variation of charge density with effective Fermi level calculated with both approaches. The two dashed lines indicate the valence (at 0) and conduction band edges. Deviation between ECA and FBA increases when the Fermi level is deep in the valence band or the conduction band. Then an increasing number of van Hove singularities plague the convergence of ECA.	21

Figure	Page
2.4 Schematic of the ultra-thin body double-gate transistors used in all transfer characteristic predictions of this work. The gate bias is controlled by $V_{GS}$ . Electron transport occurs along $x$ direction when a non-zero $V_{DS}$ is applied. The channel of the device is confined along $z$ direction and $t_{ch}$ is the channel thickness. Periodic boundary condition is assumed along $y$ direction. $L_{ch}$ is the channel (gate) length. $N_S$ and $N_D$ are doping concentrations in source and drain regions, respectively. . . . .	22
2.5 Transfer characteristics $I_D - V_{GS}$ of silicon ultra-thin body double-gate nMOSFET (a) and pMOSFET (c) of Fig. 2.4 predicted with ECA and FBA. The percentage difference is plotted in dashed lines with cross markers. (b) and (d) Band profiles of ECA and FBA corresponding to $V_{GS}$ nodes in (a) and (c), respectively. . . . .	23
2.6 Transfer characteristics $I_D - V_{GS}$ at $V_{DS} = 0.3V$ of a silicon ultra-thin body double-gate TFET. Results of FBA and ECA with three different heuristic models are shown. The maximum deviation of the three ECA results relative to their average is plotted in dashed line with cross markers. 24	24
2.7 Contour plot of the energy and position resolved density of states of the silicon TFET simulated in Fig. 2.6 at $k = \Gamma$ . The conduction and valence band edges are represented in white solid lines. The white dashed line depicts a hole/electron delimiter for $\alpha = 0.5$ in the band gap which is used to distinguish electron and hole states in ECA. . . . .	25
2.8 Transfer characteristics $I_D - V_{GS}$ at $V_{DS} = 0.3V$ of a double-gate p-GaSb/n-InAs heterojunction TFET with channel thickness $t_{ch} = 3nm$ predicted with FBA and ECA with three different heuristic models. The maximum deviation of the three ECA results relative to their average is plotted in dashed line with cross markers. The inset shows band profiles in the middle of the TFET for two gate voltages ( $V_{GS} = -0.4V$ and $V_{GS} = -0.1V$ ). . . . .	26
3.1 (a) Simulation domain considered in this work. Regions marked by 1 through 4 are three atomic layers in the middle of Si, left to the interface, right to the interface and in the middle of heavy Si, respectively. (b) Schematic of the thermal boundary resistance extraction. . . . .	34

Figure	Page
3.2 (a) Thermal conductivity of bulk Si calculated with MD and NEGF. The Büttiker probe parameters $B$ and $C$ are fitted such that the transport results of NEGF agree with MD. (b) Thermal boundary resistance as a function of the mass ratio (MR) calculated by NEGF with the fitted parameters $B$ and $C$ of (a) and MD. The error bar of MD simulation is calculated but not plotted since the standard deviation for independent simulations is only 1.6% to 4.6% of the average resistance at the Si/heavy-Si interface for different mass ratios. The dotted line is the relative difference of the two methods and is defined as $(NEGF - MD)/MD \times 100\%$ . The green curve shows the results calculated by NEGF without Büttiker probe self-energies. With increasing masses, the maximum phonon frequency and (with the frequency dependence of the Büttiker probes, Eq. B.2) the average scattering strength reduces. Consequently, the impact of scattering on $R_{interface}$ declines with increasing mass of heavy-Si. . . . .	36
3.3 Linear extrapolation of NEGF and MD for three $MR$ values. MD results show a stronger dependence on the length of the device. Open boundary conditions included in NEGF with contact self-energies give almost device size independent results. . . . .	38
3.4 (a) Energy resolved (transverse momentum integrated) density of states in Si and heavy-Si leads. (b)~(d) Energy resolved (transverse momentum integrated) current of NEGF in the heat source and sink calculated with normal, artificially weak ( $1/20\times$ ) and artificially strong ( $20\times$ ) scattering strengths, respectively. . . . .	39
3.5 The ratio of the phonon mode in regions 1 ~ 4 defined in Fig. 5.1 and by Eq. B.4 ( $\sum_{i=1}^4 R_i = 1$ ). The evenly distributed phonon modes around 22 meV (marked by the gray area) result in the highest current peak at 22 meV (see Fig. 5.4(b)) although having a lower DOS and lower group velocity compared to those at 12 meV. . . . .	41
4.1 (a)~(d) Schematic of the GaN Nanowire n-MOSFET (GaN-NW-nFET) with different cross-section shapes simulated in this work and (e) important device parameters. . . . .	44
4.2 (a) Energy dispersion relations along the transport direction of Si and GaN square cross-section nanowire ( $T_{nw} = 1.6$ nm). (b)~(d) $I_D$ vs $V_G$ curves ( $I_{OFF}$ fixed at $10^{-2}$ $\mu\text{A}/\mu\text{m}$ ), $I_{ON}$ vs $I_{OFF}$ curves (calculated at $V_{DD}=0.4$ V) and Subthreshold swing vs $I_{DS}$ curves of GaN and Si NW-nFETs with square cross-section shape. . . . .	45
4.3 Percentage of $I_{ON}$ enhancement of gate-all-around (GAA) GaN-NW-nFET with respect to tri-gate (TG) and double-gate (DG) as a function of $T_{nw}$ . . . . .	46

Figure	Page
4.4 $I_{ON}$ vs $I_{OFF}$ curves (calculated at $V_{DD}=0.4$ V) (a) normalized with respect to the perimeter and (b) with respect to the cross-section area of square cross-section GaN-NW-nFETs with different gate configurations and $T_{nw}=2.0$ nm. . . . .	47
4.5 (a) $I_{ON}$ vs $I_{OFF}$ (calculated at $V_{DD}=0.4$ V) of GaN-NW-nFETs with different geometrical shapes. (b) GaN electron effective mass ( $m_{eff}$ ) calculated using the $sp3d5s^*$ TB basis. Purple symbols are DFT calculations from Ref. [102]. (c)~(g) Cross section electron density profiles at top of the barrier for Si (c) and GaN ((d)~(g)) with different geometrical shapes at $V_{GS}=V_{DD}=0.4$ V ( $I_{OFF}$ fixed at 200 pA). . . . .	49
4.6 Energy-delay vs intrinsic gate-delay of the simulated GaN- and Si- NW-nFETs. Cross-section shape of GaN-NW is indicated by symbols: square (black), circular (yellow), hexagonal (red) and triangular (blue). The simulated devices account for a 35 % current reduction estimate due to surface roughness [95]. . . . .	50
5.1 (a) Band structure of monolayer $WTe_2$ - $ZrS_2$ heterostructure calculated with VASP. (b) Band structures calculated with VASP (lines) and NEMO5 (symbols) at $k_y=0$ along the transport direction. Both VBM of $WTe_2$ and CBM of $ZrS_2$ are folded to $k_y=0$ . (c)-(d) The top view and the side view of the monolayer $WTe_2$ - $ZrS_2$ heterostructure. The primitive unit cell is marked by a green diamond. The gray rectangle marks the unit cell that is used in all transport calculations. The transport direction is along the $x$ axis. . . . .	56
5.2 The cross-section of the n-type (a) and the p-type (b) TFET design based on the monolayer $WTe_2$ - $ZrS_2$ vertical heterojunction. . . . .	58
5.3 (a) OFF state ( $V_{tg}=0$ V) energy resolved local density of state (LDOS) at $k_y=0$ plotted along the following path: $x=0$ ( $WTe_2$ ) $\rightarrow$ $x=25$ nm ( $WTe_2$ ) $\rightarrow$ $x=25$ nm ( $ZrS_2$ ) $\rightarrow$ $x=90$ nm ( $ZrS_2$ ), as indicated by the white dashed line in the inset. Within the $WTe_2$ ( $ZrS_2$ ) layer, the $z$ coordinate is taken at the atomic plane of W (Zr). Red: high LDOS, blue: low LDOS. $VBM_{ch}$ and $CBM_{ch}$ mark the VBM of $WTe_2$ and the CBM of $ZrS_2$ in the overlap region, respectively. $CBM_{ext}$ marks the CBM of $ZrS_2$ in the drain extension region. (b) Same as (a), but at ON state ( $V_{tg}=0.4$ V). . . . .	60

Figure	Page	
5.4	Energy resolved transmission at $k_y=0$ at ON state. The three black dashed lines indicate $VBM_{ch}$ , $CBM_{ch}$ and $CBM_{ext}$ , respectively. The green, red and yellow curves represent transmission with energies above $VBM_{ch}$ , between $CBM_{ch}$ and $VBM_{ch}$ , below $CBM_{ext}$ , respectively. The blue dashed line shows the normalized cumulative current density. The inset associates the three transmission energy intervals with three different tunneling paths. The green, red and yellow arrows represent the tunneling paths of the green, red and yellow transmission curves, respectively. . . . .	61
5.5	(a)-(b) Same as Fig. 5.3(b), but with cut planes at (a) $x=17$ nm and (b) $x=33$ nm, respectively. . . . .	62
5.6	(a) OFF state ( $V_{bg}=0$ V) energy resolved local density of state (LDOS) at $k_y=0$ plotted along the following path: $x=0$ ( $WTe_2$ ) $\rightarrow$ $x=65$ nm ( $WTe_2$ ) $\rightarrow$ $x=65$ nm ( $ZrS_2$ ) $\rightarrow$ $x=90$ nm ( $ZrS_2$ ), as indicated by the white dashed line in the inset. Within the $WTe_2$ ( $ZrS_2$ ) layer, the $z$ coordinate is taken at the atomic plane of W (Zr). Red: high LDOS, blue: low LDOS. $VBM_{ch}$ and $CBM_{ch}$ mark the VBM of $WTe_2$ and the CBM of $ZrS_2$ in the overlap region, respectively. $VBM_{ext}$ marks the VBM of $WTe_2$ in the drain extension region. (b) Same as (a), but at ON state ( $V_{bg}=-0.4$ V). . . . .	64
5.7	Energy resolved transmission at $k_y=0$ at ON state. The three black dashed lines indicate $VBM_{ext}$ , $VBM_{ch}$ and $CBM_{ch}$ , respectively. The green, red and yellow curves represent transmission with energies above $VBM_{ext}$ , between $CBM_{ch}$ and $VBM_{ch}$ , below $CBM_{ch}$ , respectively. The blue dashed line shows the normalized cumulative current density. The inset associates the three transmission energy intervals with three different tunneling paths. The green, red and yellow arrows represent the tunneling paths of the green, red and yellow transmission curves, respectively. . . . .	66
5.8	(a)-(b) Same as Fig. 5.6(b), but with cut planes at (a) $x=57$ nm and (b) $x=73$ nm, respectively. . . . .	67
5.9	(a)-(c) Transfer characteristics, $SS$ vs $I_d$ and Output characteristics of the n-type TFET, respectively. The red curves include band tail states with the calibrated Urbach parameter while the blue curves consider zero band tail states. The four curves from top to bottom in (c) represent $V_{tg}=0.4$ V, 0.3 V, 0.2 V and 0.1 V, respectively. (d)-(f) Same as (a)-(c) but for the p-type TFET. The four curves from top to bottom in (f) represent $V_{bg}=-0.4$ V, $-0.3$ V, $-0.2$ V and $-0.1$ V, respectively. . . . .	68

Figure	Page
5.10 (a) Transfer characteristics of the n-type TFET with four different drain extension lengths ( $L_{ext}$ ) assuming zero band tail states. (b) n- and (c) p-type TFET output characteristics comparison of the NEGF simulation data (symbol) and the MVS-TFET model (solid line). (d) Schematic of the 25-stage ring oscillator circuit. Each inverter circuit in the chain uses the designed n- and p-type TFETs in this work. (e) Simulated waveform of the 25-stage ring oscillator circuit shown in (d) with an oscillation frequency of 1 GHz at $V_{dd}=0.4$ V. (f) Intrinsic energy versus delay, benchmarked against ultrascaled silicon MOSFETs from IRDS [151] (black data points) and TFETs based on III-V materials [152] (blue and green data points). The preferred corner is bottom left. Dashed lines: constant energy-delay products. . . . .	69
A.1 Setting data association with axis. . . . .	88
A.2 Setting one slice perpendicular to the x axis. . . . .	89
A.3 Setting two slices perpendicular to the y axis. . . . .	90
A.4 The plot after setting the slices. . . . .	90
A.5 Setting variables. . . . .	91
A.6 Setting value blanking. . . . .	92
A.7 The final figure. . . . .	92
C.1 Program flowchart of the electro-thermal coupled quantum transport model.	105

## ABSTRACT

Chu Yuanchen Ph.D., Purdue University, December 2019. Predictive Electro and Thermal Quantum Transport in Nanoscale Devices. Major Professor: Gerhard Klimeck.

Modern semiconductor devices have reached the sub-20 nm regime. Before long, it will be practically impossible to further scale down the size of Si-based MOSFETs due to short channel effects. Novel device geometries (e.g. tri-gate and gate-all-around), device concepts (e.g. TFET and NCFET) and channel materials (e.g. III-Vs and TMDs) have been proposed to achieve lower power dissipation and faster speed, allowing for higher transistor density on a chip. In the past decade, quantum transport methods (e.g. NEGF) have become the standard approaches for modeling such nanoscale devices.

In state-of-the-art quantum transport models, the dielectric constant is typically set to the material's constant, neglecting the spatial variation of the screening effects in the nanodevice structure. When applied to TFETs, hybrid states that enable band-to-band tunneling are subject to interpolation that yields model dependent charge contributions. In this work, it is exemplified that the use of different charge interpretation models brings large variability when applied to ultra-thin body transistor performance predictions. To solve these modeling challenges, an electron-only band structure model is extended to atomistic quantum transport. Performance predictions of MOSFETs and TFETs confirm the generality of the new model and its independence of additional screening models.

Secondly, as devices become smaller, their thermal resistances increase because of the reduced area under the device and thinner silicon layer in the horizontal direction. Thus, despite lower power per device, self-heating effects in digital circuits

are actually increasing which compromise both the performance and the reliability of the device. Therefore, for future electronics, an increasingly important role of energy dissipation necessitates electro-thermal co-design. As a first step, seeking to include anharmonicity in phonon related NEGF, the NEGF method with Büttiker probe scattering self-energies is proposed and its accuracy is assessed by comparing its predictions for the thermal boundary resistance with molecular dynamics (MD) simulations. For simplicity, the interface of Si/heavy-Si is considered. With Büttiker probe scattering parameters tuned against MD in homogeneous Si, the NEGF-predicted thermal boundary resistance quantitatively agrees with MD for wide mass ratios, proving that the proposed method provides an efficient and reliable way to include anharmonicity in phonon related NEGF. An algorithm to couple the electron and phonon transport in the NEGF formalism via Büttiker probes is also proposed.

Thirdly, NEGF with self-consistent Born approximation is introduced for modeling band tail and bandgap narrowing driven by LO phonons and charged impurities in III-V semiconductors. Extracted scattering rates are benchmarked against Fermi's golden rule. Urbach tail and band gap narrowing calculated in bulk III-V materials agree well with experimental results for a range of temperature and doping concentration. Predictions are made for band-tail and bandgap narrowing in confined structures.

Next, the performance of 5 nm gate length GaN nMOS nanowire field effect transistor (GaN-NW-nFET) of various geometrical shapes is investigated, around the limits of cross-sectional scalability, using atomistic quantum transport simulations. Benchmarking results with simulated Si-NW-nFET reveal large enhancement in GaN drive current in both Low Standby Power (LP) and High Performance (HP) applications. Further performance enhancement is observed with the use of non-square geometries that are akin to GaN's wurtzite crystal structure. Particularly, it is found that triangular cross-section GaN-NW-nFETs exhibit the smallest subthreshold swing, excellent drive current and superior energy-delay product compared to simulated Si-NW-nFET.

Furthermore, quantum transport simulation is applied to design of complementary van der Waal TFETs based on the monolayer p-WTe<sub>2</sub>/n-ZrS<sub>2</sub> vertical heterojunction. Non-idealities such as electron-phonon and electron-electron interactions are included via a phenomenological scattering model. Through band structure engineering and design of the electrostatics, both n- and p-type TFETs are realized with the same device configuration. Transfer and output characteristics, suitable for VLSI applications, are observed. 3D views of energy resolved local density of states (LDOS) illustrate device operating principles and identify three different tunneling paths in the device. It is found that in order to improve the electron collection at drain (source) of the n-type (p-type) TFET, efforts are needed to seek alternative material combinations with similar values of LDOS in the conduction and the valence band of the n- and the p-type materials, respectively. Circuit level simulation of the designed n- and the p-type TFETs is performed on a 25-stage ring oscillator which demonstrates the promise of the proposed designs for low-power digital VLSI applications.

Lastly, a finite element Landau-Khalatnikov equation solver is implemented in NEMO5, as a preparation to enable NCFET modeling.

# 1. INTRODUCTION

## 1.1 Need for quantum transport in nanodevices

Over the past few decades, the semiconductor industry has been driven by the goal of reducing the cost per function on a chip. One way to achieve it is through device scaling (i.e. to reduce the size of the transistor). With the footprint of each transistor being smaller, more transistors can be put onto a die of the same area, thus increasing the computational power of the chip. Nowadays, the distance between the source and the drain of a transistor has reached the sub-20 nm regime. As the channel length becomes smaller and smaller, short channel effects start to degrade the device performance. To overcome it, the fin structure, where three sides of the channel are in contact with gates, was introduced by Intel at its 22 nm node in 2011. Nevertheless, the fin structure is now facing limitations after a few generations of development. To go further, industry will soon move to the gate-all-around structure where the semiconducting channel is completely surrounded by gates. Fig. 1.1 schematically illustrates the transistor evolution mentioned above. To accurately model these devices, quantum mechanics cannot be neglected since it plays an important role in determining several key aspects of the device performance. For example, the source drain direct tunneling leakage can degrade the subthreshold performance; quantum confinement affects the charge distribution in the channel cross-section; quantum interference can result in resonance state which might have a large impact on both ON and OFF states of the device. Classical transport models like drift-diffusion equation cannot capture these quantum effects accurately. In fact, in the past decade, quantum transport methods (e.g. NEGF) have become the standard approaches to model these nanoscale devices.

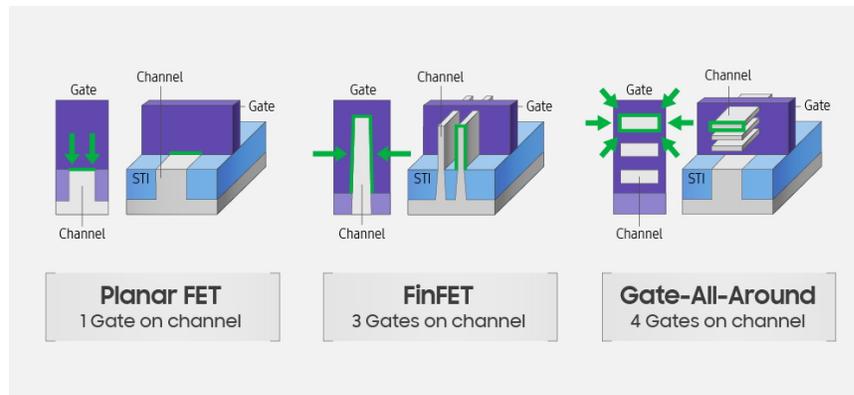


Fig. 1.1.: The evolution of semiconductor transistors. From the left to right: planar, fully depleted fin and gate-all-around structure. For each new generation, transistors become smaller, faster and consume less power. This figure is taken from Samsung newsroom [1].

The second reason why quantum transport is needed for modeling nanodevices is because of its capability to include realistic material physics. As the number of transistor on a chip increases exponentially over the past few decades, so does the power density. The dynamical power consumption, which is associated with the switching of the capacitors, is proportional to  $\kappa C V_{dd}^2 f$ , where  $C$  is the load capacitance,  $V_{dd}$  is the supply voltage,  $f$  is the CPU frequency and  $\kappa$  represents an averaged percentage time of the device being in the active switching mode. From this expression, it is obvious that reducing the supply voltage is very useful to suppress the dynamical power consumption. However, the equilibrium thermodynamics defines the fundamental lower limit of the subthreshold swing to be 60 mV/dec at room temperature, which precludes the reduction of the supply voltage and the overall power consumption of the integrated circuit. Due to this reason, we don't see CPUs with operating frequency over 4 GHz in today's PC market, since the power density will be so high that it requires extra cooling system to keep the chip temperature low. Instead, as shown in Fig. 1.2, the CPU frequency hasn't increased since about 10 years ago and the industry is embracing multicore CPUs and parallelism. On the other hand, new

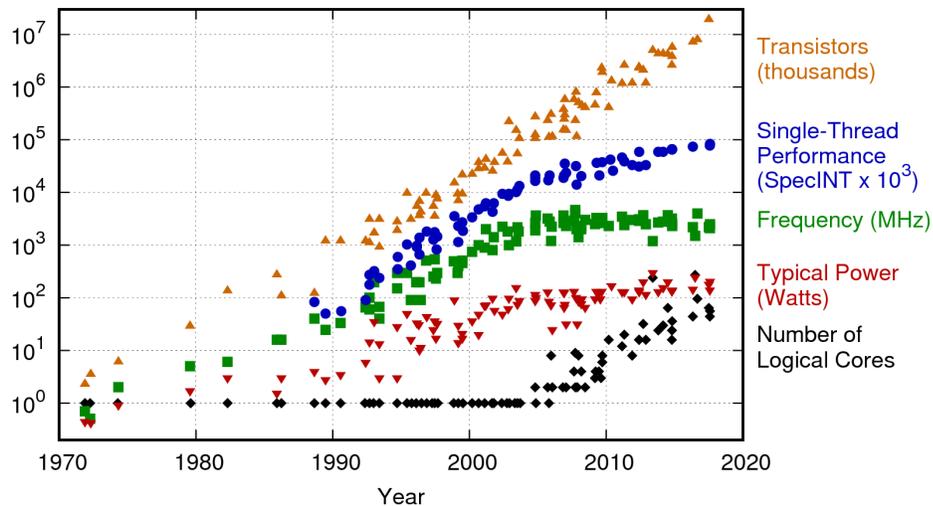


Fig. 1.2.: 42 years of microprocessor trend data. Original data up to the year 2010 collected and plotted by M.Hotowitz, F.Labonte, O.Shacham, K.Olukotun, L.Hammond and C.Batten. New plot and data collected for 2010 – 2017 by K.Rupp. This figure is taken from Dr. Karl Rupp’s personal website [2].

device concept, like tunnel FET (i.e. TFET), was proposed which can surpass the fundamental thermodynamics limit and be able to allow for further reduction of the supply voltage. Instead of thermionic current in a conventional MOSFET, the current in a TFET is the result of electrons tunneling from the valence band to the conduction band of the channel material. Therefore, large band gap material like silicon is not favored since the ON-state current will be too small to drive the circuit. III-V materials, especially III-V heterojunctions are of particular interest since they can form Type-II staggered or Type-III broken-gap band alignment, thus boost the ON-state current. In addition, van der Waal (vdW) heterostructure made by stacking one 2D material on another 2D or 3D material is also a promising candidate for TFET applications. Compared to III-V materials, the advantage of using 2D materials is that they do not have dangling bonds on their surfaces, which allows for the formation of high quality stacked vdW heterostructures with atomically sharp interfaces. Besides, the large amount of exfoliable 2D materials provides a decent design space which

allows for tailoring of the band alignment and the electrostatic control to enable ideal device performance. To model TFETs with these novel materials, it is not enough to simply represent the material with effective mass. This is because many other aspects, like band gap, coupling between valence and conduction bands, transverse  $k$  dependence and non-parabolicity can have dominating influences on the device performance. Quantum transport methods, on the other hand, solve the Schrödinger equation in the active region of the device and are able to incorporate sophisticated basis sets to accurately describe the channel material with atomistic resolution, thus providing a much more reliable solution for TFET modeling.

## **1.2 Missing elements in state-of-the-art quantum transport**

### **1.2.1 An explicit screening model**

In state-of-the-art quantum transport methods, Schrödinger equation is solved in the active device region and is coupled self-consistently with the Poisson equation. After convergence is achieved, quantities like charge distribution and current can be calculated by applying quantum mechanical operators to the wave function. In Poisson equation, to account for the screening effect, the dielectric constant is set to the material's constant. However, it is well known that in thin films, the screening close to the edges is smaller than that in the bulk of the material [3]. By neglecting the spatial variation of the screening effects in nanodevices where the thickness of the semiconducting channel is of only a few nanometers, the accuracy of the converged electrostatics is in question. In devices where the current-voltage characteristic is sensitive to the electrostatics, the predicted device performance can be with large error. Although it is possible to define a spatially varying dielectric constant in the Poisson equation, the effort to obtain experimentally measured or theoretically calculated data is very large since the problem is device structure dependent (e.g. it depends on channel thickness).

### 1.2.2 A realistic charge model for TFET

In conventional MOSFETs, only one carrier type is involved in the device operation (electrons for n-type MOSFETs and holes for p-type MOSFETs). For TFETs, since the current is a result of holes in the valence band becoming electrons in the conduction band through quantum tunneling, both carrier types need to be considered in the model. In state-of-the-art quantum transport methods, this is done by considering all states in the conduction (valence) band as electron (hole) states and applying an interpolation for states in the band gap region (i.e. above the local valence band maximum and below the local conduction band minimum). Such approach has two obvious problems. First, it yields model dependent charge distribution and second, it relies on the knowledge of local band edges which might not even exist in some situations. Both problems have direct impacts on the converged electrostatics and therefore, could bring large error to TFET performance predictions.

### 1.2.3 A sophisticated scattering model for TFET

According to a recent compilation of state-of-the-art TFET results [4], it is obvious that there exists a large discrepancy between simulation and experiment. One important source of such discrepancy is due to that most TFET simulations assume ballistic electron transport, ignoring various non-ideality effects. For instance, the Urbach tail, which is resulted from electron-electron interaction, impurities and polar optical phonon scattering, exponentially extends band edges into the band gap region, thus can greatly degrade the subthreshold performance of TFETs. Another example would be the dangling bonds at the surface of the semiconductor, which bring defect states in the band gap region. These defect states could act as hopping centers for electrons to tunnel from the valence band to the conduction band of the channel materials, therefore making it difficult to turn the device off. Although there exist phenomenological models to include non-ideality effects, to accurately model TFETs, a sophisticated and computationally efficient scattering model is in urgent need.

In an effort to understand band-tail formation and band-gap narrowing driven by polar optical phonons and charged impurities in III-V semiconductors, we introduce an NEGF method where scattering is solved for in the self-consistent Born approximation. To verify the proposed method, scattering rates due to polar optical phonon and charged impurity scattering in bulk, ultrathin body and nanowire GaAs are calculated and benchmarked against Fermi's golden rule. Urbach tails and band-gap-narrowing values are extracted for various bulk III-Vs. The extracted Urbach tails and conduction-band-gap narrowing agree well with experimental results for a range of temperatures and doping concentrations. Using the same approach, Urbach tails and band-gap-narrowing values are predicted for ultrathin body and nanowire structures. This work is published in Ref. [5].

#### 1.2.4 An electro-thermal coupled model for self-heating

Transistors dissipate active power as heat to surroundings. In planar architectures, the heat access to the bulk through lateral and vertical spreading. However, in tri-gate bulk structure, due to the confinement of the gate, the heat dissipation process sees a higher thermal resistance compared to the bulk, which eventually leads to an increase in local temperature of the device (see Fig. 1.3). The rise in device temperature due

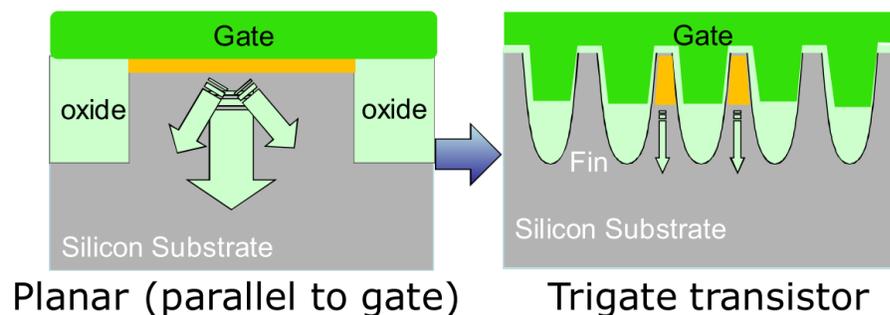


Fig. 1.3.: A schematic diagram of heat paths in planar and tri-gate bulk structures. The electron current conduction direction is perpendicular to the plane of the figure. © 2013 IEEE. Reprinted, with permission, from Ref. [6].

to self-heating effects (SHE) was reported by TSMC in their 16 nm-node bulk finfet technology [7] and by Intel in their 22 nm-node bulk finfet technology [6]. In the future, gate-all-around structure could suffer more from SHE since the channel is completely wrapped by gate oxide. And it is already reported in Si nanowire FET [8] and InGaAs nanowire FET [9]. In addition, monolayer MoS<sub>2</sub> FET is reported to suffer greatly from SHE due to small in-plane thermal conductivity and large channel-to-substrate thermal boundary resistance [10]. As device scaling goes on, despite lower power per device, SHE in digital circuits will probably increase due to worse thermal dissipation channels and higher device density, which compromises both the performance and the reliability of the device. Therefore, the increasingly important role of energy dissipation in future technology nodes necessitates electro-thermal co-design.

At the most basic level, energy dissipation begins in the ultrathin transistor channel, where a high lateral electric field energizes electrons, agitates electron-phonon interaction and heats up the lattice. On the thermal modeling side, the tiny dimension of the channel requires a quantum mechanical treatment for the heat transport, which should be done at the phonon level. The phonon transport should be coupled to the electron transport with energy exchange between electrons and phonons. Meanwhile, the electron transport should depend on the lattice temperature. This coupled electro-thermal transport should be solved self-consistently and a global energy conservation in the device should be ensured.

### 1.3 Thermal boundary resistance

Semiconductor nanodevices are typically composed of several semiconductor materials. The scattering of thermal energy carriers at the interface of two semiconductor materials results in thermal boundary resistance. Accurate prediction of the thermal boundary resistance gives important insight into the device physics and enables design improvements. This is especially true for nanoscale transistors, where

there is an increasing concern of energy dissipation. Molecular dynamics (MD) is often used to model the thermal boundary resistance. Inelastic phonon scattering is included in MD simulations through the anharmonicity of the interatomic potential. On the other hand, phonon transport in the non-equilibrium Green's function (NEGF) method has been used predominantly in the coherent (harmonic) regime due to the fact that the inclusion of incoherent scattering such as phonon-phonon decay usually requires solving polarization graphs in the self-consistent Born approximation which entails a large numerical load. It has been shown that lack of anharmonicity in NEGF simulation gives incorrect thermal boundary resistance predictions. Introducing a reliable and computationally efficient way to include anharmonicity in phonon related NEGF is of great importance. When electrons and phonons are both solved in the NEGF framework, interparticle interactions and energy and momentum transfer in e.g. self-heating or thermoelectric situations can be described on equal footing with the predictions of the respective particles' propagation.

#### 1.4 GaN nanowire nFET

Since its first introduction in 2011, the finfet structure is now facing scaling limitations after several generations of development. Gate-all-around structure is being investigated for future technology nodes due to its excellent electrostatic control and the capability to reduce the transistor footprint. Although silicon is still dominating the semiconductor industry nowadays, intrinsic properties of gallium nitride (GaN), such as its wide band gap and large saturation velocity, make it an intriguing option for gate-all-around n-MOSFET devices for digital applications beyond 7 nm technology. Besides, GaN logic devices also complement the well-known markets of GaN for RF and power applications. Accurate performance prediction of sub-10 nm nanowire transistors requires atomistic resolution. Especially, since GaN's wurtzite crystal structure has a hexagonal lattice rather than a cubic lattice, the influence on device

performance due to different cross section shapes of the device also requires careful investigations.

### 1.5 Monolayer $\text{WTe}_2\text{-ZrS}_2$ vdW TFET

In search of steep subthreshold swing (SS) devices which allow for further reduction of the supply voltage of integrated circuits, tunnel field-effect transistor (TFET) is being actively investigated due to its ability to surpass the thermionic limit and achieve steep SS. A major challenge of TFET lies in defect and band tail states around the tunnel junction since those states act as hopping centers for electrons in the valence band to tunnel to the conduction band of the channel materials, which can severely degrade the promised subthreshold performance of TFET. Recently, TFETs based on 2D/2D and 2D/3D van der Waal (vdW) heterostructures have attracted considerable attention because 2D materials do not have dangling bonds on their surfaces, thus allowing for the formation of high quality stacked vdW heterostructures with atomically sharp interfaces. Besides, the large number of exfoliable 2D materials provides a decent design space for tailoring the band alignment and electrostatic control to achieve ideal device performance. In addition, TFETs based on 2D crystal also benefit from ultra-thin body thickness which suppresses the short channel effects and leads to further scaling down of the channel length.

However, in spite of the growing interest in vdW TFETs, theoretical studies that employ quantum mechanical models necessary to accurately assess the device performance and unveil important design guidelines, are scarce. One important reason is the lack of nearest neighbor semi-empirical tight binding parameters for vdW heterostructures. Besides, computationally efficient quantum transport models that include sophisticated treatment of scattering effects and non-idealities are still under development. Nevertheless, a simulation effort to provide insights of device operation principles, unveil design guidelines, assess circuit performance, etc., will have a direct impact on a broad array of applications of ultrathin vdW heterostructure, ranging

from novel electronic and optoelectronic devices to electrical circuits and thermal devices.

## 1.6 Negative capacitance FET

The concept of negative capacitance FET (i.e. NCFET) was proposed by S. Salahuddin and S. Datta in 2008 [11]. The idea is to put a layer made of ferroelectric material between the metal gate and the oxide. When proper capacitance matching is done, the negative capacitance segment of the ferroelectric capacitor (forbidden if standalone) can be effectively stabilized. As a result, the capacitance of the overall ferroelectric-oxide-semiconductor system ( $C_{eq}$ ) will be larger than both  $C_{ox}$  and  $C_s$ , where  $C_{ox}$  and  $C_s$  are the gate oxide and the semiconductor quantum capacitance, respectively. This means that sub-60 mV/dec subthreshold swing will be possible at room temperature, and energy efficiency will also be improved since given a fixed value of charge ( $Q$ ), the intrinsic switching energy  $Q^2/C_{eq}$  will be smaller than the case where the ferroelectric layer is not present. In 2011, crystalline phases with ferroelectric behavior is found in thin films of doped hafnium oxide [12], which is an important progress since ferroelectric hafnium oxide has excellent compatibility to silicon CMOS technology. Recently, a direct measurement of steady-state negative capacitance in a ferroelectric-dielectric heterostructure was also demonstrated [13]. However, there is still a debate over the microscopic origin of the quasi-static negative capacitance regime in the ferroelectric layer and whether it can lead to sub-60 mV/dec subthreshold swing [14, 15].

