UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

# Full-Band Monte Carlo Simulations for Vertical Impact Ionization MOSFETs

Thanh Viet Dinh

Vorsitzender des Promotionsausschusses: Prof. Dr.-Ing. Walter Hansch
1. Berichter: Prof. Dr.-Ing. Christoph Jungemann
2. Berichter: PD Dr.-Ing. Habil. Rainer Kraus

Tag der Prüfung 21.09.2010

Mit der Promotion erlangter akademischer Grad:
Doktor-Ingenieur
(Dr.-Ing.)

Neubiberg, den 23. September 2010

# Contents


**1 Introduction** **1**
1.1 Motivation . . . . . 1
1.2 Scope of work . . . . . 3
1.3 Structure . . . . . 4

**2 Monte Carlo simulator for nanoscale devices** **5**
2.1 Basic equations for the electrical transport in semiconductor devices . . . . . 5
2.1.1 The Boltzmann transport equation . . . . . 5
2.1.2 The Poisson equation . . . . . 7
2.2 Monte Carlo simulator . . . . . 8
2.2.1 Introduction of the Monte Carlo method . . . . . 8
2.2.2 The Monte Carlo solver for the Boltzmann equation . . 10
2.2.3 Scattering mechanisms . . . . . 13
2.2.4 Device simulation . . . . . 15
2.3 Moment-based simulators . . . . . 16
2.4 Summary . . . . . 20

**3 Strain engineering** **22**
3.1 Stress - strain under elasticity condition . . . . . 22
3.1.1 Stress . . . . . 22
3.1.2 Strain . . . . . 23
3.1.3 Stress-strain dependence . . . . . 25
3.2 Strain effects on crystal symmetry and electronic band structures 28
3.2.1 Basic properties of diamond structures . . . . . 28
3.2.2 Strain effects on crystal symmetry . . . . . 30
3.2.3 Strain effects on electronic band structures . . . . . 30
3.3 Band structure calculation . . . . . 32
3.3.1 Pseudopotential method . . . . . 32
3.3.2 Nonlocal pseudopotential method . . . . . 34
3.3.3 Empirical pseudopotential method . . . . . 35

3.4 Band offsets and band gap for heterostructures $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ - a special case of strain . . . 37
3.4.1 Without additional uniaxial stress . . . 39
3.4.2 With additional uniaxial stress . . . 46
3.5 Summary . . . 49

**4 Impact ionization for strained materials 50**
4.1 Theory of impact ionization . . . 50
4.2 Calculation approach for impact ionization rate . . . 53
4.2.1 BZ setup . . . 54
4.2.2 Calculation of the integral $I_2$ . . . 55
4.2.3 Calculation of the integral $I_1$ . . . 59
4.2.4 Another approach for calculating $I_2$ . . . 60
4.2.5 Anisotropic and isotropic impact ionization rate . . . 62
4.3 Impact ionization in Monte Carlo simulations . . . 63
4.3.1 Maximum impact ionization rate for a tetrahedron . . 63
4.3.2 Impact ionization coefficient and quantum yield . . . 64
4.4 Results . . . 64
4.4.1 Relaxed Si . . . 64
4.4.2 Strained SiGe . . . 67
4.4.3 Uniaxially and biaxially strained Si . . . 70
4.5 Summary . . . 74

**5 Vertical impact ionization MOSFETs with strain engineering 75**
5.1 Review of IMOS transistors . . . 75
5.1.1 Lateral IMOS . . . 75
5.1.2 Self-aligned IMOS . . . 76
5.1.3 L-shaped IMOS . . . 78
5.1.4 Depletion IMOS . . . 79
5.2 Vertical IMOS transistors . . . 80
5.2.1 Device structure . . . 80
5.2.2 Vertical IMOS with a strained SiGe layer . . . 83
5.3 Device simulations . . . 85
5.3.1 Combination of MC and HD simulators . . . 85
5.3.2 Full-band Monte Carlo simulator . . . 85
5.3.3 Hydrodynamic simulator . . . 87
5.4 Results and Discussions . . . 87
5.4.1 Full-band Monte Carlo simulations . . . 87
5.4.2 Hydrodynamic simulations . . . 92
5.5 Noise investigation in vertical IMOS transistors . . . 95

5.5.1 Simulation approach . . . . . . . . . . . . . . . . . . . . . 95
5.5.2 Results and discussions . . . . . . . . . . . . . . . . . . . . 97
5.6 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101

**6 Conclusions 104**
6.1 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
6.2 Future works . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105

# List of Figures


2.1 Flowchart of the MC method for a simulation within a duration $T$ [59]. . . . 12
2.2 Flowchart of the self-consistent MC device simulation [59]. . . 15

3.1 Components of a stress tensor. . . . 23
3.2 Stress direction $[x', y', z']$ in the crystallographic coordinate system $[x, y, z]$. . . . 27
3.3 The first BZ of relaxed Si and the first irreducible wedge. . . . 29
3.4 (a) Cubic crystals under in-plane biaxial tensile stress (b) Diagrams of band splitting of Si and Ge under in-plane biaxial tensile stress. . . . 31
3.5 (a) Cubic crystals under a uniaxial compressive stress along the [110] direction (b) Diagrams of band splitting of Si and Ge under uniaxial compressive stress. . . . 32
3.6 Schematic plot of the atomic pseudopotential of Si in: (a) real space and (b) reciprocal space ($q_1 = \sqrt{3}, q_2 = \sqrt{8}, q_3 = \sqrt{11}$) [25]. 33
3.7 Flow diagram for calculating band structures with the EPM method [128]. . . . 36
3.8 Unit cell of the diamond structure with a center atom and its four nearest neighbors: (a) Unstrained crystal: $a_1 = a_2 = a_3 = a_4$; (b) Strained crystal: $a_1 \neq a_2 \neq a_3 \neq a_4$; (c) Center atom is displaced to obtain $a_1 \approx a_2 \approx a_3 \approx a_4$ [122]. . . . 37
3.9 Two types of band alignments in Si/SiGe heterostructures. . . 39
3.10 Contour plots for the minimum valence band offsets (in meV) of $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ interfaces with the [100] growth direction. 41
3.11 Contour plots for the minimum conduction band offsets (in meV) of $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ interfaces with the [100] growth direction. . . . 44
3.12 Band gap (in eV) of relaxed $Si_{1-y}Ge_y$ alloys. . . . 45
3.13 Contour plots for the band gaps (in eV) of $Si_{1-x}Ge_x$ alloys (active layer) grown pseudomorphically along the [100] direction on unstrained $Si_{1-y}Ge_y$ alloys (substrate layer). . . . 46

3.14 Band gap of the active layer under a uniaxial stress of -1.0 GPa along the [010] direction. . . . . . . . . . . . . . . . . . . . 48
3.15 Band gap of the active layer under a uniaxial stress of -1.0 GPa along the [011] direction. . . . . . . . . . . . . . . . . . . . 49

4.1 Diagram of the electron-initiated impact ionization process. . . 51
4.2 (a) Transformation from a normal BZ into a cuboid BZ (b) Break up of cuboid BZ into small cubes. . . . . . . . . . . . 54
4.3 Break up a cube into 6 similar tetrahedra. . . . . . . . . . . . 56
4.4 Intersection area of an equienergy plane with a tetrahedron: ABC or ABCD. . . . . . . . . . . . . . . . . . . . . . . . . . 57
4.5 Intersection area of an equienergy plane with a sphere in 2D visualization. . . . . . . . . . . . . . . . . . . . . . . . . . . 61
4.6 Cubes covering the tetrahedron in 2D. . . . . . . . . . . . . . 64
4.7 Isotropic impact ionization rate with the initial electron in the first conduction band for relaxed Si. N is the number of grid points between 0 - $2\pi/a$ in the cuboid BZ. . . . . . . . . . . 65
4.8 Impact ionization rate for relaxed Si in this work (solid line), compared to the results of the experiment (Cartier *et al.* [15]) and other theory calculations (Kamakura *et al.* [65], Thoma *et al.* [117] with modifications from [58]). . . . . . . . . . . . 66
4.9 Impact ionization coefficient for relaxed Si at room temperature from Full-band MC simulation and experiments (van Overstraeten *et al.* [92], Maes *et al.* [82], Slotboom *et al.* [107], Takayanagi *et al.* [112]). . . . . . . . . . . . . . . . . . . . . . 67
4.10 Quantum yield for relaxed Si at room temperature from simulations and experimental data by DiMaria *et al.* [30]. . . . . . 68
4.11 Isotropic impact ionization rate with the initial electron in the first conduction band for strained $Si_{1-x}Ge_x$ (x = 0.2) on a (001) silicon substrate. N is the number of grid points between 0 - $2\pi/a$ in the cuboid BZ. . . . . . . . . . . . . . . . . . . . . 68
4.12 Isotropic impact ionization rate with the initial electron in the first conduction band for relaxed Si and strained $Si_{1-x}Ge_x$ (x = 0.2) on a (001) silicon substrate. . . . . . . . . . . . . . . 69
4.13 Anisotropic impact ionization rate at each **k** point in the BZ for relaxed Si and strained $Si_{1-x}Ge_x$ (x = 0.2) on a (001) silicon substrate. . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
4.14 Impact ionization coefficient for strained $Si_{1-x}Ge_x$ (x = 0.2) at room temperature from Full-band MC simulations, compared to relaxed Si. Electric field is applied along two directions: $\langle 001 \rangle$ and $\langle 110 \rangle$. . . . . . . . . . . . . . . . . . . . . . . . . . 71

4.15 Quantum yield for strained $Si_{1-x}Ge_x$ (x = 0.2) at room temperature from Full-band MC simulations, compared to relaxed Si. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
4.16 Impact ionization rate for uniaxially strained Si with the stress along ⟨010⟩ and ⟨011⟩ directions. . . . . . . . . . . . . . . . . . 72
4.17 Comparison of the II rate for uniaxially/biaxially strained Si and relaxed Si. . . . . . . . . . . . . . . . . . . . . . . . . . . 73
4.18 Comparison of the II coefficient for uniaxially/biaxially strained Si and relaxed Si at room temperature from Full-band MC simulations. . . . . . . . . . . . . . . . . . . . . . . . . . . 73
4.19 Comparison of the quantum yield for uniaxially/biaxially strained Si and relaxed Si at room temperature from Full-band MC simulations. . . . . . . . . . . . . . . . . . . . . . . . . . . 74

5.1 Device structure for the n-channel Silicon-on-insulator (SOI) version of the lateral IMOS. $L_{GATE}$: i-region under the gate; $L_I$: i-region outside the gate; $t_{ox}$: oxide thickness; $t_{Si}$: Si bulk thickness [42]. . . . . . . . . . . . . . . . . . . . . . . . . 76
5.2 Device structure for: (a) Self-aligned n-channel IMOS [22] and (b) Self-align n-channel IMOS with Silicon-on-insulator (SOI) or Strain silicon-on-insulator (SSOI) [21]. $t_{ox}$ is the oxide thickness and $x_{j,se}$ is the depth of the source extension junction; TEOS: tetraethoxysilane. . . . . . . . . . . . . . . . 77
5.3 Device structure for: (a) LI-MOS with strained SiGe in the raised source/drain [118] and (b) SiGe IMOS on insulator [119]. 78
5.4 Device structure of a p-type DIMOS [88]. . . . . . . . . . . . . 79
5.5 Structure of the vertical IMOS [75]. . . . . . . . . . . . . . . . 81
5.6 Doping profile of the investigated vertical IMOS transistors: $N_{drain} = N_{source} \sim 4 \times 10^{18} cm^{-3}$; the p+ layer has a Gaussian profile with a peak value of $P_{max} = 1 \times 10^{19} cm^{-3}$ [32]. . . . . . 81
5.7 Band diagrams for the investigated vertical IMOS. . . . . . . . 82
5.8 Different operation modes of a vertical IMOS with a wide body ($d = 5\mu m$) [75]. . . . . . . . . . . . . . . . . . . . . . . . . . 82
5.9 Structure of the vertical strained-SiGe-IMOS (SSiGe-IMOS). Device dimension: d=50 nm, L=70 nm, $d_{ox}$ = 5nm; Gate: nPoly [32]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
5.10 The dashed line shows one of Ge profiles for the SSiGe-IMOS investigated in this work (the continuous line is the doping profile, which is the same as the doping profile for the RSi-IMOS in Fig. 5.6). . . . . . . . . . . . . . . . . . . . . . . . 85

5.11 Band diagrams for the RSi-IMOS and SSiGe-IMOS for the cases: (a) without external bias (b) $V_D = 2V, V_{G1} = V_{G2} = 1.2V$. Continuous lines are for the RSi-IMOS and dashed lines are for the SSiGe-IMOS. Doping and Ge profiles are shown in Fig. 5.10. 86
5.12 The simulators for the RSi-IMOS and SSiGe-IMOS. . . . . . . 86
5.13 Impact ionization rate ($1/s.cm^3$) in the RSi-IMOS within the MC simulation window at bias condition: $V_{G1} = V_{G2} = 1.2V$ and $V_D = 1.75V$. . . . . . . . . . . . . . . . . . . . . . . . . . 88
5.14 Different Ge profiles for the strained SiGe layer in the SSiGe-IMOS: (a) Group 1 and Group 2 - strained layer mostly inside the drain (b) Group 3 - strained layer in the channel. . . . . . 89
5.15 II generation current of the SSiGe-IMOS and RSi-IMOS with $V_{G1} = V_{G2} = 1.2, 1.4V$. Ge profile *GE1-2* in Fig. 5.14 is used for the SSiGe-IMOS. . . . . . . . . . . . . . . . . . . . . . . . . 90
5.16 II generation currents of the SSiGe-IMOS with different Ge profiles (in Fig. 5.14) at $V_{G1} = V_{G2} = 1.2V$. . . . . . . . . . . 91
5.17 Multiplication factor of the RSi-IMOS and SSiGe-IMOS with different Ge profiles at $V_{G1} = V_{G2} = 1.2V$: (a) With alloy scattering (b) Without alloy scattering. . . . . . . . . . . . . . 92
5.18 II generation currents from NSC-MC simulations are used to calibrate the HD model for the RSi-IMOS and SSiGe-IMOS (with Ge profile *GE3-1*) [33]. . . . . . . . . . . . . . . . . . . 93
5.19 Simulated $I_D - V_G$ curves of the RSi-IMOS and SSiGe-IMOS at $V_D = 2V$ [33]. . . . . . . . . . . . . . . . . . . . . . . . . 94
5.20 Simulated $I_D - V_D$ curves of the RSi-IMOS and SSiGe-IMOS at $V_{G1} = V_{G2} = 1.2V$ [33]. . . . . . . . . . . . . . . . . . . . . 94
5.21 Band diagrams of the RSi-IMOS and SSiGe-IMOS at $V_G = 1.2V$ and $V_D = 1.5V$. . . . . . . . . . . . . . . . . . . . . . . . . . 95
5.22 Comparison of $I_D - V_G$ curves for local and nonlocal model at different bias conditions. . . . . . . . . . . . . . . . . . . . . . 96
5.23 Spectral intensity of the drain current fluctuations for the RSi-IMOS. The case without II is also simulated for the comparison. The drain current for both cases are chosen to be the same (around 0.1A/cm). Device bias for two cases: With II: $V_D = 2V$ and $V_{G1} = V_{G2} = 1.35V$; Without II: $V_D = 2V$ and $V_{G1} = V_{G2} = 3.57V$. . . . . . . . . . . . . . . . . . . . . . . . 98
5.24 Noise small-signal equivalent circuit for the floating body [55]. 99
5.25 Spectral intensity of the drain current fluctuations of the RSi-IMOS for a gate bias of 1.4V and drain bias steps of 0.25V. . . 99
5.26 The minimum noise figure for the RSi-IMOS at a frequency of 1 GHz. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100

5.27 Comparison of the drain current noise between the SSiGe-IMOS and RSi-IMOS at the bias condition of $V_D = 2V$ and $V_G = 1.35V$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
5.28 Comparison of the drain current noise between the SSiGe-IMOS and RSi-IMOS over different gate voltages ($V_D = 2V$) at a frequency of 100 MHz. . . . . . . . . . . . . . . . . . . . . . . 101
5.29 Comparison of the drain current noise between the SSiGe-IMOS and RSi-IMOS over different drain voltages ($V_G = 1.35V$) at a frequency of 100 MHz. . . . . . . . . . . . . . . . . . . . . . . 102
5.30 Comparison of the minimum noise figure between the SSiGe-IMOS and RSi-IMOS versus gate voltage ($V_D = 2V$). . . . . . 102
5.31 Comparison of the minimum noise figure between the SSiGe-IMOS and RSi-IMOS versus drain voltage ($V_G = 1.35V$). . . . 103

# List of Tables


2.1 Approximations for matrix-valued quantities [116] . . . . . . . 19

3.1 Elastic stiffness constants $c_{ij}$ in GPa and elastic compliance constants $s_{ij}$ in $10^{-12}m^2/N$ [78]. . . . . . . . . . . . . . . . . . . 25

3.2 Number of irreducible wedges (NW) in the first BZ for different strain types. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

3.3 Deformation potentials of the valence and conduction bands for different valleys in Si and Ge (in eV), which were derived from self-consistent calculations based on a local density functional and *ab initio* pseudopotentials [29]. Spin-orbit splitting $\Delta_o$ is taken from [81] and band gap from [67]. . . . . . . . . . 47

4.1 Band gap for uniaxially and biaxially strained Si (Band gap for Si: 1.124 eV). . . . . . . . . . . . . . . . . . . . . . . . . . . 72

# List of Abbreviations

| | |
|---|---|
| BTE | Boltzmann transport equation |
| BZ | Brillouin zone |
| CME | Constant matrix element |
| CMOS | Complementary metal-oxide-semiconductor |
| CPU | Central processing unit |
| DD | Drift diffusion |
| EPM | Empirical pseudopotential method |
| FB-MC | Full-band Monte Carlo |
| Ge | Germanium |
| HBT | Heterojunction bipolar transistor |
| HD | Hydrodynamic |
| I-V | Current-Voltage |
| II | Impact Ionization |
| IMOS | Impact-Ionization MOSFET |
| ITRS | International Technology Roadmap for Semiconductors |
| MC | Monte Carlo |
| MOSFET | Metal-oxide-semiconductor field-effect transistor |
| NF | Noise figure |
| NS | Nonself-consistent |

| | |
|---|---|
| PE | Poisson equation |
| RSi-IMOS | Relaxed-Si Vertical IMOS |
| SC | Self-consistent |
| Si | Silicon |
| SiGe | Silicon-germanium alloy |
| SOI | Silicon-on-insulator |
| SS | Subthreshold slope |
| SSiGe-IMOS | Strained-SiGe Vertical IMOS |

# List of Symbols

| | |
|---|---|
| $\mathbf{a}_i$ | Basis vector of a direct lattice |
| $a_\perp$ | Perpendicular lattice constant |
| $a_{\vert\vert}$ | Parallel (in-plane) lattice constant |
| $A(\varepsilon)$ | Equienergy surface |
| $a_0$ | Lattice constant of a relaxed lattice |
| $\mathbf{b}_i$ | Basis vector of a reciprocal lattice |
| $\mathbf{B}$ | Magnetic field |
| $b, d, \Xi_d, \Xi_u$ | Deformation potentials |
| $c_{ij}$ | Elastic stiffness constant |
| $D_n$ | Electron diffusion coefficient |
| $e$ | Electron charge |
| $\mathbf{E}$ | Electric field |
| $E_g$ | Band gap |
| $\mathbf{F}_n$ | Electron-magnetic force |
| $f_n(\mathbf{r}, \mathbf{k}, t)$ | Distribution function |
| $\mathbf{G}$ | Reciprocal lattice vector |
| $h$ | Planck's constant |
| $\hbar$ | Planck's constant divided by $2\pi$ |
| $H$ | Hamiltonian operator |
| $I_1, I_2$ | Integrals of impact-ionization calculation |
| $I_{gen}$ | II generation current |
| $I_{OFF}$ | OFF current |
| $I_{ON}$ | ON current |
| $j$ | Current density |
| $\mathbf{k}$ | Wave vector |
| $k_B$ | Boltzmann constant |
| $k_x, k_y, k_z$ | Component x, y, z of a wave vector $\mathbf{k}$ |
| $L$ | Valley L |
| $L_{GATE}$ | Gate length of an IMOS |
| $L_I$ | Length of an i-region of an IMOS |
| $\hat{m}_n$ | Mass tensor |

| | |
|---|---|
| $m$ | Particle mass |
| $\mathbf{M}_D$ | Matrix element of a direct process |
| $\mathbf{M}_E$ | Matrix element of an exchange process |
| $n$ | Band index |
| $N$ | Number of grid points between 0 - $2\pi/a$ in the cuboid BZ |
| $X^{nod}$ | A quantity $X$ of a node (of a tetrahedron) |
| $n(\mathbf{r}, t)$ | Particle density |
| $p(x; y)$ | Conditional probability density |
| $P$ | Stress force |
| $P(\mathbf{k})$ | Scattering rate (with Pauli exclusion principle) |
| $q$ | Positive electron charge |
| $\mathbf{r}$ | Space vector |
| $R_{i,j,k}$ | Lattice direct vector |
| $\hat{\mathbf{S}}$ | Strain matrix |
| $s_{ij}$ | Stiffness compliance constant |
| $S(\mathbf{k})$ | Scattering rate (simplified version of $P(\mathbf{k})$) |
| $S^{II}$ | Impact ionization rate |
| $\mathbf{s}_n$ | Electron current density |
| $S_{n,n'}(\mathbf{k}, \mathbf{k}')$ | Transition rate |
| $t$ | Time |
| $X^{tet}$ | A quantity X of a tetrahedron |
| $T_{II}$ | Total matrix element of II process (also fitting factor for II rate) |
| $T_L$ | Lattice temperature |
| $T_n$ | Electron temperature |
| $T_p$ | Hole temperature |
| $T_{sim}$ | Simulation time |
| $u_n$ | Energy density |
| $\mathbf{v}$ | Group velocity of a particle's wave packet |
| $V(\mathbf{r})$ | Pseudopotential |
| $V_D$ | Drain voltage |
| $V_G$ | Gate voltage |
| $V_S$ | Source voltage |
| $w_n$ | Average energy |
| $x$ | Particle state |
| $Y$ | Fitting factor for phonon-electron scattering |
| | |
| $\alpha$ | Impact ionization coefficient |
| $\Delta E_c$ | Energy difference (between layers) at conduction bands |
| $\Delta E_v$ | Energy difference (between layers) at valence bands |
| $\Delta$ | Valley $\Delta$ |

| | |
|---|---|
| $\delta(\varepsilon - \varepsilon')$ | Delta function |
| $\epsilon$ | Dielectric constant |
| $\varepsilon_{ij}$ | Component $(ij)$ of a strain tensor |
| $\varepsilon_n(\mathbf{k})$ | Energy at a state of wave vector $\mathbf{k}$ and band index $n$ |
| $\hat{\varepsilon}$ | Strain tensor |
| $\mu$ | Mobility |
| $\Omega$ | Volume of the crystal |
| $\Psi(\mathbf{r}, t)$ | Electrostatic potential |
| $\psi_{\mathbf{k}}(\mathbf{r})$ | Pseudo wave function |
| $\Phi(\mathbf{k})$ | Weight function |
| $\varrho$ | Space charge density |
| $\hat{\sigma}$ | Stress tensor |
| $\sigma_{ij}$ | Component $(ij)$ of a stress tensor |
| $\tau$ | Relaxation time |
| $\langle\rangle$ | Statistical average |

# Acknowledgments

First of all, I would like to express my sincere thanks to Professor Christoph Jungemann, my *Doktorvater*. He has introduced me to the world of device modeling and simulation. Prof. Jungemann has been a great advisor with his profound knowledge and enthusiasm. He has always been ready for any guidance and discussion. I am also grateful to his kind support so that I could bring my family to Germany during my PhD time.

I would like to thank Professor Walter Hansch and Dr. Rainer Kraus for reading this thesis and serving on the oral examination committee. My special thank is given to Dr. Kraus as I have benefited from the cooperation work with him in modeling IMOS transistors. He provided me the configuration profile for the vertical IMOS investigated in this dissertation as well as many fruitful discussions during our work.

I would gratefully acknowledge the financial support from the Deutsche Forschungsgemeinschaft (German Research Foundation) for this work. They have covered the expenses for the conferences I have attended. Without their support, this work would not have been possible.

My three years at the Institute of Microelectronics and Circuit Theory would have been boring and much more difficult without the helps of all colleagues there. Whenever I had a problem with German language or administration, Dr. Oskar Kowarik, Dr. Roland Pfeiffer and Mr. Gregor Matz were are the first persons I asked for. During summers, I enjoyed playing tennis with Gregor, when we could chat about everything in the world. I got useful advices for my work from discussions with Dr. Sung-Min Hong. I would also thank Mrs. Jutta Maier-Lohneis for her kind support relating to the administration during my PhD time and especially for getting this thesis through administrative processes. Let me say thank to all other colleagues in the institute.

I would like to thank DAAD (German Academic Exchange Service) who offered me a two-year-scholarship for pursuing a master degree in the Technical University of Munich. This scholarship is invaluable for my education and my career.

I owe my friends their unconditional and great helps. Foremost, I would say my deep thanks to Dr. Chi Dong Nguyen, who inspired me to work in this field and always encouraged me to go ahead whenever I met any difficulty both in work and private life. Mr. Anh-Tuan Pham, with his excellent background in device simulations, gave me valuable helps. I would thank my friends, Dr. Trung Dung Vo and Ms. Brandis Keller who helped me to revise the English writing of this dissertation.

I am indebted to my parents, my aunt, my brother and sisters for their encouragement and loves to my life, my study and my career so far.

This work is devoted to my lovely wife, Mai Thi Thanh Nga, and my little son, Duc Anh. They are the joys of my life.

## Abstract

Impact ionization (II) has a strong influence on the characteristic of nanoscale transistors. In fact, when the transistors are down-scaled drastically, the electric field increases significantly, which leads to a strong effect of hot carriers in these devices. To understand these II activities, it is necessary to know the II rate of the applied material. Together with relaxed Si, uniaxially strained Si and biaxially strained SiGe are becoming popular applied materials to enhance the performance of MOSFETs and fast HBTs (heterojunction bipolar transistors), respectively. This dissertation focuses on developing a model for calculating the II rate for relaxed and strained materials. The model follows the constant-matrix-element approach and has exact energy and momentum conservation during the calculation of the six-dimensional integral in k-space. II coefficients and quantum yield for relaxed Si, strained Si and strained SiGe are calculated through Full-band Monte Carlo (FB-MC) simulations. A good agreement between simulation and experimental data for relaxed Si is obtained.

Besides the negative effects of II on nanoscale devices such as noise performance degradation or breakdown voltage reduction, it also has positive effects, e.g. in the case of Impact-Ionization MOSFET (IMOS). This device bases on a controlled avalanche breakdown to overcome the subthreshold limit of 60 mV/dec in conventional MOSFETs. However, the device requires a too high operating voltage and suffers from serious damage by hot carriers to the oxide layer. The vertical IMOS, which works on a dynamic reduction of the threshold voltage due to the floating body charged by the II process, was shown to have a better performance than the lateral IMOS. Hence, this dissertation uses numerical device simulators to investigate and optimize this device. The developed II model which is integrated into an existing MC simulator has been shown to be able to model and simulate the vertical IMOS transistors made of relaxed Si, strained Si or strained SiGe.

The Hydrodynamic model is combined with FB-MC in order to investigate I-V characteristics of the vertical IMOS. For vertical relaxed-Si-IMOS, it is proven that a good agreement between simulations and experimental measurements of the device's characteristics can be obtained by using this approach. The simulation results show that the performance of the vertical IMOS can be enhanced by a layer of strained SiGe placed between the drain and the channel. Finally the noise performance of these devices is also investigated.

**Zusammenfassung**

Stoßionisation hat einen starken Einfluss auf Transistoren der Nanogrößenordnung. Wenn Transistoren drastisch verkleinert werden, führt das stärkere elektrische Feld zu deutlich erhöhten Auswirkungen von heißen Ladungsträgern auf die Transistoreigenschaften. Um die Auswirkungen besser zu verstehen ist es notwendig die Stoßionisationsraten der verwendeten Materialien zu kennen. Zusätzlich zu unverspanntem Silizium werden häufig sowohl uniaxial verspanntes Silizium, als auch biaxial verspannte Silizium-Germanium Legierungen verwendet um die Leistung von MOSFETs und HBTs zu verbessern. Diese Dissertation beschreibt ein Model für die Berechnung der Stoßionisationsraten für unverspannte und verspannte Materialien. Dem Model liegt der Konstante-Matrix-Elemente-Ansatz zu Grunde und erfüllt Energie- und Momentumserhalt bei der Lösung des sechs-dimensionalen Integrals im k-Raum. Stoßionisationskoeffizienten und Quantenausbeute für unverspanntes Si, verspanntes Si und verspanntes SiGe werden mit der Full-Band Monte Carlo (FB-MC) berechnet. Gute Übereinstimmung zwischen Simulation und Experiment wird erreicht.

Neben den negativen Auswirkungen von Stoßionisation auf Transistoren der Nanogrößenordnung hat diese auch positive Effekte z.B. auf Impact-Ionization MOSFETs (IMOS). Dieser Transistor basiert auf einem kontrollierten Lawinendurchbruch um die subthreshold Grenze von 60 mV/Dekade in konventionellen MOSFETs zu überwinden. Allerdings braucht dieser Transistor eine zu hohe Betriebsspannung und leidet unter ernsthaften Schäden von heißen Ladungsträgern in der Oxidschicht. Der vertikale IMOS, der mit einer dynamischen Reduzierung der threshold Spannung auf Grund der Ladung des Floating Body durch Stoßionisation arbeitet, hat eine bessere Leistung als der horizontale IMOS. Deshalb wird in dieser Dissertation dieser Transistor mittels numerischer Simulation untersucht und optimiert. Die Tauglichkeit des Modells zur Modellierung und Simulation von vertikalen IMOS Transistoren wird mit der Integration des entwicklten Stoßionisationsmodel in einen bestehenden MC-Simulator gezeigt. Der IMOS besteht hierbei aus unverspanntem und verspanntem Si, sowie aus verspanntem SiGe.

Das hydrodynamische Modell wird mit FB-MC kombiniert um die Strom- und Spannungsbeziehungen des vertikalen IMOS zu untersuchen. Für den vertikalen Si-IMOS mit unverspanntem Silizium wird bewiesen, dass gute Übereinstimmung von Simulation und experimentellen Messungen der Transistoreigenschaften mit diesem Ansatz erreicht werden kann. Die Simulationsergebnisse zeigen, dass die Leistung eines vertikalen IMOS durch eine Schicht verspannten Silizium-Germaniums zwischen Drain und Channel besser wird. Abschließend wird die Rauschleistung dieser Transistoren untersucht.

# Chapter 1

# Introduction

## 1.1 Motivation

Since the declaration of CMOS nano-technology era in 1999, according to the International Technology Roadmap for Semiconductors (ITRS), CMOS transistors will be continuously down-scaled as fast as expected from Moore's law to the size of under 10 nm, which is considered as "beyond CMOS" era in 2020. Together with this aggressive down-scaling, ITRS 2009 also points out the grand challenges which semiconductor industry has to face. One of them is the management of power and leakage power consumption. This challenge is considered as the top priority to be tackled for both short (to 2017) and long term (beyond 2017) [51].

Indeed when the device becomes smaller, the supply voltage $V_D$ has to be reduced in order to decrease the dynamic power of the devices and to ensure a reliable operation. Moreover, to maintain the magnitude of saturation current, the threshold voltage $V_T$ needs to be lowered as well [68, 69, 86, 126]. It is also well known that conventional MOSFETs suffer from the limitation of the subthreshold slope $S$ due to the diffusion of carriers from the source to the channel of the device. This limit is 60 mV/dec at room temperature. Since the static leakage current $I_{LEAK}$ is related to the subthreshold slope $S$ by

$$I_{LEAK} = I_D|_{V_G=V_T} 10^{-V_T/S}, \tag{1.1}$$

$I_{LEAK}$ will rise by many orders of magnitude when the threshold voltage is reduced by just a factor of two, which leads to serious problems for short devices.

This limit can be overcome by devices which use either gate-controlled band-to-band tunneling, named Tunnel-FETs [6–8, 12, 71, 79, 113, 121], or impact ionization, named impact ionization MOSFETs (IMOS) [17, 18, 21–23,

34, 41–43, 84, 85, 88, 101, 118–120], as mechanisms for the injection or generation of charge carriers. However, the Tunnel-FET has the drawback of very low on-currents, and the lateral IMOS which uses the effect of avalanche breakdown in a gated p-i-n diode needs high supply voltages and suffers from rapid degradation due to hot carriers. Therefore, *vertical IMOS* was introduced as a promising candidate to tackle these disadvantages.

The vertical IMOS has the structure of a planar-doped barrier MOSFET with a floating body. It has been introduced and investigated extensively by experiments [2–5]. This IMOS is not based on avalanche breakdown like the lateral IMOS. Instead, the holes generated by impact ionization charge the floating p-body and cause a dynamic reduction of the threshold voltage, which leads to an extremely fast rising drain current in the subthreshold region. In [3], a very good subthreshold slope of 1.06 mV/dec at $V_D = 2.25$ V has been obtained from this IMOS. Moreover, this device offers the mechanisms for mitigating the damages by hot electrons almost completely [4] and shows the capability of working properly under high temperatures [2].

However, the shown device has also limitations including a low switching speed, remarkable hysteresis and still high supply voltage. Further improvement of the device should be done, not only by experiments but also by device simulations. Compared to experiments, simulations are more effective to explore the behavior of the device as well as offer insights into device mechanisms. For example, in [75], through 2D device simulations, it is found that double-gated vertical IMOS transistors with a very thin body region (e.g. 50 nm) can have very high $I_{on}/I_{off}$ ratios and high switching speeds. Unfortunately the necessary supply voltage is still too high. Reducing the required drain voltage in IMOS transistors is a great challenge and the topic of ongoing research (e.g. [88]).

Among numerical device simulators, Full-band Monte Carlo (FB-MC) is considered the most reliable method for a solution of the Boltzmann transport equation [54]. The results from MC simulator are often taken as a reference for other simplified models. The biggest disadvantage of this method is a requirement of a huge processing capacity and time. However, with the fast development of CPU capacity and its decreasing prices, large clusters comprising of many CPUs become feasible for small laboratories. Parallel processing is also taken advantage of dealing tons of jobs simultaneously. The Monte Carlo method, hence, is used as a main simulator in this work for simulating IMOS, together with other moment-based simulators such as the Hydrodynamic model.

For developing MC simulator applicable to investigate the activities of hot carriers in IMOS, knowledge of impact ionization (II) rates for the channel material should be known. II in relaxed Si has been calculated with *ab*

*initio* methods as well as experimentally measured by many authors, and a satisfactory agreement between theoretical calculations and experimental results has been reached [38]. However, there are few reports (both theory and experiment) of the II rates for strained Si and SiGe [32], which might be alternative materials for IMOS due to their smaller band gap (and hence expected higher ionization capability). Therefore an appropriate model which is capable of calculating II rates for materials under different stress or strain conditions should be developed.

Such an II model can also be applied to investigate hot carriers in nanoscale devices such as MOSFETs with strain channels or SiGe heterojunction bipolar transistors (HBT), which are the main innovators for semiconductor industry today [27, 40, 70]. Indeed, when the device shrinks drastically, the electric field increases significantly. This fact leads to a high number of hot carriers in the device. In MOSFETs, it is shown that the substrate current due to II is a key quantity for monitoring device degradation [50, 114]. In bipolar transistors, II is the cause for breakdown and the open-base breakdown voltage decreases with every new device generation. In addition, II degrades the noise performance [97].

## 1.2 Scope of work

From the above background and motivation, this dissertation will focus on the following tasks:

- Band structure calculation: Band structures for uniaxially and biaxially strained Si and SiGe are calculated by an empirical nonlocal pseudopotential method, which is available in-house [123]. However, an appropriate program should be developed to treat the large number of data and save individual properties of each strained material.

- Impact ionization calculation: A model for calculating II rates will be developed. Then the II rates for relaxed Si, uniaxially and biaxially strained Si and SiGe will be calculated.

- Developing Monte Carlo simulator: Such an II model will be integrated into the existing Full-band Monte Carlo simulator [59] to make this simulator capable of simulating impact ionization effects of strained Si and SiGe.

- Modeling and simulating IMOS: The developed FB-MC will be used to model and simulate IMOS. Moreover a combination with Hydro-

dynamic model from GALENE III [1] will be used to investigate this device.

- Optimizing IMOS: The developed simulators will be used to optimize IMOS by strain engineering.

- Noise investigation: One of the main problems of IMOS is its high level of noise. Therefore, investigating its noise performance is a necessary part of this dissertation.

## 1.3 Structure

This dissertation consists of six chapters. After Chapter 1 of introduction, Chapter 2 will summarize the basic equations which are considered the cornerstones for the Monte Carlo and other moment-based simulators. The overview of the Monte Carlo method, as well as its application for device simulation, is also presented here. Chapter 3 discusses the properties of strained Si and SiGe. The main focus is how the strain changes band structures and electrical properties of the material. It brings an overview of the properties of all strained types which are applied in this work. The empirical nonlocal pseudopotential method is also discussed. The model for calculating impact ionization rates is described in detail in Chapter 4. The theory background, calculation approach and the results of II for relaxed Si, strained Si and strained SiGe are also shown in this chapter. Chapter 5 focuses on impact ionization MOSFETs. Different types of IMOS, which have got the attention of the research community, are reviewed. Then a description of the investigated vertical IMOS is given. This chapter also presents the optimization of IMOS by strain engineering and the noise performance of these devices. Finally, Chapter 6 concludes the dissertation.

# Chapter 2

# Monte Carlo simulator for nanoscale devices

## 2.1 Basic equations for the electrical transport in semiconductor devices

In this section, basic equations for the electrical transport in semiconductor devices are summarized. First the Boltzmann transport equation (BTE), which is the base for the classical theory of transport, is presented. Then the Poisson equation is discussed.

### 2.1.1 The Boltzmann transport equation

The fundamental quantity in the classical transport theory is the carrier distribution function $f_n(\mathbf{r}, \mathbf{k}, t)$, which gives the occupation probability of a "state" characterized by a band index $n$, wave vector $\mathbf{k}$ and space vector $\mathbf{r}$ at time $t$. The Boltzmann transport equation which describes the transport phenomena in the semiclassical approach [53, 81]

$$\frac{\partial f_n}{\partial t} + \dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} f_n + \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f_n = \left( \frac{\partial f_n}{\partial t} \right)_{coll}, \tag{2.1}$$

where the right-hand side is the rate of change of $f_n$ due to collisions. This is the general BTE which is applied to different fields of transport theory.

