Physical Modeling of Deep-Submicron Devices

Andreas Schenk

Integrated Systems Laboratory
ETH Zürich

Outline

Introduction

Quantization models

Reduction of gate capacity and V_T shift

Direct tunneling gate leakage

Source-to-drain tunneling (?)

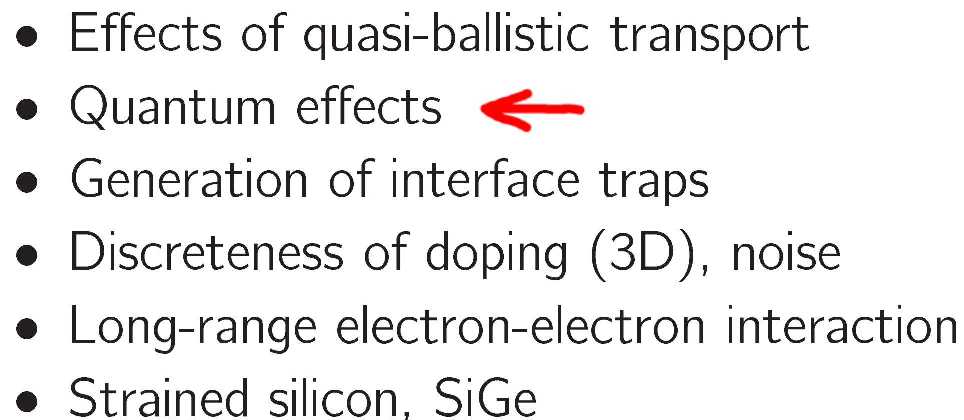
Conclusion

Introduction

Technological challenges in deep-submicron CMOS (no known solutions for some projected industry requirements), e.g.:

- Band-to-band tunneling in body-drain junction of bulk MOSFETs (GIDL)
- Direct gate tunneling leakage (off-state power consumption)
- Depletion effects (e.g. source/drain depletion in DG MOSFETs)
- Random fluctuation of device properties

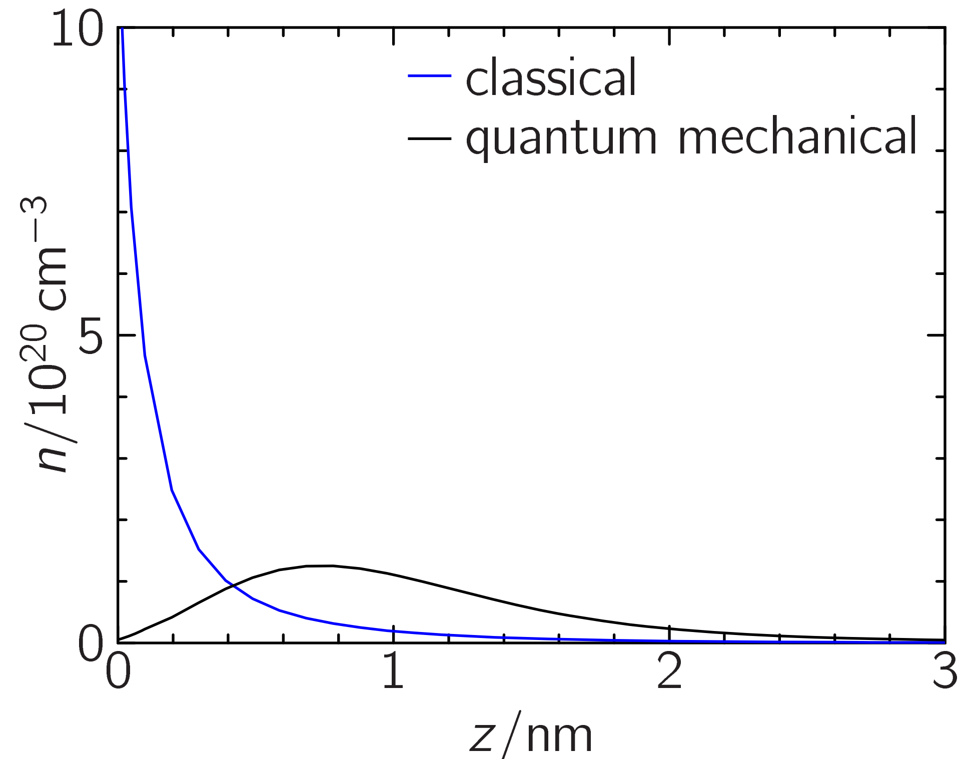
Modeling and simulation has to take up these challenges, e.g.:



- Effects of quasi-ballistic transport
- Quantum effects ←
- Generation of interface traps
- Discreteness of doping (3D), noise
- Long-range electron-electron interaction
- Strained silicon, SiGe

Quantum effects

- Confinement (density modification)
- Direct and resonant tunneling (gate current)
- B2B and defect-assisted tunneling (GIDL)
- Potential scattering (channel mobility)
- Source-to-drain tunneling



$V_G = 3\,\text{V},\ t_{ox} = 3\,\text{nm},\ N_A = 5 \cdot 10^{17}\,\text{cm}^{-3}$

Wave length of a free electron of energy $k_B T$ at 300 K: $\approx 8\,\text{nm}$

Density modification

A '**quantum potential**' Λ is introduced in the classical formula of the density (e.g. electrons and Boltzmann statistics):

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

The classical current equation reads:

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

van Dort model

Λ is a function of the local electric field $F_\perp$ normal to the interface:

$$\Lambda = a_{\text{fit}}\, h(\mathbf{d})\, (\beta \epsilon_0 \epsilon_s / 4)^{1/3} \, |F_\perp|^{2/3}$$

- Based on lowest eigenenergy of a carrier in a triangular potential well
- a_{fit} = fit factor, $h(\mathbf{d})$ = turn-off function
- Λ acts as an effective band gap widening
- E_c shift or E_v shift depending on sign of $F_\perp$
- Suited for bulk MOSFETs (for SOI and DG only, if body thickness does not reach quantum-mechanical length scales)
- Important terminal characteristics well-described, but not density
- Numerically robust

1D Schrödinger-Poisson solver

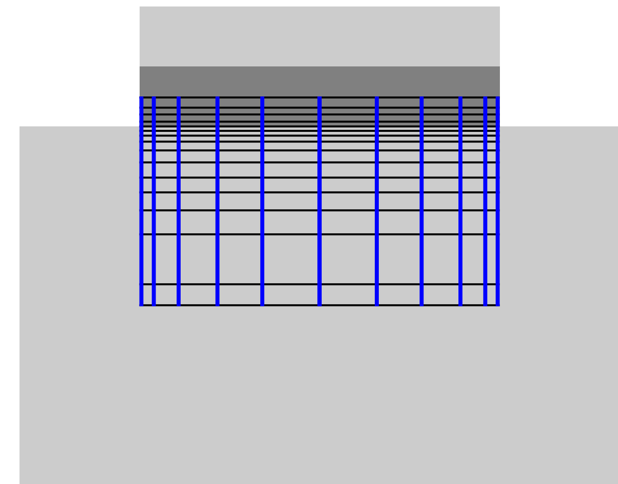
Λ follows by equating the density $n\left[\vec{R}\right]$ with the expression:

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

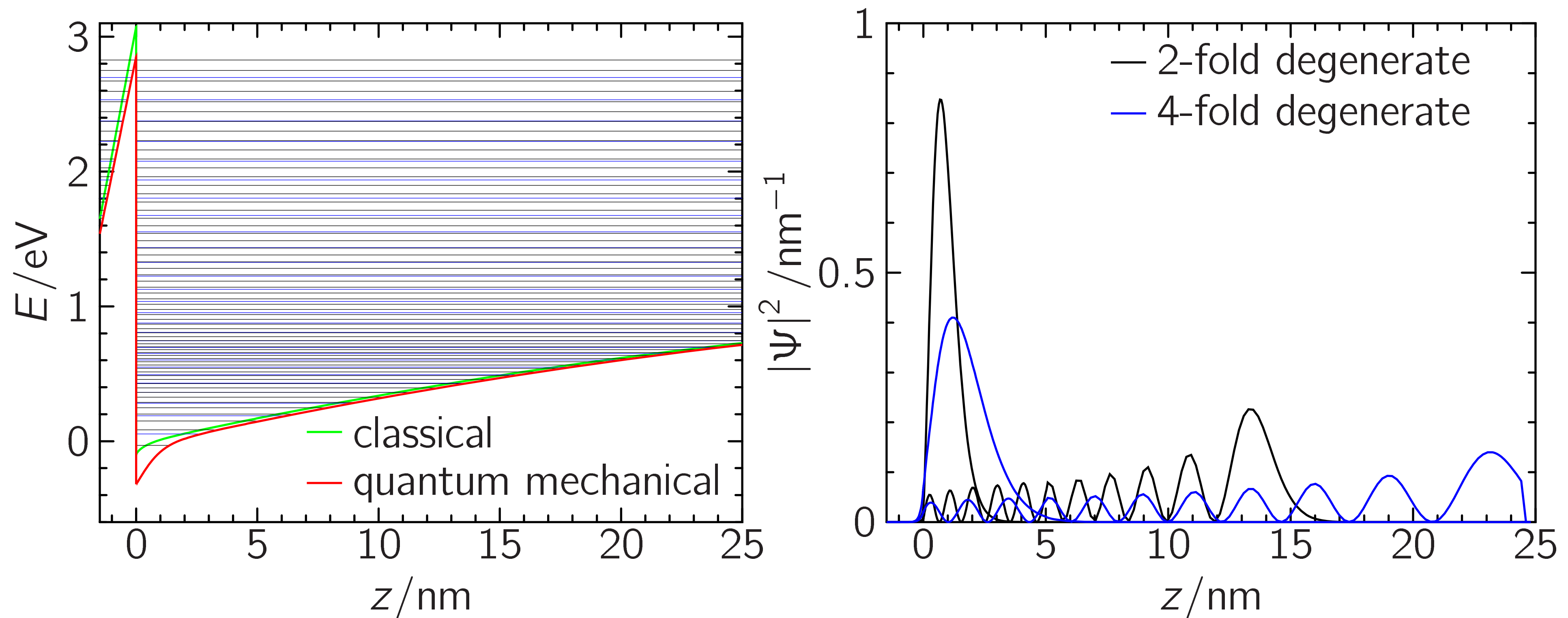
- $\Psi_j^{(\nu)}$, $E_j^{(\nu)} = j$-th eigenfunction and eigenenergy for valley ν obtained by numerical solution of the 1D Schrödinger equation in z-direction
- Boundary condition $\Psi_j^{(\nu)\prime}/\Psi_j^{(\nu)} = \pm\sqrt{2m_z^{(\nu)}|E_j^{(\nu)} - E_c|}/\hbar$ at the ends of the domain $[z_-, z_+]$ (defining the total 'quantum box')

1D Schrödinger-Poisson solver (contd.)