## 1.7 Thesis outline

The thesis is organized as follows. In Chapter 2, a computationally efficient charge model is introduced which explicitly solves the screening effects in nanoscale devices. Besides, since this model considers all states as electronic states (i.e. filled by electrons), the ambiguity in dealing with hybrid states as discussed in Section 1.2.2 is not

present in the proposed model. In Chapter 3, a reliable and computationally efficient method to include anharmonicity in phonon related NEGF is presented. The method is applied to a benchmark system of Si/heavy-Si interface. The thermal boundary resistance at the interface is calculated and benchmarked against MD simulation. Quantitative agreement between NEGF and MD is achieved. In Chapter 4, quantum transport is applied to GaN nanowire nFET with 5 nm channel length. Device performances are benchmarked against simulated Si nanowire nFET. The influence of cross-section shape on GaN nanowire nFET device performance is also examined. In Chapter 5, quantum transport is applied to design of monolayer  $\text{WTe}_2\text{-ZrS}_2$  vdW TFET. In this study, nonideality effects such as electron-phonon and electron-electron interactions are included via a phenomenological model. Device performance degradation due to band tail is also assessed. Through band structure engineering and design of electrostatics, complementary n- and p-type devices are realized, with suitable current-voltage characteristics for low power VLSI applications. In addition, device operation principles are explained, device scaling is explored and circuit level simulation is performed using the designed n- and p-type vdW TFETs.

Appendix A gives a brief tutorial on plotting local density of states (LDOS) using Tecplot 360. In appendix B, a derivation of the finite element implementation of the Landau-Khalatnikov equation is documented, which is needed to model NCFET. Appendix C presents an electro-thermal coupled model in the NEGF formalism, which can be used to model self-heating effects in nanodevices or thermoelectric situations.

## 2. FULL-BAND APPROACH FOR BAND-TO-BAND-TUNNELING DEVICES

In this chapter, sections 2.1, 2.2, 2.3, 2.4, 2.5, 2.6 and a part of section 2.8 are reproduced from Ref. [16], with the permission of AIP Publishing.

### 2.1 Motivation

As the scaling of Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) has reached sub-10 nm regime, power consumption has become a major concern [17]. The advantages of lowering the dynamic power consumption by reducing the supply voltage are fast disappearing as the static power has begun to dominate due to the exponential increase of the subthreshold leakage current [18]. Band-to-band tunneling field-effect transistor (TFET) is among the most promising candidates for future integrated circuits (ICs) due to its ability to beat the 60mV/decade limit of the subthreshold swing (SS) [19, 20]. Having a smaller subthreshold swing enables a reduction of both the supply voltage and the subthreshold leakage current, thus further lowering the power consumption of the ICs [21].

However, despite many predictions of outstanding device performance, most experimental TFETs underperform conventional MOSFETs [22]. The discrepancy between TFET simulations and experiments indicates missing or mistreated physics in TFET simulations [23–25]. Analytical models of TFETs indicate that the tunneling current has an exponential dependence on effective mass, energy band gap and electric field at the tunnel junction [26]. Obviously, the accuracy of these quantities is crucial for quantitative prediction of TFET current-voltage characteristics. This is particularly true for the electrostatic landscape, since it rules all bandstructure and tunneling properties.

The Nonequilibrium Green’s function (NEGF) formalism [27–29] is widely accepted as one of the most consistent models for electronic properties in nanodevices in the presence of quantum phenomena including quantum confinement, tunneling, interferences, etc [30, 31]. Nanotransistor properties such as charge distribution and current density are commonly deduced from the NEGF equations once they are self-consistently solved with the Poisson equation.

## 2.2 The standard excess-charge approach

The standard model (termed as excess-charge approach or ECA) to interpret the particle density in quantum transport calculations distinguishes the charge carrier type: An electron (hole) in conduction (valence) band state of n-type (p-type) MOS-FET is considered to contribute a negative (positive) unit charge. This concept limits the computational load to solving electrons (holes) in the conduction (valence) band only, i.e. a few  $k_B T$  of energy in addition to the energy range spanned by the applied bias voltage. In the Poisson equation, the dielectric constant is then typically set to the material’s constant. In tunneling devices, a particle with energy above (below) the conduction (valence) band is still considered to contribute a negative (positive) unit charge. For energies between the conduction and the valence band, various charge interpolation schemes exist. As will be shown in detail in this paper, the standard model fails in various ways, both for conventional as well as for band-to-band tunneling transistors. It turns out - consistent with previous findings in literature [3, 32] - that the electrostatic screening of valence band electrons that do not take part in transport is device physics dependent (i.e. it depends on channel thickness and varies spatially). The charge interpolation schemes required for band-to-band tunneling devices host an arbitrariness that severely limits the reliability of device performance predictions. Any such interpolation also suffers from incompatibility with the NEGF method as discussed in detail in Section 2.6. This is consistent with previous findings in broken-gap optoelectronic bandstructure calculations [33]. Therefore, the main fo-

cus of this work is the introduction of a consistent interpretation of quantum transport solutions for charge self-consistent nanodevice calculations.

### 2.3 The full-band approach

All models and results of this paper are part of the multipurpose nanodevice simulation tool NEMO5 [34]. The electronic structure is represented in atomistic nearest-neighbor tight-binding (TB) sp<sup>3</sup>d<sup>5</sup>s\* basis. Throughout the paper, empirical TB parameters for Si, GaSb and InAs are taken from Refs. [35–37], respectively. In all ultra-thin body (UTB) devices of this work, the electron transport direction ( $x$ ) and the confinement direction ( $z$ ) are atomically discretized. The periodic direction ( $y$ ) is discretized with a single unit cell and boundary conditions are applied on it. A few examples are considered with periodicity in both,  $y$ - and  $z$ - direction. The full Brillouin zone of the electronic momentum ( $k$ ) in any periodic direction is discretized. Coherent transport is assumed and no incoherent scattering is considered. If not explicitly stated otherwise, all results are obtained by iterating the quantum transport equations with the nonlinear Poisson equation (discretized on a three dimensional finite element grid) until convergence is achieved. For all devices, flat band Neumann boundary conditions are assumed for the Poisson equation except for the gate/oxide interface regions where Dirichlet conditions are used to set the applied gate voltages. Full ionization of donor and acceptor atoms is assumed.

In the state of the art model (ECA), states that are in the conduction (valence) band for all device positions are considered to be electrons (holes) and to contribute a negative (positive) unit charge. Since this is commonly the case for nMOSFETs or pMOSFETs, these devices allow limiting the calculations to conduction or valence bands, respectively. In band-to-band tunneling (BTBT) situations, both valence band and conduction band states have to be considered, since states exist that overlap with conduction and valence band simultaneously. The density of such states is translated

with a heuristic interpolation factor  $\lambda$  into their charge density contribution. The expression for the charge density contribution of each individual lead is given by

$$\begin{aligned} -qn_{ECA}(E, k) &= qp(E, k) - qn(E, k) \\ &= q[1 - f(E, \mu)]\lambda(E, k)|\psi(E, k)|^2 - qf(E, \mu)[1 - \lambda(E, k)]|\psi(E, k)|^2, \end{aligned} \quad (2.1)$$

where  $E$  is the state energy,  $\psi(E, k)$  is the injected conduction or valence band state,  $\mu$  is the lead Fermi level and  $q$  is the positive unit charge. Note that in Eq. (C.1) as in all subsequent equations, the position coordinates are omitted for better readability. The electronic states  $\psi(E, k)$  are solved with the quantum transmitting boundary method (QTBM) [38, 39]. The factor  $\lambda$  heuristically interpolates between the positive and negative charge interpretations of valence and conduction band. Therefore,  $\lambda$  depends on the electrostatic potential, the conduction and valence band edges and the chosen heuristic interpolation model. In general  $\lambda$  is a function of energy, in-plane momentum and position. In this paper, three commonly used [40–42] heuristic models for  $\lambda$  have been chosen as representatives (summarized in Table. 2.1). All three heuristic models distinguish the interpolation factor  $\lambda$  for energies  $E$  below and above a delimiter energy  $D$  given by

$$D = (1 - \alpha)E_V + \alpha E_C, \quad (2.2)$$

where  $\alpha$  is a unitless number and  $E_V$  ( $E_C$ ) is the valence (conduction) band edge. For simplicity, a function  $F(\mathcal{E})$  is defined as

$$F(\mathcal{E}) = \frac{\mathcal{E} - E}{2(\mathcal{E} - D)}, \quad (2.3)$$

For energies above  $D$ ,  $\lambda$  equals the function  $F$  evaluated for  $\mathcal{E} = E_C$  and  $\lambda = 1 - F(E_V)$  otherwise. The space charge density of the Poisson equation is obtained by summing the electron or hole charge density with the background doping density. If not stated otherwise, the Poisson equation is solved in ECA with dielectric constants of the respective material. Only for silicon UTB devices the thickness dependent dielectric constant of Ref. [3] is used. Note that the interpolation of the ECA model

only applies to band-to-band tunneling situations. This can lead to inconsistencies when both, band-to-band tunneling and intra-band transport are relevant for the device performance (e.g. p/n junctions in forward bias).

Table 2.1.: The heuristic interpolation factor  $\lambda(E)$  of the three heuristic ECA models in band-to-band tunneling devices. All given formulas and numbers are dimensionless. The interpolation factor  $\lambda(E)$  is applied when the energy is above the valence band edge and below the conduction band edge. Below (above) the valence (conduction) band edge,  $\lambda(E)$  is equal 1 (0). If transport happens exclusively in valence or conduction bands (e.g. in MOSFETs), no interpolation factor is applied in ECA.

Label	$\alpha$	$E \leq D$	$E > D$
A	0.2	$\lambda = 1 - F(E_V)$	$\lambda = F(E_C)$
B	0.5	$\lambda = 1 - F(E_V)$	$\lambda = F(E_C)$
C	0.5	1	0

As a consistent alternative to the ECA model, this work extends the charge self-consistent model (termed as full-band approach or FBA) of Refs. [33,43] to atomistic quantum transport of band-to-band tunneling devices within the NEGF formalism. For the sake of completeness, we repeat the model details here. Every state solved within the quantum transport method is considered electronic and contributes, if occupied, a negative unit charge. This is irrespective of which band that state is in. However, this model requires resolving the density contribution of all occupied states.

The standard NEGF treatment of density calculations requires to integrate the diagonal of the retarded Green's function  $G^R$  over an energy interval that covers all occupied states [44,45]. The recursive Green's function implementation of NEMO5 is applied to solve for the diagonal of  $G^R$ . Green's functions and self-energies are matrices in the position space indicated in bold font. Most of the following equations involve the diagonal of the Green's functions only which is denoted in nonbold letters.

The total electron density  $n_{FBA}$  is separated into an equilibrium  $n_{eq}$  and a nonequilibrium part  $n_{neq}$ .

$$n_{FBA} = n_{eq} + n_{neq}. \quad (2.4)$$

The equilibrium electron density contribution is dependent on one contact Fermi function [46] (e.g. the left one) and is given by

$$n_{eq} = \sum_k \int_{-\infty}^{\infty} \frac{-\text{Im} [G^R(E, k)]}{\pi} f_L(E, \mu_L) dE, \quad (2.5)$$

where  $\mu_L$  is the Fermi level of the left contact. Many atomistic models yield 10s of eV with hundreds of van Hove singularities of fully occupied valence bands [47], which are all considered within  $n_{eq}$ . To avoid resolving all these states on a real energy mesh which poses immense numerical loads [48],  $n_{eq}$  is solved with the Residual theorem [49]

$$n_{eq} = \sum_k \left\{ \int_{H+C} \frac{\text{Im} [G^R(E, k)]}{\pi} f_L(E, \mu_L) dE + i2k_B T \sum_{pole} G^R(E_{pole}, k) \right\}. \quad (2.6)$$

The poles of the integrand originate from the Fermi function of the (left) contact. These poles are located at  $E_{pole} = \mu_L + ik_B T \pi(2m + 1)$  with  $\text{Res}(E_{pole}) = -k_B T$ ,  $m \in \mathbb{N}$ . A typical integration contour [43, 50] is shown in Fig. 2.1. The integration contour consists of a semicircular part (C) whose lower bound is set about 1 eV below the lowest eigenvalue of the system. The horizontal part of the contour (H) is parallel to the real energy axis. The maximum real part of H is exceeding  $\mu_L$  by  $25 k_B T$  to include the complete tail of the contact Fermi function in the density calculation. The small contour portion that closes the integration contour beyond the horizontal section H does not have a net contribution to Eq. (C.5). When the imaginary part of the integration contour is large enough (such as indicated in Fig. 2.1), numerical solutions of Eq. (C.6) converge with few tens of contour points.

The integral of the non-equilibrium electron density must be performed along the real energy axis since the integrand is not analytic in the entire complex plane.

$$n_{neq} = \sum_k \int_{-\infty}^{\infty} \text{diag} \left\{ \mathbf{G}^R(\mathbf{E}, \mathbf{k}) \frac{\text{Im} [\boldsymbol{\Sigma}^R(\mathbf{E}, \mathbf{k})]}{\pi} \mathbf{G}^R(\mathbf{E}, \mathbf{k})^\dagger \right\} \times [f_L(E, \mu_L) - f_R(E, \mu_R)] dE. \quad (2.7)$$

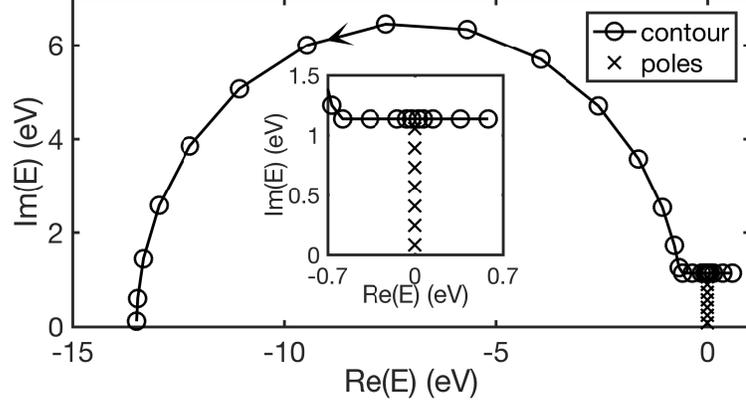


Fig. 2.1.: A typical integration contour used for  $n_{eq}$  in Eq. (C.4). The Fermi level of the left contact  $\mu_L$  is set to 0. Poles enclosed by the contour are marked by crosses and highlighted in the inset. The inset also illustrates the dense distribution of mesh points around the Fermi level that ensure a well resolved contact Fermi function. The arrow on the contour indicates the direction of the integral in the complex plane.

Equation (C.7) involves a matrix product of Green's functions and a self-energy indicated by bold letters. The integration window is restricted by the two contact Fermi functions ( $\mu_R$  being the right contact's Fermi level) and is approximately the same energy window considered in the ECA. Compared to the ECA, the extra computational load in FBA is given by a few tens of energy points for the integral of the equilibrium electron density in Eq. (C.6). This is a negligible addition given the hundreds or thousands of energy points typically needed to resolved the non-equilibrium density contribution  $n_{neq}$ .

In charge self-consistent FBA calculations, the Poisson equation requires a positive background charge  $n_{core}$  to allow the presence of electrons in the devices. This background charge is assumed to completely compensate the electronic charge density of the respective undoped device in equilibrium. The total space charge  $\rho_{FBA}$  is given by the sum of  $n_{eq}$ ,  $n_{neq}$ ,  $n_{core}$  and the doping density  $n_{doping}$ .

$$\rho_{FBA} = -qn_{FBA} + qn_{core} + qn_{doping}, \quad (2.8)$$

The definition of  $n_{core}$  and Eq. (C.8) approximate Madelung-like potentials due to charge transfer of e.g. center-device atoms and surface atoms into one reference situation. Therefore, the density  $\rho_{FBA}$  accounts for deviations of the charge density vs. the reference situation of the undoped and equilibrated device. The dielectric constant of vacuum is used in the Poisson equation for the semiconductor materials in FBA since the screening of all valence band electrons is explicitly included in the calculation [51]. As will be shown in Sec. III, this feature of FBA is important since it makes the FBA model independent of the material and device specific dielectric screening. The charge self-consistent loop with standard rank-1 updates that takes into account the response of both the equilibrium and the non-equilibrium electron density to potential changes (see Section 2.7) turns out to converge typically within 15 iterations.

## 2.4 Method verification

### 2.4.1 Convergence behavior of $n_{eq}$ calculation

The numerical convergence of solving the equilibrium electron density  $n_{eq}$  with Eqs. (C.5) and (C.6) with varying number of energy points per momentum point is compared for homogeneous 3D silicon in Figs. 2.2 (a) and (b). In both cases, the electronic Brillouin zone is resolved with 225 momentum points. The equilibrium density of Eq. (C.6) converges with only a few complex energy points (contour points and poles) per momentum point. In contrast, many and hard to resolve van Hove singularities on the real-energy axis prevent Eq. (C.5) to fully converge even with an immense number of energy points.

### 2.4.2 Transferability of the background charge

The background charge  $qn_{core}$  of the FBA model is determined for the intrinsic device. To verify the transferability of  $n_{core}$  to finite doping situations, the dependence

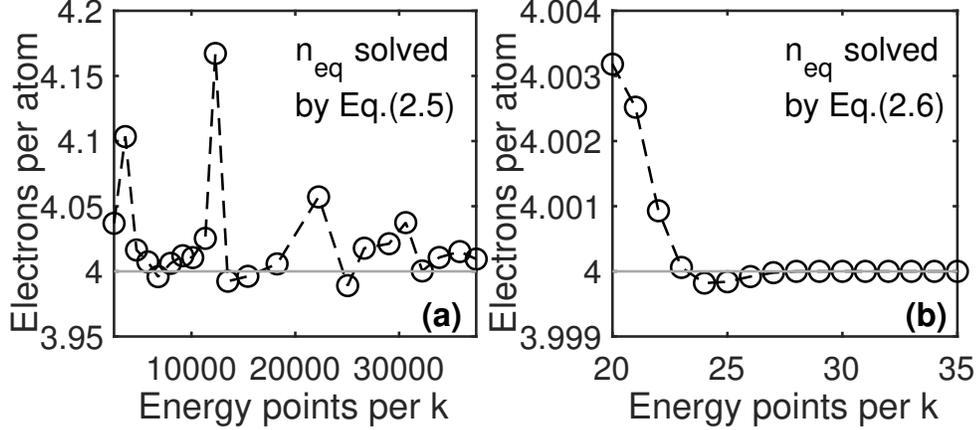


Fig. 2.2.: The atom resolved electron density  $n_{eq}$  vs number of energy points per momentum point for the real energy integral in Eq. (C.5) (a) and for the complex contour integral of Eq. (C.6) (b). The converged number of 4 valence Si electrons is indicated with a solid line to guide the eye. The  $n_{eq}$  converges with only 30 energy points in Eq. (C.6) while no convergence is observed when solved along the real energy axis even with an immense number of energy points.

of charge density on the effective Fermi level in the FBA model ( $-n_{FBA} + n_{core}$ ) is compared in homogeneous 3D silicon in equilibrium (see Fig. 2.3) with predictions of the ECA model ( $-n_{ECA}$ ). Screening in homogeneous silicon in equilibrium is well described with the material dielectric constant. Accordingly, both models agree excellently and the background charge of the FBA model applies over a large range of density variations. Deviation increases when the Fermi level is deep in the valence band or the conduction band, i.e. when the ECA model faces a lot of van Hove singularities and a converged solution of the real energy integral in Eq. (C.5) becomes numerically challenging.

### 2.4.3 Transfer characteristics of silicon MOSFETs

Transfer characteristics resulting from FBA and ECA charge self-consistent calculations of a p-type and a n-type silicon ultra-thin body double-gate MOSFET are

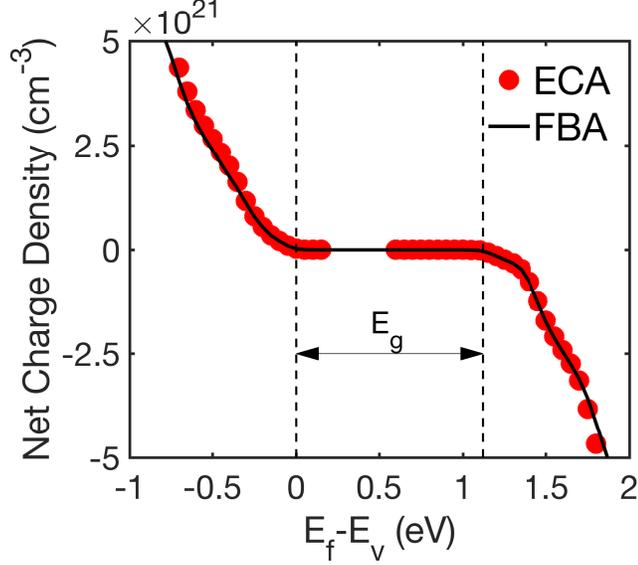


Fig. 2.3.: Variation of charge density with effective Fermi level calculated with both approaches. The two dashed lines indicate the valence (at 0) and conduction band edges. Deviation between ECA and FBA increases when the Fermi level is deep in the valence band or the conduction band. Then an increasing number of van Hove singularities plague the convergence of ECA.

shown in Fig. 2.5. Both MOSFET devices have a structure as shown in Fig. 2.4, with body thickness, channel and source/drain lengths of  $1.6\text{ nm}$ ,  $10.8\text{ nm}$  and  $11.4\text{ nm}$ , respectively. A doping concentration of  $1 \times 10^{20}\text{ cm}^{-3}$  is assumed in the source and drain regions of both transistor types. A drain-to-source voltage  $V_{DS}$  of  $0.4\text{ V}$  is applied. The equivalent oxide thickness (EOT) of top and bottom oxides is  $1\text{ nm}$ . It is reported both experimentally [32] and theoretically [3] that the dielectric constant of silicon ultra-thin films reduces with the film thickness. A dielectric constant of 9.9 is used in the ECA simulations following Ref. [3]. In FBA calculations, only the vacuum dielectric constant enters the Poisson equation.

Results of both models, for the transfer characteristics and band profiles agree very well. This is particularly true compared to TFET situations (see following subsections) that define the common ECA/FBA difference scale. The good agreement

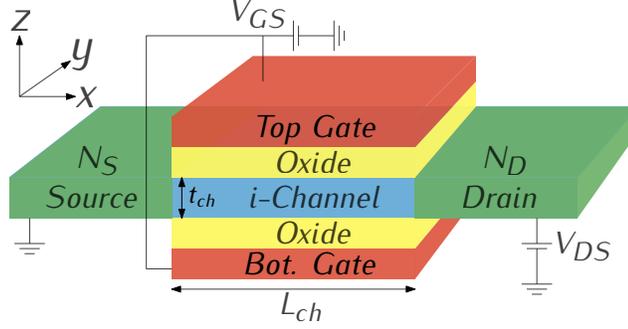


Fig. 2.4.: Schematic of the ultra-thin body double-gate transistors used in all transfer characteristic predictions of this work. The gate bias is controlled by  $V_{GS}$ . Electron transport occurs along  $x$  direction when a non-zero  $V_{DS}$  is applied. The channel of the device is confined along  $z$  direction and  $t_{ch}$  is the channel thickness. Periodic boundary condition is assumed along  $y$  direction.  $L_{ch}$  is the channel (gate) length.  $N_S$  and  $N_D$  are doping concentrations in source and drain regions, respectively.

is expected since electronlike and holelike states are clearly separated in these devices. The maximum relative difference of the drain current in the two models is below 15% and the maximum potential difference is below  $k_B T$ . Note that the dielectric constant of the ECA model could serve as a fitting degree of freedom to match the FBA results. This finding emphasizes the strength of the FBA model to explicitly handle the electrostatic response of the deep lying valence electrons - with a marginal increase in computational cost.

## 2.5 Application on TFETs

### 2.5.1 Silicon TFET

To compare the predictions of ECA and FBA for BTBT devices, the device structure of Fig. 2.4 is considered with the source (drain) doping being p-type (n-type). The source and drain doping concentrations are  $5 \times 10^{19} \text{ cm}^{-3}$  and  $2 \times 10^{19} \text{ cm}^{-3}$ , respectively. The body thickness, channel and source/drain lengths are  $1.6 \text{ nm}$ ,  $10.8 \text{ nm}$

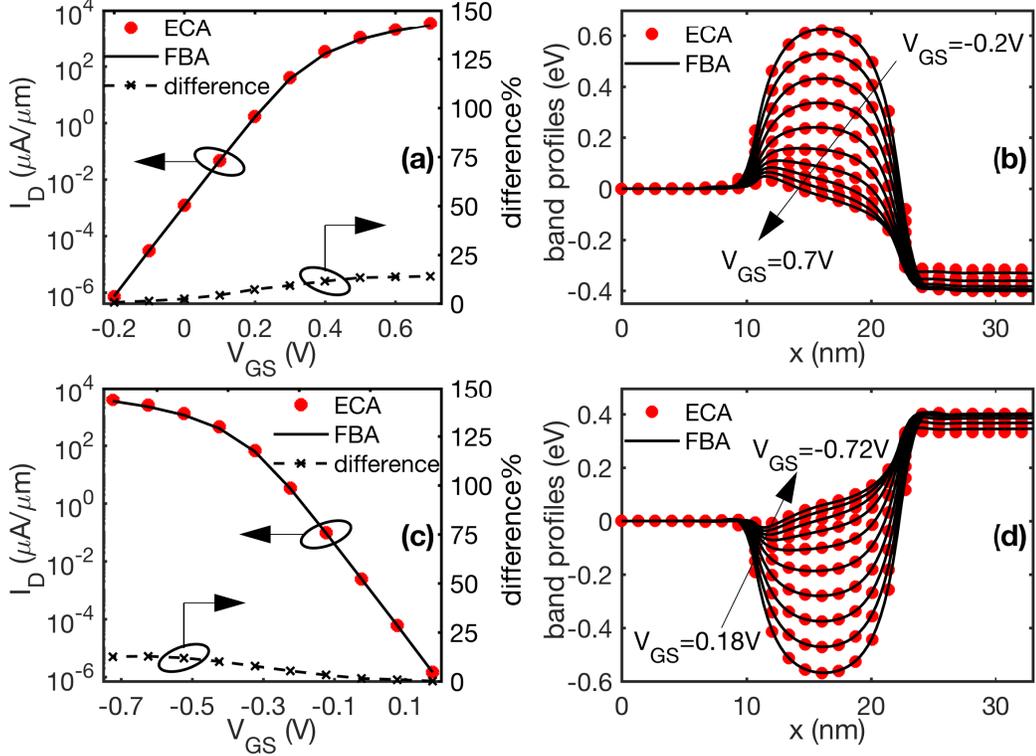


Fig. 2.5.: Transfer characteristics  $I_D - V_{GS}$  of silicon ultra-thin body double-gate nMOSFET (a) and pMOSFET (c) of Fig. 2.4 predicted with ECA and FBA. The percentage difference is plotted in dashed lines with cross markers. (b) and (d) Band profiles of ECA and FBA corresponding to  $V_{GS}$  nodes in (a) and (c), respectively.

and  $11.4\text{ nm}$ , respectively. A dielectric constant of 9.9 is used in the ECA model following Ref. [3]. The ECA utilizes heuristic models to distinguish electronlike and holelike charge contributions in BTBT devices. Three commonly used heuristic models (summarized in Table. 2.1)) are applied in ECA and results are presented and compared to the FBA result (see Fig. 2.6).

The difference in the performance predictions of ECA and FBA can be understood from Fig. 2.7. The electron density in the bandgap (at around  $10\text{ nm}$ ) is considered as electrons in FBA, whereas in ECA, a charge prefactor is assigned to it. This prefactor depends on the position of the hole/electron delimiter and the considered interpolation scheme (see Table. 2.1). A snapshot of the delimiter for  $\alpha = 0.5$  is illustrated in

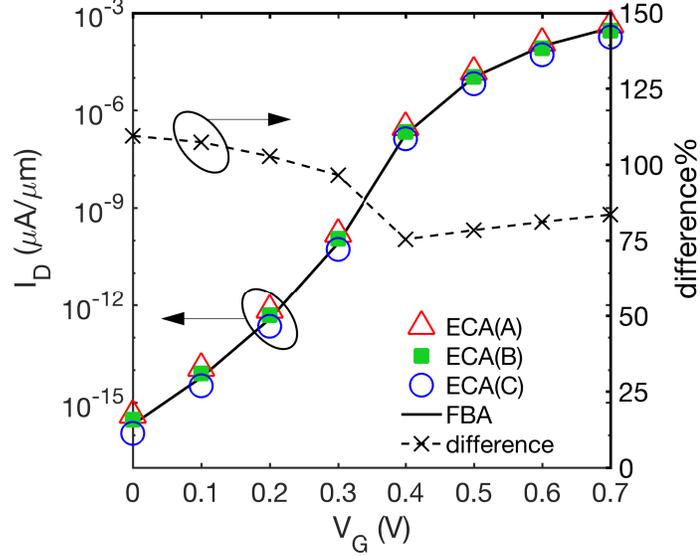


Fig. 2.6.: Transfer characteristics  $I_D - V_{GS}$  at  $V_{DS} = 0.3$  V of a silicon ultra-thin body double-gate TFET. Results of FBA and ECA with three different heuristic models are shown. The maximum deviation of the three ECA results relative to their average is plotted in dashed line with cross markers.

Fig. 2.7. Consequently, that prefactor differs in the three applied heuristic models. The different prefactors in turn impact the interpretation of ECA charge and the electrostatic potential around the tunnel junction. Thus, the TFET transfer characteristic prediction is sensitive to the chosen ECA delimiter model. This is indicated by the dashed line with cross markers in Fig. 2.6 which shows the maximum deviation among all ECA models relative to their average.

### 2.5.2 p-GaSb/n-InAs HJTFET

For completeness, the predictions of FBA and the three different ECA models are compared on a direct bandgap heterojunction TFET as well. The transfer characteristics of a double-gate p-GaSb/n-InAs heterojunction TFET with channel thickness  $t_{ch} = 3$  nm is shown in Fig. 2.8. The channel length of the considered TFET is  $L_{ch} = 20$  nm. The doping density in p-GaSb source and n-InAs drain regions are

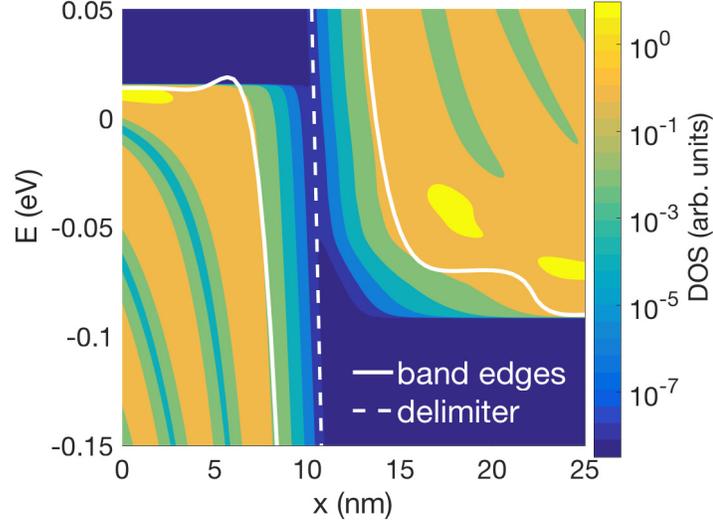


Fig. 2.7.: Contour plot of the energy and position resolved density of states of the silicon TFET simulated in Fig. 2.6 at  $k = \Gamma$ . The conduction and valence band edges are represented in white solid lines. The white dashed line depicts a hole/electron delimiter for  $\alpha = 0.5$  in the band gap which is used to distinguish electron and hole states in ECA.

$5 \times 10^{19} \text{ cm}^{-3}$  and  $2 \times 10^{19} \text{ cm}^{-3}$ , respectively. Short of dielectric constant assessments for GaSb and InAs UTBs, the respective material constants are used within ECA. A drain-to-source voltage  $V_{DS}$  of  $0.3 \text{ V}$  is applied. The equivalent oxide thickness (EOT) of top and bottom oxides is set to  $0.47 \text{ nm}$ .

The relative variation of the ECA predicted drain current densities decreases with increasing gate voltage. This can be understood with the comparison of band profiles for different gate voltages shown in the inset of Fig. 2.8. The effective tunneling distance, i.e. the spatial area in which ECA-specific models of Table 2.1 are applied, shrinks with increasing gate voltage. Then, differences in the ECA models become less relevant as well.

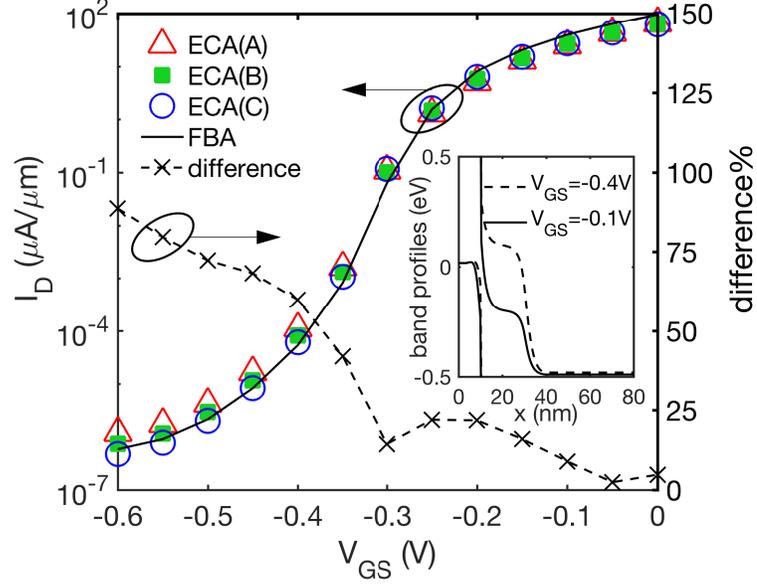


Fig. 2.8.: Transfer characteristics  $I_D - V_{GS}$  at  $V_{DS} = 0.3V$  of a double-gate p-GaSb/n-InAs heterojunction TFET with channel thickness  $t_{ch} = 3nm$  predicted with FBA and ECA with three different heuristic models. The maximum deviation of the three ECA results relative to their average is plotted in dashed line with cross markers. The inset shows band profiles in the middle of the TFET for two gate voltages ( $V_{GS} = -0.4V$  and  $V_{GS} = -0.1V$ ).

## 2.6 Incompatibility of ECA with NEGF

The following derivations show the ECA model is not applicable to the NEGF formalism since it violates fundamental NEGF equations. The retarded, advanced, lesser and greater Green functions are linearly dependent [52]

$$\mathbf{G}^R(\mathbf{E}, \mathbf{k}) - \mathbf{G}^A(\mathbf{E}, \mathbf{k}) = \mathbf{G}^>(\mathbf{E}, \mathbf{k}) - \mathbf{G}^<(\mathbf{E}, \mathbf{k}). \quad (2.9)$$

This equation holds for the full Green's function matrices, but for simplicity, the following derivations focus on the diagonals only. The real part of the spectral function  $A$  equals the density of states. It is defined as [27]

$$\begin{aligned} A(E, k) &\equiv i [G^R(E, k) - G^A(E, k)] \\ &= i [G^>(E, k) - G^<(E, k)]. \end{aligned} \quad (2.10)$$

In equilibrium, the fluctuation-dissipation theorem gives [52]

$$G^>(E, k) = -i[1 - f(E)]A(E, k), \quad (2.11)$$

$$G^<(E, k) = if(E)A(E, k). \quad (2.12)$$

Since the lesser (greater) Green's function is related to the electron (hole) density [44], the extension of the ECA interpolation of Eq. (C.1) to the NEGF framework would require Eqs. (2.11) and (2.12) to read

$$G^>(E, k) = -i[1 - f(E)]A(E, k)\lambda(E), \quad (2.13)$$

$$G^<(E, k) = if(E)A(E, k)[1 - \lambda(E)], \quad (2.14)$$

When Eqs. (A5) and (A6) are inserted into Eq. (A2), the spectral function has to fulfill

$$\begin{aligned} A(E, k) &= [1 - f(E)]A(E, k)\lambda(E) \\ &+ f(E)[1 - \lambda(E)]A(E, k). \end{aligned} \quad (2.15)$$

This requires

$$\lambda(E) - 2\lambda(E)f(E) + f(E) = 1, \quad (2.16)$$

which cannot hold true in general.