A more concrete form of the BTE equation should be derived. On the left, we can replace the rate of change of momentum $\dot{\mathbf{k}}$ and distance $\dot{\mathbf{r}}$ over time by the following relations

$$\dot{\mathbf{k}} = \frac{q}{\hbar}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = \frac{1}{\hbar}\mathbf{F}_n \tag{2.2}$$

and

$$\dot{\mathbf{r}} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon(\mathbf{k}) = \mathbf{v}, \tag{2.3}$$

where $q$ is the positive electron charge, $\hbar$ is Planck's constant divided by $2\pi$, $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic field respectively, $\mathbf{F}_n$ is the electromagnetic force and $\mathbf{v}$ is the group velocity of the particle's wave packet. Eqs. (2.2) and (2.3) are called motion equations. Then eq. (2.1) can be rewritten

$$\frac{\partial f_n}{\partial t} + \mathbf{F}_n^T \frac{1}{\hbar}\nabla_{\mathbf{k}} f_n + \mathbf{v}^T \nabla_{\mathbf{r}} f_n = \left(\frac{\partial f_n}{\partial t}\right)_{coll}. \tag{2.4}$$

The collision term on the right-hand side of eq. (2.4) is caused by the imperfections of the ideal lattice, including phonons. The total scattering at a given phase unit is calculated from the in-scattering and out-scattering processes. The scattering events are characterized by a transition rate $S_{n,n'}(\mathbf{k},\mathbf{k}')$, which is the rate of particles scattered from state $\mathbf{k}$ (band index $n$) to state $\mathbf{k}'$ (band index $n'$). After taking the Pauli exclusion principle into account, the probability of transition from state $\mathbf{k}$ to state $\mathbf{k}'$ is given by

$$P(\mathbf{k} \to \mathbf{k}') = S_{n,n'}(\mathbf{k},\mathbf{k}') f_n(\mathbf{k})(1 - f_{n'}(\mathbf{k}')). \tag{2.5}$$

The scattering rate is achieved by summing $P(\mathbf{k} \to \mathbf{k}')$ in eq. (2.5) over all possible states $\mathbf{k}'$ at all bands $n'$ as follows

$$P(\mathbf{k}) = \sum_{\mathbf{k}',n'} S_{n,n'}(\mathbf{k},\mathbf{k}') f_n(\mathbf{k})(1 - f_{n'}(\mathbf{k}')). \tag{2.6}$$

The summation over k space can be replaced by the integration over the whole Brillouin zone (BZ)

$$P_{out}(\mathbf{k}) = \frac{\Omega}{(2\pi)^3} \sum_{n'} \int_{BZ} S_{n,n'}(\mathbf{k},\mathbf{k}') f_n(\mathbf{k})(1 - f_{n'}(\mathbf{k}')) d^3k', \tag{2.7}$$

where $\Omega$ is the volume of the crystal. Here $P_{out}$ refers to the probability of scattering out of the state $\mathbf{k}$ per unit time. Similarly, we have the scattering rate $P_{in}$, which refers to the probability of scattering into the state $\mathbf{k}$ from all states $\mathbf{k}'$

$$P_{in}(\mathbf{k}) = \frac{\Omega}{(2\pi)^3} \sum_{n'} \int_{BZ} S_{n',n}(\mathbf{k}',\mathbf{k}) f_{n'}(\mathbf{k}')(1 - f_n(\mathbf{k})) d^3k'. \tag{2.8}$$

Therefore, the net charge in the distribution function due to scattering is given by the difference $(P_{in}(\mathbf{k}) - P_{out}(\mathbf{k}))$

$$\begin{aligned}\left(\frac{\partial f_n}{\partial t}\right)_{coll} = \frac{\Omega}{(2\pi)^3} \sum_{n'} \int_{BZ} & [S_{n',n}(\mathbf{k}',\mathbf{k}) f_{n'}(\mathbf{k}')(1 - f_n(\mathbf{k})) \\ & - S_{n,n'}(\mathbf{k},\mathbf{k}') f_n(\mathbf{k})(1 - f_{n'}(\mathbf{k}'))] d^3k'. \end{aligned} \tag{2.9}$$

The collision term in eq. (2.9) makes the BTE in eq. (2.4) nonlinear. The technique used to solve the nonlinear BTE is based on a self-consistent iteration scheme of a linear version of the BTE, which is derived by dropping the Pauli exclusion principle and assuming the independence of the transition rate and force from the distribution function and time [59]. The following parts, hence, limit the discussion to the linear BTE of the following expression

$$\frac{\partial f_n}{\partial t} + \mathbf{F}_n^T \frac{1}{\hbar} \nabla_{\mathbf{k}} f_n + \mathbf{v}^T \nabla_{\mathbf{r}} f_n + S_n(\mathbf{k}) f_n = \frac{\Omega}{(2\pi)^3} \sum_{n'} \int_{BZ} S_{n',n}(\mathbf{k}', \mathbf{k}) f_{n'}(\mathbf{k}') d^3k', \tag{2.10}$$

where the scattering rate $S_n(\mathbf{k})$ is obtained by simplifying the scattering rate $P_{out}$ in eq. (2.7)

$$S_n(\mathbf{k}) = \frac{\Omega}{(2\pi)^3} \sum_{n'} \int_{BZ} S_{n,n'}(\mathbf{k}, \mathbf{k}') d^3k'. \tag{2.11}$$

### 2.1.2 The Poisson equation

The Poisson equation (PE), together with the BTE, is the cornerstone of semi-classical device simulation. While the BTE describes the evolution of the distribution function within a simulation domain, the electrostatic potential is covered by the Poisson equation. In self-consistent Monte Carlo simulations, which will be presented in Sec. 2.2.4, these two equations are solved in a self-consistent manner.

In the eddy current free approximation, the quasi-static potential $\Psi(\mathbf{r}, t)$ is given by the Poisson equation (PE) [52]

$$\nabla_{\mathbf{r}}^T (\epsilon \nabla_{\mathbf{r}} \Psi) = -\varrho, \tag{2.12}$$

where $\varrho$ is the space charge density and $\epsilon$ is the dielectric constant which depends on the material type.

A device is composed of different types of materials such as Si and oxide. Therefore for solving the PE equation, the whole device is split into different regions, for example the region of semiconductor material (Si, Ge, SiGe), of oxide, of vacuum or void [59]. At the interfaces between regions, the potential must be continuous

$$\Psi^l(\mathbf{r}, t) = \Psi^r(\mathbf{r}, t), \tag{2.13}$$

where $\Psi^l$ and $\Psi^r$ are the potential on the left and right side of the interface respectively. Moreover, the displacement density must satisfy the following condition

$$\epsilon^l(\mathbf{r}) E_\perp^l(\mathbf{r}, t) + \epsilon^r(\mathbf{r}) E_\perp^r(\mathbf{r}, t) = \bar{q}, \tag{2.14}$$

where $E_\perp$ is the electric field component perpendicular to the interface with orientation into the region and $\bar{q}$ is the interface charge density.

To solve the PE, two types of boundary conditions are applied to assure an unique and physically reasonable solution inside the bounded region [52]. One is Neumann type, which is used on the surface of the simulation region, as

$$\epsilon(\mathbf{r})E_\perp(\mathbf{r},t) = 0 \tag{2.15}$$

and the other is Dirichlet type, which is used on the contacts of regions

$$\Psi(\mathbf{r},t) = V^i, \tag{2.16}$$

where $V^i$ is the inner potential of the contact.

## 2.2 Monte Carlo simulator

### 2.2.1 Introduction of the Monte Carlo method

The Monte Carlo (MC) method refers to the technique using random numbers for solving a problem. The possibility of applying the MC method to deterministic problems was noticed by Fermi, von Neumann and Ulam and then was popularized in different fields [46]. It is shown that each integral can be represented as an expected value and the problem of estimating an integral by the MC method is equivalent to the problem of estimating an unknown parameter [96].

For example, the MC method is used to evaluate the following integral

$$I = \int_a^b g(x)dx \tag{2.17}$$

The integral $I$ can be rewritten as

$$I = \int_a^b \frac{g(x)}{f_X(x)} f_X(x)dx, \tag{2.18}$$

assuming that $f_X(x)$ is any distribution function, which means $f_X(x) > 0$ when $g(x) \neq 0$ and $\int_a^b f_X(x)dx = 1$. Then the integral in eq. (2.18) is expressed by its expected value

$$I = E\left[\frac{g(X)}{f_X(X)}\right], \tag{2.19}$$

where $X$ is a random variable distributed according to $f_X(x)$.

For the sake of illustration, assuming that

$$f_X(x) = \begin{cases} 1/(b-a), & \text{if } a < x < b, \\ 0, & \text{otherwise;} \end{cases} \tag{2.20}$$

then

$$E[g(X)] = \frac{I}{b-a}, \tag{2.21}$$

and

$$I = (b-a)E[g(X)]. \tag{2.22}$$

An unbiased estimator of $I$ is its sample mean

$$\theta = (b-a)\frac{1}{N}\sum_{i=1}^{N} g(X_i). \tag{2.23}$$

The variance of $\theta$ is calculated by $E(\theta^2) - E^2(\theta)$, and the result is

$$\text{var}\theta = \frac{1}{N}\left[(b-a)\int_a^b g^2(x)dx - I^2\right]. \tag{2.24}$$

The *MC estimate* for the integral $I$ is done by the following steps:

- Generate a sequence of $N$ random numbers.
- Compute variables $X_i$, $i = 1, ..., N$.
- Compute $g(X_i)$, $i = 1, ..., N$.
- Compute the sample mean $\theta$ according to eq. (2.23).

We can see that since the factorization of the integrand is not unique, different random variables can be introduced depending on the choice of the function $f_X$. All of them have the same expected value, but different variance.

The application of the MC method in device simulations stems from the fact that the BTE can be transformed into an integral equation by a formal integration. The pioneering work in this field was done in 1966 by Kurosawa, who was the first to apply the MC method to simulate carrier transport in semiconductors [76]. Then different MC methods were successfully applied to transport calculations in a variety of semiconductors, for example in [35, 54, 87]. In the next section, the Monte Carlo solver for the Boltzmann equation is presented.

### 2.2.2 The Monte Carlo solver for the Boltzmann equation

If we call variable $x$ to stand for the particle state $[n, \mathbf{r}, \mathbf{k}]$, the integral form of the linearized BTE at (2.10) is described as [35, 53, 74]

$$p(x,t;x_0,t_0) = p_0(x,t;x_0,t_0) + \int_{t_0}^{t}\int\int p_0(x,t;x_1',t_1)S(x_1';x_1) p(x_1,t_1;x_0,t_0)dx_1'dx_1dt_1, \quad (2.25)$$

where $p_0(x,t;x_0,t_0)$ is the conditional probability density that a particle found at the time $t_0$ in the state $x_0$ appears at $t$ in the state $x$ without being scattered

$$p_0(x,t;x_0,t_0) = \delta(x - x_{New}(t;x_0,t_0))\exp\left(-\int_{t_0}^{t} S(x_{New}(\tau;x_0,t_0))d\tau\right). \quad (2.26)$$

$x_{New}$ is the solution of the equations of motion (2.2) and (2.3) at the time $t$ for a particle that was at time $t_0$ in the state $x_0$ [59].

The MC algorithm based on eq. (2.25) is called backward MC method. The usual MC method is based on the conjugate of eq. (2.25) as follows [87]

$$p(x,t;x_0,t_0) = p_0(x,t;x_0,t_0) + \int_{t_0}^{t}\int\int p(x,t;x_1',t_1)S(x_1';x_1) p_0(x_1,t_1;x_0,t_0)dx_1'dx_1dt_1. \quad (2.27)$$

If we insert eq. (2.27) into itself and repeat this step infinite times, we will get a formal solution known as the Neumann series

$$\begin{aligned} p(x,t;x_0,t_0) = {} & p_0(x,t;x_0,t_0) \\ & + \int_{t_0}^{t}\int\int p_0(x,t;x_1',t_1)S(x_1';x_1)p_0(x_1,t_1;x_0,t_0)dx_1'dx_1dt_1 \\ & + \int_{t_0}^{t}\int\int\int_{t_1}^{t}\int\int p_0(x,t;x_2',t_2)S(x_2';x_2)p_0(x_2,t_2;x_1',t_1) \\ & \qquad S(x_1';x_1)p_0(x_1,t_1;x_0,t_0)dx_2'dx_2dt_2dx_1'dx_1dt_1 \\ & + \cdots \\ & + \vdots \end{aligned} \quad (2.28)$$

The terms on the right-hand side of eq. (2.28) can be understood as follows: the first one is the probability density that a particle is not scattered during

its propagation from the initial state $x_0$ into the final state $x$, the second one is the probability density that the particle is scattered exactly once between its initial and final state, the third one includes two scattering events and so forth [59]. Therefore, the conditional probability density $p(x, t; x_0, t_0)$ can be written as a sum of all conditional probability densities by

$$p(x, t; x_0, t_0) = \sum_{i=0}^{\infty} p_i(x, t; x_0, t_0) \tag{2.29}$$

with

$$p_i(x, t; x_0, t_0) = \int_{t_0}^{t} \int \int p_{i-1}(x, t; x_1', t_1) S(x_1'; x_1) p_0(x_1, t_1; x_0, t_0) dx_1' dx_1 dt_1. \tag{2.30}$$

The probability of $i$ scattering events in the interval $[t_0, t]$ is given by the integral of $p_i$ over all final states

$$P_i(t; x_0, t_0) = \int p_i(x, t; x_0, t_0) dx. \tag{2.31}$$

A self-scattering mechanism is introduced in [35, 53] to avoid the complexity of the relation of scattering rate on the particle state $x$ in eq. (2.26) by the following form

$$S_{self}(x'; x) = (\Gamma - S(x))\delta(x' - x), \tag{2.32}$$

and hence the total scattering rate is calculated by

$$S_{tot}(x) = \int S_{self}(x', x) dx' + S(x) = \Gamma. \tag{2.33}$$

In eqs. (2.32) and (2.33), $\Gamma$ is a constant which must be larger than the scattering rate $S(x)$ for all possible states $x$. Now, the probability for $i$ scattering events in the interval $[t_0, t]$ including self-scattering can be determined independently from the initial state $x$. They have a Poisson distribution

$$P_i(t; t_0) = \frac{(\Gamma(t - t_0))^i}{i!} \exp(-\Gamma(t - t_0)). \tag{2.34}$$

It is shown that the MC method can reproduce exactly the motion of particles, which consists of a series of instantaneous scattering events and collisionless accelerations by external forces [35, 54, 76, 87], from eq. (2.29). The simulation procedure is illustrated in Fig. 2.1. Basically, there are four steps:

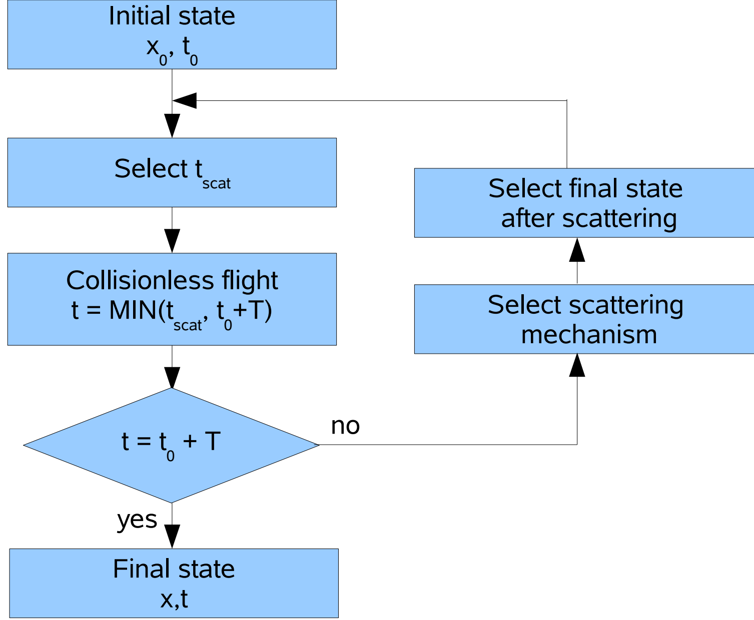


Figure 2.1: Flowchart of the MC method for a simulation within a duration $T$ [59].

- Step 1: Determine the collisionless flight by the following equation [59]

$$t_{scat} = t_0 - \frac{1}{\Gamma}\ln(1-r), \tag{2.35}$$

  where $r$ is a uniformly distributed random number within $[0,1)$.

- Step 2: Determine the particle state $x$ from eqs. (2.2) and (2.3) by

$$x = x_{New}(t_{scat}; x_0, t_0). \tag{2.36}$$

- Step 3: One of the physical scattering processes (see Sec. 2.2.3) or self-scattering is selected randomly by the direct method [91] which can be described briefly as follows. An integer random number $i$ is calculated with a discrete version of the direct method

$$i : \sum_{j=1}^{i-1} P_j < r \le \sum_{j=1}^{i} P_j, \tag{2.37}$$

  where $P_j$ is the probability of the $j^{th}$ event and $r$ is a uniformly distributed random number.

- Step 4: The final state after scattering is generated with the probability density

$$p_i^a(x') = \frac{S_i(x';x)}{S_i(x)}, \tag{2.38}$$

where the final wave vector is selected by the rejection technique [56,59], band index and wedge are determined randomly by the direct method (eq. (2.37)). The rejection method used to generate a random number $x$ with a probability density $p(x)$ ($p(x)$ is bounded) can be explained briefly as follows. First, a random number $x_1$ is generated with a probability $\tilde{p}(x)$ with

$$M\tilde{p}(x) \geq p(x), \tag{2.39}$$

where $M$ is a positive and finite constant. $\tilde{p}(x)$ is modeled in such a way that it reproduces $p(x)$ as good as possible. Moreover, a closed form of the inverse distribution function $F^{-1}(r)$ corresponding to $\tilde{p}(x)$ must exist to generate $x_1$ with a uniformly distributed random number $r_1$ ($x_1 = F^{-1}(r_1)$). Second, $x_1$ is accepted if the following condition holds

$$p(x_1) > r_2 M\tilde{p}(x_1), \tag{2.40}$$

where $r_2$ is another uniformly distributed random number. If $x_1$ is rejected, the whole procedure is repeated with two new random numbers $r_1$ and $r_2$. The expected value of the number of rejection steps is $M$.

### 2.2.3 Scattering mechanisms

Carrier transport in semiconductors is affected by scattering mechanisms. In MC simulations, the scattering effects also help the simulation to converge. The most important sources of scattering in the bulk of homogeneous crystals are phonons, impurities and other carriers [53].

Generally, the scattering rates $S_{n',n}(\mathbf{k}', \mathbf{k})$ for the scattering processes can be calculated by the Fermi golden rule [81]. However, the *ab initio* calculations are very difficult due to the complexity of the matrix element of the interaction. A number of simplified models have been suggested. Therefore, in this context, the simplified models or empirical models, which have good approximations to experimental results, will be applied.

- Phonon scattering: The transition rates of [13, 54] together with the approach from [115] are used. Moreover, the electron-phonon transition rate outside of the first conduction band is multiplied with a fitting factor so that a good fitting for both phonon scattering and impact

ionization with experimental data is achieved [37]. The details of this fitting procedure is presented in Chapter 4.

- Impurity scattering: The Brooks-Herring (BH) model is applied for the ionized impurity scattering due to its simplicity [53]. However, since this model cannot correctly describe the mobility at high doping concentration [19, 36, 73], an empirical correction of the model is used to modify the Brooks-Herring scattering rate $S^{BH}(\mathbf{k}', \mathbf{k})$ as [59]

$$S^{CBH}(\mathbf{k}', \mathbf{k}) = \zeta(N_a, N_d) S^{BH}(\mathbf{k}', \mathbf{k}), \tag{2.41}$$

  where $S^{CBH}$ is the corrected scattering rate, $\zeta(N_a, N_d)$ ($N_a$: density of acceptors, $N_d$: density of donors) is chosen in such a way that the simulated mobility $\mu$ exactly reproduces the following Caughey-Thomas (CT) expression for the minority and majority mobility of both carrier types in relaxed silicon [16]:

$$\mu = \mu_{min}^{CT} + \frac{\mu_{max}^{CT} - \mu^{CT} min}{1 + \left(\frac{N_a + N_d}{N_{ref}^{CT}}\right)^{\alpha^{CT}}}. \tag{2.42}$$

- Alloy scattering: In disordered alloys such as SiGe, alloy scattering refers to the fluctuations of composition, which is a source of scattering for electron eigenstates of the homogeneous system [53]. The scattering rate is calculated according to the model developed by Harrison and Hauser [47].

- Surface roughness scattering: When a carrier moves at the surface of a definite region, or at the interface between two different regions, the corrugations of the potential profile at the interface can cause scattering, which is called surface roughness scattering [53]. In this context, a simple approach consisting of reflective and diffusive scattering is used [89,98]. In the case of diffusive scattering, the particle is elastically scattered back into the silicon with a uniform probability distribution on an equienergy surface. The probability for a diffusive scattering event is 14% in the case of electrons and 13% in the case of holes.

- Impact ionization: This is an electron-electron or hole-hole interaction process via Coulomb forces, which creates new electrons or holes. The details of this process, together with all related calculations, are presented in Chapter 4.

### 2.2.4 Device simulation

The Monte Carlo method applied to the device simulation has two types: Nonself-consistent (NS) and self-consistent (SC). Nonself-consistent is a special case which the MC simulation is performed for a given electric field. This topic is addressed in Chapter 5. In this part, self-consistent device simulation is presented.

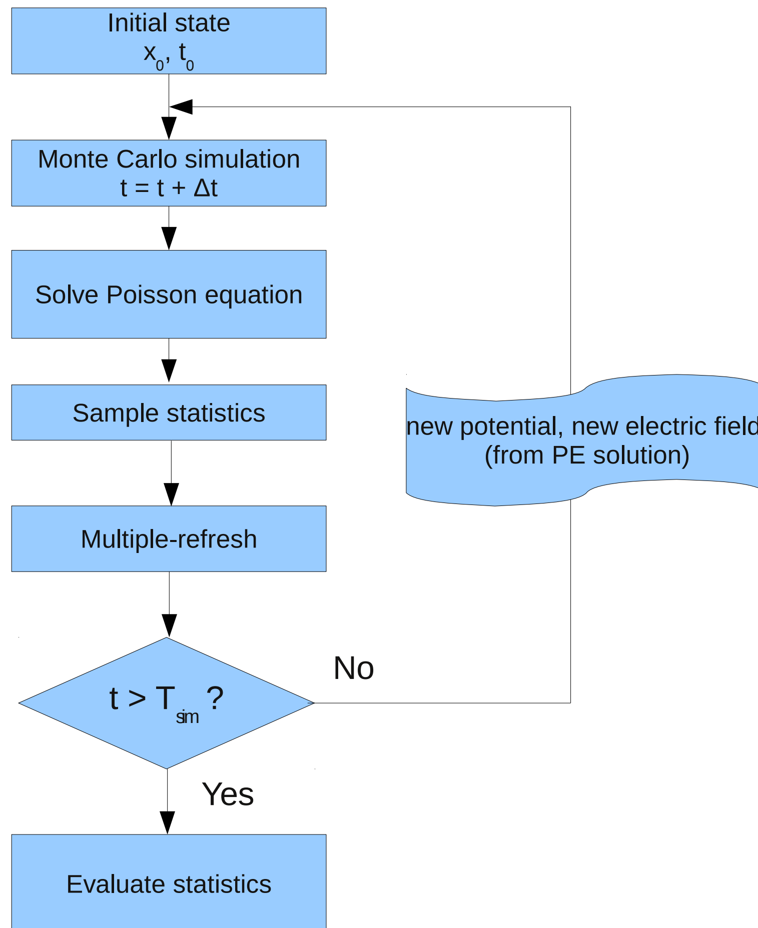


Figure 2.2: Flowchart of the self-consistent MC device simulation [59].

For SC simulations, the MC solver and the Poisson equation (PE) have to be solved self-consistently [49]. The flow chart of a SC-MC device simulation is shown in Fig. 2.2. The steps are described as follows:

- Step 1: The MC simulation is performed for a given time step length $\Delta t$.

- Step 2: After the particle simulation with a duration $\Delta t$, the PE is solved to evaluate the space charge and the electrostatic potential.
- Step 3: Calculate sample statistics, then applying the multiple-refresh technique [45, 57, 117] to speed up the simulation process and enhance the variance of the estimates.
- Step 4: If $t < T_{sim}$, continue the simulation. New electric field and potential are given to the MC simulation. It is noted that for MC simulations based on the electric field at the beginning of the time step, the method is a forward Euler scheme which is stable only for short time steps.

## 2.3 Moment-based simulators

Instead of solving the BTE directly, it is common to consider only some moments of the distribution function. A moment is obtained by multiplying the distribution function with a suitable weight function $\Phi(\mathbf{r}, \mathbf{k})$ and integrating it over k space [44, 104]

$$\langle \Phi \rangle = \frac{2}{(2\pi)^3} \frac{1}{n(\mathbf{r}, t)} \sum_n \int_{BZ} \Phi(\mathbf{r}, \mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) d^3k, \tag{2.43}$$

where $n(\mathbf{r}, t) = \frac{2}{(2\pi)^3} \sum_n \int_{BZ} f(\mathbf{r}, \mathbf{k}, t) d^3k$ is the particle density. If the BTE (2.10) is applied to $n\langle \Phi \rangle$, we can get the following balance equation [59]

$$\frac{\partial (n\langle \Phi \rangle)}{\partial t} - n\langle \mathbf{F}^T \frac{1}{\hbar} \nabla_{\mathbf{k}} \Phi \rangle + \nabla_{\mathbf{r}}^T \mathbf{j}_{\langle \Phi \rangle} - n\langle \mathbf{v}^T \nabla_{\mathbf{r}} \Phi \rangle = -n\langle S\{\Phi\} \rangle, \tag{2.44}$$

where $\mathbf{j}_{\langle \Phi \rangle}$ is a generalized flux

$$\mathbf{j}_{\langle \Phi \rangle} = n\langle \mathbf{v} \Phi \rangle \tag{2.45}$$

and the scattering term $S\{\Phi\}$ is given by

$$S\{\Phi\} = \frac{\Omega}{(2\pi)^3} \sum_{n'} \int_{BZ} S_{n',n}(\mathbf{k}', \mathbf{k})(\Phi_n - \Phi_{n'}) d^3k'. \tag{2.46}$$

In most cases, the scattering term can be presented through a relaxation time approximation

$$\langle S\{\Phi\} \rangle \approx \frac{\langle \Phi \rangle - \langle \Phi \rangle^{eq}}{\tau_\Phi}, \tag{2.47}$$

where $\langle \rangle^{eq}$ is the expectation in the case of thermal equilibrium.

The quantity $n\langle\Phi^m\rangle$ is called the m$^{\text{th}}$ moment of the distribution function and the corresponding equations of motion m$^{\text{th}}$ moment of the BTE. Based on the lower order moments, the transport models for electrons (hence the following quantities are subscripted with $n$) are derived as follows:

- m = 0: We have
$$\langle\Phi\rangle = 1, \mathbf{j}_n = n\langle\mathbf{v}\rangle, F = 0. \tag{2.48}$$
A conservation law for the *electron density* $n$ is obtained by putting the quantities of eq. (2.48) into eq. (2.44)
$$\frac{\partial n}{\partial t} + \nabla_{\mathbf{r}}\mathbf{j}_n = G^{II}, \tag{2.49}$$
where $G^{II}$ is the generation rate due to II. Eq. (2.49) is also called the continuity equation.

- m = 1: We have
$$\langle\Phi\rangle = \langle\mathbf{v}\rangle \tag{2.50}$$
and eq. (2.44) for this $\langle\Phi\rangle$ can be written as
$$\frac{\partial\mathbf{j}_n}{\partial t} - n\langle\hat{m}^{-1}\rangle\nabla_{\mathbf{r}}(E_c - q\Psi) + n\langle\hat{m}^{-1}\nabla_{\mathbf{r}}\varepsilon\rangle + \left(\nabla_{\mathbf{r}}^T[n\langle\mathbf{v}\mathbf{v}^T\rangle]\right)^T - n\langle\mathbf{v}^T\nabla_{\mathbf{r}}\mathbf{v}^T\rangle = -\frac{\mathbf{j}_n}{\tau_{\mathbf{j}_n}}, \tag{2.51}$$
where $E_c$ is the conduction band edge, which is the minimum energy of all conduction bands. To simplify this equation, some approximations have to be used. First, the group velocity is separated into position and wave vector dependent factors $\mathbf{v}(\mathbf{r},\mathbf{k}) = g(\mathbf{r})\mathbf{v}(\mathbf{k})$ [61]; hence
$$\nabla_{\mathbf{r}}\mathbf{v}^T = \frac{\nabla_{\mathbf{r}}g}{g}\mathbf{v}^T \tag{2.52}$$
and
$$\langle\mathbf{v}^T\nabla_{\mathbf{r}}\mathbf{v}^T\rangle^T = \langle\mathbf{v}\mathbf{v}^T\rangle\frac{\nabla_{\mathbf{r}}g}{g}. \tag{2.53}$$
Second, the dynamic electron temperature $T_n^\star$ is defined by
$$\frac{3}{2}k_B T_n^\star = \frac{1}{2}m_n^\star\langle\mathbf{v}^2\rangle, \tag{2.54}$$

where the mass $m_n^\star$ is calculated under equilibrium conditions with the lattice temperature $T_0$ by

$$\frac{1}{m_n^\star} = \left\langle \frac{1}{m} \right\rangle^{eq} = \frac{\langle \mathbf{v}^2 \rangle^{eq}}{3k_B T_0}. \tag{2.55}$$

The expected value of the inverse mass is eliminated by the introduction of a new relaxation time $\tau_{\mathbf{j}n}^\star$

$$\left\langle \frac{1}{m} \right\rangle \tau_{\mathbf{j}_n} = \frac{\tau_{\mathbf{j}_n}^\star}{m_n^\star}. \tag{2.56}$$

Due to the definition of the effective density of state (over all conduction bands) $N_c$ [110]

$$N_c = \sum_{n \in c.b.} \int_0^\infty D_n(\varepsilon) \exp\left( -\frac{\varepsilon}{k_B T_0} \right) d\varepsilon, \tag{2.57}$$

($D_n$: density of state) the following relation holds for the gradient of the energy under equilibrium conditions

$$\langle \nabla_{\mathbf{r}} \varepsilon \rangle^{eq} = -k_B T_0 \frac{\nabla_{\mathbf{r}} N_c}{N_c}. \tag{2.58}$$

From eq.(2.58), together with the scalar approximation for $\langle \hat{m}^{-1} \nabla_{\mathbf{r}} \varepsilon \rangle$ in Tab. 2.1, the position-dependent force is calculated by

$$\left\langle \frac{\nabla_{\mathbf{r}} \varepsilon}{m} \right\rangle \approx -\frac{k_B T_n^\star}{m_n^\star} \frac{\nabla_{\mathbf{r}} N_c}{N_c}. \tag{2.59}$$

Other scalar approximations of matrix-valued quantities (also quantities for the case $m = 2$ below) can be found in Tab. 2.1. From the above approximations, the final balance equation for the *current density* $\mathbf{j}_n$ is obtained

$$\mathbf{j}_n + \tau_{\mathbf{j}_n} \frac{\partial \mathbf{j}_n}{\partial t} = \frac{\tau_{\mathbf{j}_n}^\star}{m_n^\star} \nabla_{\mathbf{r}} (E_c - q\Psi) - \frac{\tau_{\mathbf{j}n}}{m_n^\star} \left( \nabla_{\mathbf{r}} (n k_B T_n^\star) - n k_B T_n^\star \frac{\nabla_{\mathbf{r}} N_c}{N_c} \right). \tag{2.60}$$

- m = 2: We have

$$\langle \Phi \rangle = \langle \varepsilon \rangle. \tag{2.61}$$

The average energy $w_n$ and energy current density $\mathbf{s}_n$ are defined by

$$w_n = n \langle \varepsilon \rangle \tag{2.62}$$

Table 2.1: Approximations for matrix-valued quantities [116]

| Matrix-Like | Scalar-Like |
|---|---|
| $\langle \hat{m}^{-1} \rangle = \frac{1}{\hbar^2} \nabla_{\mathbf{k}} \nabla_{\mathbf{k}}^T \varepsilon(\mathbf{k})$ | $\frac{1}{3} Tr \langle \hat{m}^{-1} \rangle$ |
| $\langle \mathbf{v}\mathbf{v}^T \rangle$ | $\frac{1}{3} \mathbf{v}^2$ |
| $\langle \hat{m}^{-1} \nabla_{\mathbf{r}} \varepsilon \rangle$ | $\langle \frac{\nabla_{\mathbf{r}} \varepsilon}{m} \rangle$ |
| $\langle \hat{m}^{-1} \varepsilon + \mathbf{v}\mathbf{v}^T \rangle$ | $\frac{1}{3} Tr \langle \hat{m}^{-1} \varepsilon \rangle + \frac{1}{3} \mathbf{v}^2$ |
| $\langle \hat{m}^{-1} \varepsilon \nabla_{\mathbf{r}} \varepsilon \rangle$ | $\langle \frac{\varepsilon \nabla_{\mathbf{r}} \varepsilon}{m} \rangle$ |
| $\langle \varepsilon \mathbf{v}\mathbf{v}^T \rangle$ | $\frac{5}{6} \frac{\langle \mathbf{v}^2 \rangle^2}{Tr \langle \hat{m}^{-1} \rangle}$ |

and

$$\mathbf{s}_n = n \langle \varepsilon \mathbf{v} \rangle. \tag{2.63}$$

The balance equation for the energy is written as

$$\frac{\partial \omega_n}{\partial t} - \mathbf{j}_n^T \nabla_{\mathbf{r}} (E_c - q\Psi) + n \langle \mathbf{v}^T \nabla_{\mathbf{r}} \varepsilon \rangle + \nabla_{\mathbf{r}}^T \mathbf{s}_n - n \langle \mathbf{v}^T \nabla_{\mathbf{r}} \varepsilon \rangle = -n \frac{\langle \varepsilon \rangle - \langle \varepsilon \rangle^{eq}}{\tau_{\omega_n}}. \tag{2.64}$$

With the assumption of a negligible energy generation by II and a new energy relaxation time $\tau_{w_n}^\star$

$$\frac{\langle \varepsilon \rangle - \langle \varepsilon \rangle^{eq}}{\tau_{w_n}} = \frac{3k_B}{2} \frac{T_n^\star - T_0}{\tau_{w_n}^\star}, \tag{2.65}$$

the final balance equation for the *energy density* $\omega_n$ becomes

$$\frac{\partial w_n}{\partial t} + \nabla_{\mathbf{r}}^T \mathbf{s}_n = \mathbf{j}_n^T \nabla_{\mathbf{r}} (E_c - q\Psi) - p \frac{3k_B}{2} \frac{T_n^\star - T_0}{\tau_{w_n}^\star}. \tag{2.66}$$

The balance equation for the energy current density is

$$\frac{\partial \mathbf{s}_n}{\partial t} - n \langle \hat{m}^{-1} \varepsilon + \mathbf{v}\mathbf{v}^T \rangle \nabla_{\mathbf{r}} (E_c - q\Psi) + n \langle \hat{m}^{-1} \varepsilon \nabla_{\mathbf{r}} \varepsilon \rangle + (\nabla_{\mathbf{r}}^T [n \langle \varepsilon \mathbf{v}\mathbf{v}^T \rangle])^T - n \langle \mathbf{v}^T \nabla_{\mathbf{r}} \varepsilon \mathbf{v}^T \rangle^T = -\frac{\mathbf{s}_n}{\tau_{\mathbf{s}_n}}. \tag{2.67}$$

Matrix-valued quantities in eq. (2.67) are also approximated by scalars (see Tab. 2.1). Then further following approximations are applied. First, the term involving the position-dependent velocity is simplified similarly to eqs. (2.52) and (2.53)

$$\langle \mathbf{v}^T \nabla_{\mathbf{r}} \varepsilon \mathbf{v}^T \rangle^T = \langle \varepsilon \mathbf{v}\mathbf{v}^T \rangle 2 \frac{\nabla_{\mathbf{r}} g}{g}. \tag{2.68}$$

Second, the position-dependent force is simplified similarly to eq. (2.59)

$$\left\langle \frac{\varepsilon \nabla_{\mathbf{r}} \varepsilon}{m} \right\rangle \approx -\frac{15}{2Tr\langle \hat{m}^{-1} \rangle} \left( \frac{k_B T_n^\star}{m_n^\star} \right)^2 \frac{\nabla_{\mathbf{r}} N_c}{N_c}. \tag{2.69}$$

In addition, a new relaxation time $\tau_{\mathbf{s}_n}$ is defined

$$\left\langle \frac{\varepsilon}{m} + \frac{\mathbf{v}^2}{3} \right\rangle \tau_{\mathbf{s}_n} = \frac{5}{6} \langle \mathbf{v}^2 \rangle \tau_{\mathbf{s}_n}^\star. \tag{2.70}$$

Moreover, if the gradient of the expected value of the inverse mass is neglected

$$\nabla_{\mathbf{r}} \frac{3}{Tr\langle \hat{m}^{-1} \rangle} \approx \mathbf{0}, \tag{2.71}$$

the final equation for the *energy current density* $\mathbf{s}_n$ is given by

$$\begin{aligned} \mathbf{s}_n + \tau_{\mathbf{s}_n} \frac{\partial \mathbf{s}_n}{\partial t} = n \frac{5 k_B T_n^\star \tau_{\mathbf{s}_n}^\star}{2 m_n^\star} \nabla_{\mathbf{r}} (E_c - q\Psi) \\ - \frac{5 \tau_{\mathbf{s}_n}^\star \tau_{\mathbf{j}_n}}{2 m_n^\star \tau_{\mathbf{j}_n}^\star} \left( \nabla_r [n (k_B T_n^\star)^2] - n (k_B T_n^\star)^2 \frac{\nabla_r N_c}{N_c} \right). \end{aligned} \tag{2.72}$$

The *Hydrodynamic (HD) model* for electrons takes into account eqs. (2.49), (2.60), (2.66) and (2.72). It means the particle density, current density, particle gas temperature and energy current density are considered. Since the first order derivatives in eqs. (2.60), (2.66) and (2.72) cause numerical problems [59], these derivatives are omitted in the GALENE simulator [1]. As the relaxation times are smaller than one picosecond, this limits the HD model to frequencies below 100 GHz, which is sufficient for modeling devices in the scope of this dissertation.

The *Drift diffusion (DD) model* assumes that carriers are always in equilibrium with the lattice, i.e. $T_n = T_p = T_L$. Therefore, the Einstein relation between diffusion coefficient $D_n$ and mobility $\mu_n$ of $D_n = k_B T \mu_n / q$ can be applied to get the well-known expression

$$\mathbf{j}_n = -q \mu_n n \nabla_{\mathbf{r}} \Psi + q D_n \nabla_{\mathbf{r}} n. \tag{2.73}$$

## 2.4 Summary

In conclusion, Monte Carlo stands on the highest level for an exact solution of the BTE. Different scattering mechanisms can be included. It can

represent well nonlocal effects, which are very essential in simulating devices with a high electric field. However, it is time and CPU-consuming. HD and DD are alternative solutions, which are much faster (of course with the loss of the accuracy, especially nonlocal effects). Therefore, depending on the requirements and devices, the appropriate model should be chosen.

One more advantage of the Monte Carlo method is that it is able to investigate the devices with strained channels when the band structures of these materials are known. As the band structure data get more realistic, the result from Monte Carlo simulations becomes more reliable. The next chapter will discuss the properties of the strained materials and the method for calculating their band structures to obtain the most realistic data.

# Chapter 3

# Strain engineering

## 3.1 Stress - strain under elasticity condition

The property of solid materials to deform under the application of an external force and to regain their original shape after the removal of the force is referred to as its elasticity. The external force applied on a specified area is known as the stress, while the amount of the deformation is called the strain. In this section, the theory of stress and strain, as well as their interdependence is discussed.

