Simulation of Quantum Effects in Nanoscale Devices

Andreas Schenk

Mathieu M. Luisier, Martin Frey, Aniello Esposito

1 ETH Zürich
2 Purdue University
Outline

- Introduction
- Quantum-mechanical confinement effects
  - quantum $V_T$-shift
  - comparison single, double, triple, and surround gate
- Simulation of quantum-ballistic ON-currents
  - Band structure, transport, electrostatics
  - ON-current for different channel orientations
  - Effect of surface roughness
- Simulation of tunneling-induced OFF-currents
  - Source-drain tunneling
  - Gate tunneling leakage
  - Band-to-band tunneling (GIDL)
- Incoherent scattering
- Conclusion
Introduction

Which quantum effects?

- source-to-drain tunneling
- band-to-band tunneling
- strain
- gate tunneling
- direct resonant
- confinement, quantization
  - quantum $V_T$ shift
  - QM mobility

Introduction
Quantum-mechanical confinement effects

Wave length of a free electron with energy $k_B T$ at 300K ≈ 8nm!

channel density profile

transfer characteristics

asymmetrical n$^+$p$^+$ DGSOI nMOSFET, $t_{Si}=5$ nm, $t_{ox}=1.5$ nm, $L_G=90$ nm
A ‘quantum potential’ $\Lambda$ is introduced in the classical formula of the density:

$$n(\vec{R}) = N_c \exp \left[ \beta \left( E_{F,n}[\vec{R}] - E_c[\vec{R}] - \Lambda[\vec{R}] \right) \right]$$

The classical current equation reads:

$$\vec{e}[\vec{R}] = -\mu k_B T \nabla n(\vec{R}) - \mu n(\vec{R}) \nabla \left( E_c[\vec{R}] + \Lambda[\vec{R}] \right)$$

$\Lambda$ follows by equating the density $n[\vec{R}]$ with the expression

$$n(z) = \frac{1}{\beta \pi \hbar^2} \sum_{j,\nu} \left| \psi_j^{(\nu)}(z) \right|^2 m_{xy}^{(\nu)}(z) \exp \left[ \beta \left( E_{F,n}(z) - E_j^{(\nu)} \right) \right]$$
Boundary condition at the ends of the domain \([z^-,z^+]\), defining the ‘quantum box’: 

\[ \frac{\psi_j^{(\nu)'}}{\psi_j^{(\nu)}} = \pm \sqrt{2m_z^{(\nu)}|E_j^{(\nu)} - E_c|/\hbar} \]

Full Newton impractical, therefore approximation

\[ \frac{\partial n_i^{qm}}{\partial \Phi_j} \approx -\delta_{ij} \frac{\partial n_i^{qm}}{\partial E_{F,i}} \]

\(\Lambda\) is given by the PDE

\[ \Lambda = -\gamma \frac{\hbar^2}{12m} \left[ \nabla^2 \log n + \frac{1}{2} (\nabla \log n)^2 \right] \]

\[ = -\gamma \frac{\hbar^2 \nabla^2 \sqrt{n}}{6m \sqrt{n}} \]
Based on method of moments for Liouville equation

\[ i\hbar \partial_t \rho = [\mathcal{H}, \rho] \]

Hierarchy closed by replacing higher-order derivatives of density matrix \( \rho \) by approx. for equilibrium density matrix \( \rho_0 \)

Assumptions: equilibrium density, isotropy of effective mass, \( \frac{\delta \Phi}{k_B T} \ll 1 \) (Born approx.)

