Three-dimensional Quantum Simulation of Silicon Nanowires

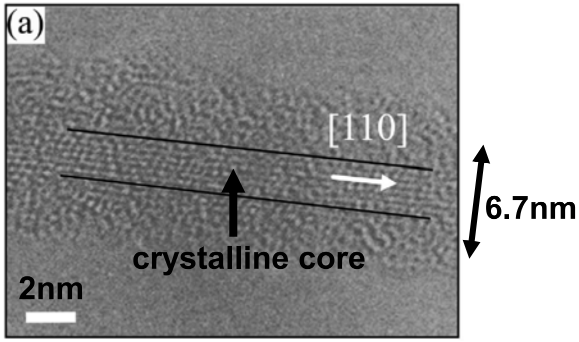
Andreas Schenk [1] and Mathieu M. Luisier [2]

[1] ETH Zürich
[2] Purdue University

PURDUE UNIVERSITY NCN IIS
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
ETH

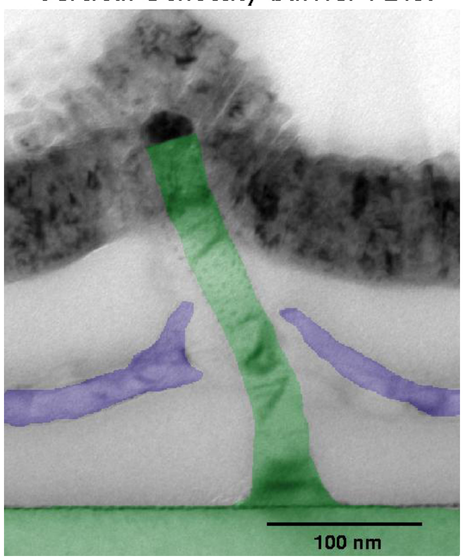
Outline

- Introduction
- Simulation method
  - Band structure, transport, electrostatics
- Quantum-ballistic currents
  - On-current for different channel orientations
  - Source-drain tunneling
  - Interface roughness
- Gate tunneling leakage
- Limit of large cross sections
- Incoherent scattering
- Conclusion



Cui et al., APL 78(15), 2214 (2001)

Vertical Schottky barrier FETs.



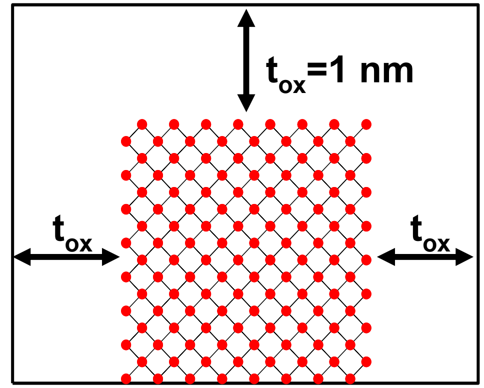
IBM Zürich, Walter Riess

- Si NWs are 'post-CMOS' candidates, both for transistors and wire connectors
- grown NWs with different crystal orientations and cross sections, working transistors demonstrated
- if cross section < 5 x 5 nm^2, strong confinement => band structure effects become important
- predictive simulation requires: accurate band structure model, quantum transport solver (OBCs), self-consistent electrostatics
- main challenges: CPU time, incoherent scattering

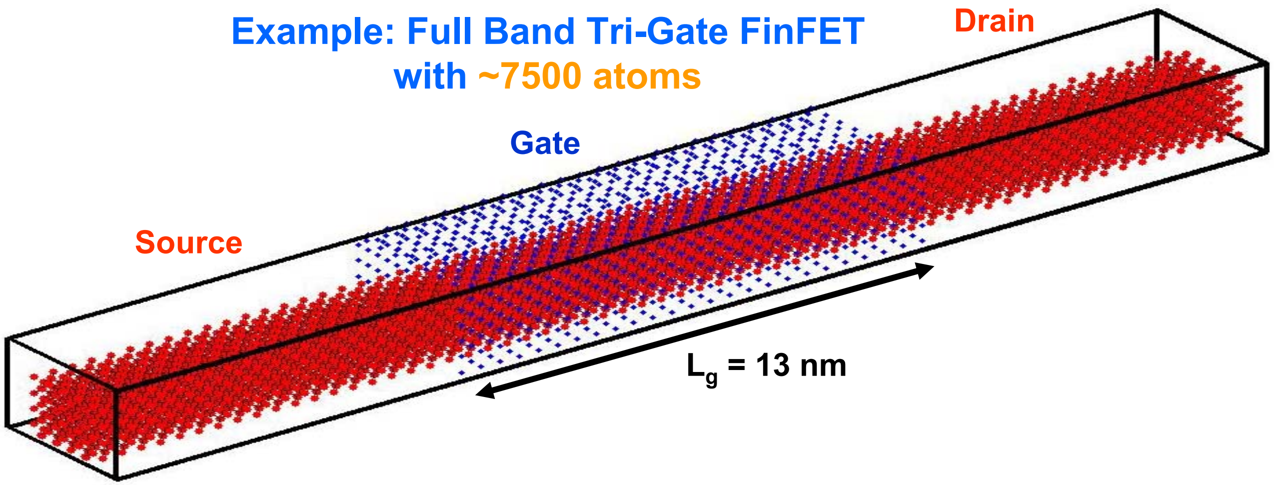
Band structure

- $sp^3d^5s^*$ 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

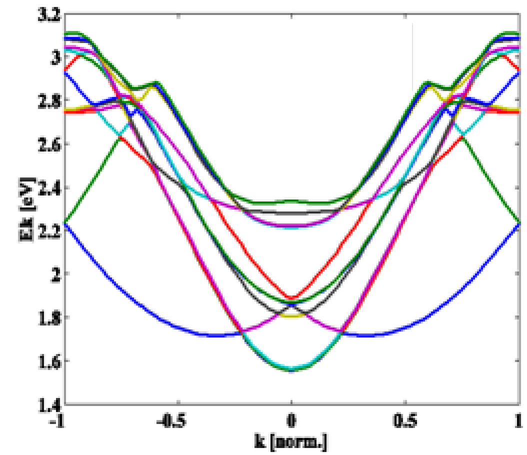
Atomistic description of a [100] nanowire with 2.1 x 2.1 nm^2 cross section



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



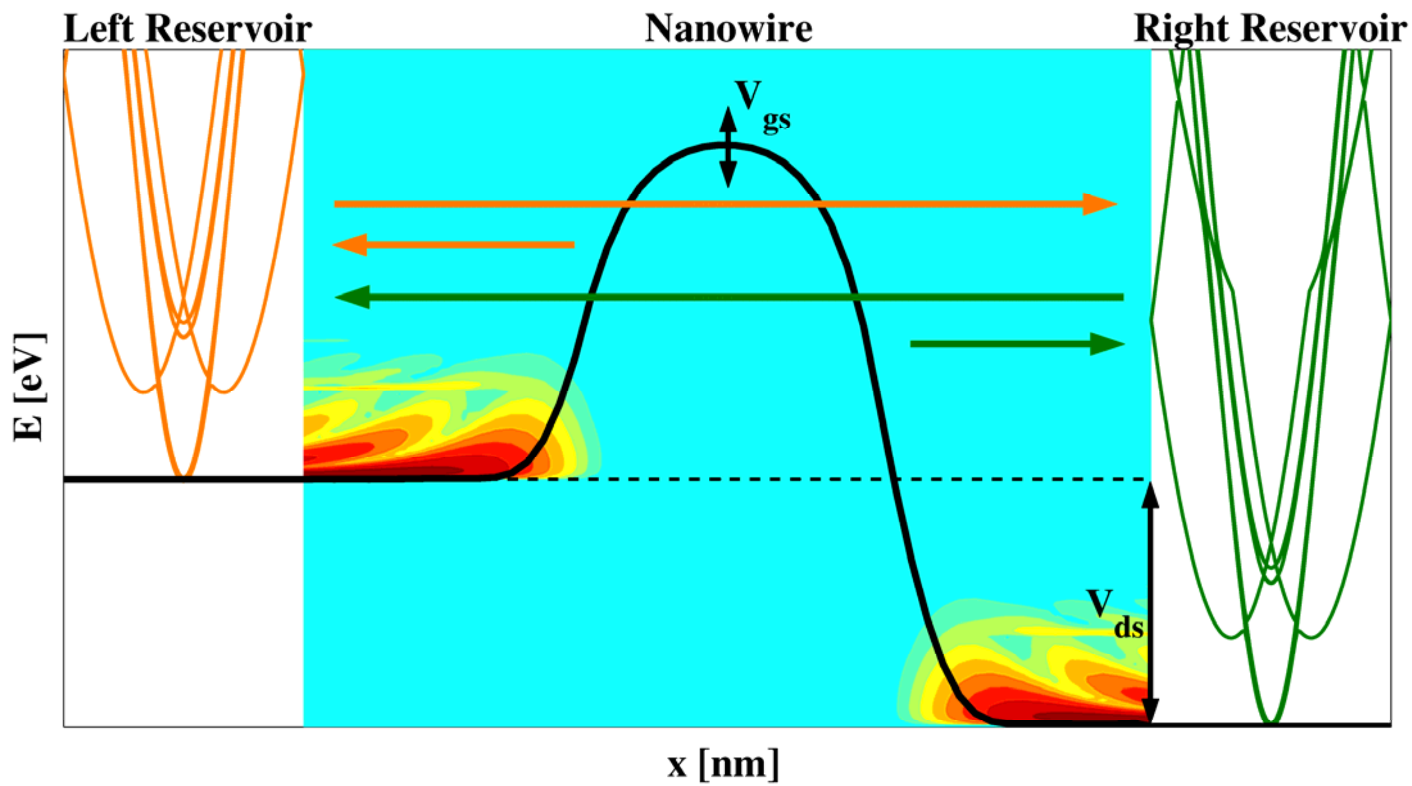
TB band structure



Simulation method

Transport

Cut along the transport direction *x* in the nanowire



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.

Transport

WF formalism

Schrödinger Equation

$$H|\psi_E\rangle = E|\psi_E\rangle$$

Tight-Binding ansatz for the wave function

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

Löwdin orbitals

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

$$\mathbf{H}_{tot}\cdot\mathbf{C}^{\sigma}_{p,n}(k) = \mathbf{I}_{0,p,n}(k)$$

Orbital-coefficient vector

Transport

Carrier and current density

$$n(x,\mathbf{r}_s) = \frac{1}{N_x}\sum_{n,p,\sigma}\sum_{i,\mathbf{R}_s}\sum_{k}|C^{\sigma}_{i,p,n}(\mathbf{R}_s,k)|^2 f(E_{p,n}(k)-\mu_p)\delta(x-x_i)\delta(\mathbf{r}_s-\mathbf{R_s})$$

$$\mathbf{J}(\mathbf{r}) = \sum_{i_1,i_2} i\frac{e}{2\hbar}\sum_{p,n_p}\frac{\Delta}{2\pi}\int \mathrm{d}E \left(H_{i_1i_2}C_{i_2,p,n_p}C^*_{i_1,p,n_p} - C_{i_1,p,n_p}C^*_{i_2,p,n_p}H_{i_2i_1}\right) \times \left|\frac{\mathrm{d}E}{\mathrm{d}k_{p,n_p}}\right|^{-1} f(E-\mu_p)(\mathbf{R}_{i_2}-\mathbf{R}_{i_1})\delta(\mathbf{r}-\mathbf{R}_{i_1})$$

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

$$T(E) = \sum_{n,m}|C_{N_s+1,p=1,n}(k_m)|^2\left|\frac{\mathrm{d}E}{\mathrm{d}k_m}\right|\left|\frac{\mathrm{d}E}{\mathrm{d}k_n}\right|^{-1}$$

NEGF formalism

$$n(\mathbf{r}) = -i\sum_j\int\frac{\mathrm{d}E}{2\pi}G^<_{jj}(E)\delta(\mathbf{r}-\mathbf{R_j})$$

$$\mathbf{J}(\mathbf{r}) = \sum_{i_1}\sum_{i_2}\frac{e}{2}\left(H_{i_1i_2}G^<_{i_2i_1} - G^<_{i_1i_2}H_{i_2i_1}\right)(\mathbf{r_{i_2}}-\mathbf{r_{i_1}})\delta(\mathbf{r}-\mathbf{r_{i_1}})$$

Transport

WF ↔ NEGF (if no incoherent scattering)

$$\mathrm{G}^{<}_{\mathrm{ij}}(E) = \sum_p \sum_n \underbrace{\mathrm{C}_{i,p}(k_n)\mathrm{C}^T_{j,p}(k_n)} \left|\frac{\mathrm{d}E(k_n)}{\mathrm{d}k_n}\right|^{-1} f(E(k_n) - \mu_p)$$

$(t_b \cdot N_A) \times (t_b \cdot N_A)$ $(t_b \cdot N_A) \times (t_b \cdot N_A)$

When NEGF? In case of incoherent scattering.