### 3.1.1 Stress

Assuming that a solid object oriented in Cartesian coordinates under the action of a number of forces along different directions in which the vector sum of these forces is zero. Take a slice orthogonal to the x-direction and define a small area on this slice as $\Delta A_x$, the total force acting on this small area can be written by

$$\Delta \mathbf{F} = \Delta F_x \cdot \hat{\mathbf{e}}_x + \Delta F_y \cdot \hat{\mathbf{e}}_y + \Delta F_z \cdot \hat{\mathbf{e}}_z, \tag{3.1}$$

where $\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z$ are the unit vectors in Cartesian coordiate system. As $\Delta \mathbf{F}$ is the total force acting only on $\Delta A_x$, the magnitude of $\Delta \mathbf{F}$ will change when $\Delta A_x$ changes. Therefore, we can define the following scalar quantities

$$\sigma_{xx} = \lim_{\Delta A_x \to 0} \frac{\Delta F_x}{\Delta A_x}, \sigma_{xy} = \lim_{\Delta A_x \to 0} \frac{\Delta F_y}{\Delta A_x}, \sigma_{xz} = \lim_{\Delta A_x \to 0} \frac{\Delta F_z}{\Delta A_x}. \tag{3.2}$$

The subscripts $i$ and $j$ in $\sigma_{ij}$ refer to the plane and the force direction, respectively. Similarly considering slices orthogonal to the y and z-directions,

we obtain

$$\sigma_{yx} = \lim_{\Delta A_y \to 0} \frac{\Delta F_x}{\Delta A_y}, \sigma_{yy} = \lim_{\Delta A_y \to 0} \frac{\Delta F_y}{\Delta A_y}, \sigma_{yz} = \lim_{\Delta A_y \to 0} \frac{\Delta F_z}{\Delta A_y}, \tag{3.3}$$

and

$$\sigma_{zx} = \lim_{\Delta A_z \to 0} \frac{\Delta F_x}{\Delta A_z}, \sigma_{zy} = \lim_{\Delta A_z \to 0} \frac{\Delta F_y}{\Delta A_z}, \sigma_{zz} = \lim_{\Delta A_z \to 0} \frac{\Delta F_z}{\Delta A_z}. \tag{3.4}$$

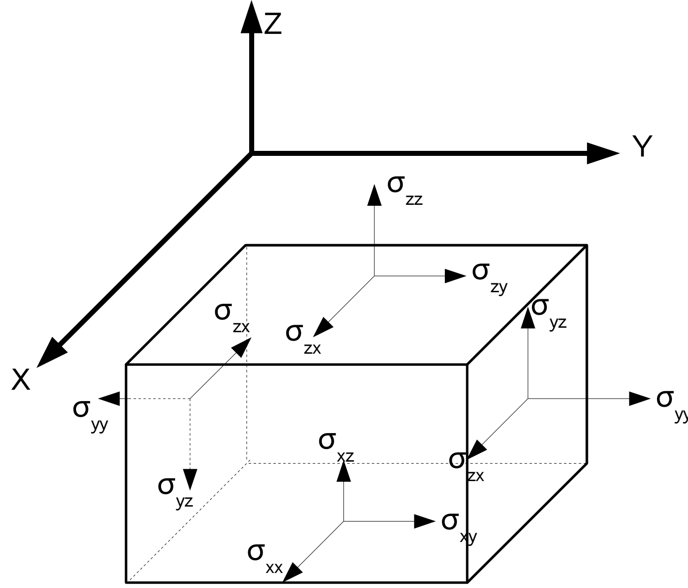


Figure 3.1: Components of a stress tensor.

Fig. 3.1 shows components of a stress tensor. For a static equilibrium, we have $\sigma_{ij} = \sigma_{ji}$. Hence, only six independent scalar quantities are left and they can be arranged in a matrix form to yield a stress tensor as

$$\hat{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \tag{3.5}$$

### 3.1.2 Strain

A body under an elastic deformation experiences an internal restoring force. The amount of deformation caused is called strain. There are two types of strain: normal strain and shear strain. Normal strain is the change in length in a given direction divided by the initial length in that direction. Shear strain is the complement of the angle between initially perpendicular segments.

When a force is applied on a solid object, the object may be transformed, rotated and deformed. For example, we have two points at locations $\mathbf{l}$ and $\mathbf{l}+d\mathbf{l}$; then under stress these points are moved to the locations $\mathbf{l}+\mathbf{u}(\mathbf{l})$ and $\mathbf{l}+d\mathbf{l}+\mathbf{u}(\mathbf{l}+d\mathbf{l})$. The absolute squared distance between the deformed points can then be obtained as [9]

$$\sum_{i=x,y,z} (dl_i + u_i(\mathbf{l}+d\mathbf{l}) - u_i(\mathbf{l}))^2 = \sum_{i=x,y,z} \left( dl_i + \sum_{j=x,y,z} \frac{\partial u_i}{\partial l_j} dl_j \right)^2 . \tag{3.6}$$

Since $d\mathbf{l}$ is considered to be a small displacement, a Taylor expansion around $\mathbf{l}$ can be performed, which gives for the the absolute squared distance

$$\sum_i \left( dl_i + \sum_j \frac{\partial u_i}{\partial l_j} dl_j \right)^2 = \sum_i dl_i^2 + 2\sum_{i,j} dl_i \frac{\partial u_i}{\partial l_j} dl_j + \sum_{i,j,k} \frac{\partial u_i}{\partial l_j} dl_j \frac{\partial u_i}{\partial l_k} dl_k, \tag{3.7}$$

where $k = x, y, z$. Since the first term in eq. (3.7) is the squared distance between the points in the relaxed system, the change in the squared distance caused by strain becomes

$$\begin{aligned} D(d\mathbf{l}) &= \sum_{i,j} dl_i \left( \frac{\partial u_i}{\partial l_j} + \frac{\partial u_j}{\partial l_i} \right) dl_j + \sum_{i,j,k} \frac{\partial u_k}{\partial l_i} \frac{\partial u_k}{\partial l_j} dl_i dl_j \\ &= \sum_{i,j} dl_i \underbrace{\left[ \left( \frac{\partial u_i}{\partial l_j} + \frac{\partial u_j}{\partial l_i} \right) + \sum_k \frac{\partial u_k}{\partial l_i} \frac{\partial u_k}{\partial l_j} \right]}_{2\varepsilon_{ij}} dl_j \\ &= 2\sum_{i,j} dl_i \varepsilon_{ij} dl_j, \end{aligned} \tag{3.8}$$

where $\varepsilon_{ij}$ are defined to be the components of a strain tensor. If the strain is small enough such that $\frac{\partial u_k}{\partial l_i} \ll 1$ holds, resulting tensor components can be simplified to

$$\varepsilon_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial l_j} + \frac{\partial u_j}{\partial l_i} \right) \tag{3.9}$$

The strain tensor is therefore symmetric, i.e. $\varepsilon_{ij} = \varepsilon_{ji}$. Consequently, the strain tensor $\hat{\varepsilon}$ comprises six independent components similarly to the case of the stress tensor

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{pmatrix} . \tag{3.10}$$

Table 3.1: Elastic stiffness constants $c_{ij}$ in GPa and elastic compliance constants $s_{ij}$ in $10^{-12}$m$^2$/N [78].

| | $c_{11}$ | $c_{12}$ | $c_{44}$ | $s_{11}$ | $s_{12}$ | $s_{44}$ |
|---|---|---|---|---|---|---|
| Si | 166.0 | 64.0 | 79.6 | 7.67 | -2.13 | 12.6 |
| Ge | 126.0 | 44.0 | 67.7 | 9.69 | -2.50 | 14.8 |

The diagonal elements are the normal strains in $x$, $y$ and $z$ directions, respectively. The off-diagonal terms are shear strains. The engineering shear strain, denoted as $\gamma_{ij}$, is defined as

$$\gamma_{ij} = \varepsilon_{ij} + \varepsilon_{ji} = 2\varepsilon_{ij}. \tag{3.11}$$

### 3.1.3 Stress-strain dependence

Hook's law of elasticity is an approximation which states that the amount by which a material body is deformed (the strain) is linearly related to the force causing the deformation (the stress). The most general relationship between stress and strain can be mathematically written as

$$\sigma_{ij} = C_{ijkl}\varepsilon_{ij}, \tag{3.12}$$

where $C_{ijkl}$ is a fourth order elastic stiffness tensor comprising 81 coefficients. However, depending on the symmetry of the crystal, the number of coefficients can be reduced. For cubic crystals such as Si and Ge, only three unique coefficients $c_{11}$, $c_{12}$ and $c_{44}$, are not equal to zero. These coefficients are known as the stiffness constants. As a result, the generalized Hook law can be written in a matrix form as

$$\begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{pmatrix} \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{xy} \\ \varepsilon_{xz} \\ 2\varepsilon_{yz} \\ 2\varepsilon_{xz} \\ 2\varepsilon_{xy} \end{pmatrix} \tag{3.13}$$

Normally we are interested in the strain which arises from a certain stress condition. The strain components can be obtained by inverting Hook's law from eq. (3.12)

$$\varepsilon_{ij} = S_{ijkl}\sigma_{ij}, \tag{3.14}$$

where $S_{ijkl}$ are called the compliance coefficients. Combining eq. (3.12) and (3.14), we have $\mathbf{S} = \mathbf{C}^{-1}$. From three independent elastic stiffness constants,

we can get three independent compliance coefficients respectively as

$$s_{11} = \frac{c_{11} + c_{12}}{c_{11}^2 + c_{11}c_{12} - 2c_{12}^2}, \tag{3.15}$$

$$s_{12} = -\frac{c_{12}}{c_{11}^2 + c_{11}c_{12} - 2c_{12}^2} \tag{3.16}$$

and

$$s_{44} = \frac{1}{c_{44}}. \tag{3.17}$$

Therefore eq. (3.13) can be rewritten as

$$\begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ 2\varepsilon_{yz} \\ 2\varepsilon_{xz} \\ 2\varepsilon_{xy} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{12} & s_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{44} \end{pmatrix} \begin{pmatrix} \sigma_{xx} \\ \sigma_{xy} \\ \sigma_{xz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix}. \tag{3.18}$$

The compliance and stiffness coefficients for pure Si and Ge are listed in Tab. 3.1. Linear interpolation is used to get respective coefficients for a SiGe alloy.

#### Coordinate transformation

It is often required to know the stress in the crystallographic coordinate system. Consider a coordinate system $[x', y', z']$ in which $x'$ axis is parallel to the stress direction. The stress in the crystallographic coordinate system $[x, y, z]$ can be calculated using the transformation matrix $\mathbf{U}$

$$\mathbf{U}(\theta, \phi) = \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{pmatrix}. \tag{3.19}$$

Here $\theta$ denotes the polar and $\phi$ the azimuthal angle of the stress direction relative to the crystallographic coordinate system, as shown in Fig. 3.2. In the system $[x', y', z']$, the stress tensor $\hat{\sigma}'$ has only one non-zero component $\sigma'_{xx}$. The stress tensor $\hat{\sigma}$ in the crystallographic coordinate system is then given by

$$\hat{\sigma} = \mathbf{U}\hat{\sigma}'\mathbf{U}^T. \tag{3.20}$$

When a non-zero stress of magnitude $P$ is applied along [100], [110] and [111] directions, the stress tensors in the principal coordinate system are

$$\hat{\sigma}_{[100]} = \begin{pmatrix} P & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.21}$$

$$\hat{\sigma}_{[110]} = \begin{pmatrix} P/2 & P/2 & 0 \\ P/2 & P/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.22}$$

and

$$\hat{\sigma}_{[111]} = \begin{pmatrix} P/3 & P/3 & P/3 \\ P/3 & P/3 & P/3 \\ P/3 & P/3 & P/3 \end{pmatrix}. \tag{3.23}$$

From eqs. (3.21)-(3.23), the corresponding strain tensors can be determined

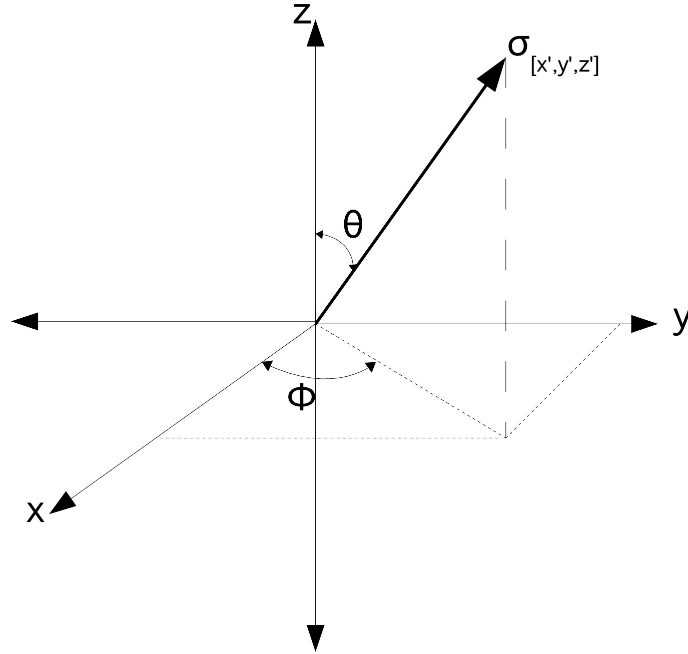


Figure 3.2: Stress direction $[x', y', z']$ in the crystallographic coordinate system $[x, y, z]$.

as in the form of eq. (3.18)

$$\hat{\varepsilon}_{[100]} = \begin{pmatrix} s_{11}P & 0 & 0 \\ 0 & s_{12}P & 0 \\ 0 & 0 & s_{12}P \end{pmatrix}, \tag{3.24}$$

$$\hat{\varepsilon}_{[110]} = \begin{pmatrix} (s_{11}+s_{12})P/2 & s_{44}P/4 & 0 \\ s_{44}P/4 & (s_{11}+s_{12})P/2 & 0 \\ 0 & 0 & s_{12}P \end{pmatrix} \tag{3.25}$$

and

$$\hat{\varepsilon}_{[111]} = \begin{pmatrix} (s_{11}+2s_{12})P/3 & s_{44}P/6 & s_{44}P/6 \\ s_{44}P/6 & (s_{11}+2s_{12})P/3 & s_{44}P/6 \\ s_{44}P/6 & s_{44}P/6 & (s_{11}+2s_{12})P/3 \end{pmatrix}. \quad (3.26)$$

## 3.2 Strain effects on crystal symmetry and electronic band structures

### 3.2.1 Basic properties of diamond structures

Popular cubic semiconductors such as Si or Ge have a diamond structure. The basis vectors of the direct lattice for such a structure are

$$\mathbf{a}_1 = \frac{a_0}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{a}_2 = \frac{a_0}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{a}_3 = \frac{a_0}{3}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad (3.27)$$

where $a_0$ is the lattice constant of a relaxed lattice. A direct lattice vector is defined as

$$\mathbf{R}_{i,j,k} = i\mathbf{a}_1 + j\mathbf{a}_2 + k\mathbf{a}_3, \quad (3.28)$$

where $i, j, k$ are integers. The basis vectors of the reciprocal lattice is related to the basis vectors of the direct lattice by $\mathbf{a}_i\mathbf{b}_i = 2\pi\delta_{i,j}$. From eq. (3.27), they can be calculated by

$$\mathbf{b}_1 = \frac{2\pi}{a_0}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{b}_2 = \frac{2\pi}{a_0}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \mathbf{b}_3 = \frac{2\pi}{a_0}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}. \quad (3.29)$$

Then a general reciprocal lattice vector has the form

$$\mathbf{G}_{i,j,k} = i\mathbf{b}_1 + j\mathbf{b}_2 + k\mathbf{b}_3. \quad (3.30)$$

The first Brillouin zone (BZ) represents the central cell of the reciprocal lattice and contains all points nearest to the enclosed reciprocal lattice vector. Their 14 faces are

$$|k_x| + |k_y| + |k_z| = \frac{3}{2}\frac{2\pi}{a_0}, |k_x| = \frac{2\pi}{a_0}, |k_y| = \frac{2\pi}{a_0}, |k_z| = \frac{2\pi}{a_0}. \quad (3.31)$$

In the case of unstrained crystal, the energy bands are invariant under eight spatial reflections and six permutations, whose combination results in 48 different symmetry operations. Therefore, the energy band information is

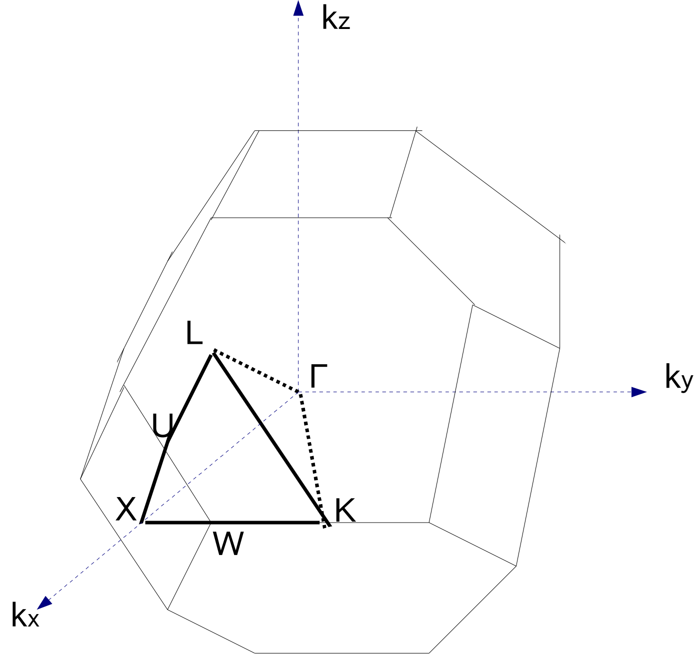


Figure 3.3: The first BZ of relaxed Si and the first irreducible wedge.

redundant in the full BZ and can hence be reduced to one wedge containing only $1/48^{\mathrm{th}}$ of the full BZ. Such wedge is called an *irreducible wedge* [59]. The first irreducible wedge has six corners:

$$\Gamma = \frac{2\pi}{a_0}\begin{pmatrix}0\\0\\0\end{pmatrix}, L = \frac{2\pi}{a_0}\begin{pmatrix}1/2\\1/2\\1/2\end{pmatrix}, K = \frac{2\pi}{a_0}\begin{pmatrix}3/4\\3/4\\0\end{pmatrix},$$
$$W = \frac{2\pi}{a_0}\begin{pmatrix}1\\1/2\\0\end{pmatrix}, X = \frac{2\pi}{a_0}\begin{pmatrix}1\\0\\0\end{pmatrix}, U = \frac{2\pi}{a_0}\begin{pmatrix}1\\1/4\\1/4\end{pmatrix}, \tag{3.32}$$

and five faces:

$$k_x + k_y + k_z = \frac{3}{2}\frac{2\pi}{a_0},$$
$$k_x = \frac{2\pi}{a_0}, k_x = k_y, k_y = k_z, k_z = 0. \tag{3.33}$$

The first BZ and the first irreducible wedge is illustrated in Fig. 3.3.

### 3.2.2 Strain effects on crystal symmetry

Under homogeneous strain, the basis vectors of the direct lattice of a crystal $\mathbf{a}'_i$ is obtained by deforming the vectors $\mathbf{a}_i$ of an unstrained crystal by

$$\mathbf{a}'_i = (\mathbf{1} + \hat{\varepsilon}) \cdot \mathbf{a}_i, \tag{3.34}$$

where $\hat{\varepsilon}$ is the strain tensor defined in eq. (3.10). Moreover, the volume $\Omega'_0$ of the strained primitive unit cell is also modified as [123]

$$\Omega'_0 = \Omega_0(1 + \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}), \tag{3.35}$$

where $\Omega_0$ is the primitive unit cell of an unstrained crystal and $\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}$ are the diagonal elements of the strain tensor. The change of the lattice vectors and cell volume under strain affects the pseudopotential factors, which are used for band structure calculation (presented in Sec. 3.3.3).

As strain reduces the symmetry of the lattice in the first BZ, the number of irreducible wedges also changes. Details of the number of irreducible wedges for different types of strain are also presented in Sec. 3.3.3.

### 3.2.3 Strain effects on electronic band structures

Any strain can be decomposed into a hydrostatic strain and two types of shear strain [128]. One type of shear strain is related to the change of lengths along the three axes and the other is related to the rotation of the axes of an infinitesimal cube. Since hydrostatic strain does not break the symmetry of crystals, it only shifts energy levels without lifting band degeneracy. It is, hence, considered undesirable because of the issues resulting from strain relaxation, band gap narrowing and MOSFET threshold voltage shifts [80].

Under in-plane (x-y plane) biaxial stress, the resulting strain contains only the first type of shear strain which lowers the crystal symmetry by shortening (biaxial tensile stress) or lengthening (biaxial compressive stress) the z-direction lattice spacing with respect to the x or y direction. Fig. 3.4(a) illustrates this type of stress. Due to the difference of the z direction with respect to the other two directions, the conduction band minima of Si along the z axis split away from those along the x and y direction. The sixfold degenerate valleys then split in energy into the $\Delta_2$ and $\Delta_4$ valleys. The case for the Ge conduction band is different. The biaxial stress has the same effect on the eight L valleys; therefore L valleys do not split. Band splitting for Si and Ge is schematically presented in Fig. 3.4(b) [109].

Under uniaxial stress, the resulting strain contains both types of shear strain. The distortion of the crystal symmetry and the band splitting under

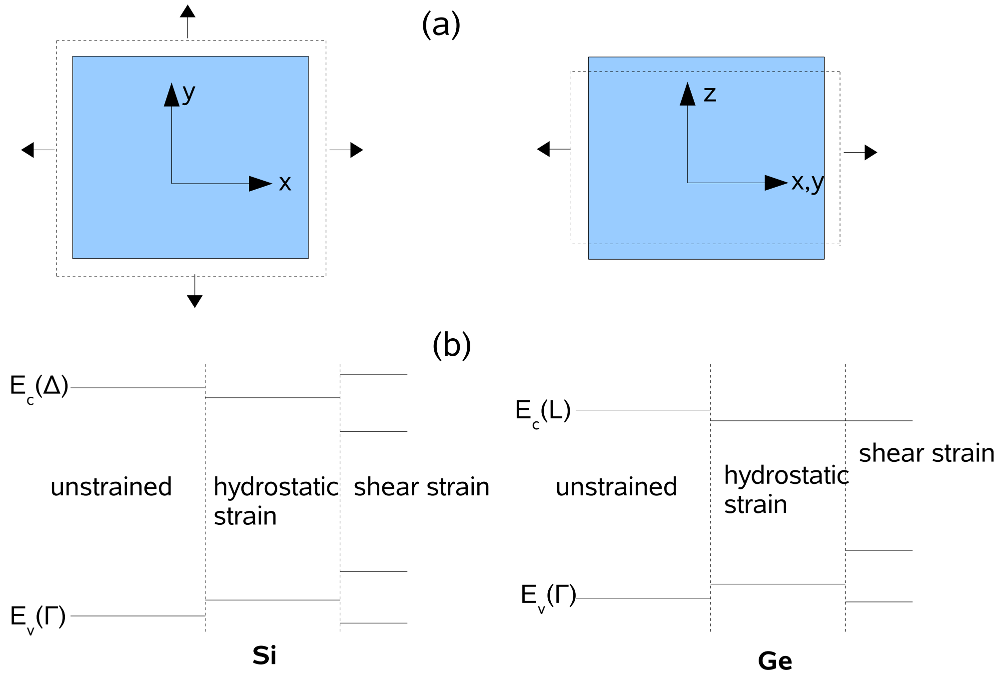


Figure 3.4: (a) Cubic crystals under in-plane biaxial tensile stress (b) Diagrams of band splitting of Si and Ge under in-plane biaxial tensile stress.

a uniaxial stress along the [110] direction are depicted in Fig. 3.5(a) and Fig. 3.5(b) respectively. Similar to biaxial stress, the Si conduction band splits into $\Delta_2$ and $\Delta_4$ valleys due to the nonequivalency between the z and x and y axes. However, the splitting for the Ge conduction band is different. It splits into two twofold L valley groups, since the uniaxial stress distinguishes the $\langle 111 \rangle$ directions according to their projections in the x-y plane [109].

Beside band lifting, band warping is another important difference between uniaxial and biaxial stress under symmetry breaking. The details of the band warping can be explored by pseudopotential methods, which are presented in Sec. 3.3.3. Applying stress along a low symmetry axis causes more destruction of crystal symmetry and results in a greater band warping than stress along a high symmetry axis. This is the reason why the valence bands of uniaxially stressed Si are more asymmetrical than those of biaxially stressed.

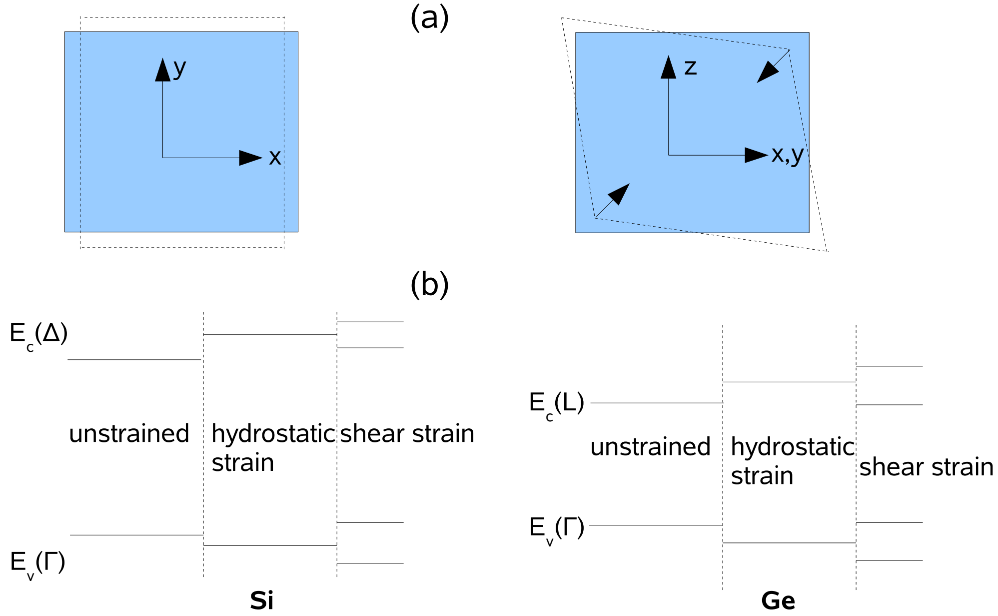


Figure 3.5: (a) Cubic crystals under a uniaxial compressive stress along the [110] direction (b) Diagrams of band splitting of Si and Ge under uniaxial compressive stress.

## 3.3 Band structure calculation

### 3.3.1 Pseudopotential method

The pseudopotential approach comes from the fact that the wave function may have strong spatial oscillations near the core (of the material atom). The wave function, hence, can be difficult to be solved in this case [128]. This difficulty can be overcome by dividing the wave function into a smooth part, called *pseudo-wave function*, and an oscillatory part. The kinetic energy from the oscillatory part provides an "effective repulsion" for the valence electrons near the core; hence we can approximate the strong true potential by a weaker "effective potential" or pseudopotential for the valence electrons. Fig. 3.6(a) shows the variation of the pseudopotential in Si along distance $r$ from the nucleus [25].

The concept of replacing the true potential with a pseudopotential can be justified mathematically. It was also shown in [25] that both the conduction and valence band states can be reproduced correctly while the cumbersome core states are eliminated. Under this approximation, we can write the one-electron Hamiltonian by

$$H = p^2/2m + V(\mathbf{r}) \tag{3.36}$$

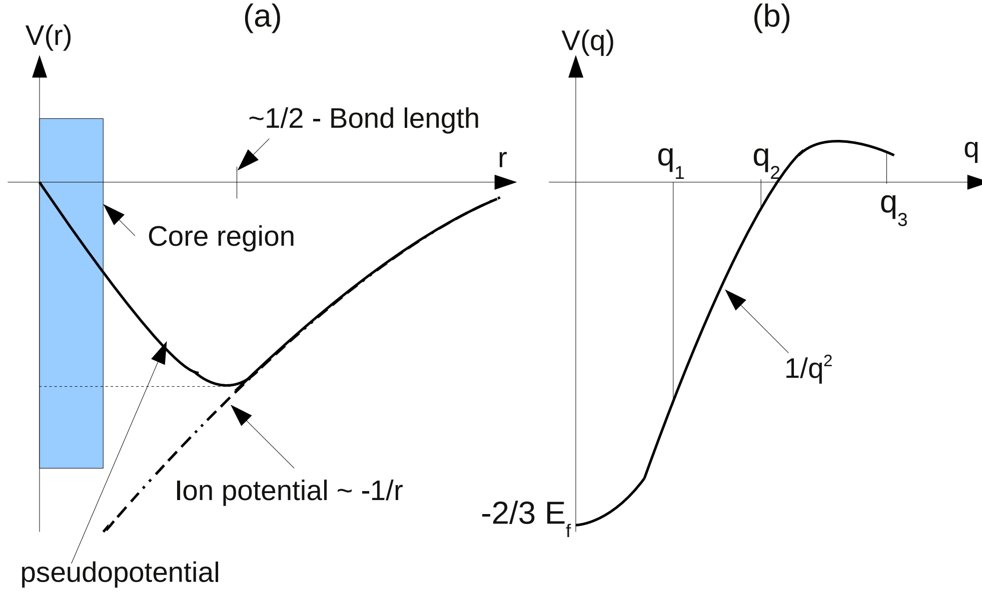


Figure 3.6: Schematic plot of the atomic pseudopotential of Si in: (a) real space and (b) reciprocal space ($\mathrm{q}_1 = \sqrt{3}, \mathrm{q}_2 = \sqrt{8}, \mathrm{q}_3 = \sqrt{11}$) [25].

and the one-electron Schrödinger equation can be replaced by the pseudo-wave equation as

$$\left[\frac{p^2}{2m} + V(\mathbf{r})\right]\psi_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}), \tag{3.37}$$

where $\psi$ is the pseudo-wave-function and $V(\mathbf{r})$ is the pseudopotential. This function can be used to calculate the physical properties of the semiconductors which depend on the valence and conduction electrons only.

The pseudopotential can be considered as a simple function of position [20]

$$V(\mathbf{r}) = \sum_{\mathbf{G}} V(\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{3.38}$$

with

$$V(\mathbf{G}) = \sum_{\alpha} S_\alpha(\mathbf{G})V_\alpha(\mathbf{G}), \tag{3.39}$$

$$S_\alpha(\mathbf{G}) = \frac{1}{N_\alpha}\sum_{\text{cell}j} e^{-i\mathbf{G}\cdot\mathbf{R}_j^\alpha}, \tag{3.40}$$

and

$$V_\alpha(\mathbf{G}) = \frac{1}{\Omega_a}\int_{\Omega_a} V^a(\mathbf{r})e^{-i\mathbf{G}\cdot\mathbf{r}}d\mathbf{r}. \tag{3.41}$$

In eqs. (3.38)-(3.41), $\mathbf{G}$ are general reciprocal lattice vectors defined in eq. (3.30), $V_\alpha(\mathbf{G})$ are the atomic form factors, $\Omega_a$ is the atomic volume, $N_\alpha$ is the number of atomic species $\alpha$ present and $\mathbf{R}_j^\alpha$ is the position of the $j^{th}$ atom of the $\alpha^{th}$ species. Fourier components $V(\mathbf{G})$ are known as the *pseudopotential form factors.*

Since pseudopotentials are only small pertubations on free electron band structures, $\psi_\mathbf{k}$ can be expanded into a sum of plane waves by

$$\psi_\mathbf{k} = e^{i\mathbf{k}\cdot\mathbf{r}} u_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{G}'} U(\mathbf{G}') e^{i\mathbf{G}'\cdot\mathbf{r}}, \tag{3.42}$$

where $\mathbf{G}'$ are general reciprocal lattice vectors and $U(\mathbf{G}')$ are the Fourier components of the periodic part $u_\mathbf{k}(\mathbf{r})$.

By substituting eqs. (3.38) and (3.42) into eq. (3.37), we can get the following equation

$$\sum_{\mathbf{G}} \left\{ \left[ \frac{\hbar^2(\mathbf{k}+\mathbf{G})^2}{2m} - E_\mathbf{k} \right] U(\mathbf{G}) + \sum_{\mathbf{G}'} V(|\mathbf{G}-\mathbf{G}'|) U(\mathbf{G}') \right\} = 0 \tag{3.43}$$

The expression given in eq. (3.43) is zero when each term in the sum is identically zero, which implies the following condition

$$\left[ \frac{\hbar^2(\mathbf{k}+\mathbf{G})^2}{2m} - E_\mathbf{k} \right] U(\mathbf{G}) + \sum_{\mathbf{G}'} V(|\mathbf{G}-\mathbf{G}'|) U(\mathbf{G}') = 0 \tag{3.44}$$

In this way, the band structure calculation is reduced to solving the eigenvalue problem specified by eq. (3.44). The matrix size and calculation accuracy depend on the number of reciprocal lattice vectors $\mathbf{G}'$ used.

### 3.3.2 Nonlocal pseudopotential method

The method discussed in the above section with such approximations is called the local pseudopotential method. It was proved in [20] that this local approximation cannot yield results which have a good agreement with experimental data. Therefore nonlocal pseudopotential should be considered.

In the nonlocal pseudopotential method, the pseudopotential $V(\mathbf{r})$ can be written as the sum of a local $V_L$ and a nonlocal $V_{NL}$ term

$$V(\mathbf{r}) = V_L(\mathbf{r}) + V_{NL}(\mathbf{r}). \tag{3.45}$$

$V_{NL}$, which depends on the angular-momentum $l$, is written as [20]

$$V_{NL}(\mathbf{r}) = \sum_{l=0}^{\infty} A_l(E) f_l(r) \mathcal{P}_l, \tag{3.46}$$

where $A_l$ is an energy-dependent well depth, $f_l$ is a function simulating the effect of core states with $l$ symmetry and $\mathcal{P}_l$ is a projection operator for the $l^{th}$ angular-momentum component. $f_l$ has a simple form

$$f_l = \begin{cases} 1 & \text{if } r < R_l \\ 0 & \text{if } r \geq R_l \end{cases}, \tag{3.47}$$

where $R_l$ is a cut-off radius. The values of $R_l$ and $\mathcal{A}_l$ for Si and Ge are presented in [93].

From this nonlocal pseudopotential, the Hamiltonian in eq. (3.36) will be adjusted by a nonlocal correction term accordingly. Then eq. (3.43) is solved to get the respective eigenvalues.

### 3.3.3 Empirical pseudopotential method

#### Empirical approach

There are two approaches for calculating band structures according to the pseudopotential method: One is the empirical pseudopotential method (EPM) and the other is the *ab initio* pseudopotential method.

In EPM, pseudopotential form factors are determined by fitting a small number of experimental data, such as the position of peaks in optical reflectivity spectra or features in the photoelectron spectra [128]. The flow diagram for calculating the band structure with EPM is shown in Fig. 3.7. The biggest disadvantage of the EPM method is that it requires experimental data.

The *ab initio* method determines the pseudopotential form factors from first principles without any experimental input. This method gives good results for the ground state properties such as the cohesive energies and charge density of the valence electrons [128]; however it gives poor results for the charge density of the conduction electrons. As a result, the EPM method is preferred.

In this dissertation, EPM also stands for *empirical nonlocal pseudopotential method* because only the nonlocal method is used thoroughly. EPM is applied to calculate the band structures for all investigated materials, including relaxed and strained ones.

#### Band structures for strained materials

The calculation of band structures of strained materials is presented in detail in Ref. [122] and is applied in this dissertation. When the strain is included, the following parameters should be adjusted when calculating band structures:

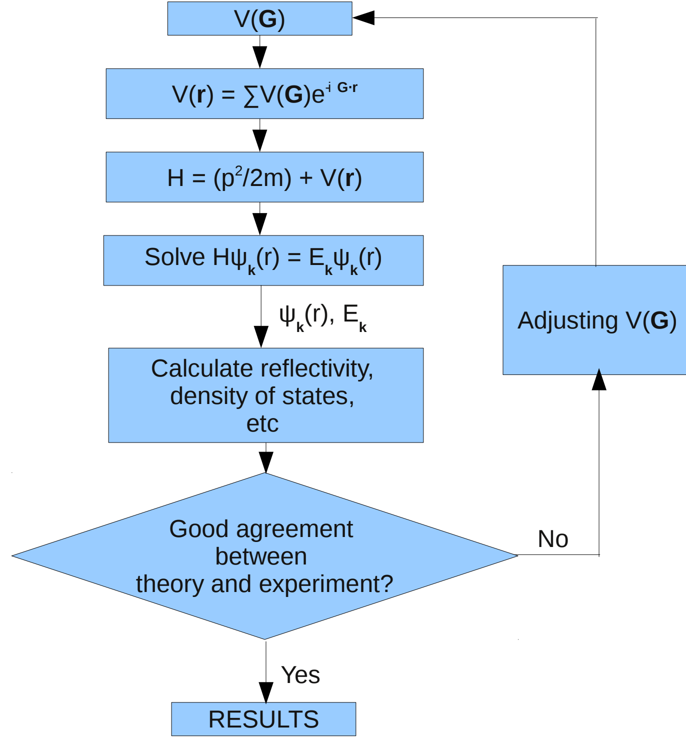


Figure 3.7: Flow diagram for calculating band structures with the EPM method [128].

- Direct lattice vectors and reciprocal lattice vectors: They are modified by eq. (3.34). These quantities are then used in the expansion of the pseudopotential (eq. (3.38)) and the pseudo-wave-function (eq. (3.42)) and the normalization of the volume of the strained unit cell (eq. (3.35)).

  The pseudopotential form factors of the strained lattice are obtained by performing a cubic spline interpolation through the pseudopotential form factors $V_0$, $V_3$, $V_8$, $V_{11}$, and $V_{3k_\mathrm{F}}$ (as in Fig. 3.6(b)). Following [93], $V_0$ is set to $-2E_\mathrm{F}/3$ and $V_{3k_\mathrm{F}}$ is set to 0, where $k_\mathrm{F}$ denotes the Fermi wave vector of the free electron gas.

- Displacement of the central atom in the primitive unit cell: Under strain, the absolute position of the central atom in the bulk primitive unit cell is undetermined; hence an additional parameter for the displacement has to be taken into account to obtain this position. This process is illustrated in Fig. 3.8.

- Number of irreducible wedges: Strain reduces the symmetry of the

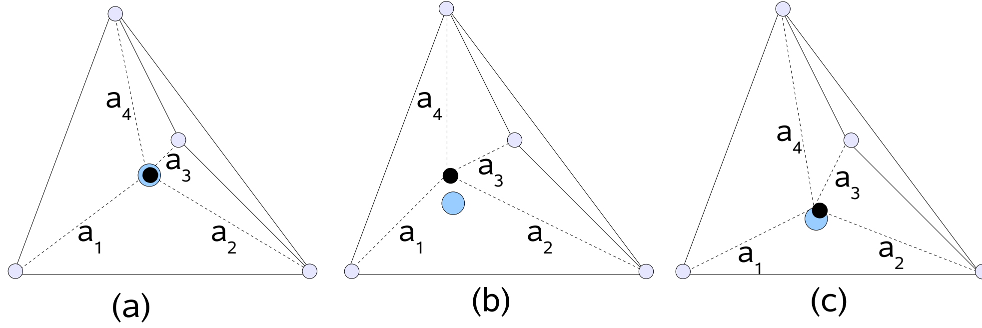


Figure 3.8: Unit cell of the diamond structure with a center atom and its four nearest neighbors: (a) Unstrained crystal: $a_1 = a_2 = a_3 = a_4$; (b) Strained crystal: $a_1 \neq a_2 \neq a_3 \neq a_4$; (c) Center atom is displaced to obtain $a_1 \approx a_2 \approx a_3 \approx a_4$ [122].

first BZ; hence the number of irreducible wedges, which are defined in Sec. 3.2.1, varies accordingly. Tab. 3.2 shows the number of irreducible wedges in the first BZ for different strain types.

## 3.4 Band offsets and band gap for heterostructures $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ - a special case of strain

Si/SiGe heterostructure has been integrated widely into the existing silicon technology - both in bipolar and MOS applications. The SiGe HBT (heterojunction bipolar transistor) has been commercialized in high frequency and low noise applications. The lattice constant mismatch between the active layer $Si_{1-x}Ge_x$ and the relaxed substrate layer $Si_{1-y}Ge_y$ causes the strain in the active layer, which affects the performance of the device. This part presents the method for calculating the band offsets and band gap of heterostructures. These parameters are necessary for other calculations (e.g. impact ionization) and device simulations.

The symbol $\Delta E$ is used to denote the energy difference between the active layer and the substrate layer $\Delta E = E(active) - E(substrate)$. For a relaxed $Si_{1-y}Ge_y$ alloy, the lattice constant is given as [93]

$$a_0(y) = a_0(Si) + 0.200326y(1-y) + [a_0(Ge) - a_0(Si)]y^2, \tag{3.48}$$

where $a_0(Si)$ and $a_0(Ge)$ are the lattice constants of relaxed Si and Ge, respectively.