Generalization for device simulation:
- DOS mass \( \rightarrow m \)
- non-equilibrium density \( \rightarrow n \)
- non-perturbative formulation of \( \Lambda \) with the smoothed potential \( \tilde{\Phi} = E_c + \Phi_m + \Lambda \)

\[ \Lambda = -\gamma \frac{\hbar^2}{12 m} \left\{ \nabla^2 \left( \beta E_{F,n} - \beta \tilde{\Phi} \right) + \frac{1}{2} \left[ \nabla \left( \beta E_{F,n} - \beta \tilde{\Phi} \right) \right]^2 \right\} \]

\( \Lambda \) is new system variable, coupled Newton

DG is multi-dimensional, pre-factor \( \gamma \) is “universal” (3.6 for Si, if \( m = m_{\text{DOS}} \))
quantum $V_T$-shift

MOS (with poly) capacitor CV curves

$t_{ox}=1.5\text{nm}$

$t_{ox}=3\text{nm}$

$N_{\text{poly}}=1\times10^{20}\text{ cm}^{-3}$, $A_G=1\ \mu\text{m}^2$
quantum V_T-shift

MOS (with poly) capacitor

eigenenergies

\[
E (\text{eV})
\]

position (nm)

\[
|\Psi|^2 (\text{nm}^{-1})
\]

position (nm)

\[
N_A=5\times10^{17} \text{ cm}^{-3}, \quad t_{ox}=3 \text{ nm}, \quad V_G=3 \text{ V}
\]
Quantum depletion at poly-SiO₂ interface

compensation of quantum shift at threshold for high poly doping (~1e20 cm⁻³)!

N_{poly}=10^{20} \text{ cm}^{-3}

N_A=5e17 \text{ cm}^{-3}, t_{ox}=3 \text{ nm}, A_G=1 \mu\text{m}^2
quantum $V_T$-shift

Electron density profile at poly-SiO$_2$ interface

- a “quantum dipole” forms as the electron waves are repelled from the poly-SiO$_2$ interface
- poly quantum depletion disappears with rising $V_G$ (smoother poly band edge curvature)
quantum $V_T$-shift

Effect on CV curves

- strength of the quantum dipole depends on doping level within the first few nanometers
- no effect on CV, if $N_{\text{poly}} < 1\text{e}19 \text{ cm}^{-3}$ at the interface
quantum $V_T$-shift

Effect of exchange-correlation potential on CV curves

$$
e_X = -\frac{q^2}{4\pi\varepsilon_0\varepsilon_r} (3n)^{1/3} = -0.909 \cdot 10^{-8} n^{1/3} \ [eV]$$

$$
e_C = -1.224 \cdot 10^{-5} \left[ \ln \left( 1 + \frac{n}{2.375 \cdot 10^{12}} \right) \right]^2 \ [eV]$$

LDA $\varepsilon_{xc}$ ($\varepsilon_c$ is continuous fit to Perdew-Zunger theory)

- Many-body effect of exchange-correlation shows up as negative $V_T$-shift (band gap narrowing)
- Image-force already taken into account by boundary condition for Hartree potential (Poisson equation)
comparison single, double, triple, and surround gate

3D quantum-drift-diffusion (QDD) simulation of a nano-scale MOSFET

corner resolution ~ 0.5 Å

$L_G = 20$ nm

source

gate

drain
Influence of mesh refinement

density @ Vgs=1V

density (parallel cut)

density (diagonal cut)
Influence of mesh refinement

comparison single, double, triple, and surround gate
Transfer characteristics

CPU time ~ 1 week

assumption: isotropic, classical mobility

• no corner effect (FD, no channel doping, 3x3 nm wire)
• only little improvement from double $\rightarrow$ surround (gate overlap, e.g. triple gate is an effective Π-gate)
Transfer characteristics (contd.)

- quantum $V_T$-shift of ~ 64 mV, independent of gate configuration
- almost perfect shift on $V_{GS}$-axis $\rightarrow$ quantum $V_T$-shift can be translated into work function difference
Si nanowires (NWs) are ‘post-CMOS’ candidates, (transistors and connectors)

transport in Si NWs is often believed to become “ballistic” (which is wrong)

all produced Si NWs have cross sections > 5 x 5 nm² => no full quantum transport simulation necessary

if (in the future) cross sections < 5 x 5 nm², strong confinement => band structure effects become important

predictive simulations then require: accurate band structure model, quantum transport solver (OBCs), self-consistent electrostatics
Band structure