- CPU time depends on number of grid lines in quantization direction (may be large)
- A tensor-product grid is needed in the quantum boxes (gate oxide + channel)
- Full Newton impractical; approximation $\partial n_i^{\mathrm{qm}} / \partial \Phi_j \approx -\delta_{ij} \partial n_i^{\mathrm{qm}} / \partial E_{\mathrm{F},i}$
- Satisfactory convergence for fixed Fermi level, problems when transport equations are to be solved
- Model more suited for calibration and validation purposes than for optimization



Example: Wave functions and eigenenergies



$V_G = 3\,\mathrm{V},\ t_{ox} = 3\,\mathrm{nm},\ N_A = 5 \cdot 10^{17}\,\mathrm{cm}^{-3}$

- Self-consistency with Poisson is essential
- Finite quantum box is no problem in inversion, some problems at flat-band

Density gradient model

Λ is given by the PDE:

$$\Lambda = -\gamma \frac{\hbar^2}{12\, m} \left[\nabla^2 \log n + \frac{1}{2} (\nabla \log n)^2 \right]$$
$$= -\gamma \frac{\hbar^2}{6m} \frac{\nabla^2 \sqrt{n}}{\sqrt{n}}$$

- Method of moments for Liouville equation $i\hbar\partial_t \rho = [\mathcal{H}, \rho]_-$
- Hierarchy closed by replacing high-order derivatives of density matrix by approximation for equilibrium density matrix ρ_0:
  - Classically: diffusion term
  - Quantum-mechanically: additional quantum terms

Density gradient model (contd.)

- γ is a material-dependent fit factor
- Preconditions: thermodynamic equilibrium, isotropy of the effective mass, $\delta\Phi / k_B T \ll 1$ (Born approximation)
- Generalization for device simulation:

  DOS mass $\longrightarrow m$

  non-equilibrium density $\longrightarrow n$

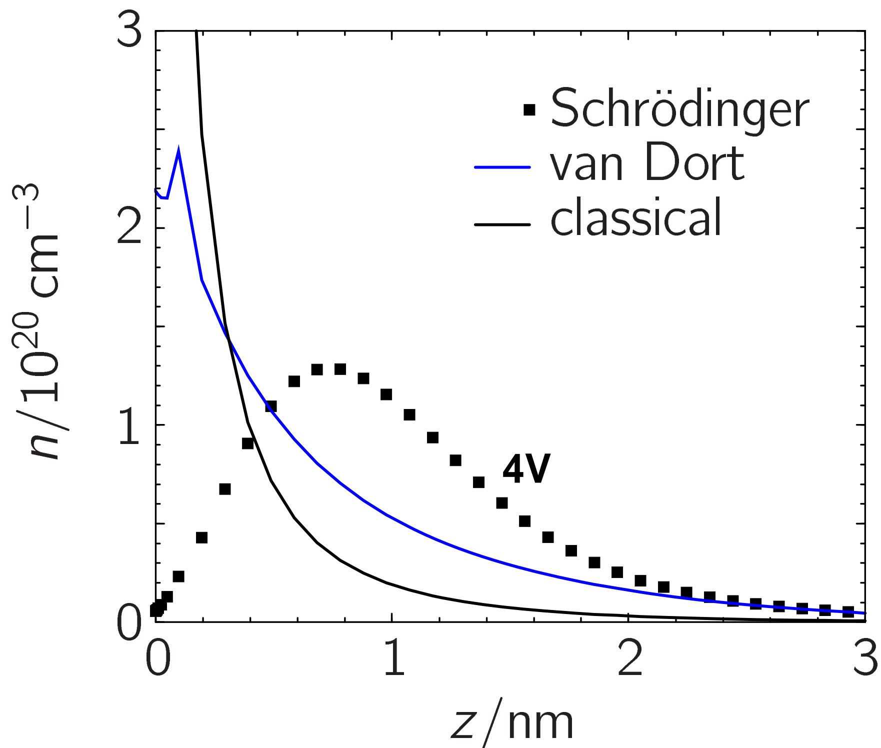
  non-perturbative formulation of Λ with the **smoothed potential** $\bar{\Phi} = E_c + \Phi_m + \Lambda$

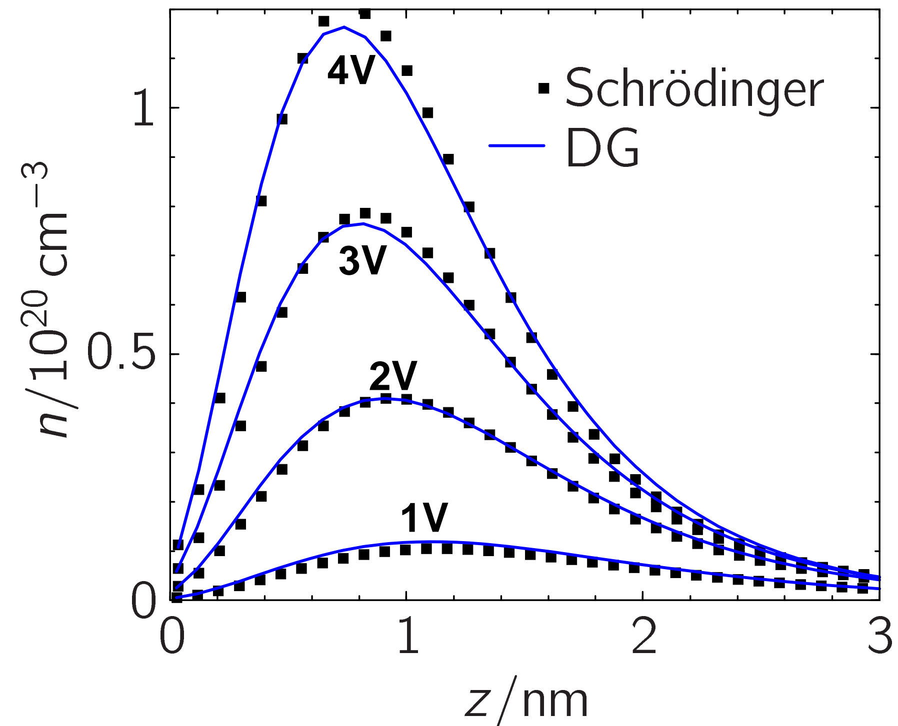
- $\Lambda = -\gamma \frac{\hbar^2}{12\,m} \left\{ \nabla^2 \left(\beta E_{F,n} - \beta\bar{\Phi} \right) + \frac{1}{2} \left[\nabla \left(\beta E_{F,n} - \beta\bar{\Phi} \right) \right]^2 \right\}$

Density gradient model (contd.)

- Discretization of Λ on unstructured grids is mandatory for professional TCAD
- Λ is treated as new system variable, sparsity structure of the Jacobian preserved
- Coupled Newton $\Rightarrow$ method is numerically robust
- DG model is multi-dimensional
- Pre-factor γ is "universal" ($\gamma = 3.6$ for silicon, if $m = m_{DOS}$)
- Many open questions (physical rigor, boundary conditions, physical models, tunneling, mass anisotropy)

