Quantum Corrections for Monte Carlo Simulation

Brian Winstead and Umberto Ravaioli

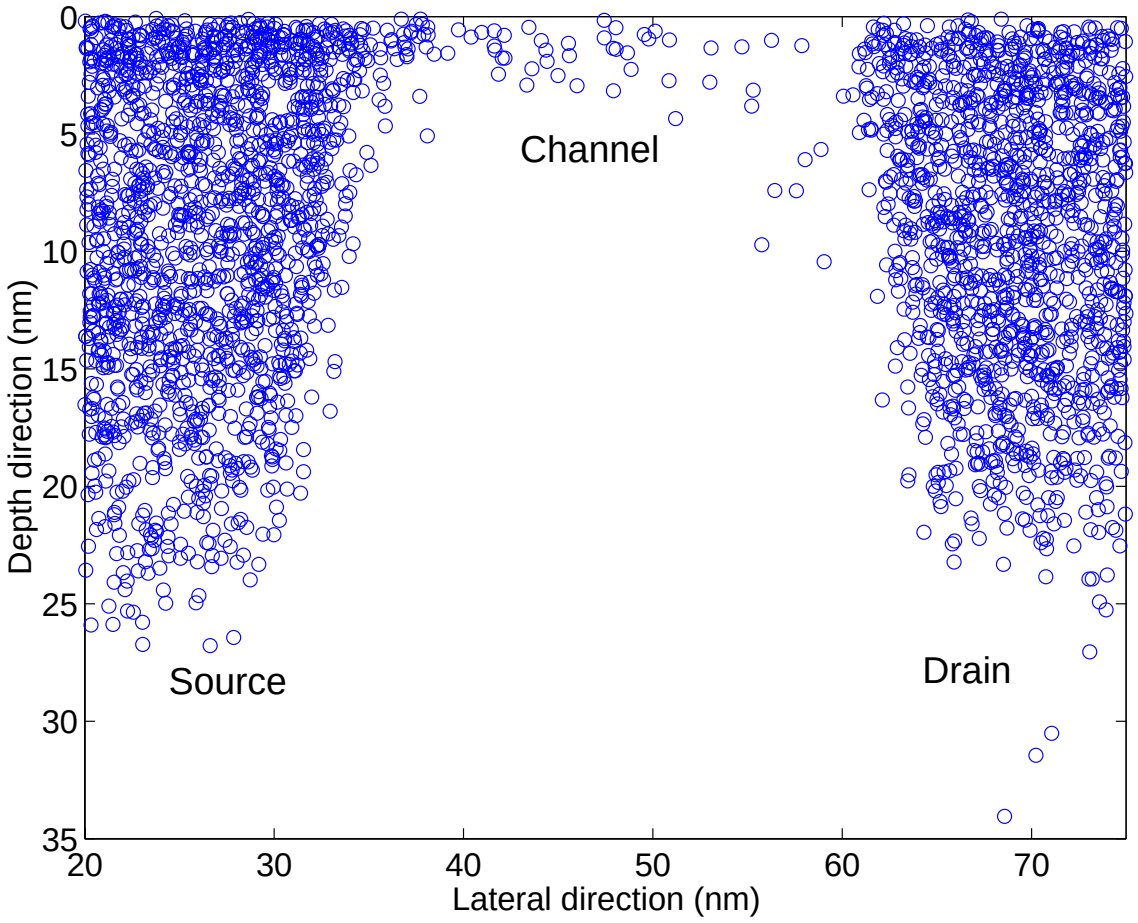
Beckman Institute

University of Illinois at Urbana-Champaign

Outline

- Quantum corrections for quantization effects
  - Effective potential
  - Wigner-based
  - Schrödinger-based
- Quantum corrections for tunneling
- Extending Schrödinger-based correction to device simulation

Monte Carlo Snapshot of a MOSFET



Motivation for Quantum Corrections

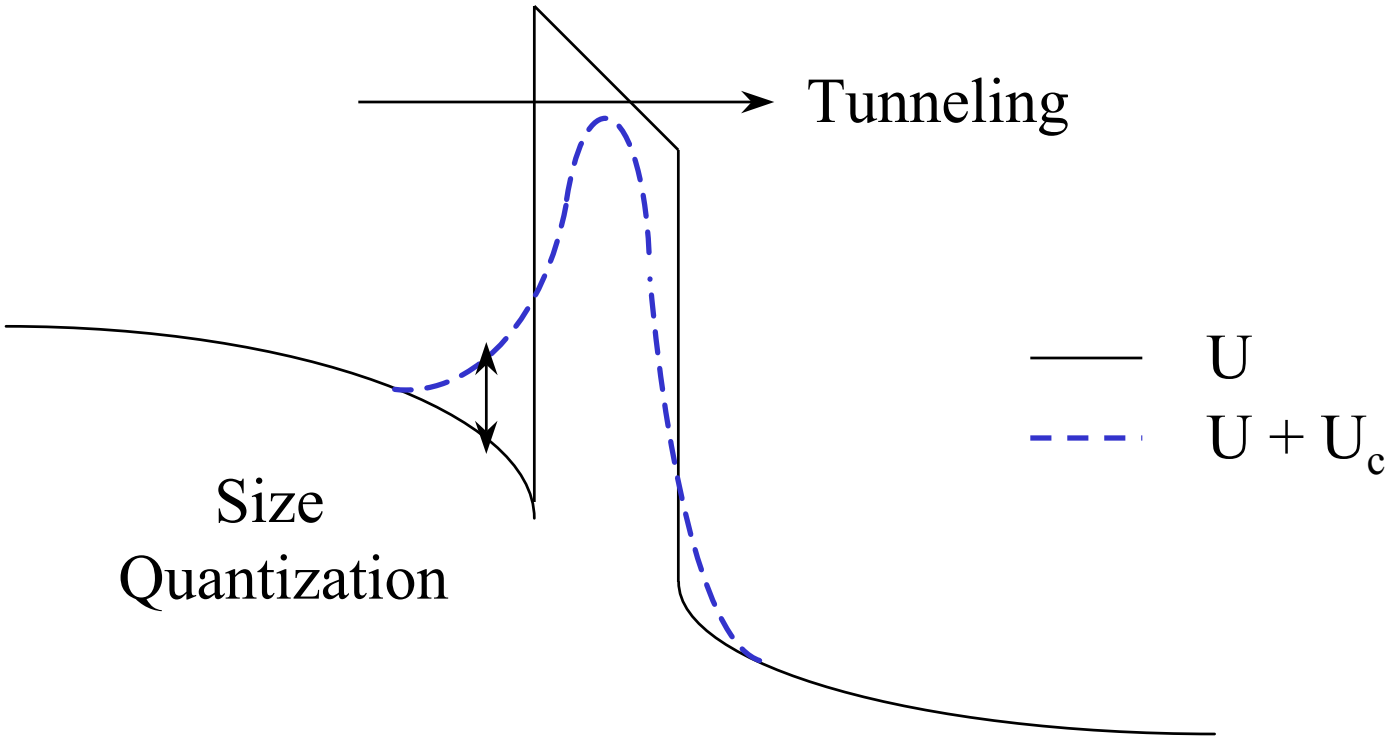
- Full-quantum transport is often impractical
- Goal is to extend the validity of semi-classical Monte Carlo to the 10-nm regime
- Quantum corrections can extend the validity of Monte Carlo in a practical way
  - Mixed quantum/classical effects are treated in a unified fashion
  - Little extra computational overhead is added in both 2D and 3D

Role of Quantum Corrections

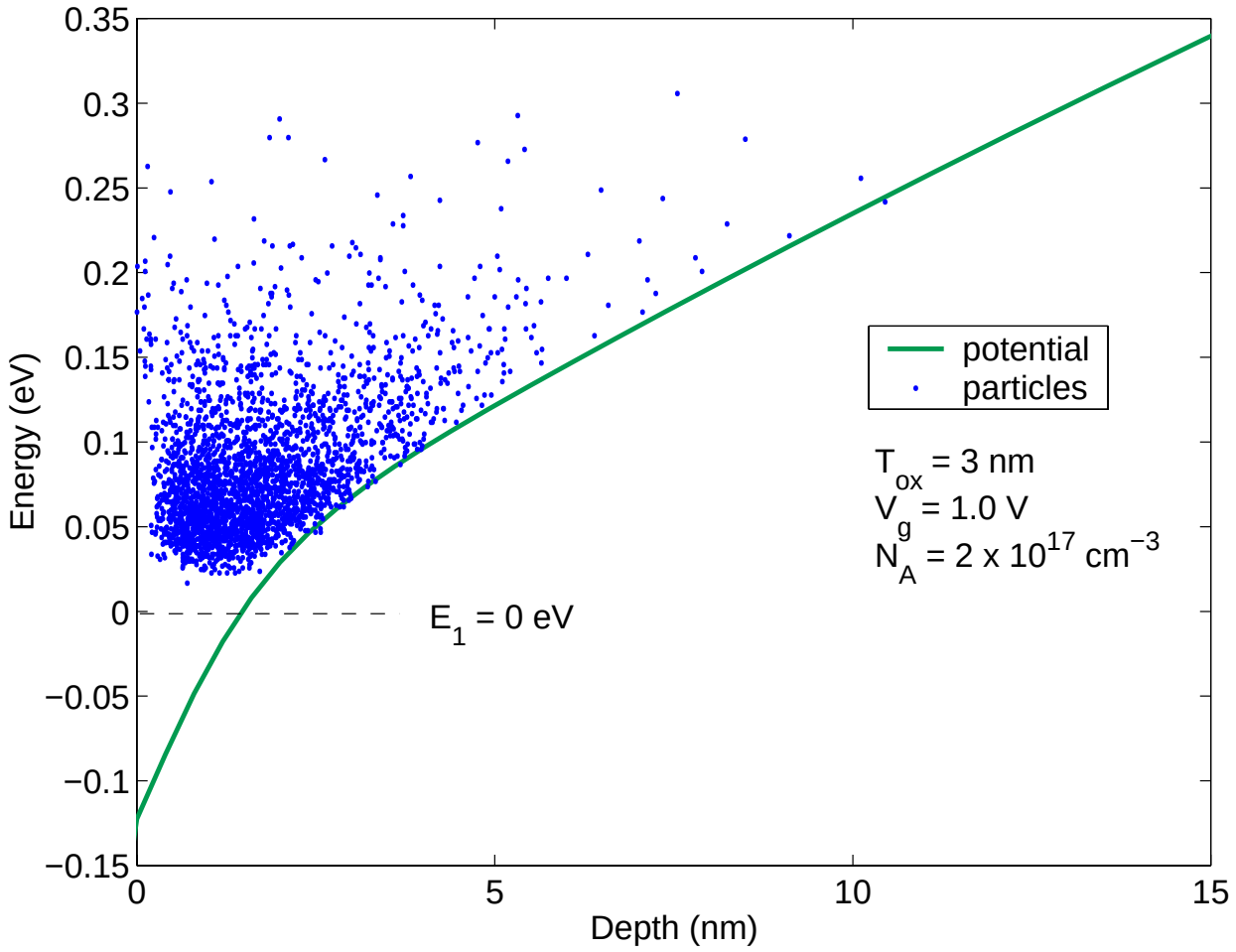
- Monte Carlo particles correctly represent the motion of wave packets centroids in the crystal
- Monte Carlo does not account for interference effects due to rapidly varying applied fields or heterojunctions
- Quantum corrections can capture non-coherent interference effects
- Coherent transport effects are small for pure silicon devices above 10-nm regime

