Simulation of Quantum Effects in Nanoscale Devices

Andreas Schenk [1]

Mathieu M. Luisier [2], Martin Frey [1], Aniello Esposito [1]

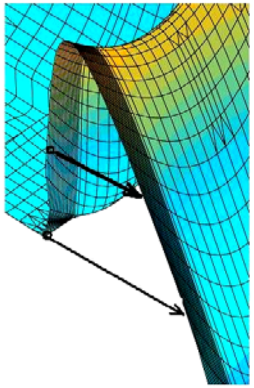
1 ETH Zürich
2 Purdue University

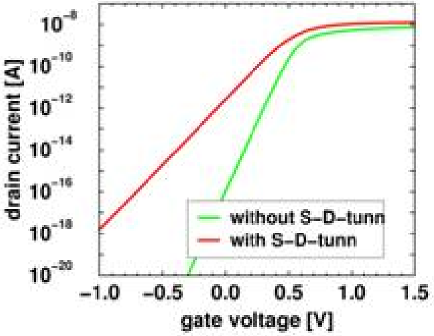
Outline

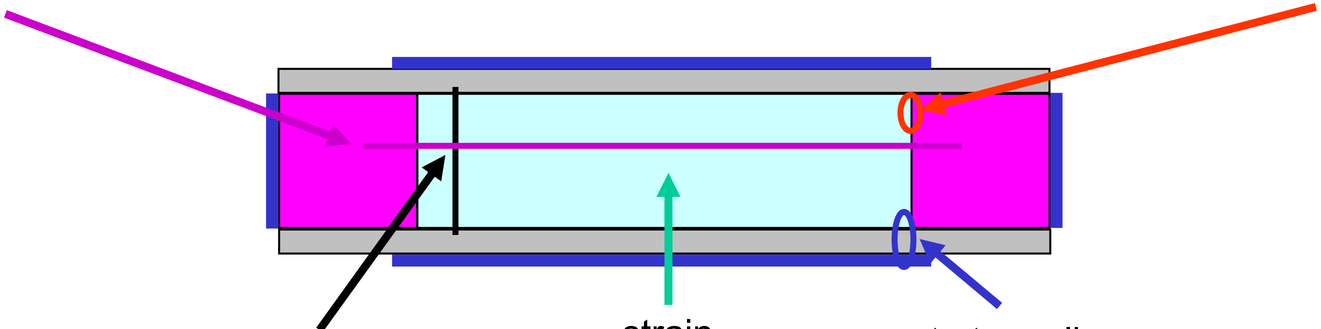
- Introduction
- Quantum-mechanical confinement effects
    - quantum V_T-shift
    - comparison single, double, triple, and surround gate
- Simulation of quantum-ballistic ON-currents
    - Band structure, transport, electrostatics
    - ON-current for different channel orientations
    - Effect of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

Which quantum effects?

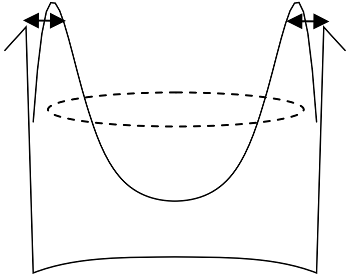
source-to-drain tunneling

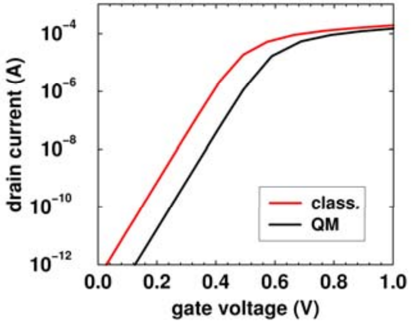




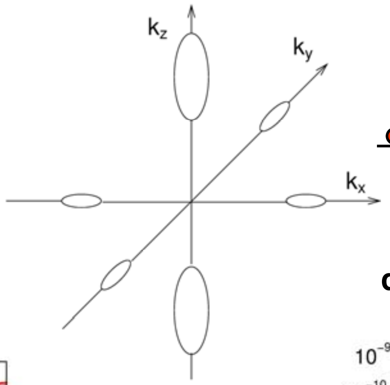


Confinement, quantization
- **quantum V_T shift**
- **QM mobility**

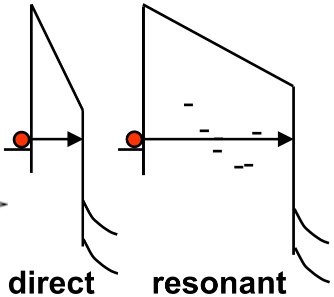




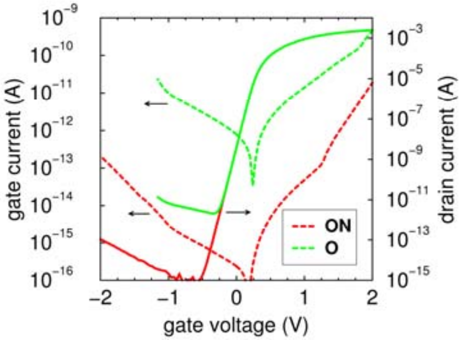
strain



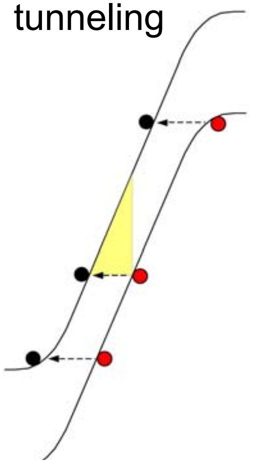
gate tunneling

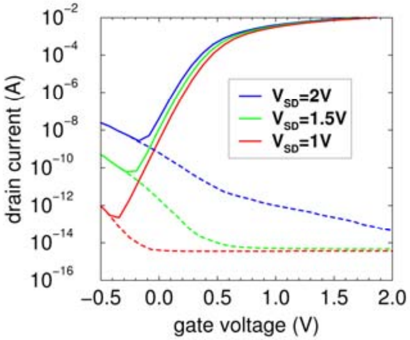


direct **resonant**



band-to-band tunneling

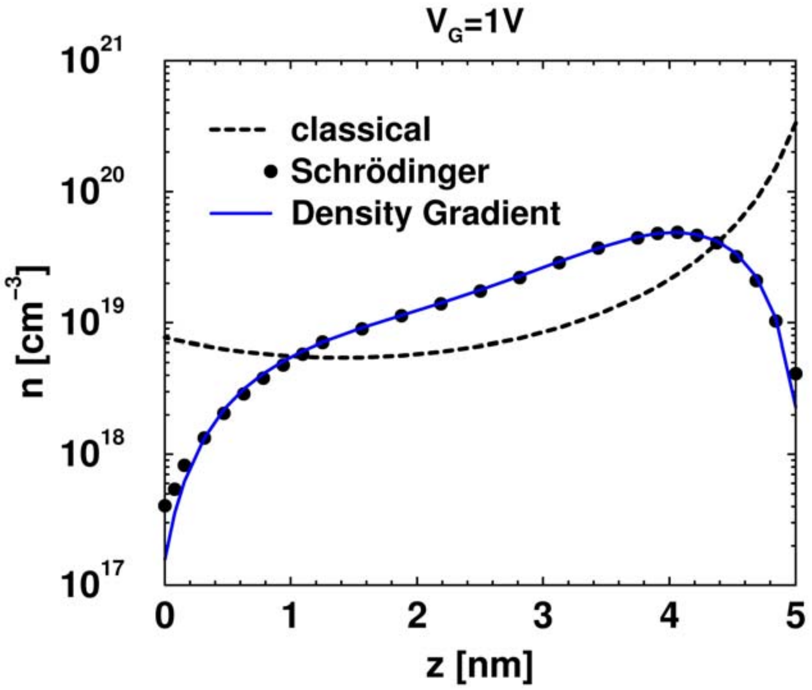




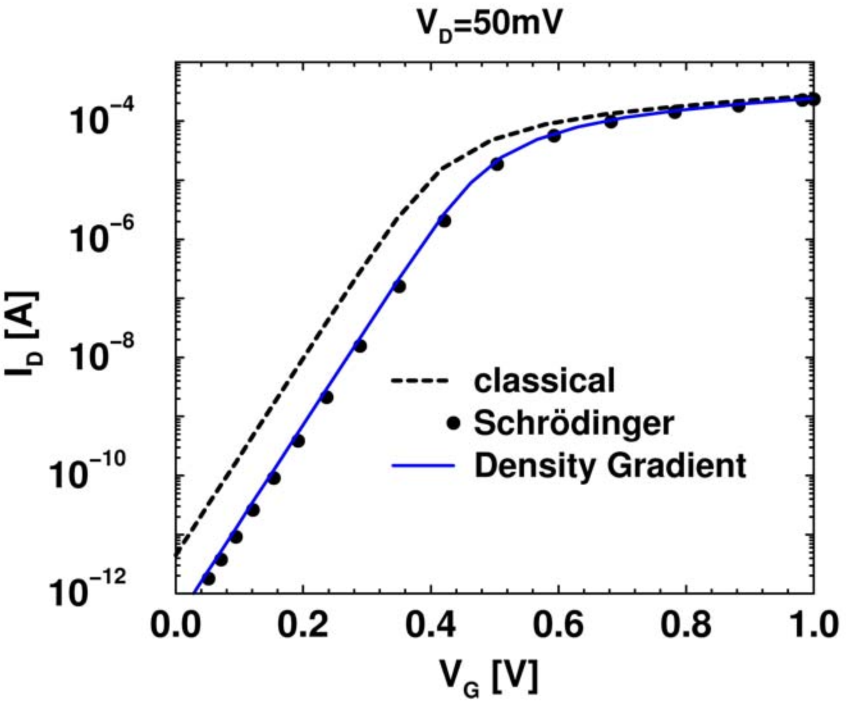
Quantum-mechanical confinement effects

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

quantum V_T-shift

A ‘*quantum potential*’ Λ 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:

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

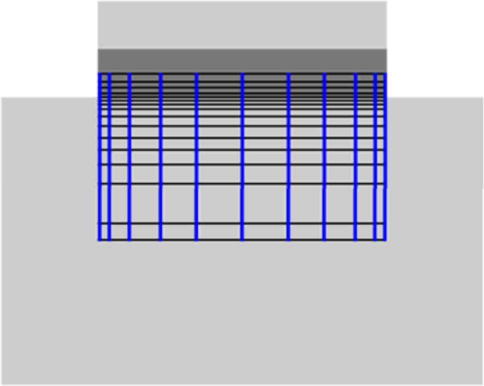
1D Schrödinger-Poisson solver

Λ follows by equating the density *n*[*R*] with the expression

$$n(z) = \frac{1}{\beta\pi\hbar^2}\sum_{j,\nu}\left|\Psi_j^{(\nu)}(z)\right|^2 m_{xy}^{(\nu)}(z)\,\exp\left[\beta\left(E_{F,n}(z) - E_j^{(\nu)}\right)\right]$$

quantum V_T-shift

Boundary condition at the ends of the domain [z^-,z^+], defining the '*quantum box*':



$$\Psi_j^{(\nu)\prime}/\Psi_j^{(\nu)} = \pm\sqrt{2m_z^{(\nu)}|E_j^{(\nu)} - E_c|}/\hbar$$

Full Newton impractical, therefore approximation

$$\partial n_i^{\mathrm{qm}}/\partial\Phi_j \approx -\delta_{ij}\,\partial n_i^{\mathrm{qm}}/\partial E_{\mathrm{F},i}$$

Density Gradient model

Λ is given by the PDE

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

$$= -\gamma\frac{\hbar^2}{6m}\frac{\nabla^2\sqrt{n}}{\sqrt{n}}$$

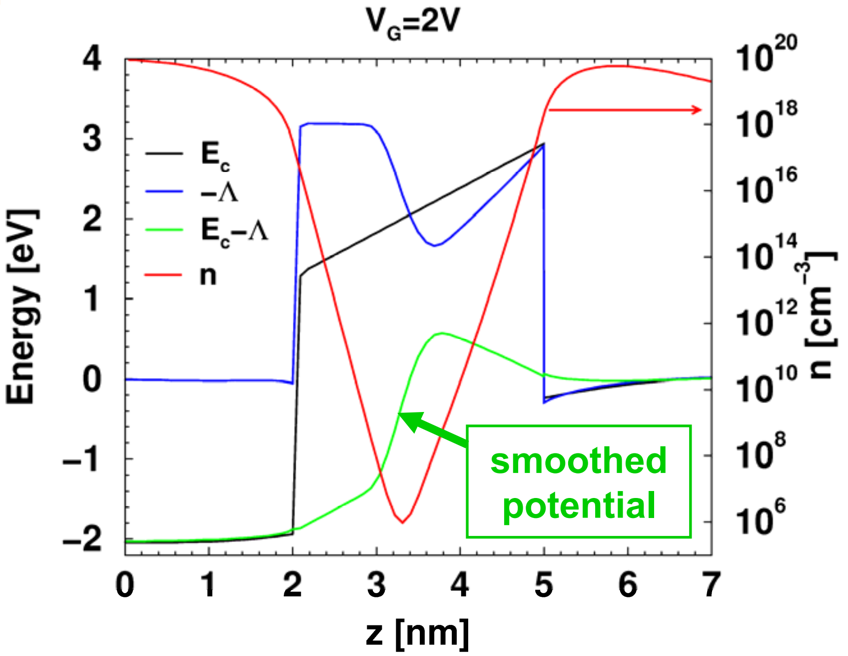
quantum V_T-shift

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 *ρ* by approx. for equilibrium density matrix ρ_0

Assumptions: equilibrium density, isotropy of effective mass, $\delta\Phi/k_BT \ll 1$ (Born approx.)



Generalization for device simulation:

- **DOS mass → *m***
- **non-equilibrium density → *n***
- **non-perturbative formulation of Λ with the *smoothed potential*** $\bar{\Phi} = E_c + \Phi_m + \Lambda$

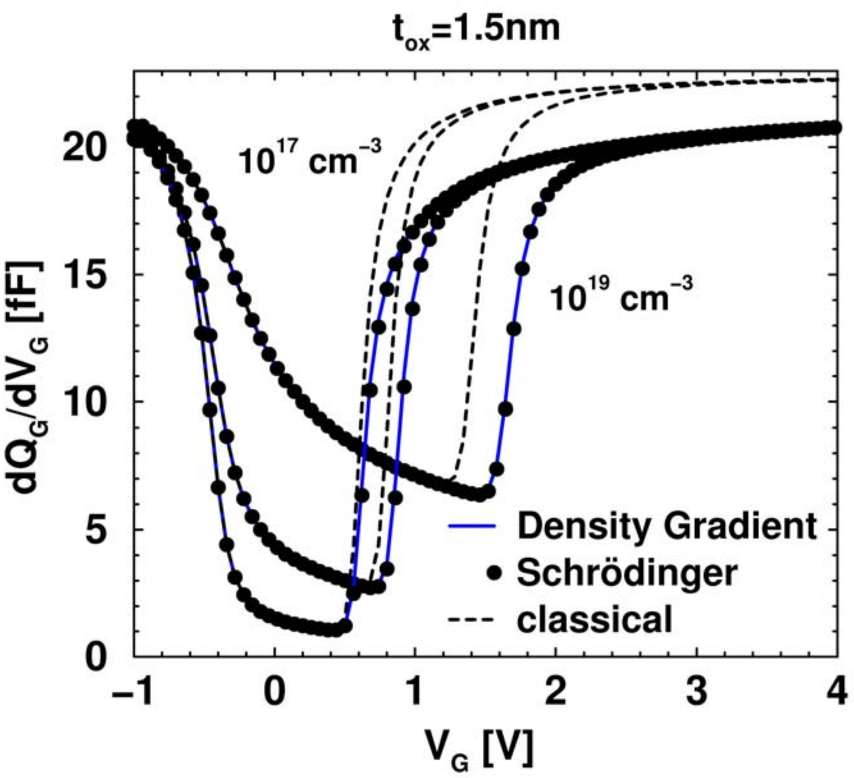
$$\Lambda = -\gamma\frac{\hbar^2}{12\,m}\left\{\nabla^2\left(\beta E_{F,n} - \beta\bar{\Phi}\right) + \frac{1}{2}\left[\nabla\left(\beta E_{F,n} - \beta\bar{\Phi}\right)\right]^2\right\}$$

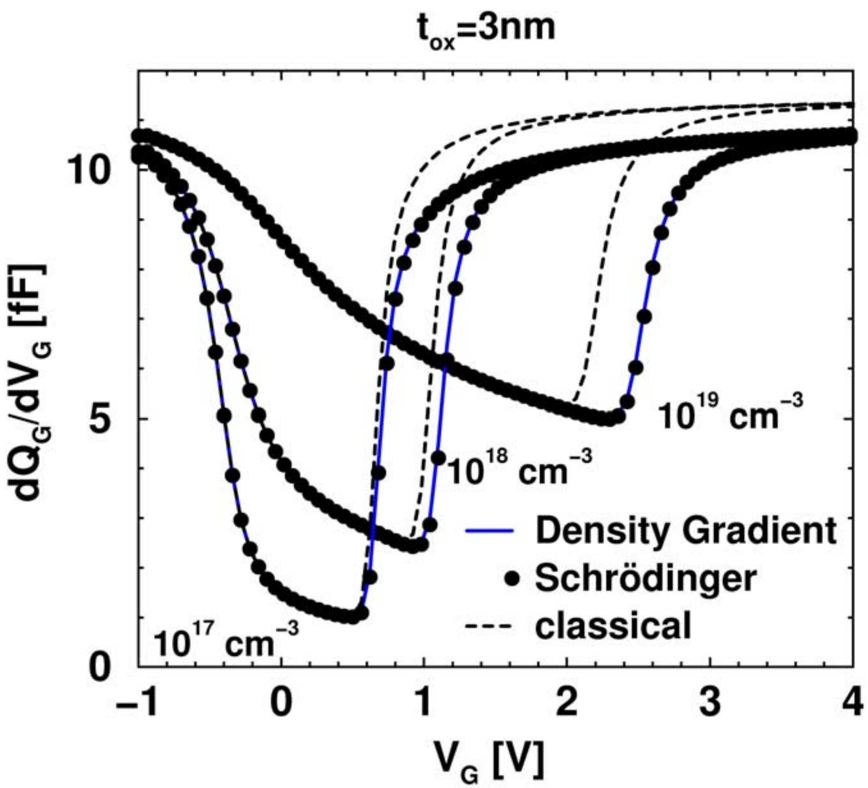
Λ is new system variable, coupled Newton

DG is multi-dimensional, pre-factor *γ* is “universal” (3.6 for Si, if $m = m_{DOS}$)