Note that altering the Green's function matrices beyond their diagonals shows another shortcoming of the ECA model. The two propagation coordinates of the Green's functions can yield very different interpolation values. It is unclear which value the nonlocal Green's function elements should be altered with.

## 2.7 Jacobian matrix for Newton iteration

The Jacobian matrix in Newton's method contains the derivative of the electron density over the electrostatic potential [53] which is difficult to calculate due to the

nonlocal response of electron density to changes in potential. We first assume that the derivative of the retarded Green's function over potential is negligible,

$$\begin{aligned} \frac{\partial n_{FBA}}{\partial \phi} &= \sum_k \int_{-\infty}^{\infty} \frac{-\text{Im} [G^R(E, k)]}{\pi} \frac{\partial f_L(E, \mu_L)}{\partial \phi} dE \\ &+ \sum_k \int_{-\infty}^{\infty} G^R(E, k) \frac{-\text{Im} [\Sigma^R(E, k)]}{\pi} G^A(E, k) \left[ \frac{\partial f_R(E, \mu_R)}{\partial \phi} - \frac{\partial f_L(E, \mu_L)}{\partial \phi} \right] dE \end{aligned} \quad (2.17)$$

then, we apply a local approximation which allows us to write

$$\begin{aligned} \frac{\partial n_{FBA}}{\partial \phi} &= \sum_k \int_{-\infty}^{\infty} \frac{-\text{Im} [G^R(E, k)]}{\pi} \frac{\partial f_L(E, \mu_L)}{-\partial(E/q)} dE \\ &+ \sum_k \int_{-\infty}^{\infty} G^R(E, k) \frac{-\text{Im} [\Sigma^R(E, k)]}{\pi} G^A(E, k) \left[ \frac{\partial f_R(E, \mu_R)}{-\partial(E/q)} - \frac{\partial f_L(E, \mu_L)}{-\partial(E/q)} \right] dE \\ &= \frac{\partial n_{eq}}{\partial \phi} + \frac{\partial n_{neq}}{\partial \phi} \end{aligned} \quad (2.18)$$

The same energy meshes used for electron density calculation can be used to calculate Eq. (2.18).

Note that  $\frac{\partial f_L(E, \mu_L)}{\partial E}$  in the equilibrium part of Eq. (2.18) has  $2^{nd}$  order poles at  $E_{pole} = \mu_L + ik_B T \pi(2m+1)$  with  $Res(E_{pole}) = 0$  where  $m \in \mathbb{N}$ . By Laurent expansion, we can write

$$\begin{aligned} F(E_{pole}) &= \left. \frac{\partial f_L(E, \mu_L)}{\partial E} \right|_{E=E_{pole}} \\ &= \sum_{n=-\infty}^{\infty} a_n (E - E_{pole})^n \\ &= \frac{a_{-2}}{(E - E_{pole})^2} + \frac{a_{-1}}{(E - E_{pole})} + a_0 + a_1 (E - E_{pole}) + \dots, \end{aligned} \quad (2.19)$$

where  $a_{-1} = 0$  according to Residue theorem and

$$a_{-2} = \lim_{E \rightarrow E_{pole}} [F(E_{pole})(E - E_{pole})^2] = k_B T. \quad (2.20)$$

The retarded Green's function is analytic in the upper complex plane, therefore, it can be written in Taylor series as

$$G^R(E, k) = G^R(E_{pole}, k) + \frac{G^R(E_{pole}, k)'}{1!} (E - E_{pole}) + \frac{G^R(E_{pole}, k)''}{2!} (E - E_{pole})^2 + \dots \quad (2.21)$$

Thus, to obtain the residue of the integrand in the equilibrium part of Eq. (2.18), we can write the Laurent series of the integrand from Eq. (2.19) and Eq. (2.21) and then calculate the coefficient  $a_{-1}$ . This gives us

$$\text{Res} \left\{ G^R(E, k) \frac{\partial f_L(E, \mu_L)}{\partial(E)} \right\} = G^R(E_{pole})' k_B T, \quad (2.22)$$

which requires numerically calculating the first derivative of the retarded Green's function at every pole enclosed in the contour.

It is critical to include the response of the equilibrium electron density to changes of potential in the Jacobian matrix. We observed that a Jacobian matrix assuming zero contribution from the equilibrium electron density delays the convergence by approximately 10 times iterations.

## 2.8 Summary

In this work, the FBA of Andlauer and Vogl is adapted to the framework of nonequilibrium Green's function based charge self-consistent quantum transport in atomically resolved nanotransistors. This model is compared to state of art ECA that heuristically interpret band-to-band tunneling states as partly electronlike and holelike. Significant ambiguity of the heuristic ECA is exemplified on various tunneling field effect transistors. The FBA lifts this ambiguity by explicitly including all valence band states and consistently interpreting them as electronlike only. For conventional MOSFETs that lack significant band to band tunneling, the ECA agrees with the FBA only if an appropriate screening constant for the valence electrons is applied. This constant is known to deviate from the respective material value. Since the FBA considers all valence electrons explicitly, no extra input of screening constants is required. In conclusion, the FBA provides a much wider application space than the conventional ECA - with a marginal increase in computational cost.

Ballistic transport is assumed in all transport simulations in this work. When scattering is included, the filling of states in the device is no longer in equilibrium with the source and drain contacts. Therefore, a close form for the Green's functions no longer

exist and instead they need to be solved self-consistently (e.g. via self-consistent Born approximation). A follow-on research can investigate the full-band approach with the inclusion of scattering. This will allow more realistic simulations of TFET since TFET performance is heavily affected due to non-ideality effects such as semiconductor/oxide interface traps, band tail states and other generation/recombination mechanics.

### 3. THERMAL BOUNDARY RESISTANCE PREDICTION WITH NEGF AND MD SIMULATIONS

At the time of the thesis deposition, Ref. [54] (arXiv) has been accepted by Applied Physics Letters and is under production. This chapter is reproduced from the accepted paper, with the permission of AIP Publishing.

#### 3.1 Motivation

Semiconductor nanodevices such as quantum cascade lasers, LEDs and thermoelectric devices are typically composed of several semiconductor materials [55–58]. Scattering of thermal energy carriers at the interface between two materials results in thermal boundary resistance [59]. The size of the thermal boundary resistance was previously reported [60] to be comparable to that of pure materials with lengths of a few to tens of nanometers. Predicting the thermal boundary resistance gives important insight into the device physics and enables design improvements. Often, molecular dynamics (MD) is used to model the thermal boundary resistance and reproduce experimental data [61]. Inelastic phonon scattering is included in MD simulations through the anharmonicity of the interatomic potential [62]. The non-equilibrium Green’s function (NEGF) method [27] is widely accepted as one of the most consistent methods for electronic quantum transport in nanodevices [30, 31]. In particular for predicting stationary device physics, NEGF is potentially more attractive than MD given that it is a spectral approach when setup in energy space, though modal methods in non-equilibrium MD have just begun to be developed [63, 64]. When electrons and phonons are both solved in the NEGF framework, interparticle interactions and energy and momentum transfer in e.g. self-heating or thermoelectric situations can be described on equal footing with the predictions of the respective particles’

propagation [65]. For phonon transport, however, the NEGF method has been used predominantly in the coherent (harmonic) regime due to the fact that the inclusion of incoherent scattering such as phonon-phonon decay usually requires solving polarization graphs in the self-consistent Born approximation which entails a large numerical load [66]. It had been shown that lack of anharmonicity in NEGF simulation give incorrect thermal boundary resistance predictions [67].

In this work, a numerically efficient method to solve phonon transport in the NEGF framework including phenomenological phonon scattering with Büttiker probes is presented and benchmarked against MD. Finite interface resistances in homogeneous structures that have plagued the Landauer approach [60, 68] is absent in the presented approach. When solved for homogeneous systems, this NEGF method yields vanishing interface resistance. The thermal boundary resistance calculated with this NEGF method shows quantitative agreement with MD simulations. The extracted spectral transport information from NEGF shows that the different phonon modal contributions play an important role in thermal transport across the interface.

### 3.2 The benchmark system

Fig. 5.1(a) shows the simulation domain considered in both MD and NEGF. The system consists of Si to the left and heavy-Si to the right of an interface at position 0. The heavy-Si differs from Si only in its atomic mass ratio vs. Si  $MR = M_{hSi}/M_{Si}$ . For all simulations in this work,  $M_{Si}$  is fixed at 28.085 u and a range of 1 to 10 is considered for  $MR$ . Transport is solved within a range  $L/2$  to the left and right of the interface. The lattice temperature for regions further to the left and right of the interface is assumed constant and to equal 320 K and 280 K, respectively. Phonon transport occurs along the  $x$  direction and the system is considered to be periodic along  $y$  and  $z$  directions. The harmonic phonon bandstructure is described with a Tersoff potential [69] and the lattice constant is set to 5.431 Å. Details of the thermal boundary resistance extraction are illustrated in Fig. 5.1(b). Firstly, linear

fits are performed on the local temperature profiles to the left and to the right of the interface, respectively. The local temperature  $T$  is obtained by minimizing the difference between the local phonon energy density and the product of a local Bose-Einstein distribution and the local phonon density of states [70]. For a given atom, the local temperature  $T$  solves the equation

$$\sum_{q_{\parallel}} \int_0^{\infty} \omega \rho(\omega) d\omega = \sum_{q_{\parallel}} \int_0^{\infty} \omega D(\omega) f_{BE}(\omega, T) d\omega, \quad (3.1)$$

where  $\omega$  is the phonon frequency and  $q_{\parallel}$  is the transverse phonon wave vector.  $\rho$  is the local phonon number density and  $D$  is the local phonon density of states. The temperatures  $T_1$  and  $T_2$  of the Si and heavy Si in the vicinity of the interface (see Fig.5.1(b)) are determined with the fitted temperature profiles. Finally, the thermal boundary resistance is calculated as  $R = (T_1 - T_2)/q$ , where  $q$  is the simulated heat flux. Following the discussion in Ref. [71], three different device lengths  $L$  are simulated for each value of  $MR$ , and  $R$  is extracted for the limit of  $1/L = 0$  by linear extrapolation.

### 3.3 Simulation approach

#### 3.3.1 MD

The LAMMPS package [72] is used for all MD simulations in this work. The lengths of both the heat source and the heat sink are  $L/10$  and the simulation timestep is 0.4 fs. Periodic boundary conditions are applied in  $y$  and  $z$  directions while the fixed boundary condition [73] is applied in the  $x$  direction. For MD calculations, the discretized device cross-section is of  $8 \times 8$  conventional unit cells, with each conventional cell containing 8 atoms. First, a canonical ensemble (NVT) is considered and run for 1.2 ns to relax the structure, allowing the system to reach thermal equilibrium at 300 K. The system is then switched to a microcanonical ensemble (NVE) and a constant heat flux is added to the heat source and extracted from the heat sink for 12 ns. After the system reaches steady state, the local temperature of each cell is

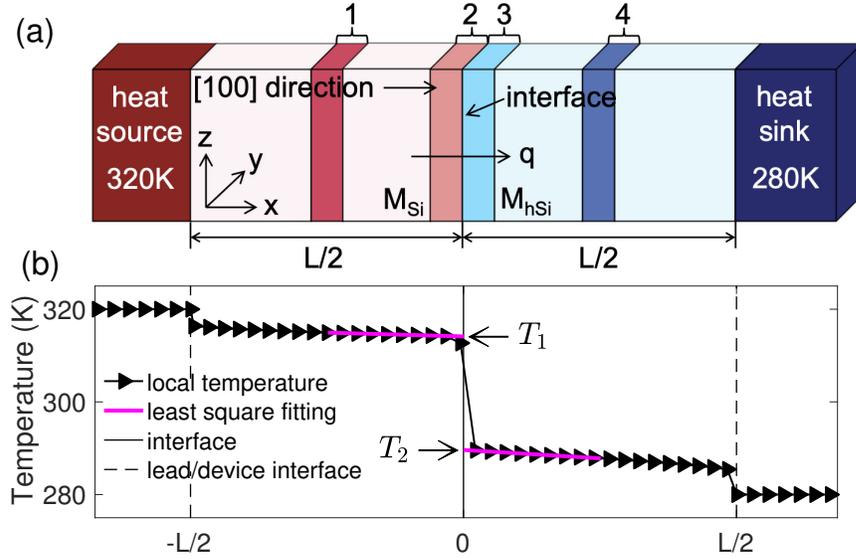


Fig. 3.1.: (a) Simulation domain considered in this work. Regions marked by 1 through 4 are three atomic layers in the middle of Si, left to the interface, right to the interface and in the middle of heavy Si, respectively. (b) Schematic of the thermal boundary resistance extraction.

obtained by averaging over ten million timesteps in the last 4 ns.  $L$  of 92, 130 and 184 unit cells is used for MD based simulation of each  $MR$  value.

### 3.3.2 NEGF

For all NEGF simulations, the nanodevice simulation tool NEMO5 [34] is used. Stationary Green's functions are solved in the energy domain, which gives spectral data without additional transformations. To calculate the harmonic interatomic force constants (IFCs), a  $3 \times 3 \times 3$  unit cell bulk Si structure is relaxed in LAMMPS using the Tersoff potential [69]. The derivatives of the forces between atoms with respect to the atom position variations give the harmonic IFCs. These values are then loaded into NEMO5 to construct the dynamical matrix [74]. Only the transport direction ( $x$ ) is discretized in real space. The periodic directions ( $y$  and  $z$ ) are represented with a single conventional unit cell. Longer-ranged periodicity is represented with

the phonon momenta in reciprocal space. Anharmonic phonon decay is included via Büttiker probes [75]. The Büttiker probe self-energies at atom  $i$  with vibrational direction  $m(x, y, z)$  are of the form

$$\Sigma_{BP(i,m)}^R(\omega) = -i \frac{2\omega\hbar^2}{\tau_{(i,m)}(\omega)}. \quad (3.2)$$

Each discretized atom in the system has a Büttiker probe applied to it. Following Ref. [70] we approximate the phonon frequency ( $\omega$ ) dependent scattering lifetime  $\tau$  as isotropic and assume it represents only the phonon-phonon Umklapp process [76]

$$\tau_{(i,m)}^{-1}(\omega) = \tau_i^{-1}(\omega) = BT_i\omega^2 e^{-C/T_i}. \quad (3.3)$$

$T_i$  is the phonon Büttiker probe temperature of the atom  $i$ . It represents the local temperature for the limit of complete phonon thermalization at each atom. The actual local temperature differs from the Büttiker probe one depending on the scattering strength [76]. To ensure energy conservation, the Büttiker probe temperature is solved iteratively by Newton's method until the integrated energy current vanishes for each Büttiker probe [70]. Since the phonon Green's functions are solved with the recursive Green's function (RGF) method [70], the device is partitioned into slabs perpendicular to transport direction. This limits the peak memory usage during computation, but requires the Büttiker probe self-energies to be equal throughout each slab. The parameters  $B$  ( $5 \times 10^{-20}$  s/K) and  $C$  (430 K) are chosen such that the NEGF prediction of the bulk Si thermal conductivity agrees with the MD solution (see Fig. 5.2(a)). The same values of  $B$  and  $C$  are used in Si and heavy-Si. In this way, the dependence of the scattering rate on the atomic mass and deviations of the anharmonicity near the interface from the one of the volume materials are ignored. This is not a fundamental limitation of the Büttiker probes. Future work on other materials and interfaces might likely require a more detailed Büttiker probe model.  $L$  of 146, 184 and 220 unit cells is used for NEGF based simulation of each  $MR$  value.

The spatial distribution of local phonon density of states (LDOS) can illustrate the contributions of different phonon modes to the heat flux in various regions. For

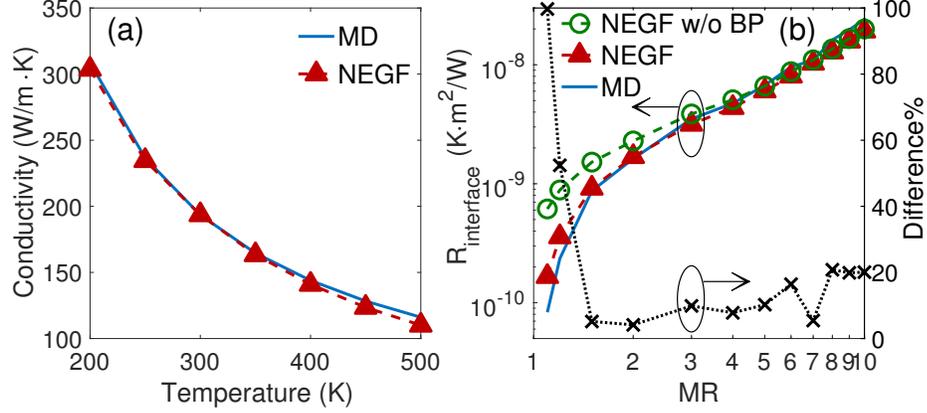


Fig. 3.2.: (a) Thermal conductivity of bulk Si calculated with MD and NEGF. The Büttiker probe parameters  $B$  and  $C$  are fitted such that the transport results of NEGF agree with MD. (b) Thermal boundary resistance as a function of the mass ratio (MR) calculated by NEGF with the fitted parameters  $B$  and  $C$  of (a) and MD. The error bar of MD simulation is calculated but not plotted since the standard deviation for independent simulations is only 1.6% to 4.6% of the average resistance at the Si/heavy-Si interface for different mass ratios. The dotted line is the relative difference of the two methods and is defined as  $(NEGF - MD)/MD \times 100\%$ . The green curve shows the results calculated by NEGF without Büttiker probe self-energies. With increasing masses, the maximum phonon frequency and (with the frequency dependence of the Büttiker probes, Eq. B.2) the average scattering strength reduces. Consequently, the impact of scattering on  $R_{interface}$  declines with increasing mass of heavy-Si.

this purpose, four regions of interest, consisting of three atomic layers each are defined in Fig. 5.1(a). The LDOS  $\phi_i(\omega)$  of any region  $i$  of these regions is averaged over all of its atoms. For a given phonon frequency, the L1-norm of the LDOS summed over all regions is defined as

$$\|\phi_{tot}(\omega)\| = \sum_{i=1}^4 \phi_i(\omega). \quad (3.4)$$

The relative contribution of LDOS of region  $i$  is then defined as

$$R_i(\omega) = \frac{\phi_i(\omega)}{\|\phi_{tot}(\omega)\|}. \quad (3.5)$$

### 3.4 Thermal boundary resistance: NEGF versus MD

Fig. 5.2(b) shows the thermal boundary resistance as a function of  $MR$  calculated with NEGF agrees quantitatively with the MD predictions. For both methods, the thermal boundary resistance increases exponentially with  $MR$  and vanishes when  $MR$  tends to unity (i.e. for a homogeneous system). Small remaining difference between the MD and NEGF results are systemic to the different treatment of phonon modes perpendicular to the transport direction: In NEGF, the phonon momentum perpendicular to transport is explicitly resolved as a parameter in the Dyson and Keldysh equations [70]. In contrast, MD calculations require as large as possible unit cells perpendicular to transport to cover as many phonon modes with long wave lengths in these directions as feasible. Another source of differences can be the open boundary condition in transport direction: Fig. 5.3 benchmarks the open boundary condition treatment in MD and NEGF, since they illustrate finite-size effects [61] with the thermal boundary resistance  $R$  as a function of the inverse system length  $1/L$ . For MD, the slope of  $R$  vs  $1/L$  increases with the mass ratio  $MR$ , i.e. with reducing average sound velocity in the device. This agrees with similar findings discussed in detail in Ref. [61]. Note that the mean free path of Si at room temperature is about 300 nm [77]. Therefore, it is practically impossible to completely eliminate the size effect along the transport direction in MD due to computational cost. In contrast to the finite sized boundary reservoirs of MD, contact self-energies in NEGF incorporate semi-infinite leads as phonon reservoirs [78]. Accordingly, we observe the slope of  $R$  vs.  $1/L$  in NEGF predictions is much smaller than that of MD. Also its increase with  $MR$  is comparably negligible. This different boundary treatment is another source of some differences between MD and NEGF seen in Fig. 5.2 (b).

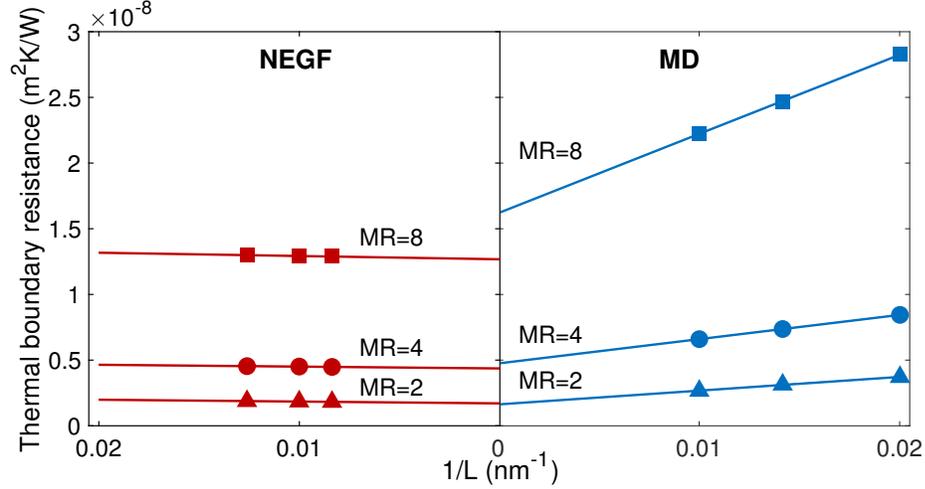


Fig. 3.3.: Linear extrapolation of NEGF and MD for three  $MR$  values. MD results show a stronger dependence on the length of the device. Open boundary conditions included in NEGF with contact self-energies give almost device size independent results.

### 3.5 Inelastic scattering in NEGF

Fig. 5.4(a) shows the phonon density of states (DOS) in homogeneous Si and heavy-Si solved with NEGF. The DOS of heavy-Si is limited to energies at or below 33.3 meV, two times less than in native Si in agreement with the applied  $MR = 4$ . Incoherent phonon scattering allows phonons with energies above this heavy-Si cutoff energy to propagate. This is illustrated in Figs. 5.4(b), (c) and (d) with the energy resolved current densities of the heat source and heat sink of the device in Fig. 5.1(a) when solved in NEGF with different scattering strengths. While the energy distribution changes with scattering, the total energy current (on the order of  $10^9$  W/m<sup>2</sup>) is conserved, and the difference of the total energy current between the source and the sink is on the order of 1 W/m<sup>2</sup>. In Fig. 5.4(b), normal scattering strength (fitted to reproduce the MD-calculated bulk thermal conductivity) is applied. The source current with energies above the energy cutoff is finite since its corresponding phonons can relax to lower energies at the heavy-Si side via inelastic scattering. In Fig. 5.4(c),

an artificially weak scattering strength is used ( $1/20\times$  normal scattering strength). Accordingly, the profiles of current in both the heat source and sink follow the profile of the heavy-Si DOS. In contrast, Fig. 5.4(d) shows the NEGF results when artificially strong scattering is used ( $20\times$  normal scattering strength). The current profiles in the heat source and heat sink follow the profiles of the DOS in the Si and the heavy-Si leads, respectively. The results show that stronger inelastic scattering brings the system closer to local thermodynamic equilibrium.

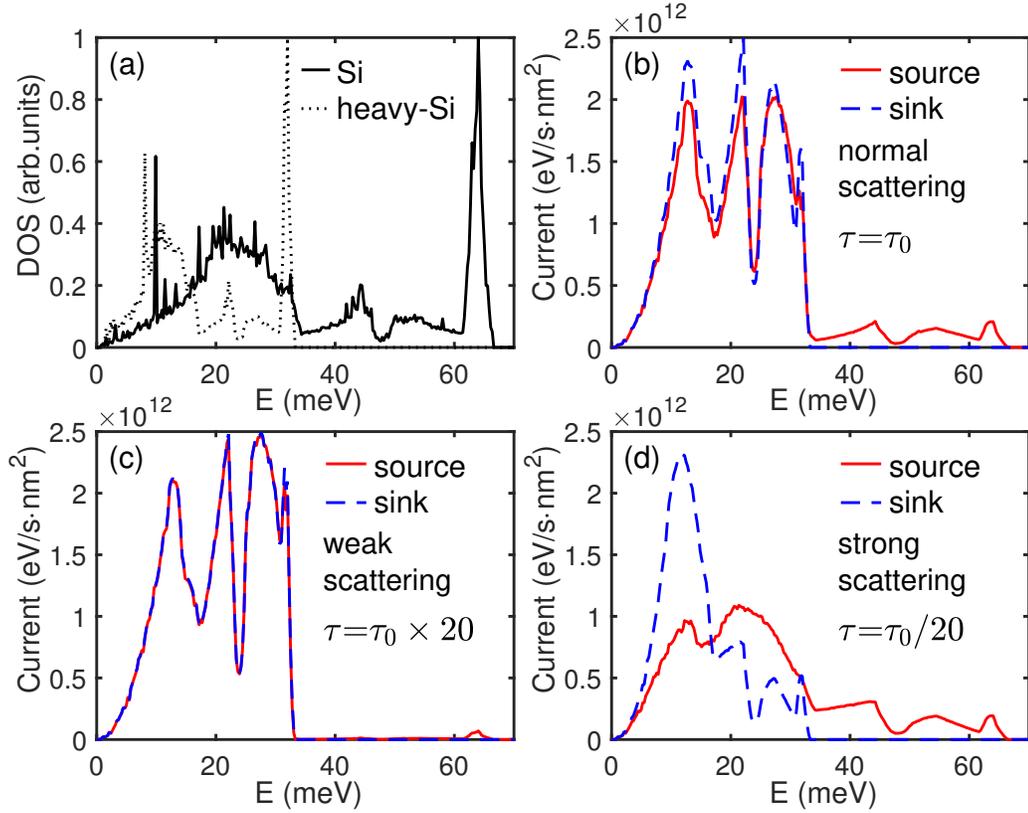


Fig. 3.4.: (a) Energy resolved (transverse momentum integrated) density of states in Si and heavy-Si leads. (b)~(d) Energy resolved (transverse momentum integrated) current of NEGF in the heat source and sink calculated with normal, artificially weak ( $1/20\times$ ) and artificially strong ( $20\times$ ) scattering strengths, respectively.

### 3.6 Role of phonon scattering in energy current across interface

In Fig. 5.4(b), four current peaks in the heavy-Si heat sink are located at 12 meV, 22 meV, 27 meV and 32 meV, respectively. They correspond to the four peaks of the heavy-Si DOS shown in Fig. 5.4(a). The relative magnitudes of the four current peaks do not follow the relative magnitudes of the four DOS peaks (i.e., the list of current peaks arranged in decreasing order of magnitude is  $22 \text{ meV} > 12 \text{ meV} > 27 \text{ meV} > 32 \text{ meV}$ , whereas the same list according to the DOS magnitude is  $32 \text{ meV} > 12 \text{ meV} > 22 \text{ meV} > 27 \text{ meV}$ ). Without interfaces involved, the current is expected to be proportional to the product of the DOS and the group velocity [79]. The results of Figs. 5.4 can be understood in Fig. 5.5, since it illustrates the different relative contributions of phonon modes in different device areas. Phonon modes around 12 meV reside nearly exclusively in the heavy-Si. Consequently, they contribute less to the overall heat current. In contrast, phonon modes around 22 meV are present in all 4 device regions considered in Fig. 5.5. Therefore, these modes can maintain a higher contribution to the total heat current.

### 3.7 Summary

In conclusion, NEGF with Büttiker probe scattering is applied to the thermal boundary resistance of the Si/heavy-Si interface. The empirical NEGF scattering parameters are tuned to reproduce the thermal conductivity of homogeneous Si predicted by MD. The scattering parameters proved to be transferable, since the NEGF results of the thermal boundary resistance quantitatively agrees with MD results for mass ratios ranging from 1 to 10. Finite interface resistances in homogeneous structures that have plagued the Landauer approach is absent in the presented approach. Besides, the present NEGF approach is found to be numerically more efficient than MD. Thanks to the open boundary conditions, NEGF shows virtually no finite-size effects compared to MD. The analysis of the NEGF spectral information shows that the scattering between different phonon modes determines the phonon energy current

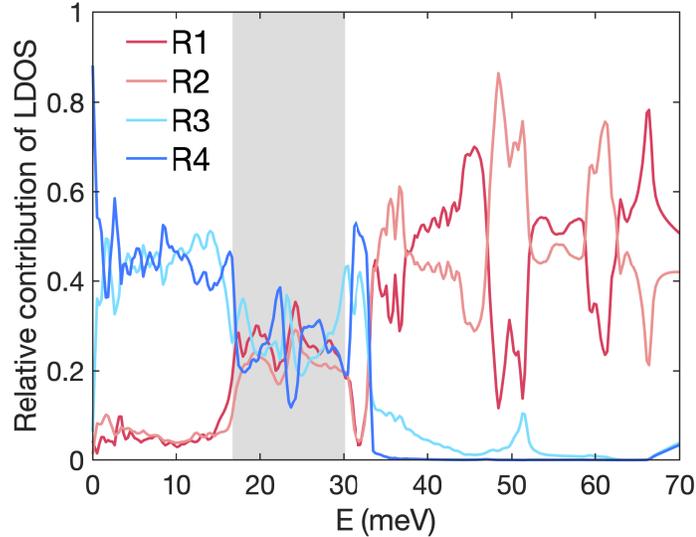


Fig. 3.5.: The ratio of the phonon mode in regions 1  $\sim$  4 defined in Fig. 5.1 and by Eq. B.4 ( $\sum_{i=1}^4 R_i = 1$ ). The evenly distributed phonon modes around 22 meV (marked by the gray area) result in the highest current peak at 22 meV (see Fig. 5.4(b)) although having a lower DOS and lower group velocity compared to those at 12 meV.

flow across interfaces. Future improvements of our approach can include a more accurate description of inelastic scattering at the interface and enable computationally efficient calculation of spectral current adjacent to the interface.

## 4. 5 nm GATE LENGTH GAN NANOWIRE NFET

© 2019 IEEE. This chapter is reprinted, with permission, from Ref. [80].

### 4.1 Motivation

As we approach the scaling limits of FinFET structure, gate-all-around structure is being investigated for future nodes due to its excellent electrostatic control [81]-[82]. Intrinsic properties of GaN, such as its wide band gap ( $\sim 3.4$  eV) and large saturation velocity ( $\sim 2 \times 10^7$  cm/s [83]), make it an intriguing option for gate-all-around n-MOSFET devices for digital applications beyond 7 nm technology, complementing the well-known markets of GaN for RF [84] and power applications [85]. It is reported [86], through simple effective-mass-based ballistic quantum transport simulations, that 5 nm channel length GaN-NW-nFET (square cross-section with  $T_{nw} = 1.6$  nm) outperforms Si ( $m_{eff} = 0.34m_0$ ), Ge ( $m_{eff} = 0.18m_0$ ), and InAs ( $m_{eff} = 0.153m_0$ ) channel materials due to its relatively lower electric permittivity and higher effective mass ( $m_{eff} = 0.37m_0$ ) which reduces source-drain direct tunneling leakage. Accurate performance prediction of sub-10 nm nanowire transistors requires atomistic resolution [87–90]. In this work, the performance of GaN-NW-nFET with different geometrical shapes, cross-section widths, and gate configurations is investigated using atomistic full-band quantum transport simulation. Results are compared with simulated 5 nm Si-NW-nFET.

### 4.2 Simulation approach

All results of this letter were obtained using the nanodevice simulation tool NEMO5 [34]. The electronic structure is represented in atomistic nearest-neighbor tight-binding

(TB)  $sp3d5s^*$  basis. Throughout the paper, empirical TB parameters for Si and GaN are taken from [35] and [34], respectively. For all nanowire devices the electron transport direction ( $x$ ) and the confinement directions ( $y$  and  $z$ ) are atomically discretized. All results are obtained by iterating the quantum transmitting boundary method [91] with the 3D nonlinear Poisson equation until convergence is achieved. Transistor benchmarking with a fixed  $I_{OFF}$  was performed by tuning the metal work function, i.e., flat-band voltage ( $V_{FB}$ ), while keeping all other device parameters the same. Because of the high optical phonon energy ( $\sim 93$  meV) [92], wide valley separation ( $> 1$  eV) and large mean free path ( $\sim 14$  nm) of GaN, ballistic transport is expected in sub-10 nm regime [93]-[94] and is assumed in all calculations, given that the channel length considered in this work is 5 nm. Ballistic transport predicts idealized device performances. To give an idea how detrimental the effect of scattering might be, surface roughness is implicitly included in the energy- and intrinsic gate-delay calculation by assuming a 35% reduction in  $I_{ON}$  in accordance with [95].

### 4.3 Results and discussions

#### 4.3.1 Device structures

Fig. 4.1 shows a schematic of the devices considered in this work. Square, circular, triangular and hexagonal cross-sections with widths varying from 1.6–2.8 nm are simulated. These widths are selected to ensure immunity to short-channel effect (design rule:  $T_{nw} < L_G/2$ ). A summary of the simulated device parameters is provided in Fig. 4.1(e). The transport directions for all GaN and Si nanowire simulations are taken along the [0001] ( $c$ -axis) and [100] directions, respectively.