Example: Density in MOS diode

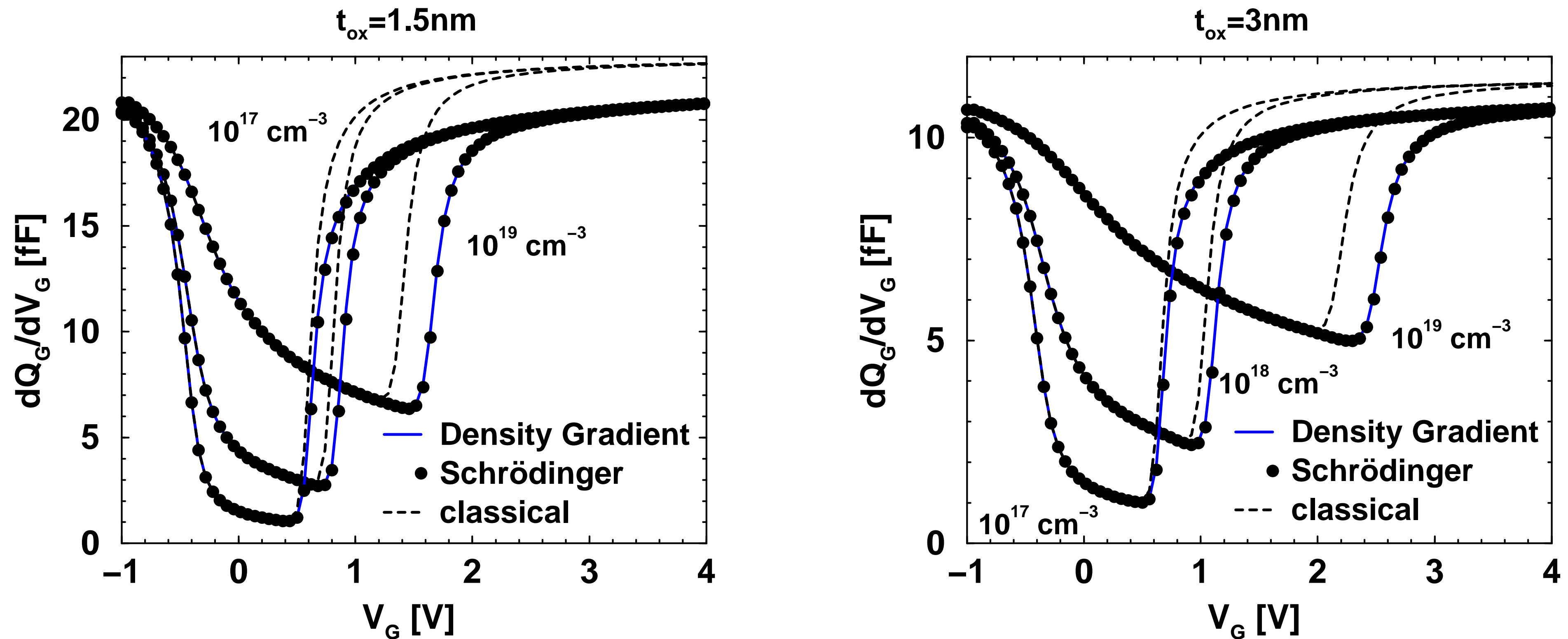




$t_{\mathrm{ox}} = 4\,\mathrm{nm},\ N_{\mathrm{poly}} = 1 \cdot 10^{20}\,\mathrm{cm}^{-3},\ N_{\mathrm{A}} = 5 \cdot 10^{17}\,\mathrm{cm}^{-3},\ T = 300\,\mathrm{K},\ \gamma_{\mathrm{ox}} = 1,$
$\gamma_{\mathrm{Si}} = 3.6$

- van Dort model does not yield the correct density
- DG density is in good agreement with "Schrödinger" for all V_{G}

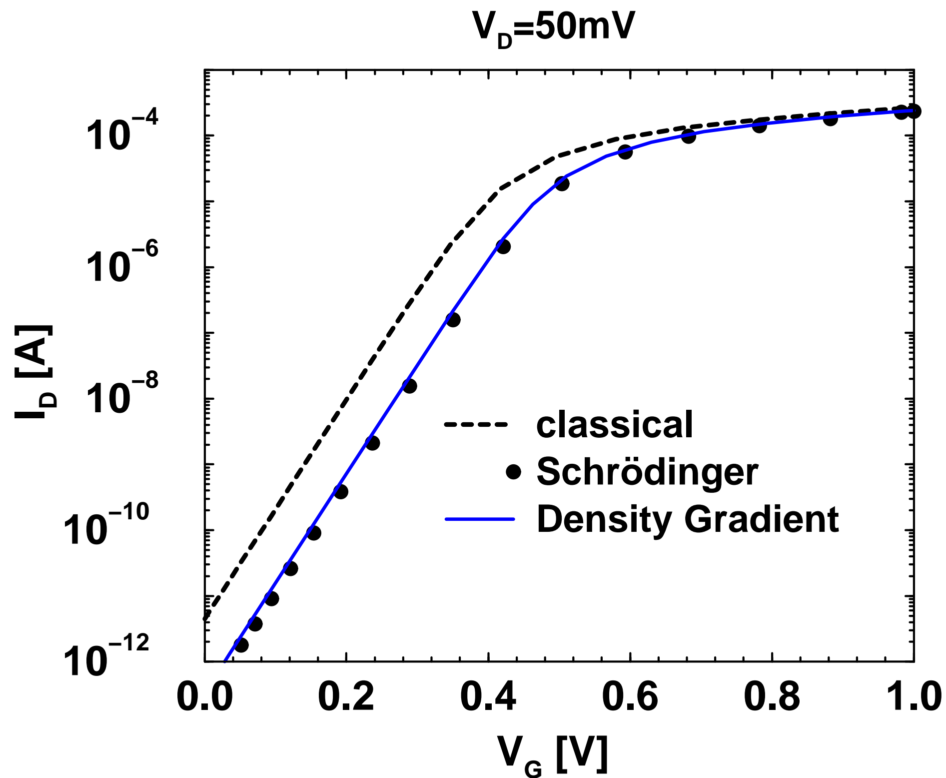
Capacity of MOS diode

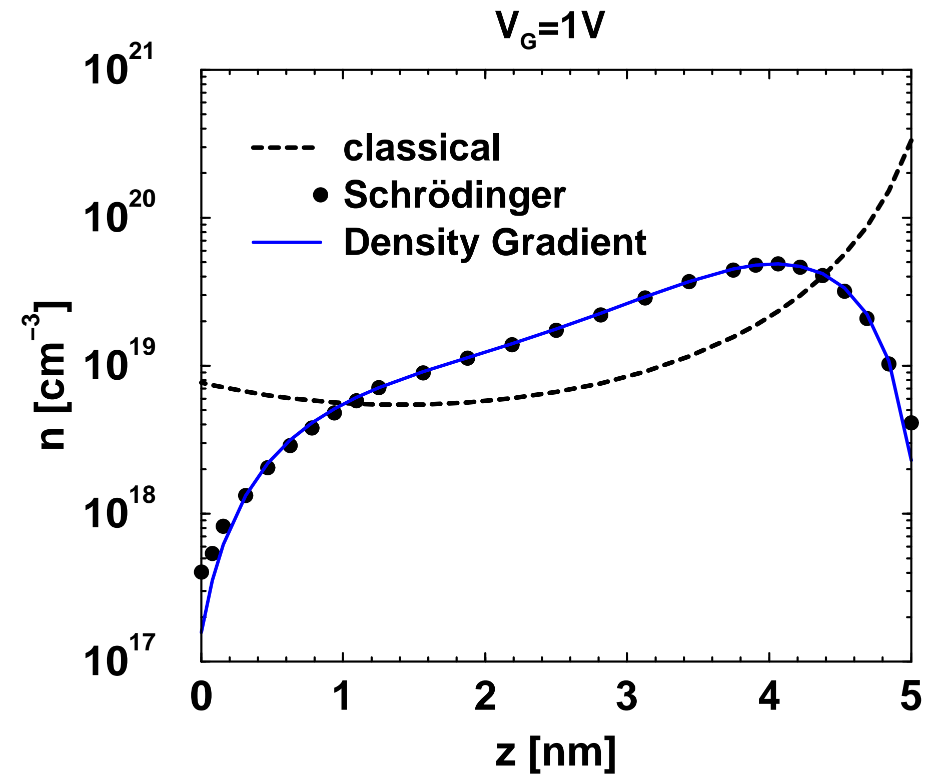


$N_{poly} = 1 \cdot 10^{20}\,\mathrm{cm}^{-3}$, $T = 300\,\mathrm{K}$, $\gamma_{ox} = 1$, $\gamma_{Si} = 3.6$, $A_G = 1\,\mu\mathrm{m}^2$

- DG model predicts the same reduction of gate capacity as "Schrödinger", independent on substrate doping and t_{ox}

Double-gated SOI MOSFET



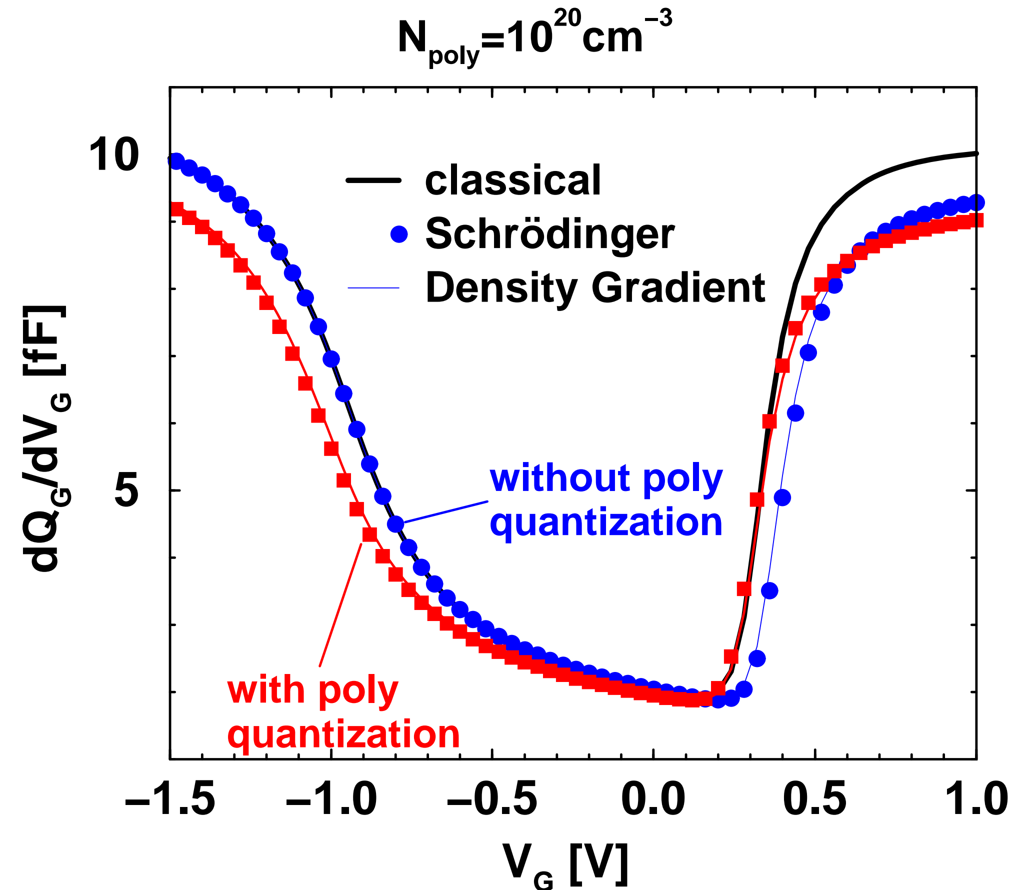


$t_{Si} = 5\,\text{nm},\ t_{ox} = 1.5\,\text{nm},\ T = 300\,\text{K}$, 1 µm channel width, 80 nm channel length