- $sp^3d^5s^*$ tight-binding method
- bulk band structure exactly reproduced
- inclusion of strain, defects, surface roughness possible
- extension to nanostructures straight-forward
- gate tunneling and b2b tunneling possible
- high computational effort required for nanostructures, since 10 bands involved without spin, 20 bands with spin
- *bulk* TB parameters, no lattice relaxation

Example: Full Band Tri-Gate FinFET with ~7500 atoms

**Atomistic description** of a [100] nanowire with 2.1 x 2.1 nm$^2$ cross section

$tox = 1$ nm

**TB band structure**
On-site and two-center integrals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Si</th>
<th>Ge</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_s$</td>
<td>−2.15168</td>
<td>−1.95617</td>
</tr>
<tr>
<td>$E_p$</td>
<td>4.22925</td>
<td>5.30970</td>
</tr>
<tr>
<td>$E_{s^*}$</td>
<td>19.11650</td>
<td>19.29600</td>
</tr>
<tr>
<td>$E_d$</td>
<td>13.78950</td>
<td>13.58060</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.01989</td>
<td>0.10132</td>
</tr>
<tr>
<td>$ss\sigma$</td>
<td>−1.95933</td>
<td>−1.39456</td>
</tr>
<tr>
<td>$s^<em>s^</em>\sigma$</td>
<td>−4.24135</td>
<td>−3.56680</td>
</tr>
<tr>
<td>$ss^*\sigma$</td>
<td>−1.52230</td>
<td>−2.01830</td>
</tr>
<tr>
<td>$sp\sigma$</td>
<td>3.02562</td>
<td>2.73135</td>
</tr>
<tr>
<td>$s^*p\sigma$</td>
<td>3.15565</td>
<td>2.68638</td>
</tr>
<tr>
<td>$sd\sigma$</td>
<td>−2.28485</td>
<td>−2.64779</td>
</tr>
<tr>
<td>$s^*d\sigma$</td>
<td>−0.80993</td>
<td>−1.12312</td>
</tr>
<tr>
<td>$pp\sigma$</td>
<td>4.10364</td>
<td>4.28921</td>
</tr>
<tr>
<td>$pp\pi$</td>
<td>−1.51801</td>
<td>−1.73707</td>
</tr>
<tr>
<td>$pd\sigma$</td>
<td>−1.35554</td>
<td>−2.00115</td>
</tr>
<tr>
<td>$pd\pi$</td>
<td>2.38479</td>
<td>2.10953</td>
</tr>
<tr>
<td>$dd\sigma$</td>
<td>−1.68136</td>
<td>−1.32941</td>
</tr>
<tr>
<td>$dd\pi$</td>
<td>2.58880</td>
<td>2.56261</td>
</tr>
<tr>
<td>$dd\delta$</td>
<td>−1.81400</td>
<td>−1.95120</td>
</tr>
</tbody>
</table>

- Optimized to reproduce bulk band structure and effective masses
- 4 different orbital types ($s$, $p$, $d$, $s^*$), => 4 on-site energy parameters ($E_s$, $E_p$, $E_{s^*}$, $E_d$)
- Spin-orbit coupling ($\lambda$)
- 14 matrix elements between orbitals with 3 different possible bonds ($\sigma$, $\pi$, and $\delta$). Symmetry + polarization not considered (Koster-Slater table)

Band structure

Energy integrals in terms of two-center integrals

Koster-Slater table
Phys. Rev. 94, 1498 (1954)
How does band structure change with increasing cross section?