#### 4.3.2 Benchmarking GAA GaN with Si NW-nFET

The assessment reported in [86] is repeated to examine how GaN compares to Si NW-nFET with the use of full-band quantum transport. The calculated energy

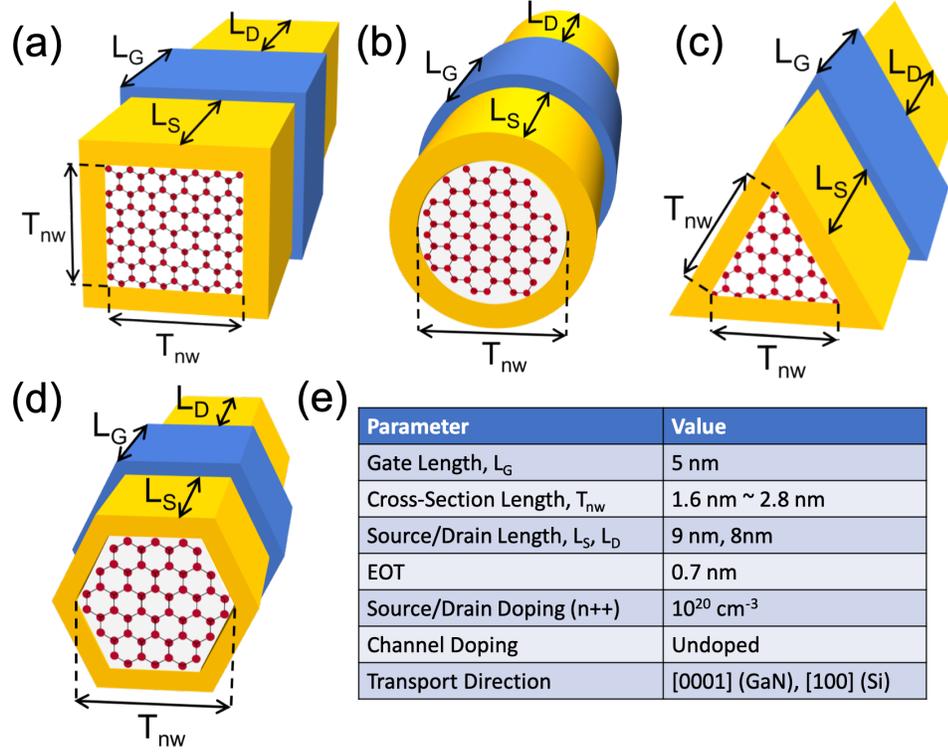


Fig. 4.1.: (a)~(d) Schematic of the GaN Nanowire n-MOSFET (GaN-NW-nFET) with different cross-section shapes simulated in this work and (e) important device parameters.

dispersion relations of Si and GaN nanowire with square cross-sections are shown in Fig. 4.2(a). Only conduction bands are shown and the minimum conduction band energy is fixed at 0. Fig. 4.2(c) compares GaN and Si square NW-nFET ON-state drive current at 2 square cross-sections:  $T_{nw} = 1.6 \text{ nm}$  and  $2.0 \text{ nm}$ . It is observed – consistent with [86] – that GaN NW-nFET outperforms Si in both LP and HP applications. The GaN-NW-nFETs show  $\sim 13\%$  ( $30\%$ ) and  $\sim 33\%$  ( $32\%$ ) better  $I_{ON}$  at  $I_{OFF} = 1 \text{ nA}/\mu\text{m}$  ( $100 \text{ nA}/\mu\text{m}$ ) than Si at  $T_{nw} = 1.6 \text{ nm}$  and  $2.0 \text{ nm}$ , respectively. GaN's higher current is attributed to its smaller effective mass,  $m_{eff}$  ( $I \propto \frac{Q_{inv}}{\sqrt{m_{eff}}}$ ). The ON/OFF performance metric  $Q = g_m/SS$  ( $Q_{GaN} = 335.07 \mu\text{S-dec/mV}$  for  $T_{nw} = 1.6 \text{ nm}$ ,  $181.73 \mu\text{S-dec/mV}$  for  $T_{nw} = 2.0 \text{ nm}$  and  $Q_{Si} = 251.33 \mu\text{S-dec/mV}$  for  $1.6 \text{ nm}$ ,

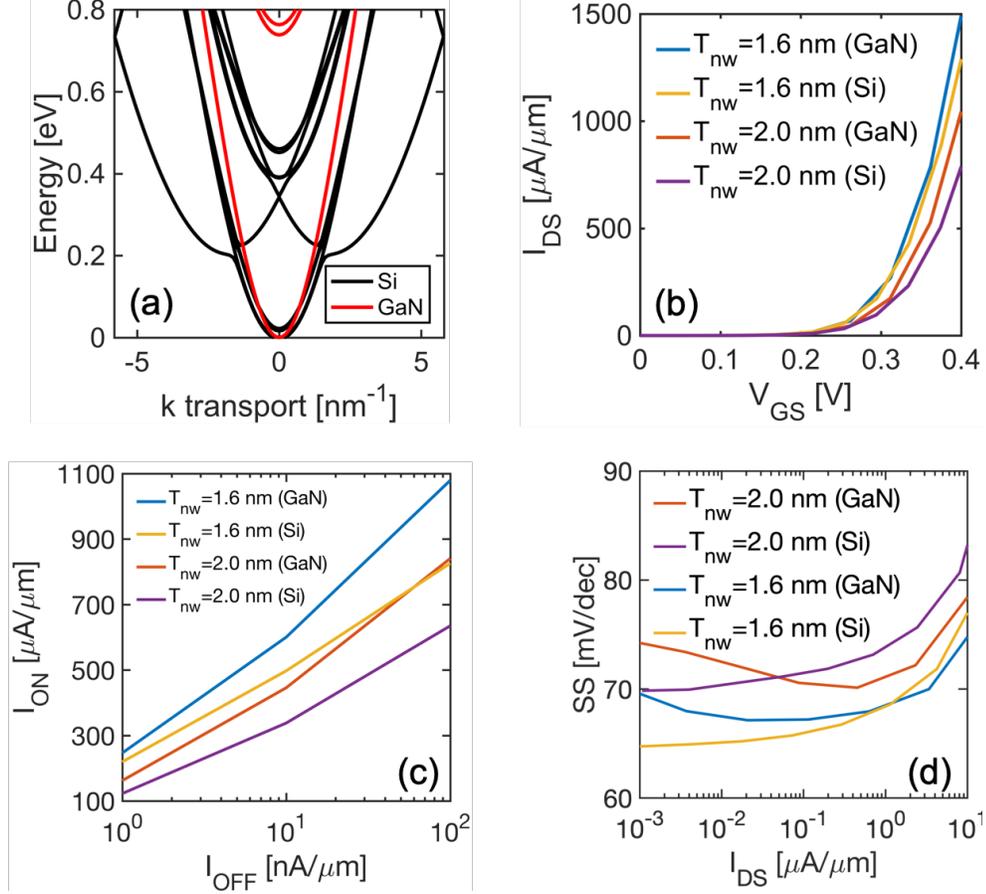


Fig. 4.2.: (a) Energy dispersion relations along the transport direction of Si and GaN square cross-section nanowire ( $T_{nw} = 1.6$  nm). (b)~(d)  $I_D$  vs  $V_G$  curves ( $I_{OFF}$  fixed at  $10^{-2}$   $\mu\text{A}/\mu\text{m}$ ),  $I_{ON}$  vs  $I_{OFF}$  curves (calculated at  $V_{DD}=0.4$  V) and Subthreshold swing vs  $I_{DS}$  curves of GaN and Si NW-nFETs with square cross-section shape.

154.86  $\mu\text{S-dec}/\text{mV}$  for  $T_{nw} = 2.0$  nm), is about 33% higher for  $T_{nw} = 1.6$  nm and 17% higher for  $T_{nw} = 2.0$  nm in GaN compared to Si at  $I_{OFF} = 10$   $\text{nA}/\mu\text{m}$ . At  $T_{nw} = 1.6$  nm (2.0 nm), the SS is 68 mV/dec (73 mV/dec) and 65 mV/dec (70 mV/dec) for GaN and Si, respectively (see Fig. 4.2(d)). The non-ideal SS mainly originates from source drain direct tunneling leakage which is worse for GaN due to its lower effective mass –  $m_{eff}$  for GaN and Si channel with  $T_{nw} = 1.6$  nm (2.0 nm) are  $0.26 m_0$  ( $0.25 m_0$ ) and  $0.43 m_0$  ( $0.35 m_0$ ).

### 4.3.3 Gate efficiency and current normalization

For large cross-sections, one expects that a multigate device with  $N$  gates to have  $N$  independent channels with an effective width given by the gate perimeter ( $N \times W_{gate}$ ). However, this independent channel assumption does not hold at small cross-sections. To illustrate this point, the relative  $I_{ON}$  gain of gate-all-around (GAA) GaN square NW-nFET with respect to tri-gate (TG) ( $I_{GAA}/I_{TG}$ ) and double-gate (DG) ( $I_{GAA}/I_{DG}$ ) as  $T_{nw}$  scales is shown in Fig. 4.3. Note that the relative gain is well below the ideal gain (33% and 100% relative to TG and DG, respectively). This is especially true for the smallest cross-section simulated ( $T_{nw} = 1.6$  nm) where the GAA enhancement with respect to TG and DG drops to  $\sim 15\%$  and  $\sim 35\%$ , respectively. The reason for the gain reduction – which is generally true for Si [96] and

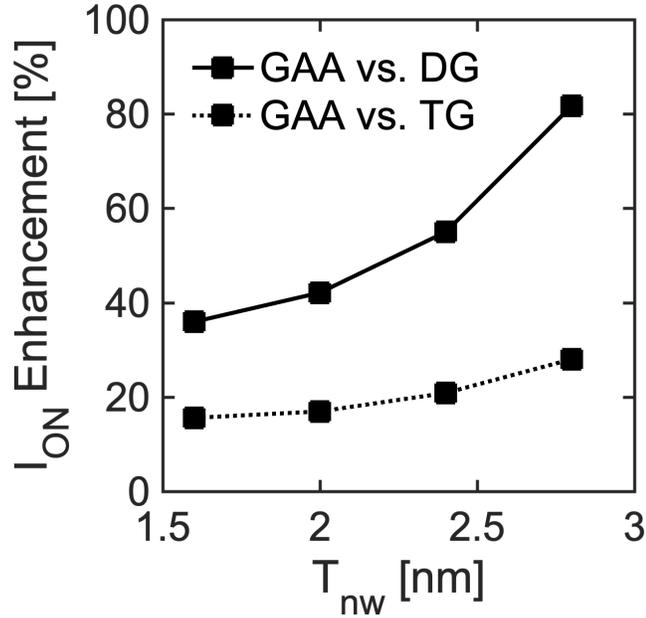


Fig. 4.3.: Percentage of  $I_{ON}$  enhancement of gate-all-around (GAA) GaN-NW-nFET with respect to tri-gate (TG) and double-gate (DG) as a function of  $T_{nw}$ .

other channel materials – is due to volume inversion, which forms a single channel instead of  $N$  independent channels. Consequently, the widely adopted gate-perimeter normalization standard is not valid at such small cross-sections. This issue is high-

lighted in Fig. 4.4(a). It is observed how such normalization misleadingly suggests that  $I_{ON}$  per gate is most efficient for the structure with the least number of gates. This can be corrected by using areal normalization (Fig. 4.4(b)) which clearly reveals the advantage of using more gates.

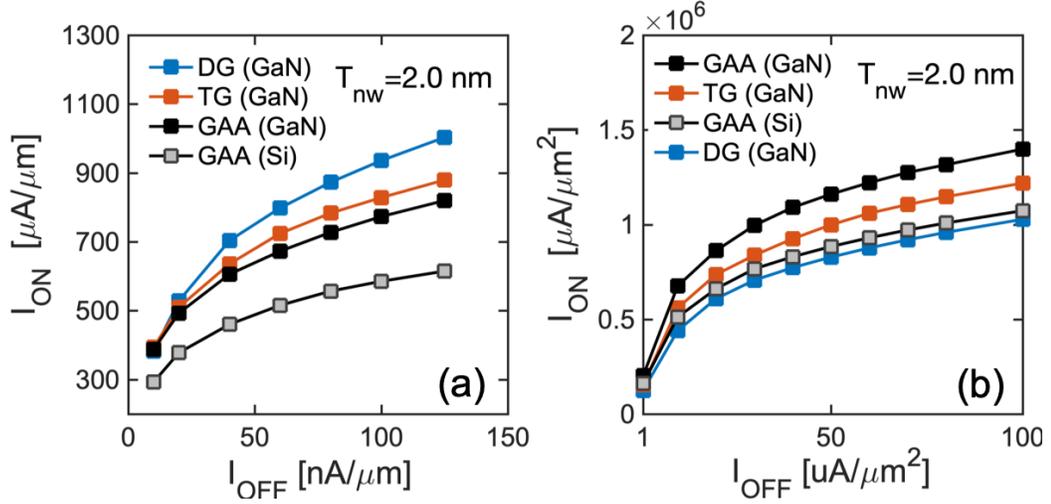


Fig. 4.4.:  $I_{ON}$  vs  $I_{OFF}$  curves (calculated at  $V_{DD}=0.4$  V) (a) normalized with respect to the perimeter and (b) with respect to the cross-section area of square cross-section GaN-NW-nFETs with different gate configurations and  $T_{nw}=2.0$  nm.

#### 4.3.4 Influence of geometrical shape

In sec. 4.3.2, it is demonstrated that the source drain direct tunneling leakage degrades the performance of 5 nm gate length GaN-NW-nFET with a square cross-section. It is also shown in sec. 4.3.3 that at the limit of small cross-sections, the relative  $I_{ON}$  enhancement by adding more gates diminishes. In this section, the goal is to understand how the geometric shape of the GaN-NW-nFET affects its performance. The study of different geometric shapes is especially relevant to GaN since GaN's wurtzite crystal structure has a hexagonal lattice rather than a cubic lattice. As such, the device structure presented thus far and that of Ref. [86], albeit being promising, is difficult to fabricate in reality, due to the large number of dangling bonds

present in square GaN cross-section. Methods for dangling bond passivation within tight binding in zincblende [97] and any other geometry [98] have been developed and implemented in NEMO5. Recent experimental work has reported that GaN nanowire of triangular shape is easier to construct experimentally than other shapes [99,100]. In this section, the performance of square, circular, triangular and hexagonal GaN-NW-nFETs are compared [39,90,101]. The circular, triangular and hexagonal shape devices were selected to assess a more realistic GaN nanowire geometry. Fig. 4.5(a) shows the  $I_{ON}$  vs  $I_{OFF}$  for the 4 geometric shapes considered with  $T_{nw} = 2.4$  nm. It is observed that the circular, triangular and hexagonal devices all have higher raw current than the square geometry despite their smaller cross-sectional area. The superior performance of circular, triangular and hexagonal cross-sections can be attributed to relatively larger effective mass (see Fig. 4.5(b)) which suppress the source drain direct tunneling leakage. A similar trend in effective mass is also observed in first principle [102] and DMol [103] calculations, with effective mass increasing as the nanowire diameter decreases. The triangular device shows the highest  $G_m$  and the lowest  $SS$  among all four geometries for the same  $T_{nw}$ , which effectively allows better cross-sectional scaling than the square device. This can be illustrated (see Table. 4.1) by comparing a square device and a triangular device of around the same area (e.g., a triangular device with  $T_{nw} = 2.4$  nm has cross-section area  $2.5 \text{ nm}^2$ ; a square device with  $T_{nw} = 1.6$  nm has cross-section area  $2.56 \text{ nm}^2$ ). It is noted that Si in the [111] plane can also be fabricated in a triangular shape. However, it is reported [104] that Si n-type NW along [111] has a much larger effective mass than [100]. A larger effective mass is good for suppressing source-drain direct tunneling leakage, but it also lowers the injection velocity and the drive current. Lastly, the energy-delay vs intrinsic gate-delay is summarized in Fig. 4.6. A reduction of 35% [95] in  $I_{ON}$  is assumed for simulated GaN and Si devices to account for surface roughness. Overall, it is observed that the GaN-NW-nFETs outperform the simulated Si NW-nFETs.

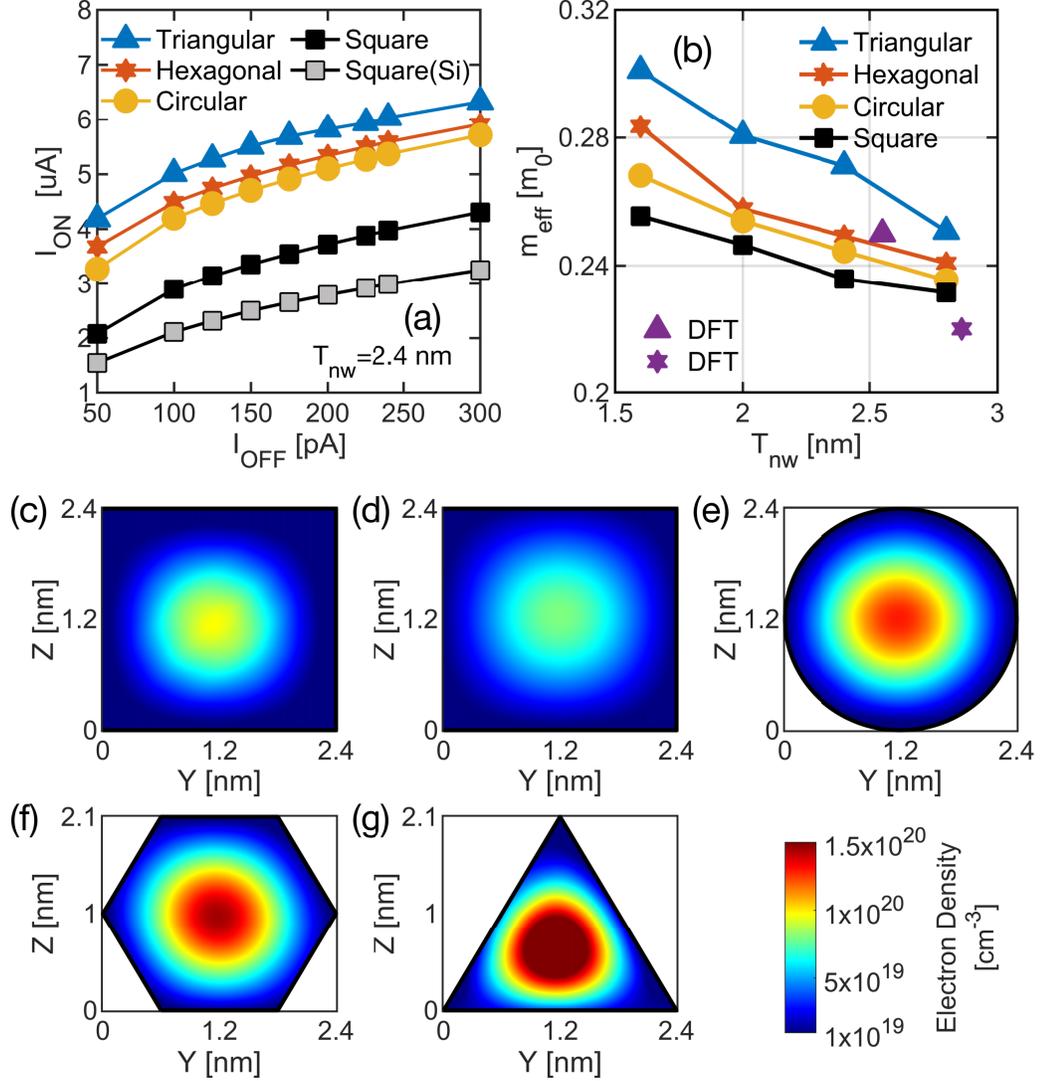


Fig. 4.5.: (a)  $I_{ON}$  vs  $I_{OFF}$  (calculated at  $V_{DD} = 0.4$  V) of GaN-NW-nFETs with different geometrical shapes. (b) GaN electron effective mass ( $m_{eff}$ ) calculated using the  $sp3d5s^*$  TB basis. Purple symbols are DFT calculations from Ref. [102]. (c)~(g) Cross section electron density profiles at top of the barrier for Si (c) and GaN ((d)~(g)) with different geometrical shapes at  $V_{GS} = V_{DD} = 0.4$  V ( $I_{OFF}$  fixed at 200 pA).

#### 4.4 Summary

This work represents a full band quantum transport simulation study of GaN-NW-nFET with different cross-sectional shapes, assessing the promise of GaN for digital

Table 4.1.: Performance benchmark of square Si-, GaN- and triangular GaN-NW-nFETs of comparable cross-sectional area.

Performance metrics	Si square 1.6 nm	GaN square 1.6 nm	GaN triangular 2.4 nm
SS(mV/dec)	65.7	67.2	63.4
Gm( $\times 10^6 \mu\text{S}/\mu\text{m}^2$ )	4.3	5.5	6.2
$I_{ON}/I_{OFF}(\times 10^6,$ $I_{OFF}=100 \mu\text{A}/\mu\text{m}^2)$	1.7	2.2	2.5

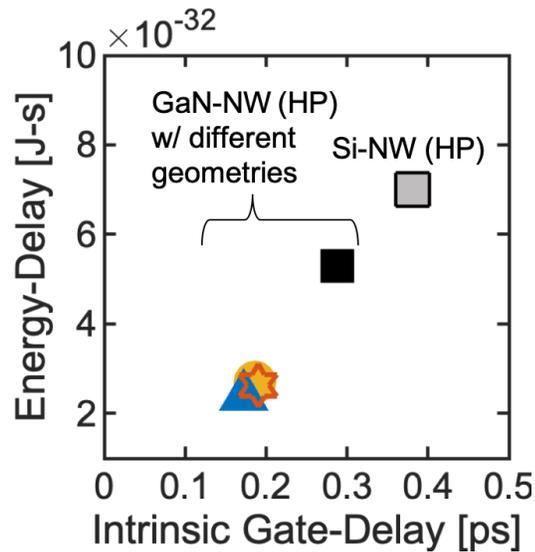


Fig. 4.6.: Energy-delay vs intrinsic gate-delay of the simulated GaN- and Si- NW-nFETs. Cross-section shape of GaN-NW is indicated by symbols: square (black), circular (yellow), hexagonal (red) and triangular (blue). The simulated devices account for a 35% current reduction estimate due to surface roughness [95].

logic applications beyond the 7 nm node. Our study reveals that GaN nanowire FETs are a strong contender for n-channel devices at the 5 nm gate length. It is found that

GaN-NW-nFET offers device performance improvement over Si-NW-nFET for both LP and HP applications, especially with the use of triangular geometry.

## 5. MONOLAYER $\text{WTe}_2$ - $\text{ZrS}_2$ VDW TFET

### 5.1 Motivation

In a conventional MOSFET, the current is determined by the thermionic emission of charge carriers over the gate-controlled potential barrier between source and drain [105]. The equilibrium thermodynamics defines the fundamental lower limit of the subthreshold swing ( $SS$ ) to be 60 mV/dec at room temperature [106], which precludes the reduction of the supply voltage and the overall power consumption of the integrated circuit [107]. Exploring novel device architectures, with alternative switching mechanisms which allow voltage scaling, is critical for a plethora of emerging device applications such as internet-of-things, wearable electronics and artificial intelligence. Band-to-band tunnel field-effect transistors (TFETs) are promising candidates for low power electronics due to their potential to surpass the thermionic limit and achieve steep  $SS$  [108, 109]. However, despite considerable efforts, TFETs with high ON-state current ( $I_{ON}$ ), low OFF-state current ( $I_{OFF}$ ) and low supply voltage have not been experimentally demonstrated [4]. In recent years, there are few theoretical TFET designs based on resonant tunneling which predict decent device performances [110, 111]. However, these designs often rely on complicated III-V semiconductor heterostructures which are difficult to fabricate [112, 113]. Moreover, interface trap states induced by dangling bonds in these quasi low-dimensional semiconductors can severely limit the switching steepness and is considered to be a major obstacle in obtaining ideal TFET characteristics [114–116]. Two-dimensional (2D) crystals, on the other hand, have no dangling bonds on their surfaces, which allows for the formation of high-quality stacked van der Waal (vdW) heterostructures with atomically sharp interfaces [117]. The stacking of one 2D material on another 2D or 3D material allows for tailoring of the band alignment and the electrostatic control

to enable ideal device performance [118–120]. In addition, TFETs with 2D crystal channel also benefit from ultra-thin body thickness which suppresses the short channel effects and leads to further scaling down of the channel length [121].

Nevertheless, in spite of the growing interest in vdW TFETs [120, 122–125], theoretical studies that employ quantum mechanical models necessary to accurately assess the device performance and unveil important design guidelines, are scarce [126, 127]. One reason is that the semi-empirical tight binding parameterization, available for 3D crystals, does not exist for 2D vdW heterojunctions [128]. In addition, quantum mechanical models must consider scattering effects because tunneling heterojunctions usually contain confined states at regions close to the tunnel junction which can significantly affect the transport characteristics of the device [110, 111, 129]. In the non-equilibrium Green’s function (NEGF) formalism [27], accurate treatment of such effects as well as non-idealities (e.g. band tailing) requires explicit modeling of electron-phonon and electron-electron interactions, which are non-local in space and couple different energies and transverse momenta [44, 66, 130]. Therefore, one has to compute and store the Green’s functions and the self-energies for a large number of energy and momentum tuples. In addition, following the self-consistent Born approximation scheme, the Green’s functions and the self-energies need to be solved self-consistently until convergence is achieved [5]. Such calculation requires the usage of the whole supercomputer and is prohibitive for device simulation at the time of writing [131].

We propose our design of complementary n- and p-type TFETs based on the monolayer  $\text{WTe}_2$ - $\text{ZrS}_2$  vertical heterojunction with a broken-gap band alignment, using atomistic quantum transport simulation with a basis derived from ab-initio density functional theory (DFT) calculations. Non-ideality effects such as electron-phonon and electron-electron interactions are taken into account through a phenomenological scattering model based on the combination of equilibrium and non-equilibrium carrier statistics [132]. This scattering model was previously applied, with nearest neighbor semi-empirical tight binding basis, to resonant tunneling diodes and TFETs

based on III-V materials [111, 132, 133] and has shown good agreement between simulation and experiment [133]. Here, for the first time, it is applied to TFETs based on a 2D material vertical heterojunction, with a basis derived from ab-initio DFT calculations. Device performance degradation due to band tailing is also assessed. Owing to band structure engineering and design of the electrostatics, decent transfer and output characteristics are observed for both the n- and the p-type TFETs –  $I_{ON}=220 \mu\text{A}/\mu\text{m}$  and  $290 \mu\text{A}/\mu\text{m}$  for the n- and the p-type TFETs, respectively, at  $I_{OFF}=0.1 \text{ nA}/\mu\text{m}$  and with a supply voltage of  $0.4 \text{ V}$  – suitable for low power VLSI applications.  $SS$  is lower than  $60 \text{ mV}/\text{dec}$  for more than 4 orders of magnitude of the drain current. The low-power VLSI circuit performance of proposed TFET devices is demonstrated by simulating a 25-stage ring oscillator circuit using the MIT Virtual Source Model for TFET (MVS-TFET) [134].

## 5.2 Simulation approach

To accurately model the electronic structure of the monolayer  $\text{WTe}_2\text{-ZrS}_2$  vertical heterostructure, the DFT tool VASP [135] is used. The applied DFT model is based on the generalized gradient approximation utilizing the Perdew-Burke-Ernzerhof functionals [136]. Firstly, the equilibrium lattice constants of hexagonal unit cells of the  $\text{WTe}_2$  monolayer (H-phase) and the  $\text{ZrS}_2$  monolayer (T-phase) are determined to be  $3.56 \text{ \AA}$  and  $3.69 \text{ \AA}$ , respectively. Then, after applying  $1.7\%$  tensile strain and  $1.9\%$  compressive strain to  $\text{WTe}_2$  and  $\text{ZrS}_2$ , respectively, a common lattice constant of  $3.62 \text{ \AA}$  is achieved. The coordinates of the basis atoms of the strained heterostructure are further relaxed until all forces acting on ions are less than  $10^{-2} \text{ eV}/\text{\AA}$  and the total energy difference between two subsequent iterations is smaller than  $10^{-5} \text{ eV}$ . A momentum mesh of  $25 \times 15 \times 1$  Monkhorst-Pack grids and an energy cut off of  $500 \text{ eV}$  are used in the electronic structure calculation without spin-orbital coupling. Lastly, the Wannier90 package is used to transform the Bloch wavefunctions into a maximally localized Wannier functions (MLWF) representation [137]. 22 energy bands around

the Fermi level are selected which cover an energy window of 16 eV. As an initial projection,  $5d$  and  $3p$  atomic orbitals are used for metal electrons and chalcogenide electrons, respectively [138, 139]. This wannierization procedure produces a Hamiltonian of the unit cell with a short interaction range of 12 Å. The relaxed atomic structure and the unit cell Hamiltonian are imported to NEMO5 simulator [140] for quantum transport analysis.

All transport problems are solved by the multi-physics domain decomposition method within the NEGF formalism [116, 141, 142]. A constant scattering potential of 10 meV [132] is assumed below the valence band maximum (VBM) and above the conduction band minimum (CBM), and is added in the monolayer regions of the device [111, 132]. In the lack of experimental data, to estimate the device performance degradation due to band tailing, an Urbach parameter of 10 meV is assumed according to theoretical predictions [5, 143, 144]. For each bias point, the NEGF equation is solved self-consistently with the nonlinear Poisson equation (discretized on a 3D finite element grid) until convergence is achieved. The dielectric tensors of monolayer  $\text{WTe}_2$  and  $\text{ZrS}_2$  are anisotropic and are obtained from DFT calculations [145, 146]. Dielectric constants in the channel region of the simulated devices are averaged between those of  $\text{WTe}_2$  [145] and  $\text{ZrS}_2$  [146]. To interpret the particle density, a heuristic model with a smooth interpolation between electron and hole states is used [16]. For all devices, flat band Neumann boundary conditions are assumed for the Poisson equation except for the gate/oxide interface regions where Dirichlet conditions are used to set the applied gate voltage. Full ionization of donor and acceptor atoms is assumed.

### 5.3 Device design

The electronic band structure of the strained heterostructure is shown in Fig. 5.1(a), with the VBM of  $\text{WTe}_2$  (located at the  $K$  point) larger than the CBM of  $\text{ZrS}_2$  (located at the  $M$  point) by 0.067 eV. Such a broken-gap band alignment predicted by DFT calculations is favorable for TFETs since a small tunneling distance can be achieved

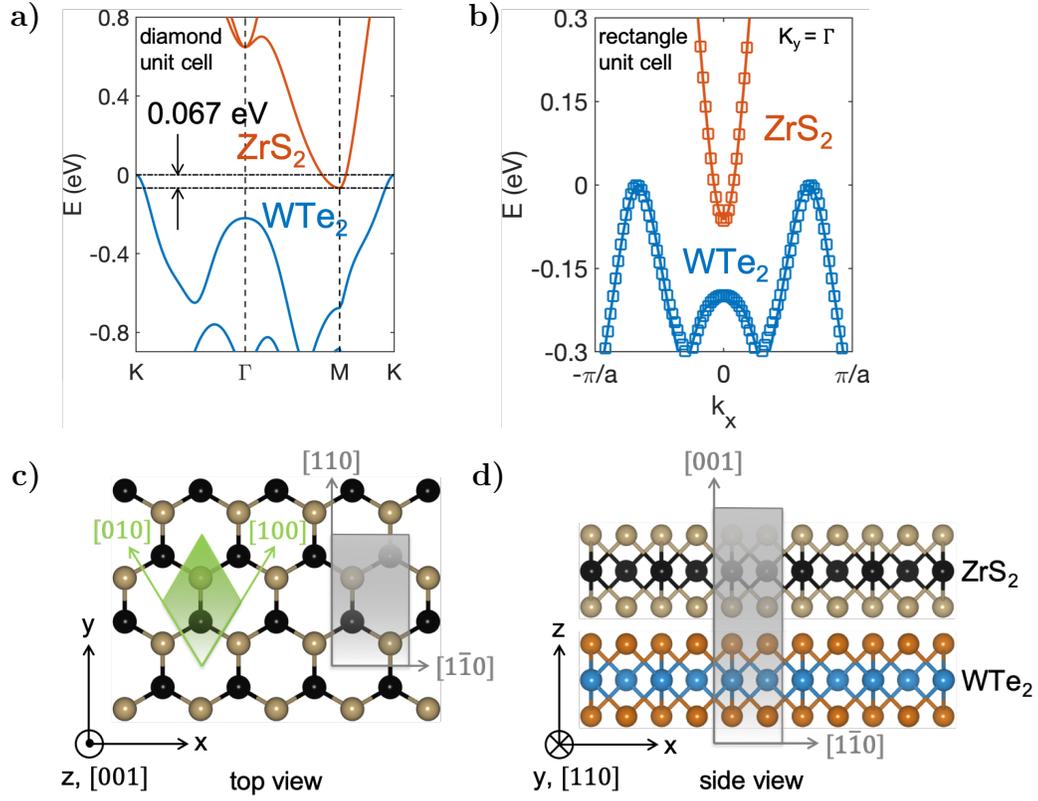


Fig. 5.1.: (a) Band structure of monolayer WTe<sub>2</sub>-ZrS<sub>2</sub> heterostructure calculated with VASP. (b) Band structures calculated with VASP (lines) and NEMO5 (symbols) at  $k_y=0$  along the transport direction. Both VBM of WTe<sub>2</sub> and CBM of ZrS<sub>2</sub> are folded to  $k_y=0$ . (c)-(d) The top view and the side view of the monolayer WTe<sub>2</sub>-ZrS<sub>2</sub> heterostructure. The primitive unit cell is marked by a green diamond. The gray rectangle marks the unit cell that is used in all transport calculations. The transport direction is along the  $x$  axis.

which leads to high ON-state current. Note that strain engineering of 2D materials is possible in principle [147–149] and is important especially for TFET applications since it allows tailoring the band alignment. According to the theoretical predictions [144, 150], the applied strain does not change the locations of the WTe<sub>2</sub> VBM and the ZrS<sub>2</sub> CBM in the reciprocal space, but it lowers the band gap of both WTe<sub>2</sub> and ZrS<sub>2</sub>. Therefore, the unstrained heterostructure would have either a Type-II stag-

gered band alignment or a Type-III broken-gap band alignment with a band offset smaller than 0.067 eV. The actual values of the applied strain, within the range that has been experimentally realized [147], are chosen to eliminate the lattice mismatch between  $\text{WTe}_2$  and  $\text{ZrS}_2$  so that a small unit cell can be used for the construction of the device structure. This consideration is critical for the simulation purpose since the computational cost (both CPU and memory) grows very rapidly with the size of the unit cell. To maximize the ON-state current, the direction in the reciprocal space which connects the VBM of  $\text{WTe}_2$  and the CBM of  $\text{ZrS}_2$  is set to be parallel to the transport direction  $x$ . In order to achieve this, the  $x$ -axis is chosen to be parallel to the  $[1\bar{1}0]$  crystallographic direction (see Fig.5.1(c)) and a periodic boundary condition is applied along the  $y$ -axis. Consequently, the device is constructed from the rectangle shape unit cell as illustrated in Figs. 5.1(c) and 5.1(d). In other words, the choice of the transport direction leads to a folded band structure shown in Fig. 1(b), where the VBM of  $\text{WTe}_2$  and the CBM of  $\text{ZrS}_2$  are both at the center of the new Brillouin zone ( $k_y=\Gamma$ ). In doing so, no change in the  $k_y$  momentum component is required for electron tunneling from the VBM of  $\text{WTe}_2$  to the CBM of  $\text{ZrS}_2$ .