Corrections in Monte Carlo

Quantum effects can be approximated in Monte Carlo by correcting the classical potential



Snapshot from MOS Capacitor



Effective Potential

- Feynman developed the effective potential in the 1960s and applied it to quantum corrections in statistical mechanics
- Particles feel nearby potential due to quantum fluctuations around classical path of least action
- V_{eff} is a non-local function of the nearby potential

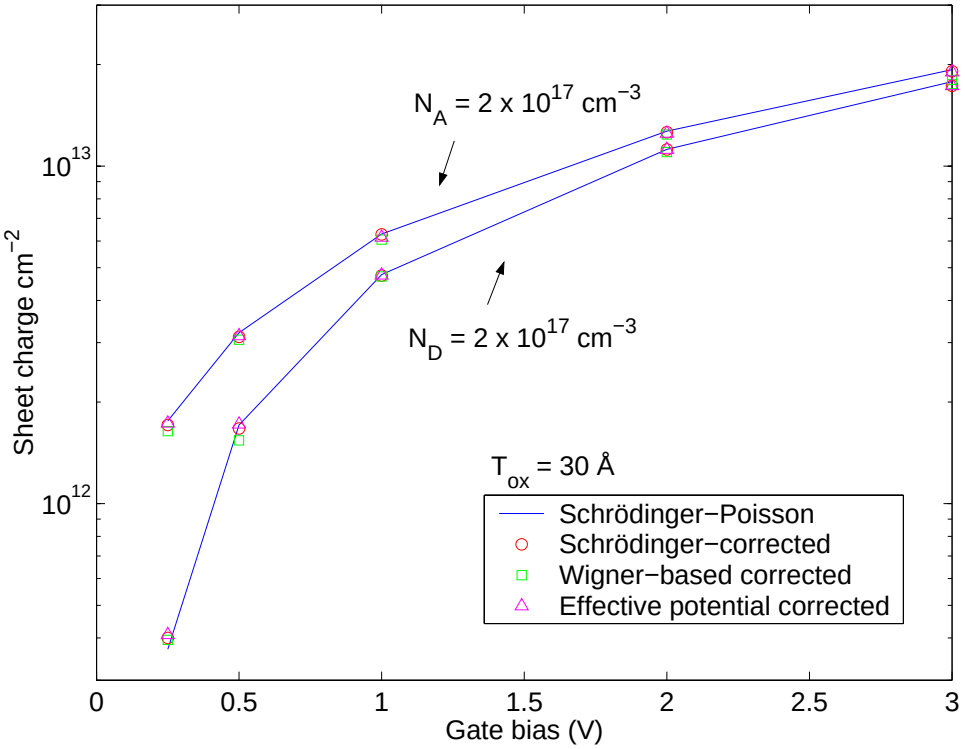
$$V_{eff}(x) = \int V(x) e^{-\frac{(x-x')^2}{2a^2}} dx' \quad , \quad a = \frac{\hbar^2}{12mkT}$$

Properties of Effective Potential

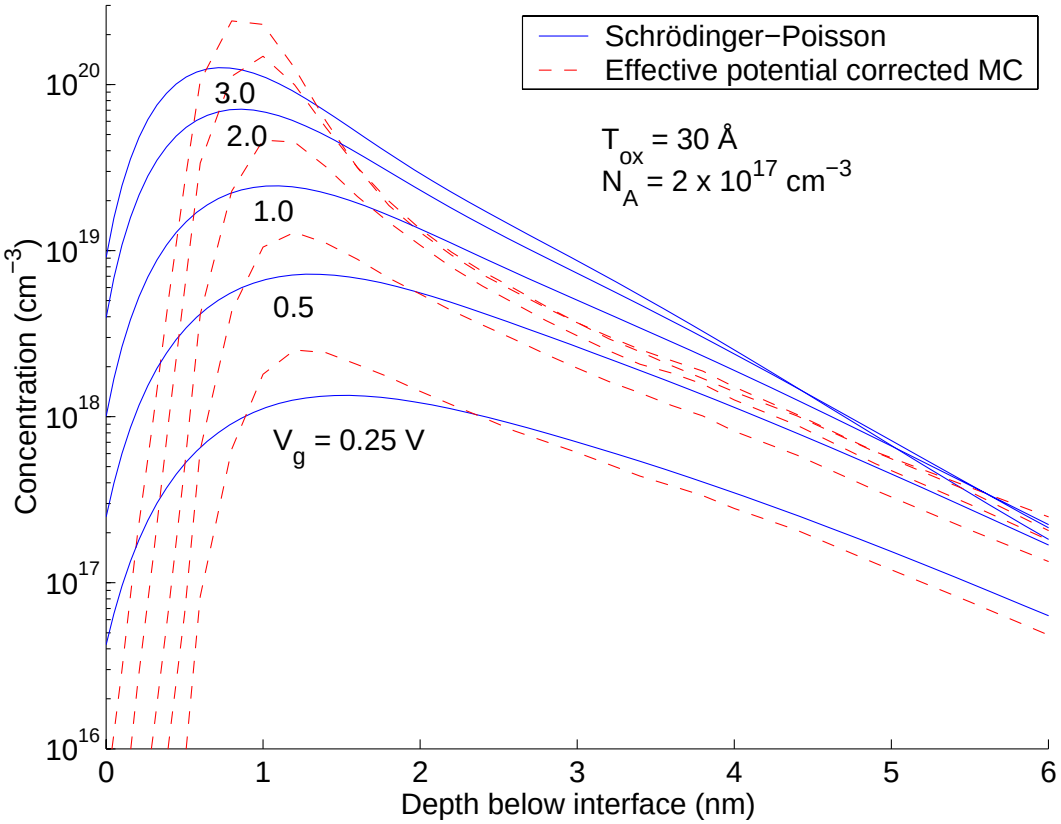
- Simple to implement and to calculate
- Not sensitive to noise from Monte Carlo
- Works best for smooth, symmetric potentials
- *a* can be treated as a fitting parameter describing the "size" of the particle
- Detailed solution near large heterojunctions is typically incorrect and cannot be fit

Effective potential for MOS

- Effective size *a* can be tuned to match sheet charge density



Effective potential for MOS



Wigner Formulation

- The Wigner method was developed in the 1930s for quantum correction to statistical mechanics
- Wigner formulation of quantum mechanics formally separates the quantum and classical contributions to the equation of motion

$$\frac{\partial f_w}{\partial t} = -\frac{\partial H}{\partial p}\frac{\partial f_w}{\partial x} + \frac{\partial H}{\partial x}\frac{\partial f_w}{\partial p}$$

← Classical Boltzmann equation

Quantum Contribution →

$$-\frac{\hbar^2}{24}\frac{\partial^3 H}{\partial x^3}\frac{\partial^3 f_w}{\partial p^3} + \frac{\hbar^4}{1920}\frac{\partial^5 H}{\partial x^5}\frac{\partial^5 f_w}{\partial p^5} + \ldots$$

- We start from the general Wigner function representation of quantum transport

$$f_w(\vec{r},\vec{p}) = \frac{1}{\hbar^3}\int dy\ e^{j\vec{p}\cdot\vec{y}/\hbar}\rho(x+y/2, x-y/2)$$

where

$$\rho(x,x')$$

is the density matrix.

- The Wigner function is the quantum equivalent of the distribution function in the semi-classical Boltzmann equation.

- The quantum transport equation of the Wigner function has the form (parabolic bands, ballistic)

$$\frac{\partial f_w}{\partial t} + \frac{\vec{p}}{m^*} \cdot \nabla f_w = \left(\frac{1}{\pi \hbar^2} \right)^3 \iint d\vec{s}\, d\vec{P}\; K(\vec{s}, \vec{P})\; f_w(\vec{r}, \vec{p} + \vec{P}, t)$$

$$K(\vec{s}, \vec{P}) = \left[U\left(\vec{r} + \frac{\vec{s}\hbar}{2} \right) - U\left(\vec{r} - \frac{\vec{s}\hbar}{2} \right) \right] \sin(\vec{s} \cdot \vec{P})$$

- Direct solution of the Wigner transport equation is still a considerable numerical challenge.
- We are interested in determining a truncated expansion of the quantum equation, that resembles the standard Boltzmann equation, so that the standard Monte Carlo technique can be applied with minor modifications.

Quantum corrections for Monte Carlo simulation

- **The complete Wigner transport equation, inclusive of collision terms, can be reformulated to resemble Boltzmann equation**

$$\frac{\partial f}{\partial t} + \vec{\mathbf{v}} \cdot \nabla_{\vec{\mathbf{r}}} f - \frac{1}{\hbar} \nabla_{\vec{\mathbf{r}}} U \cdot \nabla_{\vec{\mathbf{k}}} f$$

$$+ \sum_{\alpha=1}^{\infty} \frac{(-1)^{\alpha+1}}{\hbar 4^{\alpha} (2\alpha+1)!} \left(\nabla_{\vec{\mathbf{r}}} U \cdot \nabla_{\vec{\mathbf{k}}} f \right)^{2\alpha+1} = \left(\frac{\partial f}{\partial t} \right)_C$$

- **At first order, we truncate considering only $\alpha = 1$.**

Quantum corrections for Monte Carlo simulation

- The truncated equation has a form resembling Boltzmann equation

$$\frac{\partial f}{\partial t} + \vec{\mathbf{v}} \cdot \nabla_{\vec{\mathbf{r}}} f - \frac{1}{\hbar} F^{qc} \cdot \nabla_{\vec{\mathbf{k}}} f = \left(\frac{\partial f}{\partial t} \right)_C$$

  where F^{qc} contains the quantum correction to the forces.

- With the modified forces, the particles move as if under the influence of a classical potential, but following equivalent quantum trajectories.
- The quantum correction essentially modifies the potential energy felt by the particles.