quantum V_T-shift

MOS (with poly) capacitor CV curves



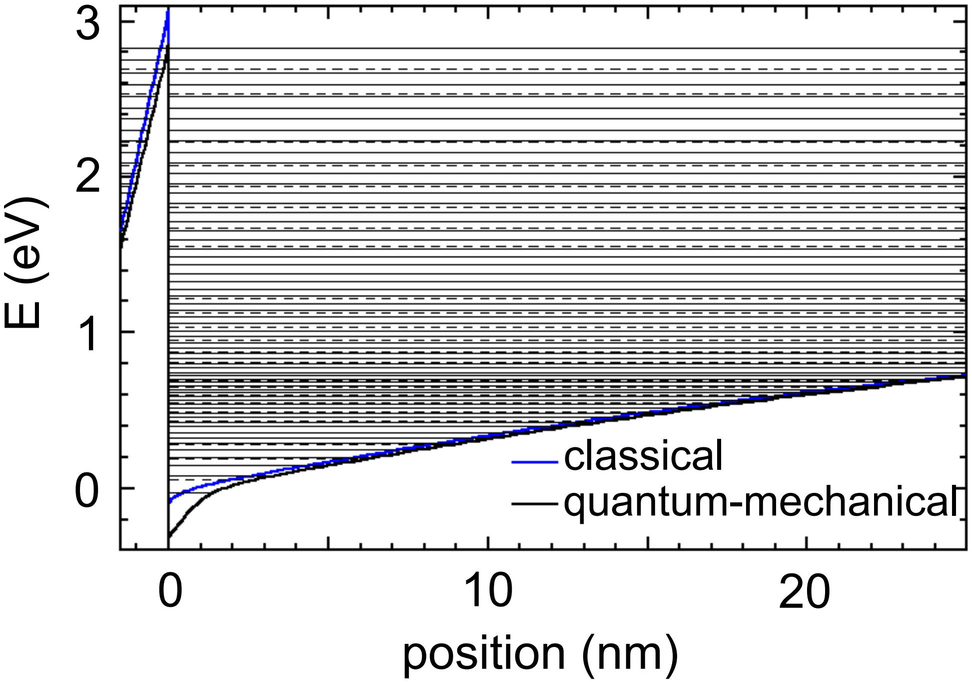


N_{poly}=1e20 cm^{-3}, A_G=1 µm^2

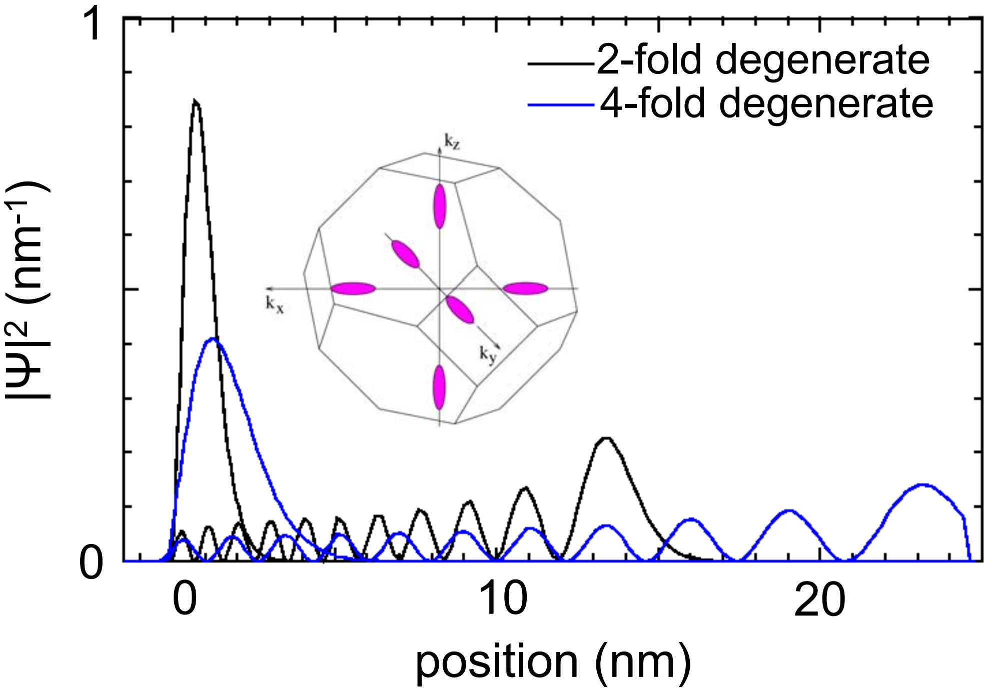
quantum V_T-shift

MOS (with poly) capacitor

eigenenergies



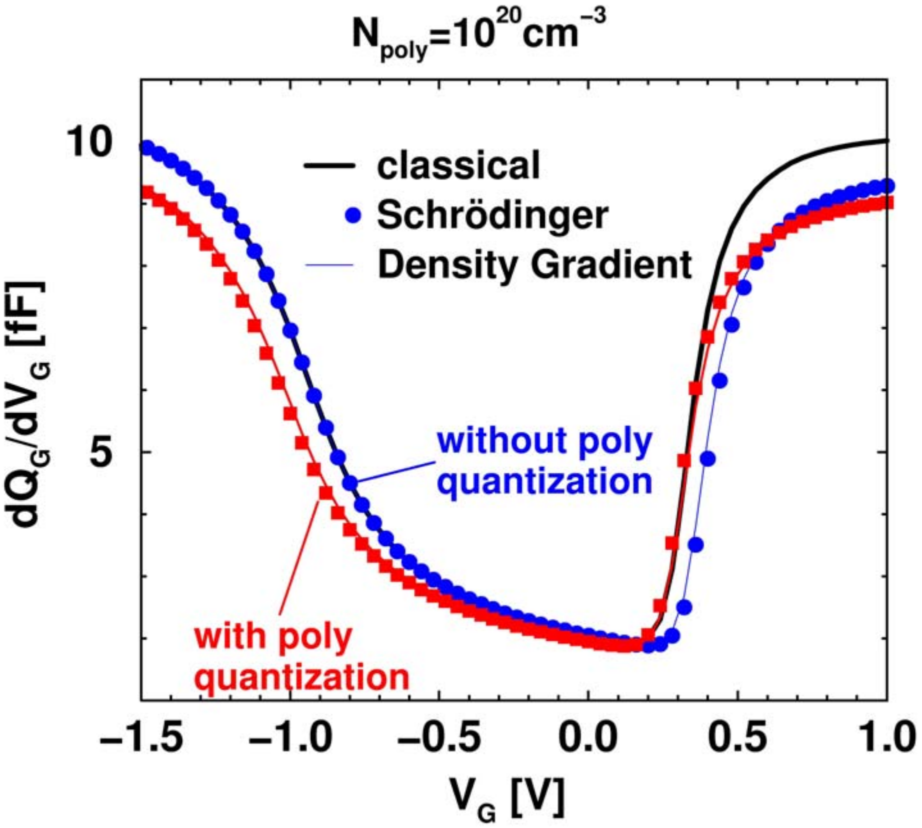
wave functions



N_A=5e17 cm^{-3}, t_{ox}=3 nm, V_G=3 V

quantum V_T-shift

Quantum depletion at poly-SiO_2 interface

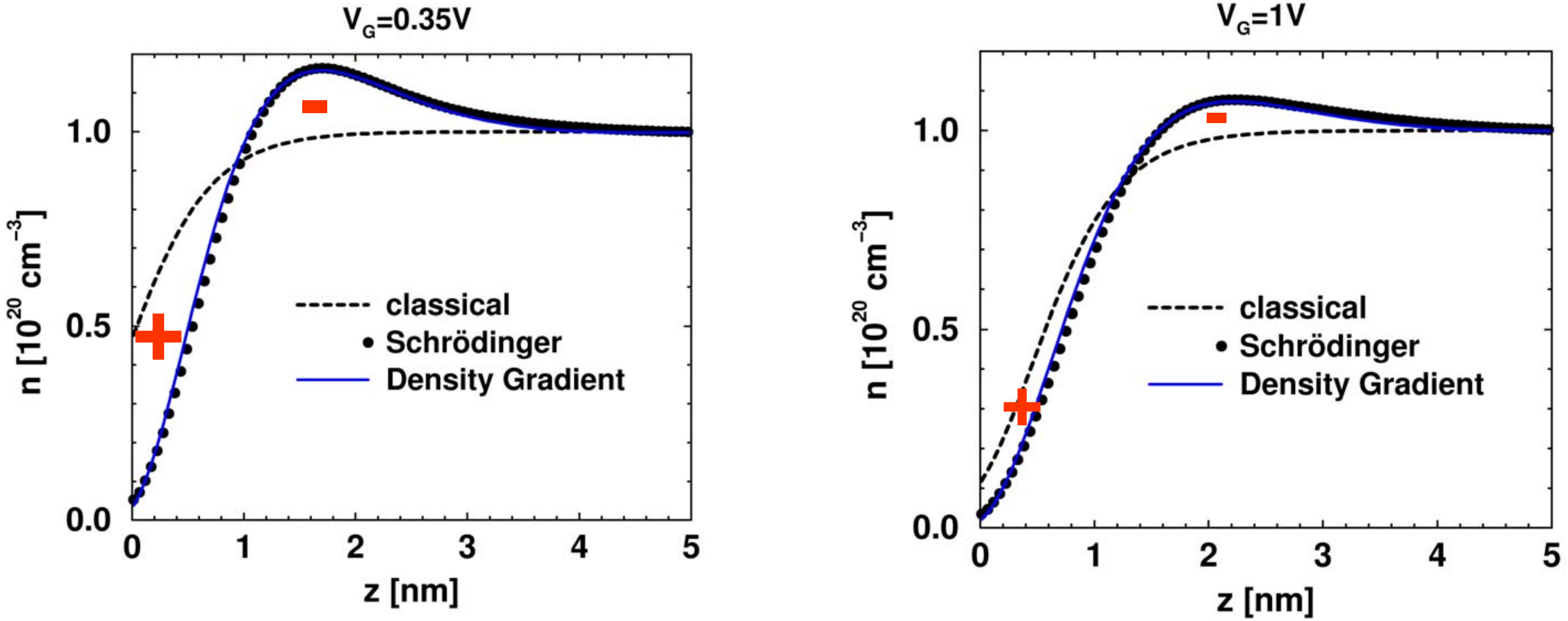


compensation of quantum shift at threshold for high poly doping (~1e20 cm^{-3})!

N_A=5e17 cm^{-3}, t_{ox}=3 nm, A_G=1 μm^2

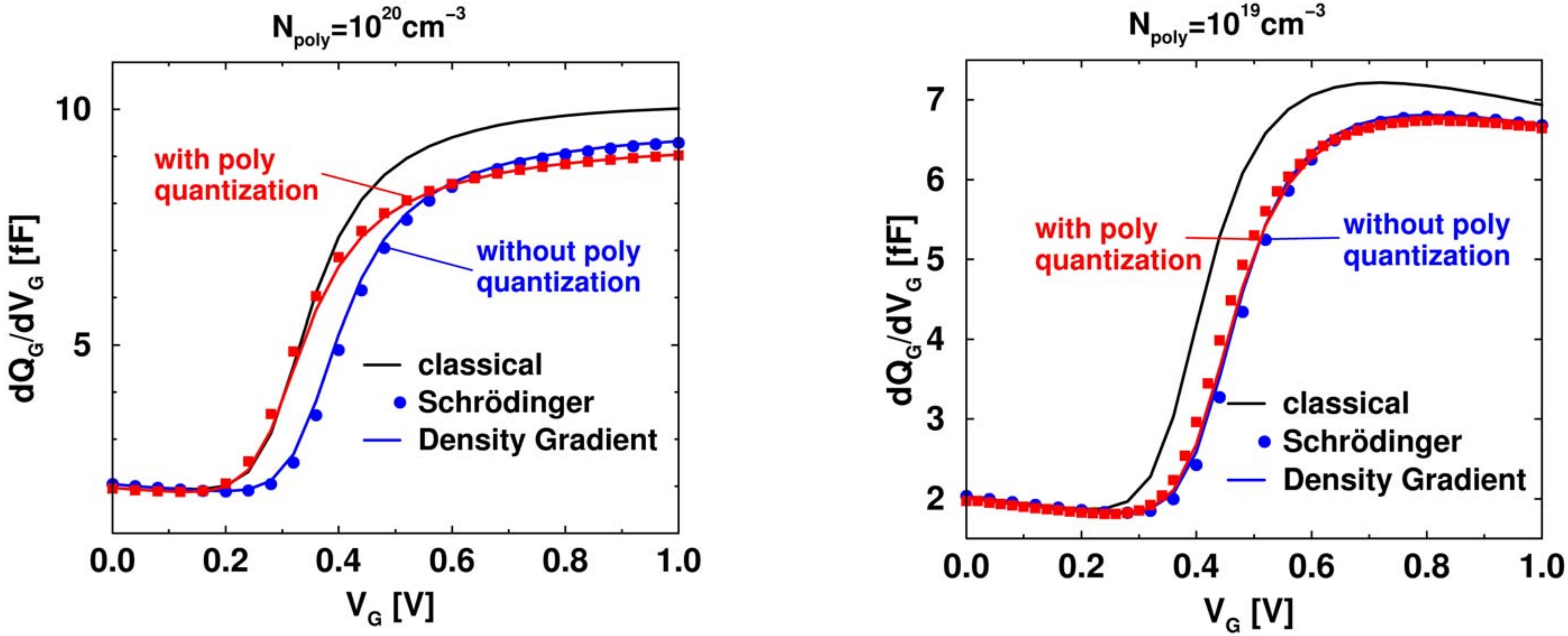
quantum V_T-shift

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)

Effect on CV curves



- **strength of the quantum dipole depends on doping level within the first few nanometers**
- **no effect on CV, if $N_{poly} < 1e19\ cm^{-3}$ at the interface**

quantum V_T-shift

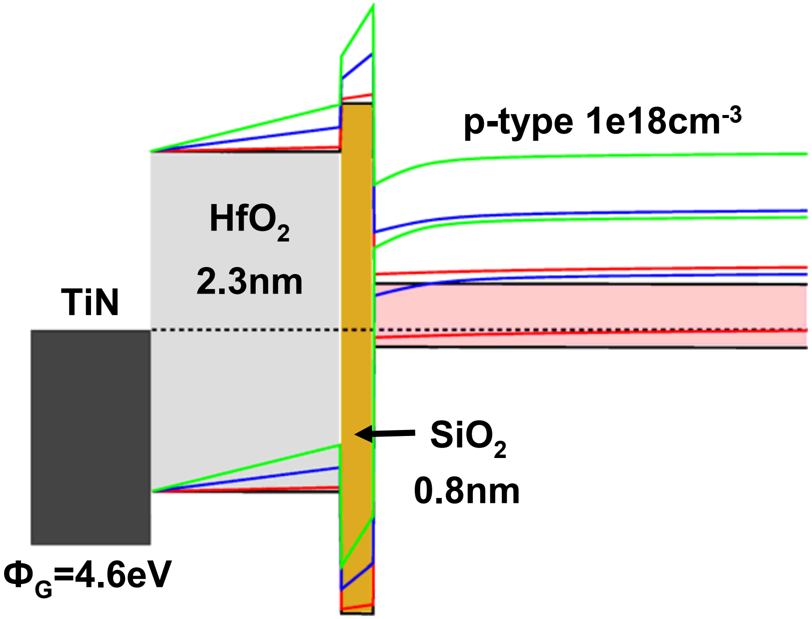
Effect of exchange-correlation potential on CV curves

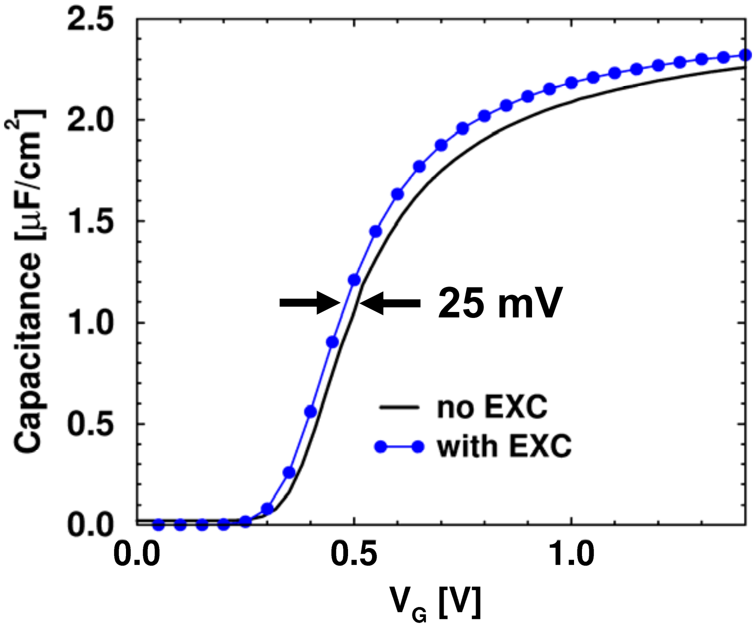
$$\epsilon_{\mathrm{X}} = -\frac{q^2}{4\pi\epsilon_0\epsilon_{\mathrm{r}}}(3\,n)^{1/3} = -0.909\cdot 10^{-8}\,n^{1/3} \quad [\mathrm{eV}]$$

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

LDA ε_{XC} (ε_C is continuous fit to Perdew-Zunger theory)

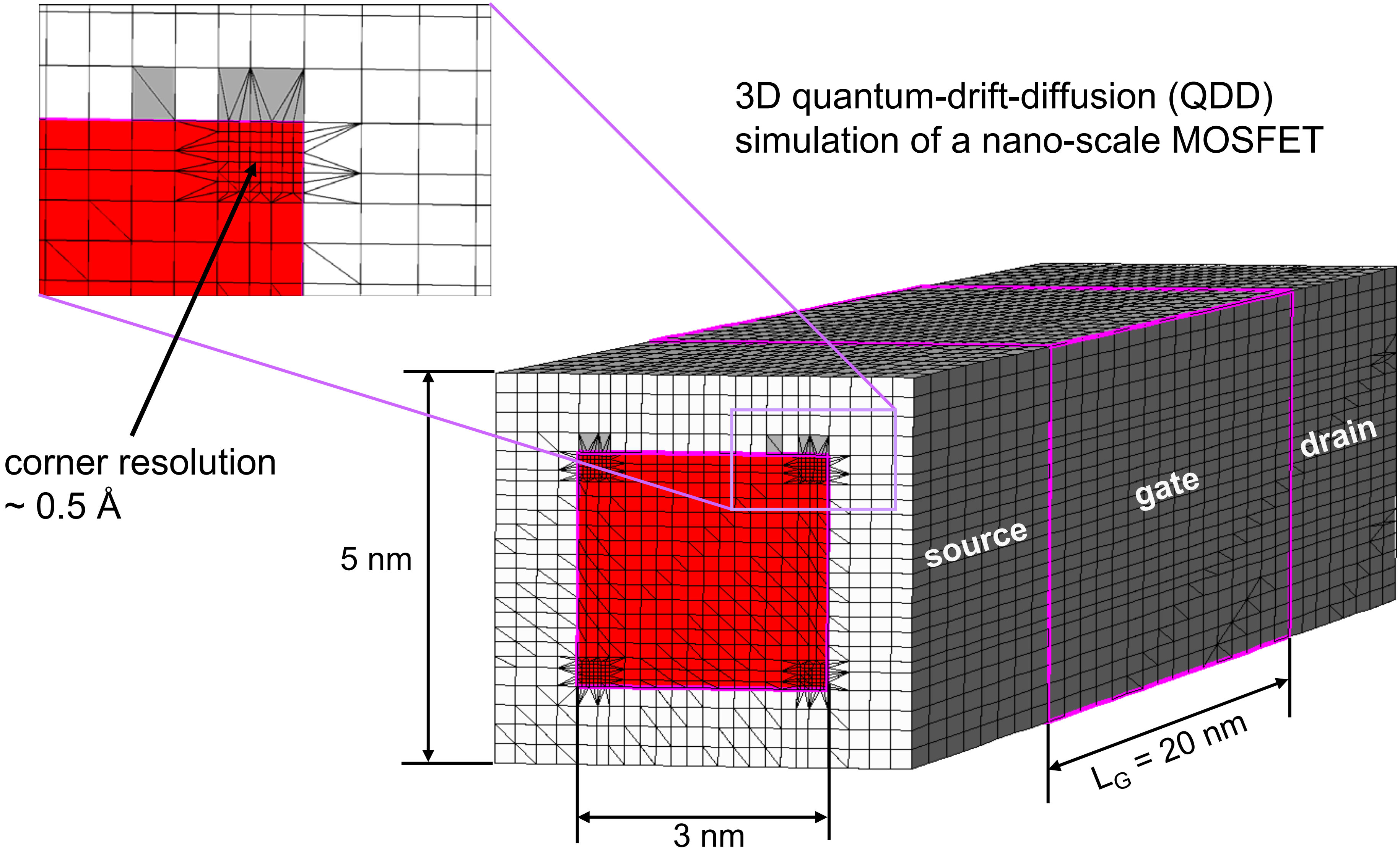
- **Many-body effect of exchange-correlation shows up as negative V_T-shift (band gap narrowing)**
- **Image-force already taken into account by boundary condition for Hartree potential (Poisson equation)**





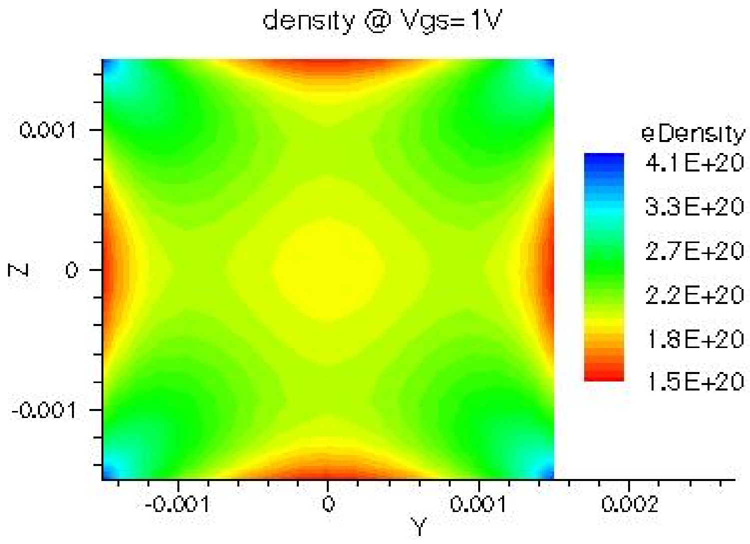
comparison single, double, triple, and surround gate

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

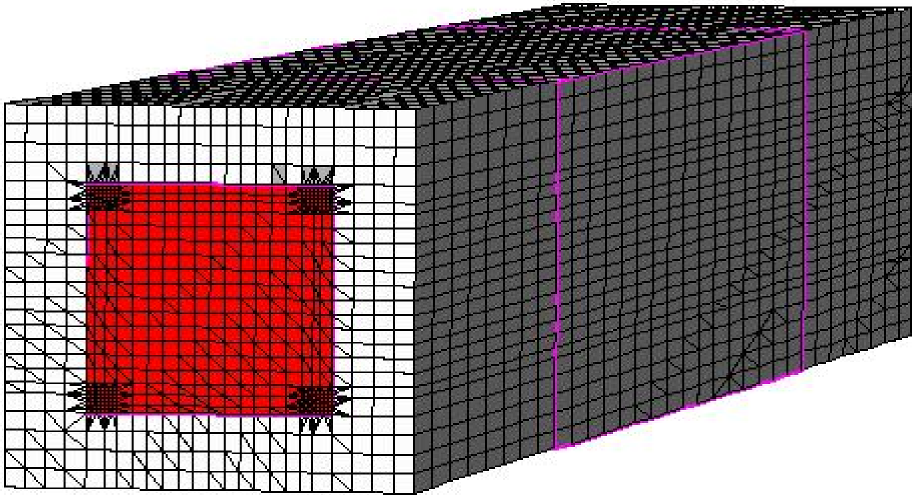


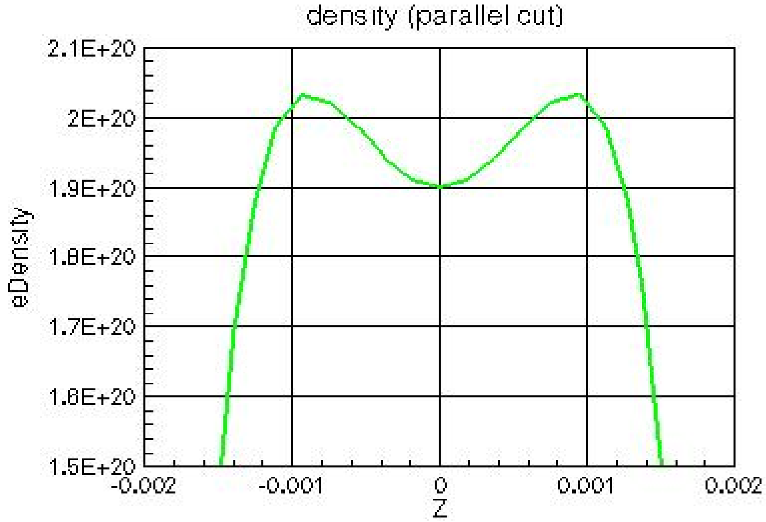
comparison single, double, triple, and surround gate

Influence of mesh refinement

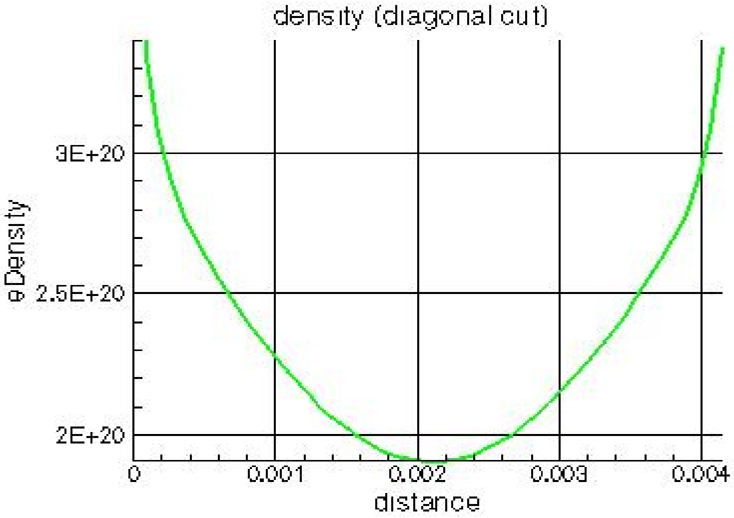


density @ Vgs=1V





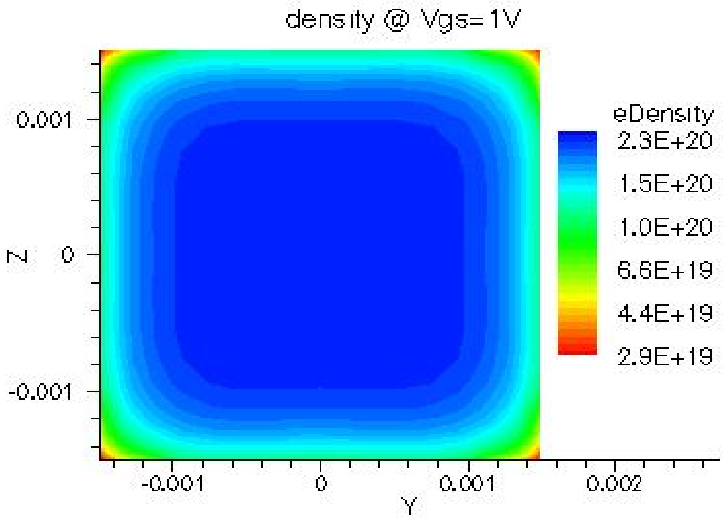
density (parallel cut)