- Threshold voltage shift due to strong quantization in a double-gated SOI with $t_{Si} = 5\,\text{nm}$ is well reproduced by DG model

Quantum depletion in poly gate

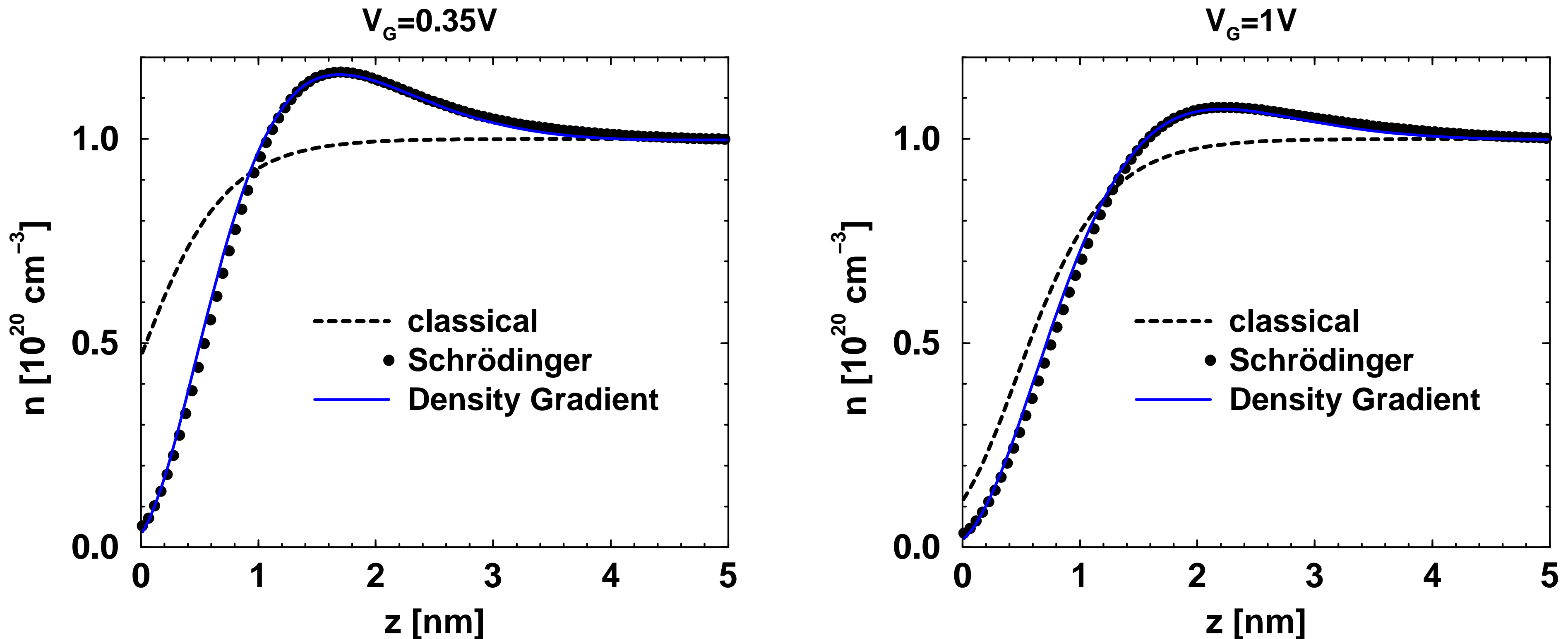
- DG sensitive to rapid changes of the potential *everywhere* in device
- $\Rightarrow$ quantum depletion in poly at poly-SiO_2 interface
- Compensation of quantum shift at threshold, if poly doping high ($\sim 1 \cdot 10^{20}\,\mathrm{cm}^{-3}$)



$t_{ox} = 3\,\mathrm{nm},\ N_A = 5 \cdot 10^{17}\,\mathrm{cm}^{-3},$

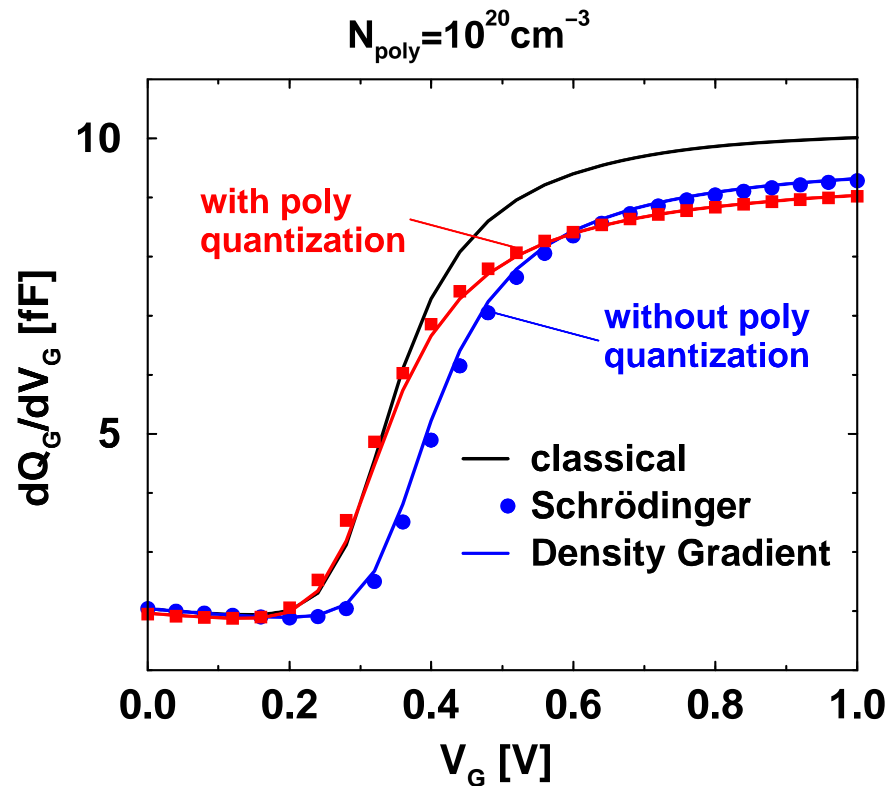
$T = 300\,\mathrm{K},\ \gamma_{ox} = 1,\ \gamma_{Si} = 3.6,\ A_G = 1\,\mu\mathrm{m}^2$

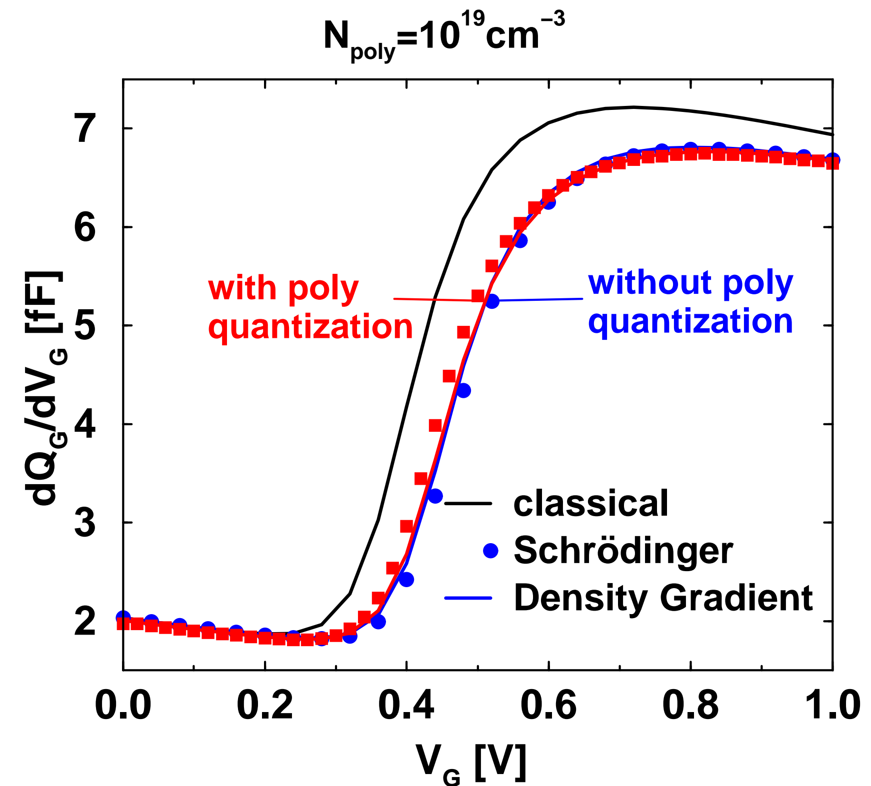
Quantum depletion in poly gate (contd.)