Quantum corrections for Monte Carlo simulation

- **The corrected forces can be evaluated by making assumptions on distribution function and bandstructure. For**

$$f = \exp\left\{-\frac{1}{k_B T}\left[E_{\mathrm{k}-\bar{\mathrm{k}}} + U(\vec{\mathrm{r}}) - E_f\right]\right\}$$

Displaced Maxwellian

$$E_{\mathrm{k}-\bar{\mathrm{k}}} = \sum_{i=x,y,z} \frac{\hbar^2 \left(k_i - \bar{k}_i\right)^2}{2m_i}$$

$$\bar{k} = \left(\bar{k}_x, \bar{k}_y, \bar{k}_z\right) = \text{momentum centroid}$$

Parabolic bands

$$F_x^{qc} = -\frac{\partial}{\partial x}\left(U - \frac{1}{24}\left\{\left[\gamma_x^2\left(k_x - \bar{k}_x\right)^2 - 3\gamma_x\right]\frac{\partial^2 U}{\partial x^2} + 3\left[\gamma_y^2\left(k_y - \bar{k}_y\right)^2 - \gamma_y\right]\frac{\partial^2 U}{\partial y^2}\right\}\right)$$

$$F_y^{qc} = -\frac{\partial}{\partial y}\left(U - \frac{1}{24}\left\{\left[\gamma_y^2\left(k_y - \bar{k}_y\right)^2 - 3\gamma_y\right]\frac{\partial^2 U}{\partial y^2} + 3\left[\gamma_x^2\left(k_x - \bar{k}_x\right)^2 - \gamma_x\right]\frac{\partial^2 U}{\partial x^2}\right\}\right)$$

$$\text{with} \quad \gamma_x = \frac{\hbar^2}{m_x k_B T}; \quad \gamma_y = \frac{\hbar^2}{m_y k_B T}$$

Quantum corrections for Monte Carlo simulation

- The forces obtained have still some practical problems in regions like sharp interfaces, where quantum effects are prominent.
- To obtain a smooth potential, we can use approximate relations obtained by integrating the displaced Maxwellian distribution with the momentum.
- We obtain these alternative expressions for the second order derivatives of the potential

$$\frac{\partial^2 U}{\partial x^2} \simeq -k_B T \frac{\partial^2 (\ln n)}{\partial x^2}; \qquad \frac{\partial^2 U}{\partial y^2} \simeq -k_B T \frac{\partial^2 (\ln n)}{\partial y^2}$$

Quantum corrections for Monte Carlo simulation

- **The "smooth" version of the quantum corrected forces is**

$$F_x^{qc} = -\frac{\partial}{\partial x}\left(U - \frac{k_B T}{24}\left\{ \left[\gamma_x^2 \left(k_x - \bar{k}_x \right)^2 - 3\gamma_x \right] \frac{\partial^2 \ln n}{\partial x^2} + 3\left[\gamma_y^2 \left(k_y - \bar{k}_y \right)^2 - \gamma_y \right] \frac{\partial^2 \ln n)}{\partial y^2} \right\} \right)$$

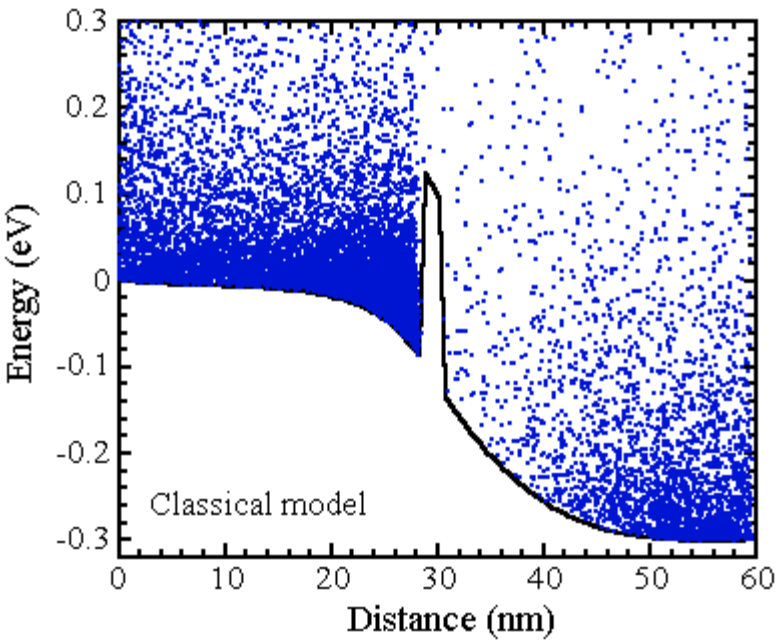
$$F_y^{qc} = -\frac{\partial}{\partial y}\left(U - \frac{k_B T}{24}\left\{ \left[\gamma_y^2 \left(k_y - \bar{k}_y \right)^2 - 3\gamma_y \right] \frac{\partial^2 \ln n}{\partial y^2} + 3\left[\gamma_x^2 \left(k_x - \bar{k}_x \right)^2 - \gamma_x \right] \frac{\partial^2 \ln n}{\partial x^2} \right\} \right)$$

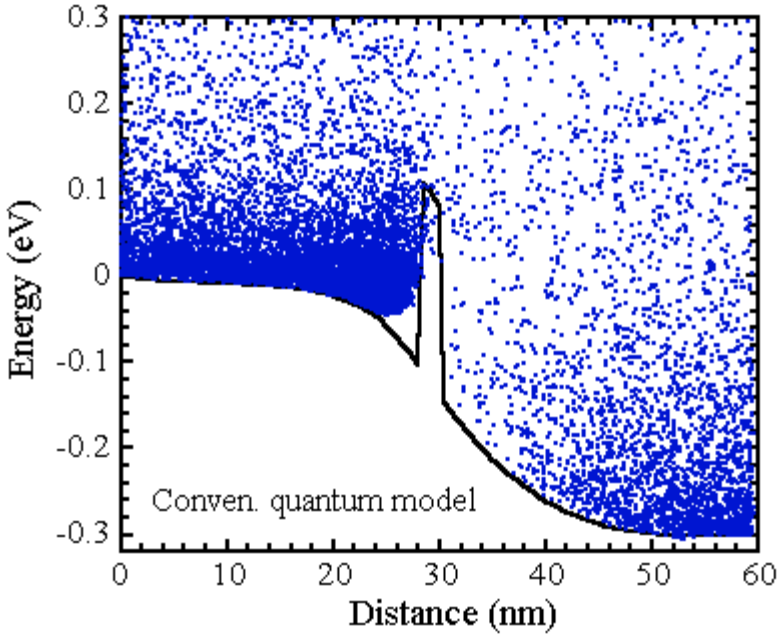
- **This formulation has explicit momentum dependence and improves upon previous results in the literature where the momentum terms were evaluated with the thermal energy**

$$\frac{\hbar^2 \left(k_i - \bar{k}_i \right)^2}{2m_i} = \frac{k_B T}{2}$$

Quantum corrections for Monte Carlo simulation

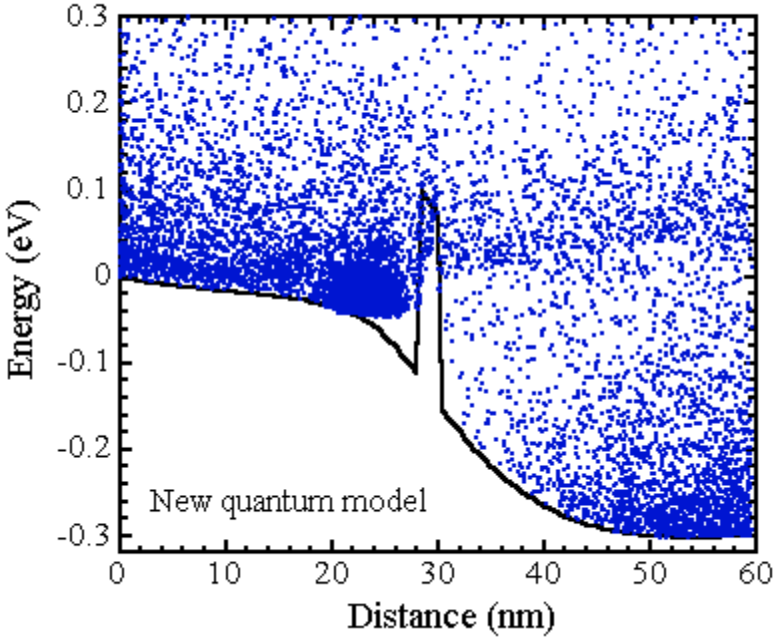
- **Test of the quantum corrections in Monte Carlo:**
  Single GaAs/AlGaAs/GaAs barrier with a fixed potential (1)