Table 3.2: Number of irreducible wedges (NW) in the first BZ for different strain types.

| Strain type | Strain tensor | NW |
|---|---|---|
| Relaxed material | $\hat{\varepsilon} = 0$ | 48 |
| Biaxial strain for {100} plane<br><br>Uniaxial strain along [100] direction | $\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{yy} \end{pmatrix}$ | 16 |
| Biaxial strain for {111} plane<br><br>Uniaxial strain along [111] direction | $\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xy} \\ \varepsilon_{xy} & \varepsilon_{xx} & \varepsilon_{xy} \\ \varepsilon_{xy} & \varepsilon_{xy} & \varepsilon_{xx} \end{pmatrix}$ | 12 |
| Biaxial strain for {110} plane<br><br>Uniaxial strain along [110] direction | $\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ \varepsilon_{xy} & \varepsilon_{xx} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}$ | 8 |
| Uniaxial strain along [120] direction | $\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ \varepsilon_{xy} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}$ | 4 |
| Uniaxial strain along other directions<br><br>(e.g. [123]) | $\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix}$ | 2 |

For a strained $\mathrm{Si_{1-x}Ge_x}$ layer, there are the parallel (in-plane) $a_{||}$ lattice constants and perpendicular (to the interface) $a_{\perp}$ lattice constants. As $a_{||}$ has to be conserved throughout the layer sequence, we have

$$a_{||} = a_0(y). \tag{3.49}$$

The lateral $\varepsilon_{||}$ of the strain tensor for strained layer $\mathrm{Si_{1-x}Ge_x}$ is defined as [93]

$$\varepsilon_{||} = \frac{a_{||}}{a_0(x)} - 1. \tag{3.50}$$

The above parameters are calculated for the case that the heterostructure is strained internally, i.e. without external forces. On some purposes, the whole structure is stressed uniaxially by an external force at certain directions (e.g. [31]). Therefore, the following part will discuss two cases of the heterostructure: Without additional stress and with external stress.

### 3.4.1 Without additional uniaxial stress

There are two main types of Si/SiGe heterostructure: One is strained SiGe over relaxed Si and another is strained Si over relaxed SiGe substrate, as illustrated in Fig. 3.9. Accordingly, there are two types of band alignments, named Type I and Type II. Type I refers to the case of a larger valence band offset and a smaller conduction band offset. In this type, band offset at the valence band is decisive to the band gap. Type II refers to the case when both the lowest conduction band level and the lowest valence band level are situated in the Si layer. Band offset and band gap for these cases are explored below.

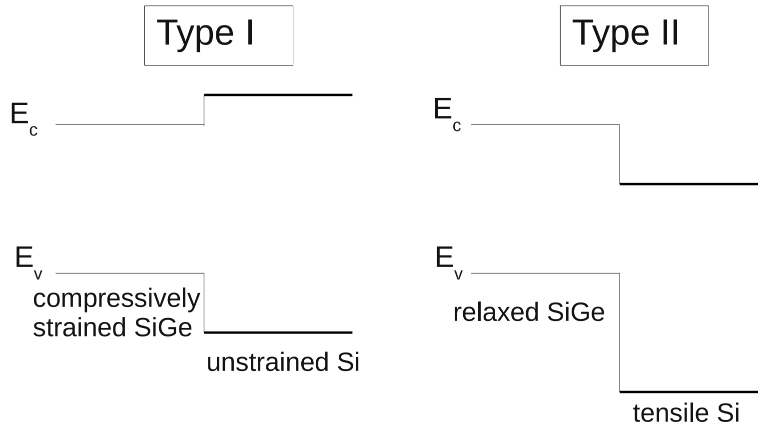


Figure 3.9: Two types of band alignments in Si/SiGe heterostructures.

#### Band offsets

**Valence bands** In [26], the offset of the *average* energy of the uppermost three valence bands has been calculated for a strained Si/Ge on a (100) substrate with the *ab initio* method. The findings there, together with the earlier calculation from Van de Walle and Martin [29], indicate that the offset is nearly independent from the strain conditions. Rieger and Vogl [93,94] have suggested an interpolation formula for calculating the average valence band offset $\Delta E_v^{av}$ at the interface between a strained $\mathrm{Si_{1-x}Ge_x}$ layer and a relaxed substrate $\mathrm{Si_{1-y}Ge_y}$ of an arbitrary growth direction by

$$\Delta E_v^{av} = (0.47 - 0.06y)(x - y). \tag{3.51}$$

This section will follow eq. (3.51) for calculating the average of the valence band offsets. For a complete description of the band alignment and band gap in SiGe heterostructures, the valence band shifting due to strain needs to be

considered as well. The following part will investigate the shift in valence bands due to different growth directions.

- **Stress along [100] direction:**

  When the growth is along the [100] direction, the strain tensor for the active layer is

$$\hat{\varepsilon}_{[100]} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}, \tag{3.52}$$

  where

$$\begin{aligned} \varepsilon_{xx} &= \varepsilon_{\perp} = -2\frac{c_{12}}{c_{11}}\varepsilon_{||}, \\ \varepsilon_{yy} &= \varepsilon_{zz} = \varepsilon_{||}. \end{aligned} \tag{3.53}$$

  $c_{ij}$ in eq. (3.53) as well as in the following equations are the stiffness constants (Tab. 3.1).

  Under the effect of uniaxial strain, the valence bands split into HH (heavy holes), LH (light holes) and SO (spin orbit) bands. Under a uniaxial stress along the [100] direction, these valence bands are shifted with respect to their weighted average as [29, 90]

$$\begin{aligned} \Delta E_{v,2} &= \frac{1}{3}\Delta_0 - \frac{1}{2}\delta E_{100}, \\ \Delta E_{v,1} &= -\frac{1}{6}\Delta_0 + \frac{1}{4}\delta E_{100} + \frac{1}{2}[\Delta_0{}^2 + \Delta_0\delta E_{100} + \frac{9}{4}(\delta E_{100})^2]^{1/2}, \\ \Delta E_{v,3} &= -\frac{1}{6}\Delta_0 + \frac{1}{4}\delta E_{100} - \frac{1}{2}[\Delta_0{}^2 + \Delta_0\delta E_{100} + \frac{9}{4}(\delta E_{100})^2]^{1/2}, \end{aligned} \tag{3.54}$$

  where $\Delta_0$ is the spin-orbit splitting for the alloy $\mathrm{Si_{1-x}Ge_x}$ and $\delta E_{100}$ is given by

$$\delta E_{100} = 2b(\varepsilon_{xx} - \varepsilon_{zz}). \tag{3.55}$$

  In eq. (3.55), $b$ is a deformation potential (Tab. 3.3). The valence band offset $\Delta E_v$ of the Si/SiGe heterostructure for this case is shown in Fig. 3.10.

- **Stress along [111] direction:**

  When the stress is along the [111] direction, the strain tensor for the active layer is

$$\hat{\varepsilon}_{[111]} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{pmatrix}, \tag{3.56}$$

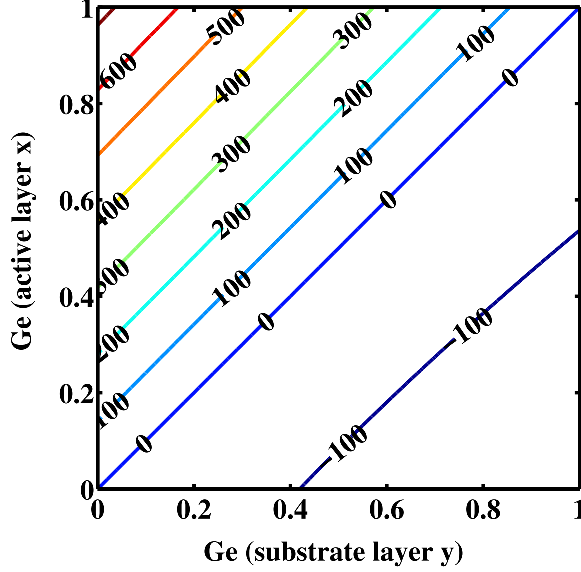


Figure 3.10: Contour plots for the minimum valence band offsets (in meV) of $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ interfaces with the [100] growth direction.

where

$$\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{zz} = \frac{4c_{44}}{c_{11} + 2c_{12} + 4c_{44}}\varepsilon||,$$
$$\varepsilon_{xy} = \varepsilon_{yz} = \varepsilon_{xz} = \frac{c_{11} + 2c_{12}}{c_{11} + 2c_{12} + 4c_{44}}\varepsilon||. \quad (3.57)$$

The analysis for the valence band splitting in this case is almost identical to that for the stress along the [100] direction. Therefore, the shift of the valence bands is [29, 90]

$$\Delta E_{v,2} = \frac{1}{3}\Delta_0 - \frac{1}{2}\delta E_{111},$$
$$\Delta E_{v,1} = -\frac{1}{6}\Delta_0 + \frac{1}{4}\delta E_{111} + \frac{1}{2}[{\Delta_0}^2 + \Delta_0\delta E_{111} + \frac{9}{4}(\delta E_{111})^2]^{1/2},$$
$$\Delta E_{v,3} = -\frac{1}{6}\Delta_0 + \frac{1}{4}\delta E_{111} - \frac{1}{2}[{\Delta_0}^2 + \Delta_0\delta E_{111} + \frac{9}{4}(\delta E_{111})^2]^{1/2}, \quad (3.58)$$

where

$$\delta E_{111} = 2\sqrt{3}d\varepsilon_{xy}. \quad (3.59)$$

$d$ in eq. (3.59) is a deformation potential (Tab. 3.3).

- **Stress along [110] direction:**

When the stress is along the [110] direction, the strain tensor for the active layer is

$$\hat{\varepsilon}_{[110]} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ \varepsilon_{xy} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}, \tag{3.60}$$

where

$$\begin{aligned} \varepsilon_{xx} &= \varepsilon_{yy} = \frac{2c_{44} - c_{12}}{c_{11} + c_{12} + 2c_{44}} \varepsilon||, \\ \varepsilon_{xy} &= -\frac{c_{11} + 2c_{12}}{c_{11} + c_{12} + 2c_{44}} \varepsilon||, \\ \varepsilon_{zz} &= \varepsilon_{||}. \end{aligned} \tag{3.61}$$

In this case, no exact expressions can be written down for the energy levels; the splitting due to strain is a consequence of a mixture of quantities $\delta E_{100}$ and $\delta E_{111}$. Then the splitting of the valence bands is approximated as [90]

$$\begin{aligned} \Delta E_{v,2} &\approx \frac{1}{3}\Delta_0 - \frac{1}{4}[(\delta E_{100})^2 + 3(\delta E_{111})^2]^{1/2} + \frac{3}{32\Delta_0}(\delta E_{100} - \delta E_{111})^2, \\ \Delta E_{v,1} &\approx \frac{1}{3}\Delta_0 + \frac{1}{4}[(\delta E_{100})^2 + 3(\delta E_{111})^2]^{1/2} + \frac{1}{4\Delta_0}(\delta E_{100} + 3\delta E_{111})^2, \\ \Delta E_{v,3} &\approx -\frac{2}{3}\Delta_0 - \frac{1}{4\Delta_0}(\delta E_{100} + 3\delta E_{111})^2 - \frac{3}{32\Delta_0}(\delta E_{100} - \delta E_{111})^2, \end{aligned} \tag{3.62}$$

where $\delta E_{100}$ and $\delta E_{111}$ are determined by

$$\delta E_{100} = 4b(\varepsilon_{xx} - \varepsilon_{zz}) \tag{3.63}$$

and

$$\delta E_{111} = 2\sqrt{3} d \varepsilon_{xy}. \tag{3.64}$$

Up to now the valence band shifting for the strained layer has been determined. For the relaxed substrate $Si_{1-y}Ge_y$, if the spin-orbit splitting is also taken into account, the valence band is always shifted by $\frac{1}{3}\Delta_0(y)$, where $\Delta_0(y)$ is the spin-orbit splitting for this $Si_{1-y}Ge_y$ alloy. Therefore, the offset between the highest valence band edges of the heterostructure is calculated by

$$\Delta E_v = \Delta E_v^{av} + max(\Delta E_{v,1}, \Delta E_{v,2}, \Delta E_{v,3}) - \frac{1}{3}\Delta_0(y). \tag{3.65}$$

**Conduction bands** A similar approach can be applied to get the band offsets for the conduction bands of the heterostructure. From the average valence band offset $\Delta E_v^{av}$, the average conduction band offset can be calculated by [29, 102]

$$\Delta E_c^{av} = (E_g(x) - E_g(y)) + \Delta E_v^{av} - (\Xi_d + \frac{1}{3}\Xi_u - a)(2\varepsilon_{||} + \varepsilon_{\perp}). \tag{3.66}$$

In eq. (3.66), $E_g(x)$ and $E_g(y)$ are the band gap of the relaxed bulk $\mathrm{Si_{1-x}Ge_x}$ and $\mathrm{Si_{1-y}Ge_y}$, respectively. The component $(\Xi_d + \frac{1}{3}\Xi_u - a)$ is the corrected factor between the conduction band deformation potentials $\Xi_d$, $\Xi_u$ and the valence band deformation potential $a$. It is noted that $(\Xi_d + \frac{1}{3}\Xi_u - a)$ is different for the $\Delta$ minimum and $L$ minimum. Hence, either $\Delta E_{c,\Delta}^{av}$ or $\Delta E_{c,L}^{av}$ should be calculated. For the energy splitting in the conduction bands, the minima at $\Delta$ and at $L$ should be considered as well[1].

- **Minima splitting at $\Delta$:** Different strain directions are considered:
  - Uniaxial strain along [111] direction: Under this strain condition, the minima at $\Delta$ are left degenerate.
  - Uniaxial strain along [100] or [110] direction: The minima under this condition are no longer equivalent. The bands along [100] and [010] will split off from the one along [001]. Therefore, we have the following results [108]

$$\begin{aligned} \Delta E_c^{[100]} &= \frac{2}{3}\Xi_u^{\Delta}\varepsilon_{xx} - \frac{1}{3}\Xi_u^{\Delta}(\varepsilon_{yy} + \varepsilon_{zz}), \\ \Delta E_c^{[010]} &= \frac{2}{3}\Xi_u^{\Delta}\varepsilon_{yy} - \frac{1}{3}\Xi_u^{\Delta}(\varepsilon_{xx} + \varepsilon_{zz}), \\ \Delta E_c^{[001]} &= \frac{2}{3}\Xi_u^{\Delta}\varepsilon_{zz} - \frac{1}{3}\Xi_u^{\Delta}(\varepsilon_{xx} + \varepsilon_{yy}). \end{aligned} \tag{3.67}$$

    The superscript $\Delta$ on $\Xi_u$ indicates the conduction-band valley considered.

- **Minima splitting at $L$:** Different strain directions are considered:
  - Uniaxial strain along [100] direction: Under this strain condition, the minima at $L$ are left degenerate.

[1] When the band structures are calculated with the EPM method, the conduction band minima are fixed at either the $\Delta$ or $L$ valleys depending on the Ge fraction.

– Uniaxial strain along [110] and [111] direction: The minima are split under these strain conditions as follows [108]

$$\Delta E_c^{[111]} = \frac{2}{3}\Xi_u^L(\varepsilon_{xy} + \varepsilon_{xz} + \varepsilon_{yz}),$$
$$\Delta E_c^{[\bar{1}11]} = \frac{2}{3}\Xi_u^L(-\varepsilon_{xy} - \varepsilon_{xz} + \varepsilon_{yz}),$$
$$\Delta E_c^{[\bar{1}\bar{1}1]} = \frac{2}{3}\Xi_u^L(\varepsilon_{xy} - \varepsilon_{xz} - \varepsilon_{yz}),$$
$$\Delta E_c^{[1\bar{1}1]} = \frac{2}{3}\Xi_u^L(-\varepsilon_{xy} + \varepsilon_{xz} - \varepsilon_{yz}). \quad (3.68)$$

Depending on the composition $x$ and $y$, we will have the conduction band offset either at the $\Delta$ or at the $L$ minimum of the active layer. Moreover, all directions should be investigated to obtain the minimum of the conduction bands. This means

$$\Delta E_{c,\Delta} = \Delta E_{c,\Delta}^{av} + min(\Delta E_c^{[100]}, \Delta E_c^{[010]}, \Delta E_c^{[001]})$$
$$\Delta E_{c,L} = \Delta E_{c,L}^{av} + min(\Delta E_c^{[111]}, \Delta E_c^{[\bar{1}11]}, \Delta E_c^{[\bar{1}\bar{1}1]}, \Delta E_c^{[1\bar{1}1]}). \quad (3.69)$$

The conduction band offsets $\Delta E_c$ of the $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ heterostructure with [100] growth direction are shown in Fig. 3.11.

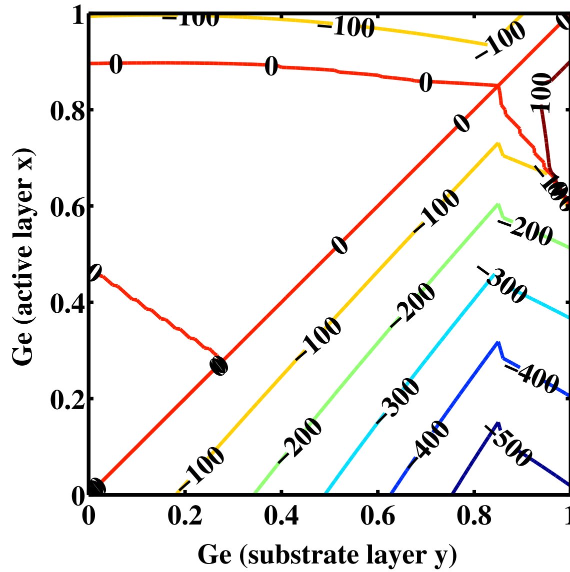


Figure 3.11: Contour plots for the minimum conduction band offsets (in meV) of $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ interfaces with the [100] growth direction.

**Band gap**

The band gap (in eV) of a relaxed $Si_{1-y}Ge_y$ alloy at the temperature of 4.2K is determined as [125]

$$E_g(y) = \begin{cases} 1.155 - 0.43y + 0.206y^2 & \text{if } y \leq 0.85 \\ 2.010 - 1.270y & \text{if } y > 0.85. \end{cases} \quad (3.70)$$

To estimate the band gap of the $Si_{1-y}Ge_y$ alloy at higher temperatures, the well-known relationship of the temperature on the band gap of Si can be applied as [111]

$$E_g(y, T) = E_g(y, 0K) - \frac{4.73 \times 10^{-4} T^2}{T + 636}, \quad (3.71)$$

where $T$ is the temperature in absolute degree $K$.

From eqs. (3.70) and (3.71) and using 1.1241 eV for the band gap of Si and 0.66 eV for Ge, the band gap for relaxed $Si_{1-y}Ge_y$ at 300K is derived

$$E_g(y) = \begin{cases} 1.1241 - 0.43y + 0.206y^2 & \text{if } y \leq 0.85 \\ 1.934 - 1.270y & \text{if } y > 0.85 \end{cases} . \quad (3.72)$$

The band gap of a relaxed SiGe alloy is shown in Fig. 3.12.

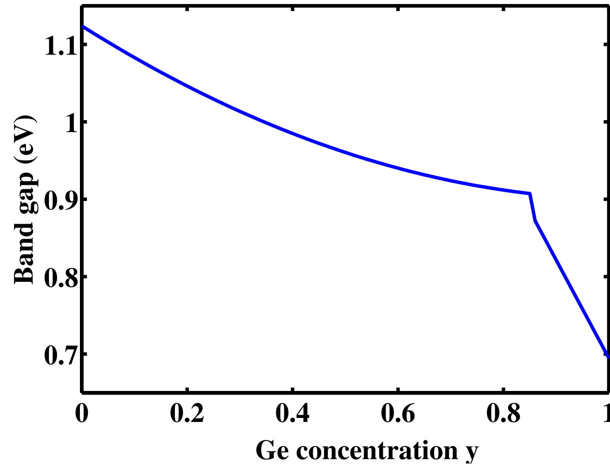


Figure 3.12: Band gap (in eV) of relaxed $Si_{1-y}Ge_y$ alloys.

With the $\Delta E_c$ and $\Delta E_v$ calculated in the previous part, we can get the band gap of an active layer of $Si_{1-x}Ge_x$ by

$$E_g(\text{active}) = E_g(x, y) = E_g(y) + \Delta E_c - \Delta E_v. \quad (3.73)$$

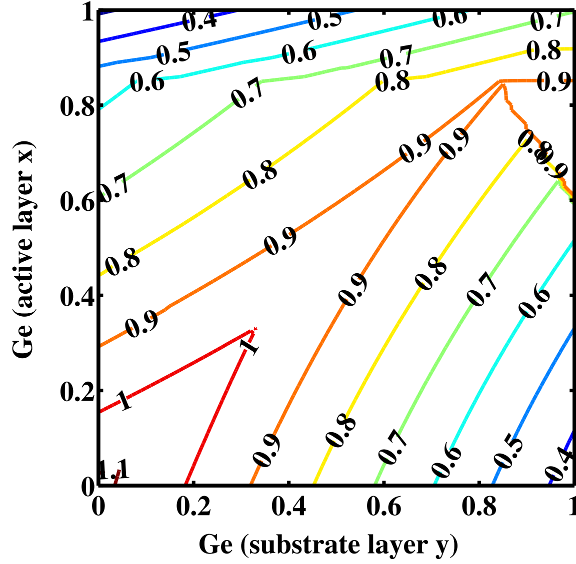


Figure 3.13: Contour plots for the band gaps (in eV) of $Si_{1-x}Ge_x$ alloys (active layer) grown pseudomorphically along the [100] direction on unstrained $Si_{1-y}Ge_y$ alloys (substrate layer).

The band gap of $Si_{1-y}Ge_y$ heterostructures with [100] growth direction is shown in Fig. 3.13.

All parameters, including deformation potentials and spin-orbit splitting $\Delta_0$, for calculating the strain effects on band structures are listed in Tab. 3.3. The values in this table are for pure Si and Ge. Linear interpolation is used to get the parameters for a SiGe alloy.

### 3.4.2 With additional uniaxial stress

In this part, the band offsets and band gap for heterostructures under an additional uniaxial stress are investigated. In fact, band offsets and band gap for this case can be calculated similarly to the case without additional stress as presented above. The only parameter still missing is the strain tensor for the active layer. Therefore, in this part, the strain tensor for each case is calculated. Assuming that the growth is along the [100] direction, uniaxial stress may be applied along either the [010] or the [011] direction. The following part will proceed with these directions.

The stress is applied to the whole device to assure the same strain level. It means that the lattice constants along the parallel direction (in-plane) are completely matched between the active and substrate layer within the whole

Table 3.3: Deformation potentials of the valence and conduction bands for different valleys in Si and Ge (in eV), which were derived from self-consistent calculations based on a local density functional and *ab initio* pseudopotentials [29]. Spin-orbit splitting $\Delta_o$ is taken from [81] and band gap from [67].

| Parameters (300K) | Si | Ge |
|---|---|---|
| Band gap Eg (eV) | 1.12 | 0.66 |
| Lattice constant (Å) | 5.43 | 5.65 |
| $\Delta_0$ (eV) | 0.044 | 0.2296 |
| $\Xi_d + \frac{1}{3}\Xi_u - a$ ($\Delta$ minimum) | 1.72 | 1.31 |
| $\Xi_d + \frac{1}{3}\Xi_u - a$ (L minimum) | -3.12 | -2.78 |
| b | -2.35 | -2.55 |
| d | -5.32 | -5.50 |
| $\Xi_u^\Delta$ | 9.16 | 9.42 |
| $\Xi_u^L$ | 16.14 | 15.13 |

device.

### Uniaxial stress along [010] direction

When the stress force is $P$ (GPa), the strain tensor will take the form of $\hat{\varepsilon}_{[100]}$ of eq. (3.52). The strain elements can be determined from eq. (3.14) and the requirement of lattice constants matching between layers (active and substrate) is satisfied through the following relations

$$\begin{aligned}\varepsilon_{yy} &= (1 + s_{11}P)\frac{a_0(y)}{a_0(x)} - 1,\\ \varepsilon_{zz} &= (1 + s_{12}P)\frac{a_0(y)}{a_0(x)} - 1,\end{aligned} \tag{3.74}$$

where $s_{11}$ and $s_{12}$ are stiffness constants (Tab. 3.1).

In order to determine $\varepsilon_{xx}$, we take the fact that there is no stress in [100] direction. This means $\sigma_{11} = 0$. Following eq. (3.13), we have

$$c_{11}\varepsilon_{xx} + c_{12}(\varepsilon_{yy} + \varepsilon_{zz}) = 0. \tag{3.75}$$

Therefore, $\varepsilon_{xx}$ is

$$\varepsilon_{xx} = -\frac{c_{12}}{c_{11}}(\varepsilon_{yy} + \varepsilon_{zz}). \tag{3.76}$$

The band gap for the case with a uniaxial stress of -1.0 GPa (compressive) along the [010] direction is shown in Fig. 3.14.

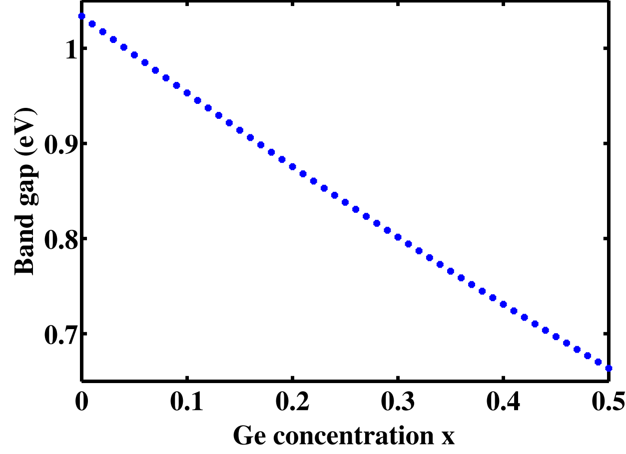


Figure 3.14: Band gap of the active layer under a uniaxial stress of -1.0 GPa along the [010] direction.

### Uniaxial stress along [011] direction

Under this stress condition, the strain tensor take the form of $\hat{\varepsilon}_{[110]}$ (eq. (3.60)); however with a different orientation

$$\hat{\varepsilon}_{[011]} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & \varepsilon_{yz} \\ 0 & \varepsilon_{yz} & \varepsilon_{zz} \end{pmatrix}. \tag{3.77}$$

The elements of this tensor can be calculated through stiffness constants $s_{11}, s_{12}, s_{44}$ similarly to the case of uniaxial stress along the [010] direction by

$$\begin{aligned} \varepsilon_{yy} &= \left(1 + (s_{11} + s_{12})\frac{P}{2}\right)\frac{a_0(y)}{a_0(x)} - 1, \\ \varepsilon_{zz} &= \varepsilon_{yy}, \\ \varepsilon_{yz} &= s_{44}\frac{Pa_0(y)}{4a_0(x)}. \end{aligned} \tag{3.78}$$

$\varepsilon_{xx}$ is calculated by $\varepsilon_{yy}$ and $\varepsilon_{zz}$ according to eq. (3.76), similarly to the case for stress along the [010] direction. The band gap for the case with a uniaxial stress of -1.0 GPa (compressive) along the [011] direction is shown in Fig. 3.15.

The above calculations are applied for two popular stress directions. The calculation for an arbitrary direction can be derived similarly.

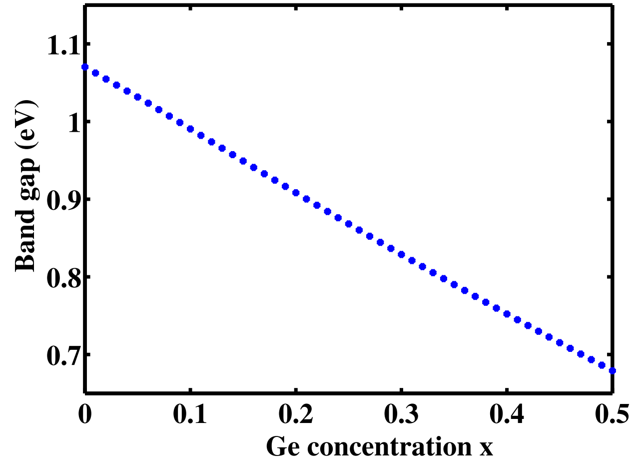


Figure 3.15: Band gap of the active layer under a uniaxial stress of -1.0 GPa along the [011] direction.

## 3.5 Summary

In this section, the stress-strain dependence is discussed. From any stress, we can calculate a strain tensor, from which the band structure of the material can be determined. Moreover, a detailed description of band structures and properties of uniaxially and biaxially strained Si and SiGe is also presented.

An analytical way for calculating the band gap of strained Si/SiGe and band edges of heterostructures under general strain conditions is also discussed. These band edges are necessary for simulating heterojunction bipolar transistors. The band gap calculated by this analytical method has been proven to be closer to the experimental data than the one calculated by EPM. Such an accurate result is very important for reliably calculating the impact ionization rates for strained Si and SiGe, which will be discussed in the next chapter.

# Chapter 4

# Impact ionization for strained materials

## 4.1 Theory of impact ionization

Impact ionization (II) is a special type of electron-electron interaction that takes place in solid materials with a band gap [117]. The interaction can be initiated by an electron or a hole. The former process is called electron-initiated impact ionization, whereas the latter one is called hole-initiated impact ionization.

Electron-initiated impact ionization happens when a primary electron (state 1) in the conduction band interacts with another electron (state 4) in the valence band via the Coulomb force and both electrons end up in the conduction bands (state 2 and state 3). This process is illustrated in Fig. 4.1. The notation *1,2,3,4* for these states are used consistently throughout this chapter. The electron in state $i$ has the wave vector $\mathbf{k}_i$, energy $E_i$ and belongs to band $n_i$.

In hole-initiated impact ionization, one hole in the valence band, after getting enough energy, interacts with another hole in the conduction band to create two new holes in the valence bands. In the scope of this dissertation, only the II process for electrons is addressed.

The band structures for the investigated materials, which include relaxed Si, strained Si and strained SiGe, are calculated by the empirical nonlocal pseudopotential method, which is presented in Sec. 3.3.3. This calculation is shown to result in realistic band structures and a reliable description of the excitation spectrum [20, 24, 93].

The ionization process happens under energy and momentum conserva-

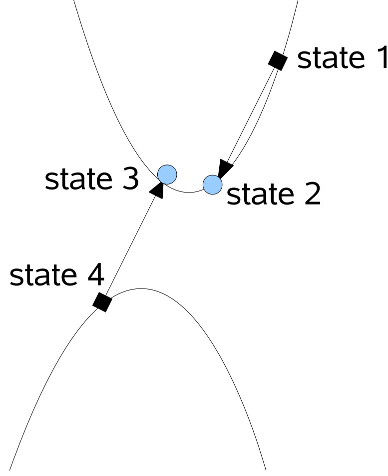


Figure 4.1: Diagram of the electron-initiated impact ionization process.

tion

$$E_1 + E_4 = E_2 + E_3 \tag{4.1}$$

and

$$\mathbf{k}_1 + \mathbf{k}_4 + \mathbf{G}_0 = \mathbf{k}_2 + \mathbf{k}_3, \tag{4.2}$$

where $\mathbf{G}_0$ is the reciprocal lattice vector which returns all wave vectors $\mathbf{k}_i$ to the first Brillouin zone. This process is called *Umklapp* process.

The Coulomb potential $V(\mathbf{r}-\mathbf{r}')$ causing the electron-electron interaction is presented by the following well-known equation

$$V(\mathbf{r}-\mathbf{r}') = \frac{e^2}{4\pi\epsilon}\frac{1}{|\mathbf{r}-\mathbf{r}'|}, \tag{4.3}$$

where $e$ is the elementary charge, $\mathbf{r}$ and $\mathbf{r}'$ are the space vectors for two interacting electrons and $\epsilon$ is the dielectric constant.

However, it is pointed out in [77] that in order to have a proper calculation of the scattering matrix elements in semiconductors, the dielectric constant should depend on the wave vectors. A dielectric function $\epsilon(\mathbf{q},\omega)$, where $\mathbf{q}$ is the momentum exchange and $\hbar\omega$ the energy transfer between two electrons during the collision, should replace the constant $\epsilon$. Therefore, the Coulomb potential in eq. (4.3), after taking the electron screening factor $\lambda$ and $\epsilon(\mathbf{q},\omega)$ into account, can be rewritten by

$$V(\mathbf{r}-\mathbf{r}') = \frac{e^2}{\epsilon(\mathbf{q},\omega)}\frac{e^{-\lambda|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}. \tag{4.4}$$

Since impact ionization involves two-particle states, the matrix element for this process consists of direct and exchange terms. When summed over

all spins except for the initial state spin, the matrix element $T_{II}$ for impact ionization is written by

$$T_{II} = (|\mathbf{M}_D|^2 + |\mathbf{M}_E|^2 + |\mathbf{M}_D - \mathbf{M}_E|^2), \tag{4.5}$$

where $\mathbf{M}_D$ and $\mathbf{M}_E$ are the matrix elements of the direct and exchange processes, respectively. They are calculated by

$$\begin{aligned}
\mathbf{M}_D &= \int\int d^3r_1 d^3r_2 \psi_1^*(\mathbf{r}_1)\psi_4^*(\mathbf{r}_2)\mathbf{V}(\mathbf{r_1}-\mathbf{r_2})\psi_2(\mathbf{r}_1)\psi_3(\mathbf{r}_2) \\
\mathbf{M}_E &= \int\int d^3r_1 d^3r_2 \psi_1^*(\mathbf{r}_1)\psi_4^*(\mathbf{r}_2)\mathbf{V}(\mathbf{r_1}-\mathbf{r_2})\psi_3(\mathbf{r}_1)\psi_2(\mathbf{r}_2),
\end{aligned} \tag{4.6}$$

where $\psi(\mathbf{r}_i)$ is the wave function for the electron of state $i$ at position $\mathbf{r}_i$. The form of this wave function for electron $i$ can be written as

$$\psi_i(\mathbf{r}_i) = \frac{1}{\sqrt{\Omega}}\sum_{\mathbf{G}} c(n_i, \mathbf{k}_i; \mathbf{G})e^{i(\mathbf{k}_i+\mathbf{G})\mathbf{r}_i}. \tag{4.7}$$

Since $\mathbf{M}_D$ and $\mathbf{M}_E$ are similar, only the calculation of $\mathrm{M}_D$ is presented here.

Taking the Coulomb potential from eq. (4.4) and wave functions from eq. (4.7), eq. (4.6) can be written as

$$\begin{aligned}
\mathbf{M}_D = \frac{1}{\Omega^2}\sum_{G_1,G_2,G_3} & c^*(n_1,\mathbf{k}_1;\mathbf{G_1})c^*(n_4,\mathbf{k}_4;\mathbf{G_4})c(n_2,\mathbf{k}_2;\mathbf{G_2})c(n_3,\mathbf{k}_3;\mathbf{G_3})\times \\
& \int\int d^3r_1 d^3r_2 e^{-i\mathbf{q}_D(\mathbf{r}_1-\mathbf{r}_2)}\frac{e^2}{\epsilon(\mathbf{q}_D,\omega_D)|\mathbf{r}_1-\mathbf{r}_2|},
\end{aligned} \tag{4.8}$$

where $\mathbf{q}_D = \mathbf{k}_1 + \mathbf{G}_1 - \mathbf{k}_2 - \mathbf{G}_2$ and $\omega_D = (E_1 - E_2)/\hbar$. The above derivation has used momentum conservation from eq. (4.2). Reciprocal lattice vector $\mathbf{G}_4$ is also calculated by momentum conservation as

$$\mathbf{G}_1 + \mathbf{G}_4 + \mathbf{G}_0 = \mathbf{G}_2 + \mathbf{G}_3. \tag{4.9}$$

To determine the integral in eq. (4.8), first the following sample integral should be calculated

$$M = \int d^3r' e^{-i\mathbf{q}\mathbf{r}'}\frac{e^{-\lambda r'}}{r'}. \tag{4.10}$$

This integral can be calculated by using spherical coordinates. Since the potential has spherical symmetry, we can choose $\mathbf{q}$ in the $z$ direction and

proceed with the integral by

$$
\begin{aligned}
M &= 2\pi \int_0^\infty r'^2 dr \int_{-1}^{1} d(cos\theta) e^{-i\mathbf{q}r'cos\theta} e^{-\lambda r'} \\
&= \frac{2\pi}{iq} \int_0^\infty (e^{-i\mathbf{q}r} - e^{i\mathbf{q}r}) e^{-\lambda r} dr \\
&= \frac{4\pi}{(\mathbf{q}^2 \ + \lambda^2)}.
\end{aligned} \tag{4.11}
$$

Applying the result of eq. (4.11) to eq. (4.8), we can get the final form as [66]

$$
\mathbf{M}_D = \frac{4\pi e^2}{\Omega} \sum_{G_1,G_2,G_3} \frac{c^*(n_1,\mathbf{k}_1;\mathbf{G_1})c^*(n_4,\mathbf{k}_4;\mathbf{G_4})c(n_2,\mathbf{k}_2;\mathbf{G_2})c(n_3,\mathbf{k}_3;\mathbf{G_3})}{\epsilon(\mathbf{q}_D,\omega_D)(\mathbf{q}_D^2 + \lambda^2)}. \tag{4.12}
$$

Similarly, the matrix element $\mathbf{M}_E$ is calculated by

$$
\mathbf{M}_E = \frac{4\pi e^2}{\Omega} \sum_{G_1,G_2,G_3} \frac{c^*(n_1,\mathbf{k}_1;\mathbf{G_1})c^*(n_4,\mathbf{k}_4;\mathbf{G_4})c(n_3,\mathbf{k}_3;\mathbf{G_3})c(n_2,\mathbf{k}_2;\mathbf{G_2})}{\epsilon(\mathbf{q}_E,\omega_E)(\mathbf{q}_E^2 + \lambda^2)}, \tag{4.13}
$$

where $\mathbf{q}_E = \mathbf{k}_1 + \mathbf{G}_1 - \mathbf{k}_3 - \mathbf{G}_3$ and $\omega_E = (E_1 - E_3)/\hbar$.

## 4.2 Calculation approach for impact ionization rate

The ionization probability per unit time for an initial electron is calculated by Fermi's golden rule [99] as follows

$$
S^{\mathrm{II}}(n_1,\mathbf{k}_1) = \frac{2\pi}{\hbar}\frac{\Omega^2}{(2\pi)^6} \sum_{n_2,n_3,n_4} \underbrace{\int d^3k_4}_{I_1} \underbrace{\int d^3k_2 T_{\mathrm{II}} \delta(E_1 + E_4 - E_2 - E_3)}_{I_2}, \tag{4.14}
$$

where $T_{II}$ is the matrix element defined in eq. (4.5), $\Omega$ is the volume of the crystal and $\hbar$ is Planck's constant divided by $2\pi$. There are three main approaches for calculating the II rates [39]: 1. *ab initio*, which accounts for energy and momentum conservation as well as the Coulomb matrix element dependence on both initial and final states. 2. Constant matrix element (CME) approximation, which uses a constant Coulomb matrix element. 3. *random-k* approximation, which relaxes momentum conservation.

It has been shown in [39, 99, 100] that the random-k approximation is not a reliable method for calculating II rates for arbitrary materials, whereas the *ab initio* approach is too complicated and CPU-time consuming. It was also proved in [99] that CME approximation is a good approach for II rate calculation.

In this dissertation, the CME approximation is used. The matrix element $T_{II}$ is treated here as a constant fitting parameter. Momentum and energy conservation are strictly kept during the calculation process. The integral in eq. (4.14) is performed over the whole Brillouin zone (BZ) including four conduction bands and three valence bands.

### 4.2.1 BZ setup

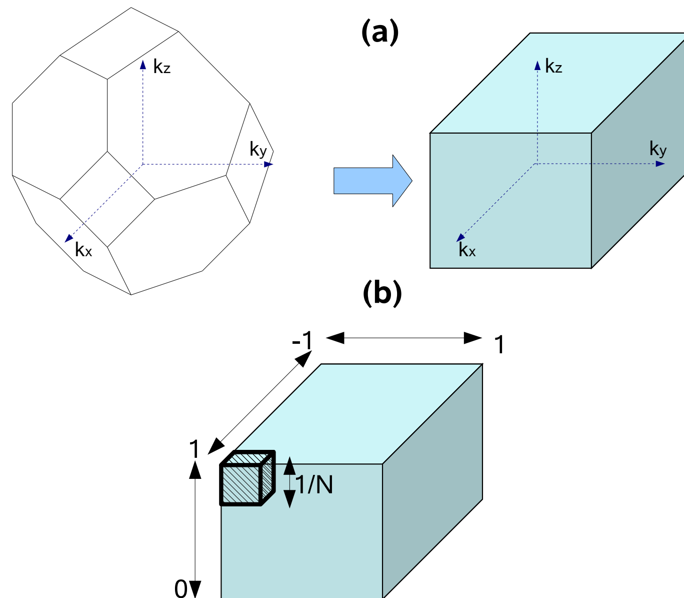


Figure 4.2: (a) Transformation from a normal BZ into a cuboid BZ (b) Break up of cuboid BZ into small cubes.