When WF? Otherwise, because CPU time is greatly reduced!

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

$(E - H - t_{10} \cdot g^R_{00} \cdot t_{01}) \cdot g^R_{00} = I$

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

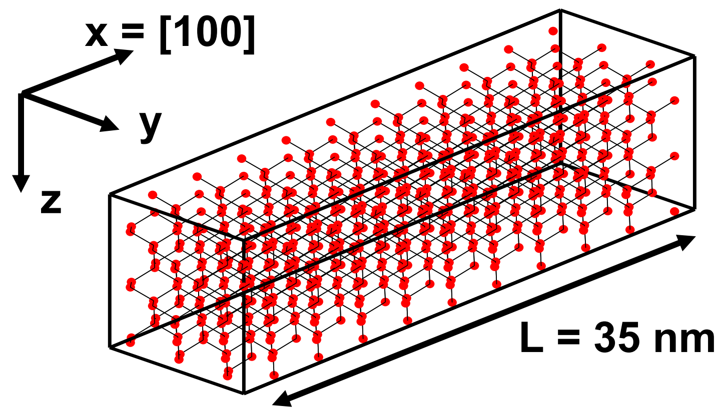
$A(E) \cdot C_n = \exp(ik_n(E)\Delta) \cdot B(E) \cdot C_n$

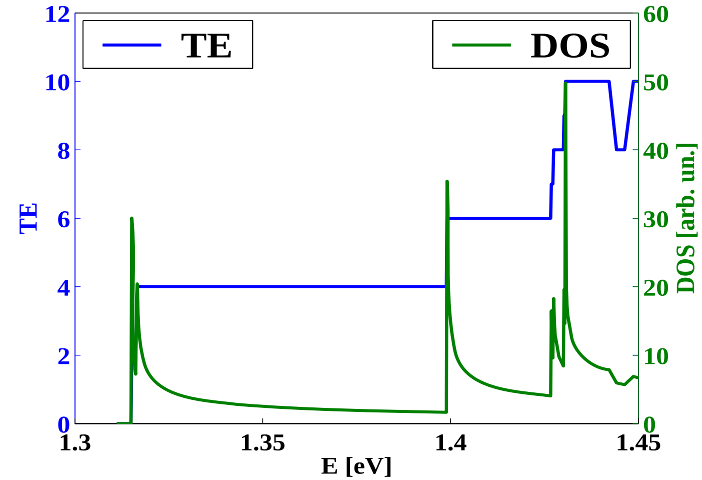
Shift-and-invert + ordinary eigenvalue problem ($N < t_b \cdot N_A$) (PRB 74, 205323 (2006))

$M(E) \cdot C_n = \lambda_n(k_n(E)) \cdot C_n$

Transport

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





First task: Open Boundary Conditions

$L_y \times L_z$ nm^2	$t_b \times N_A$	Iterative Solver	Generalized EVP	Ordinary EVP
2.5×2.5	1810	197	506	7.2
2.9×2.9	2420	462	1490	18.5
3.3×3.3	3130	1070	3930	39

All CPU times (in sec) obtained on a Sun Fire with 8×2.8 GHz AMD processors

Transport

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

Second task: Transport problem

#CPU	Umfpack	Pardiso	SuperLU_{dist}	MUMPS	Basis Compression	Recursive GF
1	406	271	560	240	105	1418
2	-	141	258	129	54	-
4	-	84	130	76	31	-
8	-	63	112	56	21	-

Wave Function (Umfpack – Basis Compression) *NEGF* (Recursive GF)

All times in sec for a 3.3×3.3×35 nm^3 NW without SO coupling

Electrostatics

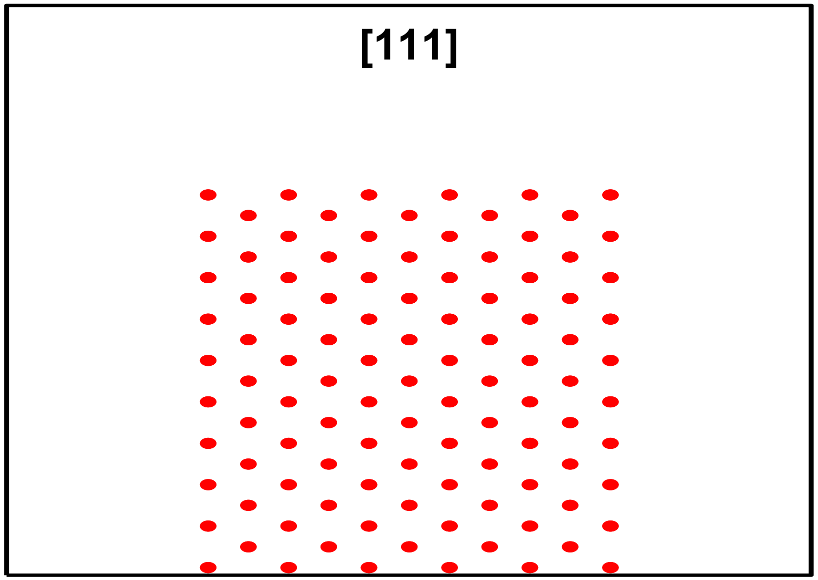
Grid Generation for Poisson equation

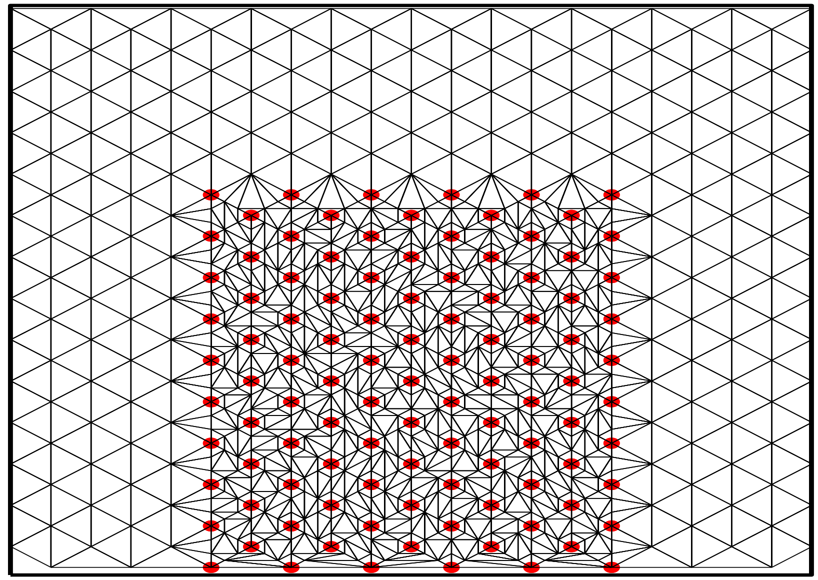
Grid must be general => Delaunay mesh: no data point (atoms) is contained in any triangle's circumcircle (2D) or in any tetrahedron's circumspheres (3D).

Carriers localized around atom positions

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

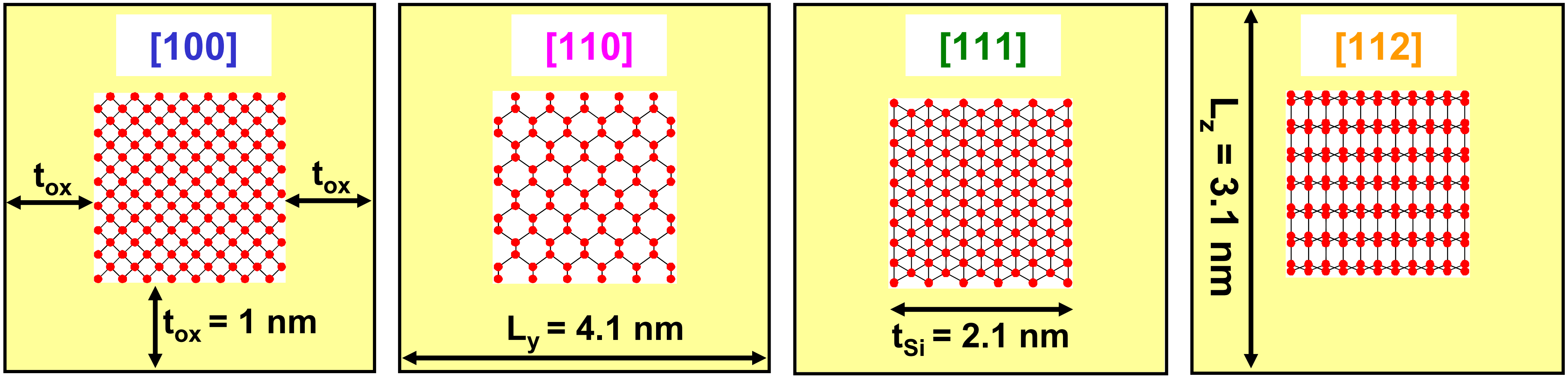
Projection of FEM mesh on cross section.
No charge in the oxide => larger elements