- 2 nm x 2 nm
- 3 nm x 3 nm
- 4 nm x 4 nm
- 5 nm x 5 nm

band structure, transport, electrostatics
Wave functions are injected from the reservoirs and either reflected or transmitted to the other side. Band structure of reservoirs can be calculated because semi-infinite. At each energy all the k-states with positive (left) or negative (right) velocity are selected for injection.
Si [100]

GaAs [111]

Transport

band structure, transport, electrostatics
Output characteristics and transmission for [100] Full Band (solid lines) vs Effective Mass (dashed lines)
Transport

**WF formalism**

**Schrödinger Equation**

\[ H |\psi_E\rangle = E |\psi_E\rangle \]

**Tight-Binding ansatz for the wave function**

\[ \langle \mathbf{r} |\psi_E\rangle = \sum_{\sigma,ijk} C_{ijk}^\sigma(E) \phi_\sigma(\mathbf{r} - \mathbf{R}_{ijk}) \]

**Scattering Boundary Conditions** => *ordinary* eigenvalue problem!!

\[ M(E, A) \phi_{k(E)}(A) = \lambda(k(E)) \phi_{k(E)}(A) \]

**Final form of the problem in the Wave Function formalism**

\[ H_{tot} \cdot C_{p,n}^\sigma(k) = I_{0,p,n}(k) \]

- **Löwdin orbitals**
- **Orbital-coefficient vector**
- **Injection matrix**
Carrier and current density

\[ n(x, r_s) = \frac{1}{N_x} \sum_{n, p, \sigma} \sum_{i, R_s} \sum_{k} |C_{i,p,n}^\sigma(R_s, k)|^2 f(E_p, n(k) - \mu_p) \delta(x - x_i) \delta(r_s - R_s) \]

\[ J(r) = \sum_{i_1, i_2} i \frac{e}{2\hbar} \sum_{p, n_p} \frac{\Delta}{2\pi} \int dE \left( H_{i_1 i_2} C_{i_2, p, n_p}^* C_{i_1, p, n_p}^* - C_{i_1, p, n_p} C_{i_2, p, n_p}^* H_{i_2 i_1} \right) x \left( \frac{dE}{dk_{p, n_p}} \right)^{-1} f(E - \mu_p)(R_{i_2} - R_{i_1}) \delta(r - R_{i_1}) \]

Alternatively, in Landauer-Büttiker formula with transmission \( T(E) \)

\[ T(E) = \sum_{n, m} |C_{N_s+1, p=1, n}(k_m)|^2 \left( \frac{dE}{dk_m} \right) \left| \frac{dE}{dk_n} \right|^{-1} \]

\[ n(r) = -i \sum_j \int \frac{dE}{2\pi} G_{jj}^{<}(E) \delta(r - R_j) \]

\[ J(r) = \sum_{i_1} \sum_{i_2} \frac{e}{2} \left( H_{i_1 i_2} G_{i_2 i_1}^{<} - G_{i_1 i_2} G_{i_2 i_1}^{<} H_{i_2 i_1} \right) (r_{i_2} - r_{i_1}) \delta(r - r_{i_1}) \]
Transport

WF ↔ NEGF (if no incoherent scattering)

\[
G_{ij}^{\leq}(E) = \sum_p \sum_n \left\{ C_{i,p}(k_n) C^T_{j,p}(k_n) \right\} \left| \frac{dE(k_n)}{dk_n} \right|^{-1} f\left( E(k_n) - \mu_p \right) \]

When NEGF? In case of incoherent scattering.
When WF? Otherwise, because CPU time is greatly reduced!

Iterative solutions \((N = t_b \cdot N_A)\)

\((E - H - t_{10} \cdot g_{00}^{R} \cdot t_{01}) \cdot g_{00}^{R} = I\)

Generalized eigenvalue problem \((N = 2 \cdot t_b \cdot N_A)\)

\[A(E) \cdot C_n = \exp(ik_n(E)\Delta) \cdot B(E) \cdot C_n\]

Shift-and-invert + ordinary eigenvalue problem \((N < t_b \cdot N_A)\) \((PRB 74, 205323 (2006))\)

\[M(E) \cdot C_n = \lambda_n(k_n(E)) \cdot C_n\]
Transport

Benchmark example: 35 nm long [100] nanowire, 1 energy point

First task: Open Boundary Conditions

<table>
<thead>
<tr>
<th>$L_y \times L_z$ nm²</th>
<th>$t_b \times N_A$</th>
<th>Iterative Solver</th>
<th>Generalized EVP</th>
<th>Ordinary EVP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5×2.5</td>
<td>1810</td>
<td>197</td>
<td>506</td>
<td>7.2</td>
</tr>
<tr>
<td>2.9×2.9</td>
<td>2420</td>
<td>462</td>
<td>1490</td>
<td>18.5</td>
</tr>
<tr>
<td>3.3×3.3</td>
<td>3130</td>
<td>1070</td>
<td>3930</td>
<td>39</td>
</tr>
</tbody>
</table>