In order to calculate the six-dimensional integral in eq. (4.14), a cuboid BZ is set up. By adding reciprocal lattice vectors, we can move the lower parts $k_z \leq 0$ of the BZ and attach them to the BZ's upper part. $\mathbf{k}$ vectors in the lower part are transformed into $\mathbf{k}'$ vectors as follows

$$\mathbf{k}' = \mathbf{k} + \begin{cases} 2\pi/a(-1;-1;1)^T & \text{if } k_x \geq 0 \text{ and } k_y \geq 0 \\ 2\pi/a(-1;1;1)^T & \text{if } k_x \geq 0 \text{ and } k_y < 0 \\ 2\pi/a(1;-1;1)^T & \text{if } k_x < 0 \text{ and } k_y \geq 0 \\ 2\pi/a(1;1;1)^T & \text{if } k_x < 0 \text{ and } k_y < 0 \end{cases}, \tag{4.15}$$

where $a$ is the lattice constant. The BZ is now transformed into a cuboid with $k_x \in [-(2\pi/a), (2\pi/a)], k_y \in [-(2\pi/a), (2\pi/a)]$ and $k_z \in [0, (2\pi/a)]$. This transformation is illustrated in Fig. 4.2(a). The cuboid BZ is then divided equally into small cubes with an edge length of $\Delta k = \frac{1}{N}(2\pi/a)$[1]. This division is illustrated in Fig. 4.2(b). $N$ decides the resolution of the cubic mesh. The wave vectors of the initial electrons (state 1 in a conduction band and state 4 in a valence band) are given by triplets $(k_x, k_y, k_z)$ on the nodes in this mesh. The II rate for all these nodes will be calculated. The integral over $\mathbf{k}$ space in eq. (4.14) is determined by running over the entire cuboid BZ.

### 4.2.2 Calculation of the integral $I_2$

For a given pair $(\mathbf{k}_1, E_1)$ and $(\mathbf{k}_4, E_4)$, at each state $(\mathbf{k}_2, E_2)$ we can determine $\mathbf{k}_3$ and $E_3$ through eqs. (4.1) and (4.2). Therefore, the integral $I_2$ from eq. (4.14), after taking out the constant $T_{II}$, can be calculated similarly to the density of states at energy $\varepsilon$ as follows

$$\begin{aligned} I_2 &= \int \delta(\varepsilon - \bar{\varepsilon}(\mathbf{k}_2)) d^3k_2 \\ &= \frac{1}{\hbar} \int_{A(\varepsilon)} \frac{d^2k_2}{|\mathbf{v}(\mathbf{k}_2)|}, \end{aligned} \tag{4.16}$$

where $\varepsilon = E_1 + E_4$, $\bar{\varepsilon} = E_2 + E_3$, $A(\varepsilon)$ is the equienergy surface for the given energy and $\mathbf{v}(\mathbf{k}_2)$ is the velocity. The energies $\bar{\varepsilon}$ are calculated and stored for all $\mathbf{k}_2$ points in the BZ.

In order to calculate the integral $I_2$ in eq. (4.16), each small cube in the BZ is divided into 6 similar tetrahedra [10] as in Fig. 4.3. Within a tetrahedron, the velocity is considered to be constant (named $\mathbf{v}^{tet}$) and the energy is linearly interpolated; hence the equienergy surface is a plane. Therefore, if energy $\varepsilon$ is inside the energy interval of the tetrahedron, this tetrahedron will

[1] In the case of strained materials, where the lattice constants along the directions of the Cartesian coordinates can be different, $a$ refers to the actual lattice constant along the specific direction. Hence $\Delta k$ can be different along $k_x, k_y, k_z$ directions.

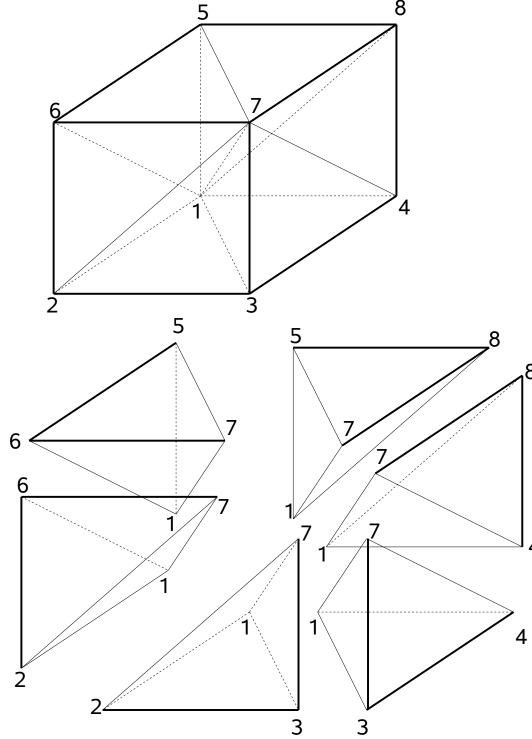

Figure 4.3: Break up a cube into 6 similar tetrahedra.

contribute to $I_2$ by $D^{tet}$ as follows [59]

$$I_2 = \sum_{BZ} D^{tet}(\varepsilon) = \sum_{BZ} \frac{1}{\hbar} \frac{A^{tet}(\varepsilon)}{|\mathbf{v}^{tet}|}, \tag{4.17}$$

where $A^{tet}(\varepsilon)$ is the area of the intersection of the equienergy surface $\varepsilon$ and the tetrahedron $i$.

As the energy is linearly interpolated within a tetrahedron, the energies at the four nodes of the tetrahedron are given by the following relations [59]

$$\begin{aligned} \varepsilon_2^{nod} &= \varepsilon_1^{nod} + (\mathbf{v}^{tet})^T \hbar (\mathbf{k}_2^{nod} - \mathbf{k}_1^{nod}) \\ \varepsilon_3^{nod} &= \varepsilon_1^{nod} + (\mathbf{v}^{tet})^T \hbar (\mathbf{k}_3^{nod} - \mathbf{k}_1^{nod}) \\ \varepsilon_4^{nod} &= \varepsilon_1^{nod} + (\mathbf{v}^{tet})^T \hbar (\mathbf{k}_4^{nod} - \mathbf{k}_1^{nod}), \end{aligned} \tag{4.18}$$

where $\mathbf{k}_i^{nod}$ and $\varepsilon_i^{nod}$ with $i = 1, 2, 3, 4$ are the wave vectors and energies of the four nodes of the tetrahedron. From eq. (4.18), the velocity $\mathbf{v}^{tet}$ can be determined by

$$\mathbf{v}^{tet} = \frac{1}{\hbar} (\mathbf{k}_2^{nod} - \mathbf{k}_1^{nod}, \mathbf{k}_3^{nod} - \mathbf{k}_1^{nod}, \mathbf{k}_4^{nod} - \mathbf{k}_1^{nod})^{-1} \begin{pmatrix} \varepsilon_2^{nod} - \varepsilon_1^{nod} \\ \varepsilon_3^{nod} - \varepsilon_1^{nod} \\ \varepsilon_4^{nod} - \varepsilon_1^{nod} \end{pmatrix}. \tag{4.19}$$

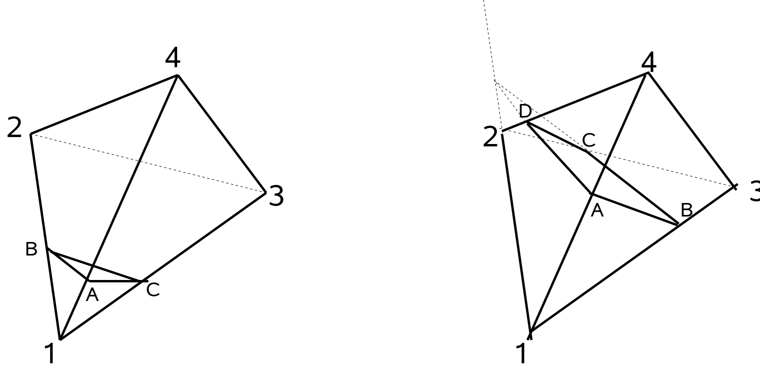


Figure 4.4: Intersection area of an equienergy plane with a tetrahedron: ABC or ABCD.

The intersection area between the equienergy surface $\varepsilon$ and the tetrahedron is illustrated in Fig. 4.4. The tetrahedron's corners are arranged in the order of ascending energies, i.e. node 1 has the smallest energy and node 4 has the highest energy. The energies and wave vectors for these nodes are denoted as $\varepsilon_j^{nod}$ and $\mathbf{k}_j^{nod}$ with $j = 1, 2, 3, 4$. If $\varepsilon$ is between $\varepsilon_1^{nod}$ and $\varepsilon_2^{nod}$ or $\varepsilon_3^{nod}$ and $\varepsilon_4^{nod}$, the intersection is a triangle. Otherwise, when $\varepsilon$ is between $\varepsilon_2^{nod}$ and $\varepsilon_3^{nod}$, the intersection is a quadrangle. Similar to the relation between the energy and wave vector at eq. (4.18), the wave vectors of the intersection points A,B,C or A,B,C,D between nodes $j$ and $k$ ($j \neq k$) of the tetrahedron $tet$ are calculated by

$$\mathbf{c}_{j,k}^{tet}(\varepsilon) = \mathbf{k}_j^{nod} + \frac{\varepsilon - \varepsilon_j^{nod}}{\varepsilon_k - \varepsilon_j} \left(\mathbf{k}_k^{nod} - \mathbf{k}_j^{nod}\right). \tag{4.20}$$

When the wave vectors for all intersection points are determined, the intersection area $A^{tet}(\varepsilon)$ can be straightforwardly calculated. The area of the triangle ABC in Fig. 4.4 is determined by

$$A^{tet} = \frac{1}{2} \left| \left(\mathbf{c}_{1,2}^{tet} - \mathbf{c}_{1,4}^{tet}\right) \times \left(\mathbf{c}_{1,3}^{tet} - \mathbf{c}_{1,4}^{tet}\right) \right|. \tag{4.21}$$

In the case of a quadrangle, it can be split into 2 triangles, and the calculation process is done similarly.

When $A^{tet}(\varepsilon)$ and $\mathbf{v}^{tet}$ have been calculated, $D^{tet}$ is also determined according to eq. (4.17), hence the calculation of the integral $I_2$ over the BZ can be done.

### Calculation of energies $\varepsilon$ and $\bar{\varepsilon}$

As mentioned in the above section, for a given pair $(\mathbf{k}_1, E_1)$ and $(\mathbf{k}_4, E_4)$, at each state $(\mathbf{k}_2, E_2)$ we can determine $\mathbf{k}_3$ and $E_3$. There is no problem with

$E_2$ as it has been calculated in advance (from band-structure calculation) and stored for every state. However, the $E_3$ which satisfies the momentum conservation has to be determined. As this process must be done for all $\mathbf{k}_2$, $\mathbf{k}_4$ and over all bands, it takes a long time to finish. Therefore, an appropriate approach to deal with this problem should be figured out.

A bigger cuboid BZ, which contains all wave vectors $\mathbf{k}_3$, is setup once at the beginning to store the energy $E_3$ for all nodes. From momentum conservation, this BZ should be defined by

$$\begin{array}{l} -3(2\pi/a) \leq k_x \leq 3(2\pi/a) \\ -3(2\pi/a) \leq k_y \leq 3(2\pi/a) \\ -3(2\pi/a) \leq k_z \leq 3(2\pi/a) \end{array} . \tag{4.22}$$

Up to now, two arrays for storing energies for all $E_2$ and $E_3$, which can correspond to any pair of $(\mathbf{k}_1, E_1)$ and $(\mathbf{k}_4, E_4)$, are available. Next the fastest way for accessing $E_2$ and $E_3$, which helps to reduce the total calculation time, is presented.

It is noticed that, with the given $\mathbf{k}_1$ and $\mathbf{k}_4$, $\mathbf{k}_2$ is different from $\mathbf{k}_3$ by a constant ($\mathbf{k}_{dis} = \mathbf{k}_1 + \mathbf{k}_4$). As the energies of the nodes $\mathbf{k}_2$ and $\mathbf{k}_3$ are stored in 2 arrays with linear indexes, the cache is utilized and the accessing process is fast if these arrays are read in a linear and sequential manner. To make it simple, all $\mathbf{k}$ vectors are converted into integers by a factor of $(a/2\pi)N$. Assuming that when $\mathbf{k}_2$ runs in the cuboid BZ for calculating the integral $I_2$ in eq. (4.17), its component $k_z$ changes fastest, then $k_y$ and $k_x$. The linear index of the arrays containing the energies for $\mathbf{k}_2$ and $\mathbf{k}_3$ increases according to this running manner of $\mathbf{k}_2$. With each pair of $k_y$ and $k_z$, the position indexes $L_2$ and $L_3$ for $E_2$ and $E_3$ can be determined by

$$L_2 = (N+1) + (2N+1)(N+k_y) + (2N+1)^2 k_z. \tag{4.23}$$

$$L_3 = (3N+1) + k_{dis,x} + (6N+1)(3N + k_{dis,y} - k_y) + (6N+1)^2(3N + k_{dis,z} - k_z). \tag{4.24}$$

When $k_x$ runs from $-N$ to $N$, the positions of $E_2$ and $E_3$ in their energy arrays can be accessed directly by indexes $(L_2+k_x)$ and $(L_3-k_x)$, respectively. By using this approach, two energy arrays are read consequentially in the increasing or decreasing order. It is very efficient for cache access and makes the calculation much faster.

### Strained Si and strained SiGe

The wave vector $\mathbf{k}$ in a strained material is related with the wave vector $\mathbf{k}^*$ in relaxed Si by

$$\mathbf{k} = \hat{\mathbf{S}}^{-1}\mathbf{k}^*, \tag{4.25}$$

where $\hat{\mathbf{S}}^{-1}$ is the inverse of the strain matrix, which is defined from the strain tensor $\hat{\varepsilon}$ by

$$\hat{\mathbf{S}} = \hat{\varepsilon} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.26}$$

For the case of SiGe, as the lattice constant of SiGe alloy is different from that of Si, the strain matrix should be changed as

$$\hat{\mathbf{S}}_{\text{SiGe}} = \frac{a(\text{SiGe})}{a_0} \hat{\mathbf{S}}, \tag{4.27}$$

where $a_0$ and $a(\text{SiGe})$ are the lattice constants of relaxed Si and SiGe respectively. For the $I_2$ calculation, we have to replace all $\mathbf{k}^*$ vectors by new strained $\mathbf{k}$ vectors in eqs. (4.18), (4.19), and (4.20).

### 4.2.3 Calculation of the integral $I_1$

The integral $I_1$ in eq. (4.14) can be rewritten with the calculated integral $I_2$

$$I_1 = \int d^3k_4 I_2. \tag{4.28}$$

Each node $\mathbf{k}_4$ is considered to be the center of a small cube with an edge length of $\Delta k = \frac{1}{N}(2\pi/a)$ as depicted in Fig. 4.2(b). Integral $I_1$ of eq. (4.28) is calculated by multiplying the value of $I_2$ with the finite volume of this small cube and summing over all $\mathbf{k}_4$ vectors in the entire BZ. When $\mathbf{k}_4$ is on the surface of the cuboid BZ, the volume $V$ is adjusted accordingly as follows

$$V(\mathbf{k}_4) = \begin{cases} (\Delta k)^3 & \text{if } \mathbf{k}_4 \text{ is inside the cuboid BZ} \\ \frac{1}{2} \times (\Delta k)^3 & \text{if } \mathbf{k}_4 \text{ is on the surfaces of the cuboid BZ} \\ \frac{1}{4} \times (\Delta k)^3 & \text{if } \mathbf{k}_4 \text{ is on the edges of the cuboid BZ} \\ \frac{1}{8} \times (\Delta k)^3 & \text{if } \mathbf{k}_4 \text{ is on the corners of the cuboid BZ} \end{cases}. \tag{4.29}$$

#### Strained Si and SiGe

In strained materials, as the wave vector is changed, the volume of the box around $\mathbf{k}_4$ is also changed. If a box created by 3 vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ in which $\mathbf{c}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$, its volume can be calculated by

$$\mathbf{V}_{box} = |\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})| = \det[\mathbf{c}, \mathbf{a}, \mathbf{b}]. \tag{4.30}$$

From this formula and the relation between wave vectors in strained materials and in relaxed Si (eq. (4.25)), we can calculate the volume $\mathbf{V}_{str}$ around the

strained $\mathbf{k}_4$ from its respective volume in relaxed Si as

$$V_{str}(\mathbf{k}_4) = V(\mathbf{k}_4)\det(\hat{\mathbf{S}}^{-1}), \tag{4.31}$$

where $\hat{\mathbf{S}}$ is the strain matrix for the strained material (in eq. (4.26)).

### 4.2.4 Another approach for calculating $I_2$

The tetrahedron method described above was shown to be a rigorous approach for calculating an integral within the BZ [10]. However, the $I_2$ calculation over six tetrahedra within each cube might be time consuming. Therefore another simpler method for calculating $I_2$ is proposed here.

Each small cube with an edge length $\Delta k = \frac{1}{N}(2\pi/a)$ is not divided into six tetrahedra. Instead, it is replaced by a sphere, which has the same volume. Therefore the radius $R^{sph}$ of this sphere is calculated by

$$R^{sph} = \sqrt[3]{\frac{3}{4\pi}}\Delta k. \tag{4.32}$$

Then eq. (4.17) can be rewritten as

$$I_2 = \sum_{BZ} D^{sph}(\varepsilon). \tag{4.33}$$

This sphere is represented by a wave vector at the sphere center, $\mathbf{k}^{sph}$, and a constant velocity, $\mathbf{v}^{sph}$, like the $\mathbf{v}^{tet}$ for a tetrahedron. These parameters can be calculated by the linear interpolation from 8 corners of the cube (in Fig 4.3) as follows

$$\mathbf{k}^{sph} = \frac{1}{8}\sum_{i=1,8} \mathbf{k}_i^{cor}, \tag{4.34}$$

and

$$\begin{aligned}
v_x^{sph} &= \frac{1}{4\Delta k}(\varepsilon_2^{cor} + \varepsilon_3^{cor} + \varepsilon_6^{cor} + \varepsilon_7^{cor} - \varepsilon_1^{cor} - \varepsilon_4^{cor} - \varepsilon_5^{cor} - \varepsilon_8^{cor}),\\
v_y^{sph} &= \frac{1}{4\Delta k}(\varepsilon_3^{cor} + \varepsilon_4^{cor} + \varepsilon_7^{cor} + \varepsilon_8^{cor} - \varepsilon_1^{cor} - \varepsilon_2^{cor} - \varepsilon_5^{cor} - \varepsilon_6^{cor}),\\
v_z^{sph} &= \frac{1}{4\Delta k}(\varepsilon_5^{cor} + \varepsilon_6^{cor} + \varepsilon_7^{cor} + \varepsilon_8^{cor} - \varepsilon_1^{cor} - \varepsilon_2^{cor} - \varepsilon_3^{cor} - \varepsilon_4^{cor}),
\end{aligned} \tag{4.35}$$

where $\mathbf{k}_i^{cor}$ and $\varepsilon_i^{cor}$ with $i = 1,..,8$ are the wave vectors and energies of the eight corners of the small cube.

The advantage of the sphere setup is that we can know the minimum and maximum energies of all states belonging to this sphere. They are $\hbar\mathbf{k}^{sph}\mathbf{v}^{sph}\pm$

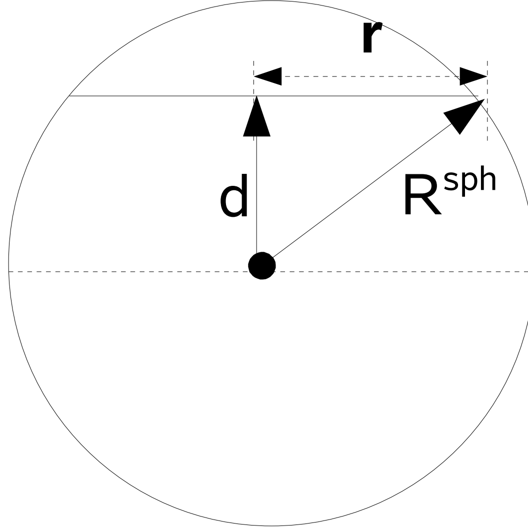


Figure 4.5: Intersection area of an equienergy plane with a sphere in 2D visualization.

$\hbar R^{sph}|\mathbf{v}^{sph}|$. Another advantage is that the intersection between a sphere and a plane is a circle. Therefore it is simple to calculate this intersection area. For each energy $\varepsilon$, we just check to see whether $\varepsilon$ is within this energy range of the sphere or not. If it is within the energy range, the radius of the intersection circle, which is illustrated in Fig 4.5, is calculated by

$$r = \sqrt{(R^{sph})^2 - d^2}, \tag{4.36}$$

while $d$ can be determined by the following energy relation

$$\hbar(\mathbf{k}^{sph}\mathbf{v}^{sph} + d|\mathbf{v}^{sph}|) = \varepsilon. \tag{4.37}$$

As a result, the radius $r$ is

$$r = \sqrt{(R^{sph})^2 - \left(\frac{\hbar\mathbf{k}^{sph}\mathbf{v}^{sph} - \varepsilon}{|\mathbf{v}^{sph}|}\right)^2}. \tag{4.38}$$

After getting $r$, the contribution of the cube to the total $I_2$ in eq. (4.33) can be calculated by

$$D^{sph} = \frac{1}{\hbar}\frac{\pi r^2}{|\mathbf{v}^{shp}|}. \tag{4.39}$$

This method has already been tested by comparing their results with those of the tetrahedron method. As expected, the difference between two methods is small. However, the improvement of the calculation speed from this sphere method is not significant. Therefore, for the sake of accuracy, the tetrahedron method is chosen for calculating II rates.

### 4.2.5 Anisotropic and isotropic impact ionization rate

#### Anisotropic II rate

Up to now, the II rate for every node in the cuboid BZ is known. However, in MC simulations where the grid can be much finer than this cubic grid, the II rate for each point should be defined. Here a logarithmic interpolation between the eight nearest corners of the cubic mesh is used to calculate the wave-vector-dependent II rate for any point in the BZ as follows [100]

$$ln[S^{II}(\mathbf{k})] = \sum_{n=1}^{8} \left[ \prod_{m=x,y,z} \left( 1 - \frac{|k_m - k_m^{(n)}|}{\Delta k_m} \right) \right] \times ln[S^{II}(\mathbf{k}^{(n)})]], \qquad (4.40)$$

where $\Delta k_m = \Delta k = \frac{1}{N}(2\pi/a)$. It was also shown in [100] that the logarithmic interpolation is able to reduce the numerical errors near the II threshold better than the linear interpolation. For the cases $S^{II}(\mathbf{k}^{(n)}) = 0$, a very small value (e.g. $10^{-10}$) is used instead.

So far the II rates for only points within the irreducible wedges, i.e. $\mathbf{k}_1$ points defined in eq. (4.14), are determined. However, from eq.( 4.40), we see that for the points on (or near) the surfaces of the irreducible wedges, the cubes containing them may have corners outside of the irreducible wedges. Therefore, the II rates for these corners should also be calculated in advance.

#### Isotropic II rate

The isotropic II rate, which depends only on the energy of the initial carrier, is also needed for 2D visualization and comparison with other authors' results. The anisotropic II rate $S^{II}(n, \mathbf{k})$ calculated from eq. (4.14) can be transformed into an isotropic II rate $S^{II}(n, \varepsilon)$ by averaging it over the density of states as follows

$$S^{II}(n, \varepsilon) = \frac{\int S^{II}(n, \mathbf{k})\delta(\varepsilon - \varepsilon_n(\mathbf{k}))d^3k}{\int \delta(\varepsilon - \varepsilon_n(\mathbf{k}))d^3k}. \qquad (4.41)$$

It is noted that the calculation in eq. (4.41) is done on a grid for MC simulations, which is much finer than the cubic grid for II calculation. The isotropic II rate is constant over an equienergy plane within a tetrahedron. Moreover it can be assumed that the II rate of the center point of a tetrahedron can be used to best represent the II rate of the whole tetrahedron. The wave vector for this center point is determined by averaging the wave vectors of all corners of the intersection area (either a triangle or a quadrangle, as in Fig. 4.4).

From eqs. (4.14), (4.41) and (4.40), the II rate for every electron state (with wave vector $\mathbf{k}$ and energy $\varepsilon$) can be determined. The II rate is isotropic in case only energy is taken into account and anisotropic in case both the energy and wave vector are considered.

## 4.3 Impact ionization in Monte Carlo simulations

### 4.3.1 Maximum impact ionization rate for a tetrahedron

In the Monte Carlo simulations, besides the II rate for the initial electron (in the case of electron-initiated impact ionization), the maximum of the II rate for each tetrahedron is also needed. There are two methods to find this parameter depending on whether an isotropic or anisotropic ionization is considered.

- Isotropic:

  When the impact ionization is considered isotropic, the maximum II rate for each tetrahedron is the rate at the maximum energy level within such a tetrahedron.

- Anisotropic:

  In order to find the maximum II rate for each tetrahedron in the case of anisotropy, the cubes (with an edge length of $\Delta k = \frac{1}{N}(2\pi/a)$ illustrated in Fig. 4.2) are setup to cover the tetrahedron. Depending on the size of the examined tetrahedron, we can have one or more than one cube, which is illustrated in Fig. 4.6. II rates at all corners of these cubes are examined to find the maximum rate. This rate is assigned to be the maximum II rate for this tetrahedron.

  This method is easy to implement and gives reasonable results. However, as the cube and the tetrahedron are not completely fitted to each other, together with the fact that the II rates are calculated for only corners of the cube, the maximum rate for each tetrahedron is often overestimated. In order to reduce this error, we need a finer cubic mesh, i.e. a smaller $\Delta k$.

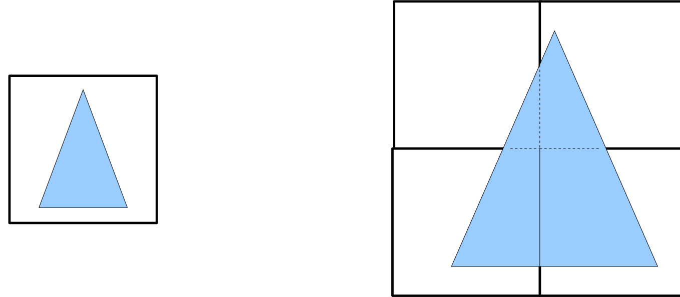

Figure 4.6: Cubes covering the tetrahedron in 2D.

### 4.3.2 Impact ionization coefficient and quantum yield

FB-MC can be used to calculate the impact ionization coefficient and quantum yield. The impact ionization coefficient is defined to be the number of secondary electrons generated by a primary electron per unit length under homogeneous field conditions.

Quantum yield is defined as the number of secondary electrons generated by a primary electron of given energy for a zero electric field. In the Monte Carlo simulator, the quantum yield is measured in a bulk system with a zero electric field into which the electrons are injected with a given initial energy.

## 4.4 Results

The simulation results are fitted to experimental data through two global fitting parameters: the constant fitting parameter $T_{\mathrm{II}}$ (in eq. (4.14)) for the II rates and $Y$ for electron-phonon scattering, where the electron-phonon transition rate outside of the first conduction band is multiplied with $Y$ [37].

Since, up to now, experimental data of the II coefficient and quantum yield of the electron-initiated II are only available for relaxed Si, relaxed Si is chosen for fitting and the obtained parameters are set to $T_{\mathrm{II}} = 1.2 \times 10^{-3}$ and $Y = 0.6$. These values will be used also for strained Si and SiGe later without any adjustment.

### 4.4.1 Relaxed Si

The II rate for relaxed Si with the initial electron in the first conduction band is shown in Fig. 4.7. Since the II rate is anisotropic, especially near the ionization threshold, a fine grid is needed to calculate the II rate at low energies. It is shown that the results converge when $N \geq 32$; and when the energy of the initial electron is larger than 3 eV, $N = 24$ is sufficient.

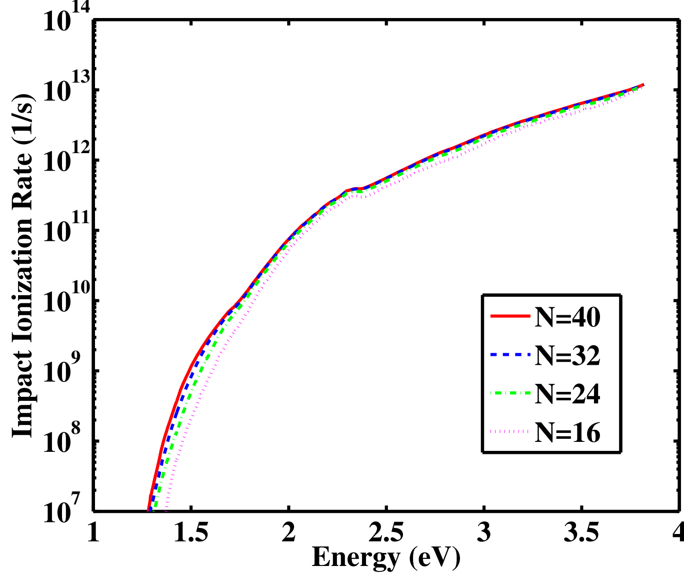


Figure 4.7: Isotropic impact ionization rate with the initial electron in the first conduction band for relaxed Si. N is the number of grid points between 0 - 2$\pi$/a in the cuboid BZ.

This II rate is compared with the rates measured by experiments or calculated by different methods reported in the literature. The comparison is shown in Fig. 4.8. First a comparison with the experimental II rates from Cartier *et al.* [15] is considered. In fact, Cartier *et al.* have performed a combined experimental and theoretical work to extract an empirical expression for the isotropic II rate $S^{II}$ as a function of the electron energy E

$$S^{II}(E) = \sum_{i=1}^{3} \Theta(E - E_{th}^{(i)}) P^{(i)} \left( \frac{E - E^{(i)}}{E_{th}^{(i)}} \right), \tag{4.42}$$

where the threshold energies are $E_{th}^{(i)} = 1.2, 1.8, 3.45$ eV and the prefactors are $P^{(i)} = 6.25 \times 10^{10}, 3.0 \times 10^{12}, 6.8 \times 10^{14}$, for $i = 1, 2, 3$, respectively. $\Theta$ is the step function. Therefore, their curve of the II rate can be considered to consist of three segments. The II rate in the present work shows a good agreement with the II rate from Cartier at the energy range from 1.8 eV to 3 eV, or in the second segment. However, in the first segment (from 1.2 to 1.8 eV), the II rate in the present work is lower. For example, at 1.3 eV, the II rate in this work is $2 \times 10^7$ (1/s) and in Cartier $4 \times 10^8$ (1/s), i.e. around one order of magnitude lower. In the third segment (more than 4 eV), the II rate in this work is around 2 times smaller.

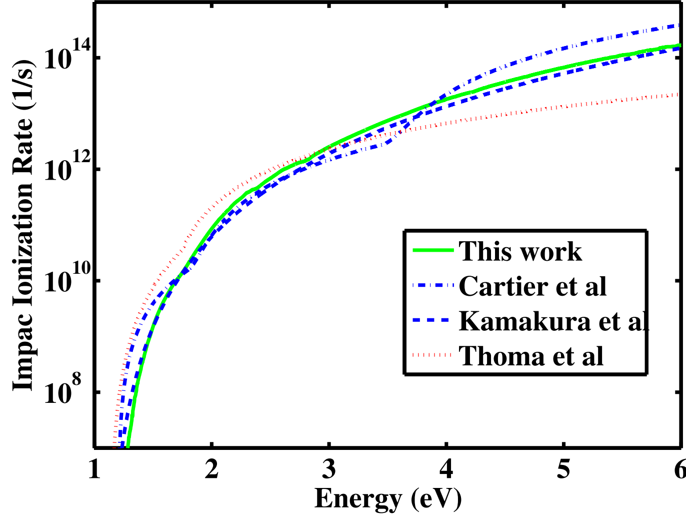


Figure 4.8: Impact ionization rate for relaxed Si in this work (solid line), compared to the results of the experiment (Cartier *et al.* [15]) and other theory calculations (Kamakura *et al.* [65], Thoma *et al.* [117] with modifications from [58]).

It is interesting that a similar observation happens with most of the *ab initio* calculations for the II rate which were published. In Fig. 4.8, the results of Kamakura *et al.* [65] is shown. Compared to the one of Cartier, the II rate of Kamakura has the same behavior as the one of this work, i.e. lower at the energies close to II threshold and higher at high energies. A good agreement over the whole curve of the II rate between Kamakura's report and the present work is observed. Such agreement is also obtained with the results of other theoretical calculations such as Kolnik *et al.* [72] or Sano and Yoshii [100]. It is also seen that the II rates of relaxed Si from *ab initio* calculations with a complex numerical band structure (often obtained by EPM) show a soft behavior. It means the ionization rate slowly increases as the energy increases. This has been explained by the indirect band gap of the material [99].

The II rate from Thoma *et al.* [117], which has been adjusted by a global scaling factor [58], fits well with the one of Cartier at low energy, from 1.2 to 1.5 eV. However, compared to all other reports the one of Thoma is too high at energies from 1.5 eV to 3 eV and too low at high energies (from 3 eV upward). This can be explained by the fact that simple band structures (with three isotropic parabolic upper bands) were used, which might lead to insufficient accuracy of the density of states.

In conclusion, the II rate depends not only on the ionization threshold of the material, but also on its band structure; hence a real band structure is needed for the calculation. And the ionization probability cannot be fitted by the simple Keldysh formula with a single set of parameters.

The II coefficients and quantum yield for relaxed Si are compared with the experimental data in Fig. 4.9 and Fig. 4.10. For II coefficients, a wide range of the electric field is investigated, from 50 to 600 kV/cm. II coefficients varying over 9 orders of magnitude are obtained. A reasonable agreement between the MC simulations and experiments is achieved.

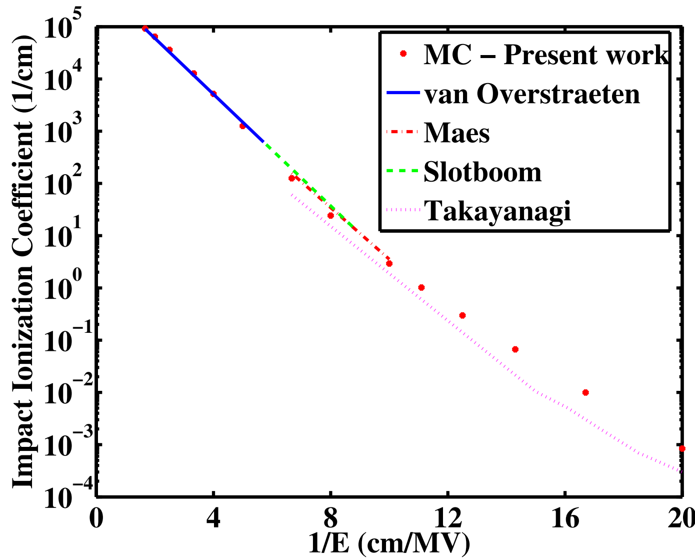


Figure 4.9: Impact ionization coefficient for relaxed Si at room temperature from Full-band MC simulation and experiments (van Overstraeten *et al.* [92], Maes *et al.* [82], Slotboom *et al.* [107], Takayanagi *et al.* [112]).

### 4.4.2 Strained SiGe

Impact ionization for biaxially strained SiGe on a (001) silicon substrate is investigated in this chapter. Fig. 4.11 shows the II rate for strained $\mathrm{Si_{1-x}Ge_x}$ (x=0.2) on relaxed Si. The band gap of this strained SiGe is $E_g = 0.961$ eV [95,102]. Convergence of the results with grid refinement is also investigated. Similar to the relaxed Si case, the results converge when $N \geq 32$ for low energies ($\leq 3$ eV) and $N = 24$ for higher energies.

Fig. 4.12 compares the II rate of strained SiGe and relaxed Si. Due to the smaller bandgap, the II rate in strained SiGe is larger than in Si. However,

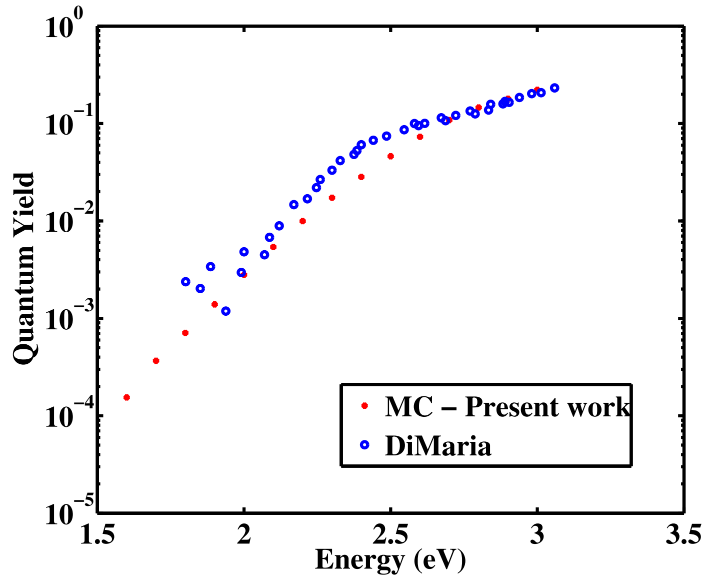


Figure 4.10: Quantum yield for relaxed Si at room temperature from simulations and experimental data by DiMaria *et al.* [30].

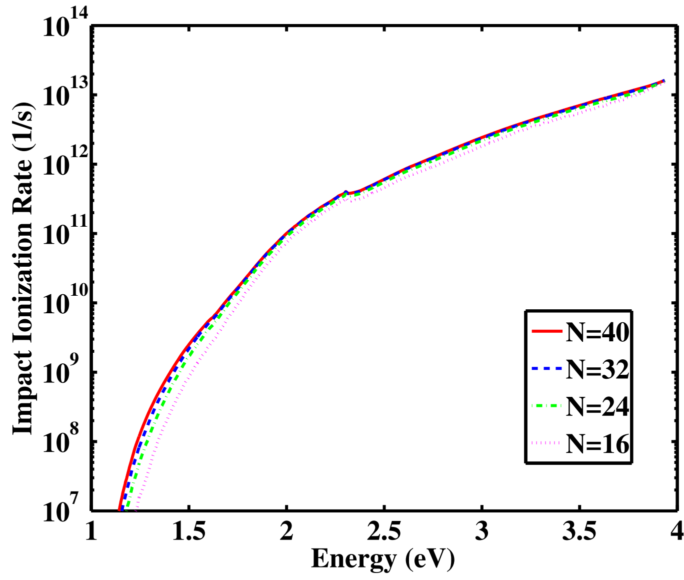


Figure 4.11: Isotropic impact ionization rate with the initial electron in the first conduction band for strained $\mathrm{Si_{1-x}Ge_x}$ (x = 0.2) on a (001) silicon substrate. N is the number of grid points between 0 - $2\pi/a$ in the cuboid BZ.

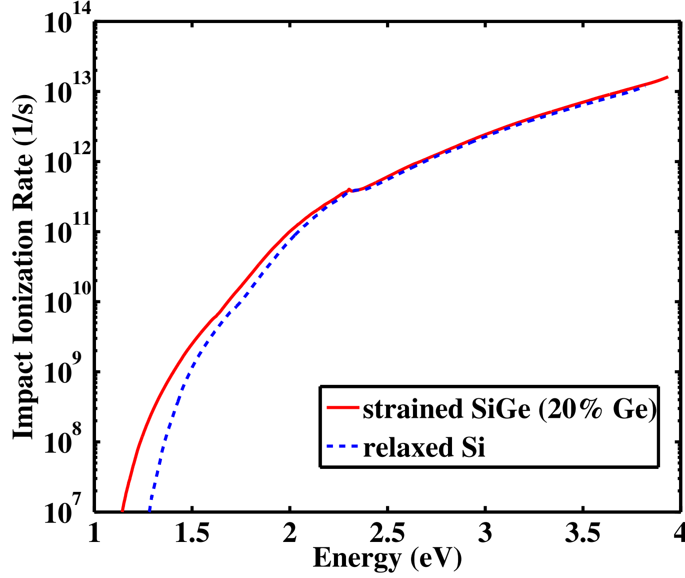


Figure 4.12: Isotropic impact ionization rate with the initial electron in the first conduction band for relaxed Si and strained $\mathrm{Si_{1-x}Ge_x}$ ($x = 0.2$) on a (001) silicon substrate.

the difference is only significant when the energy is below 2.0 eV, because at low energies the band gap is decisive for the ionization rate. At high energies, the density of states, or the availability of free states to fulfill both momentum and energy conservation, is more important than the band gap for ionization to happen.