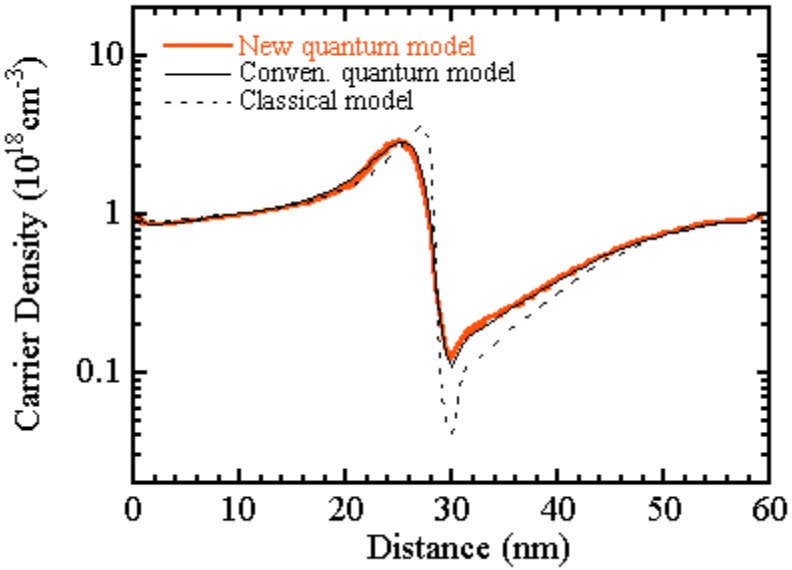




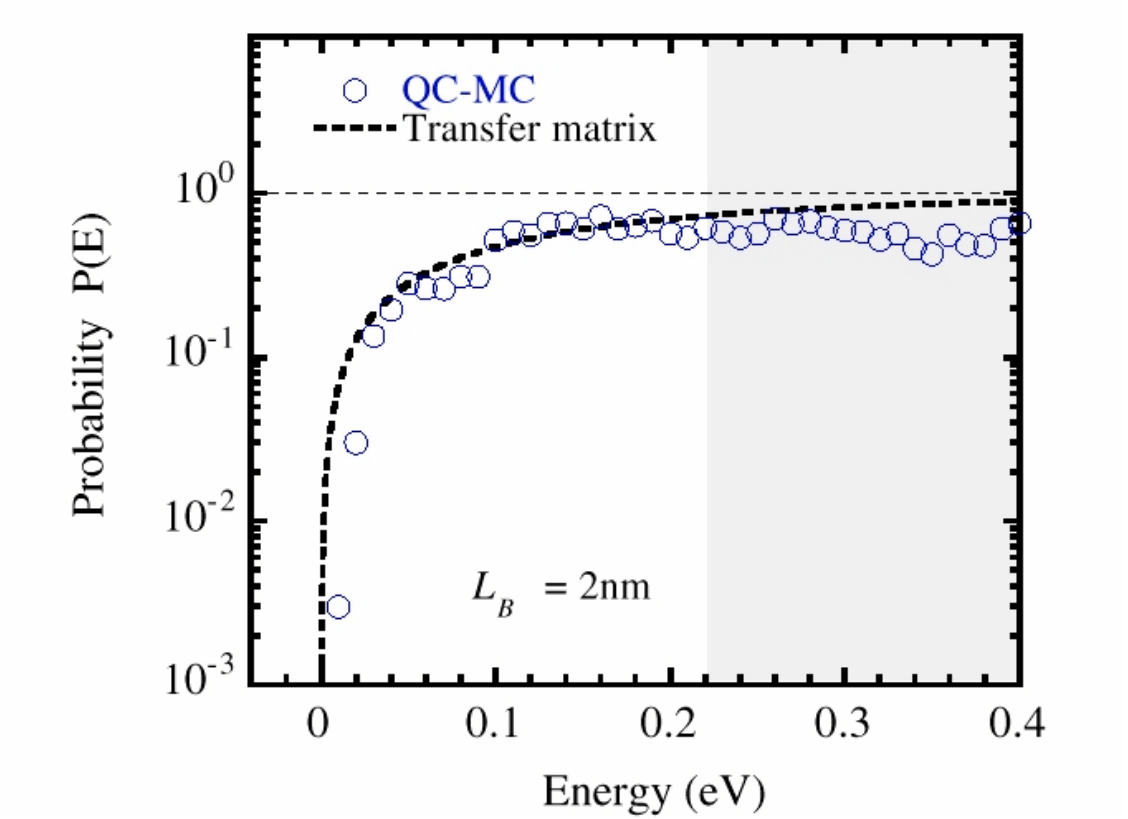
Quantum corrections for Monte Carlo simulation

- **Single GaAs/AlGaAs/GaAs barrier with a fixed potential (2)**

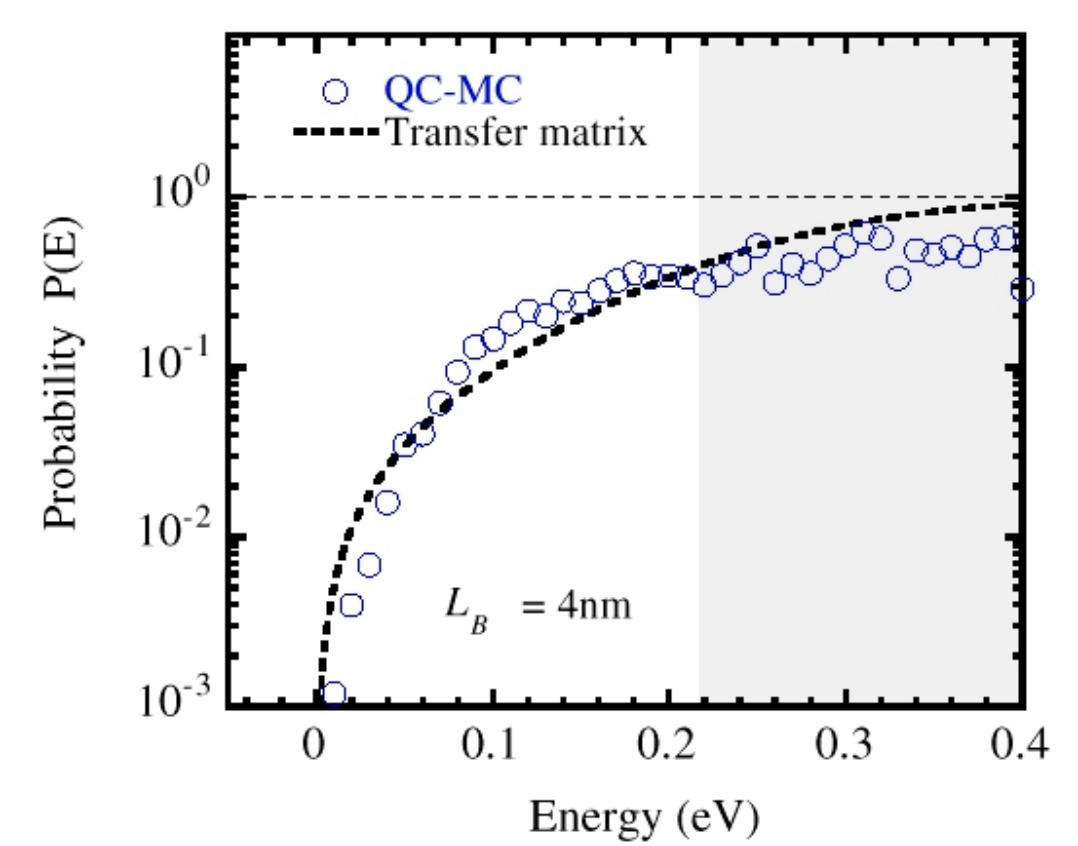




- **Single GaAs/AlGaAs/GaAs barrier with a fixed potential (3)**



- **Single GaAs/AlGaAs/GaAs barrier with a fixed potential (4)**



Wigner Correction

- Approximate quantum series
  - Truncate to the first non-classical term
  - Assume displaced maxwellian
  - Parabolic bands
- Leads to a momentum-dependent correction
- Possible to avoid last two approximations by making use of Monte Carlo full band

Simplified Wigner Correction

- Averaging out the momentum-dependence

$$V \rightarrow V - \boxed{\frac{\hbar^2}{12 m^*} \nabla^2 \ln(n)}$$

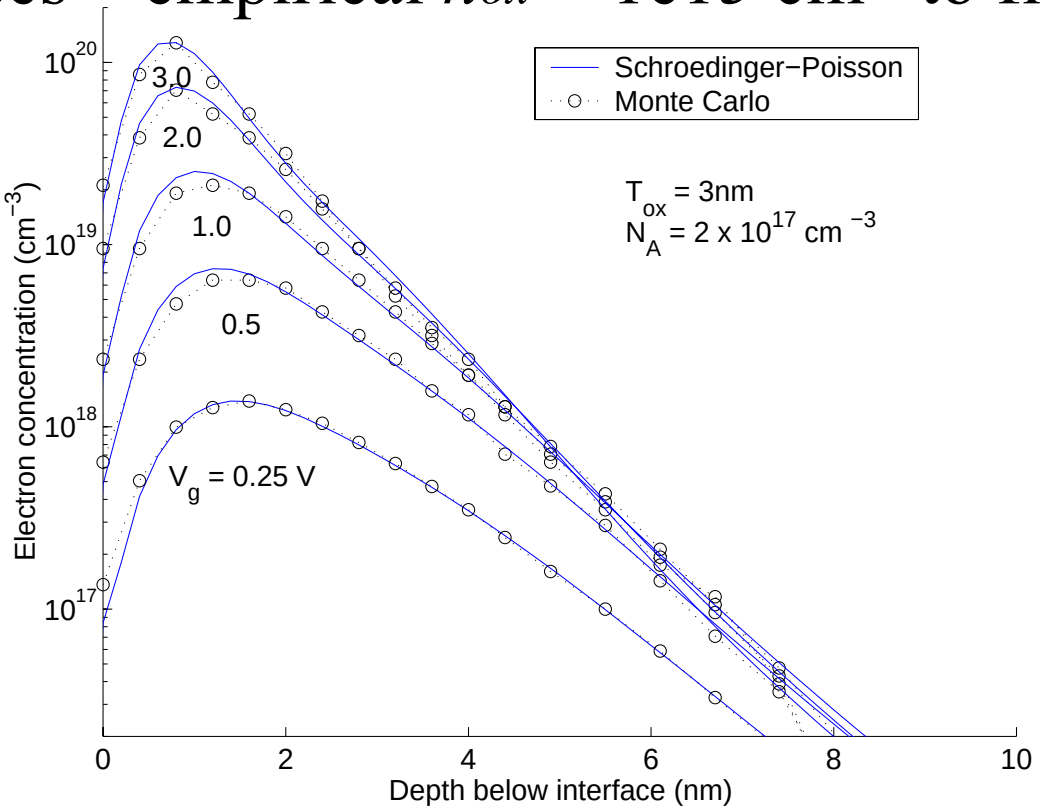
- Feynman showed this can also be derived as an approximation to the effective potential
- Further approximation leads to the density gradient method in drift-diffusion

Properties of Wigner Correction

- Requires long run times due to Monte Carlo noise in $\nabla^2 \ln(n)$ and restricts grid spacing
- Works well in drift-diffusion where noise is not an issue
- Unlike effective potential, $\nabla^2 \ln(n)$ is *local*
- For MOS, a single fitting parameter was found to adjust the correction at the oxide interface
- Requires no fitting in the silicon region

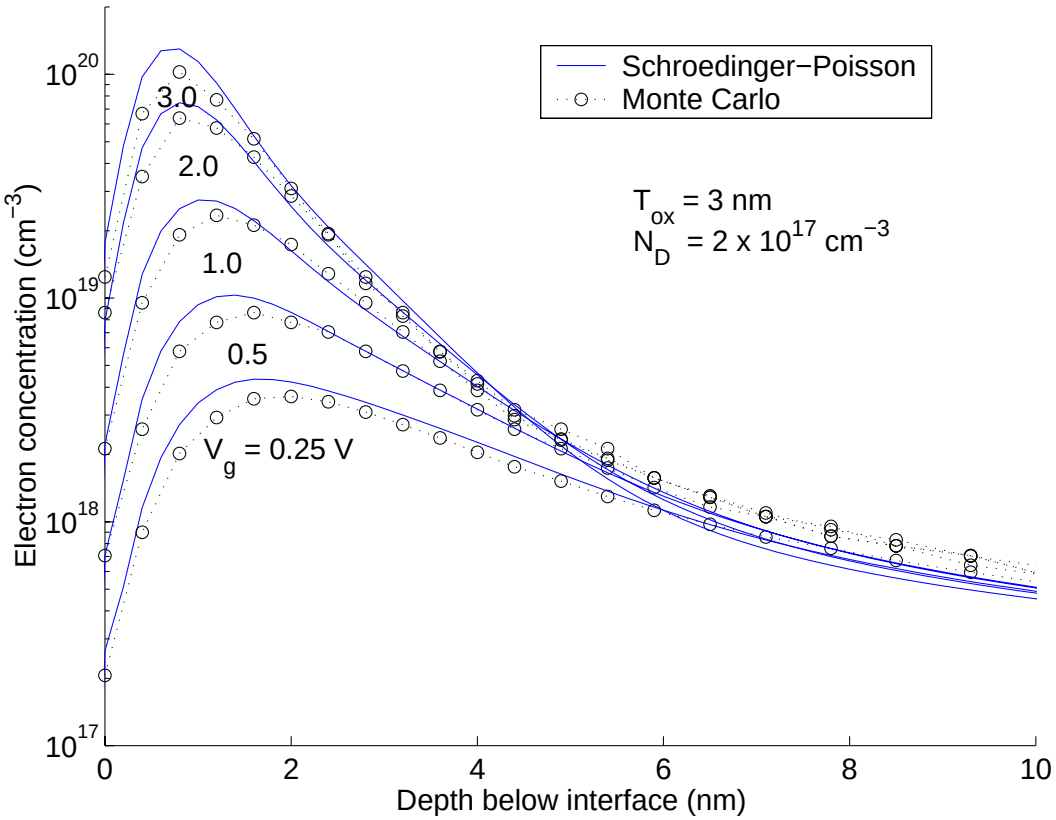
Wigner-corrected MOS Inversion

Wigner correction is accurate across range of biases—empirical n_{ox} = 1e15 cm^{-3} to fit interface