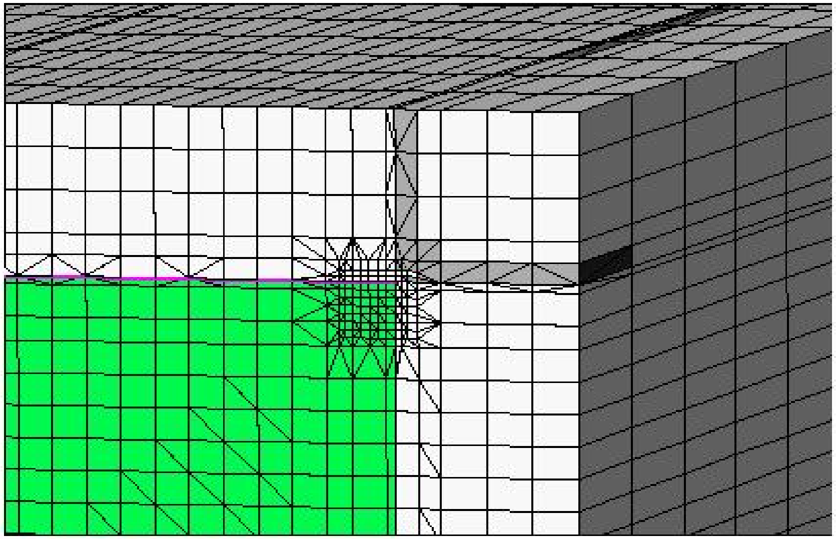
density (diagonal cut)

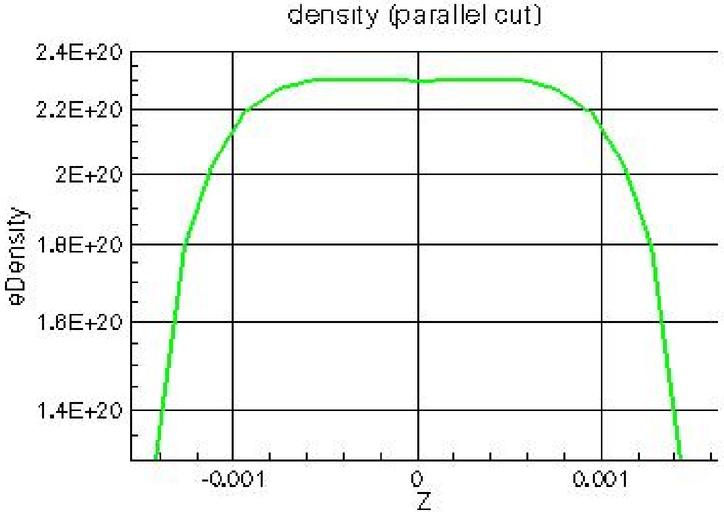
comparison single, double, triple, and surround gate

Influence of mesh refinement

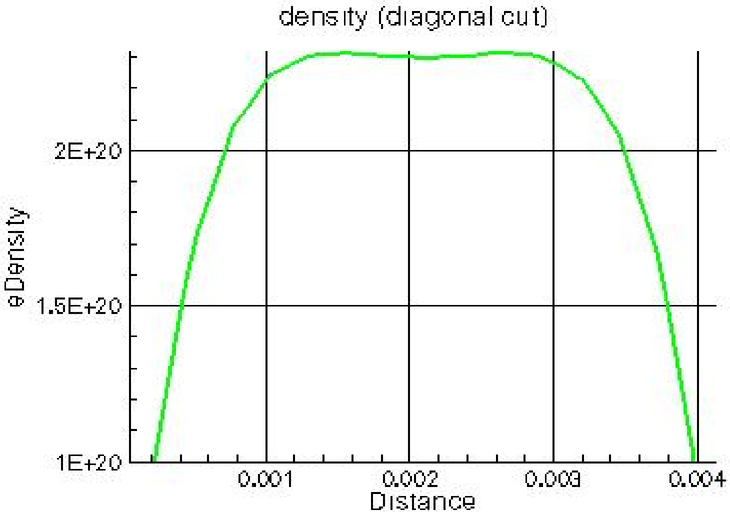


density @ Vgs= 1V





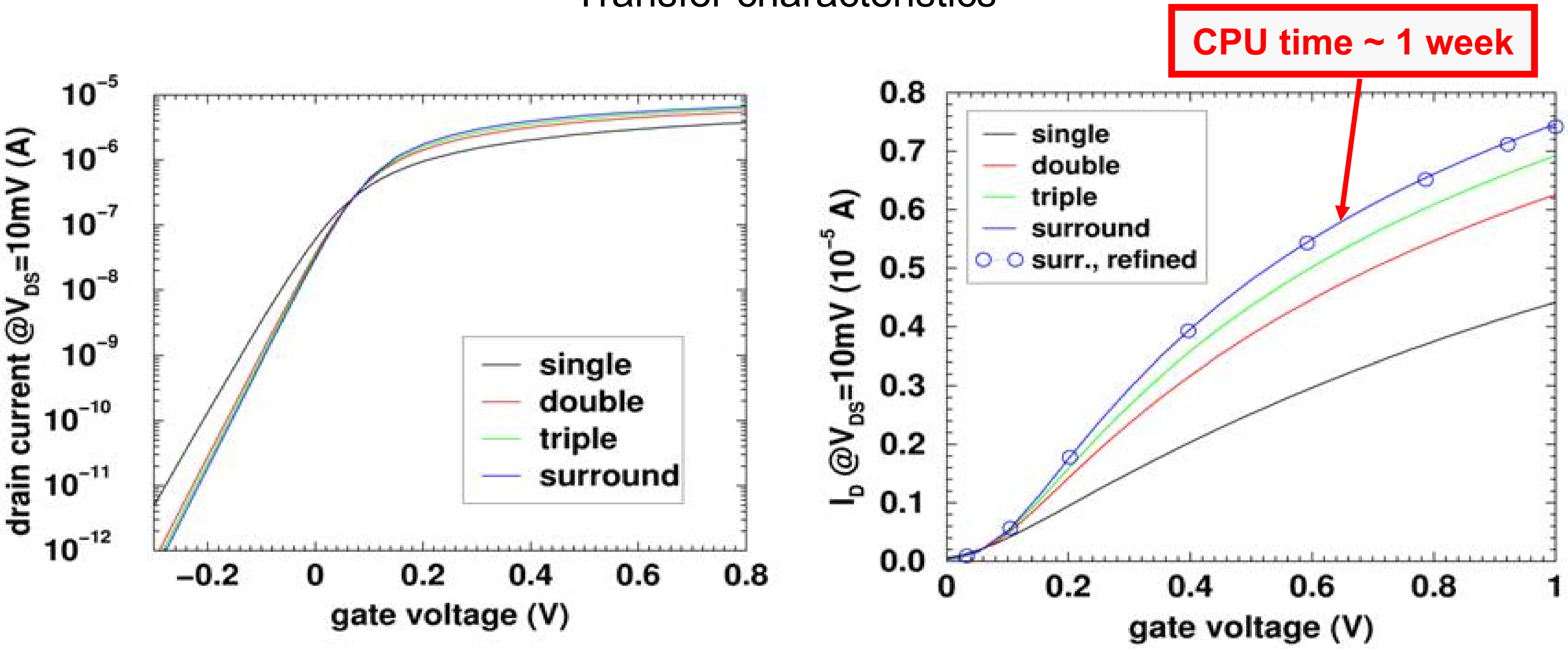
density (parallel cut)



density (diagonal cut)

comparison single, double, triple, and surround gate

Transfer characteristics



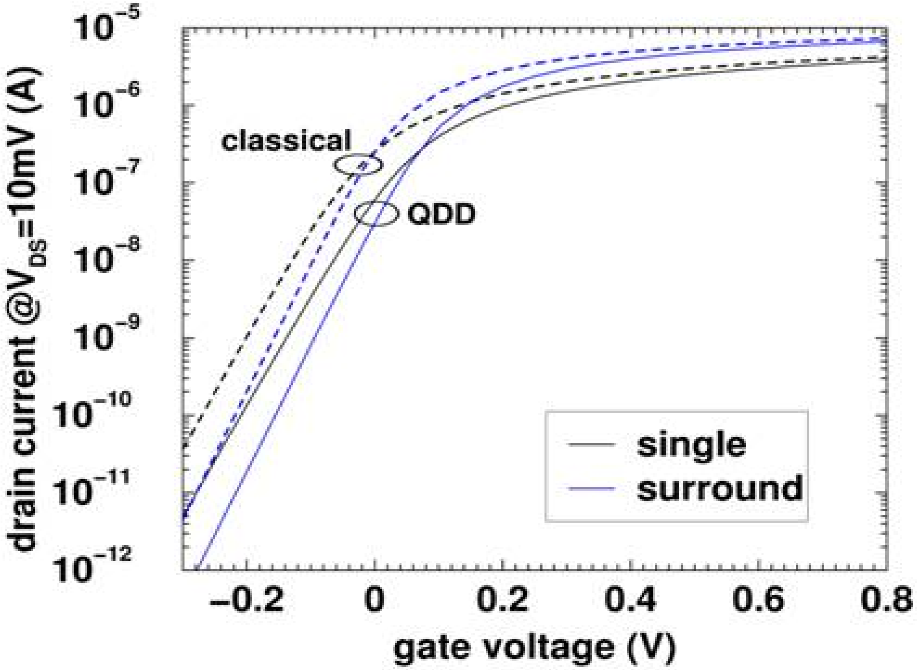
CPU time ~ 1 week

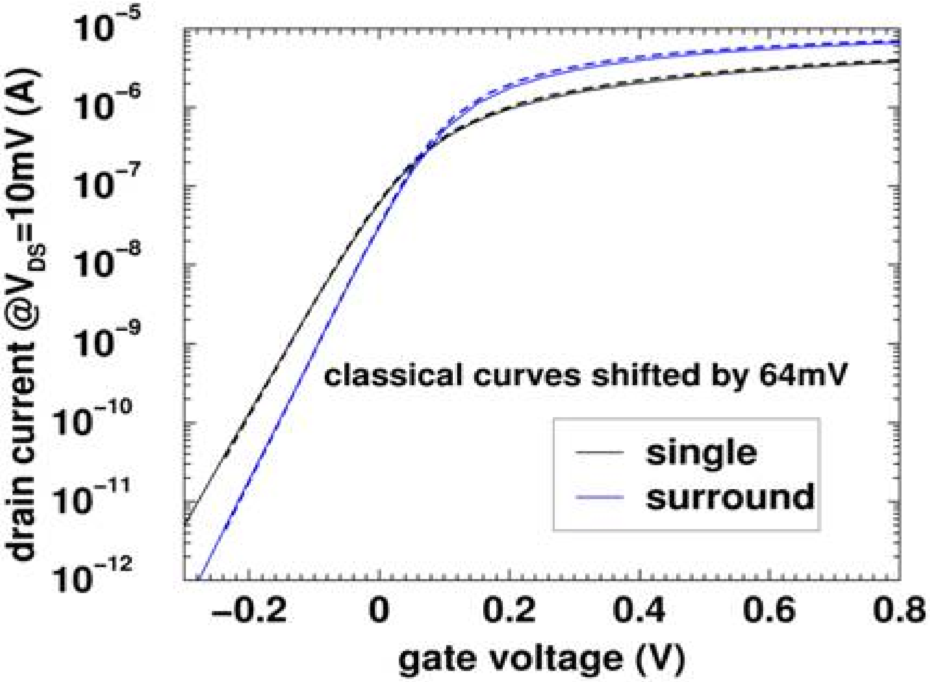
assumption: isotropic, classical mobility

- **no corner effect (FD, no channel doping, 3x3 nm wire)**
- **only little improvement from double → surround (gate overlap, e.g. triple gate is an effective Π-gate)**

comparison single, double, triple, and surround gate

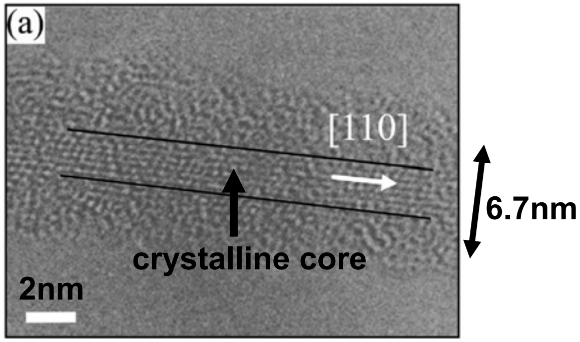
Transfer characteristics (contd.)





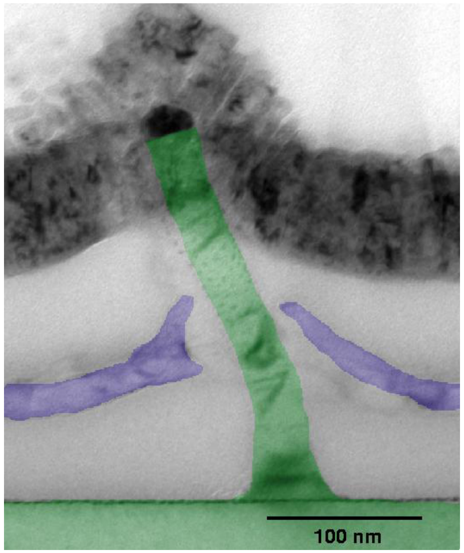
- **quantum V_T-shift of ~ 64 mV, independent of gate configuration**
- **almost perfect shift on V_{GS}-axis → quantum V_T-shift can be translated into work function difference**

Simulation of quantum-ballistic ON-currents



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

Vertical Schottky barrier FETs.



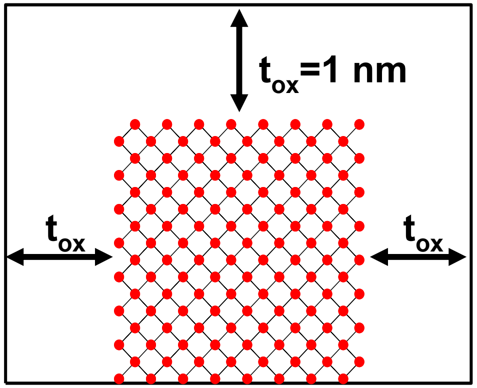
IBM Zürich, Walter Riess

- 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 nm^2 => no full quantum transport simulation necessary
- if (in the future) cross sections < 5 x 5 nm^2, strong confinement => band structure effects become important
- predictive simulations then require: accurate band structure model, quantum transport solver (OBCs), self-consistent electrostatics

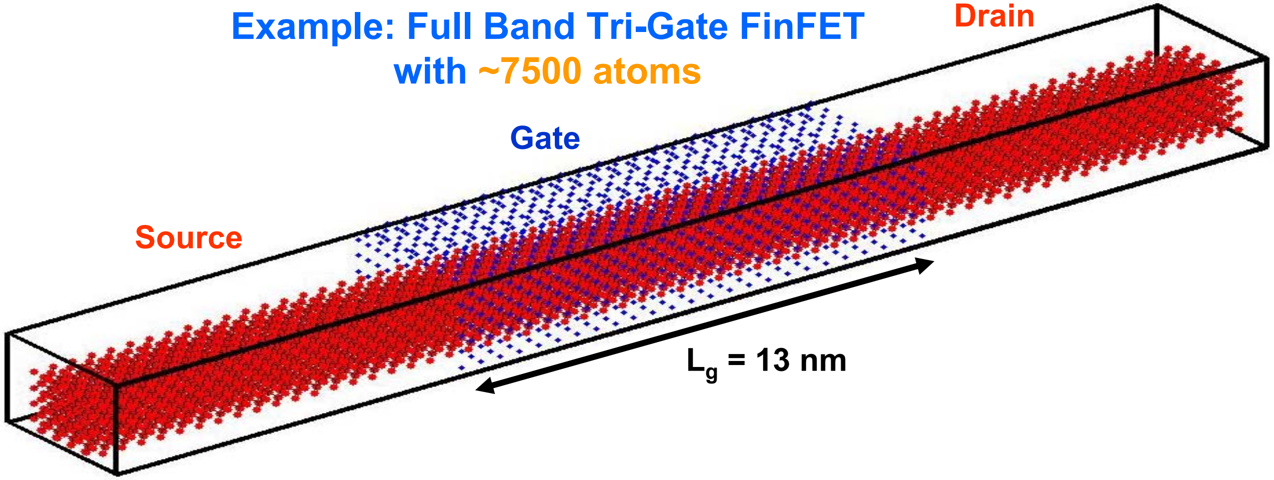
Band structure

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

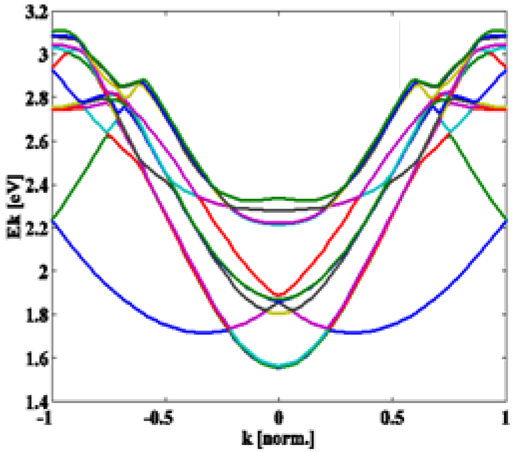
***Atomistic description* of a [100] nanowire with 2.1 x 2.1 nm^2 cross section**



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



TB band structure



Band structure

On-site and two-center integrals

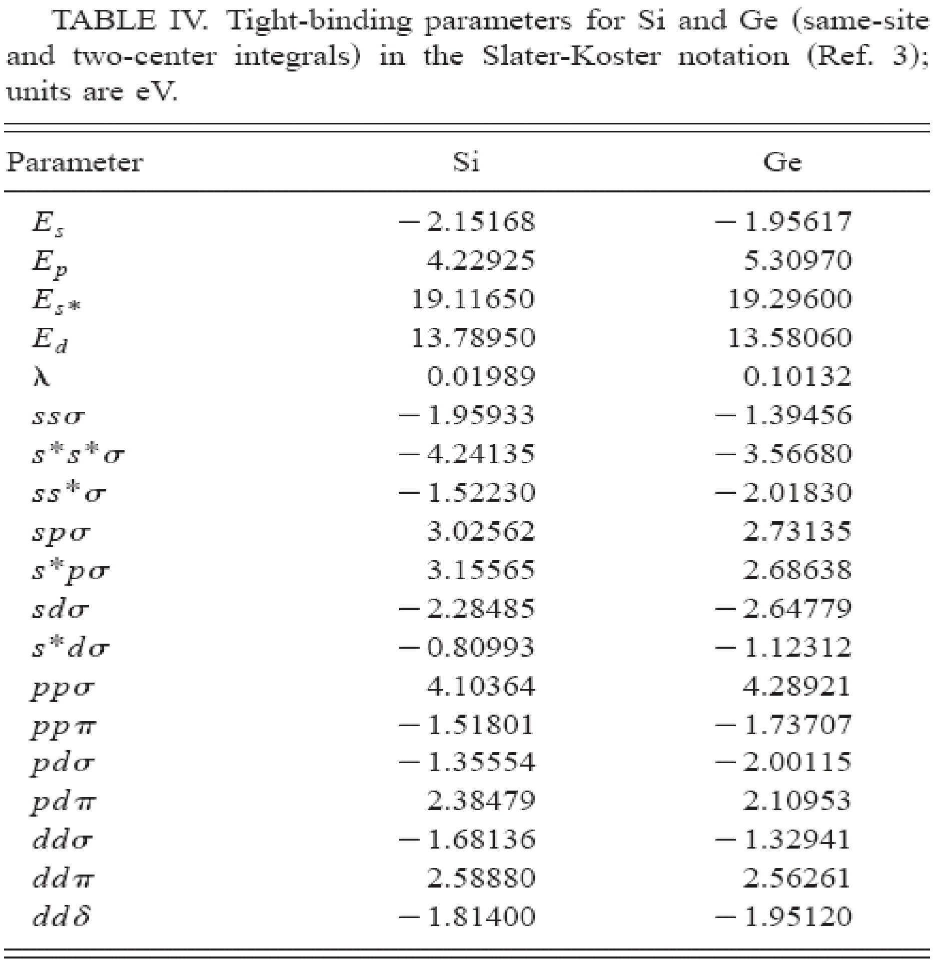


TABLE IV. Tight-binding parameters for Si and Ge (same-site and two-center integrals) in the Slater-Koster notation (Ref. 3); units are eV.

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

19 Parameters for Si and Ge:

- 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)

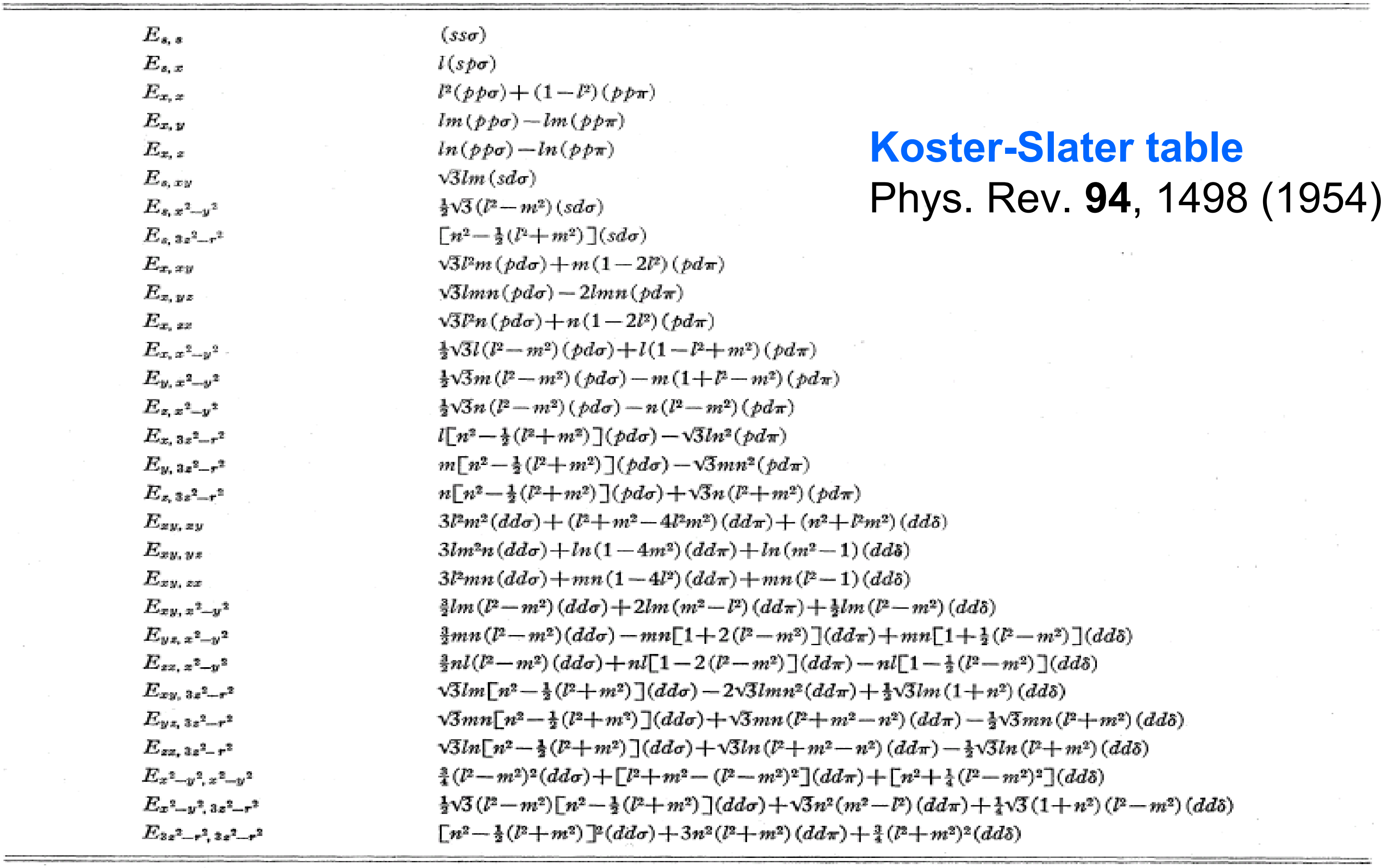
- Spin-orbit coupling (λ)