The proposed designs for both the n- and the p-type TFETs are based on a  $\text{WTe}_2/\text{n-ZrS}_2$  double-gate vertical TFET. For the n-type TFET (schematic shown in Fig. 5.2(a)), source and the bottom gate are grounded and a drain-to-source voltage ( $V_{ds}$ ) of 0.4 V is applied. The top gate voltage ( $V_{tg}$ ) is ramped to turn the device ON and OFF. Following the work of Cao et al. [127], to avoid a superlinear onset in the output characteristic, the  $\text{WTe}_2$  layer is degeneratively p-doped. The  $\text{ZrS}_2$  layer is lightly n-doped in the drain region ( $L_d$ ). The doping densities are  $N_s=10^{13} \text{ cm}^{-2}$  and  $N_d=3.3 \times 10^{12} \text{ cm}^{-2}$ , respectively. For the p-type TFET (schematic shown in Fig. 5.2(b)), drain and the top gate are grounded and a source-to-drain voltage ( $V_{sd}$ ) of 0.4 V is applied. The bottom gate voltage ( $V_{bg}$ ) is ramped to turn the device ON and OFF. The  $\text{ZrS}_2$  layer is heavily n-doped and the  $\text{WTe}_2$  layer is lightly p-doped in the drain region ( $L_d$ ). The doping densities are  $N_s=10^{13} \text{ cm}^{-2}$  and  $N_d=3.3 \times 10^{12} \text{ cm}^{-2}$ , respectively. For both the n- and the p-type TFETs, the same work function is

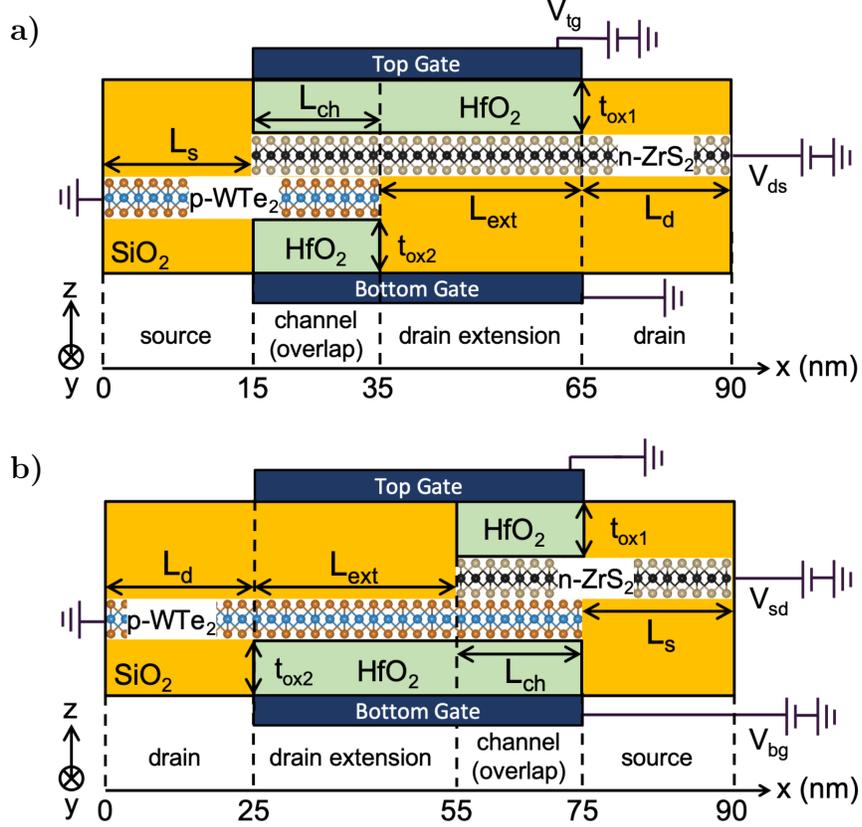


Fig. 5.2.: The cross-section of the n-type (a) and the p-type (b) TFET design based on the monolayer  $\text{WTe}_2\text{-ZrS}_2$  vertical heterojunction.

used for both the top and the bottom gates, with the work function of the p-type TFET being smaller than that of the n-type TFET by 0.43 eV. The actual values of the work functions are chosen to have  $I_{OFF}=0.1\text{ nA}/\mu\text{m}$  when the ramped gate voltage is at 0 V. The length of the source is  $L_s=15\text{ nm}$  and the length of the drain is  $L_d=25\text{ nm}$ , respectively. The drain extension has a length of  $L_{ext}=30\text{ nm}$  and the channel length is  $L_{ch}=20\text{ nm}$ . The bilayer channel region will be referred to as the overlap region hereinafter. The top and the bottom gate oxide have the same thickness of  $t_{ox1}=t_{ox2}=2.7\text{ nm}$  (EOT=0.4 nm).

#### 5.4 n-type TFET operating principle

In the proposed design of the n-type TFET, electron tunneling happens in the vertical ( $z$ ) direction and the tunneling probability is controlled by the top gate bias. Therefore, to understand the device operating principle, a 2D band diagram (i.e. one incorporating both the transport direction  $x$  and the vertical tunneling direction  $z$ ) is needed. To illustrate how the n-type TFET operates, the energy resolved local density of state (LDOS) at the OFF state ( $V_{tg}=0$  V) and the ON state ( $V_{tg}=0.4$  V) are shown in Figs. 5.3(a) and 5.3(b), respectively. Since the current contribution monotonically decays with the increase of the magnitude of  $k_y$  (i.e. the energy band gap of the heterostructure) and about 90 % of the current is carried by the first quarter of the Brillouin zone, all energy resolved LDOS figures in this letter are plotted at the center of the Brillouin zone ( $k_y=\Gamma$ ).

At the OFF state (see Fig. 5.3(a)), the VBM of  $\text{WTe}_2$  in the overlap region ( $\text{VBM}_{\text{ch}}$ ) is lower than the CBM of  $\text{ZrS}_2$  ( $\text{CBM}_{\text{ch}}$ ) due to the p-type doping in  $\text{WTe}_2$ . This creates an energy gap for the heterojunction where no vertical tunneling happens along the  $z$  direction. In addition, the tunneling path from the rightmost portion of  $\text{WTe}_2$  to the doped  $\text{ZrS}_2$  region ( $L_d$ ) is suppressed by having a long drain extension ( $L_{\text{ext}}$ ) which disconnects the LDOS in these two parts of the device.

As the top gate voltage increases, both the  $\text{WTe}_2$  valence band and the  $\text{ZrS}_2$  conduction band beneath the top gate are pulled down in energy. The bottom gate capacitance in the overlap region has a value approximately 8 times larger than the one in the drain extension region. Consequently, the band movements in the overlap region and the drain extension region are different, namely, in the drain extension region, the top gate voltage drops in both the top gate oxide and the bottom gate oxide, whereas in the overlap region, the top gate voltage drops mainly in the top gate oxide. In the overlap region, the modulation of  $\text{VBM}_{\text{ch}}$  is less pronounced than that of  $\text{CBM}_{\text{ch}}$  since the electric field of the top gate is partially screened by  $\text{ZrS}_2$  and the doping in  $\text{WTe}_2$ . As a result, a tunneling energy window is formed as  $\text{CBM}_{\text{ch}}$  becomes

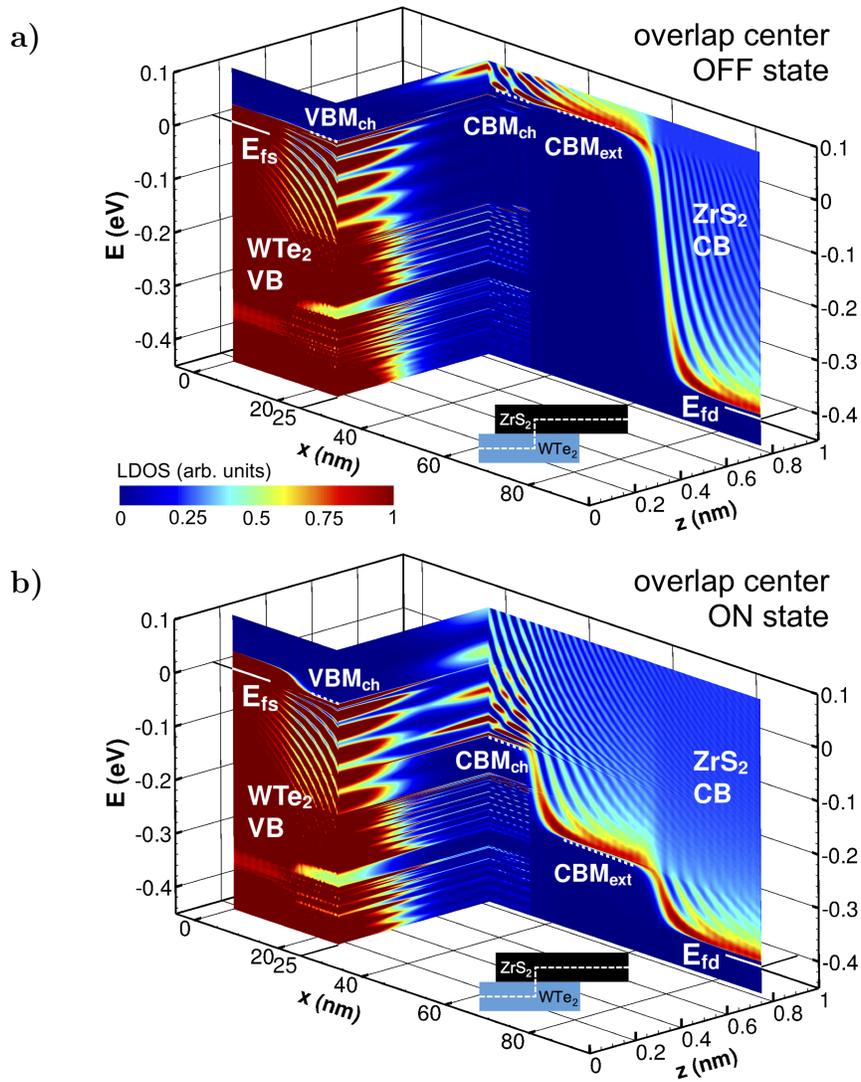


Fig. 5.3.: (a) OFF state ( $V_{tg}=0$  V) energy resolved local density of state (LDOS) at  $k_y=0$  plotted along the following path:  $x=0$  ( $\text{WTe}_2$ )  $\rightarrow x=25$  nm ( $\text{WTe}_2$ )  $\rightarrow x=25$  nm ( $\text{ZrS}_2$ )  $\rightarrow x=90$  nm ( $\text{ZrS}_2$ ), as indicated by the white dashed line in the inset. Within the  $\text{WTe}_2$  ( $\text{ZrS}_2$ ) layer, the  $z$  coordinate is taken at the atomic plane of W (Zr). Red: high LDOS, blue: low LDOS.  $\text{VBM}_{ch}$  and  $\text{CBM}_{ch}$  mark the VBM of  $\text{WTe}_2$  and the CBM of  $\text{ZrS}_2$  in the overlap region, respectively.  $\text{CBM}_{ext}$  marks the CBM of  $\text{ZrS}_2$  in the drain extension region. (b) Same as (a), but at ON state ( $V_{tg}=0.4$  V).

lower than  $VBM_{ch}$ . This enables current to flow as electrons in the source ( $L_s$ ) first propagate to the overlap region ( $L_{ch}$ ), then tunnel vertically from  $WTe_2$  to  $ZrS_2$  and continue to move towards the drain ( $L_d$ ). As the width of the tunneling energy window increases with the top gate voltage, so does the magnitude of the tunneling current. As shown in Fig. 5.3(b), at the ON state, source and drain are connected since  $CBM_{ch}$  is lower than  $VBM_{ch}$  in the overlap region which allows vertical tunneling to occur.

Energy resolved transmission and normalized cumulative current density at the ON state is shown in Fig. 5.4. As can be seen from Fig. 5.3(b), vertical tunneling

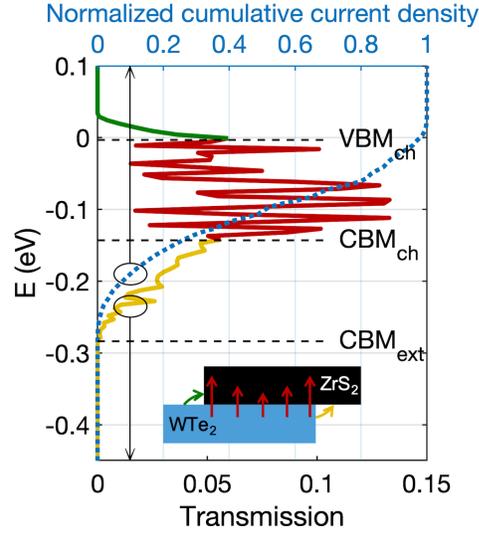


Fig. 5.4.: Energy resolved transmission at  $k_y=0$  at ON state. The three black dashed lines indicate  $VBM_{ch}$ ,  $CBM_{ch}$  and  $CBM_{ext}$ , respectively. The green, red and yellow curves represent transmission with energies above  $VBM_{ch}$ , between  $CBM_{ch}$  and  $VBM_{ch}$ , below  $CBM_{ext}$ , respectively. The blue dashed line shows the normalized cumulative current density. The inset associates the three transmission energy intervals with three different tunneling paths. The green, red and yellow arrows represent the tunneling paths of the green, red and yellow transmission curves, respectively.

in the middle of the overlap region is only allowed for energies between  $CBM_{ch}$  and  $VBM_{ch}$  (i.e. within the tunneling energy window), because for energies above (below)  $VBM_{ch}$  ( $CBM_{ch}$ ), the transmission is negligible since the wave function quickly decays

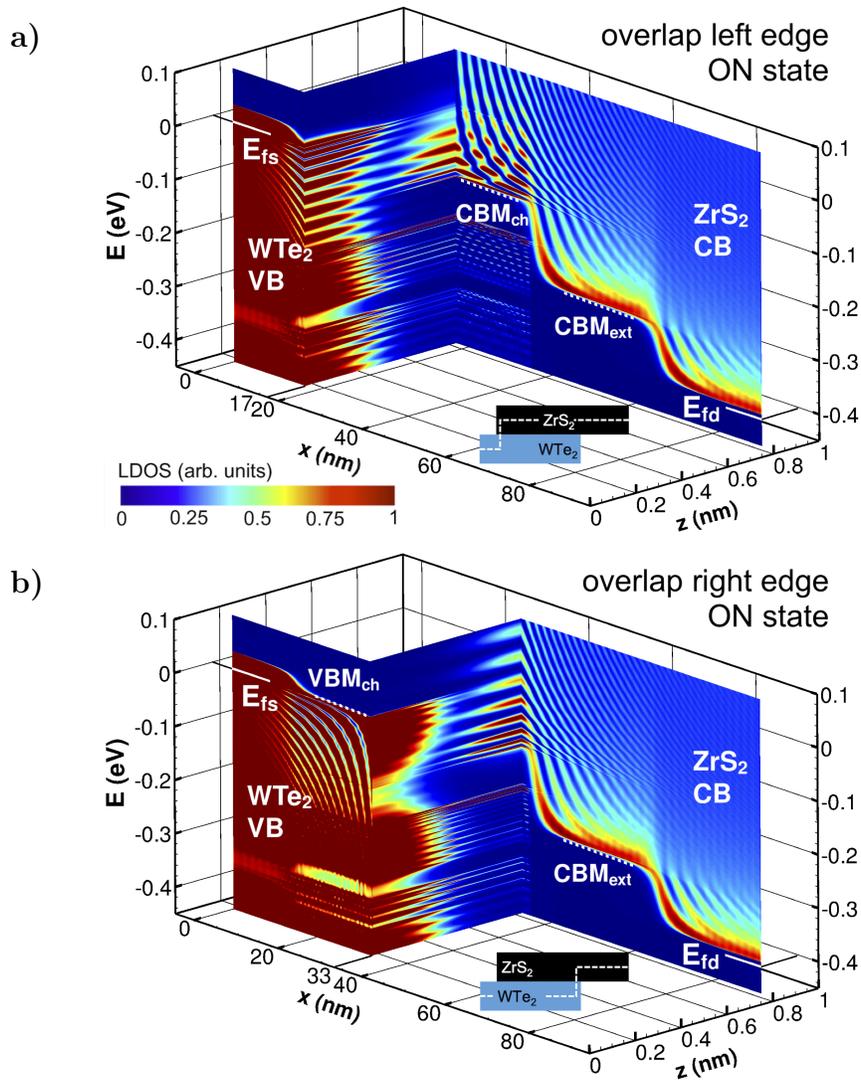


Fig. 5.5.: (a)-(b) Same as Fig. 5.3(b), but with cut planes at (a)  $x=17$  nm and (b)  $x=33$  nm, respectively.

in  $\text{WTe}_2$  ( $\text{ZrS}_2$ ) due to a lack of continuously available LDOS along the tunneling path. For transmission to appear below  $\text{CBM}_{\text{ch}}$ , the vertical tunneling has to happen near the right edge of the overlap region since the wave function of the propagating state must not die out before it enters the  $\text{ZrS}_2$  conduction band. Therefore, the responsible tunneling path is from the rightmost portion of  $\text{WTe}_2$  to  $\text{ZrS}_2$  in the drain extension region, as illustrated by the yellow arrow in the inset of Fig. 5.4. Similarly, for

transmission to appear above  $\text{VBM}_{\text{ch}}$ , the vertical tunneling has to happen near the left edge of the overlap region where LDOS in  $\text{WTe}_2$  is available, and the responsible tunneling path is from source to the leftmost portion of  $\text{ZrS}_2$ , as illustrated by the green arrow in the inset of Fig. 5.4. From the normalized cumulative current density, it can be seen that the above mentioned edge tunneling paths contribute about 28 % of the total current at  $k_y=0$ . As the magnitude of  $k_y$  increases, the percentage of edge tunneling current increases since the tunneling energy window (between  $\text{CBM}_{\text{ch}}$  and  $\text{VBM}_{\text{ch}}$ ) diminishes.

To examine the current density distribution along the overlap region, two more energy resolved LDOS plots at the ON state are shown in Figs. 5.5(a) and 5.5(b), respectively. Comparing Figs. 5.5(a) and 5.5(b) with Fig. 5.3(b), it can be seen that the tunneling channel in energy space between  $\text{WTe}_2$  and  $\text{ZrS}_2$  is denser at the two ends than that in the middle of the overlap region. This is attributed to a large value of LDOS in  $\text{ZrS}_2$  and  $\text{WTe}_2$  at the left and the right end of the overlap region, respectively, as shown in Figs. 5.5(a) and 5.5(b). Such an inhomogeneous connection between source and drain in the overlap region implies that the vertical tunneling current is likely to cluster at the two ends of the overlap region, as illustrated by the red arrows of different lengths in the inset of Fig. 5.4. This allows for the scaling down of the channel length without significantly sacrificing the ON-state current.

## 5.5 p-type TFET operating principle

Similar to the n-type TFET, a p-type TFET (schematic shown in 5.2(b)) is realized by altering the electrostatics within the device. To illustrate the operating principle of the p-type TFET, energy resolved LDOS of the p-type TFET at the OFF state ( $V_{bg}=0\text{ V}$ ) and the ON state ( $V_{bg}=-0.4\text{ V}$ ) are shown in Fig. 5.6(a) and 5.6(b), respectively. At the OFF state,  $\text{VBM}_{\text{ch}}$  is lower than  $\text{CBM}_{\text{ch}}$  due to the n-type doping in  $\text{ZrS}_2$ . Therefore, no vertical tunneling happens along the  $z$  direction due to the energy gap. Besides, the tunneling path from drain ( $L_d$ ) to the leftmost portion

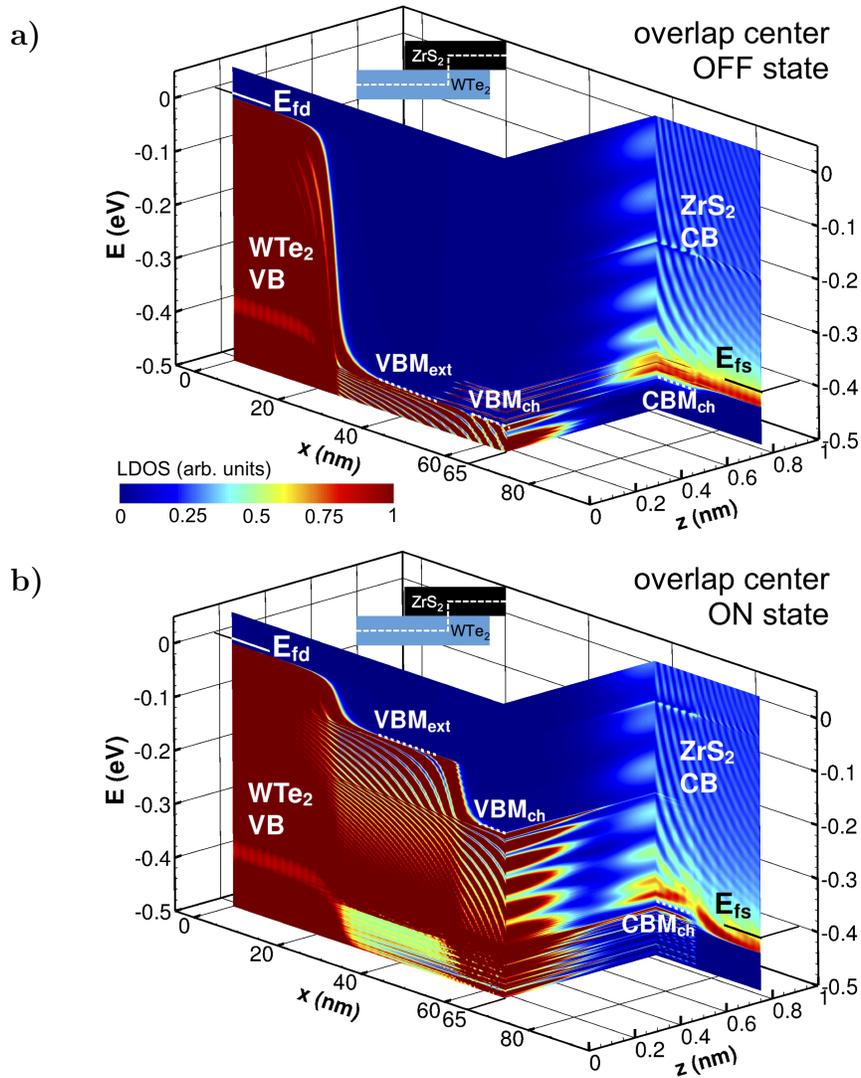


Fig. 5.6.: (a) OFF state ( $V_{bg}=0$  V) energy resolved local density of state (LDOS) at  $k_y=0$  plotted along the following path:  $x=0$  (WTe<sub>2</sub>)  $\rightarrow$   $x=65$  nm (WTe<sub>2</sub>)  $\rightarrow$   $x=65$  nm (ZrS<sub>2</sub>)  $\rightarrow$   $x=90$  nm (ZrS<sub>2</sub>), as indicated by the white dashed line in the inset. Within the WTe<sub>2</sub> (ZrS<sub>2</sub>) layer, the  $z$  coordinate is taken at the atomic plane of W (Zr). Red: high LDOS, blue: low LDOS. VBM<sub>ch</sub> and CBM<sub>ch</sub> mark the VBM of WTe<sub>2</sub> and the CBM of ZrS<sub>2</sub> in the overlap region, respectively. VBM<sub>ext</sub> marks the VBM of WTe<sub>2</sub> in the drain extension region. (b) Same as (a), but at ON state ( $V_{bg}=-0.4$  V).

of  $\text{ZrS}_2$  is suppressed by having a long drain extension ( $L_{ext}$ ). As the bottom gate voltage decreases, both the  $\text{WTe}_2$  valence band and the  $\text{ZrS}_2$  conduction band above the bottom gate are pushed up in energy. Similar to the n-type TFET, the band movements in the overlap region and the drain extension region are different due to the difference in the voltage division in these two regions. In the overlap region, the modulation of  $\text{VBM}_{ch}$  is more pronounced than that of  $\text{CBM}_{ch}$  since the electric field of the bottom gate is partially screened by  $\text{WTe}_2$  and the doping in  $\text{ZrS}_2$ . Hence, a tunneling energy window is formed as  $\text{VBM}_{ch}$  becomes higher than  $\text{CBM}_{ch}$  which enables current to flow as electrons in the drain ( $L_d$ ) first travel to the overlap region ( $L_{ch}$ ), then tunnel vertically from  $\text{WTe}_2$  to  $\text{ZrS}_2$  and get collected by source ( $L_s$ ). As the width of the tunneling energy window increases with the decrease of the bottom gate voltage, so does the magnitude of the tunneling current. As shown in Fig. 5.6(b), at the ON state,  $\text{VBM}_{ch}$  is higher than  $\text{CBM}_{ch}$  in the overlap region and the LDOS of source and drain are connected.

Energy resolved transmission at the ON state is plotted in Fig. 5.7. It is obvious from Fig. 5.6(b) that vertical tunneling in the middle of the overlap region is only allowed for energies between  $\text{CBM}_{ch}$  and  $\text{VBM}_{ch}$  (i.e. within the tunneling energy window). Therefore, the transmission below  $\text{CBM}_{ch}$  is contributed by the tunneling path from the rightmost portion of  $\text{WTe}_2$  to source, as illustrated by the yellow arrow in the inset of Fig. 5.7. For transmission above  $\text{VBM}_{ext}$ , the responsible tunneling path is from the drain extension region to the leftmost portion of  $\text{ZrS}_2$ , as illustrated by the green arrow in the inset of Fig. 5.7.

Finally, similar to the n-type TFET, an inhomogeneous connection between source and drain in the overlap region is also observed (see Figs. 5.8(a), 5.8(b) and 5.6(b)). Comparing Fig. 5.7 with Fig. 5.4, it can be seen that the p-type TFET has a larger transmission coefficient than the n-type TFET. This is because in the p-type TFET, electron tunneling happens in the energy range where the  $\text{ZrS}_2$  layer has more LDOS available than the n-type TFET, as can be seen from Figs. 5.3(b) and 5.6(b). Note that even in the p-type TFET, the transmission coefficient is still below 25%, which

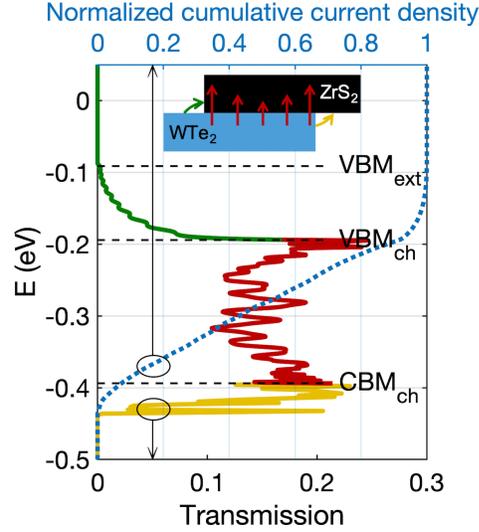


Fig. 5.7.: Energy resolved transmission at  $k_y=0$  at ON state. The three black dashed lines indicate  $VBM_{ext}$ ,  $VBM_{ch}$  and  $CBM_{ch}$ , respectively. The green, red and yellow curves represent transmission with energies above  $VBM_{ext}$ , between  $CBM_{ch}$  and  $VBM_{ch}$ , below  $CBM_{ch}$ , respectively. The blue dashed line shows the normalized cumulative current density. The inset associates the three transmission energy intervals with three different tunneling paths. The green, red and yellow arrows represent the tunneling paths of the green, red and yellow transmission curves, respectively.

is due to the LDOS mismatch between source and drain. In order to improve the electron collection at drain (source) of the n-type (p-type) TFET, one should seek for material combinations that have favorable band alignments as well as similar values of LDOS in the conduction band and the valence band of the n- and the p-type materials, respectively.

## 5.6 Device performance

Transfer characteristics of the n- and the p-type TFETs are shown in Figs. 5.9(a) and 5.9(d), respectively. OFF-state current is fixed at  $0.1 \text{ nA}/\mu\text{m}$  and a supply voltage of  $0.4 \text{ V}$  is applied. Assuming zero band tail states, ON-state current of the n-

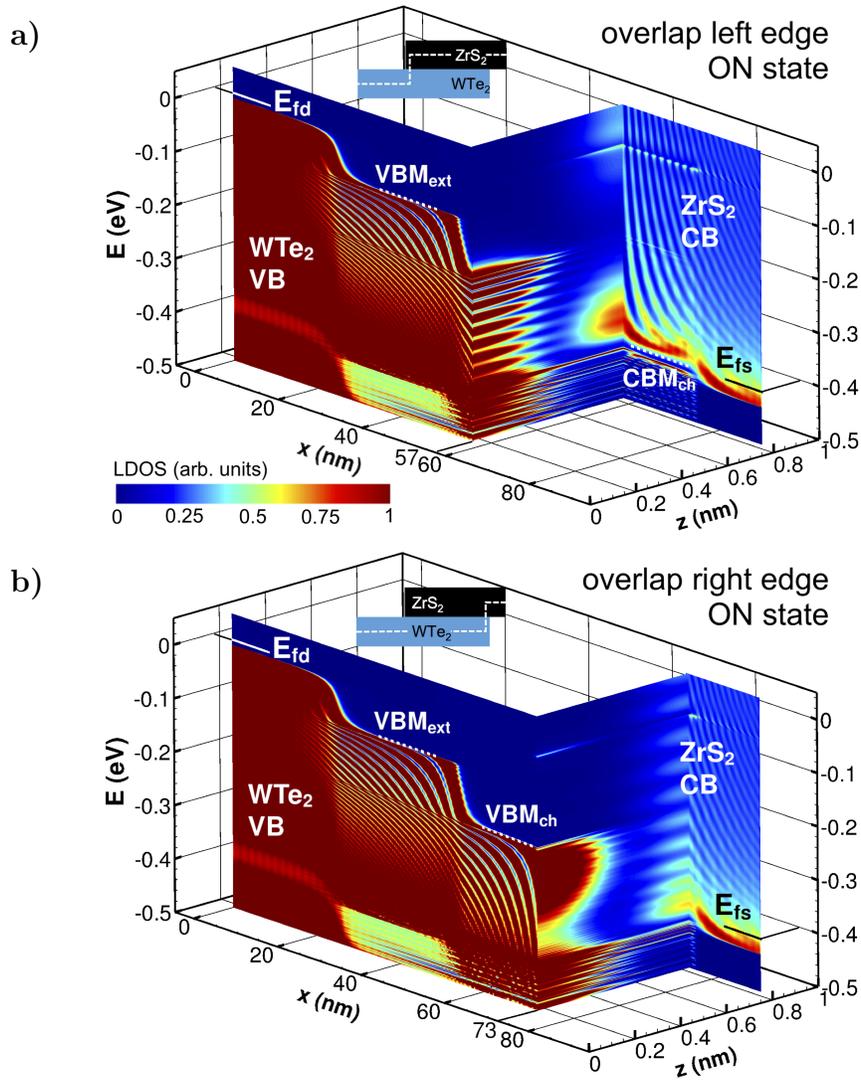


Fig. 5.8.: (a)-(b) Same as Fig. 5.6(b), but with cut planes at (a)  $x=57$  nm and (b)  $x=73$  nm, respectively.

and the p-type TFETs reach  $220 \mu\text{A}/\mu\text{m}$  and  $290 \mu\text{A}/\mu\text{m}$ , respectively. When band tailing is considered, degradation of the subthreshold region is observed as shown in Figs. 5.9(b) and 5.9(e). As a result, ON-state current of the n- and the p-type TFETs drops to  $130 \mu\text{A}/\mu\text{m}$  and  $175 \mu\text{A}/\mu\text{m}$ , respectively. Due to the heavy doping in source of both the n- and the p-type TFETs, a superlinear onset is absent in the

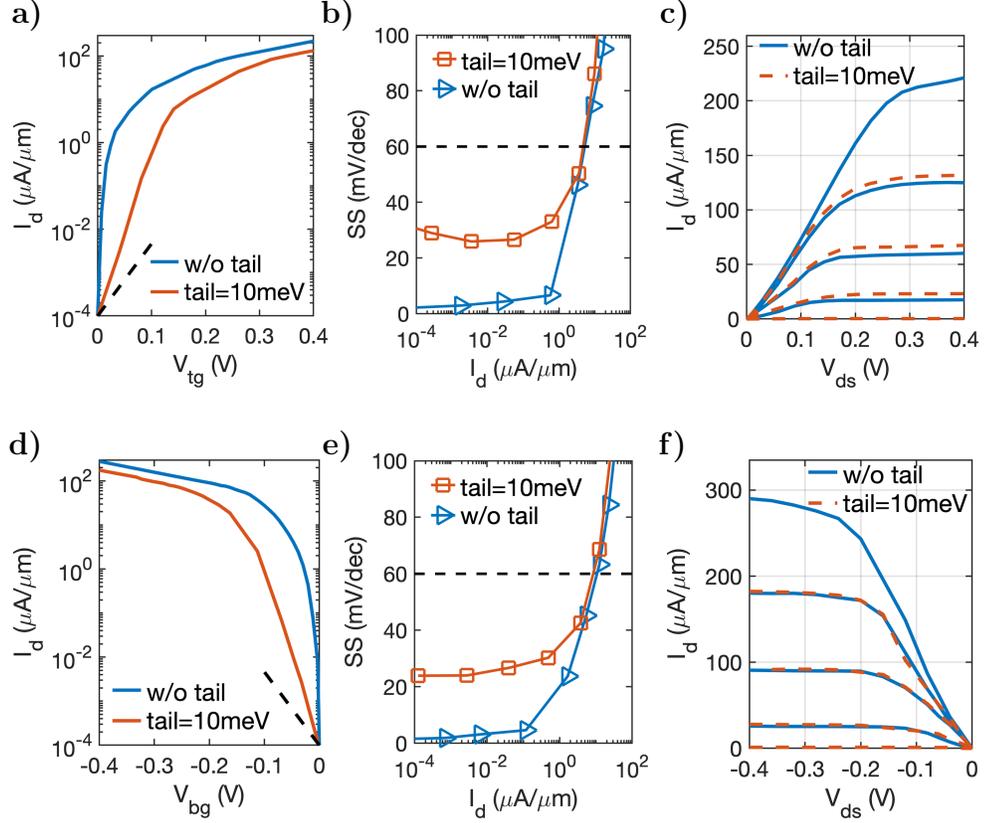


Fig. 5.9.: (a)-(c) Transfer characteristics,  $SS$  vs  $I_d$  and Output characteristics of the n-type TFET, respectively. The red curves include band tail states with the calibrated Urbach parameter while the blue curves consider zero band tail states. The four curves from top to bottom in (c) represent  $V_{tg}=0.4\text{ V}$ ,  $0.3\text{ V}$ ,  $0.2\text{ V}$  and  $0.1\text{ V}$ , respectively. (d)-(f) Same as (a)-(c) but for the p-type TFET. The four curves from top to bottom in (f) represent  $V_{bg}=-0.4\text{ V}$ ,  $-0.3\text{ V}$ ,  $-0.2\text{ V}$  and  $-0.1\text{ V}$ , respectively.

output characteristics (see Figs. 5.9(c) and 5.9(f)), which is essential for digital VLSI applications. Note that band tailing acts as if it shifts the gate voltage by  $0.1\text{ V}$ .

The scalability of  $L_{ext}$  is evaluated and the results are summarized in Fig. 5.10(a). The degradation of the subthreshold performance can be understood from Fig. 5.3(a). When  $L_{ext}=30\text{ nm}$ , the LDOS in the overlap region is well separated from those in the drain ( $L_d$ ). As  $L_{ext}$  decreases, it becomes easier for electrons to tunnel from the overlap region to the drain; hence the leakage increases. It can be seen from

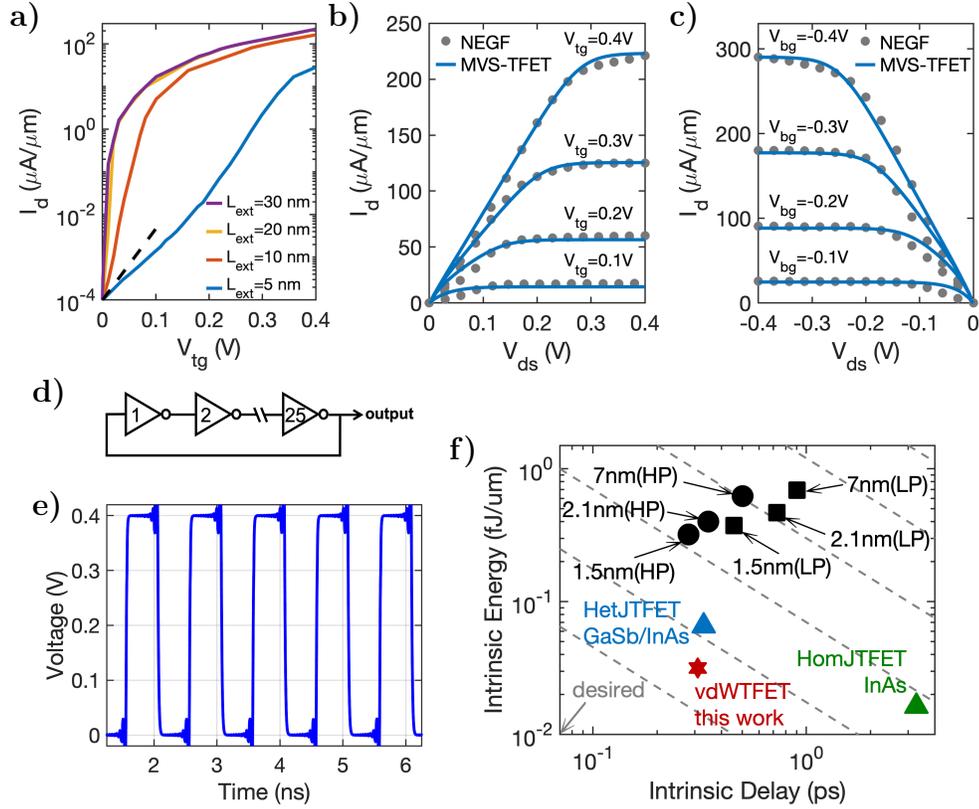


Fig. 5.10.: (a) Transfer characteristics of the n-type TFET with four different drain extension lengths ( $L_{ext}$ ) assuming zero band tail states. (b) n- and (c) p-type TFET output characteristics comparison of the NEGF simulation data (symbol) and the MVS-TFET model (solid line). (d) Schematic of the 25-stage ring oscillator circuit. Each inverter circuit in the chain uses the designed n- and p-type TFETs in this work. (e) Simulated waveform of the 25-stage ring oscillator circuit shown in (d) with an oscillation frequency of 1 GHz at  $V_{dd}=0.4$  V. (f) Intrinsic energy versus delay, benchmarked against ultrascaled silicon MOSFETs from IRDS [151] (black data points) and TFETs based on III-V materials [152] (blue and green data points). The preferred corner is bottom left. Dashed lines: constant energy-delay products.