Wigner-corrected MOS Accumulation

$n_{ox} = 1e15\ cm^{-3}$



Schrödinger-Based Correction

- Treats quantum effects in the direction perpendicular to transport
- Accurate
- No fitting parameters
- Not sensitive to noise in the Monte Carlo concentration estimator
- Efficient, additional computation time is small

Applying Schrödinger Correction

- Schrödinger equation is solved along 1D slices of the 2-D domain
- Self-consistent Monte Carlo potential is the input to Schrödinger and quantum density, n_q, is output
- Concentration is linked to the correction with a Boltzmann dependence

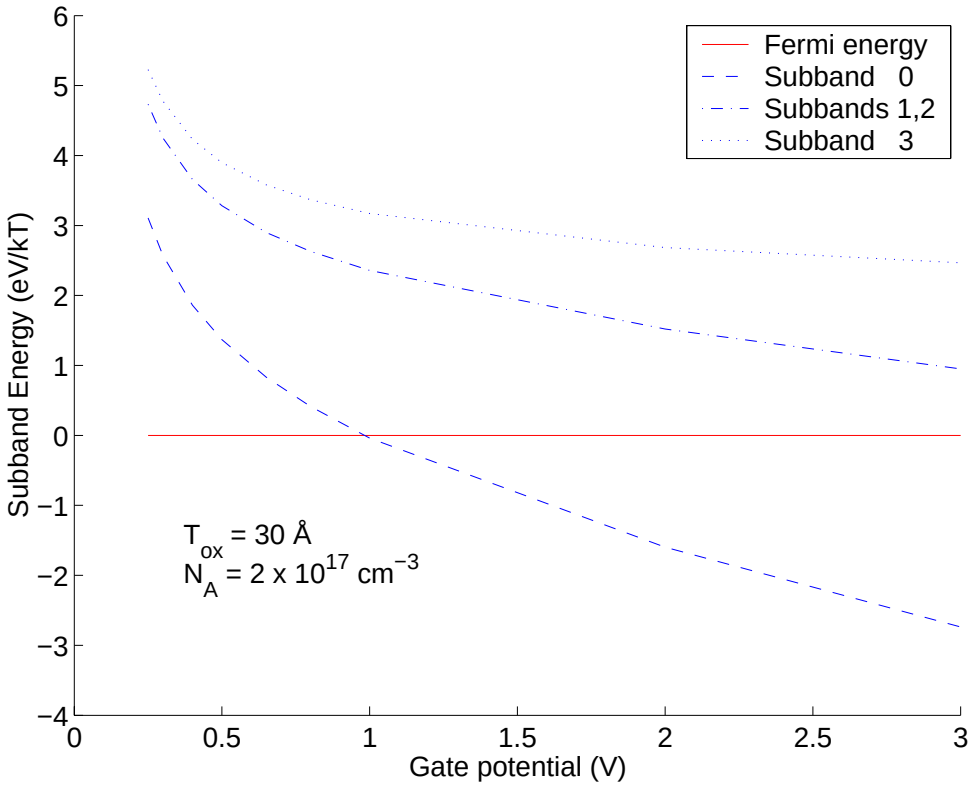
$$n_q(z) \propto e^{-\{V_p(z) + V_{qc}(z)\}/kTt}$$

Consistent with Non-equilibrium Transport

- Schrödinger energy levels/wavefunctions are filled on a Boltzmann distribution
- Within each slice, the correction forces the ***shape*** of the quantum density onto Monte Carlo
- No Fermi level is required
- Relative concentration between the slices is determined naturally by Monte Carlo transport

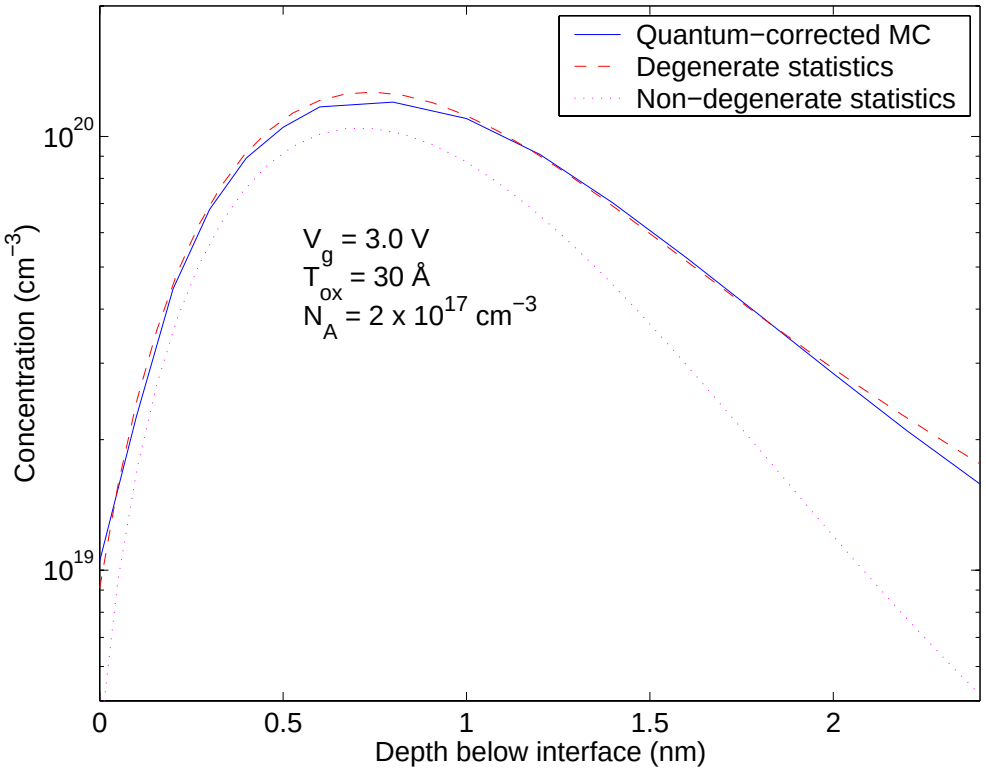
Degenerate Statistics in MOS

Degeneracy can be important in highly inverted MOS structures

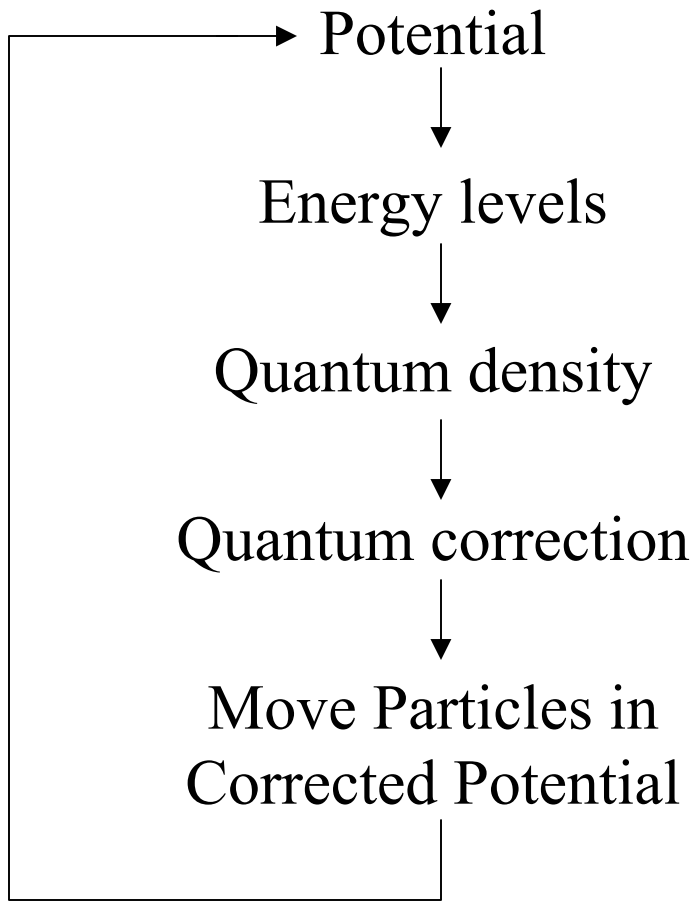


Correction Satisfies Degeneracy

$V_{qc}(z)$ is only a function of the **relative** $n_q(z)$ and the first three subbands have a similar shape