- A “quantum dipole” forms as the electron waves are repelled from the poly-SiO_2 interface
- With rising V_G the classical potential becomes smoother, $\Rightarrow$ quantum depletion in poly disappears

Quantum depletion in poly gate (contd.)





- Strength of the quantum dipole depends on doping level within the first few nanometers
- No effect on CV, if $N_{\mathrm{poly}} < 1 \cdot 10^{19}\,\mathrm{cm}^{-3}$

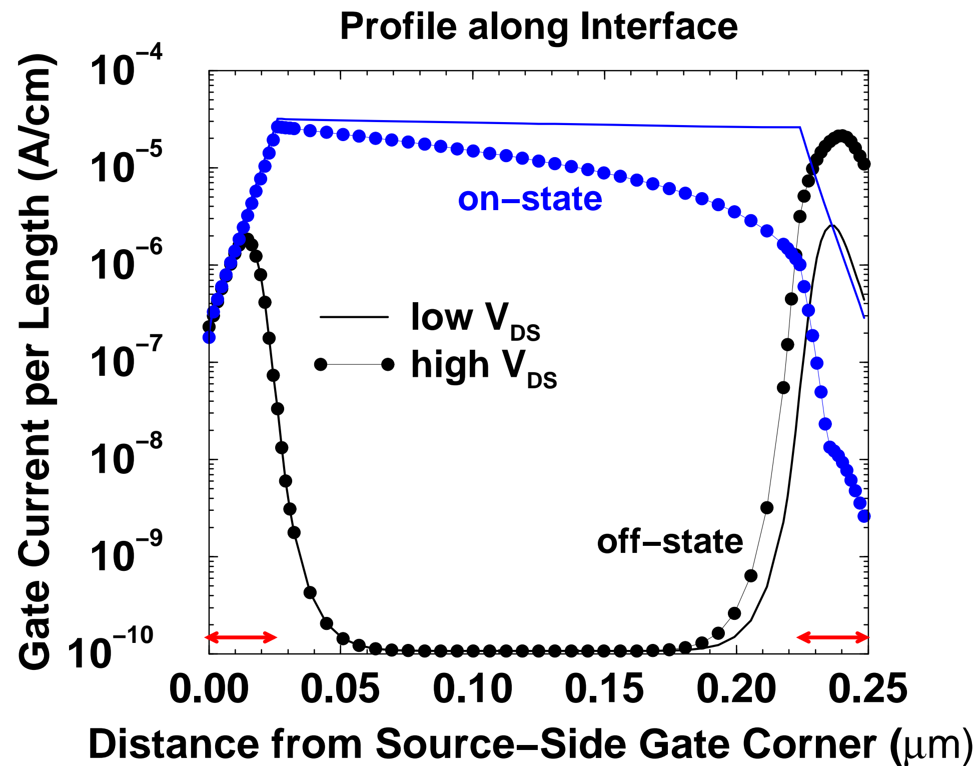
Direct tunneling gate leakage

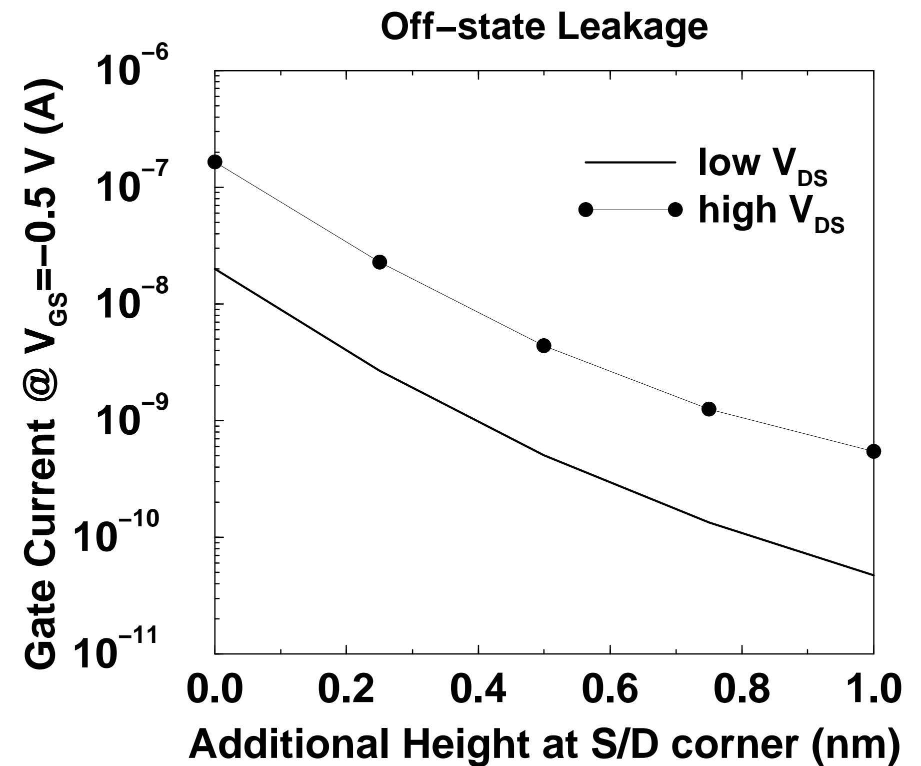
Analytical models

$$j_n = \frac{q m_c^*}{2\pi^2 \hbar^3 \beta} \int_0^\infty dE\, \mathcal{T}(E) \ln \left\{ \frac{\exp\left[\frac{E_{F,s}(d) - E_c(d) - E}{k_B T}\right] + 1}{\exp\left[\frac{E_{F,g}(0) - E_c(d) - E}{k_B T}\right] + 1} \right\}$$

- $\mathcal{T}(E)$ = transmission coefficient of potential barrier given by analytical function (limited to simple potential shapes as trapezoids)
- m_c^* = fit factor
- Weak non-locality (only Fermi level splitting)
- Self-consistent implementation as surface-recombination current
- Coupled solve with drift-diffusion equations $\Rightarrow$ link to supply from source/drain (MOSFET) or to GR in the depletion region of the substrate (MOS capacitor)
- Fast and numerically robust

Example: Gate current in MOSFET





t_{ox} : 1.5 nm $\rightarrow$ 2.0 nm in ↔

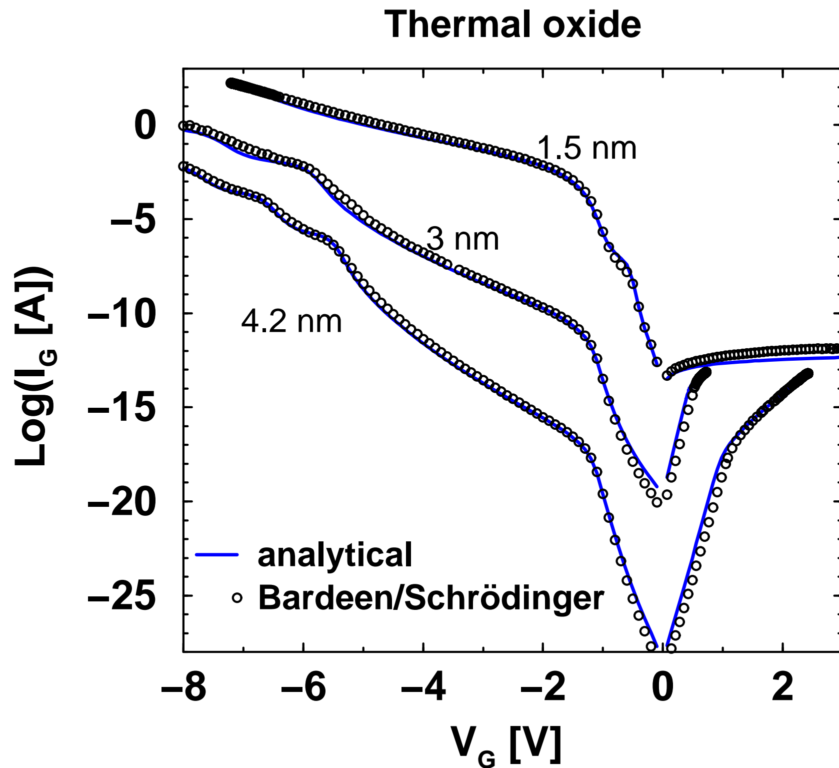
- Poly re-oxidation can strongly reduce in-tunneling (and hence off-state power consumption)