Fig. 5.10(a) that  $L_{ext}$  has to be longer than 5 nm to secure  $SS < 60$  mV/dec. On the other hand, the improvement of  $SS$  saturates when the length of  $L_{ext}$  exceeds 20 nm. Due to an inhomogeneous current distribution in the overlap region and

the LDOS mismatch between source and drain, no notable change in the ON-state current is observed when  $L_{ch}$  is reduced to 5 nm – consistent with previous findings in literature [126].

The MIT Virtual Source model for TFET (MVS-TFET) [134] – a physics-based compact model designed for TFET which is capable of capturing non-ideality effects in TFET design – is employed to assess the circuit performance of the designed TFETs. The MVS-TFET model coefficients needed for SPICE simulation are obtained by fitting the n- and the p-type TFETs DC characteristics. Figs. 5.10(b) and 5.10(c) show the n- and the p-type TFET output characteristic comparison of the MIT MVS-TFET model (fit) and the NEGF simulation (without Urbach tail), respectively. Good agreement has been achieved. The low-power VLSI circuit performance of the designed TFET devices is demonstrated by simulating a 25-stage ring oscillator circuit (RO). The circuit schematic of the TFET-based RO is illustrated in Fig. 5.10(d). Fig. 5.10(e) shows the transient waveforms of the 25-stage RO with an oscillation frequency of 1 GHz at  $V_{dd} = 0.4$  V. The delay per inverter is close to low-power CMOS performance [153, 154] but at lower  $V_{dd}$  (i.e. dynamic power). Finally, the intrinsic energy ( $C_{gate}V_{dd}^2$ ) and delay ( $C_{gate}V_{dd}/I_{on}$ ) are calculated and compared to ultrascaled high performance (HP), low standby power (LP) silicon MOSFETs [151], and to homojunction (HomJ), heterojunction (HetJ) TFETs [152] based on III-V materials, as shown in Fig. 5.10(f).

## 5.7 Summary

We design and simulate complementary TFETs based on the monolayer p-WTe<sub>2</sub>/n-ZrS<sub>2</sub> vertical heterojunction, using atomistic quantum simulation with a basis derived from ab-initio DFT calculations. Non-ideality effects such as electron-phonon and electron-electron interactions are included via a phenomenological scattering model. Through band structure engineering and design of the electrostatics, both n- and p-type TFETs are realized with the same device configuration. Decent transfer and

output characteristics are observed, suitable for VLSI digital applications. With a single control gate, a supply voltage of 0.4 V and the OFF-state current fixed at 0.1 nA/ $\mu\text{m}$ , ON-state current reach 220  $\mu\text{A}/\mu\text{m}$  and 290  $\mu\text{A}/\mu\text{m}$  for the n- and the p-type TFETs, respectively. When band tailing is considered, ON-state current of the n- and the p-type TFETs drop to 130  $\mu\text{A}/\mu\text{m}$  and 175  $\mu\text{A}/\mu\text{m}$ , respectively. At the ON state, three different tunneling paths are identified, each being responsible for a certain energy interval in the transmission spectrum. Due to an inhomogeneous current density distribution in the overlap region, channel length scaling is possible without significantly sacrificing the ON-state current. It is found that the transport direction in vdW TFETs is critical and the electron collection at drain (source) of the n-type (p-type) TFET can be improved by seeking alternative material combinations which possess favorable band alignments and have similar values of LDOS in the conduction and the valence band of the n- and the p-type materials, respectively. Finally, digital circuit simulation of 25-stage RO based on the designed TFETs show reasonable low-power performance at  $V_{dd}=0.4\text{ V}$ .

## REFERENCES

## REFERENCES

- [1] “Making Semiconductor History: Contextualizing Samsung’s Latest Transistor Technology,” <https://news.samsung.com/global/editorial-making-semiconductor-history-contextualizing-samsungs-latest-transistor-technology>.
- [2] “42 Years of Microprocessor Trend Data,” <https://www.karlsruhp.net/2018/02/42-years-of-microprocessor-trend-data/>.
- [3] S. Markov, G. Penazzi, Y. Kwok, A. Pecchia, B. Aradi, T. Frauenheim, and G. Chen, “Permittivity of Oxidized Ultra-Thin Silicon Films from Atomistic Simulations,” *IEEE Electron Device Letters*, vol. 36, no. 10, pp. 1076–1078, 2015.
- [4] H. Lu and A. Seabaugh, “Tunnel Field-Effect Transistors: State-of-the-Art,” *IEEE Journal of the Electron Devices Society*, vol. 2, no. 4, pp. 44–49, jul 2014. [Online]. Available: <http://ieeexplore.ieee.org/document/6820751/>
- [5] P. Sarangapani, Y. Chu, J. Charles, G. Klimeck, and T. Kubis, “Band-tail formation and band-gap narrowing driven by polar optical phonons and charged impurities in atomically resolved iii-v semiconductors and nanodevices,” *Phys. Rev. Applied*, vol. 12, p. 044045, Oct 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevApplied.12.044045>
- [6] C. Prasad, L. Jiang, D. Singh, M. Agostinelli, C. Auth, P. Bai, T. Eiles, J. Hicks, C. H. Jan, K. Mistry, S. Natarajan, B. Niu, P. Packan, D. Pantuso, I. Post, S. Ramey, A. Schmitz, B. Sell, S. Suthram, J. Thomas, C. Tsai, and P. Vandervoorn, “Self-heat reliability considerations on Intel’s 22nm Tri-Gate technology,” *IEEE International Reliability Physics Symposium Proceedings*, pp. 1–5, 2013.
- [7] S. E. Liu, J. S. Wang, Y. R. Lu, D. S. Huang, C. F. Huang, W. H. Hsieh, J. H. Lee, Y. S. Tsai, J. R. S. Y. Lee, and K. Wu, “Self-Heating Effect in FinFETs and Its Impact on Devices Reliability Characterization,” vol. 2, pp. 2–5, 2014.
- [8] R. Wang, J. Zhuge, C. Liu, R. Huang, D. W. Kim, D. Park, and Y. Wang, “Experimental study on quasi-ballistic transport in silicon nanowire transistors and the impact of self-heating effects,” in *Technical Digest - International Electron Devices Meeting, IEDM*. IEEE, dec 2008, pp. 1–4. [Online]. Available: <https://ieeexplore.ieee.org/document/4796806>  
<http://ieeexplore.ieee.org/document/4796806/>
- [9] S. H. Shin, M. A. Wahab, M. Masduzzaman, K. Maize, J. Gu, M. Si, A. Shakouri, P. D. Ye, and M. A. Alam, “Direct Observation of Self-Heating in III-V Gate-All-Around Nanowire MOSFETs,” *IEEE Transactions on Electron*

- Devices*, vol. 62, no. 11, pp. 3516–3523, dec 2015. [Online]. Available: <http://ieeexplore.ieee.org/document/7047088/>
- [10] E. Yalon, C. J. McClellan, K. K. H. Smithe, M. Muñoz Rojo, R. L. Xu, S. V. Suryavanshi, A. J. Gabourie, C. M. Neumann, F. Xiong, A. B. Farimani, and E. Pop, “Energy Dissipation in Monolayer MoS<sub>2</sub> Electronics,” *Nano Letters*, vol. 17, no. 6, pp. 3429–3433, 2017. [Online]. Available: <http://pubs.acs.org/doi/abs/10.1021/acs.nanolett.7b00252>
- [11] S. Salahuddin and S. Datta, “Use of negative capacitance to provide voltage amplification for low power nanoscale devices,” *Nano Letters*, vol. 8, no. 2, pp. 405–410, 2008.
- [12] T. S. Böske, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, “Ferroelectricity in hafnium oxide thin films,” *Applied Physics Letters*, vol. 99, no. 10, 2011.
- [13] A. K. Yadav, K. X. Nguyen, Z. Hong, P. García-Fernández, P. Aguado-Puente, C. T. Nelson, S. Das, B. Prasad, D. Kwon, S. Cheema, A. I. Khan, C. Hu, J. Íñiguez, J. Junquera, L. Q. Chen, D. A. Muller, R. Ramesh, and S. Salahuddin, “Spatially resolved steady-state negative capacitance,” *Nature*, vol. 565, no. 7740, pp. 468–471, 2019. [Online]. Available: <http://dx.doi.org/10.1038/s41586-018-0855-y>
- [14] J. A. Kittl, B. Obradovic, D. Reddy, T. Rakshit, R. M. Hatcher, and M. S. Rodder, “On the validity and applicability of models of negative capacitance and implications for MOS applications,” *Applied Physics Letters*, vol. 113, no. 4, 2018.
- [15] B. Obradovic, T. Rakshit, R. Hatcher, J. A. Kittl, and M. S. Rodder, “Modeling Transient Negative Capacitance in Steep-Slope FeFETs,” in *IEEE Transactions on Electron Devices*, vol. 65, no. 11, nov 2018, pp. 5157–5164. [Online]. Available: <https://ieeexplore.ieee.org/document/8501893/>
- [16] Y. Chu, P. Sarangapani, J. Charles, G. Klimeck, and T. Kubis, “Explicit screening full band quantum transport model for semiconductor nanodevices,” *Journal of Applied Physics*, vol. 123, no. 24, p. 244501, jun 2018. [Online]. Available: <http://aip.scitation.org/doi/10.1063/1.5031461>
- [17] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, “Leakage current: Moore’s law meets static power,” *Computer*, vol. 36, no. 12, pp. 68–75, Dec 2003.
- [18] J. Kao, S. Narendra, and A. Chandrakasan, “Subthreshold leakage modeling and reduction techniques,” in *Proceedings of the 2002 IEEE/ACM International Conference on Computer-aided Design*, ser. ICCAD ’02. New York, NY, USA: ACM, 2002, pp. 141–148. [Online]. Available: <http://doi.acm.org/10.1145/774572.774593>
- [19] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, “Band-to-band tunneling in carbon nanotube field-effect transistors,” *Phys. Rev. Lett.*, vol. 93, p. 196805, Nov 2004. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.93.196805>

- [20] 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 Transactions on Electron Devices*, vol. 52, no. 12, pp. 2568–2576, Dec 2005.
- [21] A. M. Ionescu and H. Riel, “Tunnel field-effect transistors as energy-efficient electronic switches,” *Nature*, vol. 479, p. 329, nov 2011. [Online]. Available: <http://dx.doi.org/10.1038/nature10679> <http://10.0.4.14/nature10679>
- [22] H. Lu and A. Seabaugh, “Tunnel field-effect transistors: State-of-the-art,” *IEEE Journal of the Electron Devices Society*, vol. 2, no. 4, pp. 44–49, July 2014.
- [23] S. Mookerjee, D. Mohata, T. Mayer, V. Narayanan, and S. Datta, *IEEE Electron Device Letters*, vol. 31, no. 6, pp. 564–566, 2010.
- [24] T. Yu, U. Radhakrishna, J. L. Hoyt, and D. A. Antoniadis, “Quantifying the impact of gate efficiency on switching steepness of quantum-well tunnel-fets: Experiments, modeling, and design guidelines,” in *2015 IEEE International Electron Devices Meeting (IEDM)*, Dec 2015, pp. 22.4.1–22.4.4.
- [25] U. E. Avci, B. Chu-Kung, A. Agrawal, G. Dewey, V. Le, R. Rios, D. H. Morris, S. Hasan, R. Kotlyar, J. Kavalieros, and I. A. Young, “Study of tfet non-ideality effects for determination of geometry and defect density requirements for sub-60mv/dec ge tfet,” in *2015 IEEE International Electron Devices Meeting (IEDM)*, Dec 2015, pp. 34.5.1–34.5.4.
- [26] A. C. Seabaugh and Q. Zhang, “Low-voltage tunnel transistors for beyond CMOS logic,” *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2095–2110, 2010.
- [27] S. Datta, “Nanoscale device modeling: the Green’s function method,” *Superlattices and Microstructures*, vol. 28, no. 4, pp. 253–278, 2000. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0749603600909200>
- [28] Z. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, “nanomos 2.5: A two-dimensional simulator for quantum transport in double-gate mosfets,” *IEEE Transactions on Electron Devices*, vol. 50, no. 9, pp. 1914–1925, Sept 2003.
- [29] M. Luisier, R. Rhyner, A. Szabo, and A. Pedersen, “Atomistic simulation of nanodevices,” in *2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sept 2016, pp. 281–286.
- [30] J. Knoch, S. Mantl, and J. Appenzeller, “Impact of the dimensionality on the performance of tunneling fets: Bulk versus one-dimensional devices,” *Solid-State Electronics*, vol. 51, no. 4, pp. 572 – 578, 2007, special Issue: Papers selected from the 2006 ULIS Conference. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038110107000573>
- [31] A. Mátyás, T. Kubis, P. Lugli, and C. Jirauschek, “Comparison between semiclassical and full quantum transport analysis of THz quantum cascade lasers,” *Physica E: Low-Dimensional Systems and Nanostructures*, vol. 42, no. 10, pp. 2628–2631, 2010.

- [32] H. G. Yoo and P. M. Fauchet, “Dielectric constant reduction in silicon nanostructures,” *Phys. Rev. B*, vol. 77, p. 115355, Mar 2008. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.77.115355>
- [33] T. Andlauer and P. Vogl, “Full-band envelope-function approach for type-II broken-gap superlattices,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 80, no. 3, pp. 1–7, 2009.
- [34] S. Steiger, M. Povolotskyi, H. H. Park, T. Kubis, and G. Klimeck, “Nemo5: A parallel multiscale nanoelectronics modeling tool,” *IEEE Transactions on Nanotechnology*, vol. 10, no. 6, pp. 1464–1474, Nov 2011.
- [35] T. B. Boykin, G. Klimeck, and F. Oyafuso, “Valence band effective-mass expressions in the  $sp^3d^5s^*$  empirical tight-binding model applied to a si and ge parametrization,” *Phys. Rev. B*, vol. 69, p. 115201, Mar 2004. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.69.115201>
- [36] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, “Empirical  $sp^3d^5s^*$  tight-binding calculation for cubic semiconductors: General method and material parameters,” *Phys. Rev. B*, vol. 57, pp. 6493–6507, Mar 1998. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.57.6493>
- [37] T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, “Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory,” *Phys. Rev. B*, vol. 66, p. 125207, Sep 2002. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.66.125207>
- [38] C. S. Lent, D. J. Kirkner, C. S. Lent, and D. J. Kirkner, “The quantum transmitting boundary method The quantum transmitting boundary method,” vol. 6353, no. 1990, 1994.
- [39] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, “Atomistic simulation of nanowires in the  $sp^3d^5s^*$  tight-binding formalism: From boundary conditions to strain calculations,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 74, no. 20, pp. 1–12, 2006.
- [40] M. Luisier and G. Klimeck, “Atomistic full-band design study of inas band-to-band tunneling field-effect transistors,” *IEEE Electron Device Letters*, vol. 30, no. 6, pp. 602–604, June 2009.
- [41] —, “Performance comparisons of tunneling field-effect transistors made of insb, carbon, and gasb-inas broken gap heterostructures,” in *2009 IEEE International Electron Devices Meeting (IEDM)*, Dec 2009, pp. 1–4.
- [42] P. Long, E. Wilson, J. Z. Huang, G. Klimeck, M. J. W. Rodwell, and M. Povolotskyi, “Design and simulation of gasb/inas 2d transmission-enhanced tunneling fet,” *IEEE Electron Device Letters*, vol. 37, no. 1, pp. 107–110, Jan 2016.
- [43] D. A. Areshkin and B. K. Nikolić, “Electron density and transport in top-gated graphene nanoribbon devices: First-principles Green function algorithms for systems containing a large number of atoms,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 81, no. 15, pp. 1–17, 2010.

- [44] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *Journal of Applied Physics*, vol. 81, no. 12, pp. 7845–7869, 1997.
- [45] M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, "Modeling of nanoscale devices," *Proceedings of the IEEE*, vol. 96, no. 9, pp. 1511–1550, Sept 2008.
- [46] N. Papior, N. Lorente, T. Frederiksen, A. García, and M. Brandbyge, "Improvements on non-equilibrium and transport green function techniques: The next-generation transiesta," *Computer Physics Communications*, vol. 212, no. Supplement C, pp. 8 – 24, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S001046551630306X>
- [47] Y. P. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, "Tight-binding analysis of si and gaas ultrathin bodies with subatomic wave-function resolution," *Phys. Rev. B*, vol. 92, p. 085301, Aug 2015. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.92.085301>
- [48] P. Kim, T. W. Odom, J.-L. Huang, and C. M. Lieber, "Electronic density of states of atomically resolved single-walled carbon nanotubes: Van hove singularities and end states," *Phys. Rev. Lett.*, vol. 82, pp. 1225–1228, Feb 1999. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.82.1225>
- [49] R. Zeller, J. Deutz, and P. Dederichs, "Application of complex energy integration to selfconsistent electronic structure calculations," *Solid State Communications*, vol. 44, no. 7, pp. 993 – 997, 1982. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0038109882903209>
- [50] M. Brandbyge, J. L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, "Density-functional method for nonequilibrium electron transport," *Physical Review B - Condensed Matter and Materials Physics*, vol. 65, no. 16, pp. 1654011–1654017, 2002.
- [51] Y. M. Niquet and C. Delerue, "Band offsets, wells, and barriers at nanoscale semiconductor heterojunctions," *Physical Review B - Condensed Matter and Materials Physics*, vol. 84, no. 7, pp. 1–9, 2011.
- [52] H. Haug and A.-P. Jauho, "Quantum kinetics in transport and optics of semiconductors," 2007.
- [53] S. E. Laux, A. Kumar, and M. V. Fischetti, "Analysis of quantum ballistic electron transport in ultrasmall silicon devices including space-charge and geometric effects," *Journal of Applied Physics*, vol. 95, no. 10, pp. 5545–5582, 2004.
- [54] Y. Chu, J. Shi, K. Miao, Y. Zhong, P. Sarangapani, T. S. Fisher, G. Klimeck, X. Ruan, and T. Kubis, "Thermal boundary resistance predictions with non-equilibrium green's function and molecular dynamics simulations," 2019.
- [55] Z. I. Alferov, "The history and future of semiconductor heterostructures," *Semiconductors*, vol. 32, no. 1, pp. 1–14, 1998. [Online]. Available: <http://link.springer.com/10.1134/1.1187350>

- [56] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, L. Albert, T. Hall, J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, “Quantum Cascade Laser Hutchinson and Alfred Y . Cho Published by : American Association for the Advancement of Science Stable URL : <http://www.jstor.org/stable/2883703> JSTOR is a not-for-profit service that helps scholars , researchers , and students disc,” vol. 264, no. 5158, pp. 553–556, 1994.
- [57] S. Nakamura, T. Mukai, and M. Senoh, “Candela-class high-brightness InGaN/AlGaIn double-heterostructure blue-light-emitting diodes,” *Applied Physics Letters*, vol. 64, no. 13, pp. 1687–1689, 1994.
- [58] M. L. Tsai, S. H. Su, J. K. Chang, D. S. Tsai, C. H. Chen, C. I. Wu, L. J. Li, L. J. Chen, and J. H. He, “Monolayer MoS<sub>2</sub> heterojunction solar cells,” *ACS Nano*, vol. 8, no. 8, pp. 8317–8322, 2014.
- [59] E. T. Swartz and R. O. Pohl, “Thermal boundary resistance,” *Reviews of Modern Physics*, vol. 61, no. 3, pp. 605–668, 1989.
- [60] E. S. Landry and A. J. H. McGaughey, “Thermal boundary resistance predictions from molecular dynamics simulations and theoretical calculations,” *Physical Review B*, vol. 80, no. 16, p. 165304, oct 2009. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.80.165304>
- [61] P. K. Schelling, S. R. Phillpot, and P. Keblinski, “Comparison of atomic-level simulation methods for computing thermal conductivity,” *Physical Review B*, vol. 65, no. 14, pp. 1–12, 2002.
- [62] P. E. Hopkins, “Multiple phonon processes contributing to inelastic scattering during thermal boundary conductance at solid interfaces,” *Journal of Applied Physics*, vol. 106, no. 1, 2009.
- [63] K. Sääskilahti, J. Oksanen, J. Tulkki, and S. Volz, “Role of anharmonic phonon scattering in the spectrally decomposed thermal conductance at planar interfaces,” *Phys. Rev. B*, vol. 90, p. 134312, Oct 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.90.134312>
- [64] T. Feng, W. Yao, Z. Wang, J. Shi, C. Li, B. Cao, and X. Ruan, “Spectral analysis of nonequilibrium molecular dynamics: Spectral phonon temperature and local nonequilibrium in thin films and across interfaces,” *Phys. Rev. B*, vol. 95, p. 195202, May 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.95.195202>
- [65] C. Stieger, A. Szabo, and M. Luisier, “Ab-initio quantum transport simulation of self-heating in single-layer 2-D materials Ab-initio quantum transport simulation of self-heating in single-layer 2-D materials,” vol. 045708, 2017.
- [66] M. Luisier, “Atomistic modeling of anharmonic phonon-phonon scattering in nanowires,” *Phys. Rev. B*, vol. 86, p. 245407, Dec 2012. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.86.245407>
- [67] J. T. Gaskins, G. Kotsonis, A. Giri, S. Ju, A. Rohskopf, Y. Wang, T. Bai, E. Sachet, C. T. Shelton, Z. Liu, Z. Cheng, B. M. Foley, S. Graham, T. Luo, A. Henry, M. S. Goorsky, J. Shiomi, J. P. Maria, and P. E. Hopkins, “Thermal

- Boundary Conductance Across Heteroepitaxial ZnO/GaN Interfaces: Assessment of the Phonon Gas Model,” *Nano Letters*, vol. 18, no. 12, pp. 7469–7477, 2018.
- [68] J. Shi, X. Yang, T. S. Fisher, and X. Ruan, “Dramatic increase in the thermal boundary conductance and radiation limit from a nonequilibrium landauer approach,” 2018.
- [69] J. Tersoff, “Modeling solid-state chemistry: Interatomic potentials for multicomponent systems,” *Phys. Rev. B*, vol. 39, pp. 5566–5568, Mar 1989. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.39.5566>
- [70] S. Sadasivam, N. Ye, J. P. Feser, J. Charles, K. Miao, T. Kubis, and T. S. Fisher, “Thermal transport across metal silicide-silicon interfaces: First-principles calculations and Green’s function transport simulations,” *Physical Review B*, vol. 95, no. 8, p. 085310, feb 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.95.085310>
- [71] D. Sellan, E. Landry, J. Turney, a. McGaughey, and C. Amon, “Size effects in molecular dynamics thermal conductivity predictions,” *Physical Review B*, vol. 81, no. 21, pp. 1–10, 2010.
- [72] S. Plimpton, “Fast parallel algorithms for short-range molecular dynamics,” pp. 1–19, 1995. [Online]. Available: <https://doi.org/10.1006/jcph.1995.1039>
- [73] B. J. Alder and T. E. Wainwright, “Studies in molecular dynamics. i. general method,” *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 1959. [Online]. Available: <https://doi.org/10.1063/1.1730376>
- [74] A. Paul, M. Luisier, and G. Klimeck, *Modified valence force field approach for phonon dispersion: From zinc-blende bulk to nanowires: Methodology and computational details*, 2010, vol. 9, no. 3-4.
- [75] M. Büttiker, “Four-terminal phase-coherent conductance,” *Phys. Rev. Lett.*, vol. 57, pp. 1761–1764, Oct 1986. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.57.1761>
- [76] K. Miao, S. Sadasivam, J. Charles, G. Klimeck, T. S. Fisher, and T. Kubis, “Büttiker probes for dissipative phonon quantum transport in semiconductor nanostructures,” *Applied Physics Letters*, vol. 108, no. 11, 2016.
- [77] Y. S. Ju and K. E. Goodson, “Phonon scattering in silicon films with thickness of order 100 nm,” *Applied Physics Letters*, vol. 74, no. 20, pp. 3005–3007, may 1999. [Online]. Available: <http://aip.scitation.org/doi/10.1063/1.123994>
- [78] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge University Press, 2005.
- [79] —, *Electronic Transport in Mesoscopic Systems*, ser. Cambridge Studies in Semiconductor Physics and Microelectronic Engineering. Cambridge University Press, 1995.

- [80] Y. Chu, S.-C. Lu, N. Chowdhury, M. Povolotskyi, G. Klimeck, M. Mohamed, and T. Palacios, "Superior Performance of 5-nm Gate Length GaN Nanowire nFET for Digital Logic Applications," *IEEE Electron Device Letters*, vol. 40, no. 6, pp. 874–877, jun 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8625493/>
- [81] N. Loubet, T. Hook, P. Montanini, C. . Yeung, S. Kanakasabapathy, M. Guilom, T. Yamashita, J. Zhang, X. Miao, J. Wang, A. Young, R. Chao, M. Kang, Z. Liu, S. Fan, B. Hamieh, S. Sieg, Y. Mignot, W. Xu, S. . Seo, J. Yoo, S. Mochizuki, M. Sankarapandian, O. Kwon, A. Carr, A. Greene, Y. Park, J. Frougier, R. Galatage, R. Bao, J. Shearer, R. Conti, H. Song, D. Lee, D. Kong, Y. Xu, A. Arceo, Z. Bi, P. Xu, R. Muthinti, J. Li, R. Wong, D. Brown, P. Oldiges, R. Robison, J. Arnold, N. Felix, S. Skordas, J. Gaudiello, T. Standardaert, H. Jagannathan, D. Corliss, M. . Na, A. Knorr, T. Wu, D. Gupta, S. Lian, R. Divakaruni, T. Gow, C. Labelle, S. Lee, V. Paruchuri, H. Bu, and M. Khare, "Stacked nanosheet gate-all-around transistor to enable scaling beyond finfet," in *2017 Symposium on VLSI Technology*, June 2017, pp. T230–T231.
- [82] G. Bae, D. Bae, M. Kang, S. M. Hwang, S. S. Kim, B. Seo, T. Y. Kwon, T. J. Lee, C. Moon, Y. M. Choi, K. Oikawa, S. Masuoka, K. Y. Chun, S. H. Park, H. J. Shin, J. C. Kim, K. K. Bhuiwarka, D. H. Kim, W. J. Kim, J. Yoo, H. Y. Jeon, M. S. Yang, S. Chung, D. Kim, B. H. Ham, K. J. Park, W. D. Kim, S. H. Park, G. Song, Y. H. Kim, M. S. Kang, K. H. Hwang, C. Park, J. Lee, D. Kim, S.-m. Jung, and H. K. Kang, "3nm GAA Technology featuring Multi-Bridge-Channel FET for Low Power and High Performance Applications," pp. 656–659, 2018.
- [83] S. Bajaj, O. F. Shoron, P. S. Park, S. Krishnamoorthy, F. Akyol, T.-H. Hung, S. Reza, E. M. Chumbes, J. Khurgin, and S. Rajan, "Density-dependent electron transport and precise modeling of gan high electron mobility transistors," *Applied Physics Letters*, vol. 107, no. 15, p. 153504, 2015. [Online]. Available: <https://doi.org/10.1063/1.4933181>
- [84] K. Shinohara, D. Regan, A. Corrion, D. Brown, S. Burnham, P. J. Willadsen, I. Alvarado-Rodriguez, M. Cunningham, C. Butler, A. Schmitz, S. Kim, B. Holden, D. Chang, V. Lee, A. Ohoka, P. M. Asbeck, and M. Micovic, "Deeply-scaled self-aligned-gate gan dh-hemts with ultrahigh cutoff frequency," in *2011 International Electron Devices Meeting*, Dec 2011, pp. 19.1.1–19.1.4.
- [85] M. Shur, "Gan based transistors for high power applications1this paper was first published in the material research society symposium proceedings, symposium e, fall, 1997.1," *Solid-State Electronics*, vol. 42, no. 12, pp. 2131 – 2138, 1998. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038110198002081>
- [86] N. Chowdhury, G. Iannaccone, G. Fiori, D. A. Antoniadis, and T. Palacios, "GaN Nanowire n-MOSFET With 5 nm Channel Length for Applications in Digital Electronics," vol. 38, no. 7, pp. 859–862, 2017.
- [87] Y. Liu, N. Neophytou, G. Klimeck, and M. S. Lundstrom, "Band-structure effects on the performance of iii–v ultrathin-body soi mosfets," *IEEE Transactions on Electron Devices*, vol. 55, no. 5, pp. 1116–1122, May 2008.

- [88] Y. Liu, N. Neophytou, T. Low, G. Klimeck, and M. S. Lundstrom, “A tight-binding study of the ballistic injection velocity for ultrathin-body soi mosfets,” *IEEE Transactions on Electron Devices*, vol. 55, no. 3, pp. 866–871, March 2008.
- [89] N. Neophytou, A. Paul, M. S. Lundstrom, and G. Klimeck, “Simulations of nanowire transistors: atomistic vs. effective mass models,” *Journal of Computational Electronics*, vol. 7, no. 3, pp. 363–366, Sep 2008. [Online]. Available: <https://doi.org/10.1007/s10825-008-0188-4>
- [90] S. Mukherjee, K. Miao, A. Paul, N. Neophytou, R. Kim, J. Geng, M. Povolotskyi, T. C. Kubis, A. Ajoy, B. Novakovic, J. Fonseca, H. Ilatikhameneh, S. Steiger, M. McLennan, M. Lundstrom, and G. Klimeck, “Band structure lab,” May 2006. [Online]. Available: <https://nanohub.org/resources/bandstrlab>
- [91] C. S. Lent and D. J. Kirkner, “The quantum transmitting boundary method,” *Journal of Applied Physics*, vol. 67, no. 10, pp. 6353–6359, 1990. [Online]. Available: <https://doi.org/10.1063/1.345156>
- [92] C. Bulutay, B. K. Ridley, and N. A. Zakhleniuk, “Full-band polar optical phonon scattering analysis and negative differential conductivity in wurtzite gan,” *Phys. Rev. B*, vol. 62, pp. 15 754–15 763, Dec 2000. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.62.15754>
- [93] E. Matioli and T. Palacios, “Room-temperature ballistic transport in iii-nitride heterostructures,” *Nano Letters*, vol. 15, no. 2, pp. 1070–1075, 2015, pMID: 25614931. [Online]. Available: <https://doi.org/10.1021/nl504029r>
- [94] M. Lundstrom, “Fundamentals of carrier transport, 2nd edn,” *Measurement Science and Technology*, vol. 13, no. 2, p. 230, 2002. [Online]. Available: <http://stacks.iop.org/0957-0233/13/i=2/a=703>
- [95] M. Luisier, A. Schenk, and W. Fichtner, “Atomistic treatment of interface roughness in si nanowire transistors with different channel orientations,” *Applied Physics Letters*, vol. 90, no. 10, p. 102103, 2007. [Online]. Available: <https://doi.org/10.1063/1.2711275>
- [96] M. Mohamed, Z. Aksamija, W. Vitale, F. Hassan, K. Park, and U. Ravaioli, “A conjoined electron and thermal transport study of thermal degradation induced during normal operation of multigate transistors,” *IEEE Transactions on Electron Devices*, vol. 61, no. 4, pp. 976–983, April 2014.
- [97] S. Lee, F. Oyafuso, P. von Allmen, and G. Klimeck, “Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures,” *Phys. Rev. B*, vol. 69, p. 045316, Jan 2004. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.69.045316>
- [98] Y. He, Y. Tan, Z. Jiang, M. Povolotskyi, G. Klimeck, and T. Kubis, “Surface passivation in empirical tight binding,” *IEEE Transactions on Electron Devices*, vol. 63, no. 3, pp. 954–958, March 2016.
- [99] T. Kuykendall, P. Pauzauskie, S. Lee, Y. Zhang, J. Goldberger, and P. Yang, “Metalorganic chemical vapor deposition route to gan nanowires with triangular cross sections,” *Nano Letters*, vol. 3, no. 8, pp. 1063–1066, 2003. [Online]. Available: <https://doi.org/10.1021/nl034422t>

- [100] J.-W. Zhao, Y.-F. Zhang, Y.-H. Li, C.-h. Su, X.-M. Song, H. Yan, and R.-Z. Wang, “A low cost, green method to synthesize gan nanowires,” *Scientific reports*, vol. 5, p. 17692, 2015.
- [101] A. Paul, M. Luisier, and G. Klimeck, “Shape and orientation effects on the ballistic phonon thermal properties of ultra-scaled si nanowires,” *Journal of Applied Physics*, vol. 110, no. 11, p. 114309, 2011. [Online]. Available: <https://doi.org/10.1063/1.3662177>
- [102] D. J. Carter, J. D. Gale, B. Delley, and C. Stampfl, “Geometry and diameter dependence of the electronic and physical properties of gan nanowires from first principles,” *Phys. Rev. B*, vol. 77, p. 115349, Mar 2008. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.77.115349>
- [103] J. C. Johnson, H.-J. Choi, K. P. Knutsen, R. D. Schaller, P. Yang, and R. J. Saykally, “Single gallium nitride nanowire lasers,” *Nature Materials*, vol. 1, p. 106, sep 2002. [Online]. Available: <https://doi.org/10.1038/nmat728>  
<http://10.0.4.14/nmat728>
- [104] T. Vo, A. J. Williamson, and G. Galli, “First principles simulations of the structural and electronic properties of silicon nanowires,” *Phys. Rev. B*, vol. 74, p. 045116, Jul 2006. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.74.045116>
- [105] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 2013. [Online]. Available: <https://books.google.com/books?id=rb2tAQAAQBAJ>
- [106] T. N. Theis and P. M. Solomon, “In quest of the next switch: Prospects for greatly reduced power dissipation in a successor to the silicon field-effect transistor,” *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2005–2014, 2010.
- [107] N. S. Kim, T. Austin, D. Blaauw, T. Mudge, K. Flautner, J. S. Hu, M. Jane Irwin, M. Kandemir, and V. Narayanan, “Leakage Current: Moore’s Law Meets Static Power,” *Computer*, vol. 36, no. 12, pp. 68–75, 2003.
- [108] J. Appenzeller, Y. M. Lin, J. Knoch, and P. Avouris, “Band-to-band tunneling in carbon nanotube field-effect transistors,” *Physical Review Letters*, vol. 93, no. 19, pp. 1–4, 2004.
- [109] K. Boucart and A. M. Ionescu, “Double-Gate Tunnel FET With High- $\kappa$  Gate Dielectric,” *IEEE Transactions on Electron Devices*, vol. 54, no. 7, pp. 1725–1733, 2007. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4252356>
- [110] P. Long, E. Wilson, J. Z. Huang, G. Klimeck, M. J. W. Rodwell, and M. Povolotskyi, “Design and Simulation of GaSb/InAs 2D Transmission-Enhanced Tunneling FETs,” *IEEE Electron Device Letters*, vol. 37, no. 1, pp. 107–110, jan 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7317763/>
- [111] J. Z. Huang, P. Long, M. Povolotskyi, H. Ilatikhameneh, T. A. Ameen, R. Rahman, M. J. Rodwell, and G. Klimeck, “A Multiscale Modeling of Triple-Heterojunction Tunneling FETs,” *IEEE Transactions on Electron Devices*, vol. 64, no. 6, pp. 2728–2735, 2017.