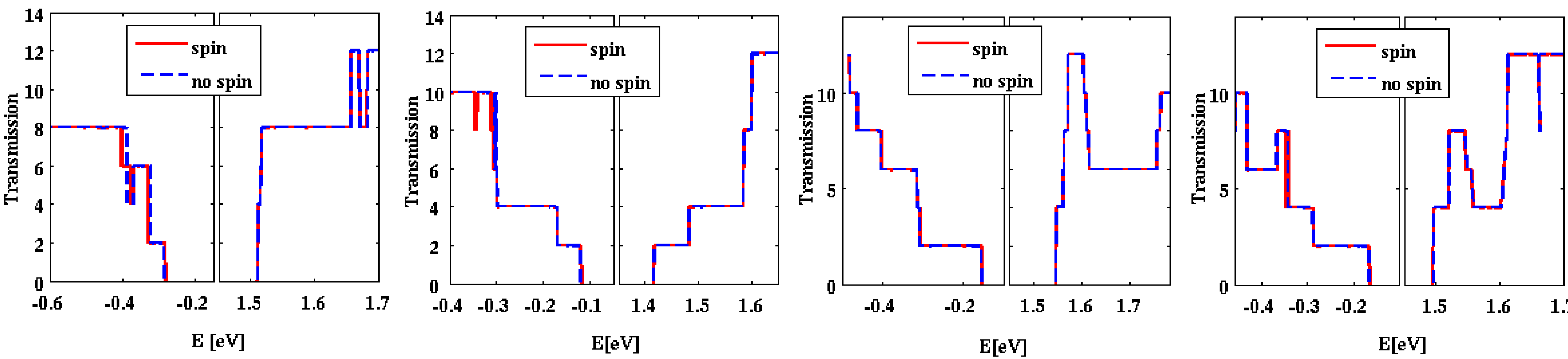




Quantum-ballistic currents

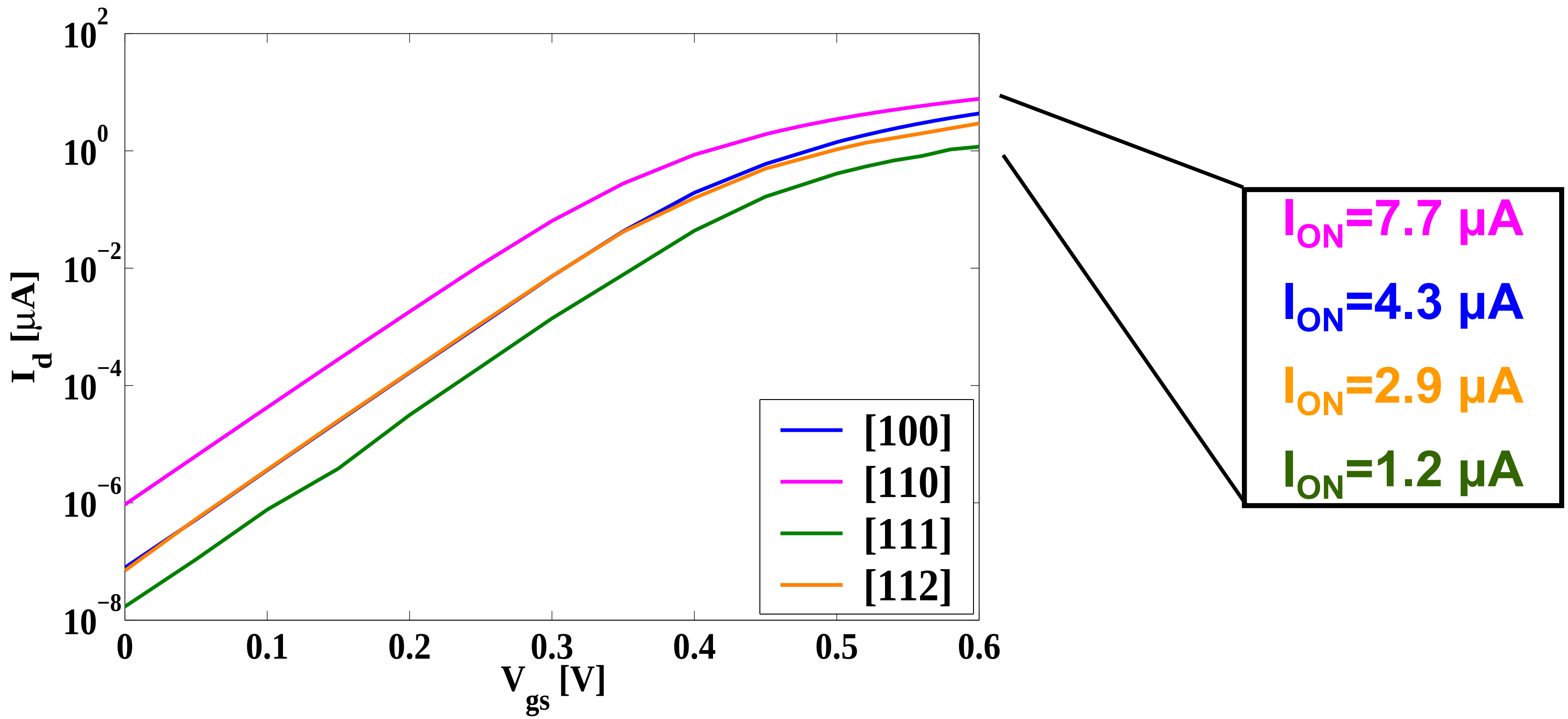
On-current for different channel orientations





On-current for different channel orientations

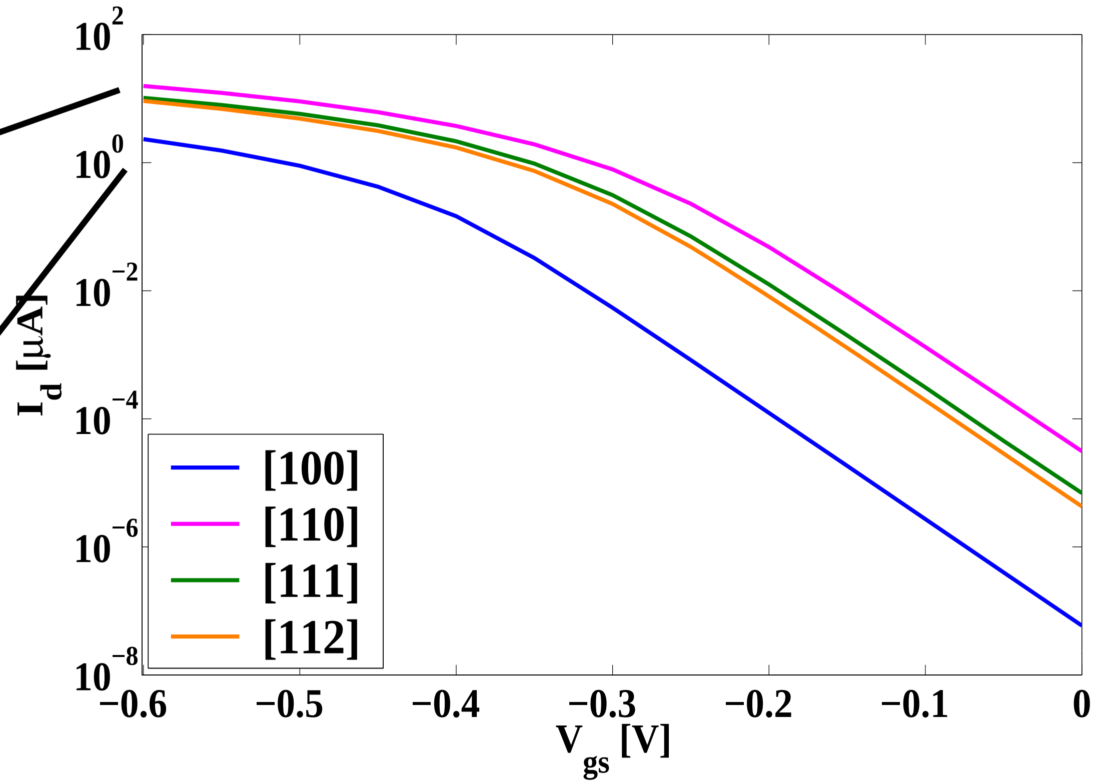
Full-band (FB) transfer characteristics: I_d-V_{gs} at V_{ds}=0.4 V
n-FET with [100], [110], [111], and [112], L_g=13 nm



On-current for different channel orientations

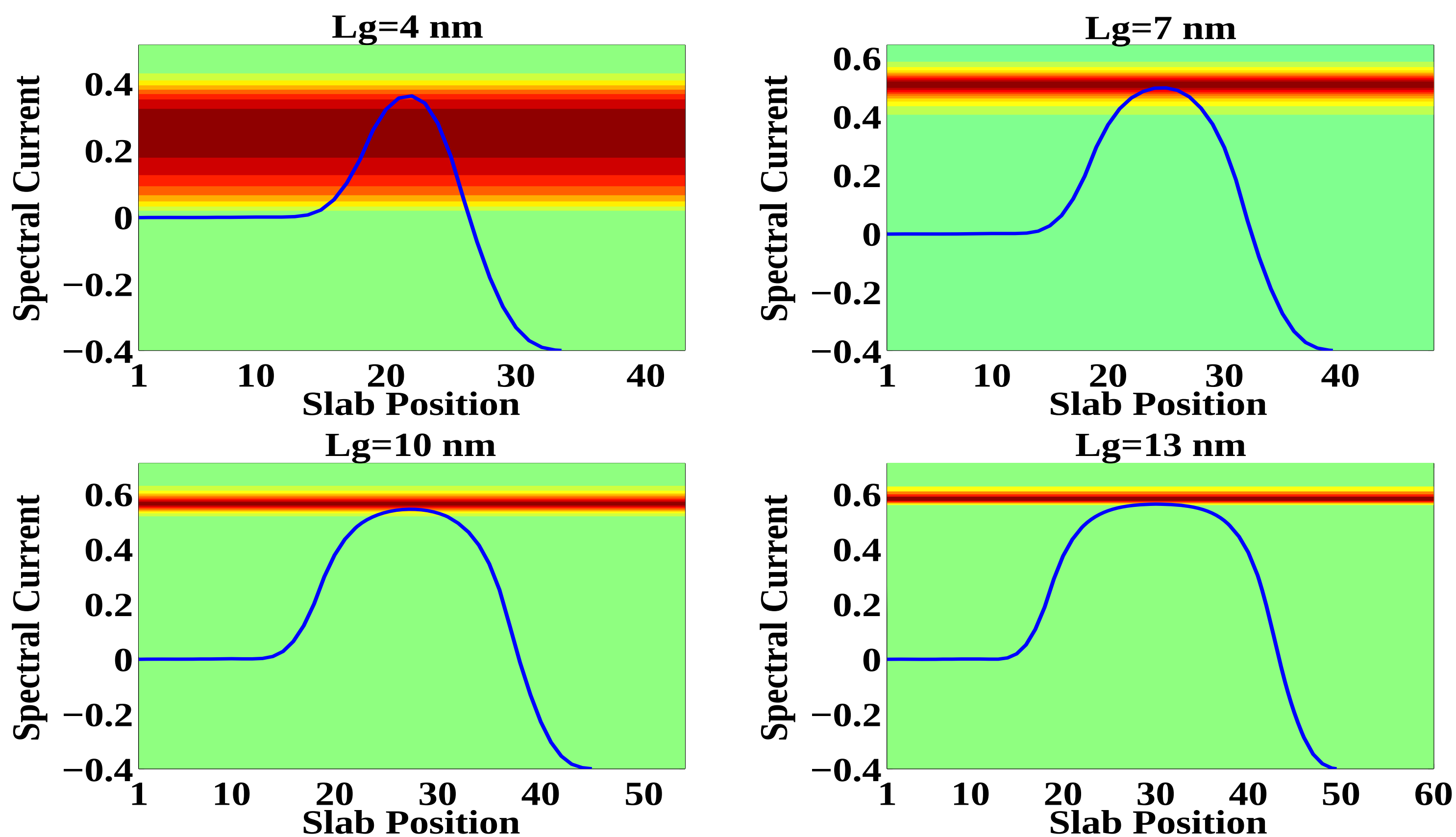
Full-band (FB) transfer characteristics: I_d-V_{gs} at V_{ds}=-0.4 V
p-FET with [100], [110], [111], and [112], L_g=13 nm

I_{ON}=15.7 µA
I_{ON}=10.3 µA
I_{ON}=9.2 µA
I_{ON}=2.3 µA



Source-drain tunneling

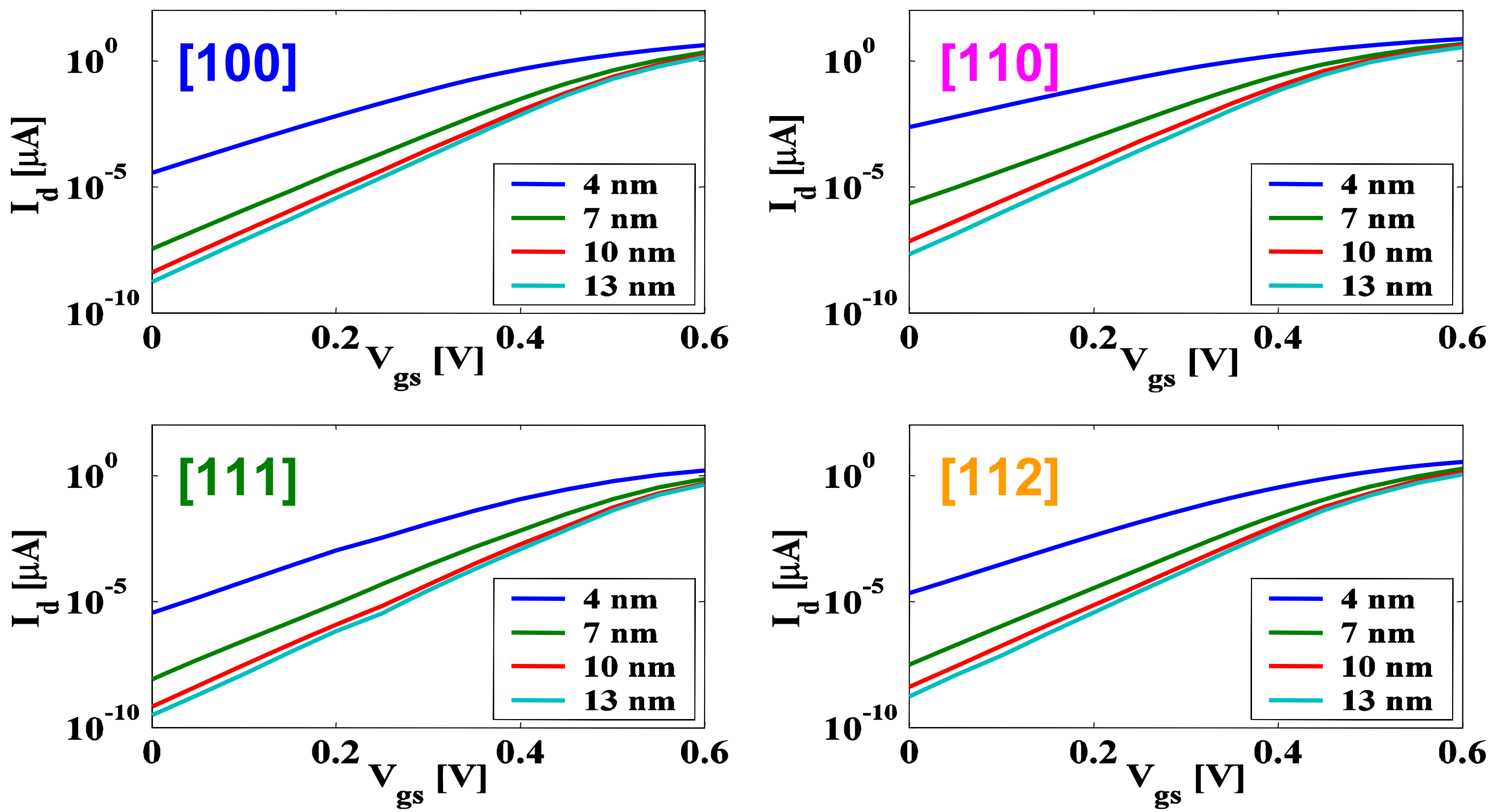
L_g influence for nanowire n-FET with [100] channel