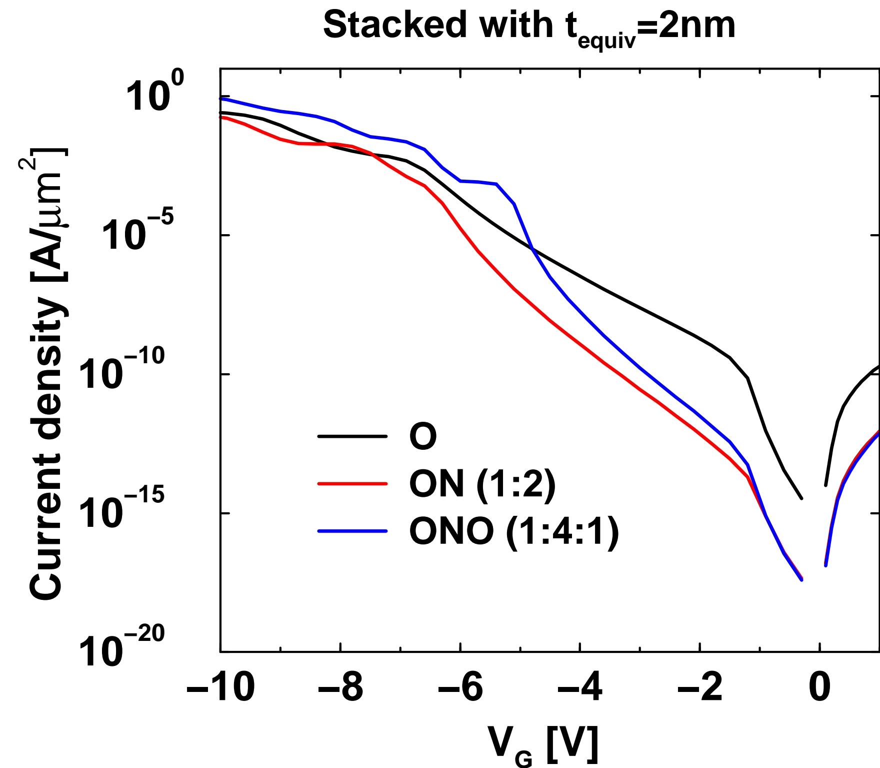
Bardeen's method

$$j_n = \frac{q\sqrt{2m_g}L}{8\pi m_{ox}^2} \sum_{i,\nu} m_{xy}^{(\nu)} \int dE\, \delta(E, E_i^{(\nu)}) \left| \Psi_{\tilde{E}} \partial_z \Psi_i^{(\nu)} - \Psi_i^{(\nu)} \partial_z \Psi_{\tilde{E}} \right|^2_{z=z_0}$$

- Perturbation scheme equivalent to ordinary first-order perturbation theory
- Allows to use numerical solutions of the Schrödinger equation $\Rightarrow$ high flexibility w.r.t. barrier potential
- Emptied channel states $\Psi_i^{(\nu)}$ refilled by SRH generation (MOS capacitor) or by supply from source/drain (MOSFET)
- Numerical problems for MOS capacitors in coupled solve due to floating "channel"
- In-tunneling requires a large number of eigensolutions on silicon side (quasi-continuum)

Example: Direct tunnel current in MOS capacitor





$$m_{\rm ox} = m_{\rm nitr} = 0.42\, m_0,\ N_{\rm A} = 1 \cdot 10^{18}\,{\rm cm}^{-3},\ T = 300\,{\rm K}$$

- Pre-factor mass m_c^* fitted once; trapezoidal barrier well described by analytical model; unknown parameters $(m_{\rm nitr})$ for stacked dielectrics

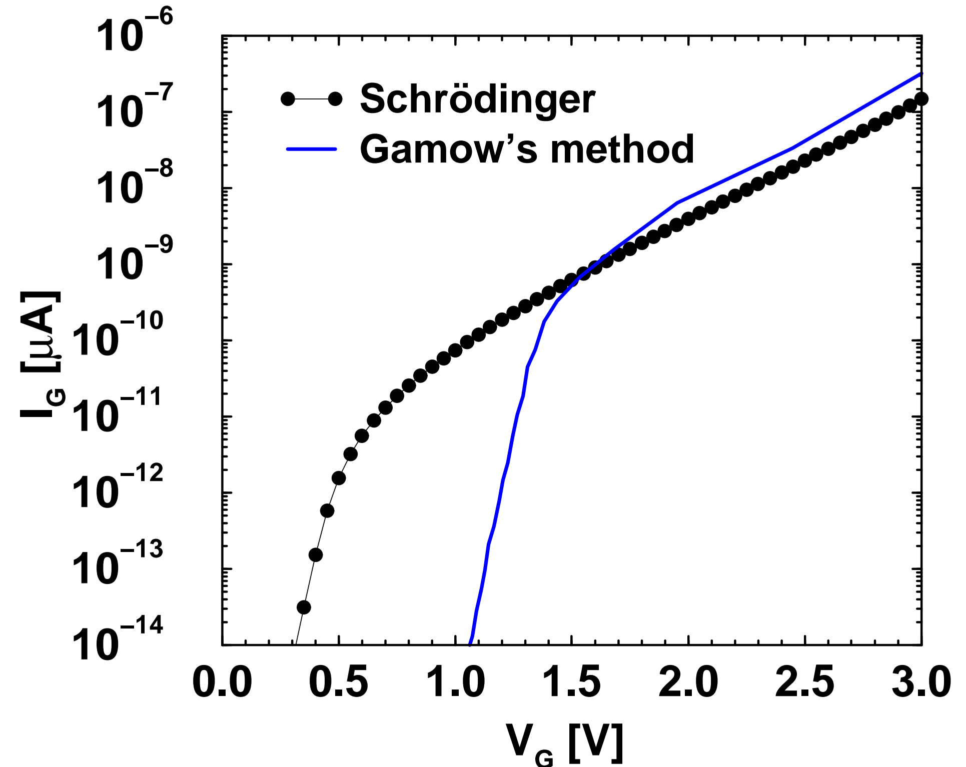
Gamow's method

$$j_n = \frac{q}{\pi \hbar^2 \beta} \sum_{i,\nu} \frac{m_{xy}^{(\nu)}}{\tau_i^{(\nu)}} \mathcal{F}\left(E_i^{(\nu)}; E_{F,s}, E_{F,g}\right)$$

- Based on Gamow's picture of nuclear decay
- $\tau_i^{(\nu)} = \hbar / 2\Gamma_i^{(\nu)}$: resonance lifetime of a quasi-bound state in the channel leaking into the gate
- Resonance width $\Gamma_i^{(\nu)}$ found numerically by solving the Schrödinger equation in a domain that covers substrate, insulator, and gate
- Precondition: $\Gamma_i^{(\nu)} \ll E_i^{(\nu)} \Rightarrow$ range of deep inversion only
- Large numerical expense to trace resonance peaks and spectral widths
- Less suited for integration into TCAD packages

Gamow's method (contd.)

- Comparison based on data interface (ESPRIT-IV 29795) between DESSIS$_{-\text{ISE}}$ and SCALPEL (IMEC)
- Identical structure and potential
- Good agreement in deep inversion (where Gamow's method is meaningful)



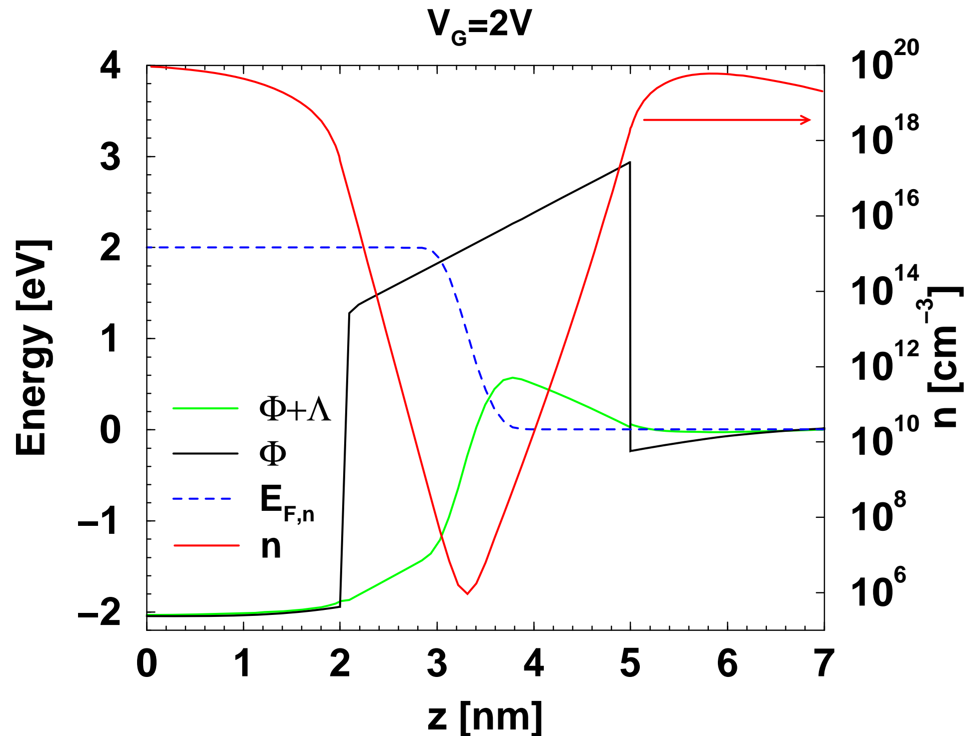
$t_{ox} = 3\,\text{nm},\ N_A = 1 \cdot 10^{15}\,\text{cm}^{-3},$
$A_G = 2 \cdot 10^5\,\mu\text{m}^2,\ T = 300\,\text{K}$