Fig. 4.13 shows that the II rate for strained SiGe is more anisotropic than for Si, especially at energies near the ionization threshold. We can see that at 1.5 eV the ionization rate of strained SiGe is scattered from $10^2$ to $10^{10}$ for different $\mathbf{k}$ points.

The II coefficients for strained SiGe are calculated for two directions of the applied electric field: in-plane and out-of-plane. The results are shown in Fig. 4.14. There are two remarkable points to be discussed here. First the difference of II coefficients between these directions is not significant, especially when the electric field is larger than 100 kV/cm. Second the II coefficient in strained SiGe is much smaller than that in relaxed Si. This is interesting because of the larger ionization rate in strained SiGe. This effect can be explained by alloy scattering [127]. In fact, the II coefficient for strained SiGe in the case without alloy scattering is also shown and is found to be larger than the one of relaxed Si.

The quantum yield for biaxially strained SiGe is also calculated and shown

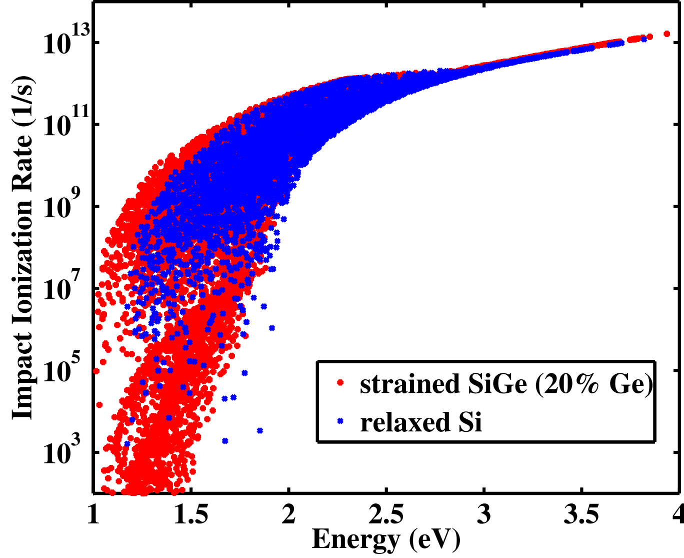


Figure 4.13: Anisotropic impact ionization rate at each **k** point in the BZ for relaxed Si and strained $\mathrm{Si_{1-x}Ge_x}$ (x = 0.2) on a (001) silicon substrate.

in Fig. 4.15. We see that the influence of the elastic alloy scattering can be neglected in this case. The results for the case with alloy scattering and without ally scattering are almost the same. The quantum yield of strained SiGe is higher than the one of relaxed Si, especially at low energies.

### 4.4.3 Uniaxially and biaxially strained Si

Impact ionization rates for uniaxially and biaxially strained Si for high stress levels of 1.5GPa, both compressive and tensile, are investigated. Substrate orientation is (100). The high stress is considered here in order to have a clear picture of the effect of stress on the II rate. Stress directions for the biaxial case are $\langle 010 \rangle$ and $\langle 001 \rangle$. For the uniaxial case, two directions are considered: $\langle 010 \rangle$ (0 degree) and $\langle 011 \rangle$ (45 degree). The details of the stress directions and band gap for strained Si are shown in Tab. 4.1.

The II rates for uniaxially strained Si are shown in Fig. 4.16. It is obvious that the II rate for uniaxial strain along $\langle 011 \rangle$ direction is higher than along $\langle 010 \rangle$, especially at low energies near the ionization threshold. It is interesting that the band gap does not play an important role here, even at low energies. Compressively-uniaxial strain along $\langle 010 \rangle$ direction has the lowest band gap (0.98eV); however its II rate is the smallest one.

Fig. 4.17 compares the II rates between uniaxial and biaxial strain. De-

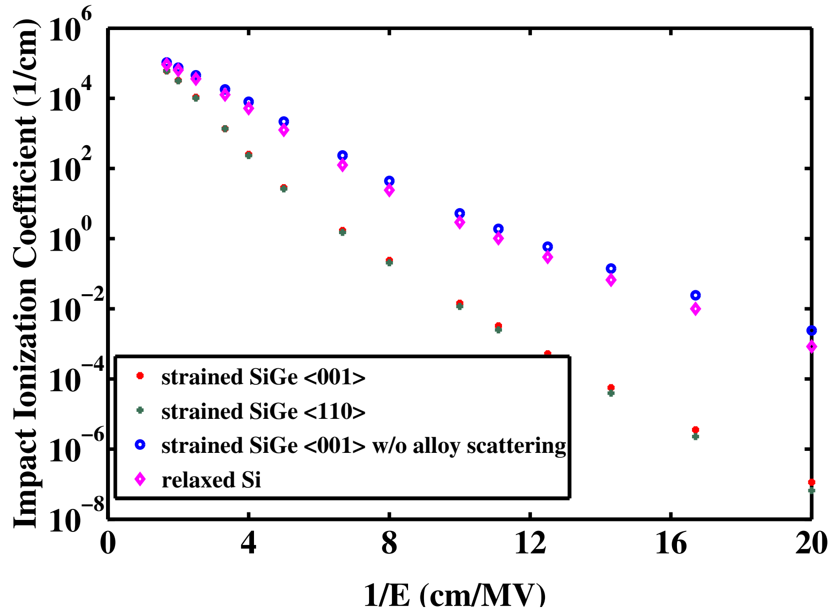


Figure 4.14: Impact ionization coefficient for strained $Si_{1-x}Ge_x$ (x = 0.2) at room temperature from Full-band MC simulations, compared to relaxed Si. Electric field is applied along two directions: $\langle 001 \rangle$ and $\langle 110 \rangle$.

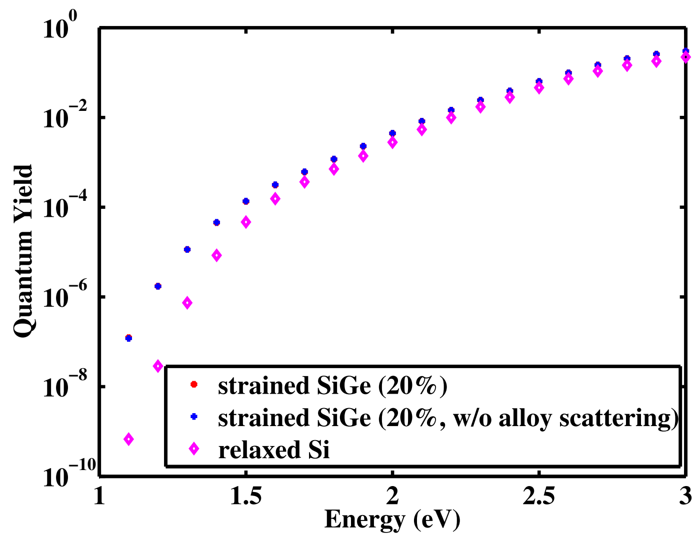


Figure 4.15: Quantum yield for strained $Si_{1-x}Ge_x$ (x = 0.2) at room temperature from Full-band MC simulations, compared to relaxed Si.

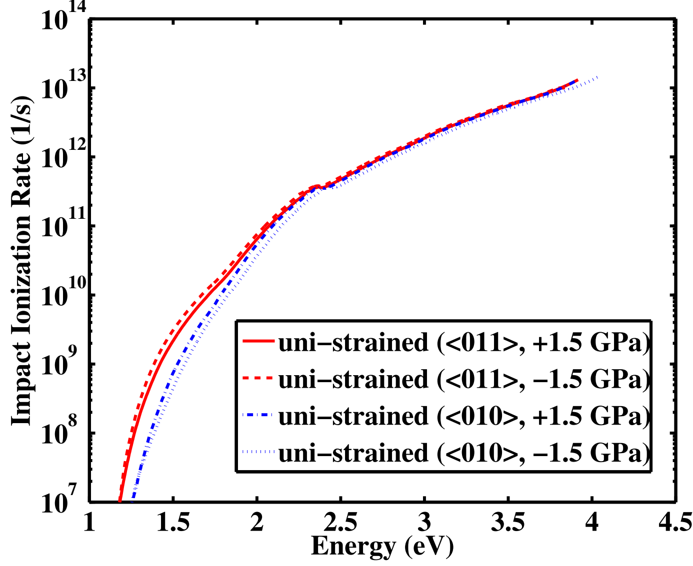


Figure 4.16: Impact ionization rate for uniaxially strained Si with the stress along ⟨010⟩ and ⟨011⟩ directions.

spite the larger band gap, uniaxial strain has a larger II rate. This feature of uniaxial strain is important for applications which require high impact ionization rates without a much reduced band gap to avoid other disadvantages such as band-to-band tunneling. The II rate for biaxial strain is almost the same as that of relaxed Si.

Table 4.1: Band gap for uniaxially and biaxially strained Si (Band gap for Si: 1.124 eV).

| Type | Stress(GPa) | Direction | Gap(eV) |
|---|---|---|---|
| Uniaxial | +1.5 | 010 | 1.04 |
| | | 011 | 1.05 |
| | -1.5 | 010 | 0.98 |
| | | 011 | 1.04 |
| Biaxial | +1.5 | 010,001 | 0.98 |
| | -1.5 | 010,001 | 1.06 |

The capability of higher ionization activities by uniaxial strain compared to biaxial strain is reconfirmed through the II coefficients and quantum yield which are shown in Fig. 4.18 and Fig. 4.19, respectively. II coefficients for uniaxial strain are from 5-10 times larger than for relaxed Si. The difference

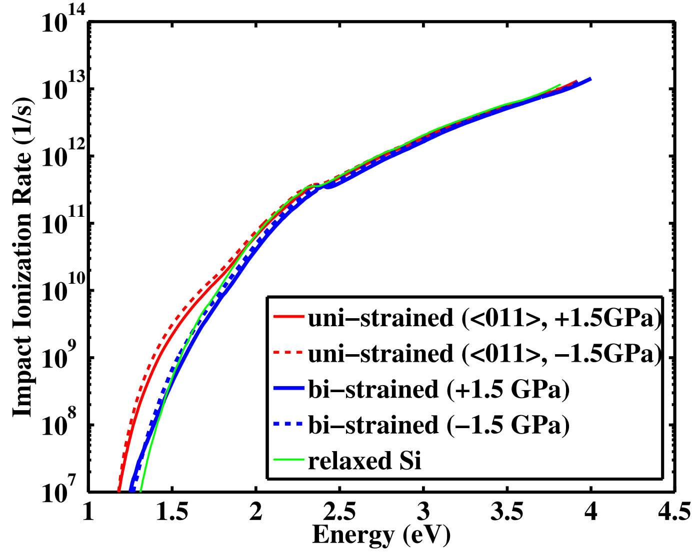


Figure 4.17: Comparison of the II rate for uniaxially/biaxially strained Si and relaxed Si.

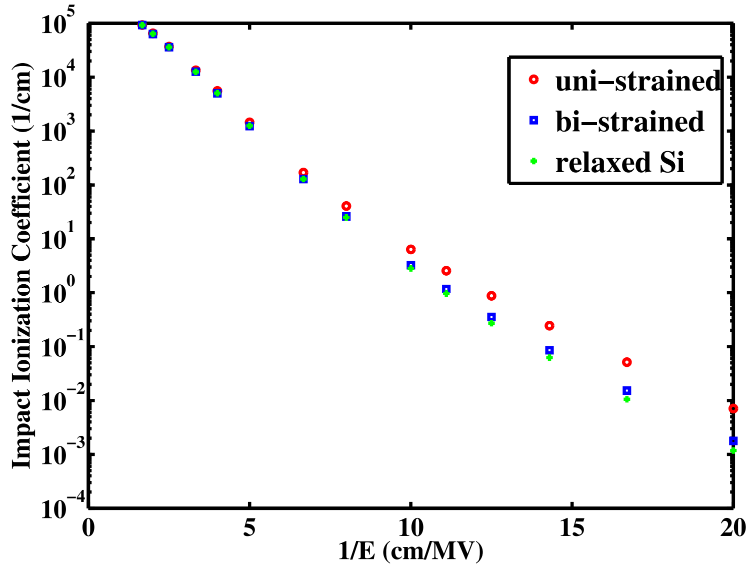


Figure 4.18: Comparison of the II coefficient for uniaxially/biaxially strained Si and relaxed Si at room temperature from Full-band MC simulations.

between biaxial strain and relaxed Si is not significant. The situation is the same when the quantum yield is considered. When the energy is lower

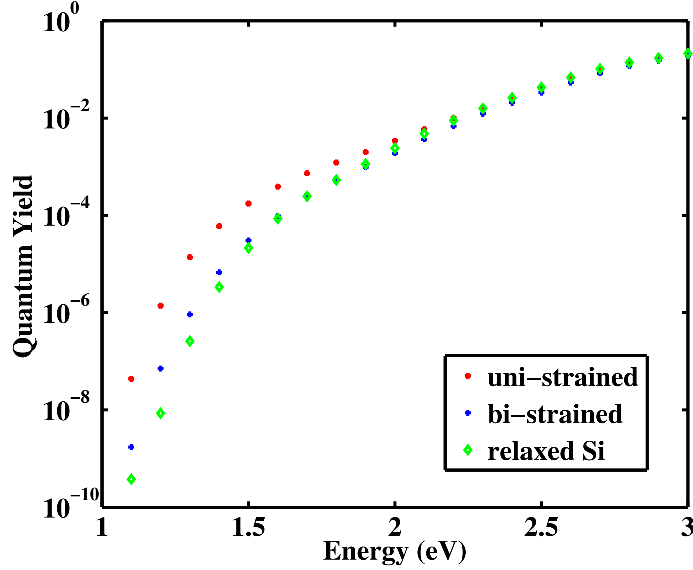


Figure 4.19: Comparison of the quantum yield for uniaxially/biaxially strained Si and relaxed Si at room temperature from Full-band MC simulations.

than 2eV, the quantum yield of uniaxially strained Si is up to 2 orders of magnitude larger than for relaxed Si.

## 4.5 Summary

A model for calculating impact ionization rates for strained Si and strained SiGe has been developed. Through this model, the characteristics of impact ionization for uniaxially/biaxially strained Si and biaxially strained SiGe have been thoroughly investigated. It is found that the band gap plays an important role only at low energies ($< 2.0$ eV), whereas the overall II rate depends on the band structure of the strained material.

For strained SiGe, alloy scattering cancels the gain from the higher II rate, hence resulting in a much smaller II coefficient compared to relaxed Si. Uniaxially strained Si appears to be the most suitable choice when a high II rate is required. Despite its larger band gap (compared to biaxial strain), it gives the highest II rate and does not suffer from alloy scattering. These results give an overview of the way strain engineering can influence the II capability in an IMOS. The next chapter will present how the performance of vertical IMOS can be enhanced by strain engineering.

# Chapter 5

# Vertical impact ionization MOSFETs with strain engineering

## 5.1 Review of IMOS transistors

### 5.1.1 Lateral IMOS

The limit of the subthreshold slope (SS) 60 mV/dec has been overcome by the impact ionization MOSFET (named IMOS), which was introduced by Gopalakrishnan *et al.* [41–43][1]. The structure of a n-channel IMOS is presented in Fig. 5.1. In this n-channel device, *p*+ is the source, *n*+ is the drain and there is a control-gate. The device is a gated p-i-n diode and works by the modulation of its channel length. This IMOS structure is named as *"original lateral IMOS"* to differentiate it with other types of IMOS introduced later.

When the gate voltage $V_G$ is low, there is no inversion layer under the gate and the effective channel length is the entire intrinsic region. At higher $V_G$, an inversion layer forms under the gate and this reduces the effective channel length of the device. When $V_G$ goes higher and higher, the voltage drop across the i-region increases and hence increases the electric field over this region. At a certain high field, avalanche breakdown happens. A very steep subthreshold slope, around 10 - 15 mV/dec was measured for this IMOS [43]. This result comes from the strong dependency of the II coefficients on the electric field and the feedback inherent in the avalanche

[1] In fact, a transistor working on the principle of avalanche breakdown was demonstrated as early as 1964 [105].

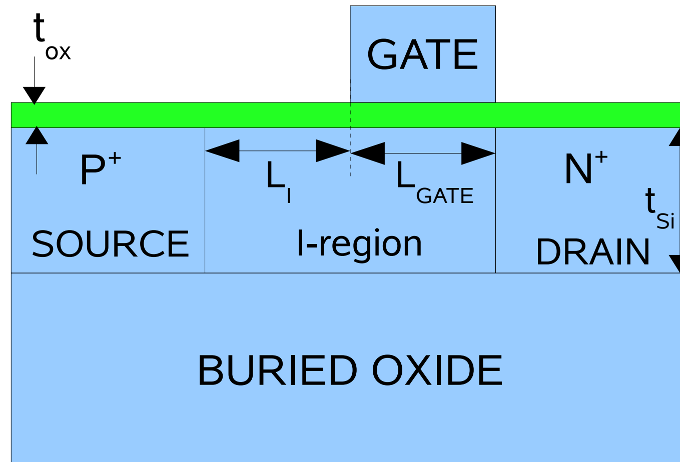


Figure 5.1: Device structure for the n-channel Silicon-on-insulator (SOI) version of the lateral IMOS. $L_{GATE}$: i-region under the gate; $L_I$: i-region outside the gate; $t_{ox}$: oxide thickness; $t_{Si}$: Si bulk thickness [42].

multiplication process [42].

One of the most serious limits of this IMOS device is that it requires a very high supply voltage. For example, with $\mathrm{L_I} = 0.3\ \mu m$, a $\mathrm{V_{DS}}$ of more than 10 V is required. Such a high voltage causes damage by hot carriers to the oxide, which leads to a tremendous shift of the threshold voltage $\mathrm{V_T}$. The experiment in [43] showed that $\mathrm{V_T}$, and hence the electrical characteristics, changed critically after only 5 times of switching. This instability limits the proper operation of the device.

Gopalakrishnan *et al.* have discussed and investigated the capability of using Ge instead of Si for the IMOS to reduce the breakdown voltage and increase the reliability of the device. The breakdown voltage in Ge-IMOS is approximately a factor of 2 - 3 lower than those in Si [43]. For example, with $\mathrm{L_I} = 0.3\ \mu m$, the breakdown voltage of a Si-IMOS is around 13 V whereas the one of a Ge-IMOS is around 6 V.

### 5.1.2 Self-aligned IMOS

Choi *et al.* have proposed a method for an easier fabrication of the IMOS. In the original process, the gate, source and drain region are formed in turn using separate masks, which means that the gate length $L_{GATE}$ and the i-region length $L_I$ are limited by the alignment error and that the fabrication cost is increased due to additional photolithography steps. To reduce the fabrication cost, a self-aligned n-channel IMOS was introduced [22],

in which the i- and the source regions are formed simultaneously using a TEOS (tetraethoxysilane) sidewall spacer. The device structure is presented in Fig. 5.2(a).

Besides the cost, the purpose of the work was to accelerate scaling down the size and enhance the performance of IMOS. It was reported that at room temperature a SS of 7.5 mV/dec at a bias of $V_S = -5.5$ V and $V_D = -1.0$ V was obtained for this n-channel IMOS device. Compared to the original IMOS, this result is a significant improvement in term of the supply voltage. However, the ratio between ON- and OFF-current is 81.1/2.8 $\mu A$ per $\mu m$, which is far below the respective ratio of more than 5 orders of magnitude reported in [43].

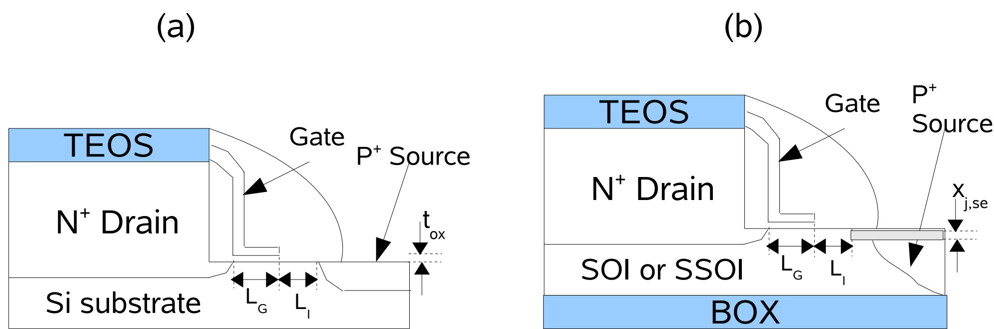


Figure 5.2: Device structure for: (a) Self-aligned n-channel IMOS [22] and (b) Self-align n-channel IMOS with Silicon-on-insulator (SOI) or Strain silicon-on-insulator (SSOI) [21]. $t_{ox}$ is the oxide thickness and $x_{j,se}$ is the depth of the source extension junction; TEOS: tetraethoxysilane.

Another attempt was also introduced by Choi *et al.* in [23] to further enhance the performance of the self-aligned IMOS. By using Silicon-on-insulator (SOI) and modifying the mask layout, the ON/OFF current ratio was shown to increase by a factor of 1000 compared to that of the old structure. However, there was no improvement in the supply voltage. Also in this work, an approach of combining a 70 nm IMOS and a 70 nm tunneling-field-effect transistors (TFET) was introduced to implement both high-performance and low-power functionality on one chip.

In the latest publication [21], Choi indicated the possibility of using strain techniques to reduce the breakdown voltage [21]. The proposed structure is presented in Fig. 5.2(b). In this work, different device parameters such as i-region length, source extension junction depth and gate oxide thickness were varied to get the lowest breakdown voltage. However, only simulation results have been shown.

### 5.1.3 L-shaped IMOS

Toh *et al.* [118] proposed another structure for the IMOS with elevated source/drain regions, named LI-MOS. The special thing of this structure is the L-shape of the impact ionization region, which comprises an elevated i-region and a horizontal i-region. The structure is presented in Fig. 5.3(a). Here the L-shaped i-region has the total length $L_I = T_I + L_S$.

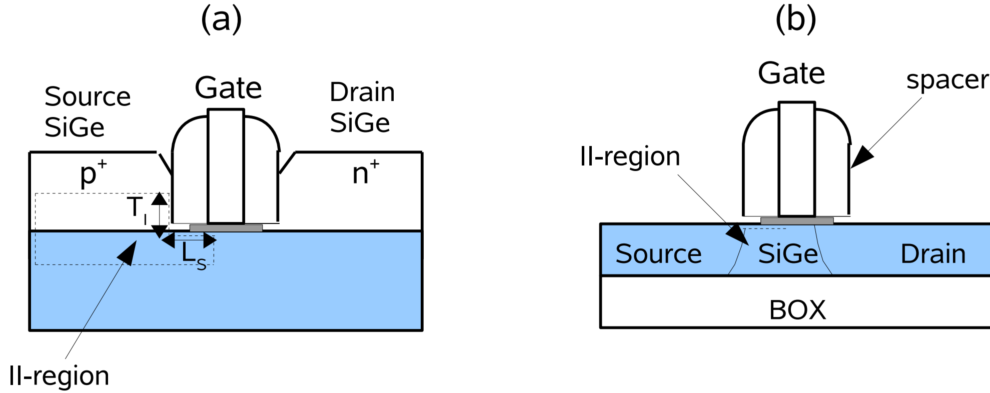


Figure 5.3: Device structure for: (a) LI-MOS with strained SiGe in the raised source/drain [118] and (b) SiGe IMOS on insulator [119].

One advantage of the LI-MOS transistor is that it allows the engineering of the band structure in the vertical i-region to enhance the II activity as well as relocate it away from the gate dielectric for better reliability. Strained $\mathrm{Si_{1-x}Ge_x}$ has been adopted in this device. With $L_I = 60$ nm ($T_I = 30$ nm and $L_S = 30$ nm), $L_G = 60$ nm and a strained raised source/drain of a $\mathrm{Si_{0.75}Ge_{0.25}}$ alloy, the breakdown voltage was 8.0 V, i.e. around 10% lower than that of the relaxed-Si structure. At $V_S = -7.8$ V, a subthreshold slope of 3.28 mV/dec and 4.46 mV/dec was obtained for LI-MOS with and without this SiGe alloy, respectively [118].

In another publication [119], Toh *et al.* have demonstrated other IMOS structures for reducing operating voltages and enhancing performances by incorporating relaxed SiGe, presented in Fig. 5.3(b). A breakdown voltage of 6.35 V and 5.05 V was obtained for a n-channel IMOS without Ge and with 40%Ge, respectively ($L_G = 50$ nm and $L_I = 40$ nm). A SS of 5 - 6 mV/dec was achieved for these devices. This structure was also optimized by different doping and dimensioning techniques in [120]. However, no significant improvement of supply voltage was obtained.

### 5.1.4 Depletion IMOS

Another effort of improving the performance of lateral IMOS was published by Onal *et al.* [88] recently. The authors have introduced a depletion-IMOS (DIMOS). The basic device of a DIMOS is similar to the original IMOS, except that the region under the gate is not intrinsic but doped to a moderate-to-high level of the drain conductivity type. This device has to be implemented in a fully depleted structure in a SOI or Fin-type configuration. Fig. 5.4 presents the structure of the DIMOS.

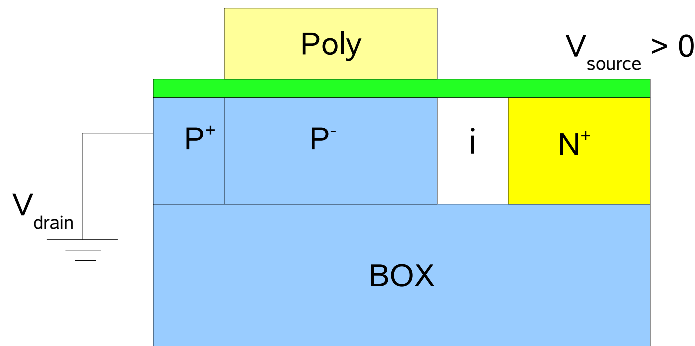


Figure 5.4: Device structure of a p-type DIMOS [88].

In the DIMOS, the gate is used to deplete the underlying doped region to switch the normally ON-device to the OFF state. For a p-channel DIMOS as in Fig. 5.4, the n+ (source) terminal is biased positive with respect to the p+ (drain) which leads to an avalanche breakdown of the lateral p-i-n diode when $V_G = V_{fb}$ ($V_{fb}$: flatband voltage). A subsequent sweep of the gate voltage to the positive direction depletes the region underneath the gate of free carriers. This, in turn, lowers the lateral field that causes avalanche breakdown; hence, the device turns off with a SS steeper than $kT/q$.

The main advantage of the DIMOS is that the avalanche breakdown occurs uniformly in the bulk, instead of concentrating only near the surface as in the original IMOS, which helps to increase II and reduce the damage by hot carriers to the oxide. Experimental results have shown the improvement of the DIMOS over the original IMOS, especially regarding the reliability and operating voltage. SS of the DIMOS is kept steep (under 20 mV/dec) and $V_T$ decreases lightly (by 0.72 V) after 50 switching cycles (at a supply voltage $V_{DS} = 11.5$ V), which is a significant improvement. A 40% lower voltage operation and an 80% improvement for the ON current were achieved. It is also expected here that the drain-to-source voltages of DIMOS can be scaled down to the simulated bulk p-i-n avalanche breakdown limits for Si ($\approx 3$ V) and Ge ($\leq 1$ V).

## 5.2 Vertical IMOS transistors

Though the above mentioned lateral IMOS structures have significant improvement over the original one, they still suffer from high operating voltages and damages by hot electrons to the oxide. The lowest supply voltage reported so far is 4.8 V [119]. In an effort to overcome these limits, the vertical IMOS, which is a planar-doped barrier MOSFET with a floating body, has been introduced and investigated extensively by experiments [2–5]. The device structure is shown in Fig. 5.5. This vertical IMOS is not based on avalanche breakdown like the lateral IMOS. Instead, the holes generated by impact ionization charge the floating p-body and cause a dynamic reduction of the threshold voltage, which leads to an extremely fast rising drain current in the subthreshold region.

One of the most important advantages of the device is that it offers the mechanisms for mitigating the damages by hot electrons almost completely [4] and shows the capability of working properly under high temperatures [2]. These characteristics will be discussed in details later.

### 5.2.1 Device structure

The structure of the standard vertical IMOS made solely of relaxed silicon, named RSi-IMOS, is shown (Fig. 5.5). This structure comprises source and drain regions with $n^+$ doping, an intrinsic channel with a highly doped $\delta p^+$ layer in the middle and two gates. With this structure, an arbitrary choice of doping levels for this delta layer, which is above a certain minimum level necessary to obtain a significant potential barrier, can be made. Therefore, it is possible to achieve high electric fields in the intrinsic zone near the drain without applying a very high drain-source voltage. Moreover, the static leakage current is very low, in the range of pA [3] thanks to this barrier. The doping profile of the device is shown in Fig. 5.6 [32].

The operation of this device can be explained through its band structure at different bias conditions in Fig. 5.7. This device has a triangular-shaped potential barrier with the peak at the delta layer. The potential difference between the drain and source directly results in an electric field in the intrinsic region near the drain. A positive gate voltage is also needed to lower the barrier to let electrons flow from source to drain for turning on the device. At low drain biases, for example in this case $V_D = 1V, V_{G1} = V_{G2} = 1.2V$, the electric field is not high enough to cause significant impact ionization and the device, hence, behaves like a conventional MOSFET. The electric field is increased by increasing the drain voltage. At a certain condition, for example in this case $V_D = 2V, V_{G1} = V_{G2} = 1.2V$, II is significant so that the

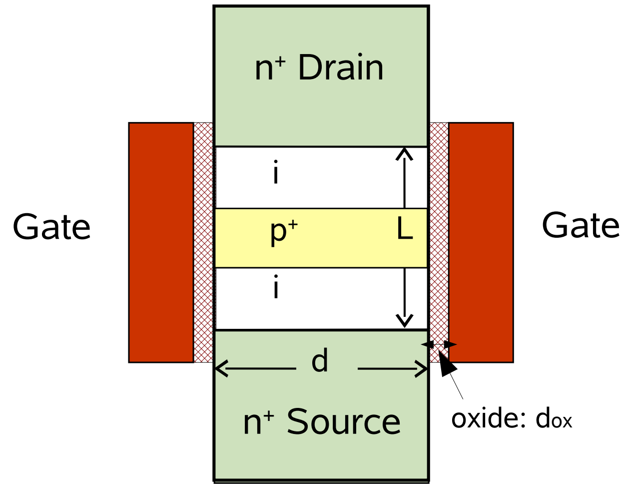


Figure 5.5: Structure of the vertical IMOS [75].

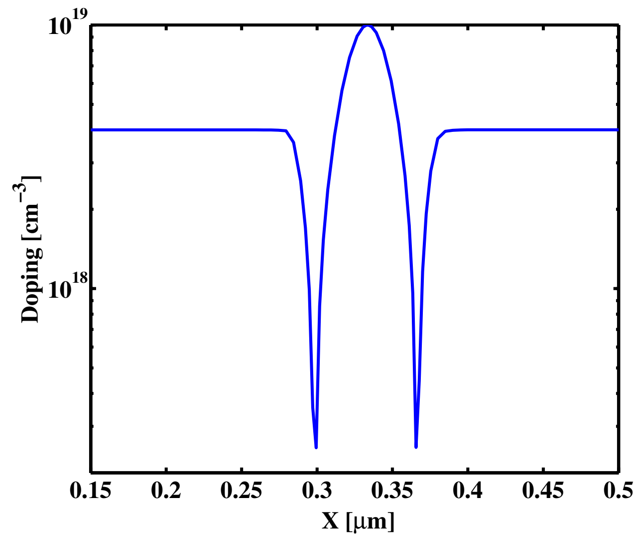


Figure 5.6: Doping profile of the investigated vertical IMOS transistors: $N_{drain} = N_{source} \sim 4 \times 10^{18}cm^{-3}$; the p+ layer has a Gaussian profile with a peak value of $P_{max} = 1 \times 10^{19}cm^{-3}$ [32].

threshold voltage is reduced drastically. Therefore, a very steep subthreshold slope is achieved. Fig. 5.8 shows operation modes of the device through device simulations [75]. These simulation results agree well with the experimental data published in [3].

We see that since this device needs impact ionization to charge the floating body, it requires lower electric field and hence much lower drain voltage

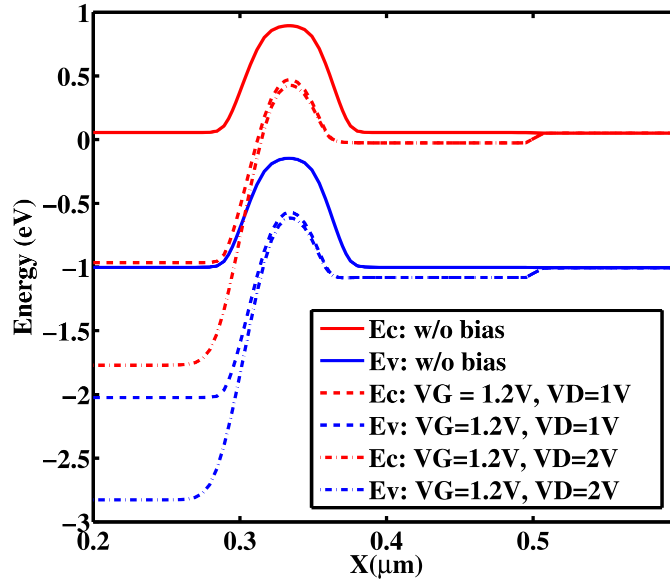


Figure 5.7: Band diagrams for the investigated vertical IMOS.

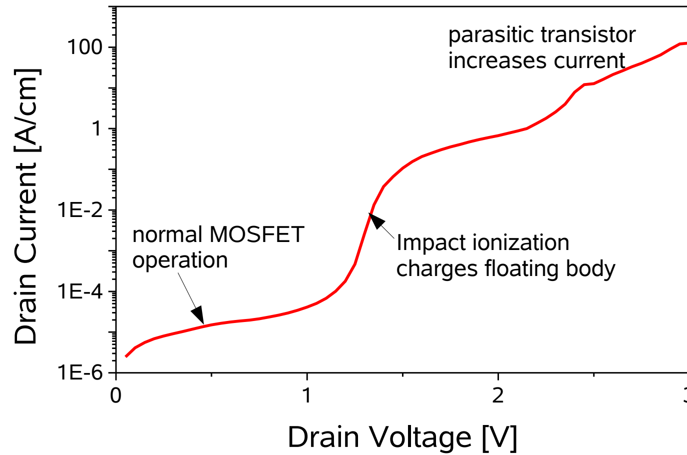


Figure 5.8: Different operation modes of a vertical IMOS with a wide body (d = $5\mu$m) [75].

than the lateral IMOS. Moreover, the hot carriers are spread into the bulk, not only confined to the silicon-oxide interface. As a result, the device can reduce the damages by hot electrons almost completely and has a stable operation. The device was also shown to work well up to 250°C with a stable threshold voltage and leakage current in the nA range [2]. Higher temperatures, up to 350°C, are also expected for these IMOS. This is a very

significant improvement of the working reliability for IMOS transistors.

However, the device suffers from hysteresis. This phenomenon can be explained by the parasitic bipolar transistors formed at high drain voltage, as in Fig. 5.8, with the $n^+$ drain region as emitter, $n^+$ source region as collector and the delta layer as base. When the current originating from this bipolar transistor exceeds a certain level, the device can no longer be turned off by reducing the gate source voltage. At that time, the bipolar transistor itself delivers the electrons which provide the holes for its base current via impact ionization. In this case, only a reduction of drain voltage below the level for impact ionization or a negative gate source voltage applied to the device can switch it off again [3].

Another disadvantage of the device is its low switching speed. To switch the device completely off, the floating body has to be discharged completely via the body source junction. However during the discharging process, electrons are further injected from the source until this junction is no longer forward biased. Recombination can also contribute to the discharging; however the recombination rate is low due to the low electron density in the p-body [75]. This effect is well known from SOI devices with a floating body. Finding a suitable way to increase the switching speed of the device, up to the GHz range, is very challenging.

Device simulations have been applied to find the optimum parameters for these vertical IMOS [75]. It was shown that a very thin body, such as 50 nm, can help to increase the switching speed of the IMOS to the GHz range. However, reducing the supply voltage and minimizing or avoiding hysteresis is still an open issue.

### 5.2.2 Vertical IMOS with a strained SiGe layer

In the paper by Savio *et al.* [101], through a review of all I-MOS structures made of Si reported by different groups and by a combination of simulations and experiments for their own lateral IMOS structure, it was concluded that the supply voltage of silicon IMOS devices saturates at a certain length and does not fall under about 4.5 V. The threshold minimum of short transistors cannot be reduced by a further reduction of the electrical length. Therefore any further improvement of the IMOS should go with Ge structures. In this dissertation, for the purpose of reducing the supply voltage, a vertical IMOS with a strained SiGe layer placed between the channel and the drain is investigated.

The vertical IMOS with a strained SiGe layer, named in this work SSiGe-IMOS, is depicted in Fig. 5.9. The pseudomorphically-grown SiGe layer is biaxially strained because of the lattice mismatch with the relaxed bulk

Si [28]. Due to its smaller band gap, strained SiGe has higher II rates than relaxed Si, especially at low energies near the ionization threshold as shown in Fig. 4.12. Therefore, strained SiGe is expected to give higher generation rates than Si if a strained layer is implemented in the II region of the IMOS.

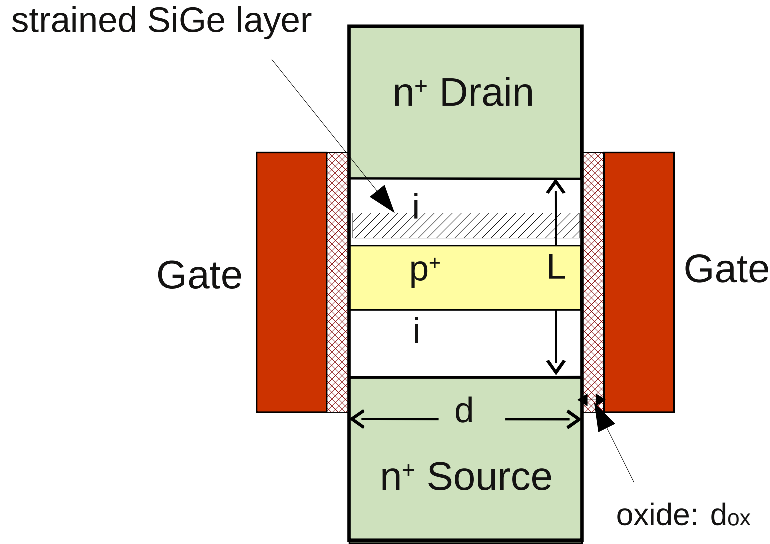


Figure 5.9: Structure of the vertical strained-SiGe-IMOS (SSiGe-IMOS). Device dimension: d=50 nm, L=70 nm, $d_{ox}$ = 5nm; Gate: nPoly [32].

Since the II rates depend on the location within the device, the position of the strained SiGe layer will be determined carefully to obtain the highest rates. The threshold voltage of the device also changes depending on the layer's position. Moreover, the shape and thickness of the SiGe layer affects the performance of the device. Abrupt changes in the Ge content can result in spurious barriers, which could choke the channel. Therefore, various Ge profiles are investigated. An example of a Ge profile for the SSiGe-IMOS is illustrated in Fig. 5.10.

The band diagrams of the channel region for the RSi-IMOS and SSiGe-IMOS are presented in Fig. 5.11. When no bias is applied, as in Fig. 5.11(a), the high barrier at the $\delta p^+$ prevents electrons to flow from the source to the drain. However, when the drain and gate voltage are large enough, as in Fig. 5.11(b), the potential barrier is lowered and the electric field in the intrinsic region near the drain increases. Consequently, electrons from the source can get into the intrinsic region with a high electric field and impact ionization happens.