- 14 matrix elements between orbitals with 3 different possible bonds (σ, π, and δ). Symmetry + polarization not considered (Koster-Slater table)

Phys. Rev. B **69**, 115201 (2004)

Band structure

Energy integrals in terms of two-center integrals

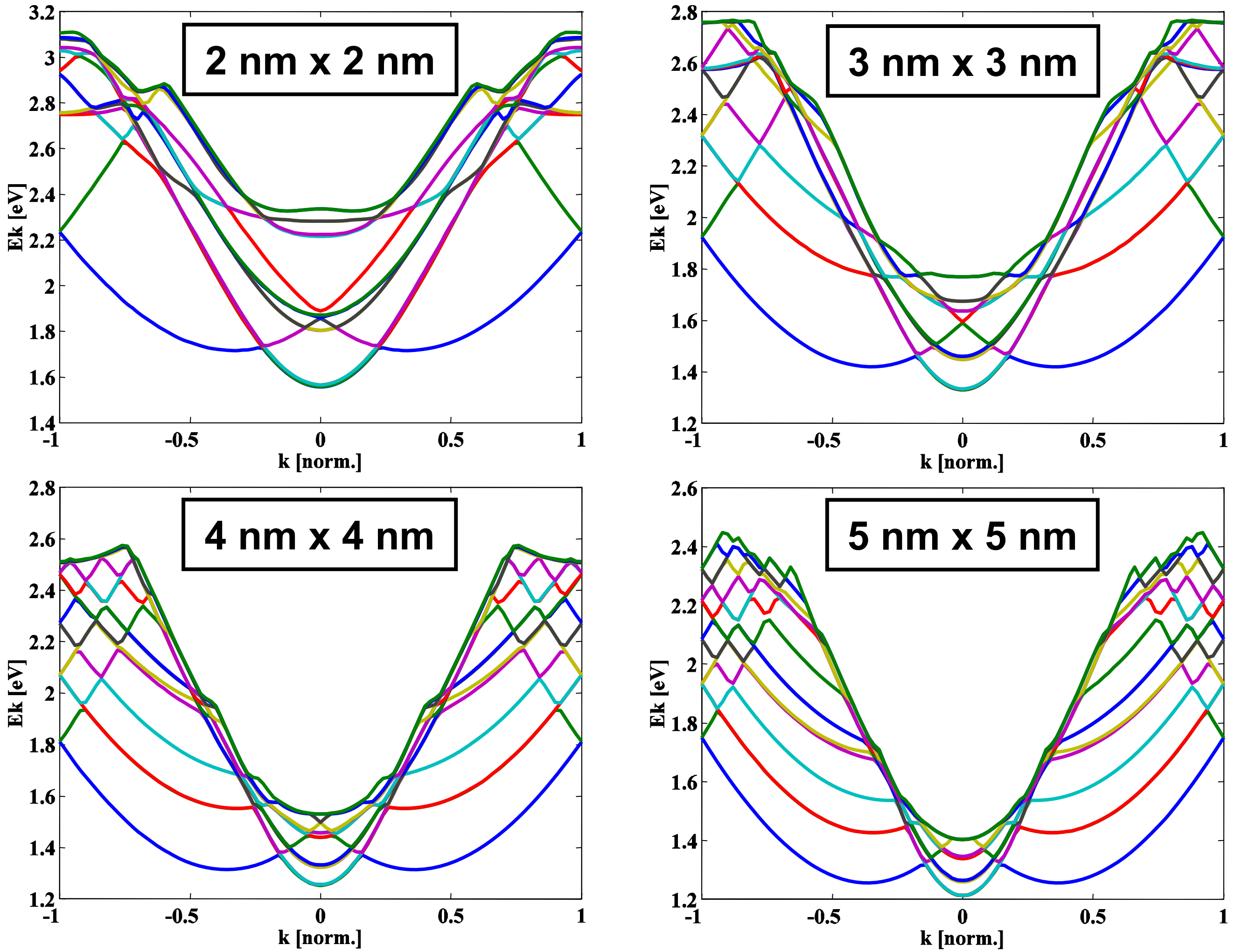


$E_{s,s}$	$(ss\sigma)$
$E_{s,x}$	$l(sp\sigma)$
$E_{x,x}$	$l^2(pp\sigma)+(1-l^2)(pp\pi)$
$E_{x,y}$	$lm(pp\sigma)-lm(pp\pi)$
$E_{x,z}$	$ln(pp\sigma)-ln(pp\pi)$
$E_{s,xy}$	$\sqrt{3}lm(sd\sigma)$
E_{s,x^2-y^2}	$\frac{1}{2}\sqrt{3}(l^2-m^2)(sd\sigma)$
$E_{s,3z^2-r^2}$	$[n^2-\frac{1}{2}(l^2+m^2)](sd\sigma)$
$E_{x,xy}$	$\sqrt{3}l^2m(pd\sigma)+m(1-2l^2)(pd\pi)$
$E_{x,yz}$	$\sqrt{3}lmn(pd\sigma)-2lmn(pd\pi)$
$E_{x,zx}$	$\sqrt{3}l^2n(pd\sigma)+n(1-2l^2)(pd\pi)$
E_{x,x^2-y^2}	$\frac{1}{2}\sqrt{3}l(l^2-m^2)(pd\sigma)+l(1-l^2+m^2)(pd\pi)$
E_{y,x^2-y^2}	$\frac{1}{2}\sqrt{3}m(l^2-m^2)(pd\sigma)-m(1+l^2-m^2)(pd\pi)$
E_{z,x^2-y^2}	$\frac{1}{2}\sqrt{3}n(l^2-m^2)(pd\sigma)-n(l^2-m^2)(pd\pi)$
$E_{x,3z^2-r^2}$	$l[n^2-\frac{1}{2}(l^2+m^2)](pd\sigma)-\sqrt{3}ln^2(pd\pi)$
$E_{y,3z^2-r^2}$	$m[n^2-\frac{1}{2}(l^2+m^2)](pd\sigma)-\sqrt{3}mn^2(pd\pi)$
$E_{z,3z^2-r^2}$	$n[n^2-\frac{1}{2}(l^2+m^2)](pd\sigma)+\sqrt{3}n(l^2+m^2)(pd\pi)$
$E_{xy,xy}$	$3l^2m^2(dd\sigma)+(l^2+m^2-4l^2m^2)(dd\pi)+(n^2+l^2m^2)(dd\delta)$
$E_{xy,yz}$	$3lm^2n(dd\sigma)+ln(1-4m^2)(dd\pi)+ln(m^2-1)(dd\delta)$
$E_{xy,zx}$	$3l^2mn(dd\sigma)+mn(1-4l^2)(dd\pi)+mn(l^2-1)(dd\delta)$
E_{xy,x^2-y^2}	$\frac{3}{2}lm(l^2-m^2)(dd\sigma)+2lm(m^2-l^2)(dd\pi)+\frac{1}{2}lm(l^2-m^2)(dd\delta)$
E_{yz,x^2-y^2}	$\frac{3}{2}mn(l^2-m^2)(dd\sigma)-mn[1+2(l^2-m^2)](dd\pi)+mn[1+\frac{1}{2}(l^2-m^2)](dd\delta)$
E_{zx,x^2-y^2}	$\frac{3}{2}nl(l^2-m^2)(dd\sigma)+nl[1-2(l^2-m^2)](dd\pi)-nl[1-\frac{1}{2}(l^2-m^2)](dd\delta)$
$E_{xy,3z^2-r^2}$	$\sqrt{3}lm[n^2-\frac{1}{2}(l^2+m^2)](dd\sigma)-2\sqrt{3}lmn^2(dd\pi)+\frac{1}{2}\sqrt{3}lm(1+n^2)(dd\delta)$
$E_{yz,3z^2-r^2}$	$\sqrt{3}mn[n^2-\frac{1}{2}(l^2+m^2)](dd\sigma)+\sqrt{3}mn(l^2+m^2-n^2)(dd\pi)-\frac{1}{2}\sqrt{3}mn(l^2+m^2)(dd\delta)$
$E_{zx,3z^2-r^2}$	$\sqrt{3}ln[n^2-\frac{1}{2}(l^2+m^2)](dd\sigma)+\sqrt{3}ln(l^2+m^2-n^2)(dd\pi)-\frac{1}{2}\sqrt{3}ln(l^2+m^2)(dd\delta)$
$E_{x^2-y^2,x^2-y^2}$	$\frac{3}{4}(l^2-m^2)^2(dd\sigma)+[l^2+m^2-(l^2-m^2)^2](dd\pi)+[n^2+\frac{1}{4}(l^2-m^2)^2](dd\delta)$
$E_{x^2-y^2,3z^2-r^2}$	$\frac{1}{2}\sqrt{3}(l^2-m^2)[n^2-\frac{1}{2}(l^2+m^2)](dd\sigma)+\sqrt{3}n^2(m^2-l^2)(dd\pi)+\frac{1}{4}\sqrt{3}(1+n^2)(l^2-m^2)(dd\delta)$
$E_{3z^2-r^2,3z^2-r^2}$	$[n^2-\frac{1}{2}(l^2+m^2)]^2(dd\sigma)+3n^2(l^2+m^2)(dd\pi)+\frac{3}{4}(l^2+m^2)^2(dd\delta)$

Koster-Slater table

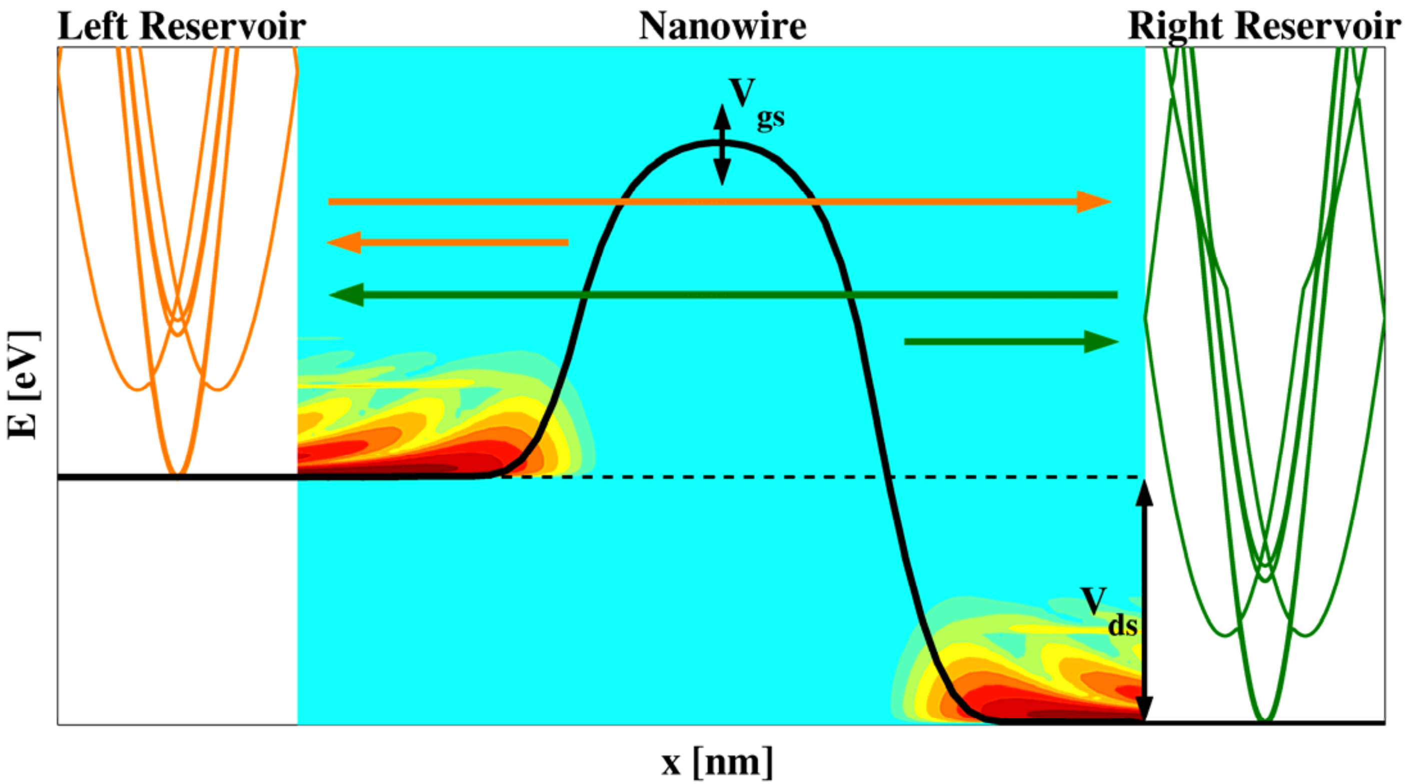
Phys. Rev. **94**, 1498 (1954)

How does band structure change with increasing cross section?



Transport

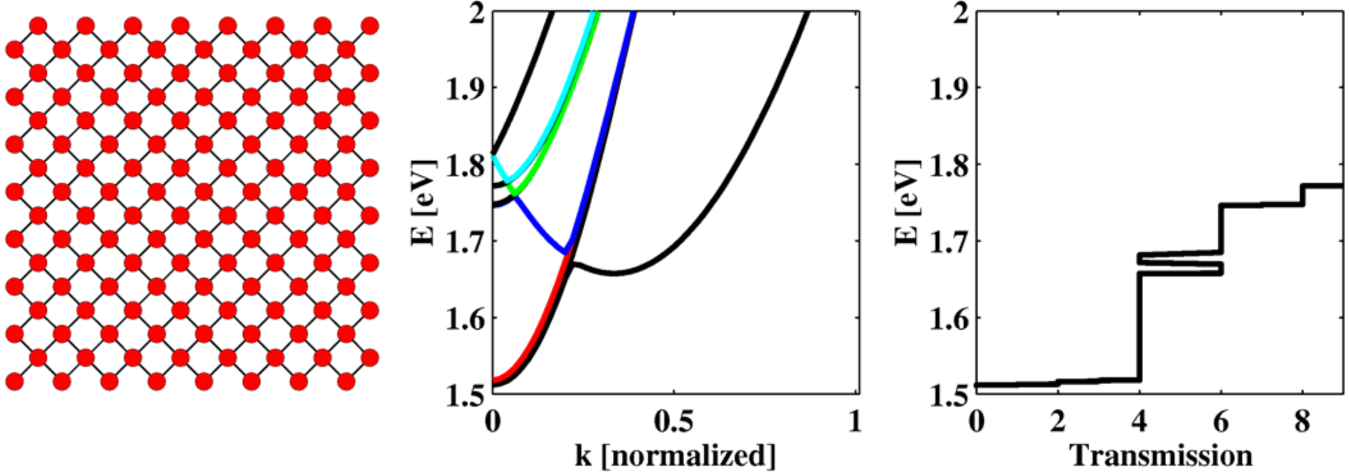
Cut along the transport direction *x* in the nanowire



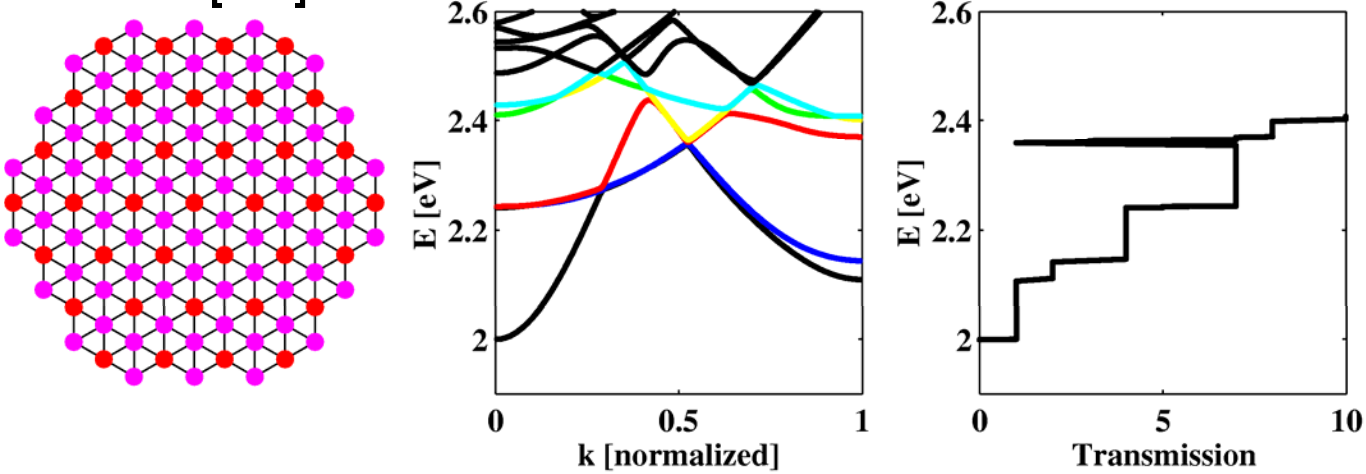
Wave functions are injected from the reservoirs and either **reflected** or **transmitted** to the other side. Band structure of reservoirs can be calculated because semi-infinite. At each energy all the k-states with **positive** (left) or **negative** (right) **velocity** are selected for injection.

Transport

Si [100]

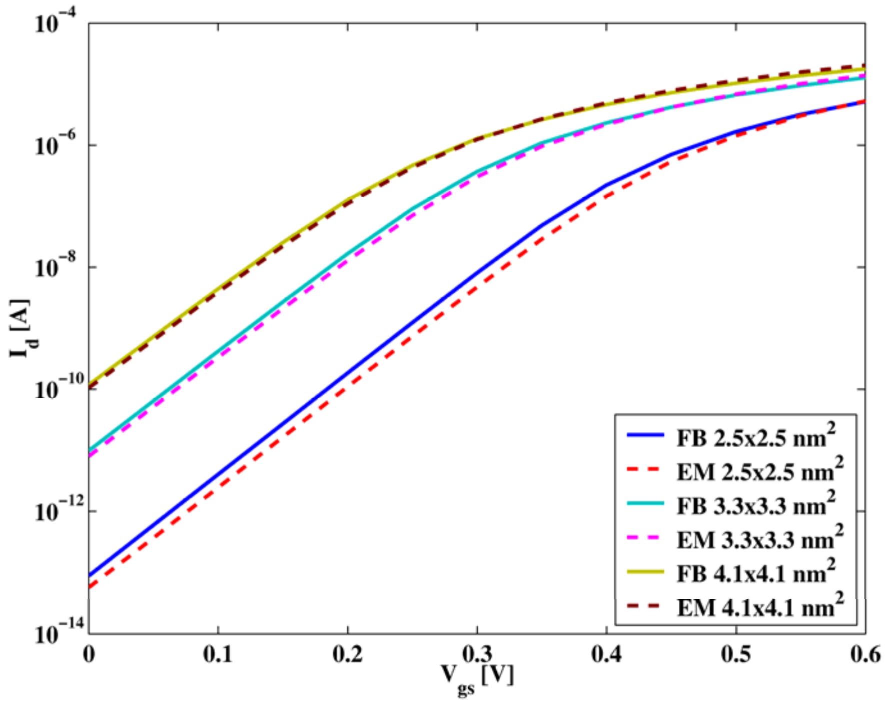


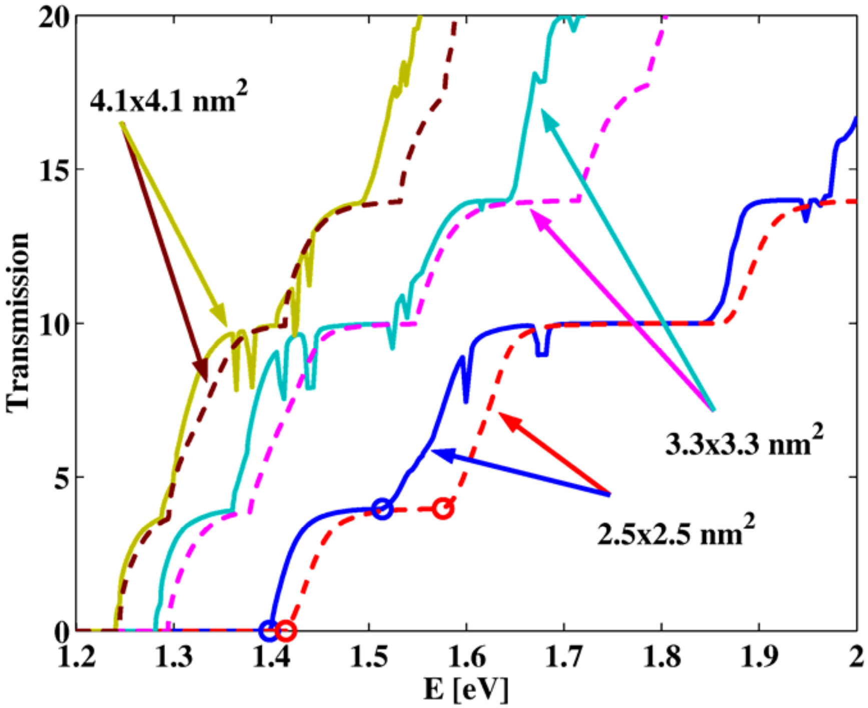
GaAs [111]



Transport

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

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

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

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

Injection matrix

Orbital-coefficient vector

Transport

Carrier and current density

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

$$\mathbf{J}(\mathbf{r}) = \sum_{i_1,i_2} i\frac{e}{2\hbar}\sum_{p,n_p}\frac{\Delta}{2\pi}\int \mathrm{d}E\left(H_{i_1i_2}C_{i_2,p,n_p}C^{*}_{i_1,p,n_p} - C_{i_1,p,n_p}C^{*}_{i_2,p,n_p}H_{i_2i_1}\right)\text{x}$$

$$\text{x}\left|\frac{\mathrm{d}E}{\mathrm{d}k_{p,n_p}}\right|^{-1} f(E-\mu_p)(\mathbf{R}_{i_2}-\mathbf{R}_{i_1})\delta(\mathbf{r}-\mathbf{R}_{i_1})$$

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

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

$$\mathbf{J}(\mathbf{r}) = \sum_{i_1}\sum_{i_2}\frac{e}{2}\left(H_{i_1i_2}G^{<}_{i_2i_1} - G^{<}_{i_1i_2}H_{i_2i_1}\right)(\mathbf{r}_{\mathbf{i}_2}-\mathbf{r}_{\mathbf{i}_1})\delta(\mathbf{r}-\mathbf{r}_{\mathbf{i}_1})$$

Transport

WF ↔ NEGF (if no incoherent scattering)

$$G^{<}_{ij}(E) = \sum_{p}\sum_{n} \underbrace{C_{i,p}(k_n) C^{T}_{j,p}(k_n)}_{(t_b \cdot N_A) \times (t_b \cdot N_A)} \left| \frac{dE(k_n)}{dk_n} \right|^{-1} f(E(k_n) - \mu_p)$$

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

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 $(N = 2\, t_b \cdot N_A)$

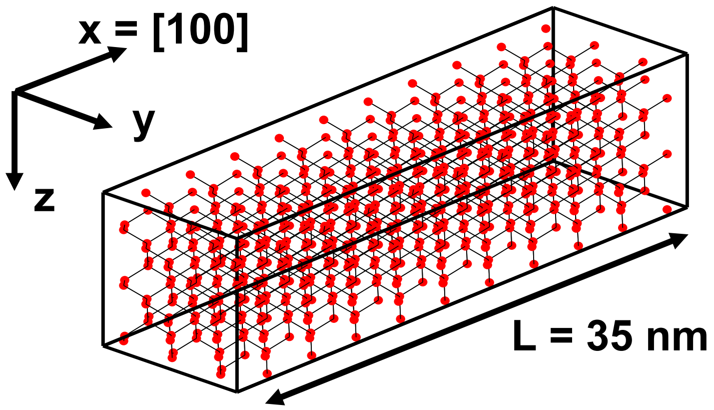
$A(E) \cdot C_n = \exp(ik_n(E)\Delta) \cdot B(E) \cdot C_n$