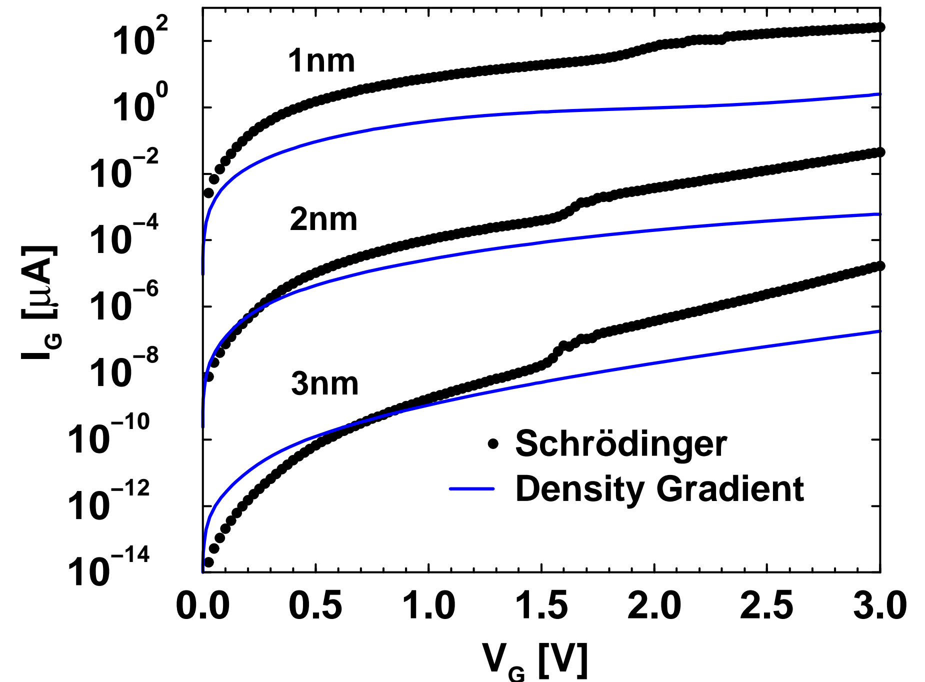
Density gradient method

$$j_n = j_{DD} \qquad \text{(ordinary drift-diffusion current)}$$

- Barriers strongly reduced by quantum potential Λ
- Barrier regions treated as wide-gap semiconductor with oxide parameters
- Current equation solved in the barrier with certain mobility value
- Quasi-Fermi level drops in the barrier region (at the minimum of the carrier density)
- $\mu_n = 0.05\,\mathrm{cm}^2/\mathrm{Vs}$ from fit to 'Schrödinger' in the case $t_{\mathrm{ox}} = 2\,\mathrm{nm}$

Density gradient method (contd.)



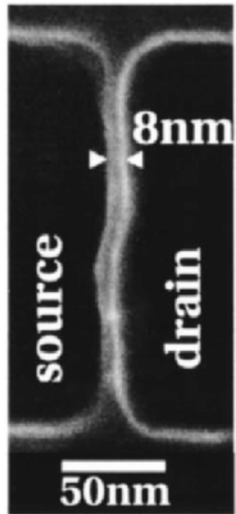


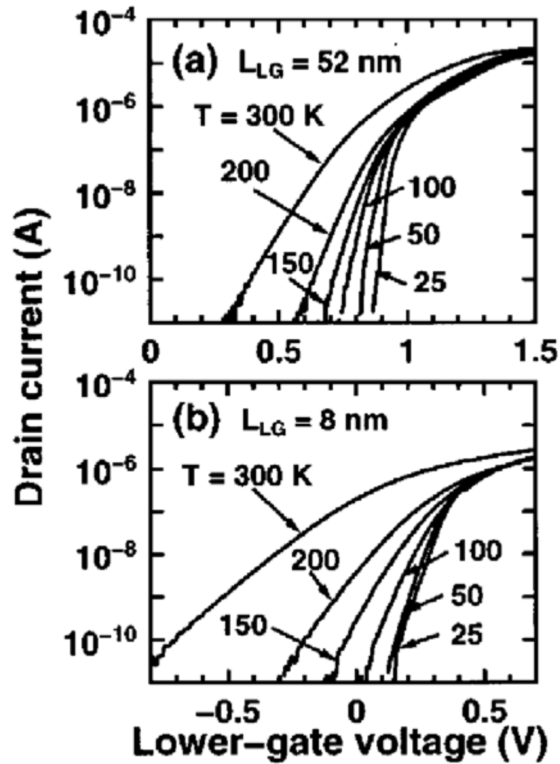
NMOSFET with $L_{\mathrm{G}} = 300\,\mathrm{nm}$, $W = 1\,\mu\mathrm{m}$, $\gamma_{\mathrm{ox}} = 1$, $\gamma_{\mathrm{Si}} = 3.6$, $\mu_{\mathrm{ox}} = 0.05\,\mathrm{cm}^2/\mathrm{Vs}$, $T = 300\,\mathrm{K}$

Source-to-drain tunneling (?)

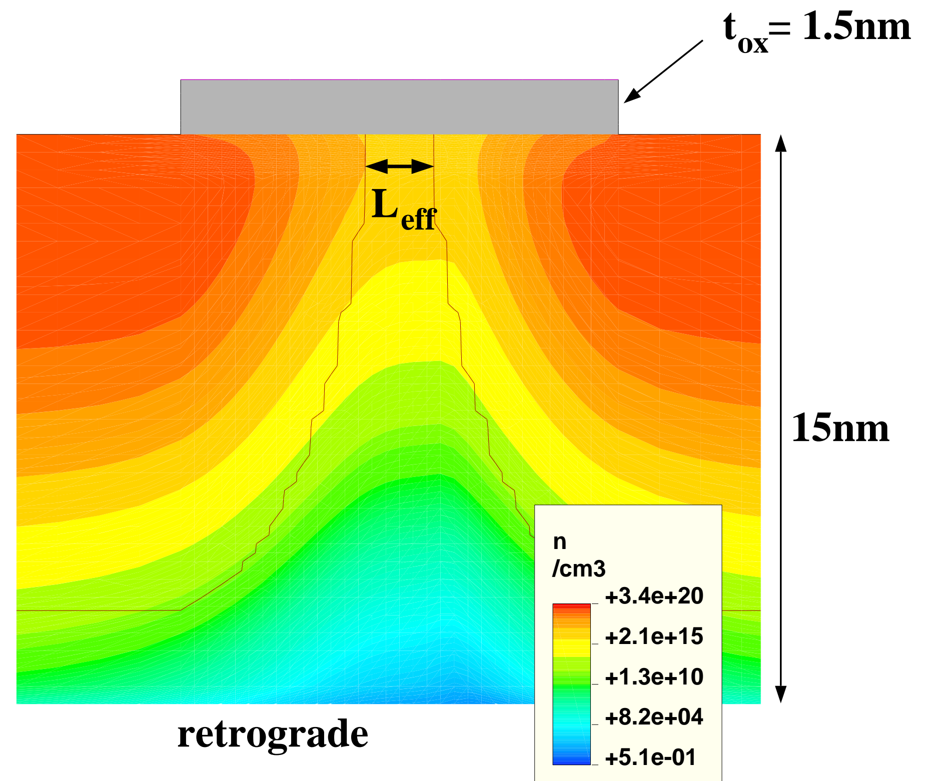
Experiment:

Kawaura et al., APL 76 (25), 3810 (2000)