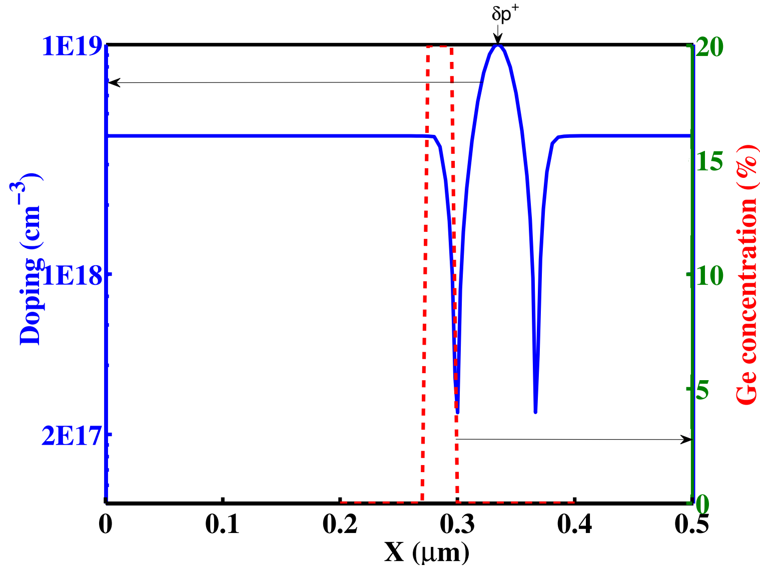


Figure 5.10: The dashed line shows one of Ge profiles for the SSiGe-IMOS investigated in this work (the continuous line is the doping profile, which is the same as the doping profile for the RSi-IMOS in Fig. 5.6).

## 5.3 Device simulations

### 5.3.1 Combination of MC and HD simulators

HD simulations are combined with MC simulations to investigate the devices. Since the HD simulator and the Lucky-Electron (LE) model for II [63] cannot capture all details of the anisotropic band structure and the avalanche mechanism, the LE-II model is calibrated to match the MC results. First, the IMOS is simulated with HD but without II for different contact voltages. The resulting electrostatic potentials, charge carrier distributions, and current densities provide the input for nonself-consistent MC simulations (NSC-MC). Then the obtained generation rates are used to adjust the parameters of the LE-II model in the HD simulator. This approach which assures consistent simulations between the two simulators is outlined in Fig. 5.12.

### 5.3.2 Full-band Monte Carlo simulator

#### Impact ionization

Simulation of the II activities in devices by MC requires the knowledge of the scattering rate of II for the investigated materials. In this work, we

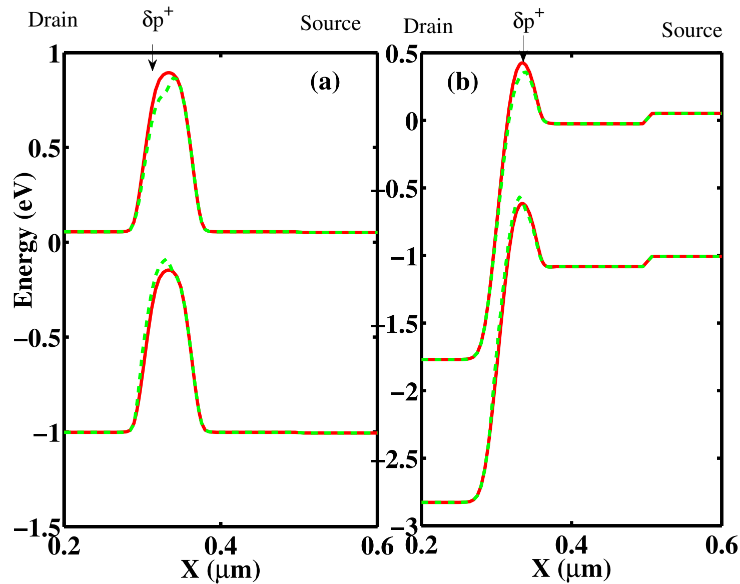


Figure 5.11: Band diagrams for the RSi-IMOS and SSiGe-IMOS for the cases: (a) without external bias (b) $V_D = 2V, V_{G1} = V_{G2} = 1.2V$. Continuous lines are for the RSi-IMOS and dashed lines are for the SSiGe-IMOS. Doping and Ge profiles are shown in Fig. 5.10.

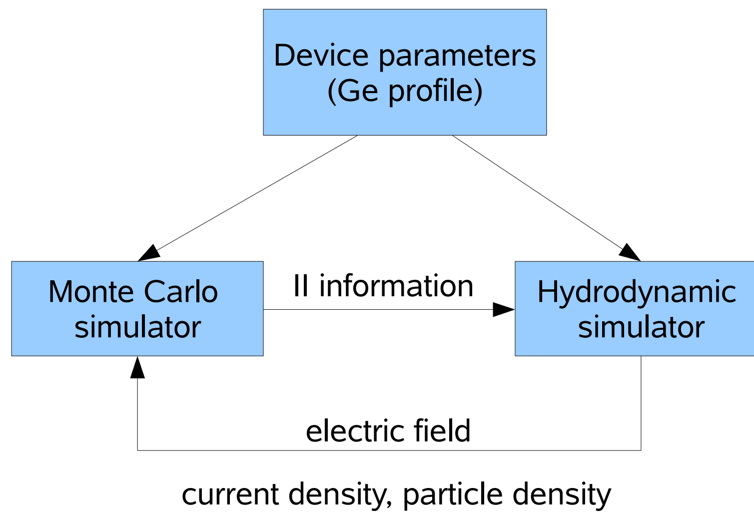


Figure 5.12: The simulators for the RSi-IMOS and SSiGe-IMOS.

use the II rates for strained SiGe which were presented in Chapter 4. This II model takes into account the anisotropy of the ionization rates, i.e. the II rate depends not only on the energy of the initial carrier, but also on its wave

vector. This is important, because the II rate of strained SiGe is strongly anisotropic in the k-space, especially when the energies of the initial carriers are close to the ionization threshold (see Fig. 4.13).

**Non-self consistent simulations**

The nonself-consistent Monte Carlo (NSC-MC) approach has been shown to be an efficient and reliable tool for hot electron simulations [14, 64]. The CPU-time of NSC-MC simulations is orders of magnitude less than that of self-consistent ones, whereas both methods yield the same results for hot electron quantities. Therefore, in this work, NSC simulations are used to investigate carrier generation by II in the vertical IMOS.

NSC-MC also allows us to restrict the MC simulation window domain to a small window of the active device region and to neglect most parts of the high particle density areas of the source and drain regions [48]. Fig. 5.13 shows the MC simulation window within the investigated RSi-IMOS. It is clear that all hot-electron activities are confined to this area of the device.

### 5.3.3 Hydrodynamic simulator

As has been shown in [75], HD simulations are able to give very reliable results compared with experimental data, whereas the simpler drift-diffusion model cannot reproduce the typical effects of the vertical IMOS with sufficient accuracy. Therefore HD is used to obtain the electrical characteristics of the RSi-IMOS and SSiGe-IMOS. But it is only necessary to solve the energy balance equation for electrons; holes can be considered with constant hole temperature. The reason is that the effect of impact ionization originates mainly from electrons. The simulation results remain nearly identical if the energy balance equation of the holes is included, but simulation time and convergence problems increase significantly [33].

The results which are presented in the following have been obtained with the device simulator GALENE III [1]. For impact ionization, a nonlocal model which is the soft threshold LE model [63] has been used. Band-to-band tunneling has also been taken into account with a nonlocal model [124].

## 5.4 Results and Discussions

### 5.4.1 Full-band Monte Carlo simulations

Besides the II rate, the II coefficient and quantum yield are important measures of II activities. From Sec. 4.4.2 in Chapter 4, we already know that

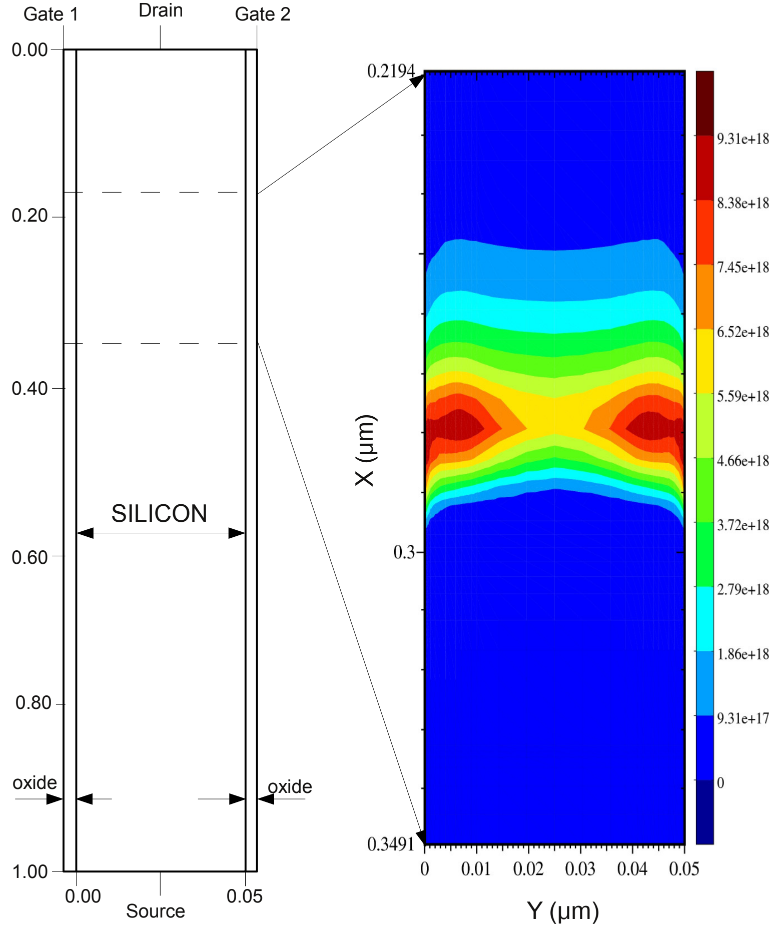


Figure 5.13: Impact ionization rate ($1/s.cm^3$) in the RSi-IMOS within the MC simulation window at bias condition: $V_{G1} = V_{G2} = 1.2V$ and $V_D = 1.75V$.

II coefficients for biaxially strained SiGe (with 20% Ge) are much smaller than for relaxed Si (see Fig. 4.14), whereas the quantum yield in strained SiGe is larger than the one in relaxed Si (see Fig. 4.15). The difference of these two quantities leads us to investigate to see whether II in the SSiGe-IMOS happens under conditions which are closer to the II coefficient experiment or quantum yield experiment.

Various Ge profiles with a maximum concentration of 20% are investigated. As we know, there is a tradeoff between the layer Ge fraction and the thickness to maintain the layer strain. Hence, the maximum thickness of the strained layer is kept below about 50 nm to avoid any relaxation [28]. In addition, we see from Fig. 5.13 that the II generation rate in the RSi-IMOS

is highest in the range from 0.25 - 0.30 $\mu$m along the vertical direction of the device. Therefore, the following Ge profiles which can be divided into three groups (Fig. 5.14) will be investigated:

- Group 1: The strained SiGe layer covers the highest II area and extends over the junction to the channel (Profiles: GE1-1, GE1-2).
- Group 2: The strained SiGe layer covers only the highest II area, i.e. from 0.26 - 0.29 $\mu$m (Profiles: GE2-1, GE2-2).
- Group 3: The strained SiGe layer is placed in the channel region, between the peak of the p-doping and the junction to the drain (Profiles: GE3-1, GE3-2).

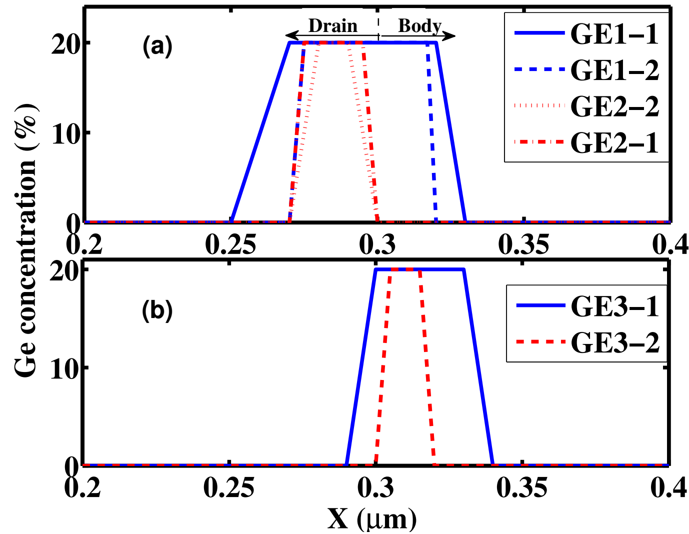


Figure 5.14: Different Ge profiles for the strained SiGe layer in the SSiGe-IMOS: (a) Group 1 and Group 2 - strained layer mostly inside the drain (b) Group 3 - strained layer in the channel.

The generation current due to II in the RSi-IMOS and SSiGe-IMOS is investigated to have a first look at the impact of the strained SiGe layer. Fig. 5.15 shows the obtained II generation current for various bias conditions. The Ge profile *GE1-2* of Fig. 5.14 is used. For all drain and gate bias conditions, the vertical IMOS with a strained SiGe layer shows an about 10 times higher II generation current than the RSi-IMOS. This result seems to be related to the quantum yield of strained SiGe and relaxed Si (Fig. 4.15). For the same II generation current, the drain voltage in the SSiGe-IMOS can be lower by 0.4-0.7V than in the RSi-IMOS.

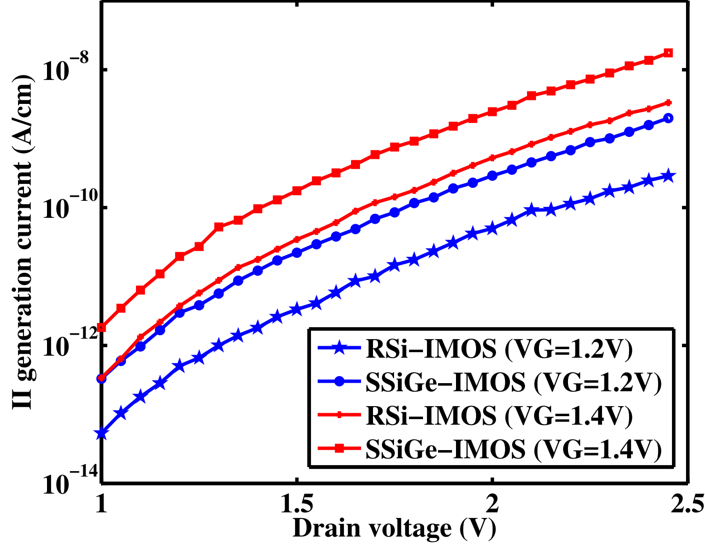


Figure 5.15: II generation current of the SSiGe-IMOS and RSi-IMOS with $\mathrm{V_{G1}} = \mathrm{V_{G2}} = 1.2, 1.4\mathrm{V}$. Ge profile *GE1-2* in Fig. 5.14 is used for the SSiGe-IMOS.

Next the dependence of the II generation current on Ge profile is shown in Fig. 5.16. There are two remarkable points to be discussed from these results. First, the II generation current rises when the thickness of the strained SiGe layer increases, regardless of the position of Ge layer. The current of GE1-1, GE2-1 and GE3-1 is larger than that of GE1-2, GE2-2 and GE3-2, respectively. Second, it is surprising that the Ge profiles of group 3 give the highest II generation currents even though all strained layers are placed in the channel, not in the highest II region in the drain. When group 1 and group 3 are compared at the same thickness, the Ge profiles of group 3 prevail. The Ge profile *GE3-1* of 35 nm thickness can give a higher II generation current than the Ge profile *GE1-1* of 50 nm.

For a better understanding of the origins of these results, the avalanche multiplication factor *M-1* for these devices is calculated. This factor is defined as

$$M - 1 = \frac{I_{gen}}{I_{inj}}, \tag{5.1}$$

where $I_{gen}$ is the II generation current and $I_{inj}$ is the injection current of the NSC-MC simulation, which is determined by the HD simulator. From the results in Fig. 5.17(a), we see that the multiplication factors in all investigated SSiGe-IMOS are smaller than those of the RSi-IMOS (in this figure, only

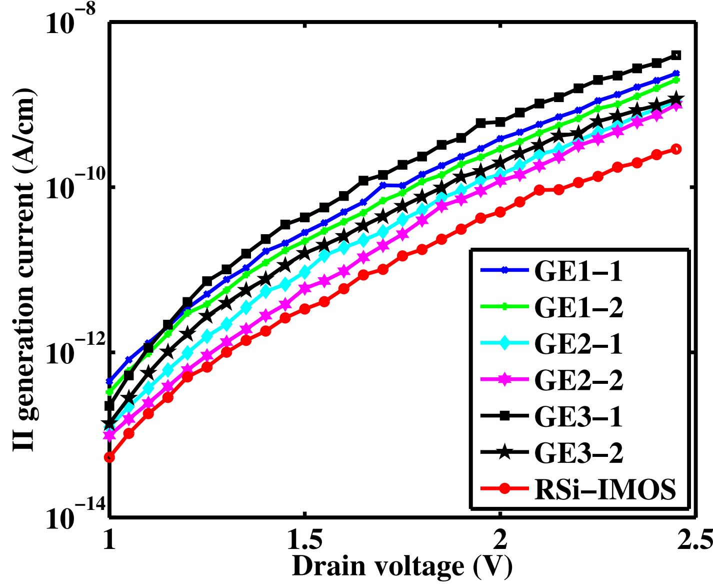


Figure 5.16: II generation currents of the SSiGe-IMOS with different Ge profiles (in Fig. 5.14) at $V_{G1} = V_{G2} = 1.2V$.

one representative of each Ge profile group is shown for easy visualization). However if the alloy scattering effect is turned off, the multiplication factors in the SSiGe-IMOS with the Ge profiles of group 1 and group 2 are equal to or slightly higher than the factor in the RSi-IMOS, which is shown in Fig. 5.17(b). This is similar to the case of II coefficients in Fig. 4.14, where strained SiGe has much lower coefficients than relaxed Si in the case with alloy scattering. When alloy scattering is not taken into account, the II coefficient of strained SiGe is larger.

It is also observed that although the Ge profiles of group 3 produce the highest II generation current for the SSiGe-IMOS, their respective multiplication factors are the smallest. Alloy scattering does not make a significant contribution here. The multiplication factors for both cases, with and without alloy scattering, are almost the same and they are smaller than their counterpart in the RSi-IMOS.

From these observations, it can be concluded that II events in the vertical IMOS occur under conditions which are closer to the II coefficient experiment than the quantum yield experiment. The increase in II generation current is not due to the increasing capability of II activities in strained SiGe layer, i.e. II efficiency does not increase. Therefore, other reasons should be considered, such as the threshold voltage of the device. The simulation results from the

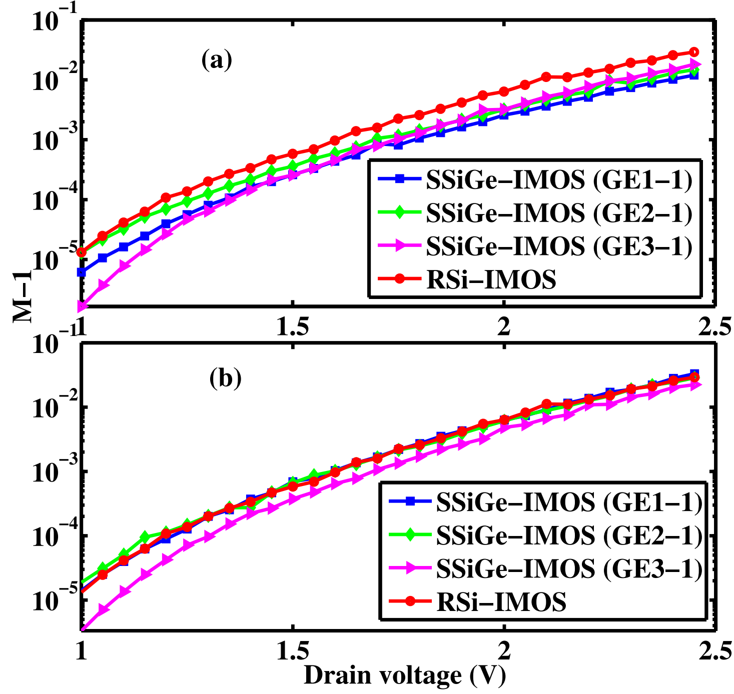


Figure 5.17: Multiplication factor of the RSi-IMOS and SSiGe-IMOS with different Ge profiles at $V_{G1} = V_{G2} = 1.2V$: (a) With alloy scattering (b) Without alloy scattering.

HD simulator will provide more insight into the SSiGe-IMOS devices.

### 5.4.2 Hydrodynamic simulations

The results of the II generation obtained by NSC-MC simulations are used to calibrate the lucky electron model which is used to calculate the avalanche generation with the HD simulator. A comparison of the II generation currents resulting from NSC-MC and HD simulations is shown in Fig. 5.18. The same parameter set of the lucky electron model is used for the RSi-IMOS and the SSiGe-IMOS. The agreement between MC and the lucky electron model is good; therefore HD simulations are sufficiently accurate to investigate the influence of relaxed Si and strained SiGe on the electrical characteristics of the IMOS transistor.

The following results have been obtained by self-consistent solutions which consider the interactions of avalanche generation and the electrostatic potential in the p-body determining charge carrier densities and currents [33].

In Fig. 5.19, the $I_D - V_G$ characteristics of two different SSiGe-devices

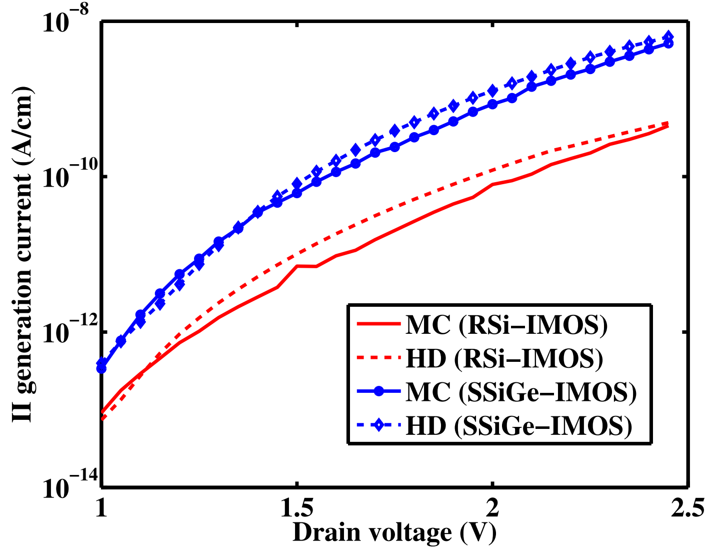


Figure 5.18: II generation currents from NSC-MC simulations are used to calibrate the HD model for the RSi-IMOS and SSiGe-IMOS (with Ge profile *GE3-1*) [33].

and a RSi-IMOS transistor are compared. There is nearly no difference between the RSi-IMOS and the IMOS with the strained SiGe layer in the drain (*GE2-2*); whereas with such a layer inside the p-body region (*GE3-1*), the threshold voltage of the SSiGe-IMOS is reduced by approximately 130 mV. However, this threshold shift has no influence on the leakage current at zero gate voltage as well as on the on-current at high gate voltages. They remain at the same level; therefore the good $I_{on}/I_{off}$ ratio is not reduced. Also the subthreshold swing does not change.

Fig. 5.20 shows the drain current as a function of the drain voltage at $V_G = 1.2$ V. With a strained SiGe layer in the drain region (profile *GE2-2*), a distinct change of the device characteristic is again not achieved. However, if this layer is inside the p-body (*GE3-1*), the current of the SSiGe-IMOS is significantly larger than the current of the RSi-IMOS at any drain bias. The transition from normal MOSFET operation to IMOS operation also occurs at a lower drain voltage. This is due to the fact that the higher current level in the SSiGe-IMOS leads to a larger amount of generated holes which, in turn, charge up the body region [33]. At low drain voltages, however, the avalanche generation is too small to influence the device behavior. Therefore, the current increase due to the SiGe layer must be caused by a different effect which is investigated below.

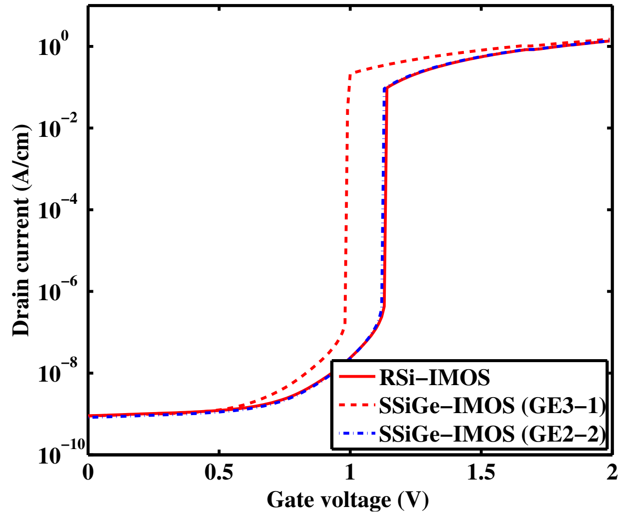


Figure 5.19: Simulated $I_D - V_G$ curves of the RSi-IMOS and SSiGe-IMOS at $V_D = 2V$ [33].

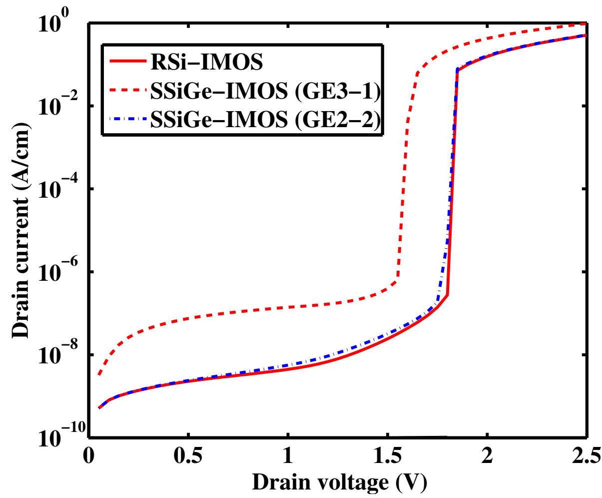


Figure 5.20: Simulated $I_D - V_D$ curves of the RSi-IMOS and SSiGe-IMOS at $V_{G1} = V_{G2} = 1.2V$ [33].

Fig. 5.21 shows the band diagrams of the RSi-IMOS and SSiGe-IMOS with different Ge profiles. It can be seen that a Ge-layer inside the p-body (*GE3-1*) lowers the energy barrier between the source and drain of electrons in the conduction band. This has the consequence that the electron current in the MOSFET channel increases. Due to such a lower barrier, the electric field is also reduced and the effect of II is weaker. The reduction of these

quantities also explains why smaller multiplication factors are observed for SSiGe-devices with a Ge layer in the body (see Fig. 5.17). Nevertheless, the II generation in these devices is higher due to the higher electron current. For the Ge layer in the drain region, a small reduction of the energy barrier in the conduction band can also be noticed. But this layer simultaneously lowers the barrier for holes in the p-body and the charging up of the body is reduced. Both effects compensate each other [33].

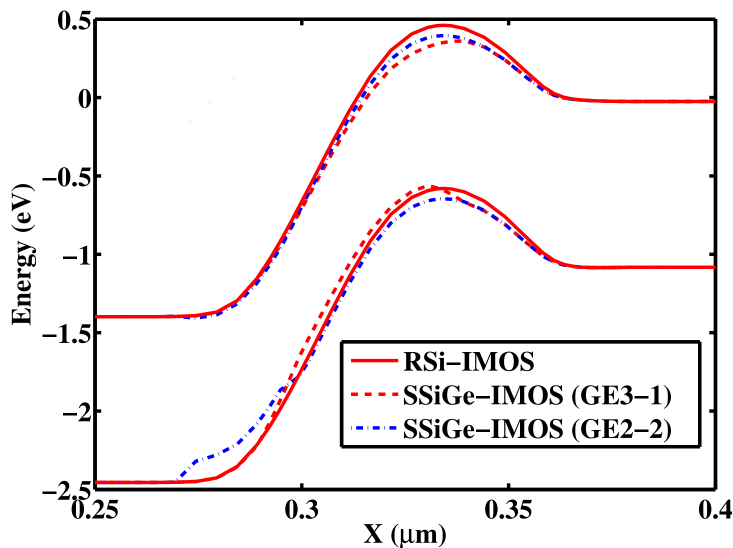


Figure 5.21: Band diagrams of the RSi-IMOS and SSiGe-IMOS at $V_G = 1.2V$ and $V_D = 1.5V$.

## 5.5 Noise investigation in vertical IMOS transistors

### 5.5.1 Simulation approach

Impact ionization degrades the noise performance in nanoscale devices. This fact is shown in [97] for heterojunction bipolar transistor (HBT). The excess noise attributed to II has also been shown for SOI MOSFETs by many groups, which was reviewed in [106]. For lateral IMOS, the avalanche breakdown produces orders of magnitude more drain noise than a conventional MOSFET [43]. The vertical IMOS is expected to produce less noise than the lateral one because II is used to charge the floating body only. However, it is

necessary to investigate the noise in these devices to see whether this noise is detrimental for digital and analog applications or not.

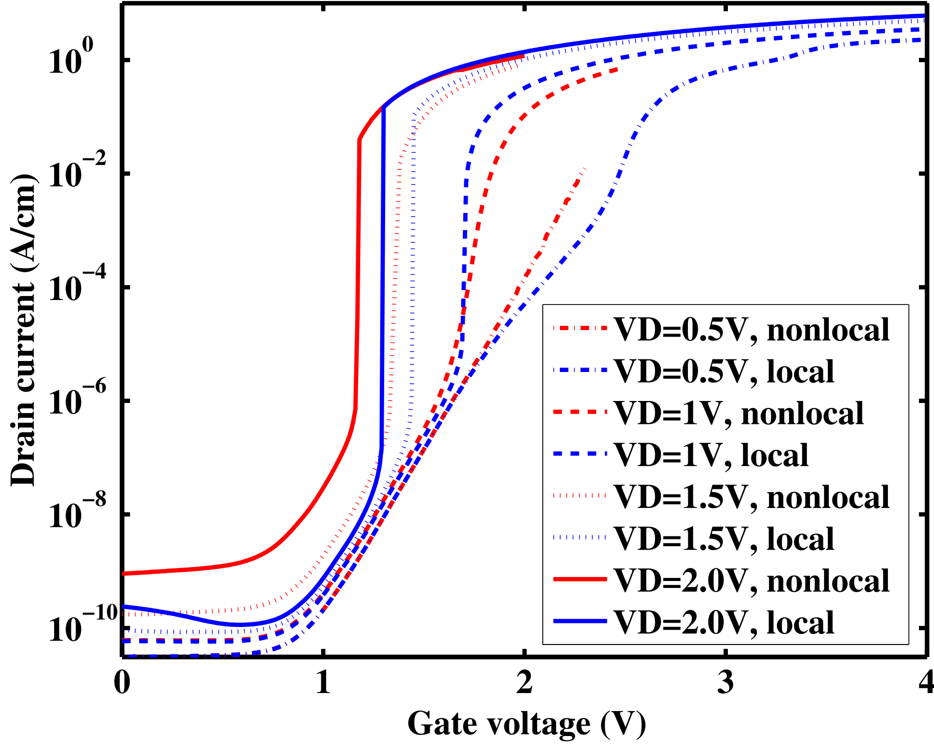


Figure 5.22: Comparison of $I_D - V_G$ curves for local and nonlocal model at different bias conditions.

Hydrodynamic simulations have been shown to be able to capture correctly the impact of the floating body effect on noise in SOI devices [62]. In principle, the vertical IMOS here behaves like these SOI devices. Therefore, it is expected that its noise performance can also be investigated by HD simulations. This work will follow the model mentioned in [60, 62] for noise investigation.

The simulations include noise due to carrier scattering (so-called diffusion noise), II and Shockley-Read-Hall recombination [11, 103]. The local temperature model for II [83] is used instead of the nonlocal one. The reason is that nonlocal models destroy the band structure of the Jacobian required for noise calculations [62]. It was also shown in [62] that the local model can well represent the noise behavior of SOI devices.

The I-V characteristics of the IMOS device obtained from local and nonlocal models are shown in Fig. 5.22. We can see that II effect in the local

model happens later than in the nonlocal one. At low drain bias, the difference between the two models is significant. However, at higher drain bias, which is the operating voltage of the investigated IMOS device, the local model can give acceptable results, especially at $I_{on}$ and $I_{off}$ ranges. For the noise simulation, band-to-band tunneling is also turned off.

### 5.5.2 Results and discussions

#### Results for the RSi-IMOS

First of all, the spectral intensity of the drain current fluctuations is obtained for a wide range of frequencies. The result is shown in Fig. 5.23. The contribution of different types of noise to the overall noise performance of the vertical IMOS is presented in this figure. The noise due to hole diffusion is the largest among the types of noise. The second largest is the generation-recombination noise, followed by the electron diffusion noise. The II noise is the smallest one.

As discussed above, since II does not contribute directly to the drain current but generates holes that accumulate in the floating body, noise from II is not as significant as the other noise sources. In contrast, the noise of the generated hole current from II is the largest. In fact, also in Fig. 5.23, we see that when II is turned off, the main noise comes from electron diffusion. The total noise of the RSi-IMOS with II is two orders of magnitude larger than without II.

It is observed that the noise of the vertical IMOS has a well-known Lorentzian shape, which is similar to the noise in other SOI-MOSFETs with a floating body. The noise is characterized by a plateau value in the power spectral density, followed by a $1/f^2$ decrease above a characteristic corner frequency $f_c$. This behavior is caused by a first-order low pass filter, which is shown in [55]. The small-signal equivalent circuit of the noise for the floating body is illustrated in Fig. 5.24.

The equivalent body-ground capacitance $C_{eq}$ can be modeled as the sum of all the capacitance seen from the body, which includes the body to depletion capacitance $C_{bb}$, the body to gate (two gates) capacitance $C_{bg}$ and the body to source and drain capacitance $C_{js}$ and $C_{jd}$. The $C_{eq}$ can also be calculated by the change of the body charge over the body potential as $C_{eq} \approx \partial Q_b/\partial V_b$. The equivalent body-ground resistor is calculated by $r_{eq} = dV_b/dI_b$. The cut-off frequency of the Lorentzian noise spectrum is determined by

$$f_c = \frac{1}{2\pi r_{eq} C_{eq}} \tag{5.2}$$

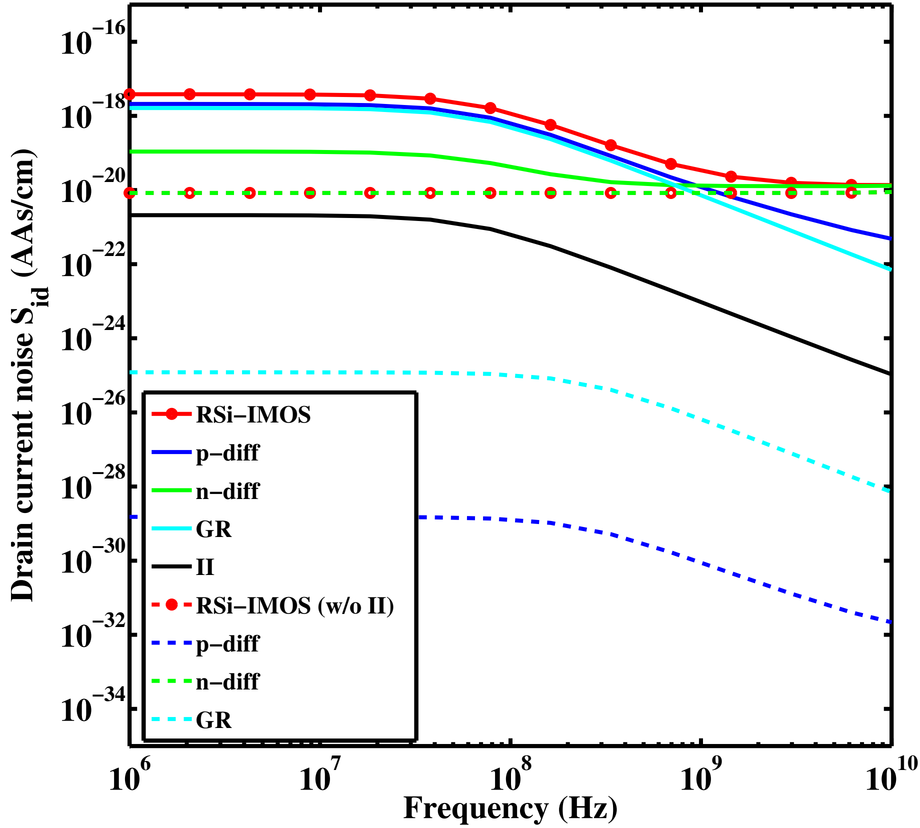


Figure 5.23: Spectral intensity of the drain current fluctuations for the RSi-IMOS. The case without II is also simulated for the comparison. The drain current for both cases are chosen to be the same (around 0.1A/cm). Device bias for two cases: With II: $V_D = 2V$ and $V_{G1} = V_{G2} = 1.35V$; Without II: $V_D = 2V$ and $V_{G1} = V_{G2} = 3.57V$.

Fig. 5.25 shows that the change of the noise performance with the increasing drain bias is also similar to the noise behavior of SOI-MOSFETs, which has been measured [55] and shown through HD simulations [62]. Increasing drain bias results in a reduction of $r_{eq}$ and leads to a low amplification gain by the floating-body to the shot noise. Therefore, the noise magnitude decreases and the cutoff frequency increases (due to eq. (5.2)).

The minimum noise figure for the RSi-IMOS is also calculated, which is presented in Fig. 5.26. We see that the noise figure increases when the drain voltage increases. This behavior has already been seen in Fig. 5.25. When II is turned off, $NF_{min}$ is almost independent of the drain voltage, which is

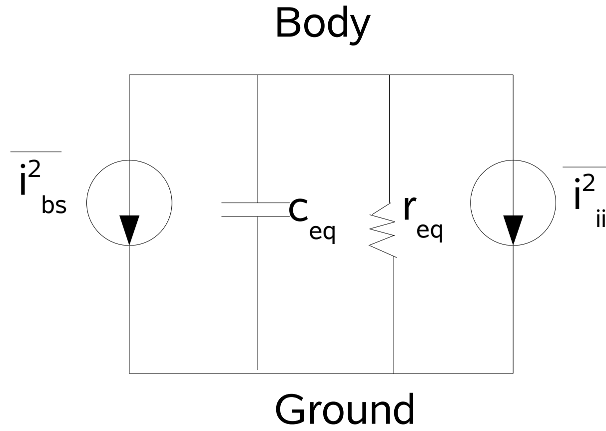


Figure 5.24: Noise small-signal equivalent circuit for the floating body [55].

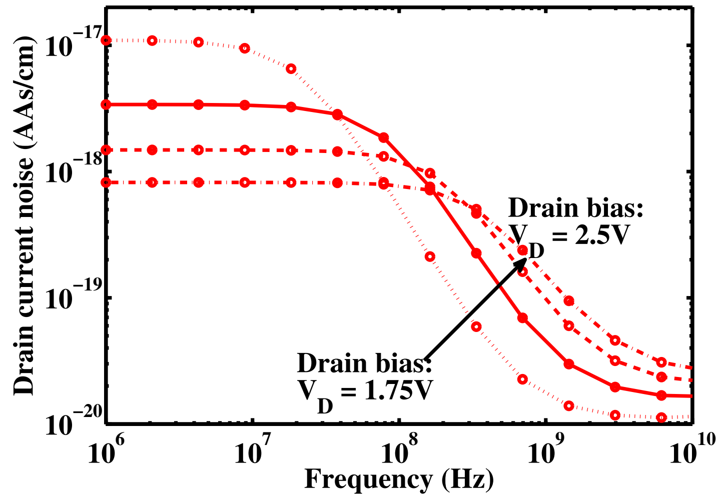


Figure 5.25: Spectral intensity of the drain current fluctuations of the RSi-IMOS for a gate bias of 1.4V and drain bias steps of 0.25V.

also mentioned in [97].