Quantum-ballistic currents

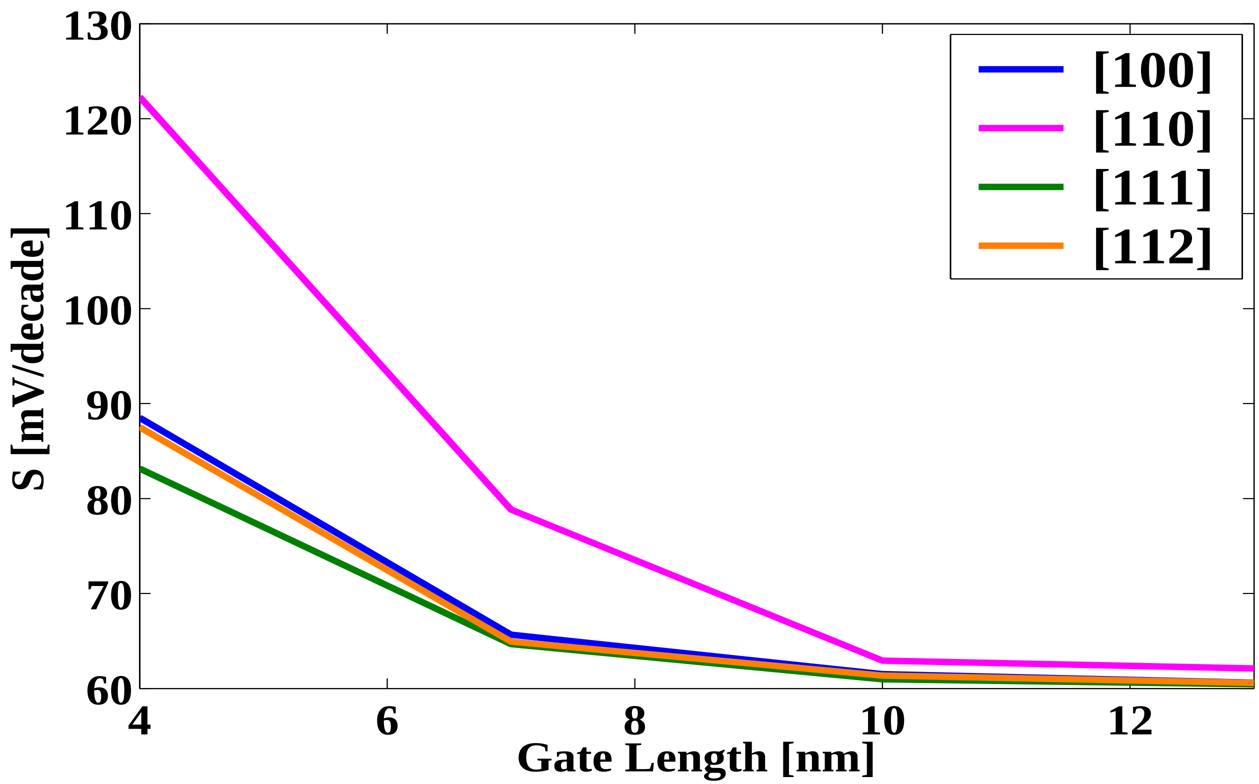
Source-drain tunneling

I_d-V_{gs} @ V_{ds}=0.4 V for 4 different channel lengths (4 nm, 7 nm, 10 nm, and 13 nm) and for [100], [110], [111], [112]



Source-drain tunneling

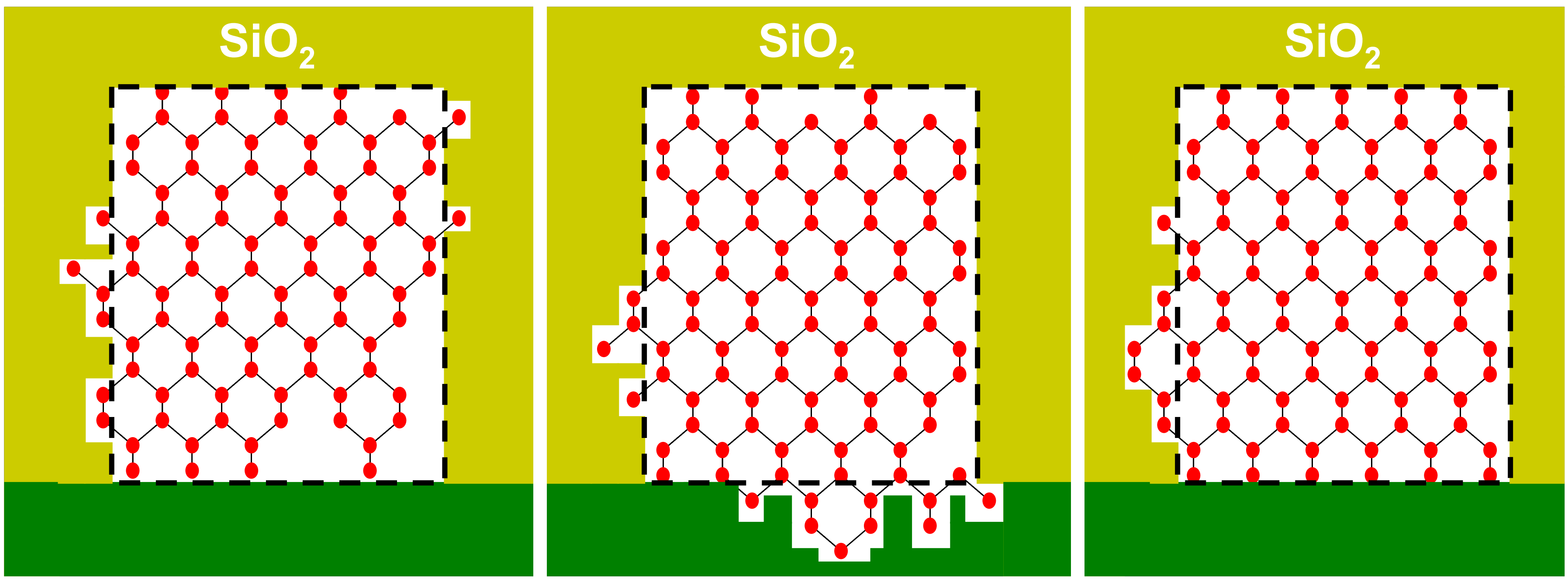
Sub-threshold swing *S* as function of gate length L_g
Low effective mass => high on-current / strong tunneling