All CPU times (in sec) obtained on a Sun Fire with 8×2.8 GHz AMD processors
**Second task: Transport problem**

<table>
<thead>
<tr>
<th>#CPU</th>
<th>Umfpack</th>
<th>Pardiso</th>
<th>SuperLU\textsubscript{dist}</th>
<th>MUMPS</th>
<th>Basis Compression</th>
<th>Recursive GF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>406</td>
<td>271</td>
<td>560</td>
<td>240</td>
<td>105</td>
<td>1418</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>141</td>
<td>258</td>
<td>129</td>
<td>54</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>84</td>
<td>130</td>
<td>76</td>
<td>31</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>63</td>
<td>112</td>
<td>56</td>
<td>21</td>
<td>-</td>
</tr>
</tbody>
</table>

All times in **sec** for a 3.3×3.3×35 nm\(^3\) NW without SO coupling
Electrostatics

Grid Generation for Poisson equation

Grid must be general => Delaunay mesh: no data point (atoms) is contained in any triangle’s circumcircle (2D) or in any tetrahedron’s circumspheres (3D).

Carriers localized around atom positions

\[ n(r) = \sum_{i} n_i \delta(r - r_i) \]

Projection of FEM mesh on cross section. No charge in the oxide => larger elements
Electrostatics

Poisson Equation

$$\nabla \varepsilon \nabla V(\mathbf{r}) = -q \left( N_D^+ (\mathbf{r}) - n(\mathbf{r}) \right)$$

Carrier Density

$$n(\mathbf{r}) = \sum_i n_i \delta(\mathbf{r} - \mathbf{r}_i)$$

Solution: Finite Element Method (FEM)

$$\int dV \psi(\mathbf{r}) \nabla \varepsilon \nabla V(\mathbf{r}) = -q \int dV \psi(\mathbf{r}) \left( N_D^+ (\mathbf{r}) - n(\mathbf{r}) \right)$$

$$\int dV \psi(\mathbf{r}) n(\mathbf{r}) = \sum_i n_i \psi(\mathbf{r}_i)$$
band structure, transport, electrostatics

- MPI
- Blas/Lapack
- Umfpack
- PARDISO
- SuperLU\textsubscript{dist}
- MUMPS
- Basis Comp.
- RGF
- Aztec

Wire Generator → Hamiltonian Generator → Boundary Conditions → Transport

Poisson Equation → OK?

MPA → MB → MBU → MBUPSMBR
On-current for different channel orientations

How does transmission change with channel orientation?

For [100] orientation:
- $t_{ox} = 1$ nm

For [110] orientation:
- $L_y = 4.1$ nm

For [111] orientation:
- $t_{Si} = 2.1$ nm

For [112] orientation:
- $L_z = 3.1$ nm

The diagrams show the transmission as a function of energy for spin and no spin cases for each orientation.
On-current for different channel orientations

How does current depend on channel orientation?

Full-band (FB) transfer characteristics: $I_d - V_{gs}$ at $V_{ds} = 0.4$ V
n-FET with [100], [110], [111], and [112], $L_g = 13$ nm

- $I_{ON} = 7.7 \mu A$
- $I_{ON} = 4.3 \mu A$
- $I_{ON} = 2.9 \mu A$
- $I_{ON} = 1.2 \mu A$

Graph showing the $I_d - V_{gs}$ characteristics for different channel orientations.
Full-band (FB) transfer characteristics: $I_d-V_{gs}$ at $V_{ds}=-0.4$ V
p-FET with [100], [110], [111], and [112], $L_g=13$ nm

$I_{ON}=15.7$ μA
$I_{ON}=10.3$ μA
$I_{ON}=9.2$ μA
$I_{ON}=2.3$ μA
Spectral currents at $V_{gs} = 0.4 \text{ V}$, $V_{ds} = 0.11 \text{ V}$ (left) and $V_{gs} = 0.5 \text{ V}$, $V_{ds} = 1.1 \text{ V}$ (right).