Quantum-corrected Simulation Flow



$$\nabla \cdot \varepsilon \nabla V_p = -\rho$$

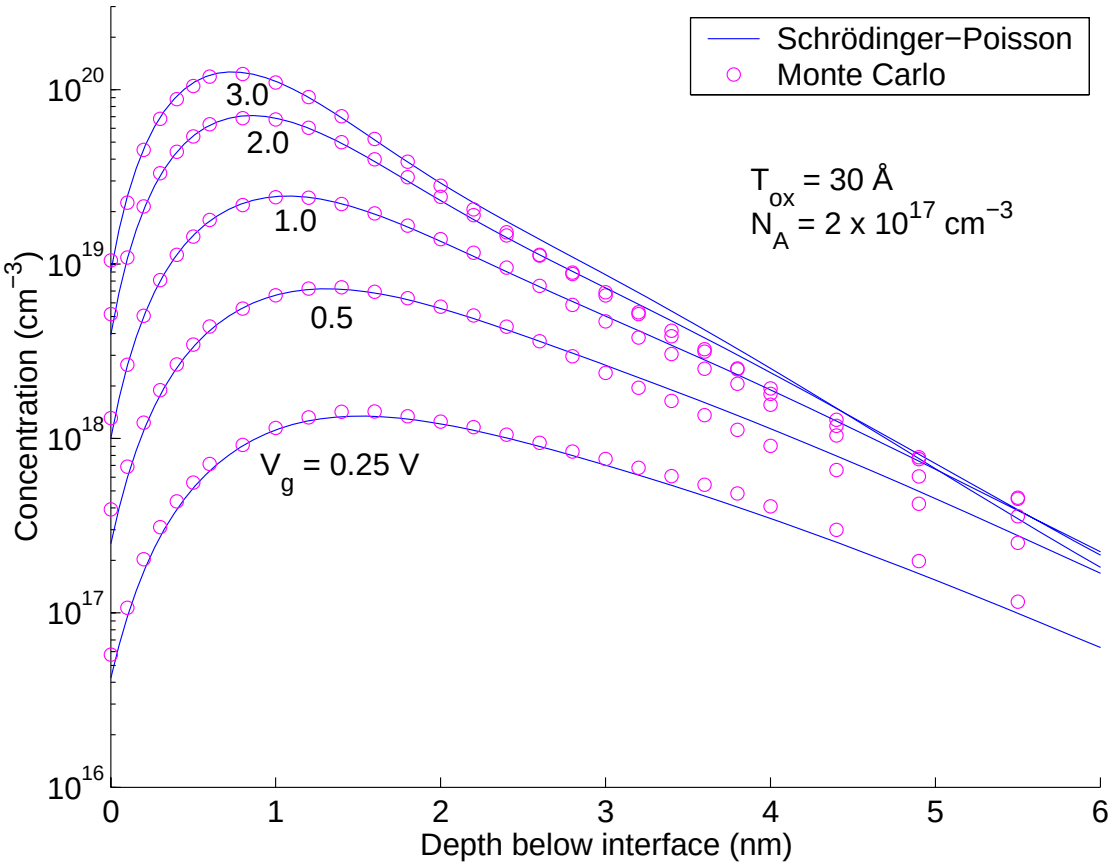
$$-\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m^*}\frac{d\psi}{dz}\right) + V_p\psi = E\psi$$

$$n_q(z) \propto \sum_E \left|\psi(E,z)\right|^2 e^{-E/kT_t}$$

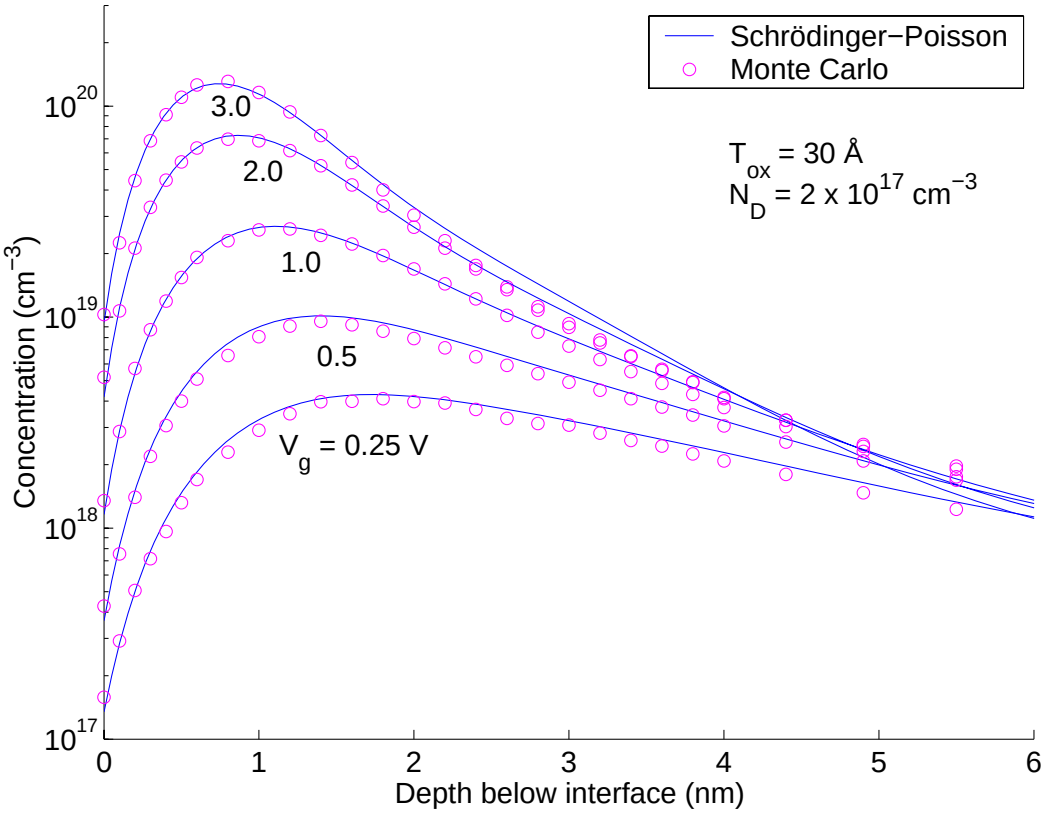
$$-kT_t \log\left(n_q(z)\right) - V_p(z) + V_o$$