Interface roughness

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

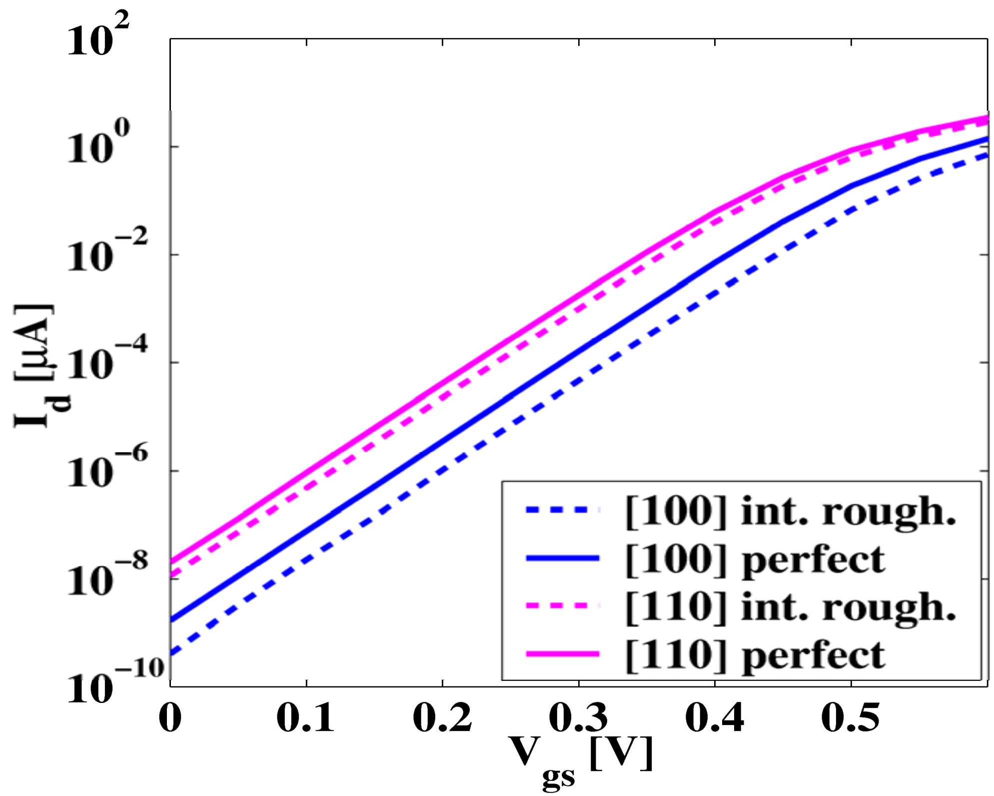
x=14 nm x=16 nm x=26 nm



interface roughness scattering: $S(x)=\Delta^2 \exp(-|x|/L_m)$

Interface roughness

FB I_d-V_{gs} at V_{ds}=0.4 V for one possible interface realization

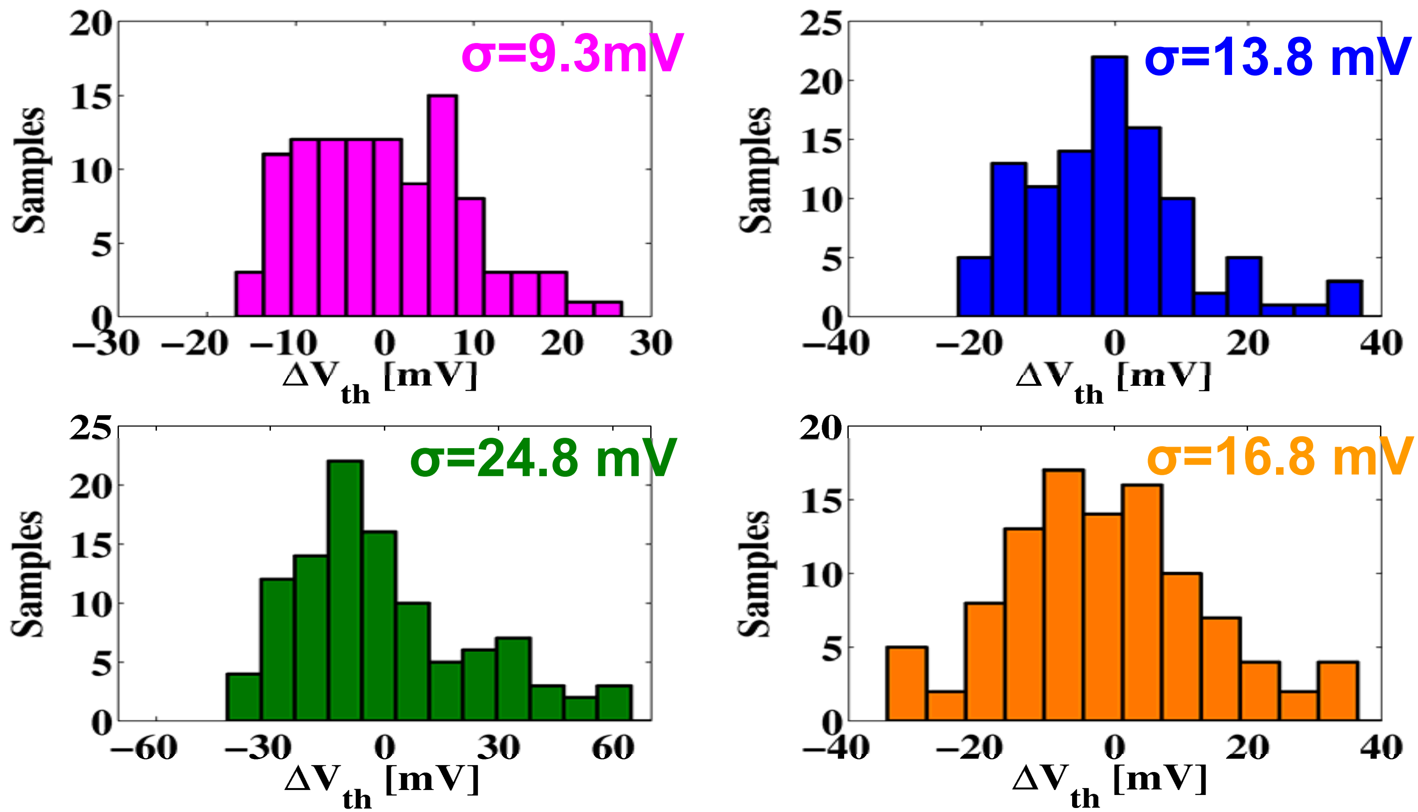


1) Sub-threshold swing remains constant: S ≈ 60 mV/dec.
2) Threshold voltage V_{th} ↗, drain current I_d ↘

Interface roughness

Variation of the threshold voltage V_{th} at V_{ds}=0.4 V

[110], [100], [111], and [112]

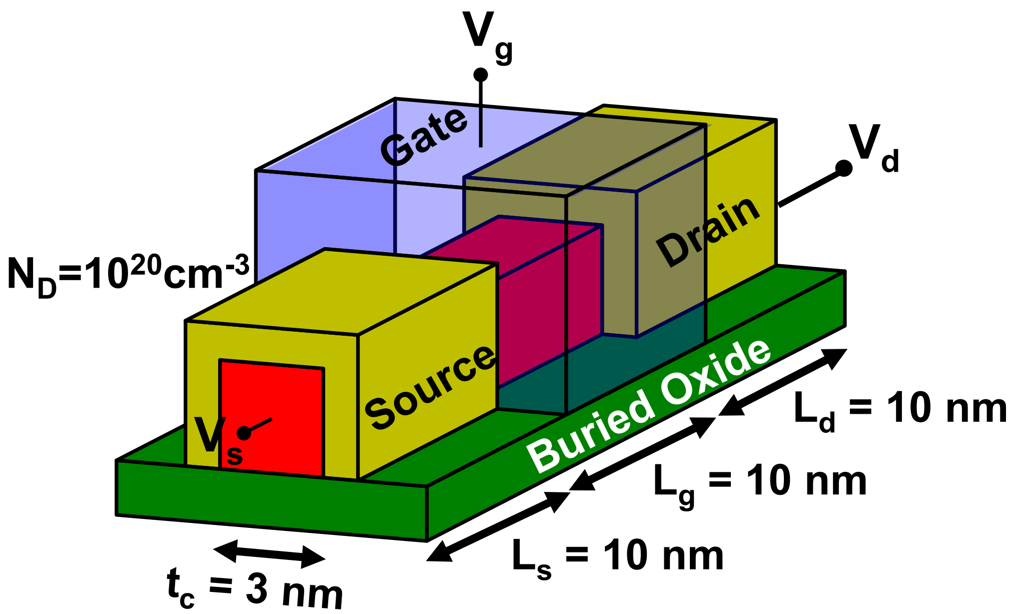


Performance summary: best channel orientation?

	n-I_{on} [μA]	p-I_{on} [μA]	σ [mV]	S @ L_g=4 nm
[110]	7.7	15.7	9.3	122.3
[100]	4.3	2.3	13.8	88.5
[112]	2.9	9.2	16.8	87.5
[111]	1.2	10.3	24.8	83.2

[110] has the highest on-current, is the least sensitive to interface roughness, but suffers the most from source-to-drain tunneling

Triple-gate nanowire FET with poly-Si or TiN contacts



EOT = 1nm

- 1 nm SiO_2
- 0.5 nm SiO_2 + 3.5 nm HfO_2

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

$$<\mathbf{r}\,|\,\psi_E> \;=\; \sum_{ijk} C_{ijk}(E)\,\delta(\mathbf{r}-\mathbf{R}_{ijk})$$

3D sparse linear problem Ax=b

$$(E-H-\Sigma_S-\Sigma_D-\Sigma_G)\cdot C = S_{Inj}+D_{Inj}+G_{Inj}$$

BCs and injection mechanism

$$M\cdot C_B = 2\cdot\cos(k_B)\cdot C_B$$

3D carrier and current densities

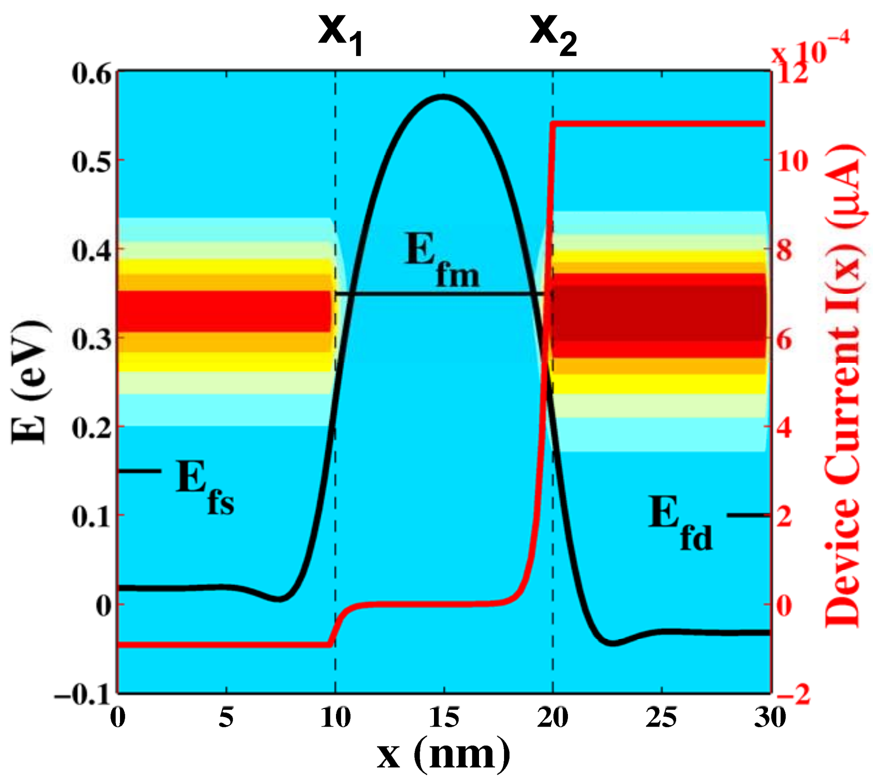
C_{ijk} → $n(x_i,y_j,z_k)$ → Poisson equ.