### Results for the SSiGe-IMOS

First the spectral density of the drain current fluctuations for the SSiGe-IMOS at a gate voltage of 1.35 V and a drain voltage of 2 V is simulated, in comparison with the RSi-IMOS at the same bias. The result is shown in Fig. 5.27. We can see that the GR noise is the cause for the noise difference between these two devices. This can be explained that more holes and also more electrons, as discussed in Sec. 5.2.2, are generated for the SSiGe-

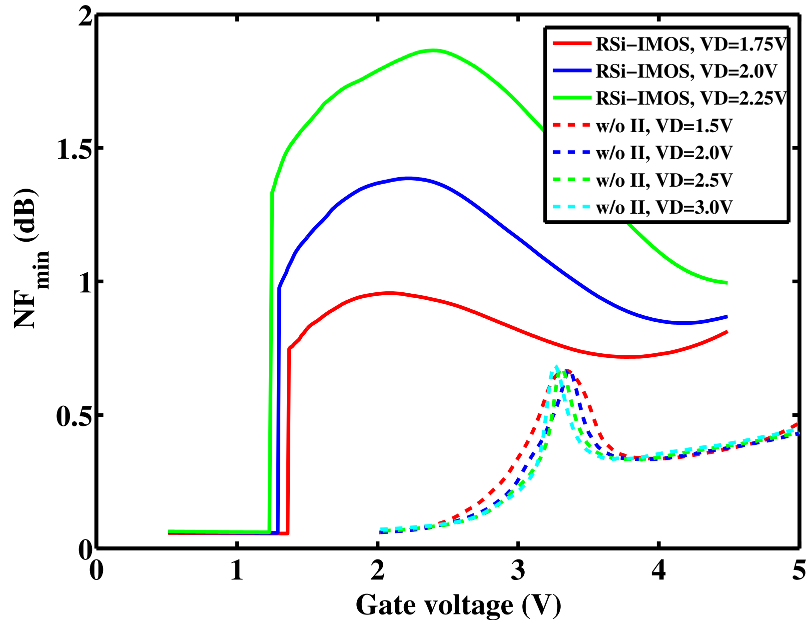


Figure 5.26: The minimum noise figure for the RSi-IMOS at a frequency of 1 GHz.

IMOS than for the RSi-IMOS; hence the process of generation-recombination happens also to be stronger.

Second the drain current noise over the varied drain and gate voltages is investigated. Fig. 5.28 and Fig. 5.29 show respectively the spectral intensity of the drain current fluctuations over $V_G$ and $V_D$ at a frequency of 100 MHz. The noise is relative to the magnitude of the drain current, as presented in Fig. 5.19 and Fig. 5.20. In general, the difference between these two structures is not significant.

Finally the minimum noise figure is investigated. Fig. 5.30 compares $NF_{min}$ at different frequencies versus gate voltage. The SSiGe-IMOS has a higher $NF_{min}$ than the RSi-IMOS does at lower frequencies; however at higher frequencies the situation is opposite. This can be explained from the fact that the cut-off frequency of the SSiGe-IMOS is higher than that of the RSi-IMOS, which was shown in Fig. 5.27 above.

The minimum noise figure of these devices under varied drain voltages is presented in Fig. 5.31. The situation is similar to the case of varied gate voltages as in Fig. 5.31. One remarkable point to be discussed here is that at very high drain voltages, even at high frequencies, the SSiGe-IMOS gives worse $NF_{min}$ than the RSi-IMOS.

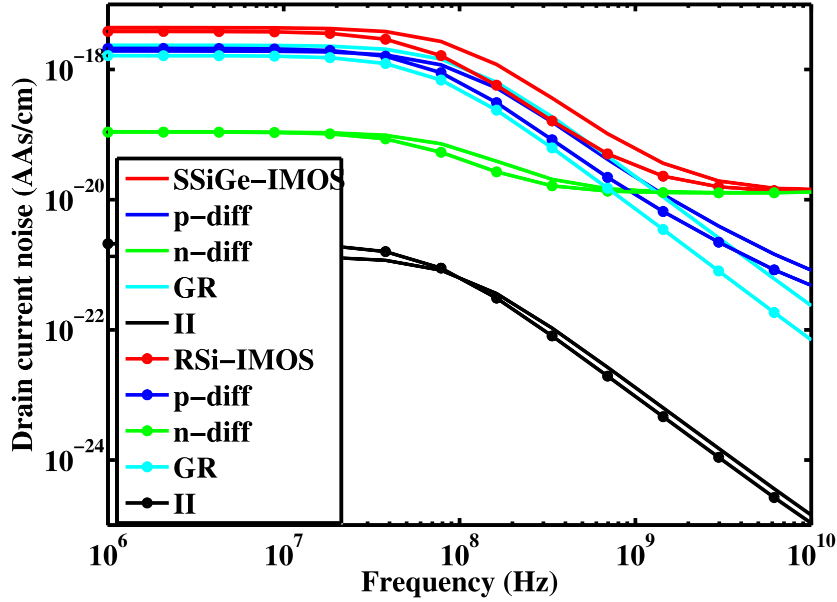


Figure 5.27: Comparison of the drain current noise between the SSiGe-IMOS and RSi-IMOS at the bias condition of $V_D = 2V$ and $V_G = 1.35V$.

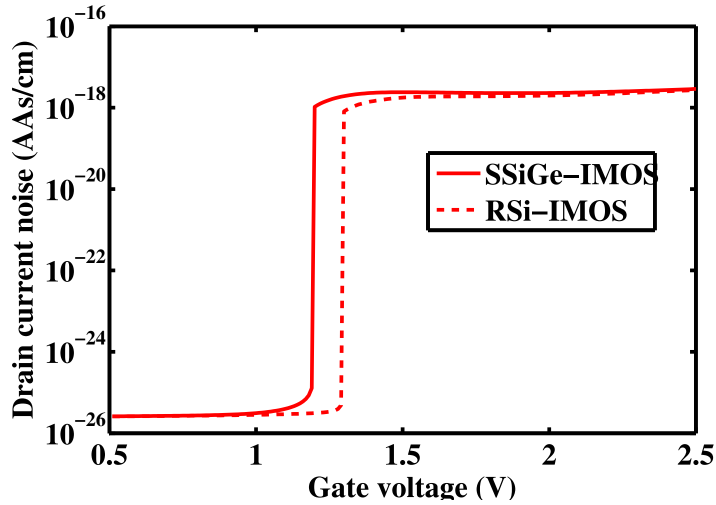


Figure 5.28: Comparison of the drain current noise between the SSiGe-IMOS and RSi-IMOS over different gate voltages ($V_D = 2V$) at a frequency of 100 MHz.

## 5.6 Summary

A vertical IMOS transistor with a strained SiGe layer in the drain-sided ionization region and in the channel has been investigated by 2D device

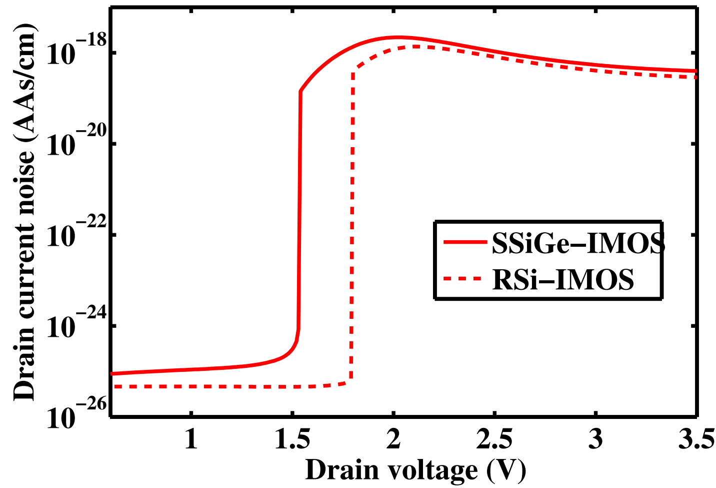


Figure 5.29: Comparison of the drain current noise between the SSiGe-IMOS and RSi-IMOS over different drain voltages ($V_G = 1.35V$) at a frequency of 100 MHz.

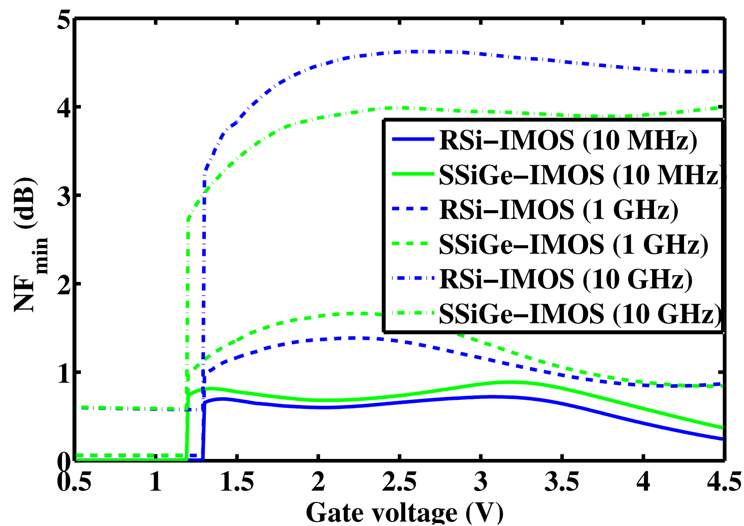


Figure 5.30: Comparison of the minimum noise figure between the SSiGe-IMOS and RSi-IMOS versus gate voltage ($V_D = 2V$).

simulations. The results show that it is possible to improve the characteristic of the vertical IMOS transistor with a strained SiGe-layer. However the position of this layer inside the device must be chosen very carefully since such a layer can have various consequences. The avalanche multiplication factor is not increased by strained SiGe, therefore a layer inside the drain

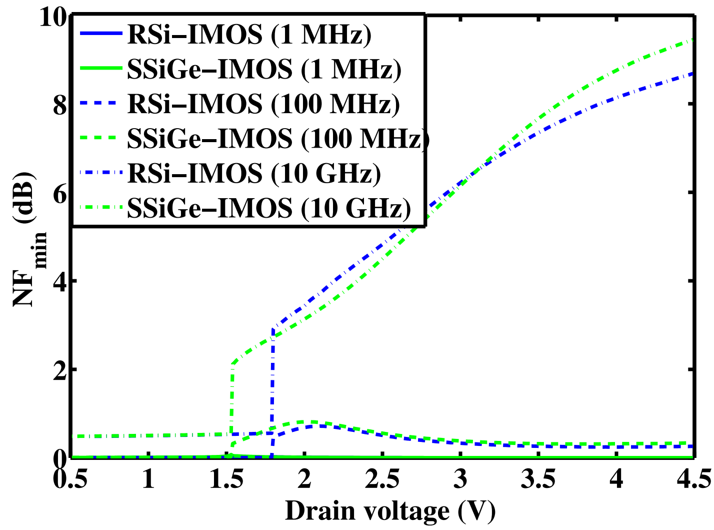


Figure 5.31: Comparison of the minimum noise figure between the SSiGe-IMOS and RSi-IMOS versus drain voltage ($V_G = 1.35$V).

region has little influence. But if the layer is incorporated in the p-body, it can reduce the threshold of the turn-on of the IMOS.

Noise performance of the IMOS devices has been investigated. Generally, the spectral intensity of the drain current fluctuations in these IMOS is around two orders of magnitude larger than the case without II. Therefore, the noise performance should be considered carefully for any application of this IMOS device.

# Chapter 6

# Conclusions

## 6.1 Conclusions

A model for calculating impact ionization rates for strained Si and strained SiGe has been developed. The model follows the approach of the constant matrix element approximation. The calculation is done on the basis of exact energy and momentum conservation together with realistic full band structures. The wave-vector space is discretized with a very fine grid [32], which has never been done before to my knowledge. A good agreement between Full-band Monte Carlo simulations and experimental data for II coefficient and quantum yield of relaxed Si is obtained. For strained Si and SiGe, the II rates are shown to be converged for the fine **k** space grid, which still can be processed with current CPU capacity. This fact makes the results reliable in situations where there is no experimental data for comparison.

From this model, the characteristics of II for uniaxially/biaxially strained Si and biaxially strained SiGe have been explored. It is found that the band gap plays an important role only at low energies ($< 2.0$ eV), whereas the overall II rate depends on the band structure of the strained material. Therefore it is common that the II rate of strained Si and SiGe is remarkably higher than that of relaxed Si only at low energies. The difference at high energies is not significant. For strained SiGe, alloy scattering cancels the gain from the higher II rate, hence resulting in a much smaller II coefficient compared to relaxed Si.

Such an II model has been integrated into the Full-band Monte Carlo simulator for simulating relaxed-Si-IMOS and strained-SiGe-IMOS devices. The II results from FB-MC simulations can be used to calibrate the Hydrodynamic model which is more efficient for simulating the devices especially for getting I-V curves. This hierarchical simulation approach, which is a combi-

nation of FB-MC and Hydrodynamic simulations, can offer insight into this device, which cannot be obtained by either method alone.

The developed simulator has been used to thoroughly investigate the vertical IMOS in order to optimize it. Strain engineering has been applied as an attempt to improve its performance. A thin layer of biaxially strained SiGe, of from 20 - 50 nm, is placed inside the device. However such a thin layer still suffers severely from alloy scattering. II coefficients in this strain layer are, hence, smaller than those in relaxed Si. As a result, the avalanche multiplication factor is not increased by strained SiGe. The possibility of applying biaxially strained SiGe to improve the performance of vertical IMOS seems to be very limited.

Finally, the noise performance for the vertical IMOS has been investigated. Although this noise is smaller than in the lateral IMOS, it is still more than two orders of magnitude larger than without II. This fact limits the application of the IMOS.

## 6.2 Future works

In the developed model, II rates are adjusted by a constant factor which is obtained by fitting a pair of II coefficient and quantum yield from simulations with the results from experiments. Up to now, such a pair of experimental data is only available for electrons. For hole-initiated II, no experimental data is available for fitting. Therefore, in the scope of this dissertation, II rates for holes have not been presented. This missing result should be calculated in a future work, when the experimental data is available or when there is another more suitable model for calculating such rates.

Throughout this dissertation, II rates of strained Si and SiGe are fitted by the global fitting parameters obtained from the relaxed Si case. Due to the shortage of experimental data, no comparison or proof of the results of this model for strained material is done. This is another future work.

Furthermore, only biaxially strained SiGe IMOS has been investigated because the Hydrodynamic model cannot handle uniaxial strain. Future work should investigate the case of a uniaxial strained IMOS.

# Bibliography

[1] *GALENE III User's Guide.* BST, Technische Universität Braunschweig, 3.2 edition, 2006.

[2] U. Abelein, A. Assmuth, P. Iskra, M. Reinl, M. Schlosser, T. Sulima, and I. Eisele. Vertical 40 nm impact ionization MOSFET (I-MOS) for high temperature applications. In *Proc. 26th International Conference on Microelectronics (MIEL '2008)*, pages 287–290, Niš (Serbia), 2008.

[3] U. Abelein, A. Assmuth, P. Iskra, M. Schindler, T. Sulima, and I. Eisele. Doping profile dependence of the vertical impact ionization MOSFET's (I-MOS) performance. *Solid–State Electron.*, 51:1405–1411, 2007.

[4] U. Abelein, M. Born, K.K. Bhuwalka, M. Schindler, M. Schlosser, T. Sulima, and I. Eisele. Improved reliability by reduction of hot-electron damage in the vertical impact-ionization MOSFET (I-MOS). *IEEE Electron Device Lett.*, 28(1):65–67, Jan. 2007.

[5] U. Abelein, M. Born, K.K. Bhuwalka, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele. A novel vertical impact ionization MOSFET (I-MOS) concept. In *Proc. 25th International Conference on Microelectronics (MIEL '2006)*, pages 127–129, Niš (Serbia), 2006.

[6] S. Banerjee, W. Richardson, J. Coleman, and A. Chatterjee. A new three-terminal tunnel device. *IEEE Electron Device Lett.*, 8(8):347–349, 1987.

[7] K. K. Bhuwalka, J. Schulze, and I. Eisele. Scaling the vertical tunnel FET with tunnel bandgap modulation and gate workfunction engineering. *IEEE Trans. Electron Devices*, 52(5):909–917, May 2005.

[8] K. K. Bhuwalka, S. Sedlmaier, A. K. Ludsteck, C. Tolksdorf, J. Schulze, and I. Eisele. Vertical tunnel field-effect transistor. *IEEE Trans. Electron Devices*, 51(2):279–282, 2004.

[9] G. L. Bir and G. E. Pikus. *Symmetry and Strain-Induced Effects in Semiconductors*. John Wiley & Sons, New York, 1974.

[10] P. E. Bloechl, O. Jepsen, and O. K. Andersen. Improved tetrahedron method for brillouin-zone integrations. *Phys. Rev. B*, 49(23):16223–16234, 1994.

[11] F. Bonani and G. Ghione. Generation–recombination noise modelling in semiconductor devices through population or approximate equivalent current density fluctuations. *Solid–State Electron.*, 43:285–295, 1999.

[12] K. Boucart and A. M. Ionescu. Double-gate tunnel FET with high-k gate dielectric. *IEEE Trans. Electron Devices*, 54(7):1725–1733, 2007.

[13] R. Brunetti, C. Jacoboni, F. Nava, L. Reggiani, G. Bosman, and R. J. J. Zijlstra. Diffusion coefficient of electrons in silicon. *J. Appl. Phys.*, 52:6713–6722, 1981.

[14] Jeff D. Bude and Marco Mastrapasqua. Impact ionization and distribution functions in sub-micron nMOSFET technologies. *IEEE Electron Device Lett.*, 16(10):439–441, 1995.

[15] E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely. Impact ionization in silicon. *Appl. Phys. Lett.*, 62:3339–3341, 1993.

[16] D. M. Caughey and R. E. Thomas. Carrier mobilities in silicon empirically related to doping and field. *Proc. IEEE*, 55:2192–2193, 1967.

[17] C. Charbuillet, E. Dubois, S. Monfray, P. Bouillon, and T. Skotnicki. Fabrication and analysis of CMOS fully-compatible high conductance impact-ionization MOS (I-MOS) transistors. In *Proc. ESSDERC*, pages 299–302, 2006.

[18] C. Charbuillet, S. Monfray, E. Dubois, P. Bouillon, F. Judong, and T. Skotnicki. High current drive in ultra-short impact-ionization MOS (I-MOS) devices. In *IEEE Tech. Dig. IEDM*, pages 1–4, 2006.

[19] D. Chattopadhyay and H. J. Queisser. Electron scattering by ionized impurities in semiconductors. *Rev. Mod. Phys.*, 53:745–768, 1981.

[20] J. R. Chelikowsky and M. L Cohen. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. *Phys. Rev. B*, 14:556–582, 1976.

[21] W. Y. Choi. Effect of device parameters on the breakdown voltage of impact-ionization metal-oxide-semiconductor devices. *Japanese Journal of Applied Physics*, 48:040203–040203–3, 2009.

[22] W. Y. Choi, J. Y. Song, J. D. Lee, Y. J. Park, and B.-G. Park. 100-nm n-/p-channel I-MOS using a novel self-aligned structure. *IEEE Electron Device Lett.*, 26(4):261–263, 2005.

[23] W. Y. Choi, J. Y. Song, J. D. Lee, Y. J. Park, and B.-G. Park. 70-nm impact-ionization metal-oxide-semiconductor (I-MOS) devices integrated with tunneling field-effect transistors (TFETs). *IEEE Tech. Dig. IEDM*, pages 955–958, 2005.

[24] M. L Cohen and T.K. Bergstresser. Band structure and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures. *Phys. Rev.*, 141:789–796, 1966.

[25] M. L. Cohen and J. R. Chelikowsky. *Electronic Structure and Optical Properties of Semiconductors.* Springer, New York, 2nd edition, 1989.

[26] L. Colombo, R. Resta, and S. Baroni. Valence-band offsets at strained Si/Ge interfaces. *Phys. Rev. B*, 44:5572–5579, 1991.

[27] J. D. Cressler. SiGe HBT technology: A new contender for Si–based RF and microwave circuit applications. *IEEE Trans. Microwave Theory and Techniques*, 46(5):572–589, 1998.

[28] J. D. Cressler and G. Niu. *Silicon-Germanium Heterojunction Bipolar Transistors.* Artech House, Boston, London, 2003.

[29] C. G. Van de Walle and R. M. Martin. Theoretical calculations of heterojunction discontinuities in the Si/Ge system. *Phys. Rev. B*, 34:5621–5634, 1986.

[30] D. J. DiMaria, T. N. Theis, J. R. Kirtley, F. L. Pesavento, and D. W. Dong. Electron heating in silicon dioxide and off-stoichiometric silicon dioxide films. *J. Appl. Phys.*, 57:1214–1238, 1985.

[31] T. V. Dinh, S.-M. Hong, and C. Jungemann. Improving the high frequency performance of SiGe HBTs by a global additional uniaxial stress. In *International SiGe Technology and Device Meeting*, volume 5, pages –, Stockholm, Sweden, 2010.

[32] T. V. Dinh, R. Kraus, and C. Jungemann. Investigation of the performance of strained-SiGe vertical IMOS-transistors. In *Proc. ESSDERC*, pages 165–168, 2009.

[33] T. V. Dinh, R. Kraus, and C. Jungemann. Investigation of the performance of strained-SiGe vertical IMOS-transistors. *Solid–State Electron.*, -:–, 2010.

[34] V. Dobrovolsky, S. Pavljuk, V. Rossokhaty, and S. Cristoloveanu. Silicon-on-insulator control impact-ionization-avalanche transistor. *Appl. Phys. Lett.*, 88:073502, 2006.

[35] W. Fawcett, A. D. Boardman, and S. Swain. Monte Carlo determination of electron transport properties in gallium arsenide. *J. Phys. Chem. Solids*, 31:1963–1990, 1970.

[36] M. V. Fischetti. Effect of the electron-plasmon interaction on the electron mobility in silicon. *Phys. Rev. B*, 44:5527–5534, 1991.

[37] M. V. Fischetti and S. E. Laux. Monte Carlo analysis of electron transport in small semiconductor devices including band–structure and space–charge effects. *Phys. Rev. B*, 38:9721–9745, 1988.

[38] M. V. Fischetti, S. E. Laux, and E. Crabbe. Understanding hot-electron transport in silicon devices: Is there a shortcut? *J. Appl. Phys.*, 78:1058–1087, 1995.

[39] M. V. Fischetti, N. Sano, S. E. Laux, and K. Natori. Full-band Monte Carlo simulation of high-energy transport and impact ionization of electrons and holes in ge, si and gaas. In *Proc. SISPAD*, volume 1, pages 43–44, Tokyo, 1996.

[40] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr. A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors. In *IEEE Tech. Dig. IEDM*, pages 978–980, 2003.

[41] K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer. I-MOS: A novel semiconductor device with a substhreshold slope lower than kt/q. In *IEEE Tech. Dig. IEDM*, pages 289–292, 2002.

[42] K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer. Impact ionization MOS (I-MOS)-Part I: device and circuit simulations. *IEEE Trans. Electron Devices*, 52(1):69 – 76, 2005.

[43] K. Gopalakrishnan, R. Woo, C. Jungemann, P. B. Griffin, and J. D. Plummer. Impact ionization MOS (I-MOS)-Part II: experimental results. *IEEE Trans. Electron Devices*, 52(1):77 – 84, 2005.

[44] T. Grasser, T-W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE*, 91(2):251–274, 2003.

[45] M. G. Gray, T. E. Booth, T. J. T. Kwan, and C. M. Snell. A multi-comb variance reduction scheme for Monte Carlo semiconductor simulators. *IEEE Trans. Electron Devices*, 45(4):918–924, 1998.

[46] J. M. Hammarsley and D. C. Handscomb. *Monte Carlo Methods*. Methuen/Chapman and Hall, London, 1964.

[47] J. W. Harrison and J. R. Hauser. Theoretical calculations of electron mobility in ternary III-V compounds. *J. Appl. Phys.*, 47:292–300, 1976.

[48] Jack M. Higman, Karl Hess, Chang G. Hwang, and Robert W. Dutton. Coupled Monte Carlo-drift diffusion analysis of hot-electron effects in MOSFET's. *IEEE Trans. Electron Devices*, 36(5):930–937, 1989.

[49] R.W. Hockney and J.W. Eastwood. *Computer Simulation Using Particles*. Institute of Physics Publishing, Bristol, Philadelphia, 1988.

[50] C. Hu, S.C. Tam, F.-C. Hsu, P.-K. Ko, T.-Y. Chan, and K. W. Terrill. Hot-electron induced MOSFET degradation - model, monitor, and improvement. *IEEE Trans. Electron Devices*, 32(2):375–385, 1985.

[51] International Roadmap Committee. The international technology roadmap for semiconductors. `public.itrs.net`, 2009.

[52] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, 2nd edition, 1975.

[53] C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation*. Springer, Wien, 1989.

[54] C. Jacoboni and L. Reggiani. The Monte Carlo method for the solution of charge transport in semiconductors with application to covalent materials. *Rev. Mod. Phys.*, 55:645–705, 1983.

[55] W. Jin, P. C. Chan, S. K. H. Fung, and P. K. Ko. Shot-noise-induced excess low-frequency noise in floating-body partially depleted SOI MOSFET's. *IEEE Trans. Electron Devices*, 46(7):1180–1185, 1999.

[56] C. Jungemann, M. Bartels, S. Keith, and B. Meinerzhagen. Efficient methods for Hall factor and transport coefficient evaluation for electrons and holes in Si and SiGe based on a full-band structure. In *Proc. IWCE International Workshop on Computational Electronics*, pages 104–107, Osaka (Japan), 1998.

[57] C. Jungemann, S. Decker, R. Thoma, W.-L. Engl, and H. Goto. Phase space multiple refresh: a general purpose statistical enhancement technique for Monte Carlo device simulation. *IEEE J. Tech. Comp. Aided Design*, (2), 1997.

[58] C. Jungemann, S. Keith, F. M. Bufler, and B. Meinerzhagen. Effects of band structure and phonon models on hot electron transport in silicon. *Electrical Engineering*, 79:99–101, 1996.

[59] C. Jungemann and B. Meinerzhagen. *Hierarchical Device Simulation: The Monte-Carlo Perspective*. Computational Microelectronics. Springer, Wien, New York, 2003.

[60] C. Jungemann and B. Meinerzhagen. Numerical simulation of RF noise in Si devices. In *Proc. SISPAD*, pages 87–94, 2006.

[61] C. Jungemann, B. Neinhüs, and B. Meinerzhagen. Investigation of the local force approximation in numerical device simulation by full-band Monte Carlo simulation. *VLSI Design*, 13(1-4):281–285, 2001.

[62] C. Jungemann, B. Neinhüs, and B. Meinerzhagen. Hydrodynamic modeling of RF noise for silicon-based devices. *Internationa Journal of high speed electronics*, 13(3):823–848, 2003.

[63] C. Jungemann, R. Thoma, and W. L. Engl. A soft threshold lucky electron model for efficient and accurate numerical device simulation. *Solid–State Electron.*, 39:1079–1086, 1996.

[64] C. Jungemann, S. Yamaguchi, and H. Goto. On the accuracy and efficiency of substrate current calculations for sub-$\mu$m n-MOSFET's. *IEEE Electron Device Lett.*, 17(10):464–466, 1996.

[65] Y. Kamakura, H. Mizuno, M. Yamaji, M. Morifuji, K. Taniguchi, C. Hamaguchi, T. Kunikiyo, and M. Takenaka. Impact ionization

model for full band Monte Carlo simulation. *J. Appl. Phys.*, 75:3500–3506, 1994.

[66] E. O. Kane. Electron scattering by pair production in silicon. *Phys. Rev.*, 159(3):624–631, 1967.

[67] E. Kasper and K. Lyutovich. *Properties of silicon germanium and SiGe:Carbon*. INSPEC, London, 2000.

[68] A. Khakifirooz and D. A. Antoniadis. MOSFET performance scaling - part I: Historical trends. *IEEE Trans. Electron Devices*, 55:1391–1400, 2008.

[69] A. Khakifirooz and D. A. Antoniadis. MOSFET performance scaling - part II: Future directions. *IEEE Trans. Electron Devices*, 55:1401–1408, 2008.

[70] M. Khater, J.-S. Rieh, T. Adam, A. Chinthakindi, J. Johnson, R. Krishnasamy, M. Meghelli, F. Pagette, D. Sanderson, C. Schnabel, K. T. Schonenberg, P. Smith, K. Stein, A. Stricker, S.-J. Jeng, D. Ahlgren, and G. Freeman. SiGe HBT technology with $f_{max}/f_T = 350/300$GHz and gate delay below 3.3 ps. In *IEEE Tech. Dig. IEDM*, pages 247–250, 2004.

[71] M. Kimura and T. Matsudate. A new type of Schottky tunnel transistor. *IEEE Electron Device Lett.*, 15(10):412–414, 1994.

[72] J. Kolnik, Y. Wang, I. H. Oguzman, and K. F. Brennan. Theoretical investigation of wave-vector-depedent analytical and numerical formulations of the interband impact-ionization transition rate for electrons in bulk silicon and GaAs. *J. Appl. Phys.*, 76:3542–3551, 1994.

[73] H. Kosina and G. Kaiblinger-Grujin. Ionized-impurity scattering of majority electrons in silicon. *Solid–State Electron.*, 42:331–338, 1998.

[74] H. Kosina, M. Nedjalkov, and S. Selberherr. Theory of the Monte Carlo method for semiconductor device simulation. *IEEE Trans. Electron Devices*, 47(10):1898–1908, 2000.

[75] R. Kraus and C. Jungemann. Investigation of the vertical IMOS-transistor by device simulation. In *Proc. ULIS, International Conference on Ultimate Integration of Silicon*, volume 10, pages 281–284, 2009.

[76] T. Kurosawa. Monte Carlo calculation of hot electron problems. *J. Phys. Soc. Jap.*, 21:424–426, 1966.

[77] D. B. Laks, G. F. Neumark, A. Hangleiter, and S. T. Pantelides. Theory of interband auger recombination in n-type silicon. *Phys. Rev. Lett.*, 61:1229–1232, 1988.

[78] M. Levinstein, S. Rumyantsev, and M. Shur. *Handbook Series on Semiconductor Parameters*. World Scientific, London, 1999.

[79] K. K. Likharev. Layered tunnel barriers for nonvolatile memory devices. *Appl. Phys. Lett.*, 73:2137–2139, 1998.

[80] J-S. Lim, S. E. Thompson, and J. G. Fossum. Comparison of threshold-voltage shifts for uniaxial and biaxial tensile-stressed n-MOSFETs. *IEEE Electron Device Lett.*, 25(11):731–733, 2004.

[81] O. Madelung. *Introduction to Solid State Theory*. Springer, Berlin, 1978.

[82] W. Maes, K. de Meyer, and R. Van Overstraeten. Impact ionization in silicon: A review and update. *Solid–State Electron.*, 33:705–718, 1990.

[83] R. K. Mains, G. I. Haddad, and P. A. Blakey. Simulation of GaAs IMPATT diodes including energy and velocity transport equations. *IEEE Trans. Electron Devices*, 30(10):1327–1337, 1983.

[84] F. Mayer, C. Le Royer, G. Le Carval, L. Clavelier, and S. Deleonibus. Static and dynamic TCAD analysis of IMOS performance : From the single device to the circuit. *IEEE Trans. Electron Devices*, 53:1852–1857, 2006.

[85] F. Mayer, C. Le Royer, G. Le Carval, C. Tabone, L. Clavelier, and S. Deleonibus. Co-integration of 2 mv/dec subthreshold slope impact ionization MOS (I-MOS) with CMOS. In *Proc. ESSDERC*, pages 303–306, 2006.

[86] E. Morifuji, T. Yoshida, M. Kanda, S. Matsuda, S. Yamada, and F. Matsuoka. Supply and threshold-voltage trends for scaled logic and SRAM MOSFETs. *IEEE Trans. Electron Devices*, 53:1427–1432, 2006.

[87] M. Nedjalkov and P. Vitanov. Iteration approach for solving the Boltzmann equation with the Monte Carlo method. *Solid–State Electron.*, 32:893–896, 1989.

[88] C. Onal, R. Woo, H.-Y.S. Koh, P.B. Griffin, and J.D. Plummer. A novel depletion-IMOS (DIMOS) device with improved reliability and reduced operating voltage. *IEEE Electron Device Lett.*, 30(1):64–67, 2009.

[89] H.-J. Peifer. Monte–Carlo Simulation des Hochenergietransports von Elektronen in submikron MOS–Strukturen. Doctor thesis, RWTH Aachen, Aachen, 1992. Augustinus Buchhandlung.

[90] F. H. Pollak and M. Cardona. Piezo-electroreflectance in Ge, GaAs, and Si. *Phys. Rev.*, 172:816–837, 1968.

[91] P. J. Price. Monte Carlo calculation of electron transport in solids. *Semiconductors and Semimetals*, 14:249–309, 1979.

[92] R. van Overstraeten and H. de Man. Measurement of the ionization rates in diffused silicon p-n junctions. *Solid–State Electron.*, 13:583–608, 1970.

[93] M. M. Rieger and P. Vogl. Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates. *Phys. Rev. B*, 48:14276–14287, 1993.

[94] M. M. Rieger and P. Vogl. Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates. *Phys. Rev. B*, 50:8138, 1994. Erratum.

[95] D. J. Robbins, L. T. Canham, S. J. Barnett, A. D. Pitt, and P. Calcott. Near-band-gap photoluminescence from pseudomorphic $Si_{1-x}Ge_x$ single layers on silicon. *J. Appl. Phys.*, 71:1407–1414, 1992.

[96] R. Y. Rubinstein. *Simulation and the Monte Carlo method.* Wiley series in probability and mathematical statistics. John Wiley & Sons, New York, 1981.

[97] P. Sakalas, M. Ramonas, M. Schroeter, C. Jungemann, A. Shimukovitch, and W. Kraus. Impact ionization noise in SiGe HBTs: Comparison of device and compact modeling with experimental results. *IEEE Trans. Electron Devices*, 56(2):328–336, February 2009.

[98] E. Sangiorgi and M. R. Pinto. A semi-empirical model of surface scattering for Monte Carlo simulation of silicon n-MOSFET's. *IEEE Trans. Electron Devices*, 39(2):356–361, 1992.

[99] N. Sano and A. Yoshii. Impact ionization theory consistent with a realistic band structure of silicon. *Phys. Rev. B*, 45:4171–4179, 1992.

[100] N. Sano and A. Yoshii. Impact ionization rate near thresholds in Si. *J. Appl. Phys.*, 75:5102–5105, 1994.

[101] A. Savio, S. Monfray, C. Charbuillet, and T. Skotnicki. On the limitations of silicon for I-MOS integration. *IEEE Trans. Electron Devices*, 56(5):1110–1117, 2009.

[102] F. Schäffler. High–mobility Si and Ge structures. *Semicond. Sci. Technol.*, 12:1515–1549, 1997.

[103] D. L. Scharfetter. Measured dependence of lifetime upon defect density and temperature in depletion layers of epitaxial silicon diodes. In *Solid-State Dev. Res. Conf.*, Santa Barbara, USA, 1967.

[104] A. Schenk. *Advanced Physical Models for Silicon Device Simulation*. Computational Microelectronics. Springer, Wien, New York, 1998.

[105] W. Shockley and W. W. Hooper. In *Western Electronic Show and Convention/Los Angeles*, 1964.

[106] E. Simoen, A. Mercha, C. Claeys, and N. B. Lukyanchikova. Low-frequency noise in silicon-on-insulator devices and technologies. *Solid–State Electron.*, 51:16–37, 2007.

[107] J. W. Slotboom, G. Streutker, G. J. T. Davids, and P. B. Hartog. Surface impact ionization in Silicon devices. In *IEDM*, pages 494–497, 1987.

[108] S. Smirnov, H. Kosina, and S. Selberherr. Substrate orientation-dependence of electron mobility in strained SiGe layers. In *Proc. SISPAD*, pages 55–58, 2003.

[109] Y. Sun, S. E. Thompson, and T. Nishida. Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors. *J. Appl. Phys.*, 101, 2007.

[110] S. M. Sze. *Physics of Semiconductors Devices*. Wiley, New York, 1981.

[111] S. M. Sze. *Semiconductor Devices: Physics and Technology*. Wiley, New York, 1985.

[112] I. Takayanagi, K. Matsumoto, and J. I. Nakamura. Measurement of electron impact ionization coefficient in bulk silicon under a low-electric field. *J. Appl. Phys.*, 72:1989–1992, 1992.

[113] E. Takeda, H. Matsuoka, Y. Igura, and S. Asai. A band to band tunneling MOS device ($B^2T - MOSFET$). In *IEEE Tech. Dig. IEDM*, pages 402–405, 1988.

[114] E. Takeda and N. Suzuki. An empirical model for device degradation due to hot-carrier injection. *IEEE Electron Device Lett.*, 4(4):111–113, 1983.

[115] J. Y. Tang and K. Hess. Impact ionization of electrons in silicon (steady state). *J. Appl. Phys.*, 54:5139–5144, 1983.

[116] R. Thoma, A. Emunds, B. Meinerzhagen, H. J. Peifer, and W. L. Engl. Hydrodynamic equations for semiconductors with nonparabolic bandstructures. *IEEE Trans. Electron Devices*, 38(6):1343–1352, 1991.

[117] R. Thoma, H. J. Peifer, W. L. Engl, W. Quade, R. Brunetti, and C. Jacoboni. An improved impact–ionization model for high–energy electron transport in Si with Monte Carlo simulation. *J. Appl. Phys.*, 69:2300–2311, 1991.

[118] E.-H. Toh, G. H. Wang, L. Chan, G.-Q Lo, G. Samudra, and Y.-C. Yeo. Strain and materials engineering for the I-MOS transistor with an elevated impact-ionization region. *IEEE Trans. Electron Devices*, 54(10):2778–2785, October 2007.

[119] E.-H. Toh, G. H. Wang, L. Chan, G. Samudra, and Y.-C. Yeo. Reduction of impact-ionization threshold energies for performance enhancement of complementary impact-ionization metal-oxide-semiconductor transistors. *Appl. Phys. Lett.*, 91:153501–153501–3, 2007.

[120] E.-H. Toh, G. H. Wang, L. Chan, G. Samudra, and Y.-C. Yeo. A double-spacer I-MOS transistor with shallow source junction and lightly doped drain for reduced operating voltage and enhanced device performance. *IEEE Electron Device Lett.*, 29(2):189–191, 2008.

[121] J. R. Tucker, C. Wang, and P. S. Carney. Silicon field-effect transistor based on quantum tunneling. *Appl. Phys. Lett.*, 65:618–620, 1994.

[122] E. Ungersboeck. Advanced modeling of strained cmos technology. Doctor thesis, TU Wien, Wien, 2007.

[123] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr. The effect of general strain on the band structure and electron mobility of silicon. *IEEE Trans. Electron Devices*, 54(9):2183–2190, 2007.

[124] A. v. Schwerin, W. Bergner, and H. Jacobs. Self-consistent simulation of hot-carrier damage enhanced gate induced drain leakage. In *IEEE Tech. Dig. IEDM*, pages 543–546, 1992.

[125] J. Weber and M. I. Alonso. Near–band–gap photoluminescence of Si–Ge alloys. *Phys. Rev. B*, 40:5683–5693, 1989.

[126] H.-S. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. J. Wesler. Nanoscale CMOS. *Proc. IEEE*, 87(4):407–410, 1999.

[127] K. Yeom, J. Hinckley, and J. Singh. Theoretical study on threshold energy and impact ionization coefficient for electrons in $Si_{1-x}Ge_x$. *Appl. Phys. Lett.*, 64:2985–2987, 1994.

[128] P. Y. Yu and M. Cardona. *Fundamentals of Semiconductors: Physics and Materials Properties.* Springer, Berlin, 3rd edition, 2001.