$$\frac{d\vec{k}}{dt} = \frac{q}{\hbar}\nabla_{\vec{r}}(V_p + V_{qc})$$

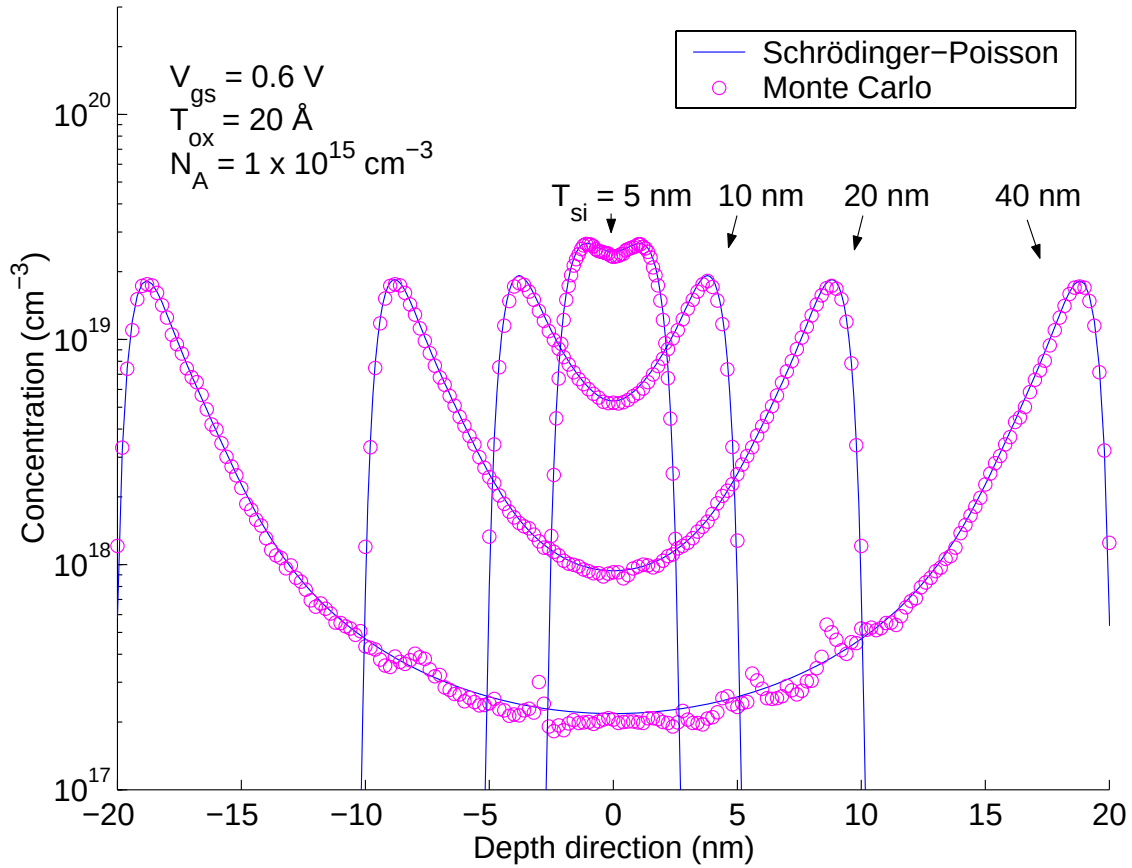
Schrödinger-corrected MOS Inversion



Schrödinger-corrected MOS Accumulation

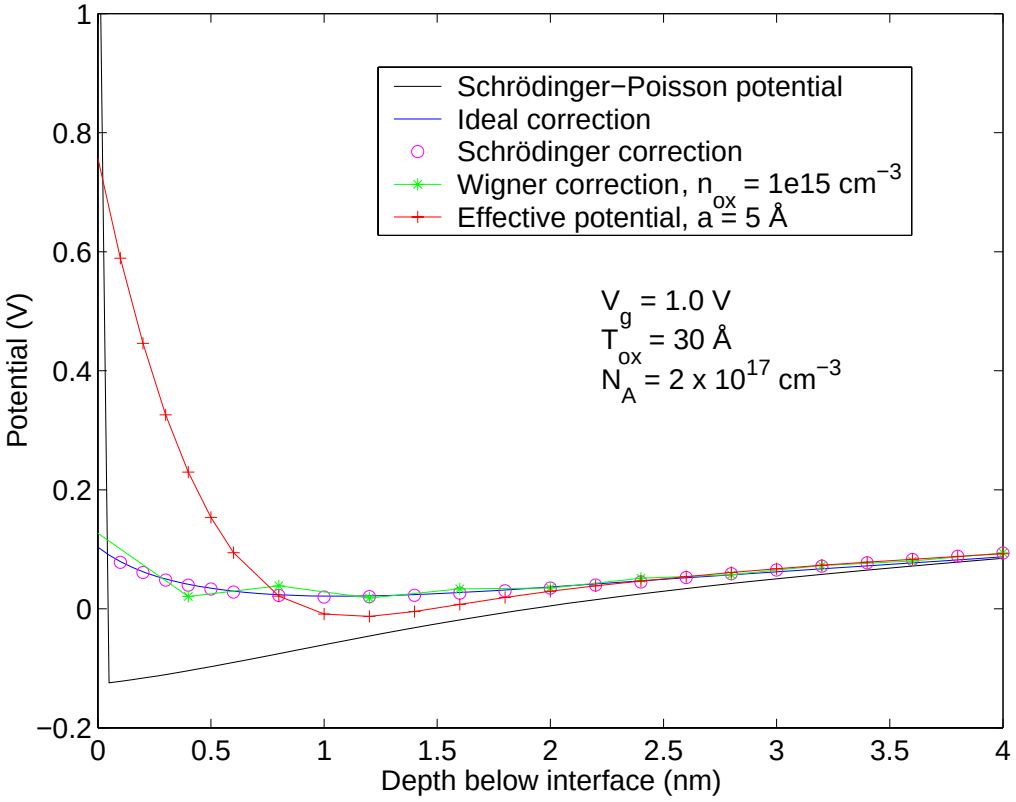


Schrödinger-corrected Double-gate

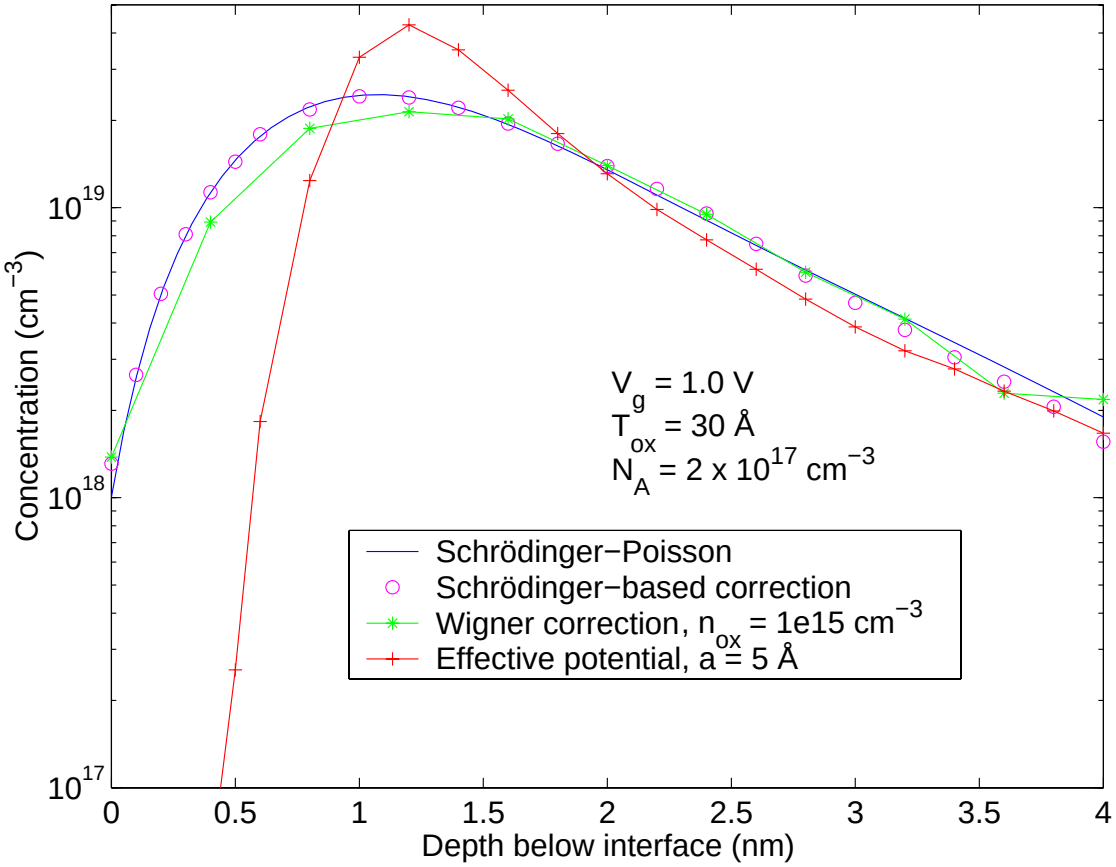


Schrödinger vs Other Corrections for MOS

- Typical behavior of different corrections for MOS



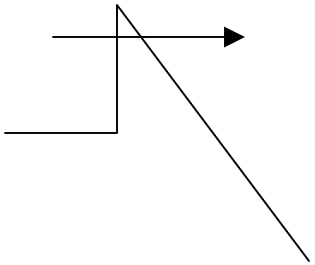
Schrödinger vs Other Corrections for MOS

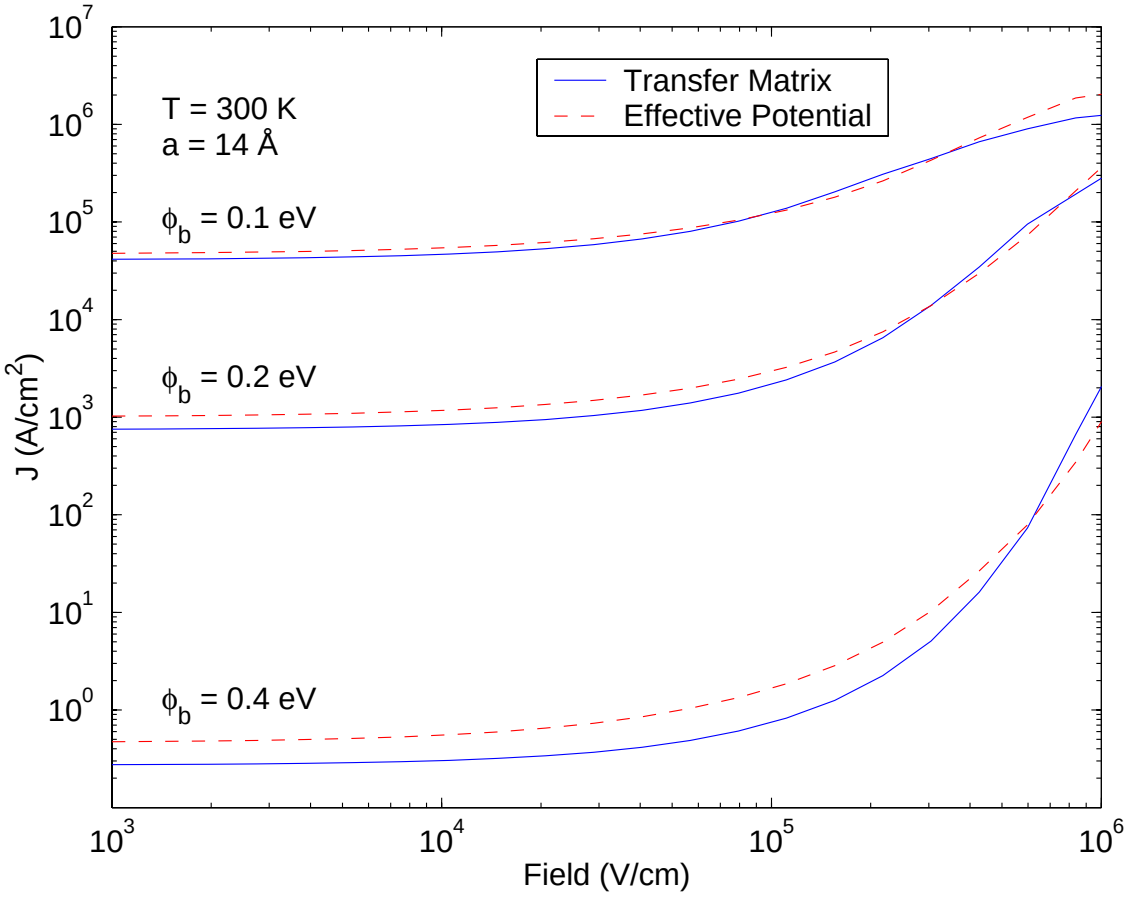


Adding Tunneling to Monte Carlo

- Calculation of transfer matrix tunneling probability for each Monte Carlo particle is accurate
- Implementation of transfer matrix is cumbersome and can be inefficient
- Effective potential is simple and gives reasonable results for current through a small barrier
- Useful in current project to extend MOCA to SiGe or could be applied to source/channel tunneling

Effective Potential for Tunneling



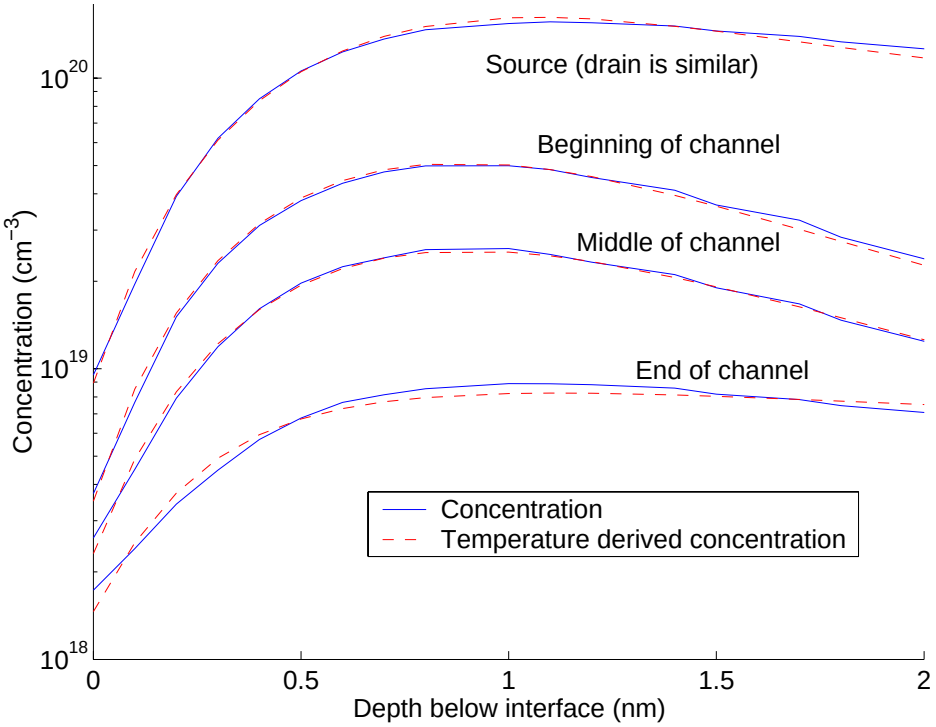


Extending Schrödinger Correction to Devices

- Heating occurs in the direction $\perp$ to transport
- Cannot make use of an electron "temperature" because it is not well-defined for non-equilibrium
- Define a "transverse" temperature T_t to describe the variation of the concentration with potential in the $\perp$ direction
- T_t is validated if a single temperature at each point along the transport path accurately describes the variation