C_{ijk} → $J(x_i,y_j,z_k)$ → Observable

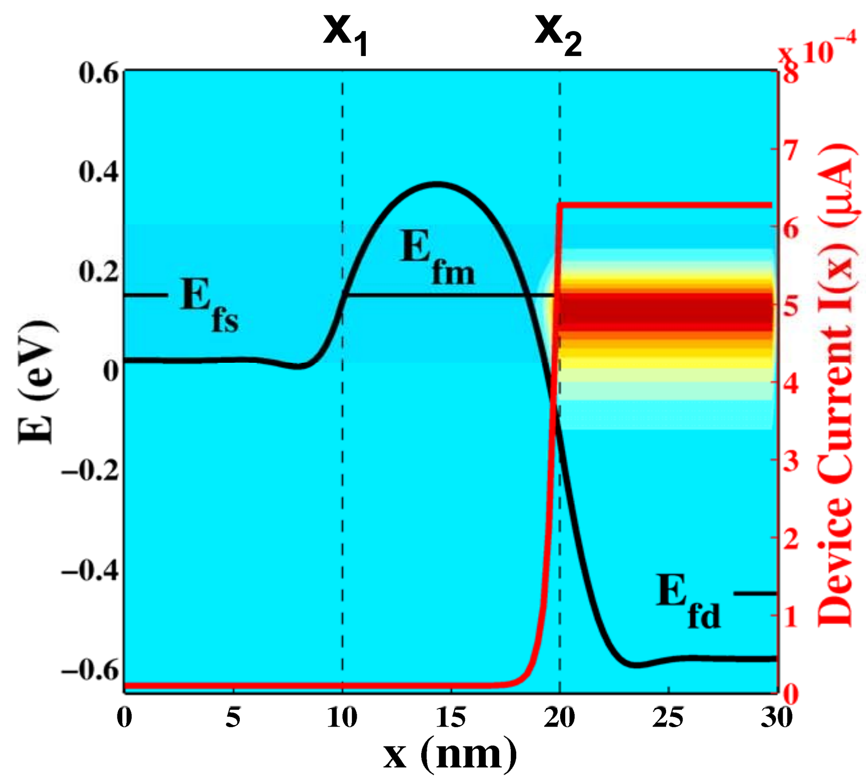
Poisson equ. ↔ (SC) ↔ Observable

Spectral gate current and total device current along the *x*-axis

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



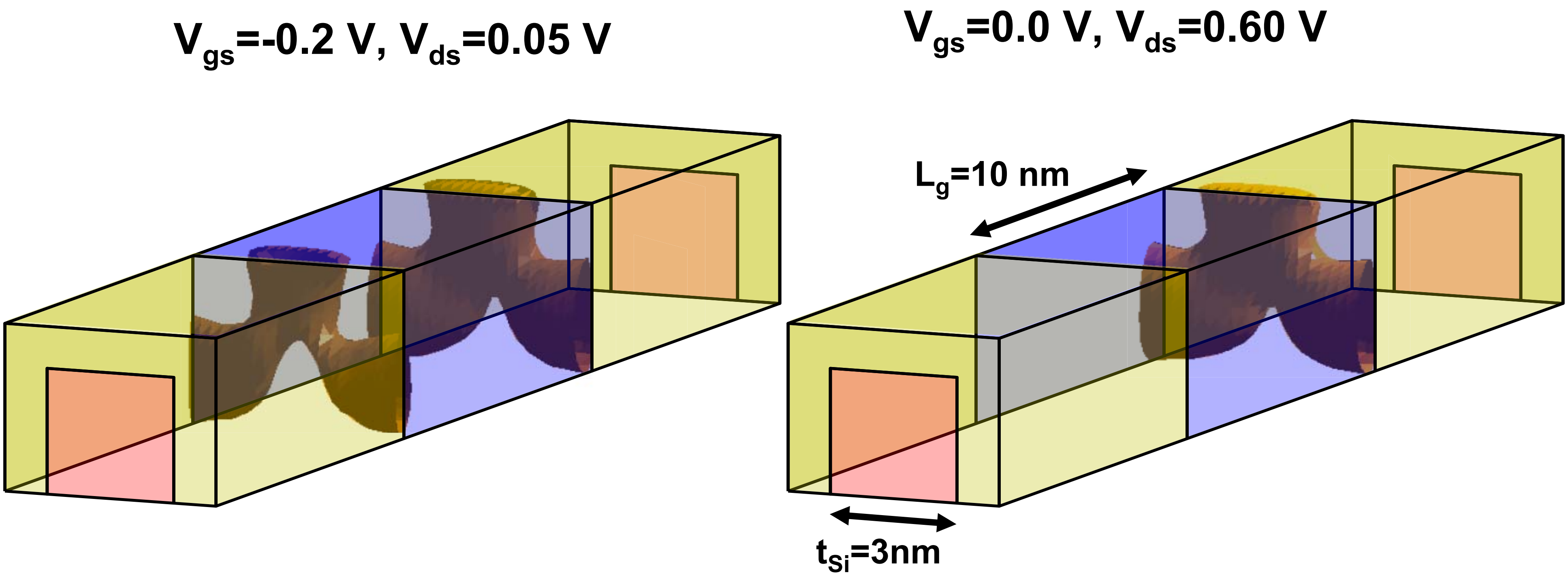
V_{gs}=0.0 V, V_{ds}=0.60 V



current conservation: $I(x_2)-I(x_1) = I_G$

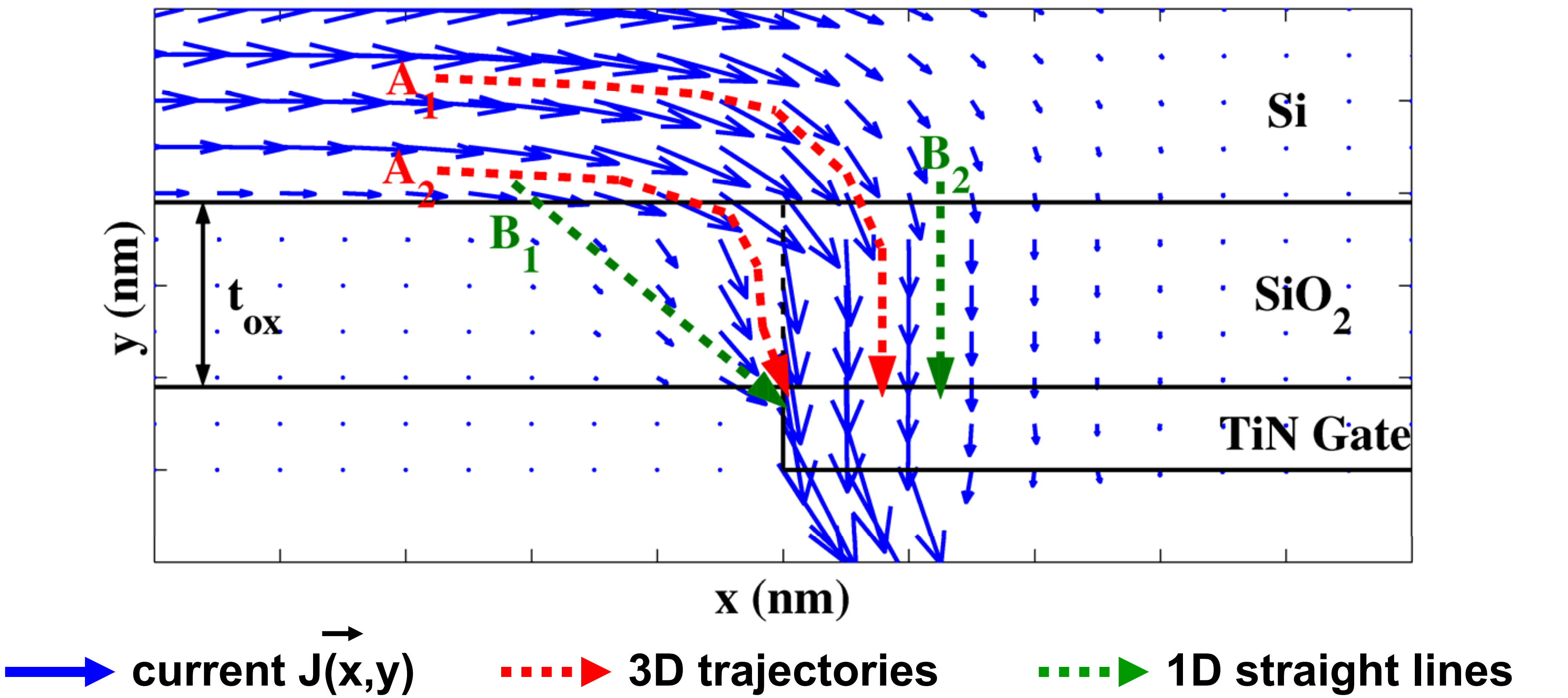
Isosurfaces of the gate current for a triple-gate structure SiO_2 dielectric layer + TiN metal contact

Current escapes at the gate corners



Current trajectories around gate corner

1D approximation vs full 3D (projected)

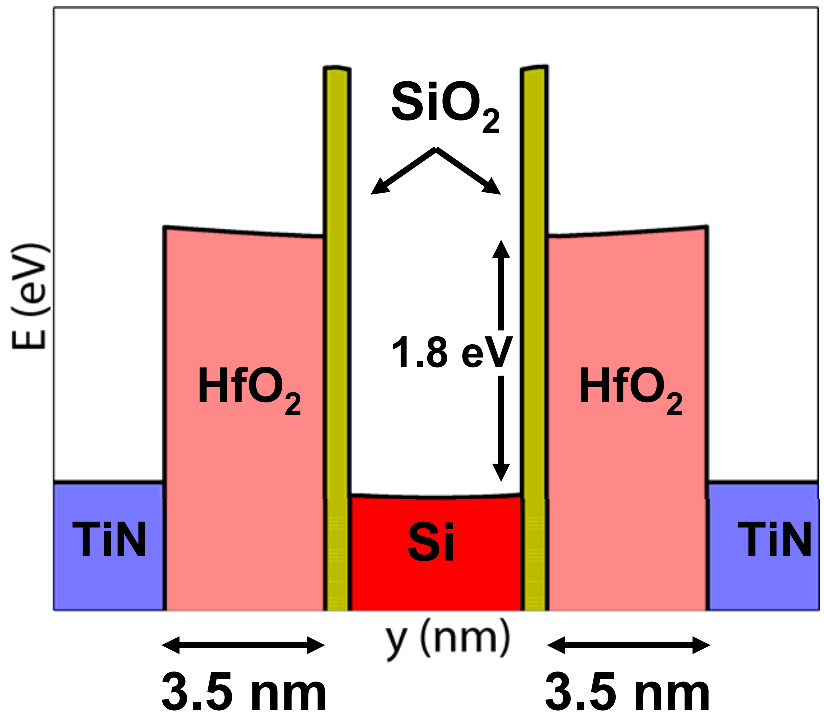


Gate tunneling leakage

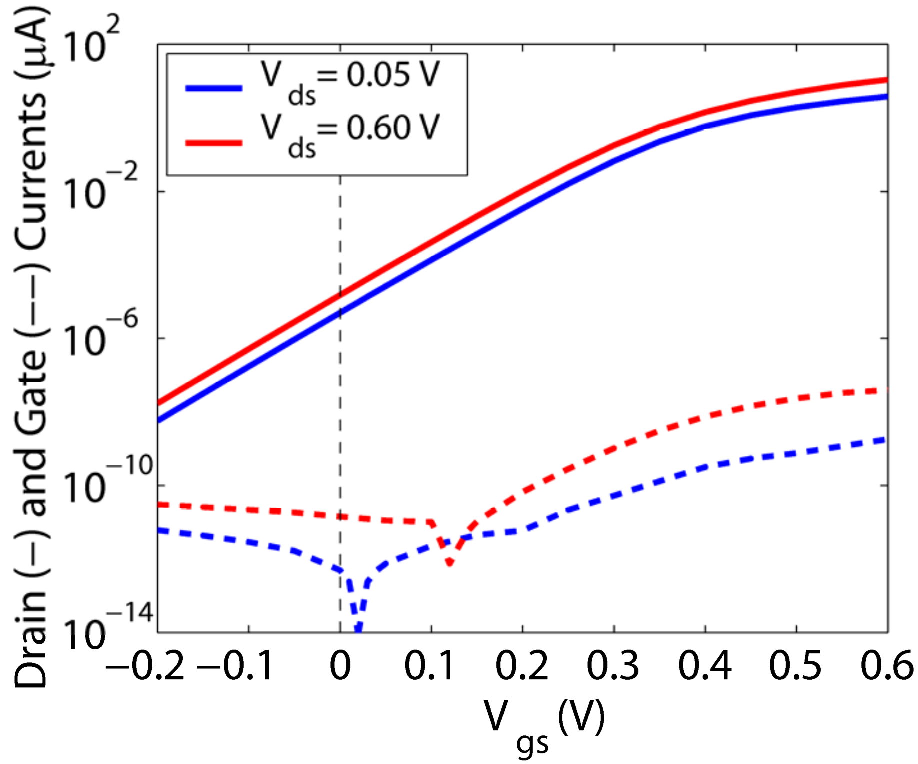
Gate stack: SiO_2 (0.5 nm) + HfO_2 (3.5 nm, $m^* = 0.2\ m_0$, $\varepsilon_R = 25$)