- [112] G. Dewey, B. Chu-Kung, J. Boardman, J. M. Fastenau, J. Kavalieros, R. Kotlyar, W. K. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. W. Then, and R. Chau, "Fabrication, characterization, and physics of III-V heterojunction tunneling field effect transistors (H-TFET) for steep sub-threshold swing," *Technical Digest - International Electron Devices Meeting, IEDM*, vol. 3, pp. 33.6.1–33.6.4, 2011.
- [113] J. A. Del Alamo, "Nanometre-scale electronics with III-V compound semiconductors," *Nature*, vol. 479, no. 7373, pp. 317–323, 2011.
- [114] A. Vandooren, D. Leonelli, R. Rooyackers, A. Hikavy, K. Devriendt, M. Demand, R. Loo, G. Groeseneken, and C. Huyghebaert, "Analysis of trap-assisted tunneling in vertical Si homo-junction and SiGe hetero-junction Tunnel-FETs," *Solid-State Electronics*, vol. 83, pp. 50–55, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.sse.2013.01.026>
- [115] Yingxin Qiu, Runsheng Wang, Qianqian Huang, and Ru Huang, "A Comparative Study on the Impacts of Interface Traps on Tunneling FET and MOSFET," *IEEE Transactions on Electron Devices*, vol. 61, no. 5, pp. 1284–1291, 2014.
- [116] P. Long, J. Z. Huang, M. Povolotskyi, P. Sarangapani, G. A. Valencia-Zapata, T. Kubis, M. J. Rodwell, and G. Klimeck, "Atomistic modeling trap-assisted tunneling in hole tunnel field effect transistors," *Journal of Applied Physics*, vol. 123, no. 17, 2018.
- [117] K. S. Novoselov, A. Mishchenko, A. Carvalho, and A. H. Castro Neto, "2D materials and van der Waals heterostructures," *Science*, vol. 353, no. 6298, 2016.
- [118] F. M. Ginzburg and L. Model, "Ginzburg – Landau Model Using the Ginzburg – Landau Model," pp. 801–804, 2017.
- [119] X. Wang and F. Xia, "Stacked 2D materials shed light," *Nature Materials*, vol. 14, no. 3, pp. 264–265, 2015.
- [120] D. Sarkar, X. Xie, W. Liu, W. Cao, J. Kang, Y. Gong, S. Kraemer, P. M. Ajayan, and K. Banerjee, "A subthermionic tunnel field-effect transistor with an atomically thin channel," *Nature*, vol. 526, no. 7571, pp. 91–95, oct 2015. [Online]. Available: <http://www.nature.com/articles/nature15387>
- [121] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling theory for double-gate SOI MOSFET's," *IEEE Transactions on Electron Devices*, vol. 40, no. 12, pp. 2326–2329, 1993. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/249482> <http://ieeexplore.ieee.org/document/249482/>
- [122] X. Cao, M. Tosun, Y.-Z. Chen, J. Guo, D.-h. Lien, P. Zhao, H. Fang, A. Javey, Y.-L. Chueh, and T. Roy, "Dual-Gated MoS<sub>2</sub>/WSe<sub>2</sub> van der Waals Tunnel Diodes and Transistors," *ACS Nano*, vol. 9, no. 2, pp. 2071–2079, 2015.
- [123] T. Roy, M. Tosun, M. Hettick, G. H. Ahn, C. Hu, and A. Javey, "2D-2D tunneling field-effect transistors using WSe<sub>2</sub>/SnSe<sub>2</sub> heterostructures," *Applied Physics Letters*, vol. 108, no. 8, 2016.

- [124] A. Nourbakhsh, A. Zubair, M. S. Dresselhaus, and T. Palacios, “Transport Properties of a MoS<sub>2</sub>/WSe<sub>2</sub> Heterojunction Transistor and Its Potential for Application,” *Nano Letters*, vol. 16, no. 2, pp. 1359–1366, 2016. [Online]. Available: <http://pubs.acs.org/doi/10.1021/acs.nanolett.5b04791>
- [125] Y. Balaji, Q. Smets, C. J. Lockhart De La Rosa, A. K. A. Lu, D. Chiappe, T. Agarwal, D. H. Lin, C. Huyghebaert, I. Radu, D. Mocuta, and G. Groeseneken, “Tunneling Transistors Based on MoS<sub>2</sub>/MoTe<sub>2</sub> Van der Waals Heterostructures,” *IEEE Journal of the Electron Devices Society*, vol. 6, no. February, pp. 1018–1055, 2018.
- [126] Á. Szabó, S. J. Koester, M. Luisier, and A. Band-to band, “Ab-Initio Simulation of van der Waals MoTe<sub>2</sub>–SnS<sub>2</sub> Heterotunneling FETs for Low-Power Electronics,” vol. 36, no. 5, pp. 514–516, 2015.
- [127] J. Cao, A. Cresti, D. Esseni, and M. Pala, “Quantum simulation of a heterojunction vertical tunnel FET based on 2D transition metal dichalcogenides,” *Solid-State Electronics*, vol. 116, no. February, pp. 1–7, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.sse.2015.11.003>
- [128] F. Zahid, L. Liu, Y. Zhu, J. Wang, and H. Guo, “A generic tight-binding model for monolayer, bilayer and bulk MoS<sub>2</sub>,” *AIP Advances*, vol. 3, no. 5, 2013.
- [129] R. C. Bowen, G. Klimeck, R. K. Lake, W. R. Frensley, and T. Moise, “Quantitative simulation of a resonant tunneling diode,” *Journal of Applied Physics*, vol. 81, no. 7, pp. 3207–3213, apr 1997. [Online]. Available: <http://aip.scitation.org/doi/10.1063/1.364151>
- [130] T. Kubis, C. Yeh, P. Vogl, A. Benz, G. Fasching, and C. Deutsch, “Theory of nonequilibrium quantum transport and energy dissipation in terahertz quantum cascade lasers,” *Phys. Rev. B*, vol. 79, p. 195323, May 2009. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.79.195323>
- [131] R. Andrawis, J. D. Bermeo, J. Charles, J. Fang, J. Fonseca, Y. He, G. Klimeck, Z. Jiang, T. Kubis, D. Mejia, D. Lemus, M. Povolotskyi, S. A. P. Rubiano, P. Sarangapani, and L. Zeng, “NEMO5: Achieving High-end Internode Communication for Performance Projection Beyond Moore’s Law,” *2015 Gordon Bell Prize submission*, oct 2015. [Online]. Available: <http://arxiv.org/abs/1510.04686>
- [132] T. A. Ameen, H. Ilatikhameneh, J. Z. Huang, M. Povolotskyi, R. Rahman, and G. Klimeck, “Combination of Equilibrium and Nonequilibrium Carrier Statistics into an Atomistic Quantum Transport Model for Tunneling Heterojunctions,” *IEEE Transactions on Electron Devices*, vol. 64, no. 6, pp. 2512–2518, 2017.
- [133] G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, T. S. Moise, G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, and S. Ted, “Quantum device simulation with a generalized tunneling formula Simulation of Quantum Tunneling in an Open System Quantum device simulation with a generalized tunneling formula,” vol. 2539, no. 1995, 2006.

- [134] R. N. Sajjad, U. Radhakrishna, and D. A. Antoniadis, “A tunnel fet compact model including non-idealities with verilog implementation,” *Solid-State Electronics*, vol. 150, pp. 16 – 22, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038110118303125>
- [135] G. Kresse and J. Furthmüller, “Software vasp, vienna (1999),” *Phys. Rev. B*, vol. 54, no. 11, p. 169, 1996.
- [136] J. Klimeš, D. R. Bowler, and A. Michaelides, “Chemical accuracy for the van der waals density functional,” *Journal of Physics: Condensed Matter*, vol. 22, no. 2, p. 022201, dec 2009. [Online]. Available: <https://doi.org/10.1088%2F0953-8984%2F22%2F2%2F022201>
- [137] A. A. Mostofi, J. R. Yates, Y. S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, “wannier90: A tool for obtaining maximally-localised Wannier functions,” *Computer Physics Communications*, vol. 178, no. 9, pp. 685–699, 2008.
- [138] C. H. Lee, E. C. Silva, L. Calderin, M. A. T. Nguyen, M. J. Hollander, B. Bersch, T. E. Mallouk, and J. A. Robinson, “Tungsten Ditelluride: A layered semimetal,” *Scientific Reports*, vol. 5, pp. 1–8, 2015. [Online]. Available: <http://dx.doi.org/10.1038/srep10013>
- [139] A. Ghafari, A. Boochani, C. Janowitz, and R. Manzke, “Electronic structure of ZrSxSe2-x by Tran-Blaha modified Becke-Johnson density functional,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 84, no. 12, pp. 1–7, 2011.
- [140] S. Steiger, M. Povolotskyi, H.-H. Park, T. Kubis, and G. Klimeck, “NEMO5: A Parallel Multiscale Nanoelectronics Modeling Tool,” *IEEE Transactions on Nanotechnology*, vol. 10, no. 6, pp. 1464–1474, nov 2011. [Online]. Available: <http://ieeexplore.ieee.org/document/6069914/>
- [141] G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, and T. S. Moise, “Quantum device simulation with a generalized tunneling formula,” *Applied Physics Letters*, vol. 67, no. 17, pp. 2539–2541, 1995. [Online]. Available: <https://doi.org/10.1063/1.114451>
- [142] J. Geng, P. Sarangapani, K.-C. Wang, E. Nelson, B. Browne, C. Wordelman, J. Charles, Y. Chu, T. Kubis, and G. Klimeck, “Quantitative multi-scale, multi-physics quantum transport modeling of gan-based light emitting diodes,” *physica status solidi (a)*, vol. 215, no. 9, p. 1700662, 2018. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pssa.201700662>
- [143] P. Sarangapani, Y. Chu, K. C. Wang, D. Valencia, J. Charles, and T. Kubis, “Nonequilibrium Green’s function method: Transport and band tail predictions in transition metal dichalcogenides,” in *2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, vol. 2018-Septe. IEEE, sep 2018, pp. 38–39. [Online]. Available: <https://ieeexplore.ieee.org/document/8551748/>
- [144] M. O. Li, D. Esseni, D. Jena, and H. G. Xing, “Lateral transport in two-dimensional heterojunction interlayer tunneling field effect transistor (Thin-TFET),” *Device Research Conference - Conference Digest, DRC*, vol. 074508, pp. 17–18, 2014.

- [145] A. Kumar and P. K. Ahluwalia, “Tunable dielectric response of transition metals dichalcogenides  $MX_2$  ( $M=Mo, W$ ;  $X=S, Se, Te$ ): Effect of quantum confinement,” *Physica B: Condensed Matter*, vol. 407, no. 24, pp. 4627–4634, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.physb.2012.08.034>
- [146] A. Laturia, M. L. Van de Put, and W. G. Vandenberghe, “Dielectric properties of hexagonal boron nitride and transition metal dichalcogenides: from monolayer to bulk,” *npj 2D Materials and Applications*, vol. 2, no. 1, p. 6, 2018. [Online]. Available: <http://www.nature.com/articles/s41699-018-0050-x>
- [147] R. Roldán, A. Castellanos-Gomez, E. Cappelluti, and F. Guinea, “Strain engineering in semiconducting two-dimensional crystals,” *Journal of Physics Condensed Matter*, vol. 27, no. 31, 2015.
- [148] X. He, H. Li, Z. Zhu, Z. Dai, Y. Yang, P. Yang, Q. Zhang, P. Li, U. Schwingenschlöggl, and X. Zhang, “Strain engineering in monolayer  $WS_2$ ,  $MoS_2$ , and the  $WS_2/MoS_2$  heterostructure,” *Applied Physics Letters*, vol. 109, no. 17, 2016. [Online]. Available: <http://dx.doi.org/10.1063/1.4966218>
- [149] T. Shen, A. V. Penumatcha, and J. Appenzeller, “Strain Engineering for Transition Metal Dichalcogenides Based Field Effect Transistors,” *ACS Nano*, vol. 10, no. 4, pp. 4712–4718, 2016.
- [150] B. Amin, T. P. Kaloni, and U. Schwingenschlöggl, “Strain engineering of  $WS_2$ ,  $WSe_2$ , and  $WTe_2$ ,” *RSC Advances*, vol. 4, no. 65, p. 34561, 2014. [Online]. Available: <http://xlink.rsc.org/?DOI=C4RA06378C>
- [151] “International Roadmap for Devices and Systems (IRDS) 2017 Edition,” <https://irds.ieee.org/>.
- [152] D. E. Nikonov and I. A. Young, “Overview of Beyond-CMOS Devices and a Uniform Methodology for Their Benchmarking,” *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2498–2533, dec 2013. [Online]. Available: <http://ieeexplore.ieee.org/document/6527325/>
- [153] M. Meijer, B. Liu, R. van Veen, and J. P. de Gyvez, “Post-silicon tuning capabilities of 45nm low-power cmos digital circuits,” in *2009 Symposium on VLSI Circuits*, June 2009, pp. 110–111.
- [154] K. von Arnim, E. Augendre, C. Pacha, T. Schulz, K. T. San, F. Bauer, A. Nackaerts, R. Rooyackers, T. Vandeweyer, B. Degroote, N. Collaert, A. Dixit, R. Singanamalla, W. Xiong, A. Marshall, C. R. Cleavelin, K. Schrufer, and M. Jurczak, “A low-power multi-gate fet cmos technology with 13.9ps inverter delay, large-scale integrated high performance digital circuits and sram,” in *2007 IEEE Symposium on VLSI Technology*, June 2007, pp. 106–107.
- [155] “Tecplot 360, a visualization analysis software tool developed by Tecplot, Inc., which is headquartered in Bellevue, Washington.” <https://www.tecplot.com/products/tecplot-360/>.
- [156] J. T. Lü and J. S. Wang, “Coupled electron and phonon transport in one-dimensional atomic junctions,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 76, no. 16, pp. 1–9, 2007.

- [157] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, M. S. Lundstrom, R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. S. Lundstrom, “A simple quantum mechanical treatment of scattering in nanoscale transistors A simple quantum mechanical treatment of scattering in nanoscale transistors,” vol. 5613, no. 2003, 2016.
- [158] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, “Two-dimensional quantum mechanical modeling of nanotransistors,” *Journal of Applied Physics*, vol. 91, no. 3, pp. 2343–2354, 2002.

## APPENDICES

## A. PLOT LOCAL DENSITY OF STATES USING TEC PLOT

### A.1 Introduction

This section gives a brief tutorial on plotting local density of states (LDOS) using Tecplot 360 [155] (referred to as Tecplot hereafter). Comparing with MATLAB contour plot which is commonly seen in quantum transport literature, Tecplot provides a better visualization when it comes to vdWTFETs, where electron tunnel vertically in the channel region, and a 2D band diagram (i.e. one incorporating both the transport direction and the vertical tunneling direction) is needed to understand the device operating principle.

### A.2 Data preparation

In quantum transport simulation, the raw data of LDOS is atom or orbital resolved. To prepare data in vtk format which can be loaded by Tecplot, one can generate a FEM mesh and interpolate data onto FEM nodes. Then, the vtk output method can be called to save data in a .vtu file. Usually, both the interpolation method and the output method are available in standard FEM library like libmesh, deal.ii, etc.

### A.3 Plot figure in Tecplot

Firstly, load the .vtu file into Tecplot. This will take a few minutes since the .vtu file is usually quite big in size. Note that it is optional to unclick "Auto redraw" at the bottom of the left panel. When "Auto redraw" is selected, Tecplot automatically

redraws the figure when a change is made. When "Auto redraw" is not selected, the redrawing is triggered manually by clicking the "Redraw" button.

Secondly, if needed, choose "Plot→assign XYZ" to modify the data association with each axis. For example, in the data file, the transport direction is x, the vertical direction is z and the energy is set to be along the y direction. Therefore, in order to have energy plotted along the z axis, one should set the axis as shown in Fig. A.1.

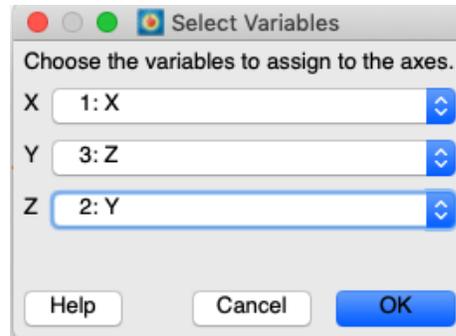


Fig. A.1.: Setting data association with axis.

Next, on the left panel, click "details" right to "contour", this opens up a window where one can configure the color map, color bar, min and max of the field data, etc.

To cut and show slices of the data, select "details" next to "Slices" on the left panel. Here, one can configure the slice by setting which axis the slice is perpendicular to and the cross point coordinate. Fig. A.2 shows the settings for a slice perpendicular to the x axis at  $x = 25$ . Fig. A.3 shows the settings for two slices perpendicular to the y axis at  $y = 0.172$  and  $y = 0.83$ , respectively. Now, by selecting "Slices" in the left panel and clicking "Redraw", one will get the figure as shown in Fig. A.4, where data on the three selected slices are plotted only.

The final step is to blank out some parts of the plot to improve visibility. This is done by first setting some variables and then choosing value blanking. Choose "Data→Alter→Specify Equations", and type in variables, as shown in Fig. A.5. Then, choose "Plot→Blanking→Value blanking", and choose to blank data when these vari-

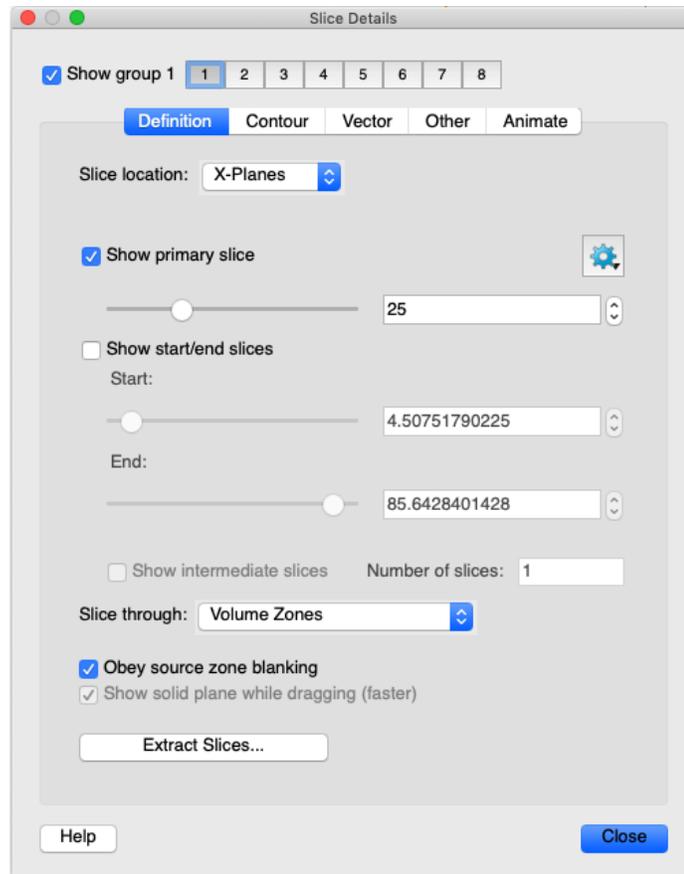


Fig. A.2.: Setting one slice perpendicular to the x axis.

ables are equal to 1, as shown in Fig. A.6. Now, click "Redraw" and one will get the figure as shown in Fig. A.7. Finally, save the layout.

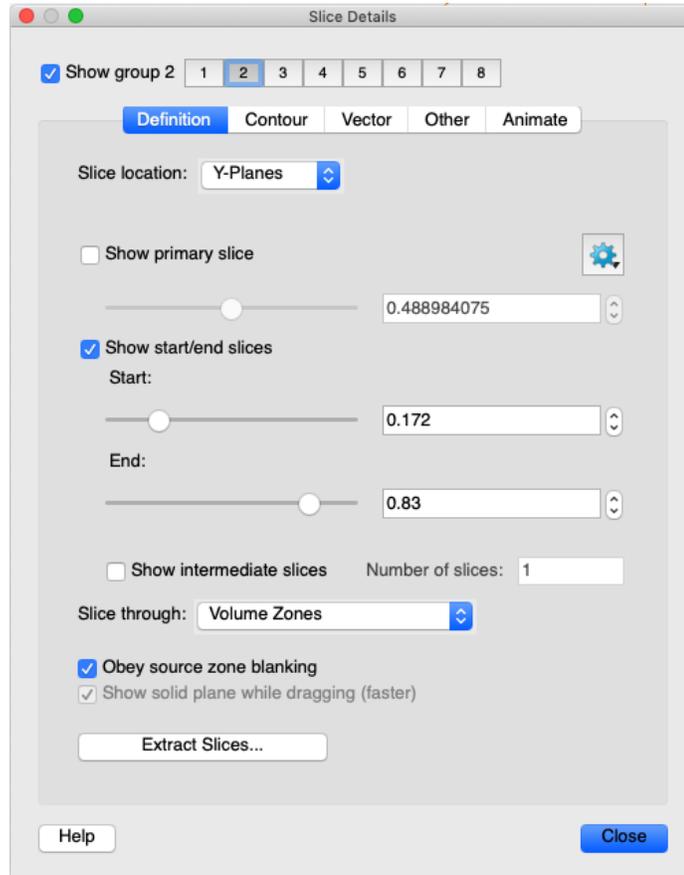


Fig. A.3.: Setting two slices perpendicular to the y axis.

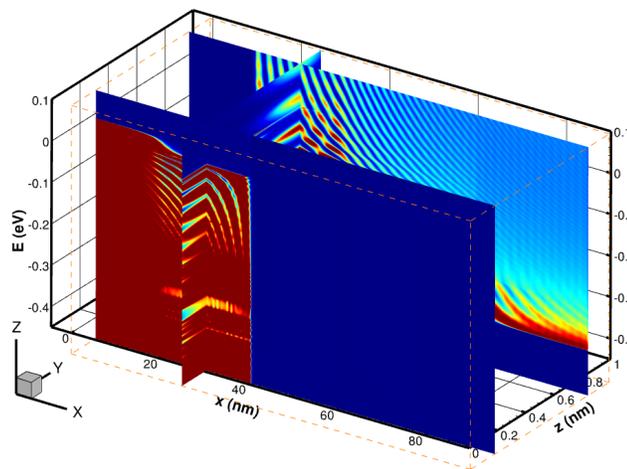


Fig. A.4.: The plot after setting the slices.

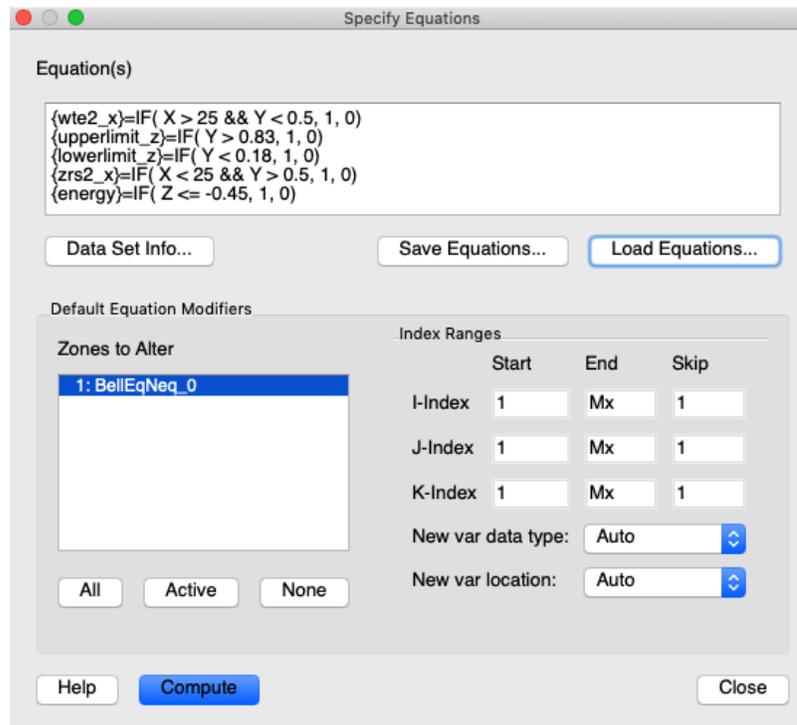


Fig. A.5.: Setting variables.

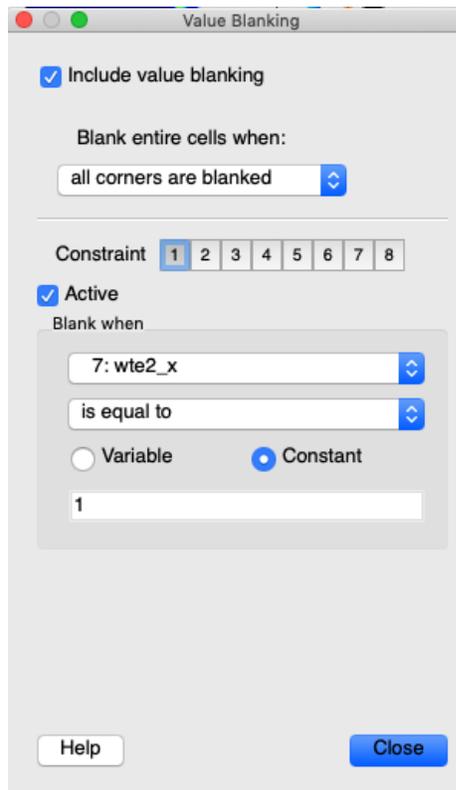


Fig. A.6.: Setting value blanking.

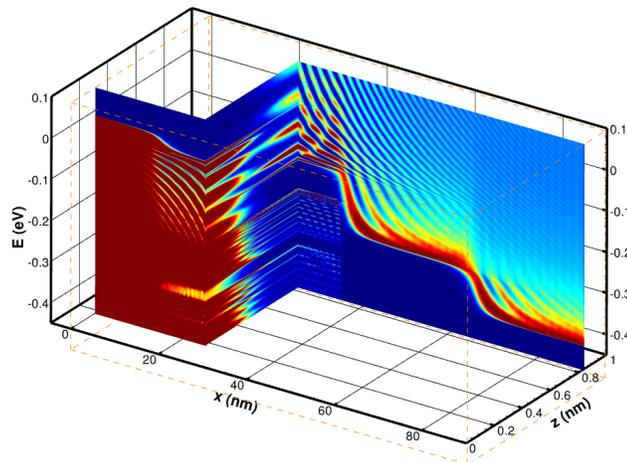


Fig. A.7.: The final figure.

## B. FINITE ELEMENT IMPLEMENTATION OF LANDAU-KHALATNIKOV EQUATION

### B.1 Introduction

In order to model ferroelectric field-effect transistors (FE-FETs), we need to solve the Landau-Khalatnikov (L-K) equation along the ferroelectric axis of the ferroelectric material

$$E_z = \alpha P_z + \beta P_z^3 + \gamma P_z^5 - \kappa_x \frac{\partial^2 P_z}{\partial x^2} - \kappa_y \frac{\partial^2 P_z}{\partial y^2} - \kappa_z \frac{\partial^2 P_z}{\partial z^2}, \quad (\text{B.1})$$

where  $z$  aligns with the ferroelectric axis. We solve Eq. B.1 on the ferroelectric material volume  $\Omega$ .

### B.2 Equation discretization

First we need to derive the weak form of the equation above. We can obtain the weak form by multiplying the equation by a test function  $u$  from the left and integrate over the domain  $\Omega$

$$\begin{aligned} \int_{\Omega} u E_z dV &= \alpha \int_{\Omega} u P_z dV + \beta \int_{\Omega} u P_z^3 dV + \gamma \int_{\Omega} u P_z^5 dV \\ &\quad - \kappa_x \int_{\Omega} u \frac{\partial^2 P_z}{\partial x^2} dV - \kappa_y \int_{\Omega} u \frac{\partial^2 P_z}{\partial y^2} dV - \kappa_z \int_{\Omega} u \frac{\partial^2 P_z}{\partial z^2} dV. \end{aligned} \quad (\text{B.2})$$

We seek approximation solution

$$P_z = \sum_i P_{zi} \varphi_i, \quad (\text{B.3})$$

where the  $P_{zi}$  are unknown expansion coefficients we need to determine (the “degree of freedom” of this problem), and  $\varphi_i$  are the finite element shape functions we will use. We can also write

$$P_z^3 = \left( \sum_i P_{zi} \varphi_i \right)^3, \quad (\text{B.4})$$

$$P_z^5 = \left( \sum_i P_{zi} \varphi_i \right)^5, \quad (\text{B.5})$$

$$E_z = \sum_i E_{zi} \varphi_i. \quad (\text{B.6})$$

For the last 3 terms in Eq. B.2, we have

$$u \frac{\partial^2 P_z}{\partial x^2} = \frac{\partial}{\partial x} u \frac{\partial P_z}{\partial x} - \frac{\partial u}{\partial x} \frac{\partial P_z}{\partial x}. \quad (\text{B.7})$$

$$u \frac{\partial^2 P_z}{\partial y^2} = \frac{\partial}{\partial y} u \frac{\partial P_z}{\partial y} - \frac{\partial u}{\partial y} \frac{\partial P_z}{\partial y}. \quad (\text{B.8})$$

$$u \frac{\partial^2 P_z}{\partial z^2} = \frac{\partial}{\partial z} u \frac{\partial P_z}{\partial z} - \frac{\partial u}{\partial z} \frac{\partial P_z}{\partial z}. \quad (\text{B.9})$$

Bring Eq. B.3 to Eq. B.7~B.9, we have

$$\begin{aligned} u \frac{\partial^2 P_z}{\partial x^2} &= \frac{\partial}{\partial x} \varphi_i \frac{\partial \sum_j P_{zj} \varphi_j}{\partial x} - \frac{\partial \varphi_i}{\partial x} \frac{\partial \sum_j P_{zj} \varphi_j}{\partial x} \\ &= \sum_j P_{zj} \frac{\partial}{\partial x} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) - \sum_j P_{zj} \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} \\ &= \sum_j P_{zj} \left[ \frac{\partial}{\partial x} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) - \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} \right]. \end{aligned} \quad (\text{B.10})$$

$$\begin{aligned} u \frac{\partial^2 P_z}{\partial y^2} &= \frac{\partial}{\partial y} \varphi_i \frac{\partial \sum_j P_{zj} \varphi_j}{\partial y} - \frac{\partial \varphi_i}{\partial y} \frac{\partial \sum_j P_{zj} \varphi_j}{\partial y} \\ &= \sum_j P_{zj} \frac{\partial}{\partial y} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) - \sum_j P_{zj} \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} \\ &= \sum_j P_{zj} \left[ \frac{\partial}{\partial y} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) - \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} \right]. \end{aligned} \quad (\text{B.11})$$

$$\begin{aligned} u \frac{\partial^2 P_z}{\partial z^2} &= \frac{\partial}{\partial z} \varphi_i \frac{\partial \sum_j P_{zj} \varphi_j}{\partial z} - \frac{\partial \varphi_i}{\partial z} \frac{\partial \sum_j P_{zj} \varphi_j}{\partial z} \\ &= \sum_j P_{zj} \frac{\partial}{\partial z} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) - \sum_j P_{zj} \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} \\ &= \sum_j P_{zj} \left[ \frac{\partial}{\partial z} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) - \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} \right]. \end{aligned} \quad (\text{B.12})$$

So the last 3 terms in Eq. B.2 can be written as

$$\begin{aligned}\kappa_x \int_{\Omega} u \frac{\partial^2 P_z}{\partial x^2} dV &= \kappa_x \int_{\Omega} \sum_j P_{zj} \left[ \frac{\partial}{\partial x} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) - \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} \right] dV \\ &= \kappa_x \sum_j P_{zj} \int_{\Omega} \frac{\partial}{\partial x} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) dV - \kappa_x \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV.\end{aligned}\quad (\text{B.13})$$

$$\begin{aligned}\kappa_y \int_{\Omega} u \frac{\partial^2 P_z}{\partial y^2} dV &= \kappa_y \int_{\Omega} \sum_j P_{zj} \left[ \frac{\partial}{\partial y} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) - \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} \right] dV \\ &= \kappa_y \sum_j P_{zj} \int_{\Omega} \frac{\partial}{\partial y} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) dV - \kappa_y \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV.\end{aligned}\quad (\text{B.14})$$

$$\begin{aligned}\kappa_z \int_{\Omega} u \frac{\partial^2 P_z}{\partial z^2} dV &= \kappa_z \int_{\Omega} \sum_j P_{zj} \left[ \frac{\partial}{\partial z} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) - \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} \right] dV \\ &= \kappa_z \sum_j P_{zj} \int_{\Omega} \frac{\partial}{\partial z} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) dV - \kappa_z \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV.\end{aligned}\quad (\text{B.15})$$

By divergence theorem, the first term in Eq. B.13~B.15 can be written as

$$\kappa_x \sum_j P_{zj} \int_{\Omega} \frac{\partial}{\partial x} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) dV = \kappa_x \sum_j P_{zj} \int_{\partial\Omega} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) \vec{x} \cdot d\vec{S}.\quad (\text{B.16})$$

$$\kappa_y \sum_j P_{zj} \int_{\Omega} \frac{\partial}{\partial y} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) dV = \kappa_y \sum_j P_{zj} \int_{\partial\Omega} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) \vec{y} \cdot d\vec{S}.\quad (\text{B.17})$$

$$\kappa_z \sum_j P_{zj} \int_{\Omega} \frac{\partial}{\partial z} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) dV = \kappa_z \sum_j P_{zj} \int_{\partial\Omega} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) \vec{z} \cdot d\vec{S}.\quad (\text{B.18})$$

At the surface of the ferroelectric material, we apply Neumann boundary condition.