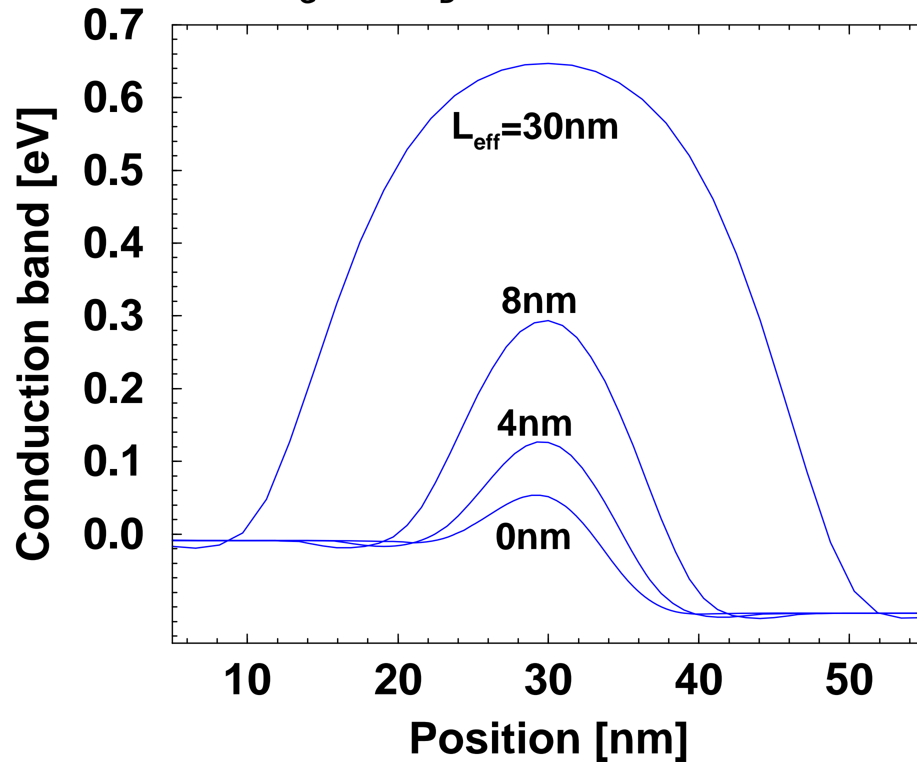


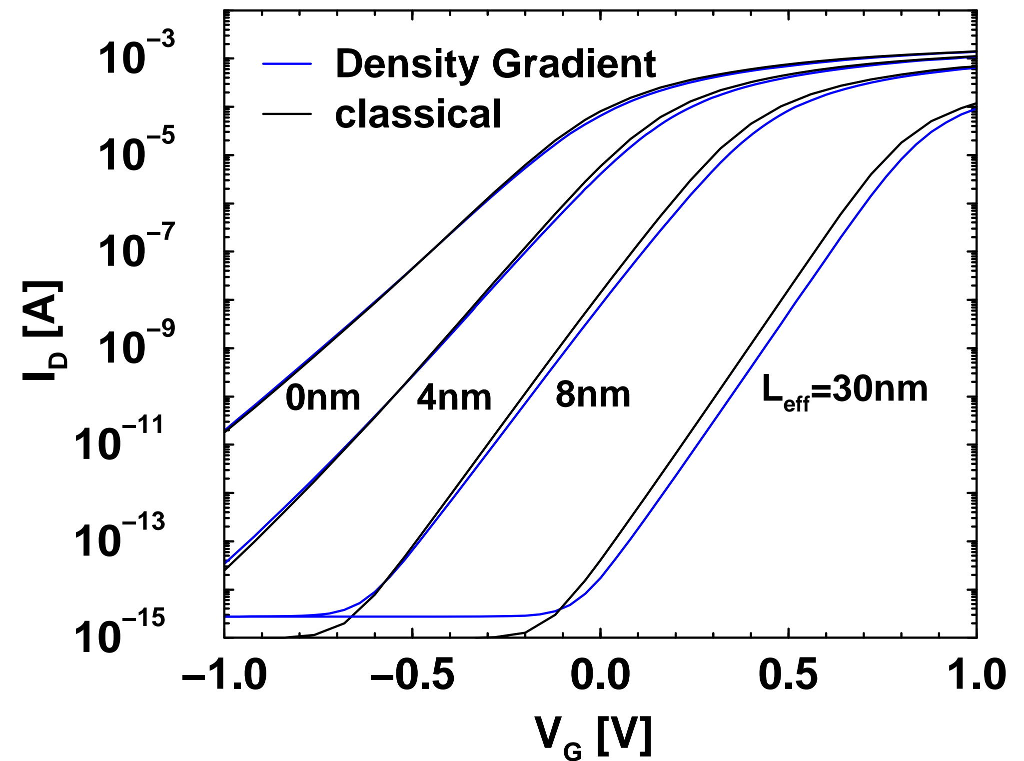
Simulation: reduction of L_{eff},



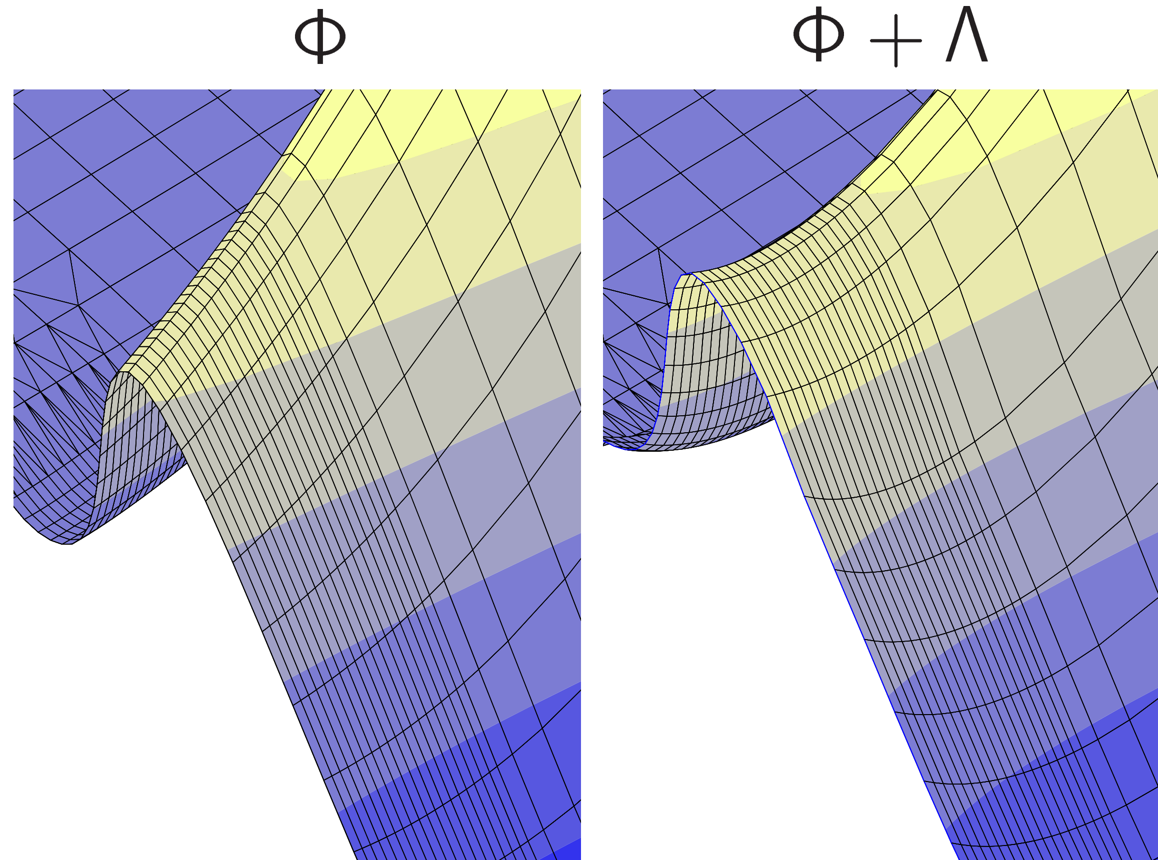
application of DG method

Potential barriers and transfer characteristics



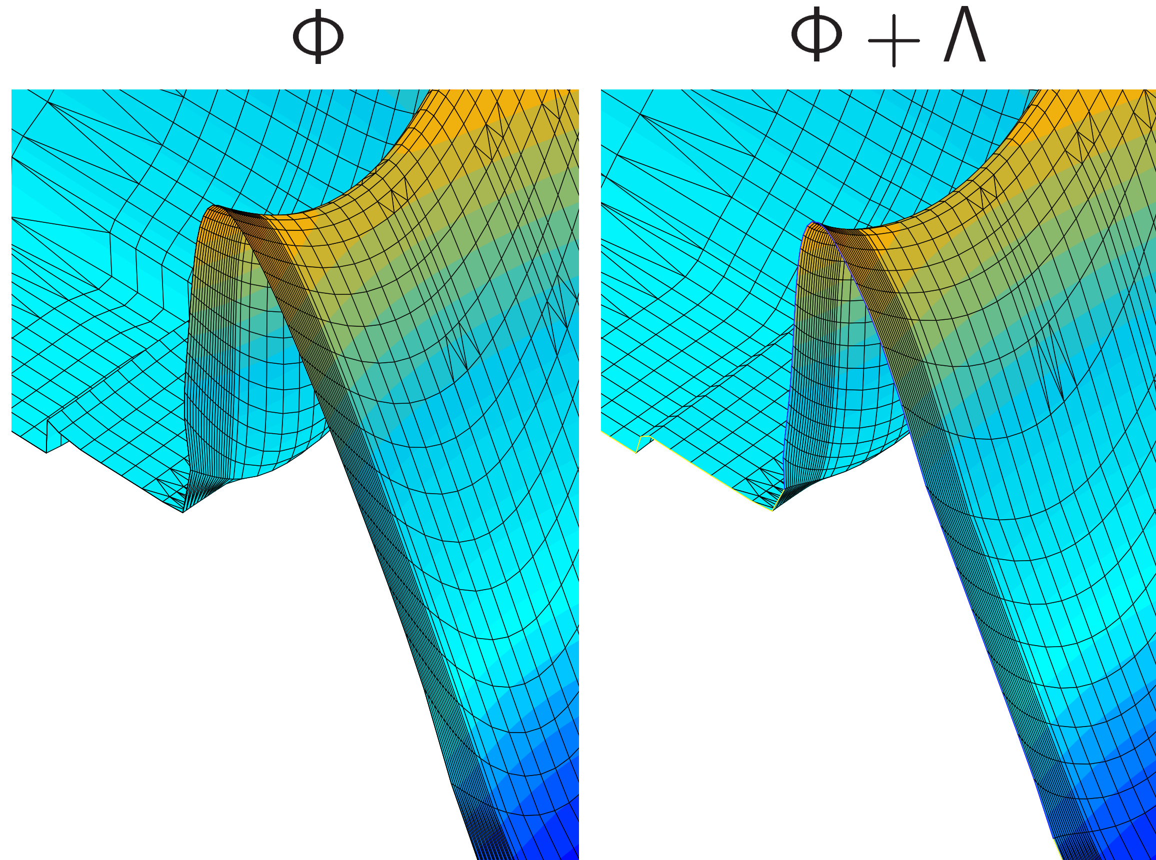


Band edge profile



NMOSFET with $L_{\text{eff}} = 4\,\text{nm}$, $W = 1\,\mu\text{m}$, $V_G = 0\,\text{V}$, $V_D = 0.1\,\text{V}$, $\gamma_{ox} = 1$, $\gamma_{Si} = 3.6$, $T = 300\,\text{K}$

Band edge profile



NMOSFET with $L_{\mathrm{eff}} = 4\,\mathrm{nm}$, $W = 1\,\mu\mathrm{m}$, $V_{\mathrm{G}} = -1\,\mathrm{V}$, $V_{\mathrm{D}} = 0.5\,\mathrm{V}$, $\gamma_{\mathrm{ox}} = 1$, $\gamma_{\mathrm{Si}} = 3.6$, $T = 300\,\mathrm{K}$

Conclusion

- Modeling of quantum effects in CMOS now established in state-of-the-art TCAD
- Schrödinger-Poisson most sound, but expensive and often convergence problems
- Density gradient method most promising alternative; quantum-corrected dissipative transport scheme, full Newton, multi-dimensional
- Direct tunneling: all the different methods give essentially the same; predictability remains limited
- Are quantum effects really important?

Acknowledgment

Collaboration with

Andreas Wettstein (ISE AG Zurich)

Eugeny Lyumkis (ISE Corp. San Jose)

Oleg Penzin (ISE Corp. San Jose)

is gratefully acknowledged.