Performance: Good threshold voltage V_{th}, low off-current I_{OFF}

Dielectric + Contact Structure

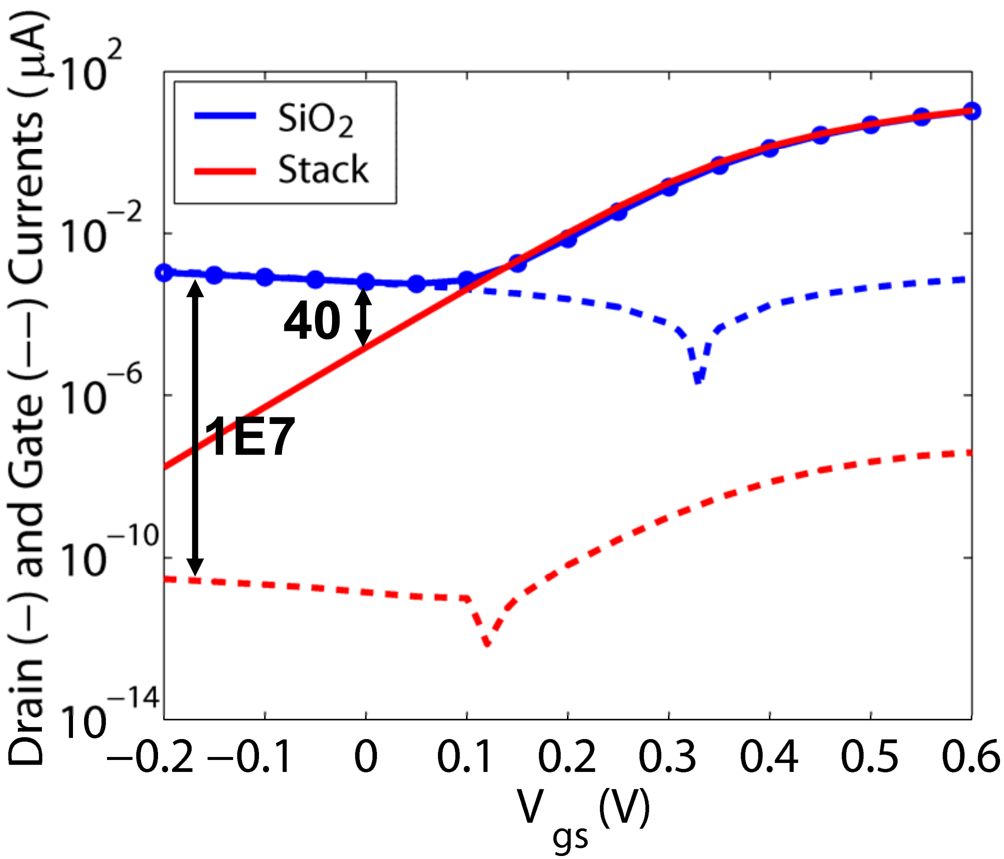


Transfer Characteristics

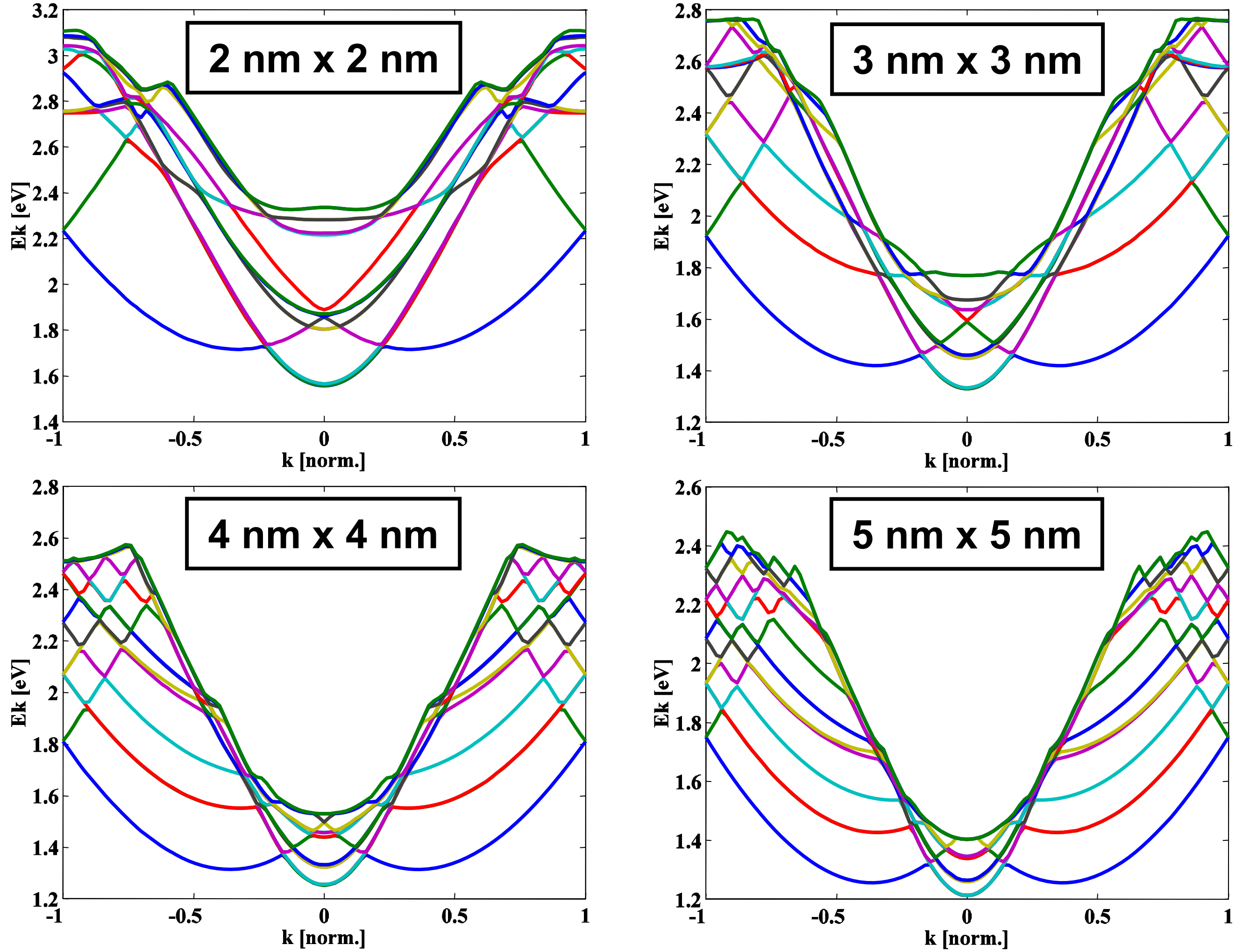


Gate tunneling leakage

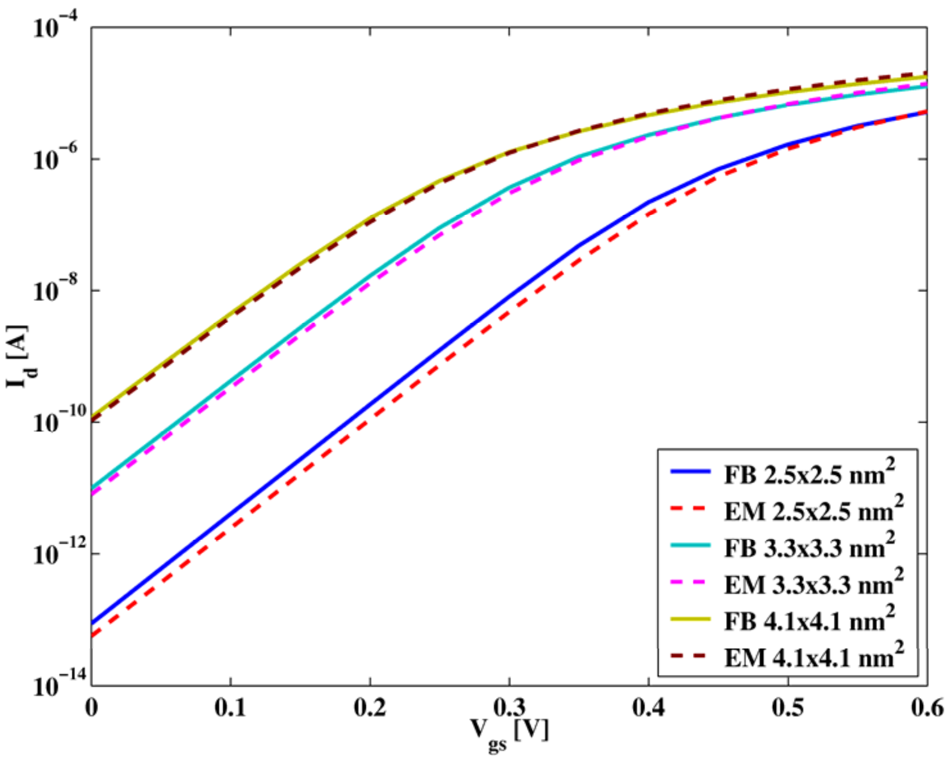
Stack reduces gate current by 7 orders of magnitude at low V_{gs}, off-current by a factor of 40, and keeps the same on-current

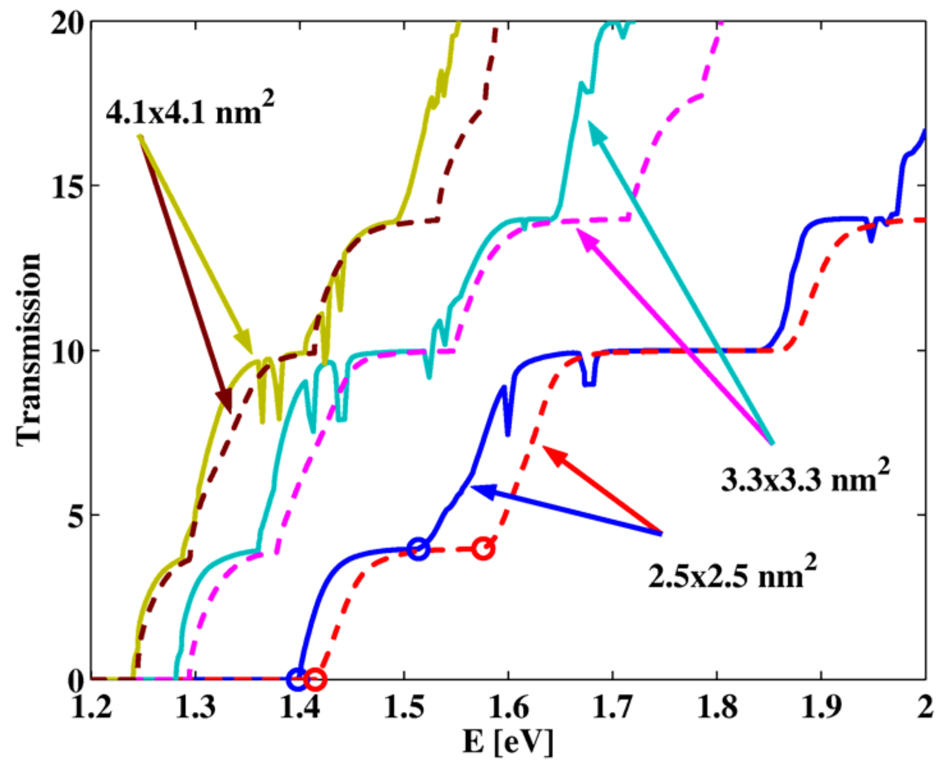


How does band structure change with increasing cross section?

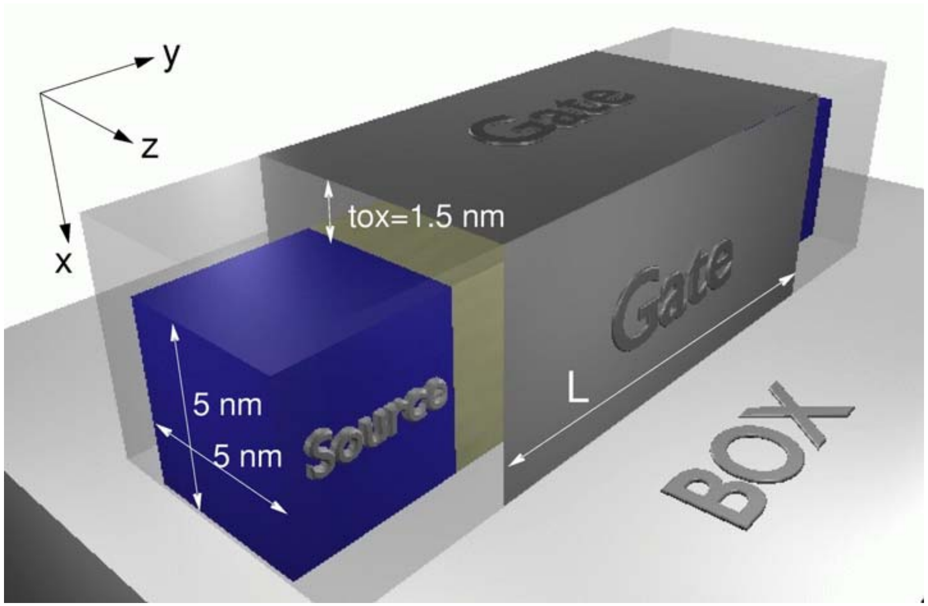


Output characteristics and transmission for [100]
Full Band (solid lines) *vs* Effective Mass (dashed lines)