There is a strong injection-induced tunneling part of the ON-current. This is a quantum-ballistic artifact (traveling states from source carry their Fermi level to the drain, negligible back-scattered states, too low density, Poisson equation shifts the conduction band down, deformation of S-D barrier, strong tunneling).

Can we trust the high quantum-ballistic ON-currents?
How does interface roughness influence the current?

Process variations => cross section variations
Example: [110] nanowire

interface roughness scattering: \( S(x) = \Delta^2 \exp(-|x|/L_m) \)
FB $I_d-V_{gs}$ at $V_{ds}=0.4$ V for one possible interface realization

1) Sub-threshold swing remains constant: $S \approx 60$ mV/dec.
2) Threshold voltage $V_{th}$, drain current $I_d$
Variation of the threshold voltage $V_{th}$ at $V_{ds}=0.4$ V

$[\text{110}], \text{ [100], [111], and [112]}$
### Performance summary: ON-current, OFF-current, roughness

<table>
<thead>
<tr>
<th></th>
<th>$n$-$I_{on}$ [$\mu$A]</th>
<th>$p$-$I_{on}$ [$\mu$A]</th>
<th>$\sigma$ [mV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[110]</td>
<td>7.7</td>
<td>15.7</td>
<td>9.3</td>
</tr>
<tr>
<td>[100]</td>
<td>4.3</td>
<td>2.3</td>
<td>13.8</td>
</tr>
<tr>
<td>[112]</td>
<td>2.9</td>
<td>9.2</td>
<td>16.8</td>
</tr>
<tr>
<td>[111]</td>
<td>1.2</td>
<td>10.3</td>
<td>24.8</td>
</tr>
</tbody>
</table>

[110] has the highest on-current, is the least sensitive to interface roughness.
Simulation of tunneling-induced OFF-currents

$L_g$ influence for nanowire n-FET with [100] channel

![Graphs showing spectral current for different $L_g$ values](image)
$I_d - V_{gs} @ V_{ds} = 0.4 \text{ V for 4 different channel lengths (4 nm, 7 nm, 10 nm, and 13 nm)}$ and for [100], [110], [111], [112]
Sub-threshold swing $S$ as function of gate length $L_g$

Low effective mass $\Rightarrow$ high on-current / strong tunneling

![Graph showing sub-threshold swing as a function of gate length with different crystallographic orientations.](image)
### Performance summary: ON-current, OFF-current, roughness, S-D leakage

<table>
<thead>
<tr>
<th></th>
<th>n-I\textsubscript{on} [μA]</th>
<th>p-I\textsubscript{on} [μA]</th>
<th>σ [mV]</th>
<th>S @ L\textsubscript{g}=4 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>[110]</td>
<td>7.7</td>
<td>15.7</td>
<td>9.3</td>
<td>122.3</td>
</tr>
<tr>
<td>[100]</td>
<td>4.3</td>
<td>2.3</td>
<td>13.8</td>
<td>88.5</td>
</tr>
<tr>
<td>[112]</td>
<td>2.9</td>
<td>9.2</td>
<td>16.8</td>
<td>87.5</td>
</tr>
<tr>
<td>[111]</td>
<td>1.2</td>
<td>10.3</td>
<td>24.8</td>
<td>83.2</td>
</tr>
</tbody>
</table>

[110] has the highest on-current, is the least sensitive to interface roughness, but suffers the most from source-to-drain tunneling.
Popular 3D mode-space approx. not suited => multi-terminal real space simulator (eff. mass!)