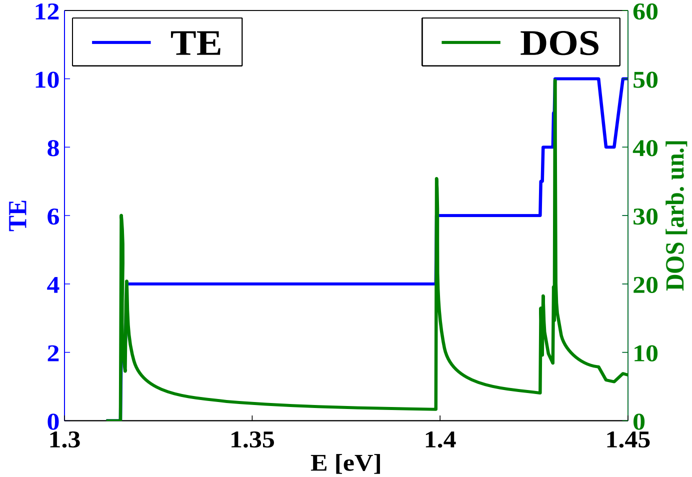
Shift-and-invert + ordinary eigenvalue problem $(N < t_b \cdot N_A)$ (PRB 74, 205323 (2006))

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

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	$\text{SuperLU}_{\text{dist}}$	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*** ↔ ***NEGF***

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

Electrostatics

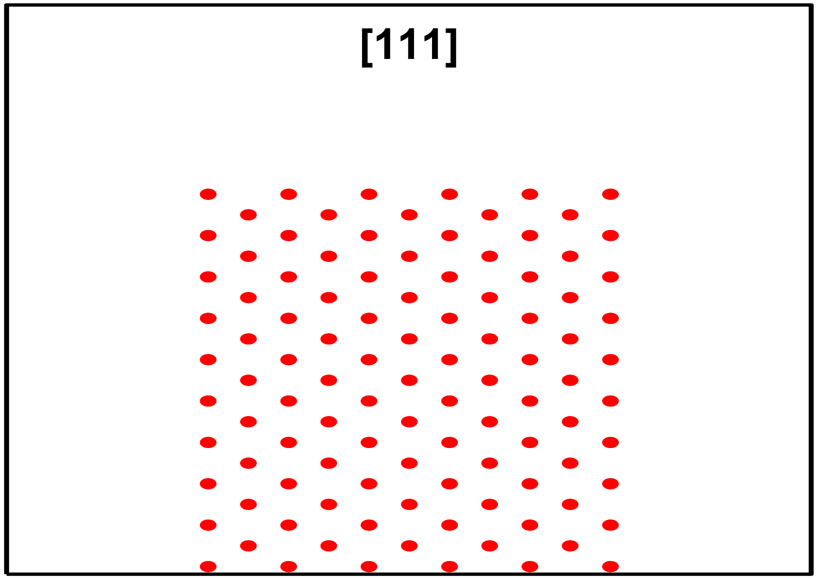
Grid Generation for Poisson equation

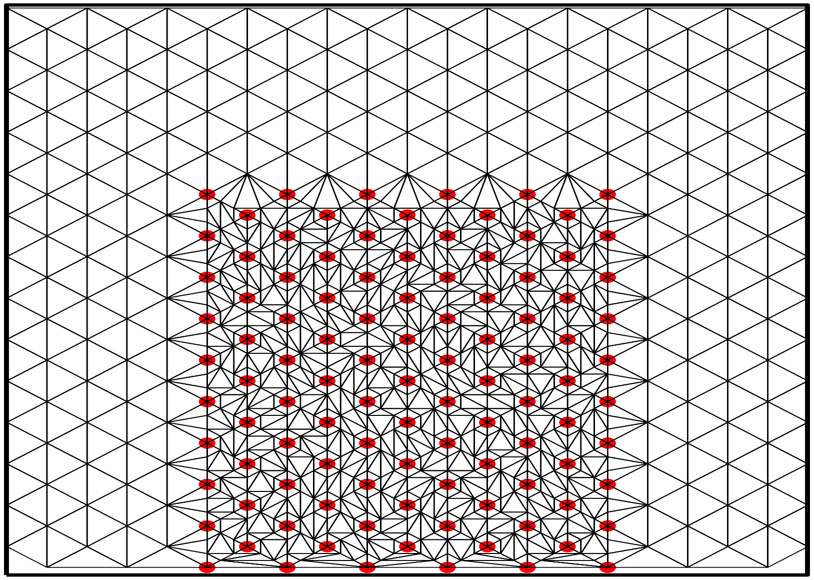
Grid must be general => Delaunay mesh: no data point (atoms) is contained in any triangle's circumcircle (2D) or in any tetrahedron's circumspheres (3D).

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





Electrostatics

Poisson Equation

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

Carrier Density

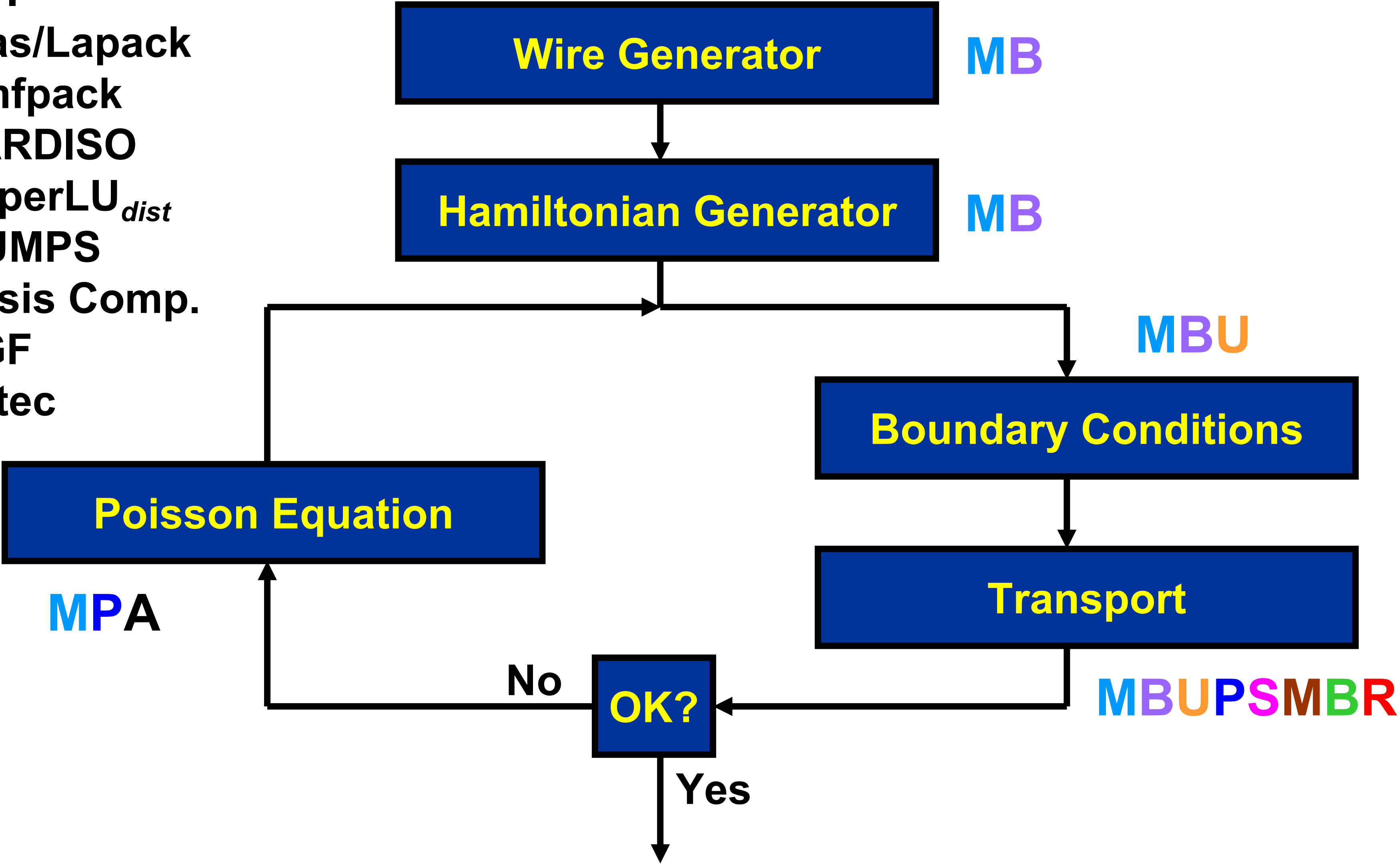
$$n(\mathbf{r}) = \sum_i n_i \delta(\mathbf{r} - \mathbf{r}_i)$$

Solution: Finite Element Method (FEM)

$$\int \mathrm{d}V\,\psi(\mathbf{r})\nabla\epsilon\nabla V(\mathbf{r}) \quad = \quad -q\int \mathrm{d}V\,\psi(\mathbf{r})\left(N_D^+(\mathbf{r}) - n(\mathbf{r})\right)$$

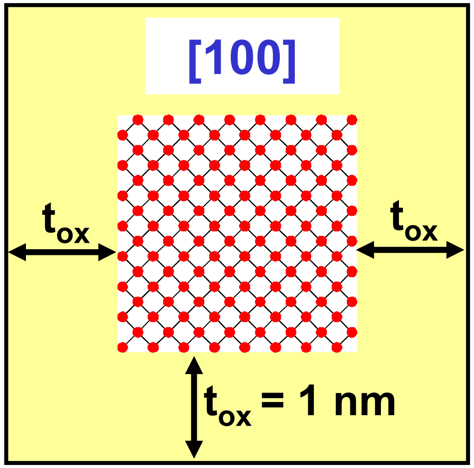
$$\int \mathrm{d}V\,\psi(\mathbf{r})n(\mathbf{r}) \quad = \quad \sum_i n_i\psi(\mathbf{r}_i)$$

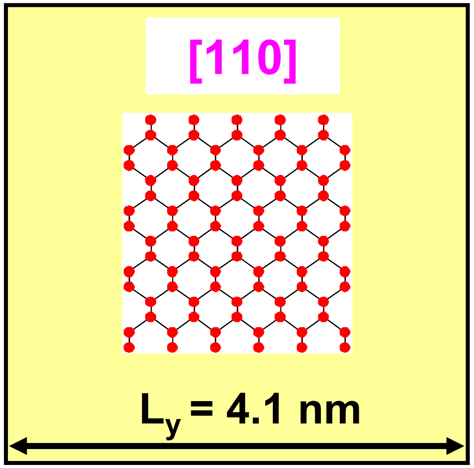
- **MPI**
- **Blas/Lapack**
- **Umfpack**
- **PARDISO**
- **SuperLU$_{dist}$**
- **MUMPS**
- **Basis Comp.**
- **RGF**
- **Aztec**

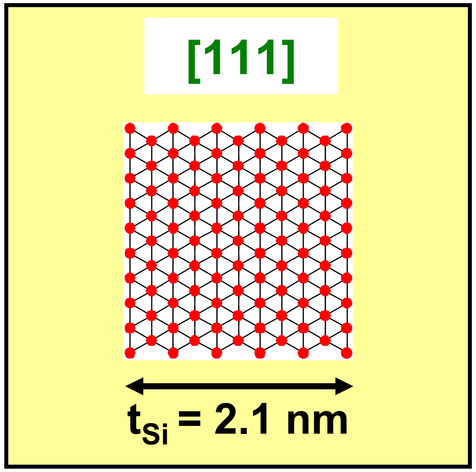


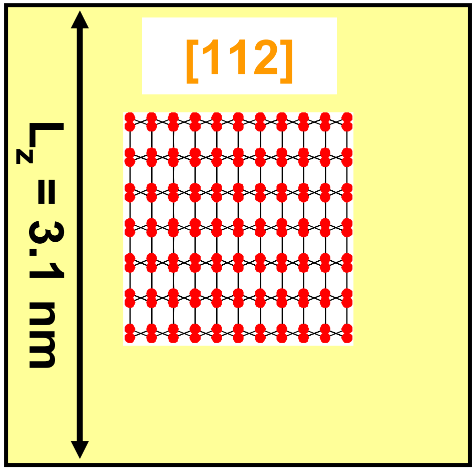
On-current for different channel orientations

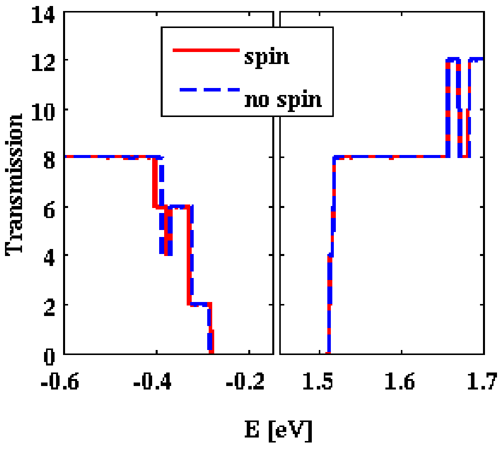
How does transmission change with channel orientation?

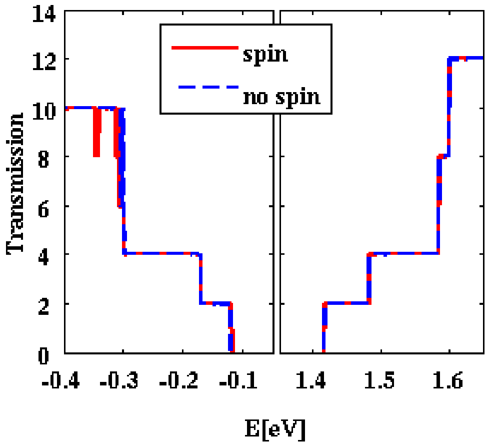


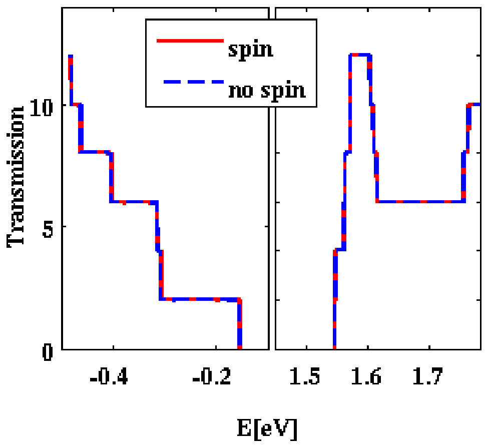


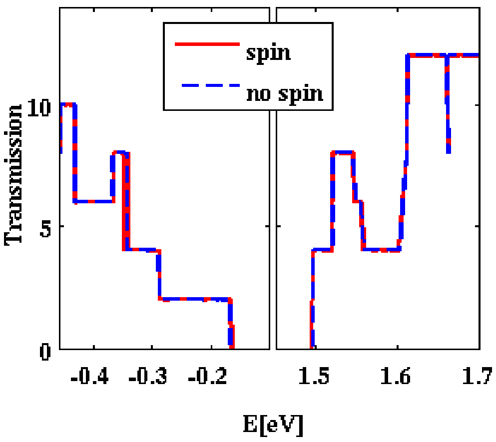






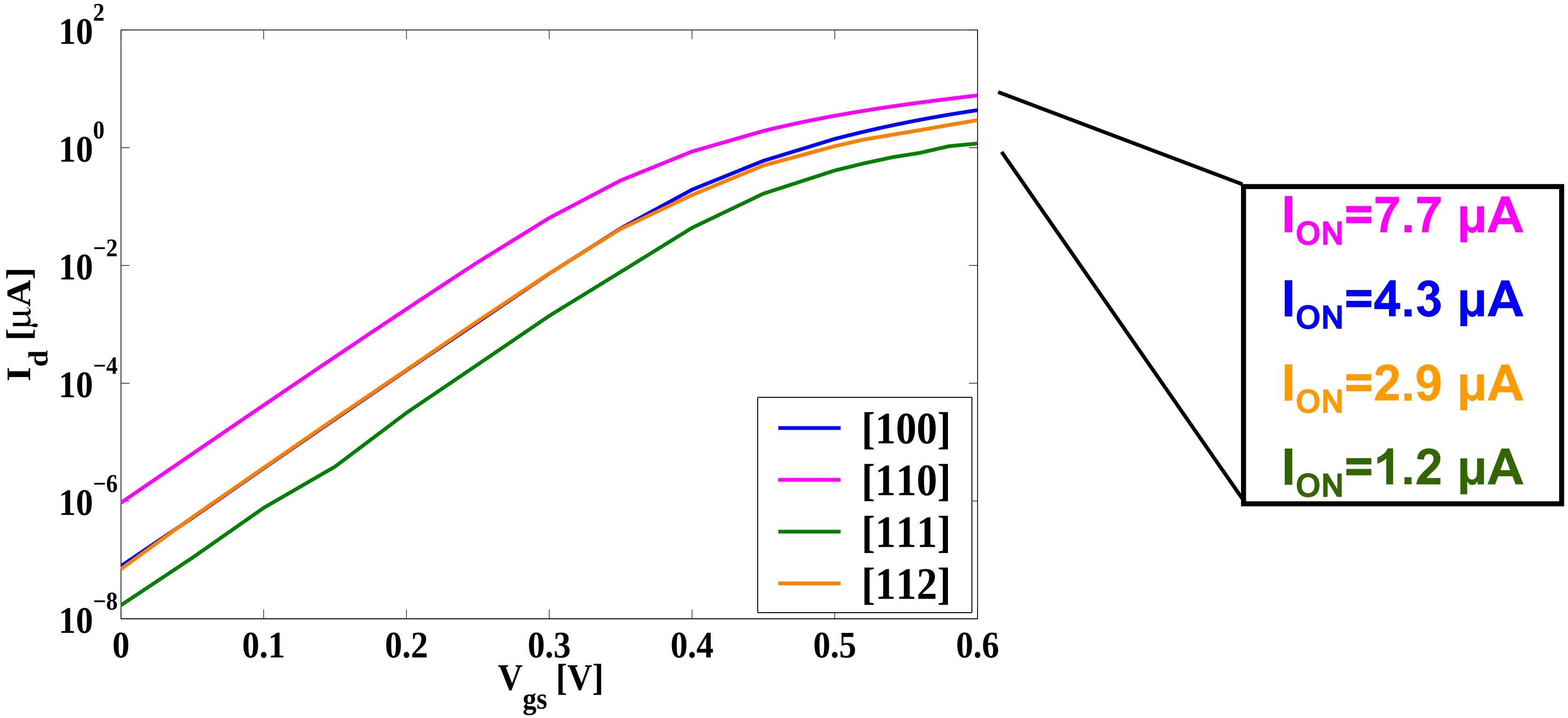






How does current depend on channel orientation?

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 μA

I_{ON}=4.3 μA

I_{ON}=2.9 μA

I_{ON}=1.2 μA

On-current for different channel orientations

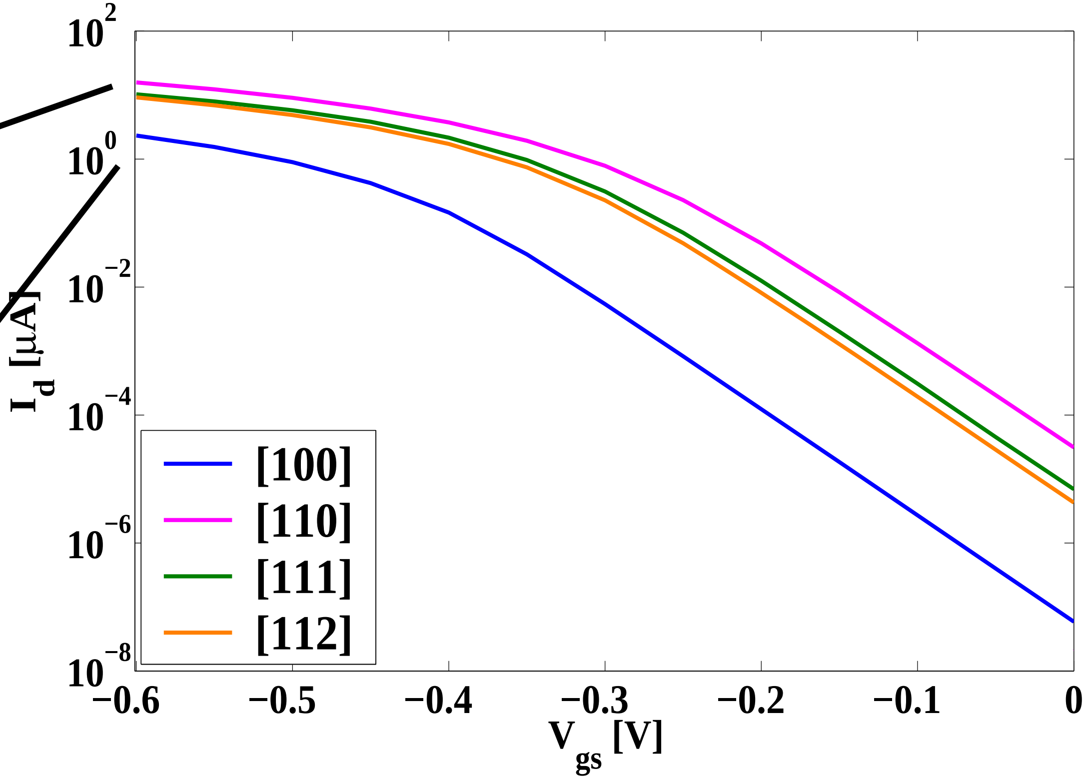
Full-band (FB) transfer characteristics: I_d-V_{gs} at V_{ds}=-0.4 V
p-FET with [100], [110], [111], and [112], L_g=13 nm

I_{ON}=15.7 μA

I_{ON}=10.3 μA

I_{ON}=9.2 μA

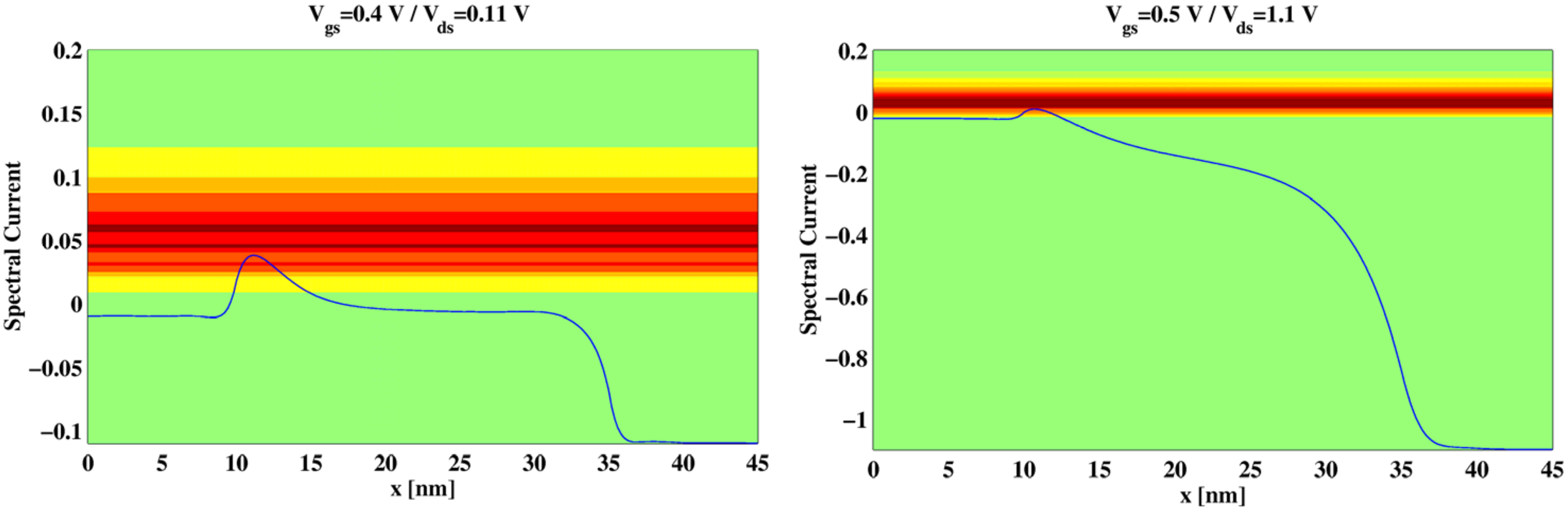
I_{ON}=2.3 μA



On-current for different channel orientations

Quantum-ballistic spectral currents

Can we trust the high quantum-ballistic ON-currents?