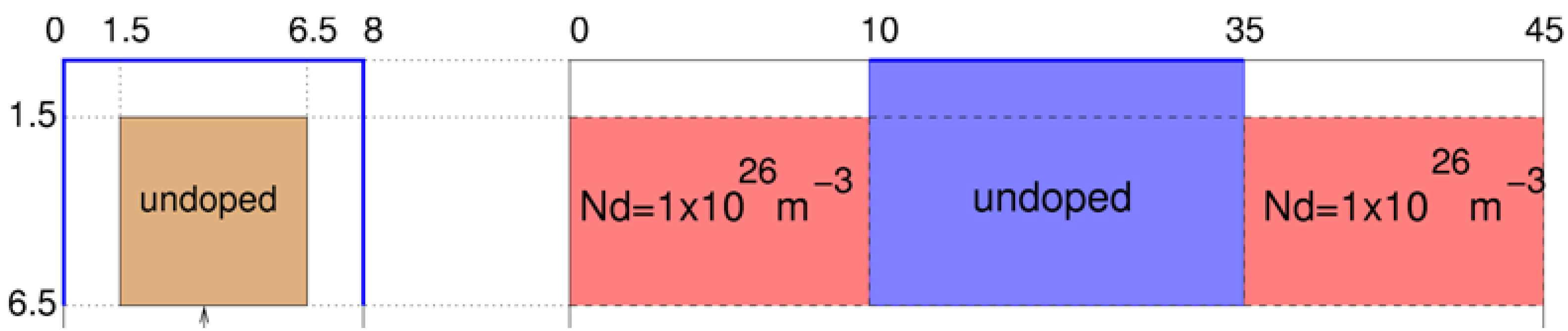




Example: Triple-gate 5nm x 5nm NW FET

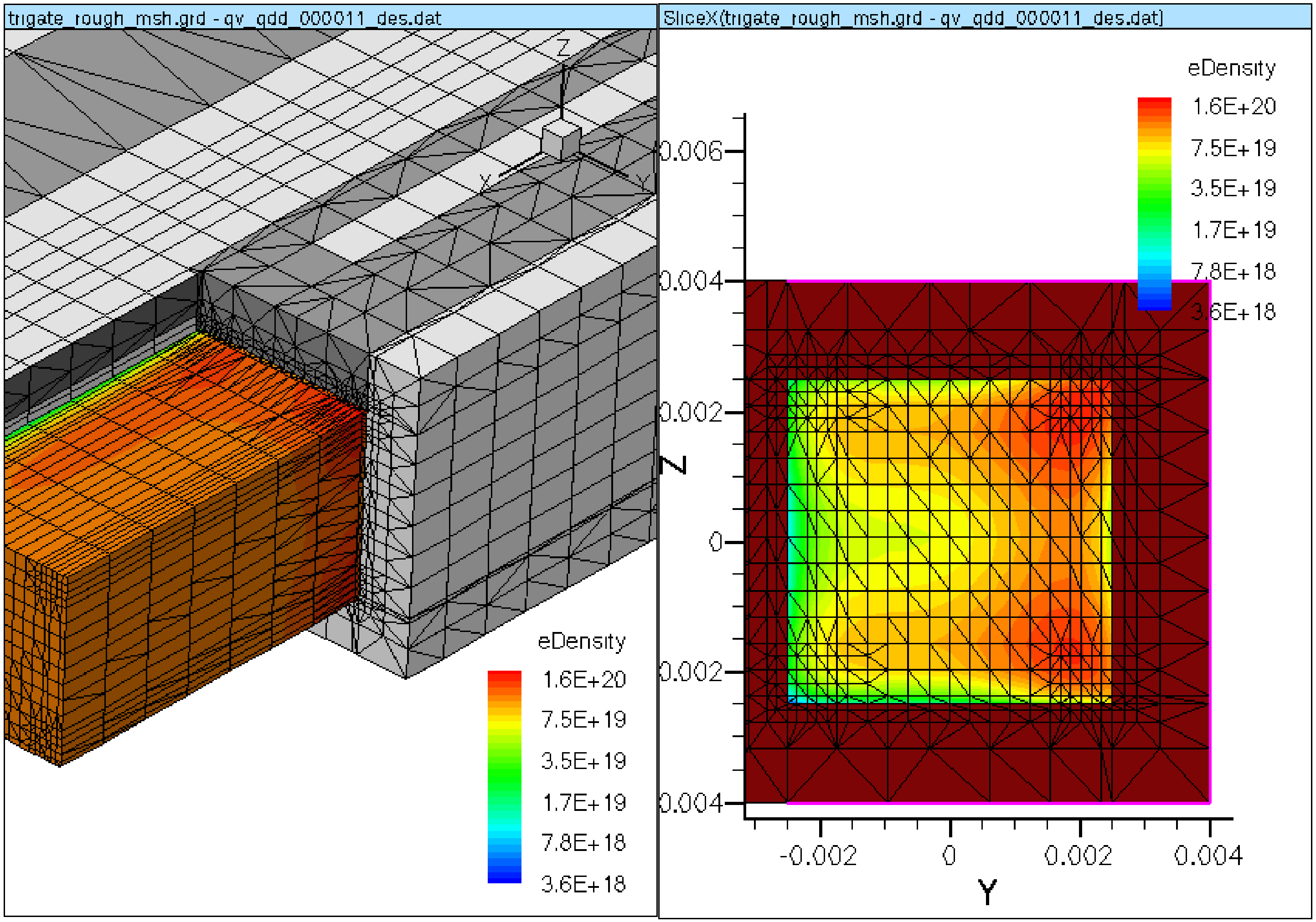


TGNW-FET (courtesy EU SINANO project)
Gate length: 25 nm (65 nm technology node).
Channel Cross-Section : Square (5x5 nm^2).
Source/drain extensions: 10 nm.
Oxide parameters: material is SiO_2 (k~3.9).
Field Oxide Thickness : 1.5 nm.
Buried Oxide Thickness : 150 nm.
Gate electrode work function: 4.1 eV.
von Neumann boundary conditions at S/D ends.
<u>Doping specifications:</u>
Substrate undoped, source and drain: $N_d=10^{20}$ cm^{-3}.

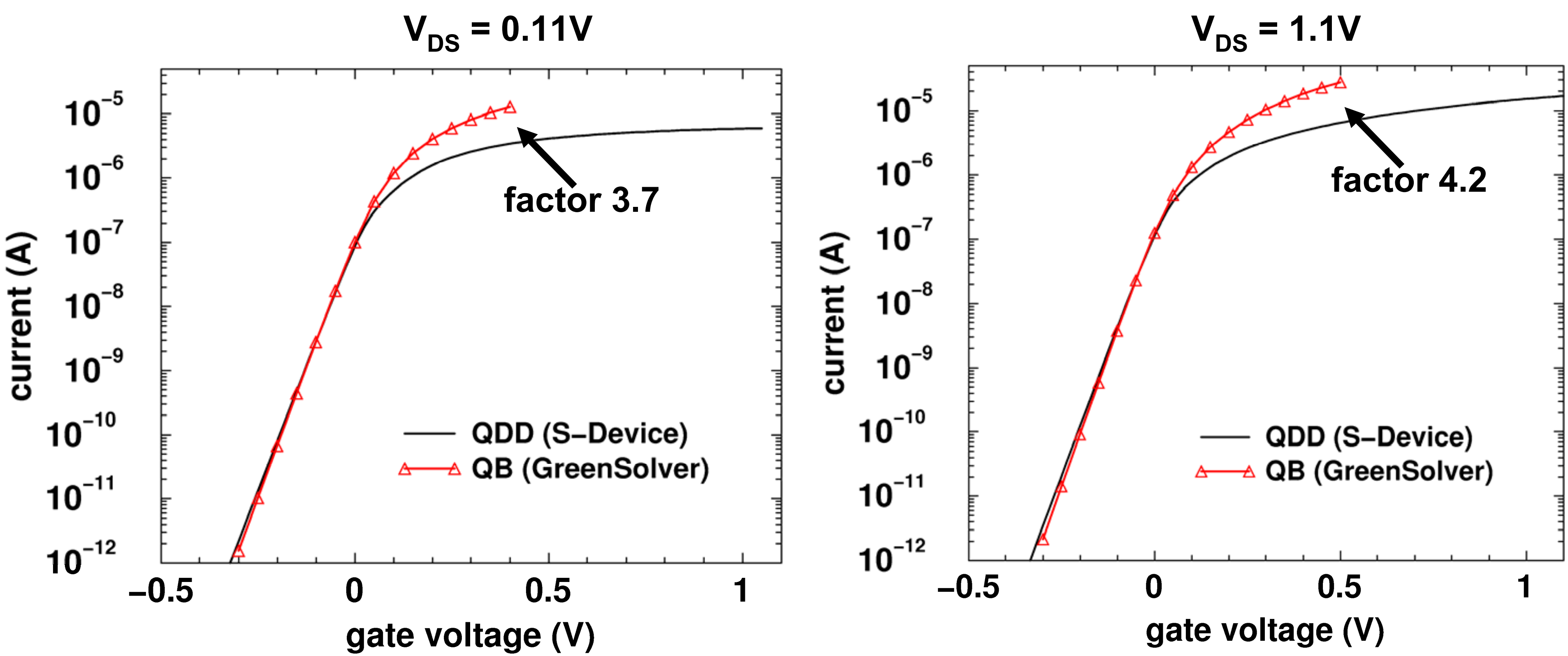


Limit of large cross sections

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



Comparison of currents



Incoherent scattering

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 $\hbar\omega_q = c_s q$

→ self-energy $\Sigma^<_{ac}$ becomes local in space

$$\Sigma^<_{ac}(r, r', E) = \frac{\Xi^2 k_B T}{\rho c_s^2} G^<(r, r', E)\delta(r - r')$$

$G^<$ requires G^R (full size): $G^< = G^R(\Sigma^<_S + \Sigma^<_D + \Sigma^<_{ac})G^A$

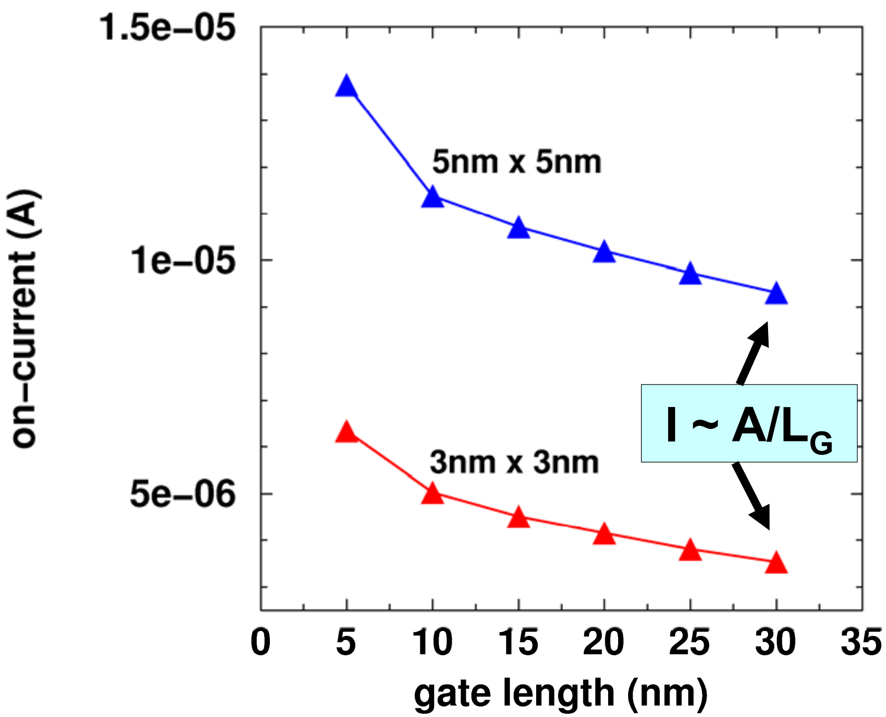
$$(E - H - \Sigma^R)G^R = \mathbb{1}$$

Elastic appr. $E \pm \hbar\omega_q \approx E$ → Σ^R simplifies to

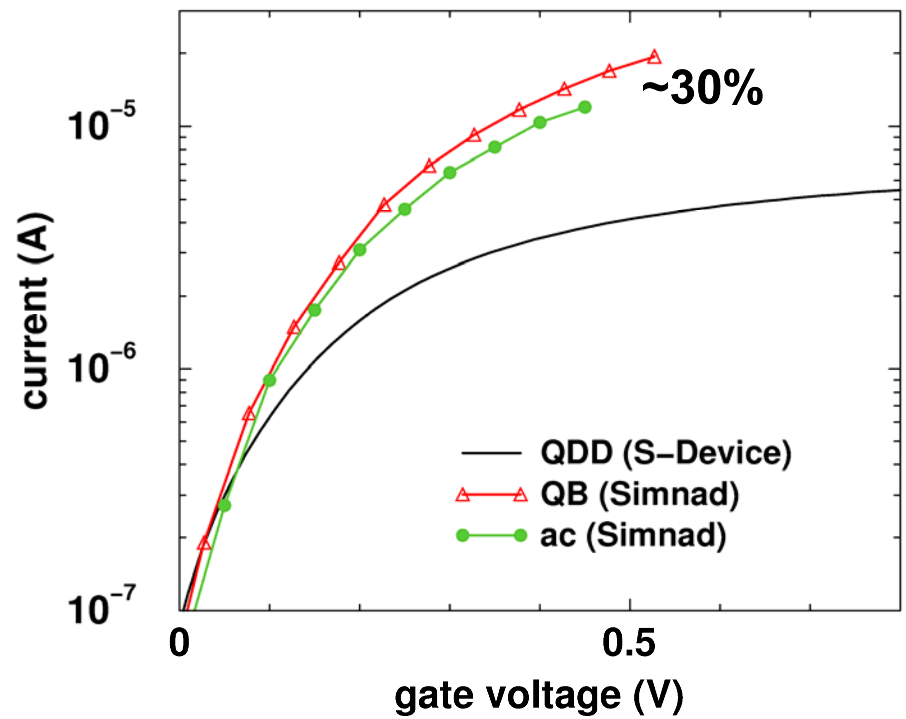
$$\Sigma^R(r, r', E) = \frac{\Xi^2 k_B T}{\rho c_s^2} G^R(r, r', E)\delta(r - r')$$

Incoherent scattering

Scaling @ V_{GS}=0.5V, V_{DS}=50mV



TG 5 x 5 x 25 nm³ NW FET



Conclusion

- **Atomistic, full-band approach to simulate Si NW FETs is possible (and justified) up to 5x5 nm^2 cross sections**
- **A variety of effects (channel orientation, strain, surface roughness, S-D tunneling, gate tunneling) can be studied in the quantum-ballistic limit**
- **For *ballistic* transport use WF formalism and not NEGF, because much more efficient!**
- **Challenges: Incoherent scattering, CPU time**