Thus for surface elements we have

$$\frac{\partial P_{zj} \varphi_j}{\partial x} = 0, \quad \frac{\partial P_{zj} \varphi_j}{\partial y} = 0, \quad \frac{\partial P_{zj} \varphi_j}{\partial z} = 0.\quad (\text{B.19})$$

Eq. B.16-B.18 can be written as

$$\kappa_x \sum_j P_{zj} \int_{\partial\Omega} \left( \varphi_i \frac{\partial \varphi_j}{\partial x} \right) \vec{x} \cdot d\vec{S} = \kappa_x \sum_j \int_{\partial\Omega} \left( \varphi_i \frac{\partial P_{zj} \varphi_j}{\partial x} \right) \vec{x} \cdot d\vec{S},\quad (\text{B.20})$$

$$\kappa_y \sum_j P_{zj} \int_{\partial\Omega} \left( \varphi_i \frac{\partial \varphi_j}{\partial y} \right) \vec{y} \cdot d\vec{S} = \kappa_y \sum_j \int_{\partial\Omega} \left( \varphi_i \frac{\partial P_{zj} \varphi_j}{\partial y} \right) \vec{y} \cdot d\vec{S},\quad (\text{B.21})$$

$$\kappa_z \sum_j P_{zj} \int_{\partial\Omega} \left( \varphi_i \frac{\partial \varphi_j}{\partial z} \right) \vec{z} \cdot d\vec{S} = \kappa_z \sum_j \int_{\partial\Omega} \left( \varphi_i \frac{\partial P_{zj} \varphi_j}{\partial z} \right) \vec{z} \cdot d\vec{S}.\quad (\text{B.22})$$

thus they vanish and Eq. B.13-B.15 becomes

$$\kappa_x \int_{\Omega} u \frac{\partial^2 P_z}{\partial x^2} dV = -\kappa_x \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV. \quad (\text{B.23})$$

$$\kappa_y \int_{\Omega} u \frac{\partial^2 P_z}{\partial y^2} dV = -\kappa_y \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV. \quad (\text{B.24})$$

$$\kappa_z \int_{\Omega} u \frac{\partial^2 P_z}{\partial z^2} dV = -\kappa_z \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV. \quad (\text{B.25})$$

Bring Eq. B.3-B.6 and Eq. B.23-B.25 to Eq. B.2, we get

$$\begin{aligned} \int_{\Omega} \varphi_i \sum_j E_{zj} \varphi_j dV &= \alpha \int_{\Omega} \varphi_i \sum_j P_{zj} \varphi_j dV + \beta \int_{\Omega} \varphi_i \left( \sum_j P_{zj} \varphi_j \right)^3 dV \\ &+ \gamma \int_{\Omega} \varphi_i \left( \sum_j P_{zj} \varphi_j \right)^5 dV + \kappa_x \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV \\ &+ \kappa_y \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV + \kappa_z \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV, \end{aligned} \quad (\text{B.26})$$

$$\begin{aligned} \sum_j E_{zj} \int_{\Omega} \varphi_i \varphi_j dV &= \alpha \sum_j P_{zj} \int_{\Omega} \varphi_i \varphi_j dV + \beta \int_{\Omega} \varphi_i \left( \sum_j P_{zj} \varphi_j \right)^3 dV \\ &+ \gamma \int_{\Omega} \varphi_i \left( \sum_j P_{zj} \varphi_j \right)^5 dV + \kappa_x \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV \\ &+ \kappa_y \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV + \kappa_z \sum_j P_{zj} \int_{\Omega} \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV. \end{aligned} \quad (\text{B.27})$$

Here we define the notation

$$(a, b) = \int_{\Omega} ab dV, \quad (\text{B.28})$$

Thus we can write

$$\int_{\Omega} \varphi_i \varphi_j dV = (\varphi_i, \varphi_j), \quad (\text{B.29})$$

$$\int_{\Omega} \varphi_i \left( \sum_j P_{zj} \varphi_j \right)^3 dV = (\varphi_i, \left( \sum_j P_{zj} \varphi_j \right)^3), \quad (\text{B.30})$$

$$\int_{\Omega} \varphi_i \left( \sum_j P_{zj} \varphi_j \right)^5 dV = (\varphi_i, \left( \sum_j P_{zj} \varphi_j \right)^5), \quad (\text{B.31})$$

$$\int_{\Omega} \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV = \left( \frac{\partial \varphi_i}{\partial x}, \frac{\partial \varphi_j}{\partial x} \right). \quad (\text{B.32})$$

$$\int_{\Omega} \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV = \left( \frac{\partial \varphi_i}{\partial y}, \frac{\partial \varphi_j}{\partial y} \right). \quad (\text{B.33})$$

$$\int_{\Omega} \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV = \left( \frac{\partial \varphi_i}{\partial z}, \frac{\partial \varphi_j}{\partial z} \right). \quad (\text{B.34})$$

Bring Eq. B.29-B.34 to Eq. B.27, we have

$$\begin{aligned} \sum_j E_{zj}(\varphi_i, \varphi_j) &= \alpha \sum_j P_{zj}(\varphi_i, \varphi_j) + \beta(\varphi_i, \left( \sum_j P_{zj} \varphi_j \right)^3) + \gamma(\varphi_i, \left( \sum_j P_{zj} \varphi_j \right)^5) \\ &+ \kappa_x \sum_j P_{zj} \left( \frac{\partial \varphi_i}{\partial x}, \frac{\partial \varphi_j}{\partial x} \right) + \kappa_y \sum_j P_{zj} \left( \frac{\partial \varphi_i}{\partial y}, \frac{\partial \varphi_j}{\partial y} \right) + \kappa_z \sum_j P_{zj} \left( \frac{\partial \varphi_i}{\partial z}, \frac{\partial \varphi_j}{\partial z} \right). \end{aligned} \quad (\text{B.35})$$

Now we can define the weak form of the discrete problem: find  $P_{zj}$  so that Eq. B.35 is satisfied, where  $i = 0, 1, 2, 3, \dots, N-1$  and  $N$  is the number of shape function. The problem reads: find vector  $X, Y, U, V, C$  and  $G$  so that

$$KA = KX + U + V + MY + CN + GL, \quad (\text{B.36})$$

given vector  $A_j = E_{zj}$ , and

$$K_{ij} = (\varphi_i, \varphi_j), \quad (\text{B.37})$$

$$M_{ij} = \left( \frac{\partial \varphi_i}{\partial x}, \frac{\partial \varphi_j}{\partial x} \right), \quad (\text{B.38})$$

$$N_{ij} = \left( \frac{\partial \varphi_i}{\partial y}, \frac{\partial \varphi_j}{\partial y} \right), \quad (\text{B.39})$$

$$L_{ij} = \left( \frac{\partial \varphi_i}{\partial z}, \frac{\partial \varphi_j}{\partial z} \right), \quad (\text{B.40})$$

$$X_j = \alpha P_{zj}, \quad (\text{B.41})$$

$$U_j = \beta(\varphi_j, \left( \sum_m P_{zm} \varphi_m \right)^3), \quad (\text{B.42})$$

$$V_j = \gamma(\varphi_j, \left( \sum_m P_{zm} \varphi_m \right)^5), \quad (\text{B.43})$$

$$Y_j = \kappa_x P_{zj} \quad (\text{B.44})$$

$$C_j = \kappa_y P_{zj} \quad (\text{B.45})$$

$$G_j = \kappa_z P_{zj}. \quad (\text{B.46})$$

### B.3 Solve with Newton's method

This is a non-linear equation and we can solve it iteratively using Newton's method

$$P_z^{n+1} = P_z^n - J^{-1} \mathcal{F}, \quad (\text{B.47})$$

where

$$\mathcal{F}_i = \sum_j K_{ij} A_j - \sum_j K_{ij} X_j - U_i - V_i - \sum_j M_{ij} Y_j - \sum_j N_{ij} C_j - \sum_j L_{ij} G_j, \quad (\text{B.48})$$

and

$$\begin{aligned} J_{ik} &= \frac{\partial \mathcal{F}_i}{\partial P_{zk}} \\ &= - \sum_j K_{ij} \frac{\partial X_j}{\partial P_{zk}} - \frac{\partial U_i}{\partial P_{zk}} - \frac{\partial V_i}{\partial P_{zk}} - \sum_j M_{ij} \frac{\partial Y_j}{\partial P_{zk}} - \sum_j N_{ij} \frac{\partial C_j}{\partial P_{zk}} - \sum_j L_{ij} \frac{\partial G_j}{\partial P_{zk}} \\ &= -\alpha \sum_j K_{ij} \delta_{jk} - \beta \frac{\partial(\varphi_i, (\sum_m P_{zm} \varphi_m)^3)}{\partial P_{zk}} - \gamma \frac{\partial(\varphi_i, (\sum_m P_{zm} \varphi_m)^5)}{\partial P_{zk}} - \\ &\quad \kappa_x \sum_j M_{ij} \delta_{kj} - \kappa_y \sum_j N_{ij} \delta_{kj} - \kappa_z \sum_j L_{ij} \delta_{kj}. \end{aligned} \quad (\text{B.49})$$

The second term in Eq. B.49 can be written as

$$\begin{aligned}
\beta \frac{\partial(\varphi_i, (\sum_m P_{zm}\varphi_m)^3)}{\partial P_{zk}} &= \beta \frac{\int_{\Omega} \varphi_i (\partial \sum_m P_{zm}\varphi_m)^3 dV}{\partial P_{zk}} \\
&= \beta \int_{\Omega} \varphi_i \frac{\partial(\sum_m P_{zm}\varphi_m)^3}{\partial P_{zk}} dV \\
&= \beta \int_{\Omega} \varphi_i \cdot 3(\sum_m P_{zm}\varphi_m)^2 \cdot \frac{\partial(\sum_m P_{zm}\varphi_m)}{\partial P_{zk}} dV \\
&= \beta \int_{\Omega} \varphi_i \cdot 3(\sum_m P_{zm}\varphi_m)^2 \cdot (\sum_m \varphi_m \delta_{mk}) dV \\
&= 3\beta \int_{\Omega} \varphi_i \varphi_k (\sum_m P_{zm}\varphi_m)^2 dV. \tag{B.50}
\end{aligned}$$

Similarly, the third term in Eq. B.49 can be written as

$$\begin{aligned}
\gamma \frac{\partial(\varphi_i, (\sum_m P_{zm}\varphi_m)^5)}{\partial P_{zk}} &= \gamma \frac{\int_{\Omega} \varphi_i (\partial \sum_m P_{zm}\varphi_m)^5 dV}{\partial P_{zk}} \\
&= \gamma \int_{\Omega} \varphi_i \frac{\partial(\sum_m P_{zm}\varphi_m)^5}{\partial P_{zk}} dV \\
&= \gamma \int_{\Omega} \varphi_i \cdot 5(\sum_m P_{zm}\varphi_m)^4 \cdot \frac{\partial(\sum_m P_{zm}\varphi_m)}{\partial P_{zk}} dV \\
&= \gamma \int_{\Omega} \varphi_i \cdot 5(\sum_m P_{zm}\varphi_m)^4 \cdot (\sum_m \varphi_m \delta_{mk}) dV \\
&= 5\gamma \int_{\Omega} \varphi_i \varphi_k (\sum_m P_{zm}\varphi_m)^4 dV. \tag{B.51}
\end{aligned}$$

Bring Eq. B.50 and Eq. B.51 to Eq. B.49, we have

$$\begin{aligned}
J_{ik} &= -\alpha K_{ik} - 3\beta \int_{\Omega} \varphi_i \varphi_k (\sum_m P_{zm}\varphi_m)^2 dV - 5\gamma \int_{\Omega} \varphi_i \varphi_k (\sum_m P_{zm}\varphi_m)^4 dV - \\
&\quad \kappa_x M_{ik} - \kappa_y N_{ik} - \kappa_z L_{ik}. \tag{B.52}
\end{aligned}$$

#### B.4 Solve with iterative method I

For a given electrical field, there can be more than one polarization fields that satisfy the L-K equation. In order to capture the hysteresis in the simulation, we need to take the polarization field which is close to the one in the last step. It is possible that Newton's method converges to the unwanted solution, therefore, alternatively, we can

solve the L-K equation using an iterative scheme starting with the polarization in the last step. In the  $n^{\text{th}}$  iteration, the linear equation is solved for  $S^{n+1}$ ,

$$\alpha K S^{n+1} = KA - U^n - V^n - MY^n - NC^n - LG^n, \quad (\text{B.53})$$

where  $U^n$ ,  $V^n$  and  $Y^n$  are given by

$$S_j^{n+1} = P_{zj}^{n+1}, \quad (\text{B.54})$$

$$U_j^n = \beta(\varphi_j, (\sum_m P_{zm}^n \varphi_m)^3), \quad (\text{B.55})$$

$$V_j^n = \gamma(\varphi_j, (\sum_m P_{zm}^n \varphi_m)^5), \quad (\text{B.56})$$

$$Y_j^n = \kappa_x P_{zj}^n \quad (\text{B.57})$$

$$C_j^n = \kappa_y P_{zj}^n \quad (\text{B.58})$$

$$G_j^n = \kappa_z P_{zj}^n. \quad (\text{B.59})$$

In the first iteration, we can take  $S^0$  from the converged solution of the last step.

#### B.4.1 Computing with trapezoidal integration

By using trapezoidal integration, the quadrature points coincide with the FEM nodes, thus the shape functions become orthogonal. The scheme is known as the discontinu-

ous Galerkin scheme since discontinuous test functions are used who only have values defined at FEM nodes. We first split integrals over  $\Omega$  into integrals over all cells,

$$K_{ij} = (\varphi_i, \varphi_j) = \sum_{R \in T} \int_R \varphi_i \varphi_j dV, \quad (\text{B.60})$$

$$M_{ij} = \left( \frac{\partial \varphi_i}{\partial x}, \frac{\partial \varphi_j}{\partial x} \right) = \sum_{R \in T} \int_R \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV, \quad (\text{B.61})$$

$$N_{ij} = \left( \frac{\partial \varphi_i}{\partial y}, \frac{\partial \varphi_j}{\partial y} \right) = \sum_{R \in T} \int_R \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV, \quad (\text{B.62})$$

$$L_{ij} = \left( \frac{\partial \varphi_i}{\partial z}, \frac{\partial \varphi_j}{\partial z} \right) = \sum_{R \in T} \int_R \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV, \quad (\text{B.63})$$

$$\begin{aligned} U_j^n &= \beta(\varphi_j, (\sum_m P_{zm}^n \varphi_m)^3) = \beta \sum_{R \in T} \int_R \varphi_j (\sum_m P_{zm}^n \varphi_m)^3 dV \\ &= \beta \sum_{R \in T} \int_R \varphi_j (P_{zj}^n \varphi_j)^3 dV, \end{aligned} \quad (\text{B.64})$$

$$\begin{aligned} V_j^n &= \gamma(\varphi_j, (\sum_m P_{zm}^n \varphi_m)^5) = \gamma \sum_{R \in T} \int_R \varphi_j (\sum_m P_{zm}^n \varphi_m)^5 dV \\ &= \gamma \sum_{R \in T} \int_R \varphi_j (P_{zj}^n \varphi_j)^5 dV, \end{aligned} \quad (\text{B.65})$$

and then approximate each cell's contribution by quadrature

$$K_{ij}^R = \int_R \varphi_i \varphi_j dV \approx \sum_q \varphi_i(r_q^R) \cdot \varphi_j(r_q^R) \cdot w_q^R, \quad (\text{B.66})$$

$$M_{ij}^R = \int_R \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV \approx \sum_q \frac{\partial \varphi_i(r_q^R)}{\partial x} \cdot \frac{\partial \varphi_j(r_q^R)}{\partial x} \cdot w_q^R, \quad (\text{B.67})$$

$$N_{ij}^R = \int_R \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV \approx \sum_q \frac{\partial \varphi_i(r_q^R)}{\partial y} \cdot \frac{\partial \varphi_j(r_q^R)}{\partial y} \cdot w_q^R, \quad (\text{B.68})$$

$$L_{ij}^R = \int_R \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV \approx \sum_q \frac{\partial \varphi_i(r_q^R)}{\partial z} \cdot \frac{\partial \varphi_j(r_q^R)}{\partial z} \cdot w_q^R, \quad (\text{B.69})$$

$$U_j^{nR} = \beta \int_R \varphi_j (P_{zj}^n \varphi_j)^3 dV \approx \beta \sum_q [\varphi_j(r_q^R)]^4 \cdot (P_{zj}^n)^3 \cdot w_q^R, \quad (\text{B.70})$$

$$V_j^{nR} = \gamma \int_R \varphi_j (P_{zj}^n \varphi_j)^5 dV \approx \gamma \sum_q [\varphi_j(r_q^R)]^6 \cdot (P_{zj}^n)^5 \cdot w_q^R, \quad (\text{B.71})$$

where  $r_q^R$  is the  $q^{th}$  quadrature point on cell  $R$ , and  $w_q^R$  is the  $q^{th}$  quadrature weight. Since matrix  $K$  is a diagonal matrix, we will calculate the following equation for  $S^{n+1}$  in every iteration,

$$\alpha K S^{n+1} = K A - U^n - V^n - M Y^n - N C^n - L G^n. \quad (\text{B.72})$$

#### B.4.2 Computing with high order integration

For high order integration, the quadrature points also appear at locations other than the FEM nodes and we can no longer assume the shape functions to be orthogonal. Eqs. B.60-B.63 still hold, Eqs. B.64 and B.65 become

$$U_j^n = \beta(\varphi_j, (\sum_m P_{zm}^n \varphi_m)^3) = \beta \sum_{R \in T} \int_R \varphi_j (\sum_m P_{zm}^n \varphi_m)^3 dV, \quad (\text{B.73})$$

$$V_j^n = \gamma(\varphi_j, (\sum_m P_{zm}^n \varphi_m)^5) = \gamma \sum_{R \in T} \int_R \varphi_j (\sum_m P_{zm}^n \varphi_m)^5 dV. \quad (\text{B.74})$$

Then, Eqs. B.66-B.69 still hold, Eqs. B.70 and B.71 become

$$U_j^{nR} = \beta \int_R \varphi_j (\sum_m P_{zm}^n \varphi_m)^3 dV \approx \beta \sum_q \left[ \varphi_j(r_q^R) \cdot (\sum_m P_{zm}^n \varphi_m(r_q^R))^3 \cdot w_q^R \right], \quad (\text{B.75})$$

$$V_j^{nR} = \gamma \int_R \varphi_j (\sum_m P_{zm}^n \varphi_m)^5 dV \approx \gamma \sum_q \left[ \varphi_j(r_q^R) \cdot (\sum_m P_{zm}^n \varphi_m(r_q^R))^5 \cdot w_q^R \right]. \quad (\text{B.76})$$

The summation over  $m$  in Eqs. B.75 and B.76 loops over every quadrature points of the element where the  $j^{th}$  degree of freedom is at. Obviously, Eq. B.72 still holds.

#### B.5 Solve with iterative method II

It is found that when using the method in Sec. B.4, the iterative scheme converges to the meta-stable state when the initial guess is close to it, otherwise, the iterative scheme explodes. To achieve convergence to the stable states, we found that the following method works. Again, the trapezoidal integration rule is used. In the  $n^{th}$  iteration, the linear equation is solved for  $Q^{n+1}$ ,

$$\beta D Q^{n+1} = K A - \alpha K S^n - \gamma F^n - \kappa_x M S^n - \kappa_y N S^n - \kappa_z L S^n, \quad (\text{B.77})$$

where  $D$ ,  $S$  and  $F$  are given by

$$D_{ii} = (\varphi_i, \varphi_i^3), \quad (\text{B.78})$$

$$S_j^n = P_{zj}, \quad (\text{B.79})$$

$$F_j^n = (\varphi_j, \varphi_j^5) \cdot (S_j^n)^5. \quad (\text{B.80})$$

In the  $(n + 1)^{th}$  iteration, we use

$$S_j^{n+1} = \pm \sqrt{\frac{Q_j^{n+1}}{S_j^n}}, \quad (\text{B.81})$$

where the plus or minus sign gives us either one of the two stable states. We first split integrals over  $\Omega$  into integrals over all cells,

$$K_{ij} = (\varphi_i, \varphi_j) = \sum_{R \in T} \int_R \varphi_i \varphi_j dV, \quad (\text{B.82})$$

$$M_{ij} = \left( \frac{\partial \varphi_i}{\partial x}, \frac{\partial \varphi_j}{\partial x} \right) = \sum_{R \in T} \int_R \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV, \quad (\text{B.83})$$

$$N_{ij} = \left( \frac{\partial \varphi_i}{\partial y}, \frac{\partial \varphi_j}{\partial y} \right) = \sum_{R \in T} \int_R \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV, \quad (\text{B.84})$$

$$L_{ij} = \left( \frac{\partial \varphi_i}{\partial z}, \frac{\partial \varphi_j}{\partial z} \right) = \sum_{R \in T} \int_R \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV, \quad (\text{B.85})$$

$$D_{ii} = (\varphi_i, \varphi_i^3) = \sum_{R \in T} \int_R \varphi_i \varphi_i^3 dV, \quad (\text{B.86})$$

$$F_j^n = (\varphi_j, \varphi_j^5) \cdot (S_j^n)^5 = \sum_{R \in T} \int_R \varphi_j \varphi_j^5 dV \cdot (S_j^n)^5, \quad (\text{B.87})$$

and then approximate each cell's contribution by quadrature

$$K_{ij}^R = \int_R \varphi_i \varphi_j dV \approx \sum_q \varphi_i(r_q^R) \cdot \varphi_j(r_q^R) \cdot w_q^R, \quad (\text{B.88})$$

$$M_{ij}^R = \int_R \frac{\partial \varphi_i}{\partial x} \frac{\partial \varphi_j}{\partial x} dV \approx \sum_q \frac{\partial \varphi_i(r_q^R)}{\partial x} \cdot \frac{\partial \varphi_j(r_q^R)}{\partial x} \cdot w_q^R, \quad (\text{B.89})$$

$$N_{ij}^R = \int_R \frac{\partial \varphi_i}{\partial y} \frac{\partial \varphi_j}{\partial y} dV \approx \sum_q \frac{\partial \varphi_i(r_q^R)}{\partial y} \cdot \frac{\partial \varphi_j(r_q^R)}{\partial y} \cdot w_q^R, \quad (\text{B.90})$$

$$L_{ij}^R = \int_R \frac{\partial \varphi_i}{\partial z} \frac{\partial \varphi_j}{\partial z} dV \approx \sum_q \frac{\partial \varphi_i(r_q^R)}{\partial z} \cdot \frac{\partial \varphi_j(r_q^R)}{\partial z} \cdot w_q^R, \quad (\text{B.91})$$

$$D_{ii}^R = \int_R \varphi_i \varphi_i^3 dV \approx \sum_q [\varphi_i(r_q^R)]^4 \cdot w_q^R, \quad (\text{B.92})$$

$$F_j^{nR} = \int_R \varphi_j \varphi_j^5 dV \cdot (S_j^n)^5 \approx \sum_q [\varphi_j(r_q^R)]^6 \cdot w_q^R \cdot (S_j^n)^5, \quad (\text{B.93})$$

where  $r_q^R$  is the  $q^{\text{th}}$  quadrature point on cell  $R$ , and  $w_q^R$  is the  $q^{\text{th}}$  quadrature weight.

## C. AN ELECTRO-THERMAL COUPLED QUANTUM TRANSPORT MODEL

### C.1 Introduction

The nonequilibrium Green's function (NEGF) formalism is widely accepted as one of the most consistent models for transport properties in nanodevices in the presence of quantum phenomena including quantum confinement, tunneling, interferences, etc [30,31]. However, a fully quantum mechanical model to study the coupled electron-phonon transport via NEGF is numerically expensive. Therefore, its application is normally limited to small systems composed of 100 atoms or less [156].

In this work, an electro-thermal quantum transport model is developed in the NEGF formalism. Scattering in the electron and phonon transport is treated by a phenomenological model using so-called Büttiker probes (BP) [75,157]. To avoid solving the full Green's functions, the Büttiker probe algorithm is altered to be compatible with recursive Green's function (RGF) [44]. The Program flowchart of the electro-thermal quantum transport model is shown in Fig. C.1.

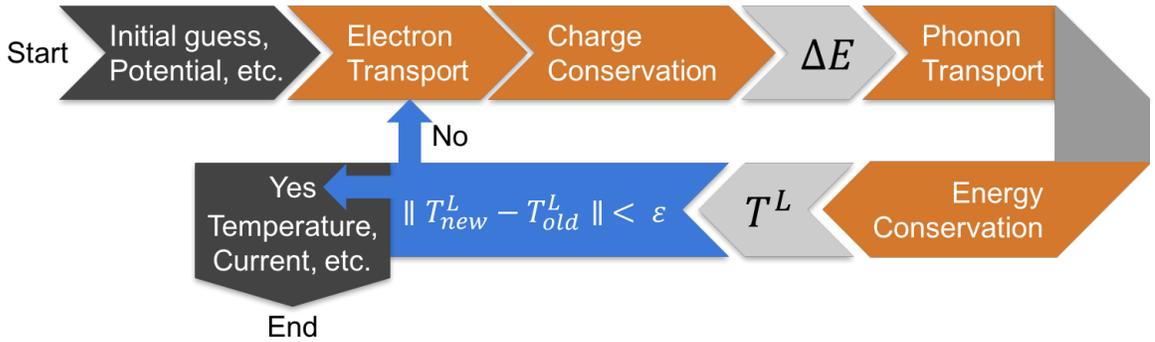


Fig. C.1.: Program flowchart of the electro-thermal coupled quantum transport model.

## C.2 Büttiker probe RGF algorithm

The RGF methodology for solving the block diagonals of  $G^r$  and  $G^<$  can be found in Ref. [44,158]. Here, the modifications needed to combine RGF with Büttiker probe (BP) are provided. All Büttiker probes within a RGF block are assumed to have the same local Fermi level and temperature for electrons and phonons, respectively.

In the forward loop, the equation for left-connected retarded Green's function  $g_{ii}^r$  ( $i$  denotes the RGF block index and  $i = 1, 2, \dots, N$ , where  $i = 1$  is the first block right to the source contact and  $N$  is the first block left to the drain contact) is given by

$$g_{i,i}^r = (D_{i,i} - t_{i,i-1}g_{i-1,i-1}^r t_{i-1,i})^{-1}, \quad (\text{C.1})$$

where  $D_{i,i} = (EI - H_{i,i} - \Sigma_S - \Sigma_D - \Sigma_{BP})$  for electrons and  $D_{i,i} = (E^2I - H_{i,i} - \Sigma_S - \Sigma_D - \Sigma_{BP})$  for phonons.  $H_{i,i}$  is the onsite Hamiltonian or Dynamical matrix for electrons and phonons, respectively.  $t_{i,i-1}$  and  $t_{i-1,i}$  are the inter-block coupling matrices. The boundary condition for the block with index  $i = 1$  is

$$g_{1,1}^r = D_{1,1}^{-1}. \quad (\text{C.2})$$

Similarly, the equation for left-connected lesser Green's function  $g_{ii}^<$  is given by

$$g_{i,i}^< = g_{i,i}^r [\Sigma_{i,i}^< + t_{i,i-1}g_{i-1,i-1}^< t_{i-1,i}] g_{i,i}^{r\dagger}, \quad (\text{C.3})$$

where  $\Sigma_{i,i}^< = i\Gamma_{i,i}\mathfrak{S}_i$ .  $\Gamma_{i,i} = i[\Sigma_{i,i} - \Sigma_{i,i}^\dagger]$  and  $\Sigma_{i,i}$  is the sum of the contact self energy and the BP self energy.  $\mathfrak{S}_i$  is the RGF block resolved equilibrium distribution of the  $i^{\text{th}}$  block for electrons or phonons. Equilibrium distribution of Electrons and phonons is assumed in Büttiker probes. Thus,  $\mathfrak{S}_i$  takes shape of the Fermi distribution or the Bose-Einstein distribution for electrons or phonons, respectively. The boundary condition for the block with index  $i = 1$  is

$$g_{1,1}^< = g_{1,1}^r \Sigma_{1,1}^< g_{1,1}^{r\dagger}. \quad (\text{C.4})$$

The forward loop runs over  $i = 1, 2, \dots, N$ .

In the backward loop, the equation for the retarded Green's function is given by

$$G_{i,i}^r = g_{i,i}^r (I + t_{i,i+1} G_{i+1,i+1}^r t_{i+1,i} g_{i,i}^r), \quad (\text{C.5})$$

$$G_{i,i+1}^r = -g_{i,i}^r t_{i,i+1} G_{i+1,i+1}^r, \quad (\text{C.6})$$

with the boundary condition for the block with index  $i = N$  being

$$G_{N,N}^r = g_{N,N}^r. \quad (\text{C.7})$$

the equation for the lesser Green's function is given by

$$G_{i,i}^< = g_{i,i}^< + g_{i,i}^r (t_{i,i+1} G_{i+1,i+1}^< t_{i+1,i}) g_{i,i}^{r\dagger} - g_{i,i}^< t_{i,i+1} G_{i+1,i+1}^r - G_{i+1,i+1}^r t_{i+1,i} g_{i,i}^<, \quad (\text{C.8})$$

with the boundary condition for the block with index  $i = N$  being

$$G_{N,N}^< = g_{N,N}^<. \quad (\text{C.9})$$

The backward loop runs over  $i = N, N - 1, \dots, 1$ .

For current conservation, the derivative of the diagonal blocks of  $G^<$  with respect to the Büttiker probe Fermi levels or temperatures are also needed. Therefore, in the forward loop, the derivative of  $g_{ii}^<$  with respect to the Büttiker probe Fermi levels or temperatures is given by

$$\frac{\partial g_{i,i}^<}{\partial L_j} = \begin{cases} g_{i,i}^r t_{i,i-1} \frac{\partial g_{i-1,i-1}^<}{\partial L_j} t_{i-1,i} g_{i,i}^{r\dagger}, & \text{if } j < i, \\ i g_{i,i}^r \Gamma_{BP,i,i} g_{i,i}^{r\dagger} \frac{\partial \mathfrak{S}_i}{\partial L_j}, & \text{if } j = i, \\ 0, & \text{if } j > i, \end{cases} \quad (\text{C.10})$$

where  $L_j$  is probe Fermi level or temperature of the  $j^{\text{th}}$  block for electrons or phonons, respectively. To reduce the computational intensity, only the diagonal blocks of Eq. C.10 ( $j = i$ ) are solved. The boundary condition for the block with index  $i = 1$  is

$$\frac{\partial g_{1,1}^<}{\partial L_j} = \begin{cases} 0, & \text{if } j \neq 1, \\ i g_{1,1}^r \Gamma_{BP,1,1} g_{1,1}^{r\dagger} \frac{\partial \mathfrak{S}_1}{\partial L_1}, & \text{if } j = 1, \end{cases} \quad (\text{C.11})$$

In the backward loop, the derivative of  $G_{i,i}^<$  with respect to the Büttiker probe Fermi levels or temperature is given by

$$\frac{\partial G_{i,i}^<}{\partial L_j} = \frac{\partial g_{i,i}^<}{\partial L_j} + g_{i,i}^r \left( t_{i,i+1} \frac{\partial G_{i+1,i+1}^<}{\partial L_j} t_{i+1,i} \right) g_{i,i}^{r\dagger} - \frac{\partial g_{i,i}^<}{\partial L_j} t_{i,i+1} G_{i,i+1}^{r\dagger} - G_{i,i+1}^r t_{i+1,i} \frac{\partial g_{i,i}^<}{\partial L_j}, \quad (\text{C.12})$$

with the boundary condition for the block with index  $i = N$  being

$$\frac{\partial G_{N,N}^<}{\partial L_j} = \frac{\partial g_{N,N}^<}{\partial L_j}. \quad (\text{C.13})$$

For computational burden, only the diagonal blocks of Eq. C.12 ( $j = i$ ) are solved. The current density of the probe with index  $i$  is calculated using the current operator [79]. The equation is given by

$$I_i = \frac{q^2}{h} \sum_k \int Tr \left( \Gamma_{i,i} A_{i,i} \mathfrak{S}_i + i \Gamma_{i,i} G_{i,i}^< \right) dE \quad (\text{C.14})$$

for electron charge current density, and

$$Q_i = \frac{q}{h} \sum_k \int E Tr \left( \Gamma_{i,i} A_{i,i} \mathfrak{S}_i + i \Gamma_{i,i} G_{i,i}^< \right) dE \quad (\text{C.15})$$

for phonon energy current density, where  $A_{i,i}$  is the spectral function and is defined as

$$A_{i,i} = i \left( G_{i,i}^r - G_{i,i}^{r\dagger} \right). \quad (\text{C.16})$$

The diagonal element  $(i, i)^{th}$  in the Jacobian matrix is given by

$$J_{i,i} = \frac{q^2}{h} \sum_k \int Tr \left( \Gamma_{i,i} A_{i,i} \frac{\partial \mathfrak{S}_i}{\partial L_i} + i \Gamma_{i,i} \frac{\partial G_{i,i}^<}{\partial L_i} \right) dE \quad (\text{C.17})$$

for electrons, and

$$J_{i,i} = \frac{q}{h} \sum_k \int E Tr \left( \Gamma_{i,i} A_{i,i} \frac{\partial \mathfrak{S}_i}{\partial L_i} + i \Gamma_{i,i} \frac{\partial G_{i,i}^<}{\partial L_i} \right) dE \quad (\text{C.18})$$

for phonons. To obtain the off-diagonal elements in the Jacobian matrix, it is required to calculate off-diagonal blocks of Eq. C.10 and Eq. C.12. Although the off-diagonal elements of the Jacobian matrix should help to achieve convergence with a smaller number of iterations, it increases the computational burden dramatically. Therefore,

there is a trade-off in whether to calculate the off-diagonal elements of the Jacobian matrix.

The algorithm for decoupled electron or phonon transport proceeds as follows:

(1) Start with an initial guess for Büttiker probe Fermi levels (electron) or temperatures (phonon).

(2) For each energy-momentum combination, calculate the forward and backward loop to get  $G_{i,i}^r$ ,  $G_{<i,i}$  and  $\frac{\partial G_{i,i}^<}{\partial L_i}$  using Eq. C.5, Eq. C.8 and Eq. C.12.

(3) Calculate charge current density (electron) or energy current density (phonon) in every Büttiker probe using Eq. C.14 and Eq. C.15.

(4) Calculate the Jacobian matrix using Eq. C.17 and Eq. C.18.

(5) Update the Büttiker probe Fermi levels (electron) or temperatures (phonon) using Newton step, which is given by

$$E_{F_{new}} = E_{F_{old}} - J^{-1} \times I_{old} \quad (\text{C.19})$$

for electrons, and

$$T_{new}^L = T_{old}^L - J^{-1} \times Q_{old} \quad (\text{C.20})$$

for phonons.

(6) If  $\|E_{F_{new}} - E_{F_{old}}\|$  or  $\|T_{new}^L - T_{old}^L\|$  is larger than the convergence threshold, go back to step (1) and proceed with the new Büttiker probe Fermi levels or temperatures, otherwise, the algorithm is considered to be converged.

In order to simulate self-heating effects in the device, the energy exchange between electrons and phonons is calculated by assuming that electrons leave the Büttiker probe at the Fermi level of the probe. Therefore, the loss of electron energy at the  $i^{th}$  block is given by

$$\Delta E_i = \frac{1}{2\pi} \sum_k \int (E - E_{F_i}) A_{i,i} dE. \quad (\text{C.21})$$

The energy is then absorbed by phonons at the  $i^{th}$  block which affects the phonon distribution in the device. The local lattice temperature of the  $i^{th}$  block  $T_i^L$  is obtained

by equating the local phonon energy density to the product of the local phonon density of states with phonon energy and a Bose-Einstein distribution:

$$\frac{1}{\pi} \sum_k \int -i\hbar\omega G_{i,i}^{<}(\hbar\omega) = \frac{1}{\pi} \sum_k \int \hbar\omega A_{i,i} f_{BE}(\omega, T_i^L) d(\hbar\omega). \quad (\text{C.22})$$

Once the local lattice temperature is obtained, it will be used in the Fermi distribution function for electrons in the electron Büttiker probe. It is obvious that the electron and phonon transport must be solved iteratively until self-consistency is achieved since Eq. C.21 depends on the lattice temperature (output of the phonon transport) and Eq. C.22 depends on the energy exchange between electrons and phonons (output of the electron transport).

VITA

## VITA

Yuanchen Chu was born in Shanghai, China on the 30th of May, 1991. From 2009 to 2013, he studied electronic engineering in Tongji University in Shanghai, China, and in Polytechnic University of Milan in Milan, Italy. In July 2013, he received dual Bachelor's degrees from Tongji University and Polytechnic University of Milan. In the same year, he joined Purdue university, as a direct PhD student in the department of Electrical and Computer Engineering, and became a research assistant in the NEMO group. His research interests are in the development of electron and phonon quantum transport models and its application in nanoscale devices.