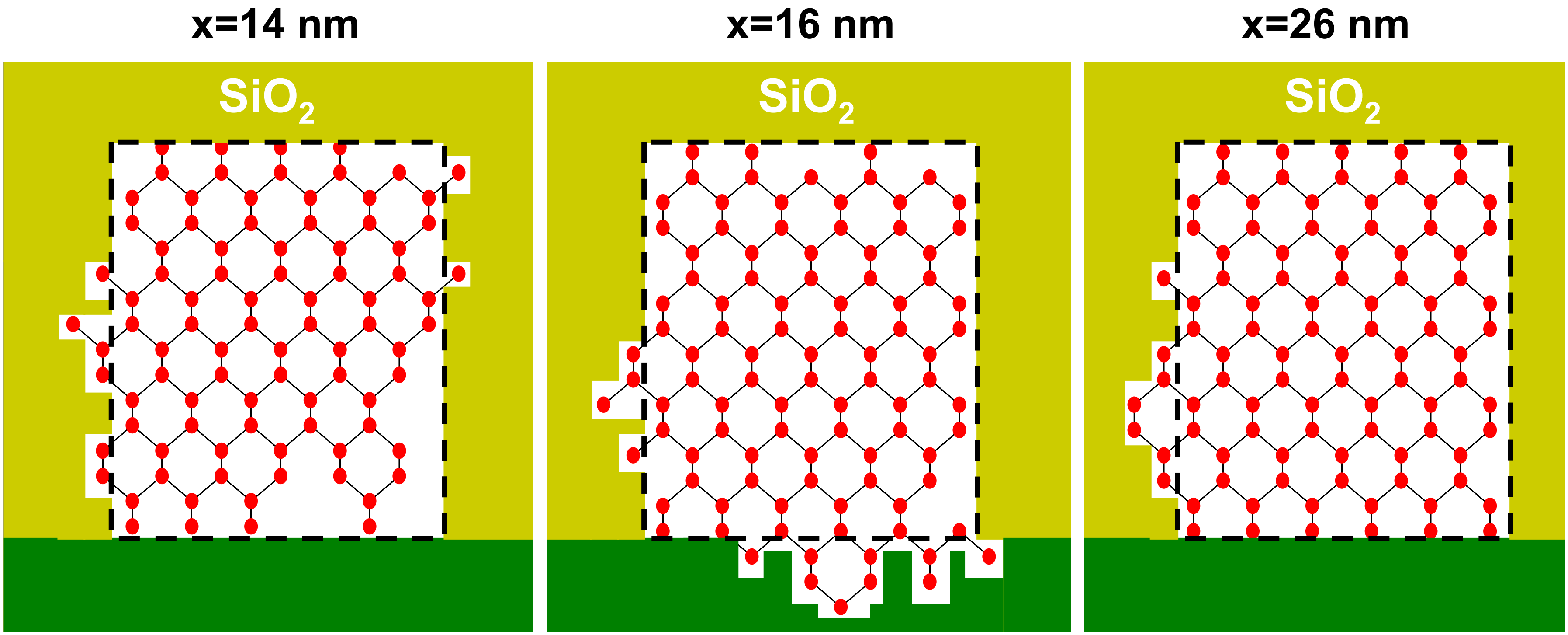
Spectral currents at V_{gs} = 0.4 V, V_{ds} = 0.11 V (left) and V_{gs} = 0.5 V, V_{ds} = 1.1 V (right).

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).

How does interface roughness influence the current?

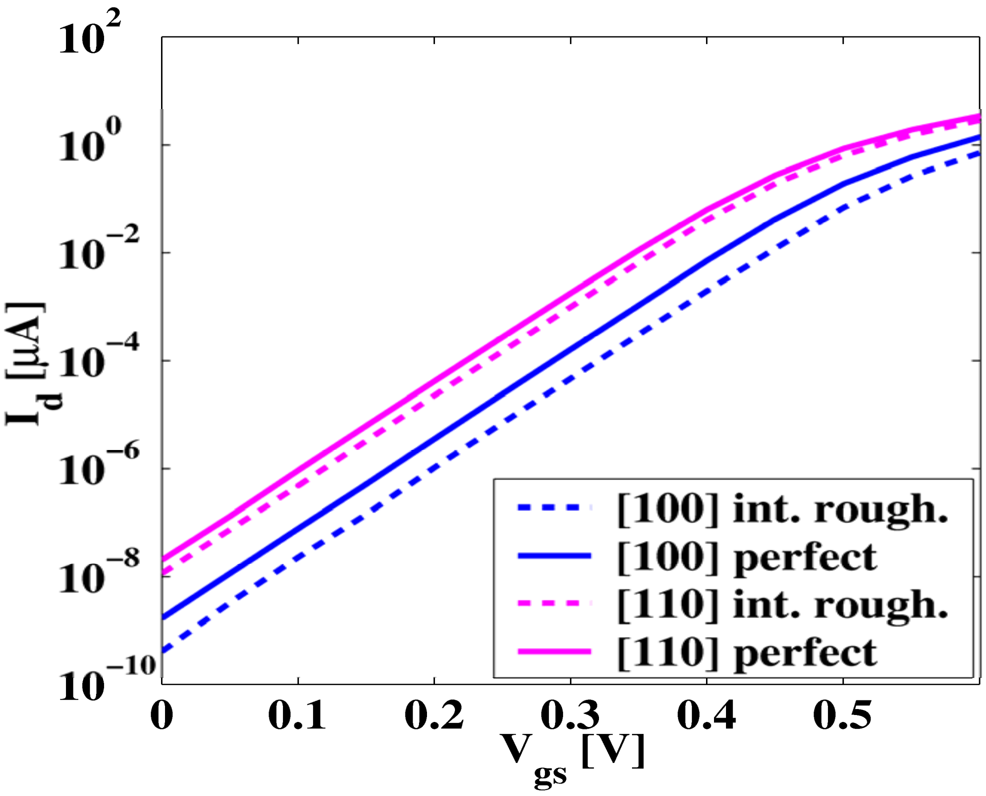
Process variations => cross section variations

Example: [110] nanowire



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

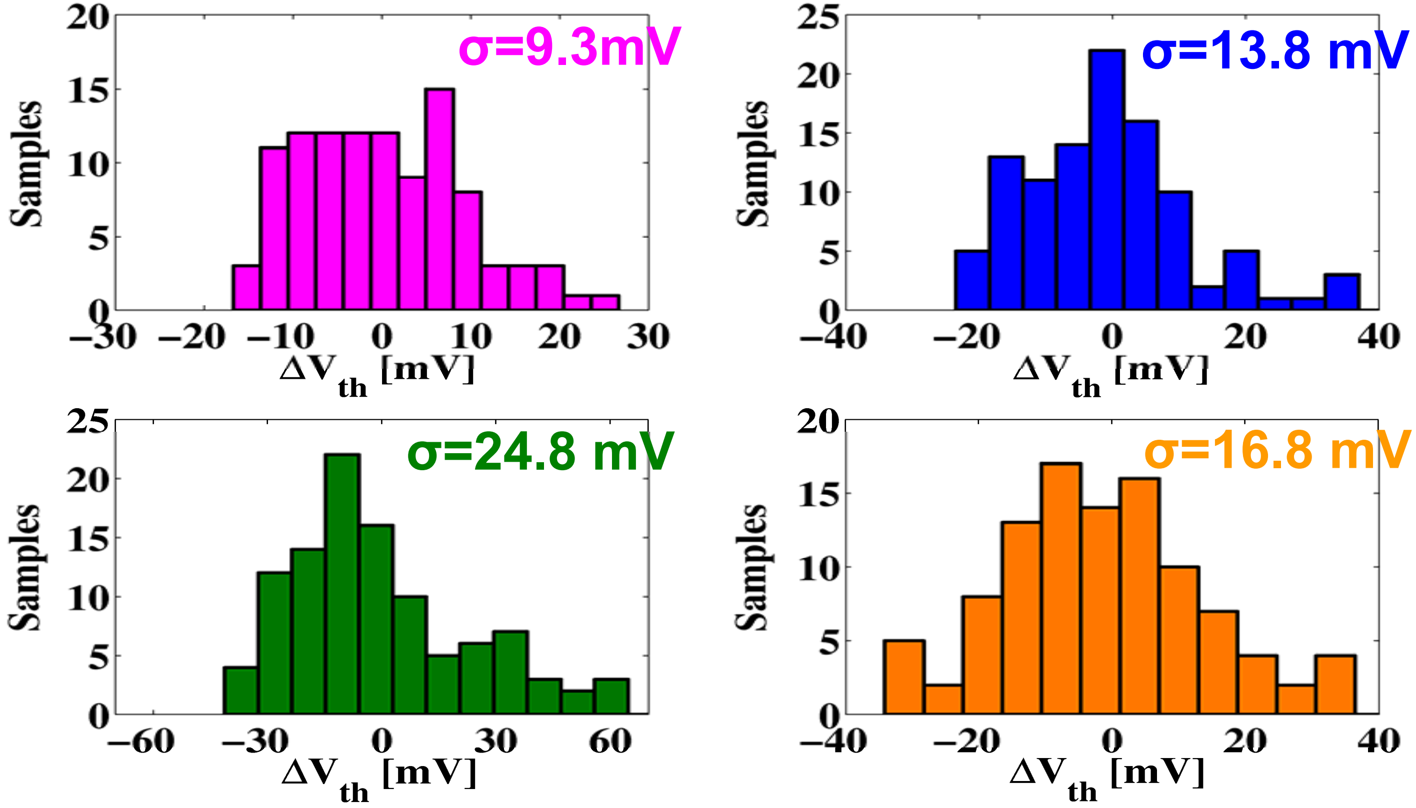
FB I_d-V_{gs} at V_{ds}=0.4 V for one possible interface realization



1) **Sub-threshold swing remains constant: S ≈ 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

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



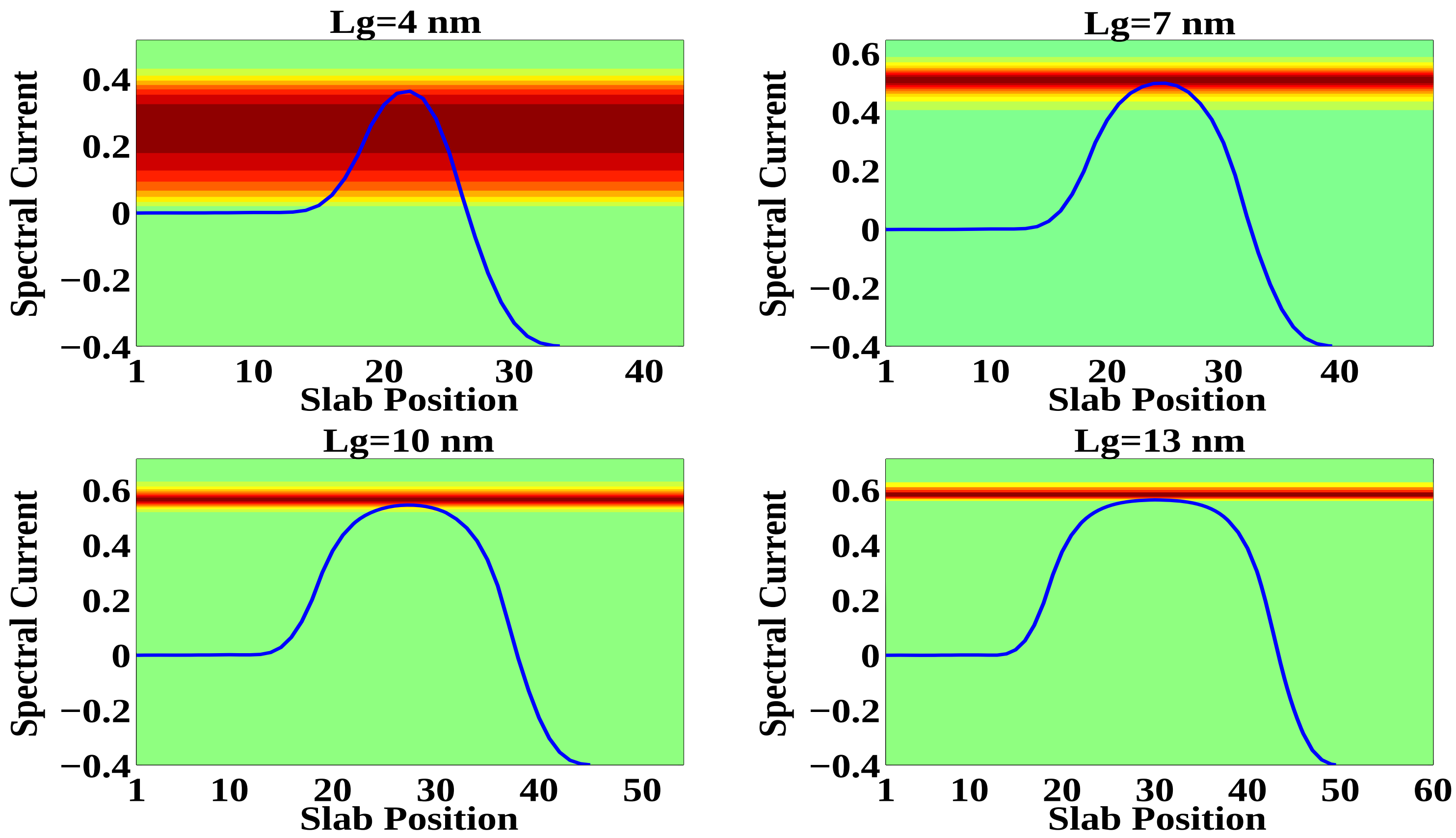
Performance summary: ON-current, OFF-current, roughness

	n-I_{on} [μA]	p-I_{on} [μA]	σ [mV]
[110]	7.7	15.7	9.3
[100]	4.3	2.3	13.8
[112]	2.9	9.2	16.8
[111]	1.2	10.3	24.8

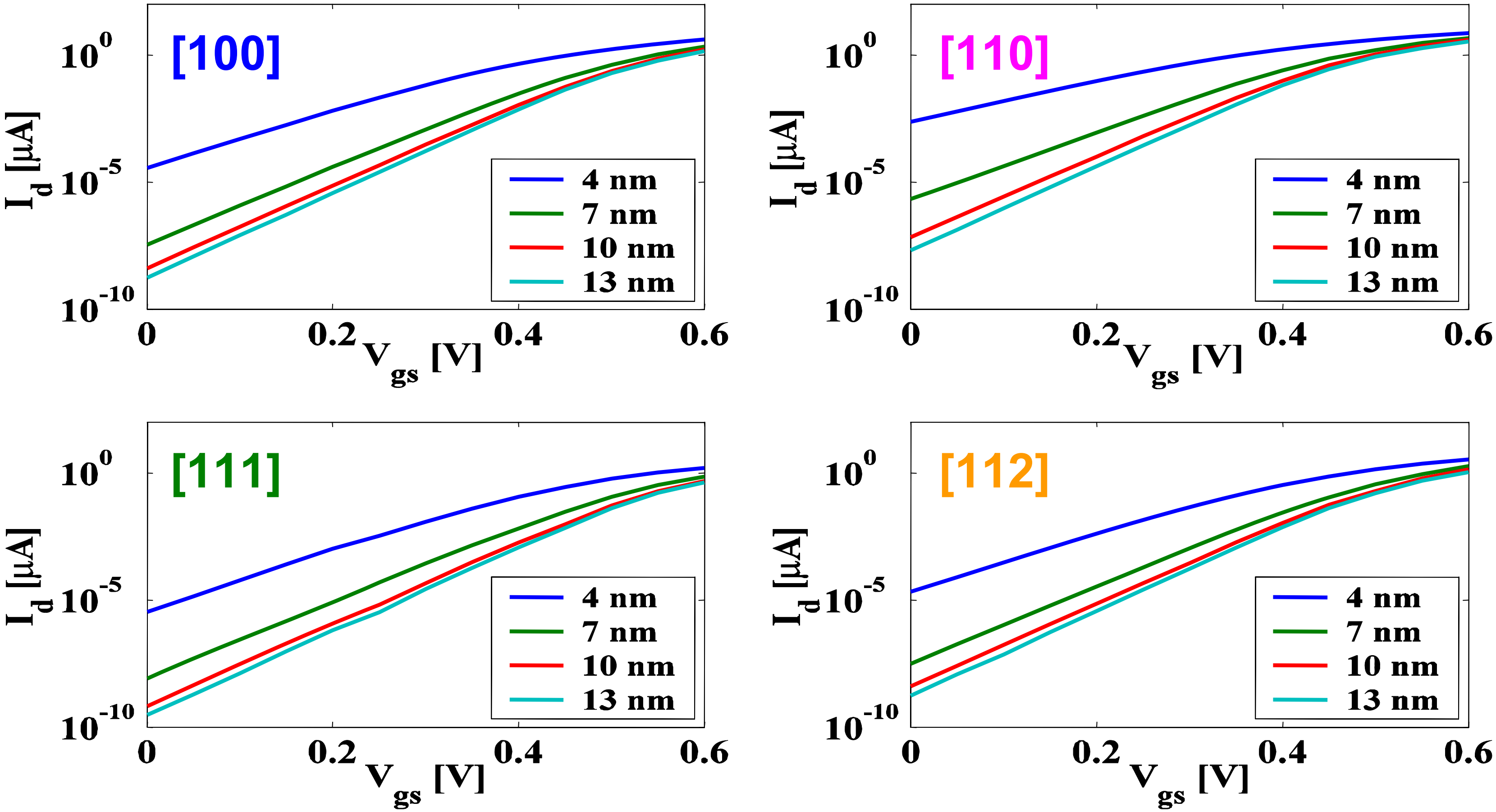
[110] has the highest on-current, is the least sensitive to interface roughness

Simulation of tunneling-induced OFF-currents

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

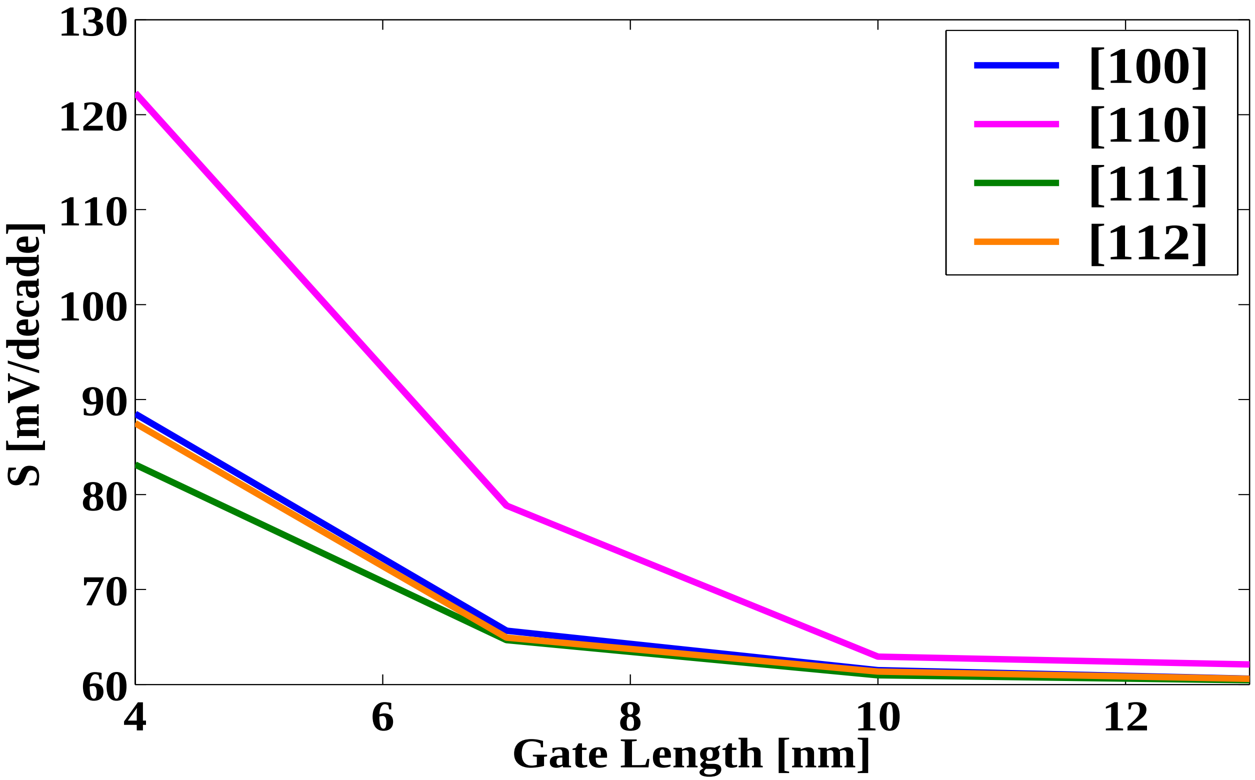


I_d-V_{gs} @ V_{ds}=0.4 V for 4 different channel lengths (4 nm, 7 nm, 10 nm, and 13 nm) and for [100], [110], [111], [112]



Sub-threshold swing *S* as function of gate length L_g

Low effective mass => high on-current / strong tunneling



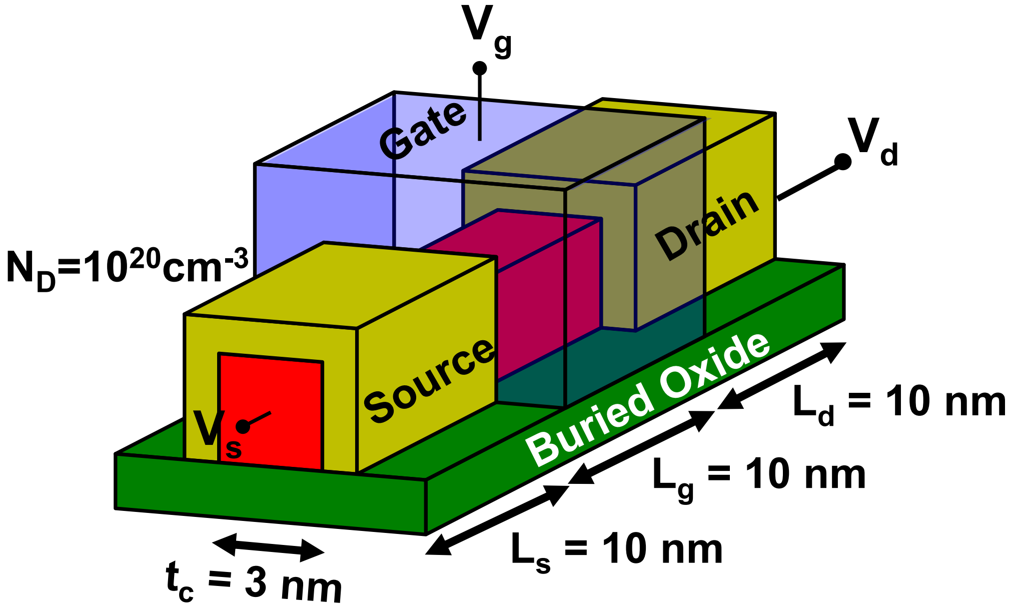
Performance summary: ON-current, OFF-current, roughness, S-D leakage