3D Schrödinger equation
\[ H | \psi_E > = E | \psi_E > \]
eff. mass approx. + finite difference
\[ < r | \psi_E > = \sum_{ijk} C_{ijk}(E) \delta(r - R_{ijk}) \]

3D sparse linear problem \( Ax = b \)
\[
(E-H-\Sigma_S-\Sigma_D-\Sigma_G) \cdot C = S_{\text{Inj}} + D_{\text{Inj}} + G_{\text{Inj}}
\]

BCs and injection mechanism
\[ M \cdot C_B = 2 \cdot \cos(k_B) \cdot C_B \]

3D carrier and current densities
\[ n(x_i,y_j,z_k) \xrightarrow{\text{C}_{ijk}} J(x_i,y_j,z_k) \xrightarrow{\text{SC}} \text{Poisson equ.} \xrightarrow{\text{Observable}} \]
Spectral gate current and total device current along the $x$-axis

$V_{gs} = -0.2$ V, $V_{ds} = 0.05$ V

$V_{gs} = 0.0$ V, $V_{ds} = 0.60$ V

current conservation: $I(x_2) - I(x_1) = I_G$
Isosurfaces of the gate current for a triple-gate structure
\( \text{SiO}_2 \) dielectric layer + \( \text{TiN} \) metal contact

**Current escapes at the gate corners**

\[
V_{gs} = -0.2 \text{ V, } V_{ds} = 0.05 \text{ V}
\]

\[
V_{gs} = 0.0 \text{ V, } V_{ds} = 0.60 \text{ V}
\]

- \( L_g = 10 \text{ nm} \)
- \( t_{Si} = 3 \text{ nm} \)
Current trajectories around gate corner

1D approximation vs full 3D (projected)

gate tunneling leakage

current $J(x,y)$
3D trajectories
1D straight lines

$\text{Si}$
$\text{SiO}_2$
TiN Gate

$t_{\text{ox}}$
Gate stack: SiO₂ (0.5 nm) + HfO₂ (3.5 nm, m* = 0.2 m₀, εᵣ = 25)
Performance: Good threshold voltage V<sub>th</sub>, low off-current I<sub>OFF</sub>

Dielectric + Contact Structure

Transfer Characteristics

Gate tunneling leakage
Stack reduces **gate current** by 7 orders of magnitude at low $V_{gs}$, **off-current** by a factor of 40, and keeps the same **on-current**.
Phonon-assisted band-to-band tunneling is an important leakage mechanism in steep pn-junctions (with a doping level of $10^{19}$ cm$^{-3}$ or more on both sides) or in high normal electric fields of MOS structures.

$$ R_{\text{net}}^{bb} = A F^{7/2} \frac{\tilde{n} \tilde{p} - n_i^{2, \text{eff}}}{(\tilde{n} + n_{i, \text{eff}})(\tilde{p} + n_{i, \text{eff}})} \left[ \left( \frac{F_C^+}{F} \right)^{-3/2} \exp \left( \frac{F_C^+}{F} \right) + \left( \frac{F_C^-}{F} \right)^{-3/2} \exp \left( -\frac{F_C^-}{F} \right) \right] \frac{\exp \left( \frac{\hbar \omega}{kT} \right) - 1}{1 - \exp \left( -\frac{\hbar \omega}{kT} \right)} $$

Non-local nature of the B2B rate can be modeled in a simple way:

$$ \tilde{n} = n \left( \frac{n_{i, \text{eff}}}{N_C} \right) \frac{\left| \nabla F_{F,n} \right|}{F} $$

Non-locality is crucial, as it e.g. prevents tunneling where no final states are available. In a MOSFET this usually happens close to the gate oxide interface, i.e. in a region where the electric field $F$ in the semiconductor becomes maximal. The “critical field strengths” are given by

$$ F_C^{\pm} = B (E_{g, \text{eff}}^{\pm} \hbar \omega)^{3/2} $$
band-to-band tunneling (GIDL)

120nm bulk MOSFET

- No direct experimental verification of the B2B rate in Si exists
- Calibrations based on GIDL data rely on correct modeling of lateral dopant diffusion
The non-local B2B model accounts for the "dark space" near the oxides. This reduces the rate compared to a local B2B model.