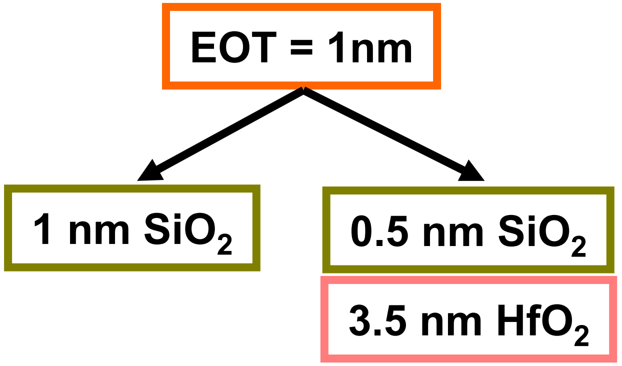
	$n\text{-}I_{on}$ [μA]	$p\text{-}I_{on}$ [μA]	σ [mV]	S @ L_g=4 nm
[110]	7.7	15.7	9.3	122.3
[100]	4.3	2.3	13.8	88.5
[112]	2.9	9.2	16.8	87.5
[111]	1.2	10.3	24.8	83.2

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

gate tunneling leakage

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

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

eff. mass approx. + finite difference

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

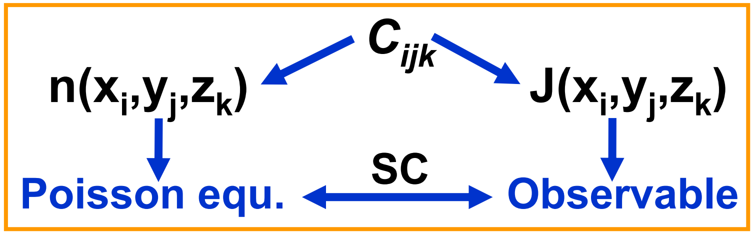
3D sparse linear problem Ax=b

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

BCs and injection mechanism

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

3D carrier and current densities



C_{ijk}

$n(x_i,y_j,z_k)$

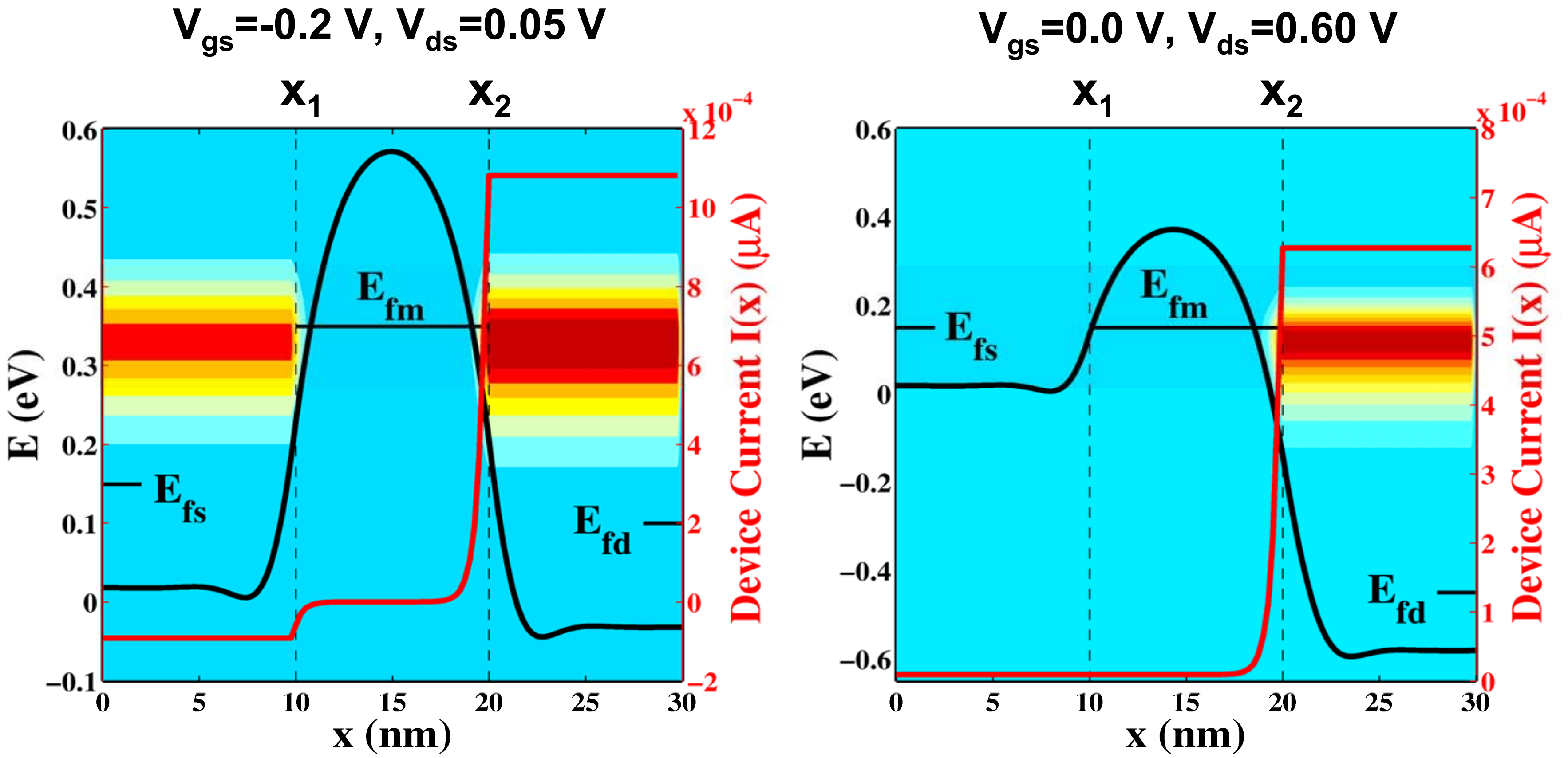
$J(x_i,y_j,z_k)$

Poisson equ.

SC

Observable

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



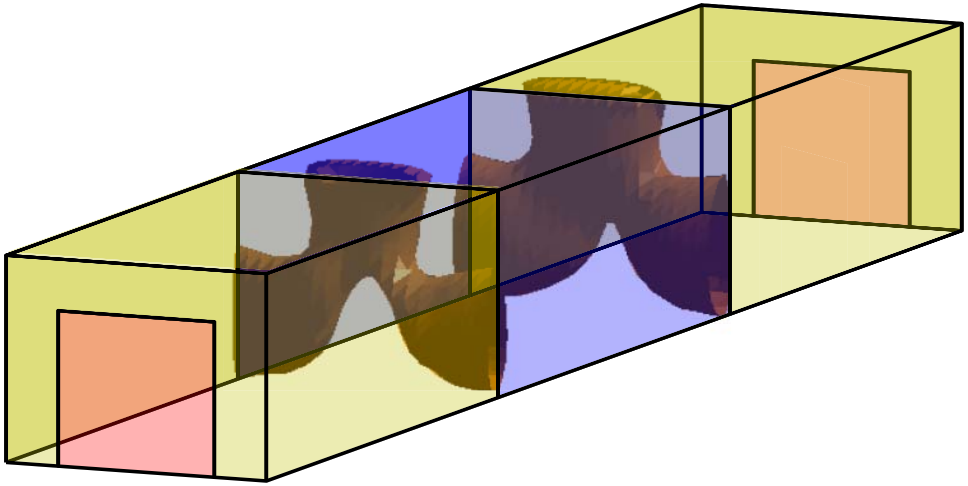
current conservation: $I(x_2)-I(x_1) = I_G$

gate tunneling leakage

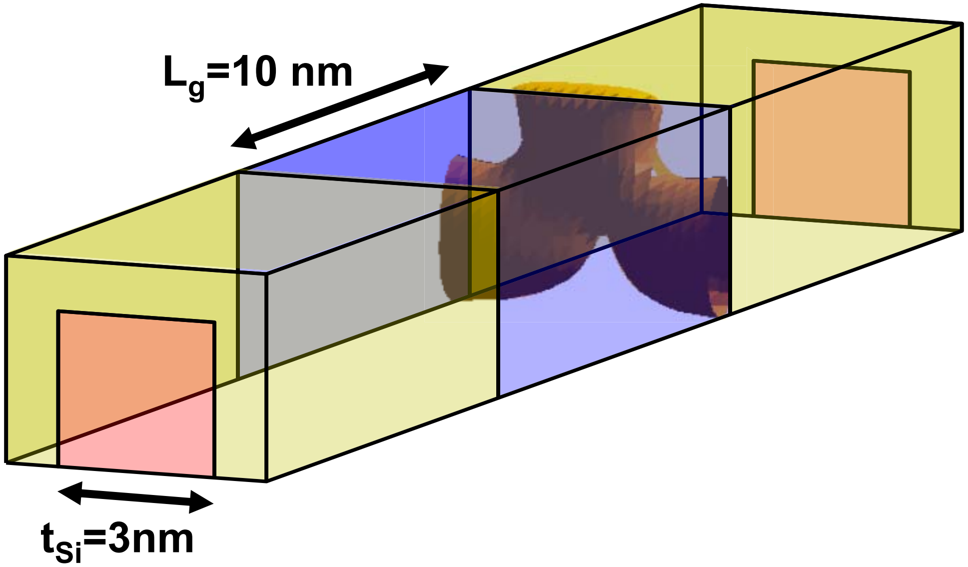
Isosurfaces of the gate current for a triple-gate structure
SiO_2 dielectric layer + TiN metal contact

Current escapes at the gate corners

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

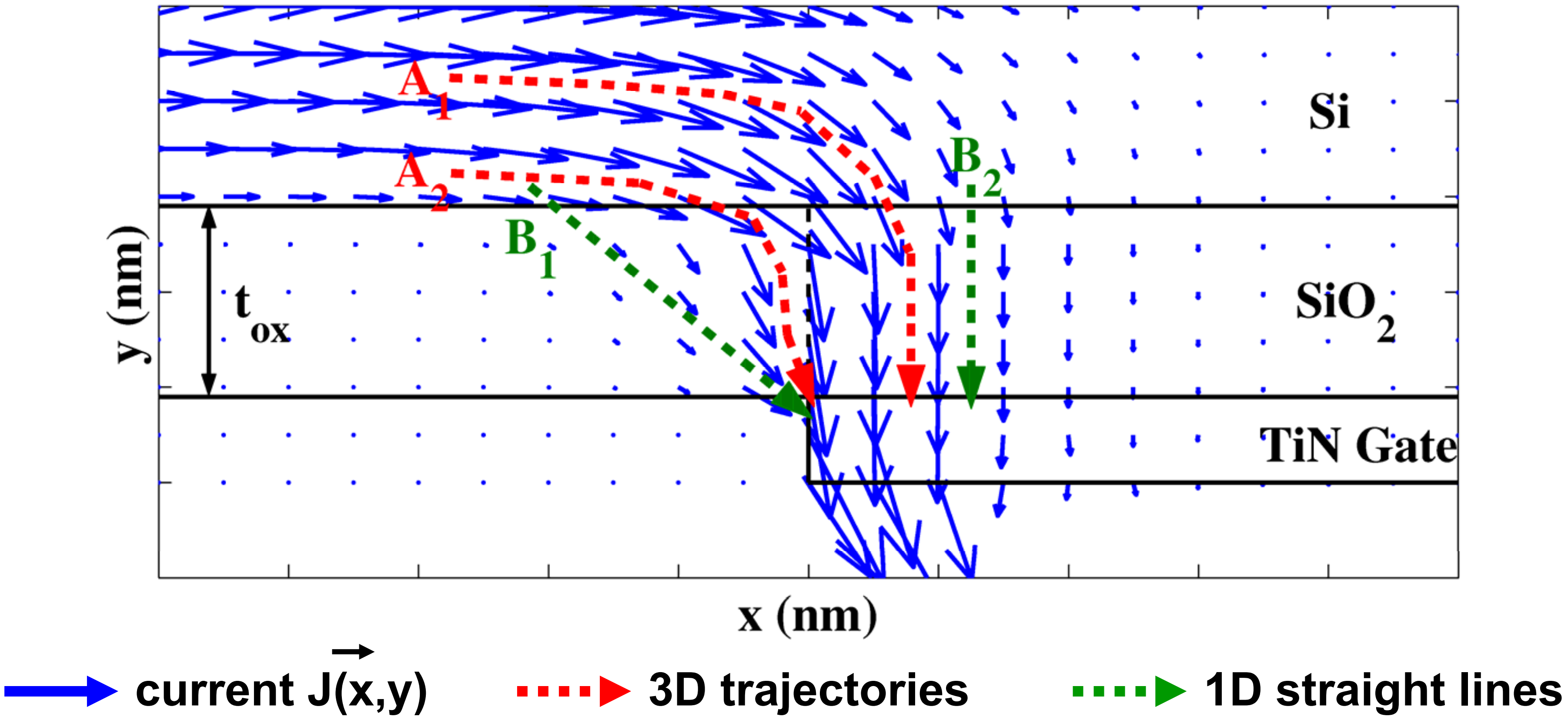


V_{gs}=0.0 V, V_{ds}=0.60 V



Current trajectories around gate corner

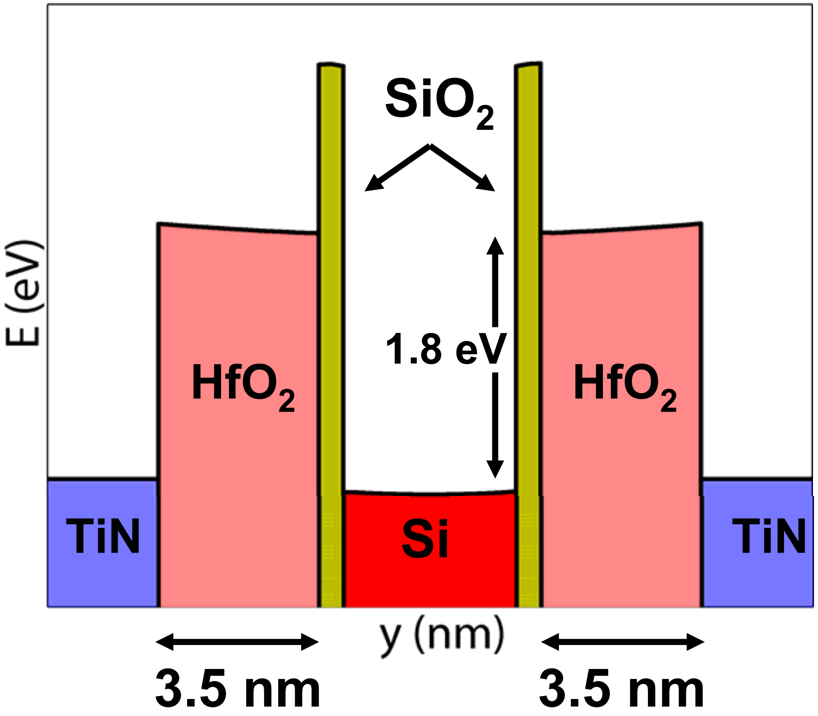
1D approximation vs full 3D (projected)



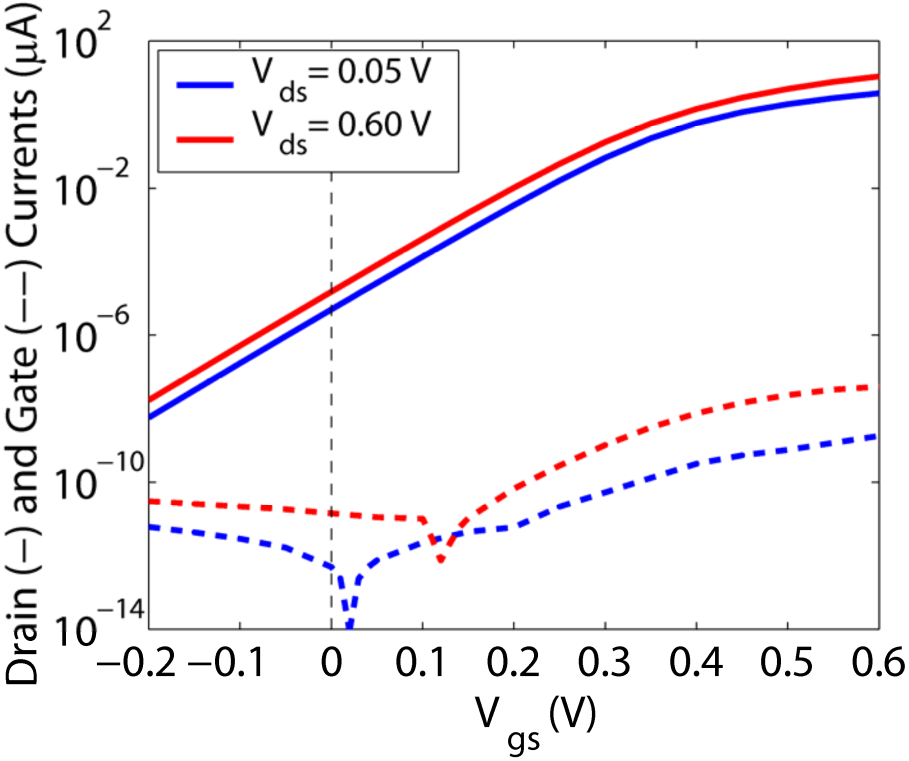
gate tunneling leakage

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

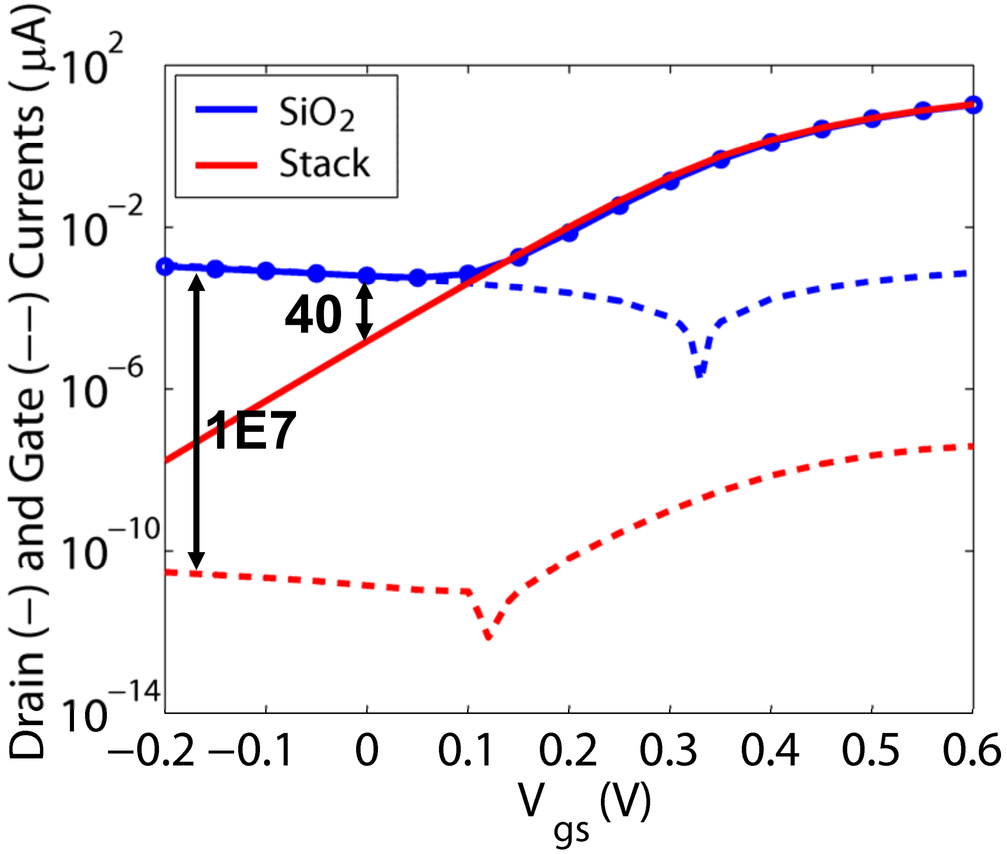


Transfer Characteristics



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



band-to-band tunneling (GIDL)

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.

$$R_{\mathrm{net}}^{\mathrm{bb}} = AF^{7/2}\frac{\tilde{n}\tilde{p}-n_{\mathrm{i,\,eff}}^{2}}{(\tilde{n}+n_{\mathrm{i,eff}})(\tilde{p}+n_{\mathrm{i,eff}})}\left[\frac{(F_{\mathrm{C}}^{\mp})^{-3/2}\exp\left(-\frac{F_{\mathrm{C}}^{\mp}}{F}\right)}{\exp\left(\frac{\hbar\omega}{kT}\right)-1}+\frac{(F_{\mathrm{C}}^{\pm})^{-3/2}\exp\left(-\frac{F_{\mathrm{C}}^{\pm}}{F}\right)}{1-\exp\left(-\frac{\hbar\omega}{kT}\right)}\right]$$

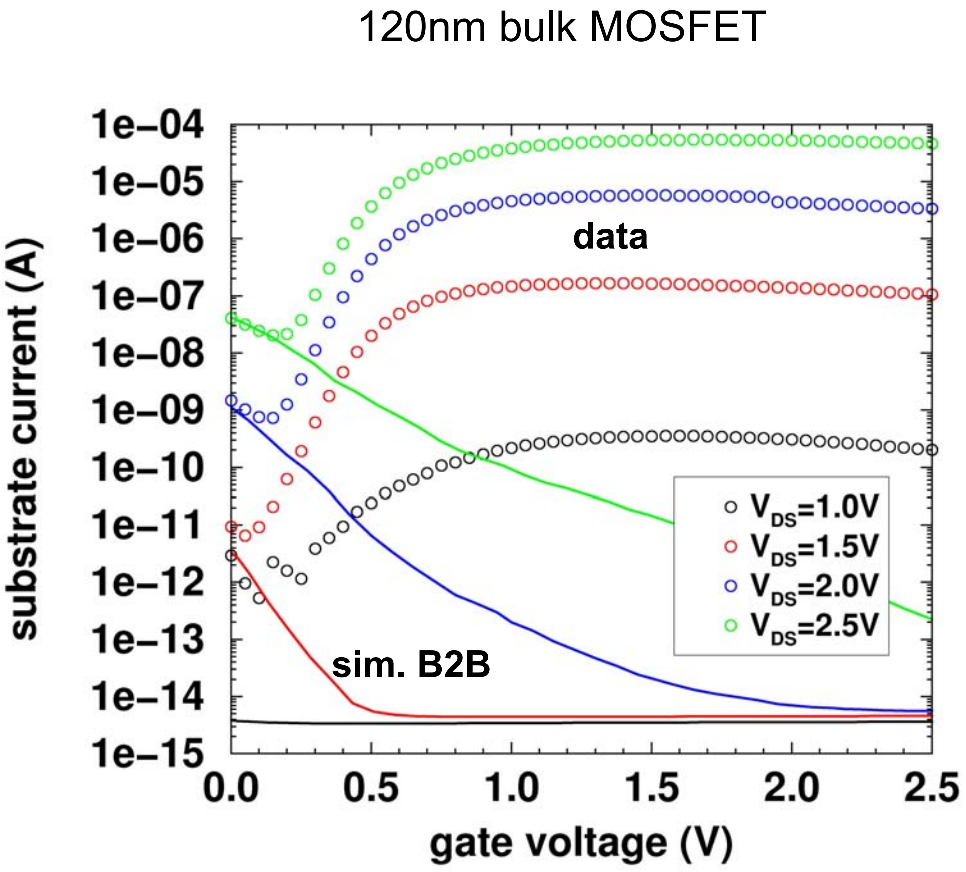
Non-local nature of the B2B rate can be modeled in a simple way:

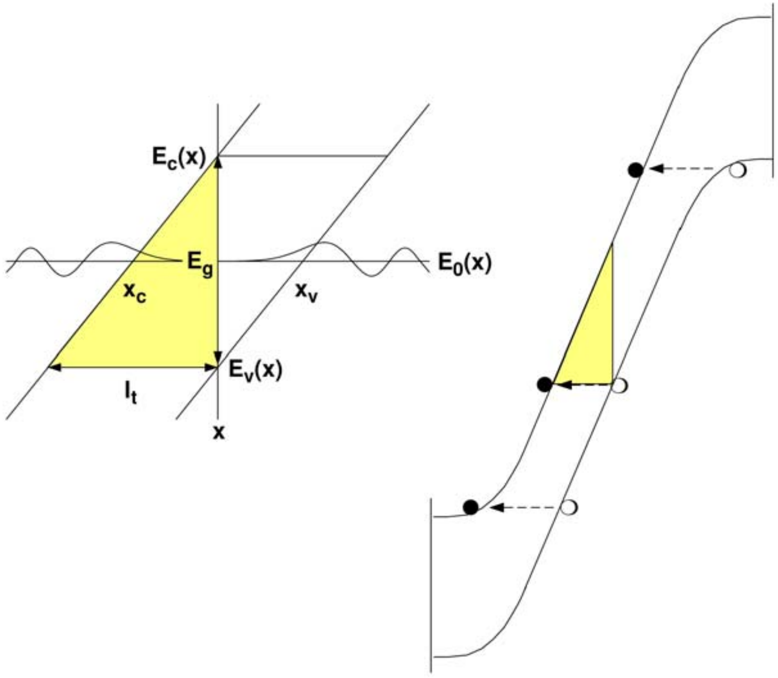
$$\tilde{n} = n\left(\frac{n_{\mathrm{i,eff}}}{N_{\mathrm{C}}}\right)^{\frac{|\nabla E_{\mathrm{F},n}|}{F}}$$

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

$$F_{\mathrm{C}}^{\pm} = B(E_{\mathrm{g,eff}}\pm\hbar\omega)^{3/2}$$

band-to-band tunneling (GIDL)

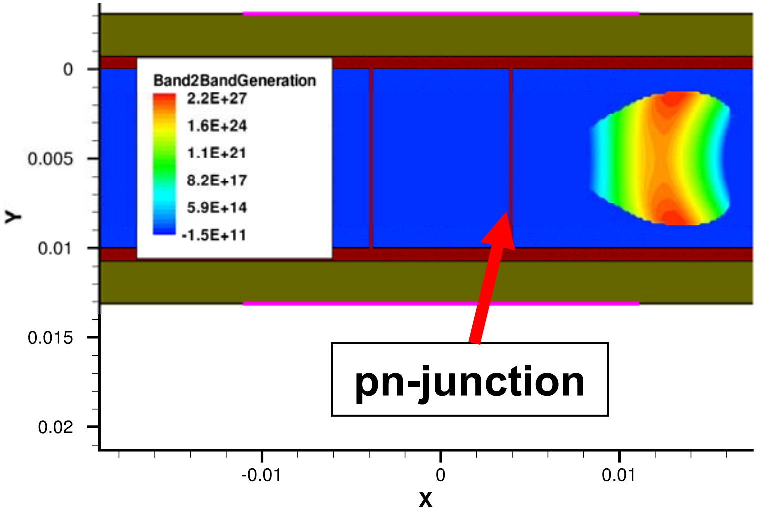




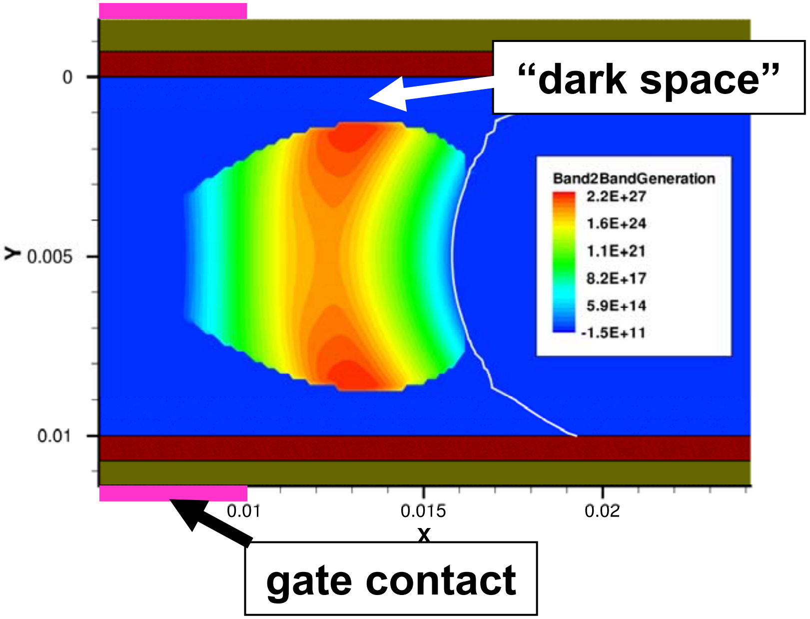
- **No direct experimental verification of the B2B rate in Si exists**
- **Calibrations based on GIDL data rely on correct modeling of lateral dopant diffusion**

band-to-band tunneling (GIDL)

22nm UTB DG SOI (*PULLNANO* template)



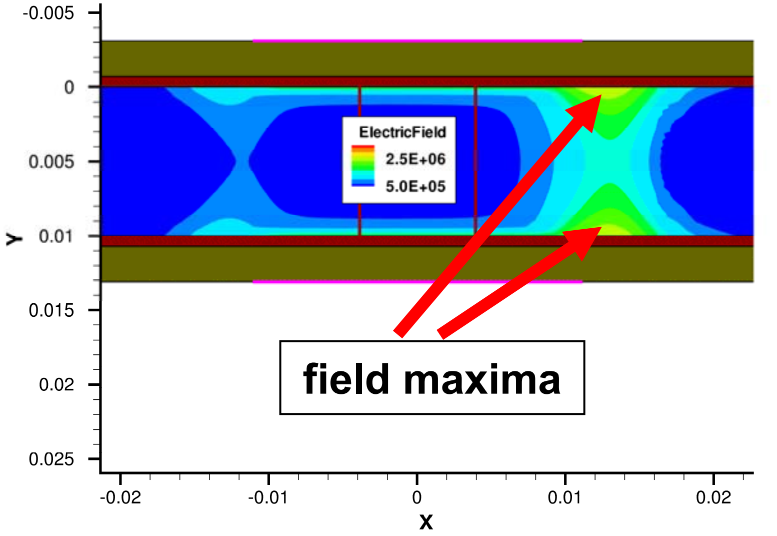
pn-junction



“dark space”

gate contact

- **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.**



field maxima

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

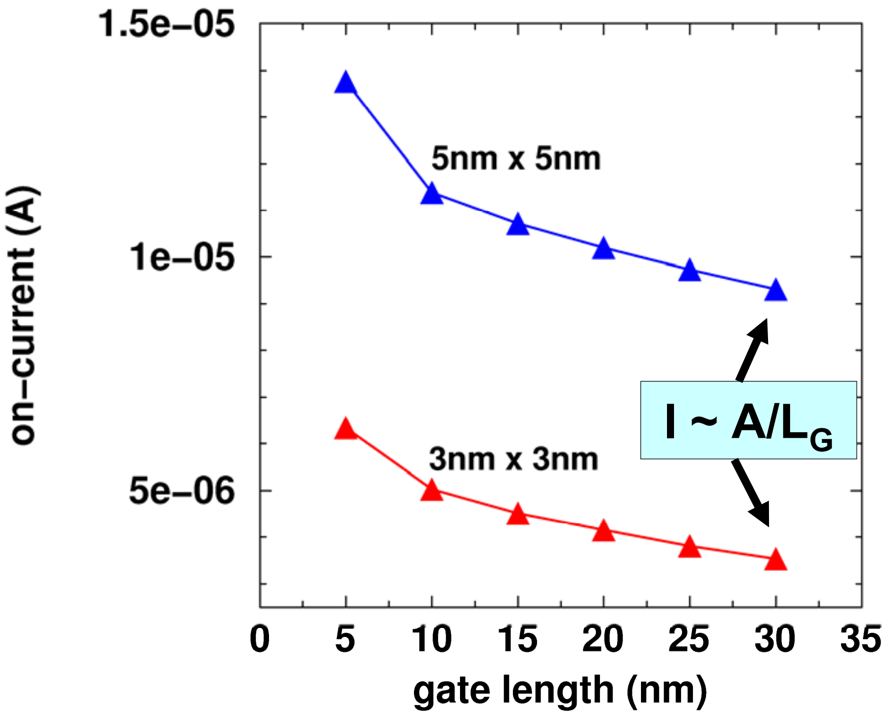
$$(E - H - \Sigma^R)G^R = \mathbb{1}$$

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

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

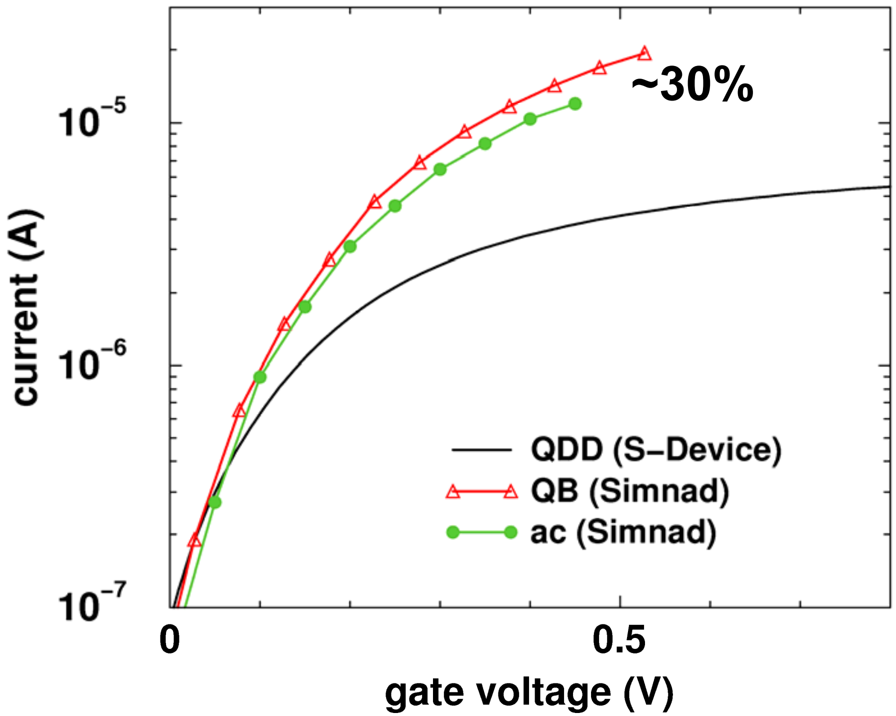
Incoherent scattering

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



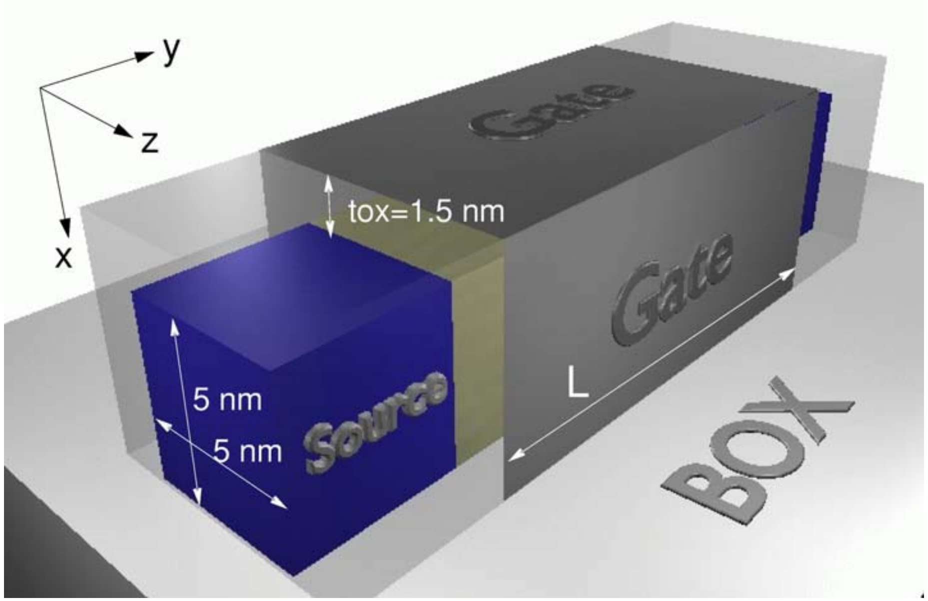
I ~ A/L_G

TG 5 x 5 x 25 nm^3 NW FET

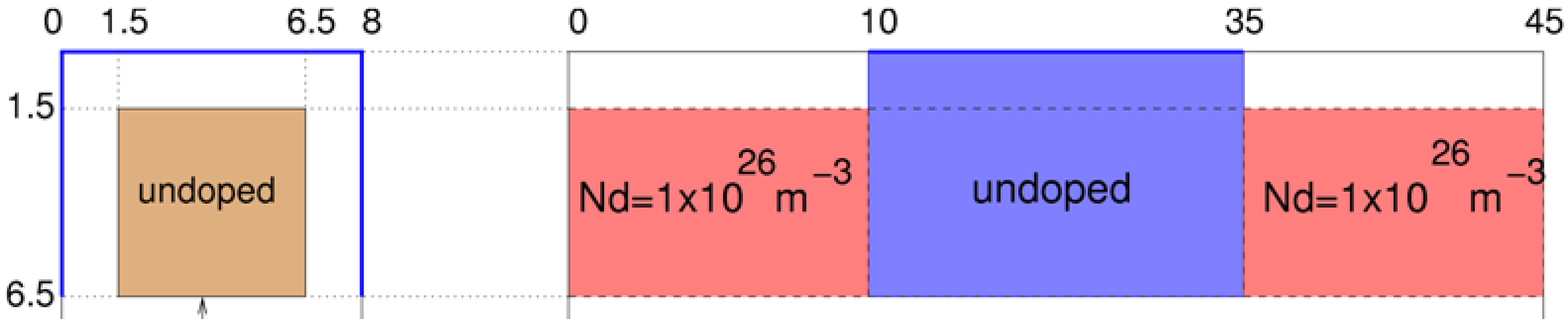


~30%

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

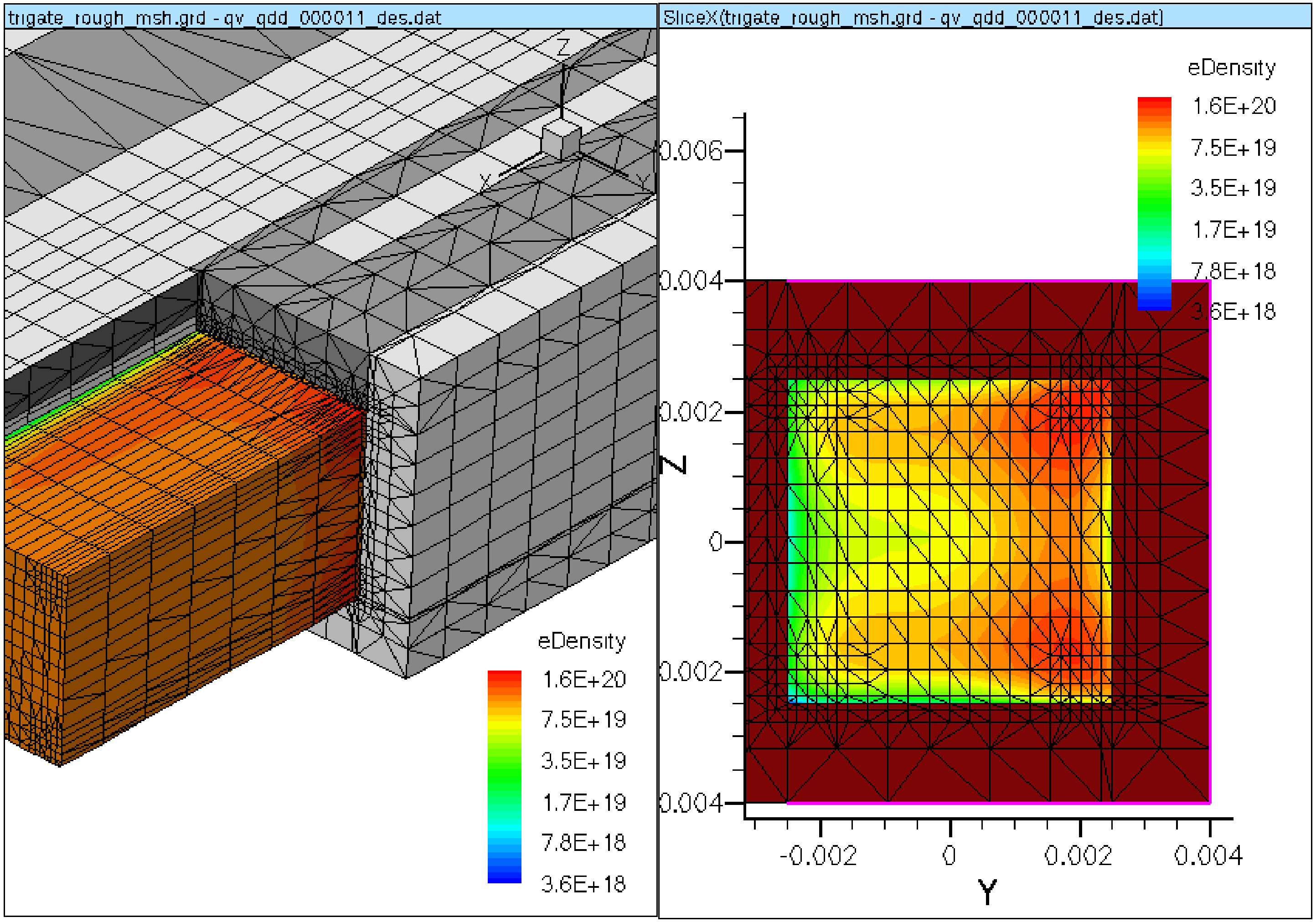


TGNW-FET (courtesy EU SINANO project)
Gate length: 25 nm (65 nm technology node).
Channel Cross-Section : Square (5x5 nm^2).
Source/drain extensions: 10 nm.
Oxide parameters: material is SiO_2 (k~3.9).
Field Oxide Thickness : 1.5 nm.
Buried Oxide Thickness : 150 nm.
Gate electrode work function: 4.1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



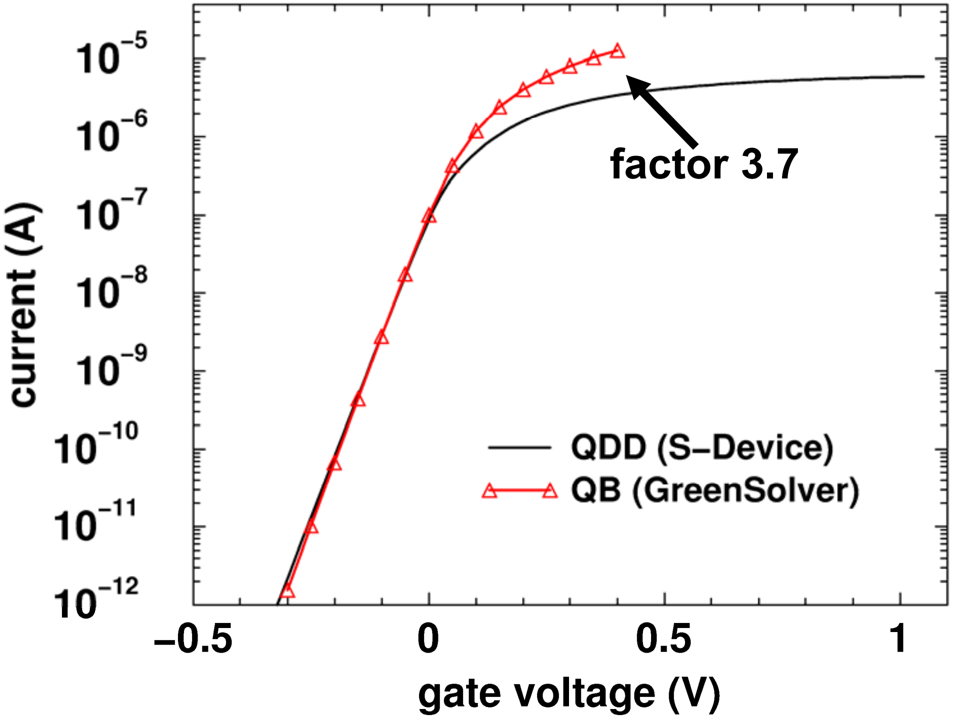
Incoherent scattering

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



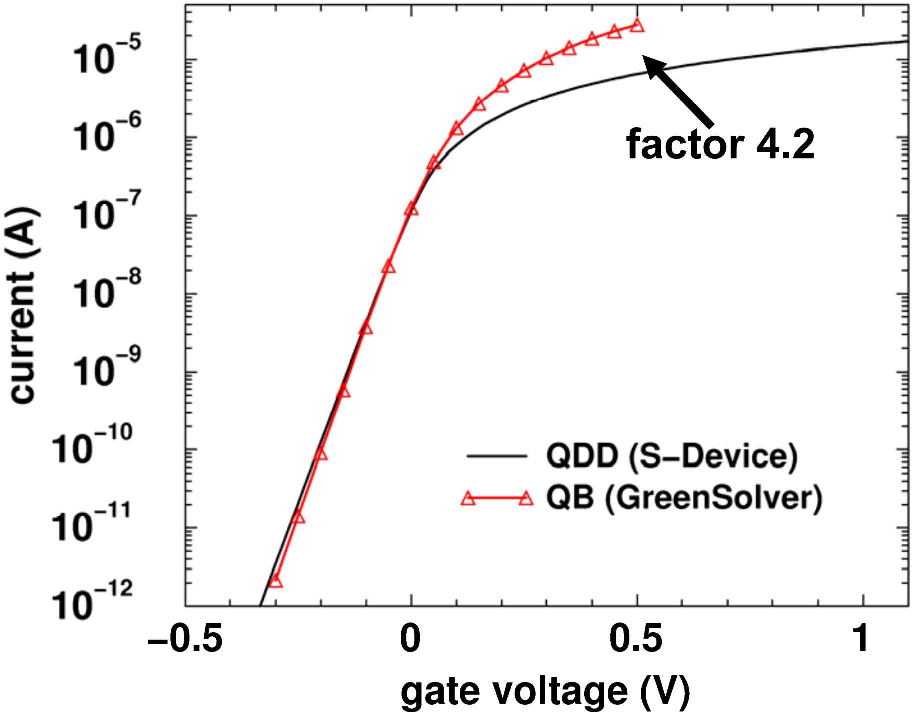
Comparison of currents

V_{DS} = 0.11V



factor 3.7

V_{DS} = 1.1V



factor 4.2

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 nm^2 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**