Maxima of the electric field do NOT occur in the pn-junction, but right to the drain-side gate corners (largest voltage drop).

=> the B2B rate is also located right to the gate corners and NOT at the metallurgical pn-junction.

The B2B rate cannot be changed much by changing the steepness of the pn-junction.
Example: acoustic phonon scattering

Assumptions: Bulk phonon dispersion, bulk coupling constants, EMA

Simplifications: High-T appr. \( \hbar \omega_q \ll k_B T \), lin. dispersion \( \hbar \omega_q = c_s q \)

\[ \rightarrow \text{self-energy } \sum_{ac}^< \text{ becomes local in space} \]

\[ \sum_{ac}^<(r, r', E) = \frac{\Xi^2 k_B T}{\rho c_s^2} G^<(r, r', E) \delta(r - r') \]

\( G^< \) requires \( G^R \) (full size):

\[ G^< = G^R(\sum_S^< + \sum_D^< + \sum_{ac}^<)G^A \]

\( (E - H - \sum^R)G^R = 1 \)

Elastic appr. \( E \pm \hbar \omega_q \approx E \rightarrow \sum^R \) simplifies to

\[ \sum^R(r, r', E) = \frac{\Xi^2 k_B T}{\rho c_s^2} G^R(r, r', E) \delta(r - r') \]
Incoherent scattering

Scaling \( V_{GS} = 0.5V \), \( V_{DS} = 50mV \)

\[ I \sim \frac{A}{L_G} \approx 30\% \]

TG 5 x 5 x 25 nm\(^3\) NW FET

\[ \text{on-current (A)} \]

\[ \text{on-current (A)} \]

\[ \text{gate length (nm)} \]

\[ \text{gate length (nm)} \]

\[ \text{gate voltage (V)} \]

\[ \text{gate voltage (V)} \]

QDD (S–Device)

QB (Simnad)

ac (Simnad)
Incoherent scattering

Example: Triple-gate 5nm x 5nm NW FET

TGNW-FET (courtesy EU SINANO project)
Gate length: 25 nm (65 nm technology node).
Channel Cross-Section: Square (5x5 nm²).
Source/drain extensions: 10 nm.
Oxide parameters: material is SiO$_2$ (k~3.9).
Field Oxide Thickness: 1.5 nm.
Buried Oxide Thickness: 150 nm.
Gate electrode work function: 4.1 eV.
von Neumann boundary conditions at S/D ends.
Doping specifications:
Substrate undoped, source and drain: $N_d=10^{20}$ cm$^{-3}$.
Incoherent scattering

**S-DEVICE mesh for the TGNW-FET and electron density at \( V_{GS} = 1.1V \)**

![Graph showing S-DEVICE mesh for the TGNW-FET and electron density at \( V_{GS} = 1.1V \)]
Incoherent scattering

Comparison of currents

- $V_{DS} = 0.11\text{V}$
  - Factor 3.7
- $V_{DS} = 1.1\text{V}$
  - Factor 4.2

![Graphs showing comparison of currents for $V_{DS} = 0.11\text{V}$ and $V_{DS} = 1.1\text{V}$]
Conclusion

• DG and 1D-SP are now state-of-the-art TCAD tools for the modeling of quantum-mechanical confinement effects
• DG is most practical method, because: (i) quantum-corrected dissipative transport scheme, (ii) full Newton, (iii) multi-dimensional
• Atomistic, full-band approach to simulate Si nano FETs is possible (and justified) up to 5x5 nm² cross sections (wire) or 5 nm body thickness (UTB)
• A variety of effects (channel orientation, strain, surface roughness, S-D tunneling, gate tunneling) can be studied in the quantum-ballistic limit
• Quantum-ballistic treatment overestimates the ON-current, because (i) incoherent scattering remains important, (ii) source-drain tunneling artifact
• For ballistic transport use WF formalism and not NEGF, because much more efficient!
• Challenges: Incoherent scattering, CPU time