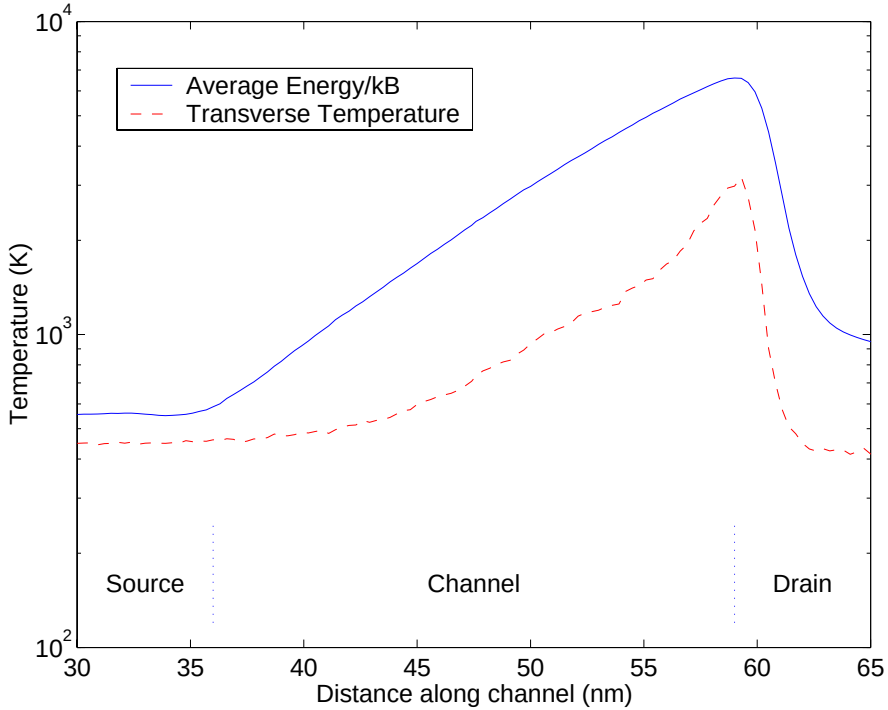
Validating Transverse Temperature

$T_t(z)$ accurately describes potential→concentration for highly non-equilibrium transport in 25nm FET

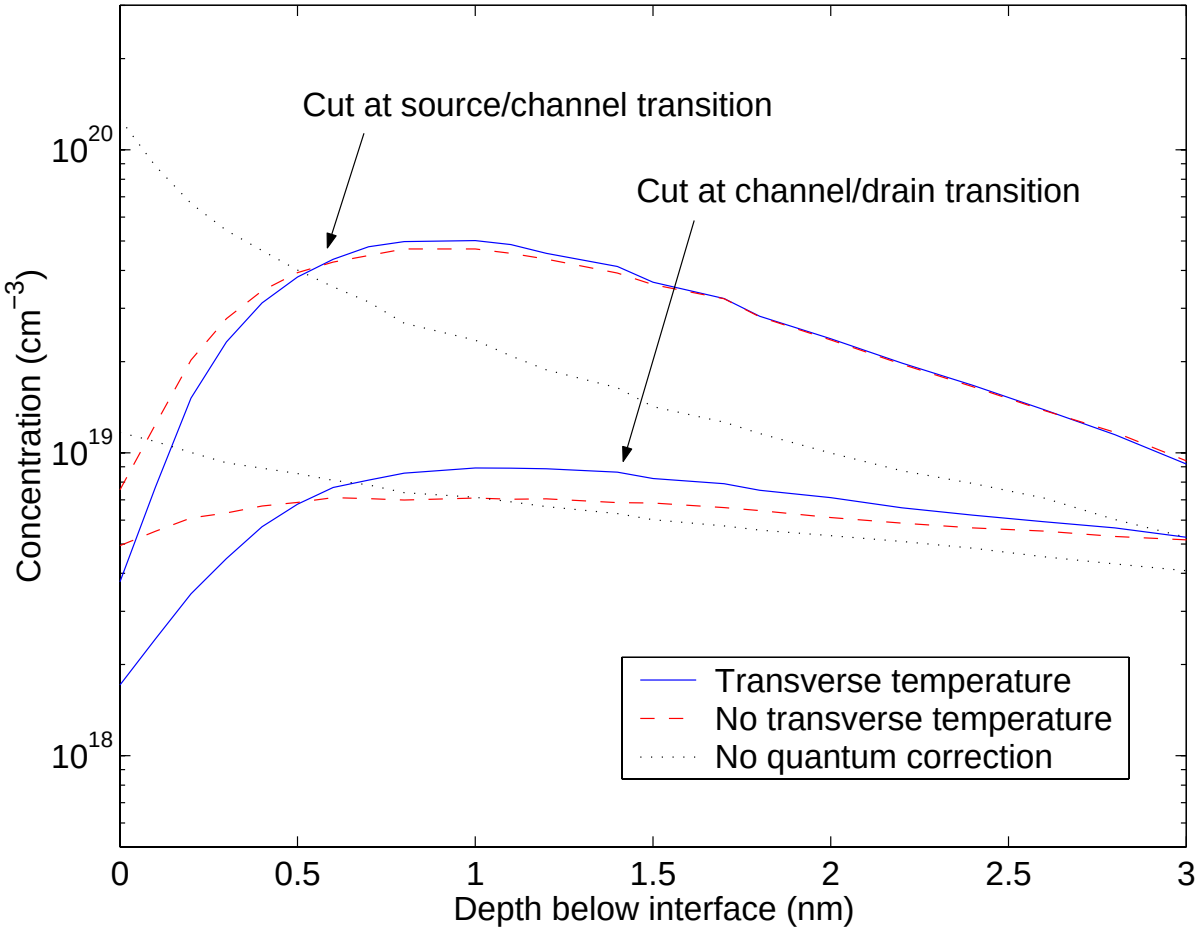


Typical Transverse Temperature

- Transverse temperature for a 25-nm MOSFET in saturation bias

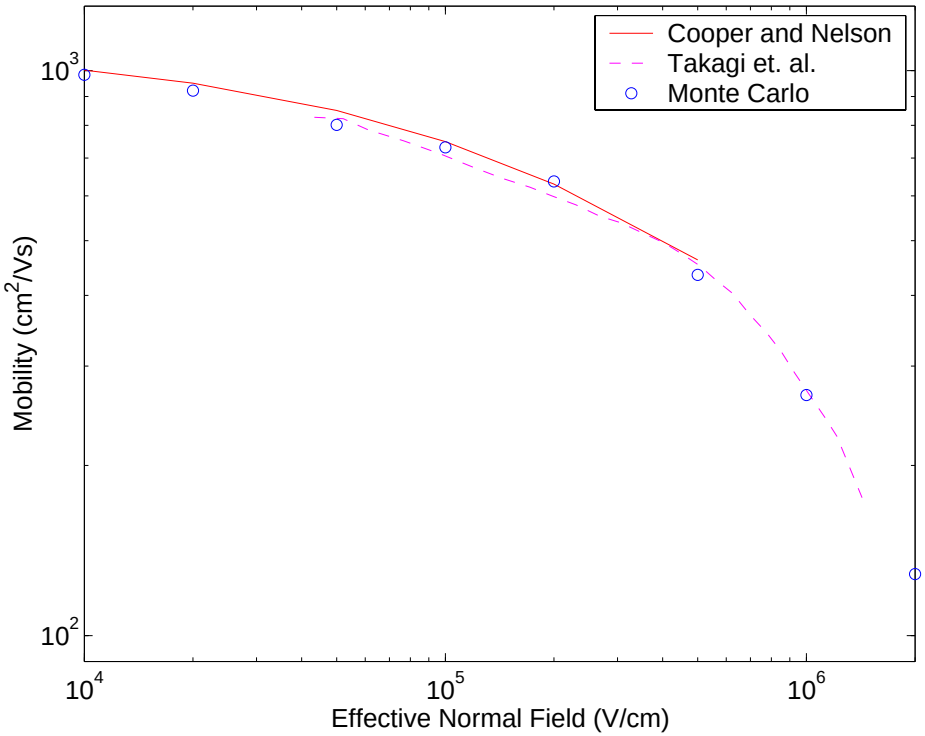


Effect of Heating in the ⊥ Direction

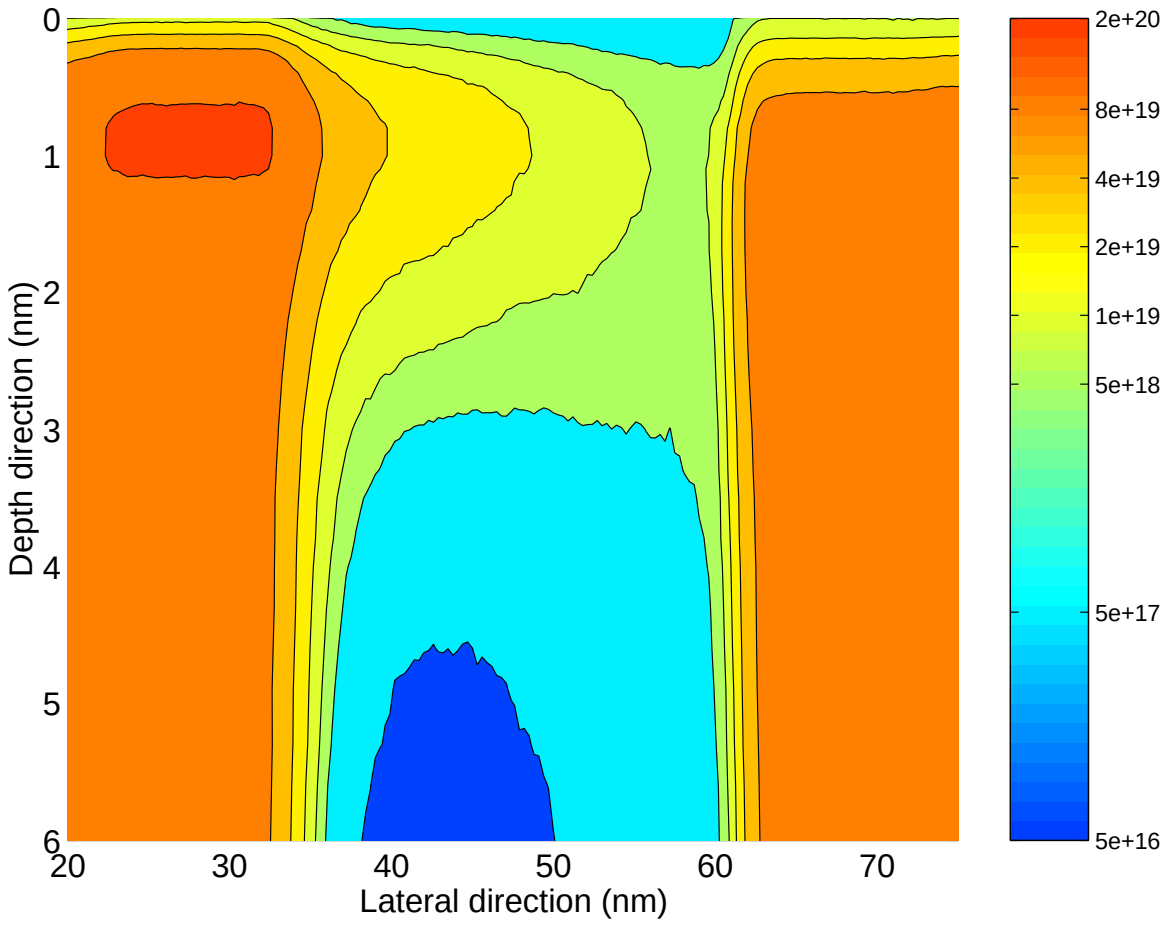


Surface Scattering

Surface scattering model of Yamakawa, using roughness parameters obtained from experiment

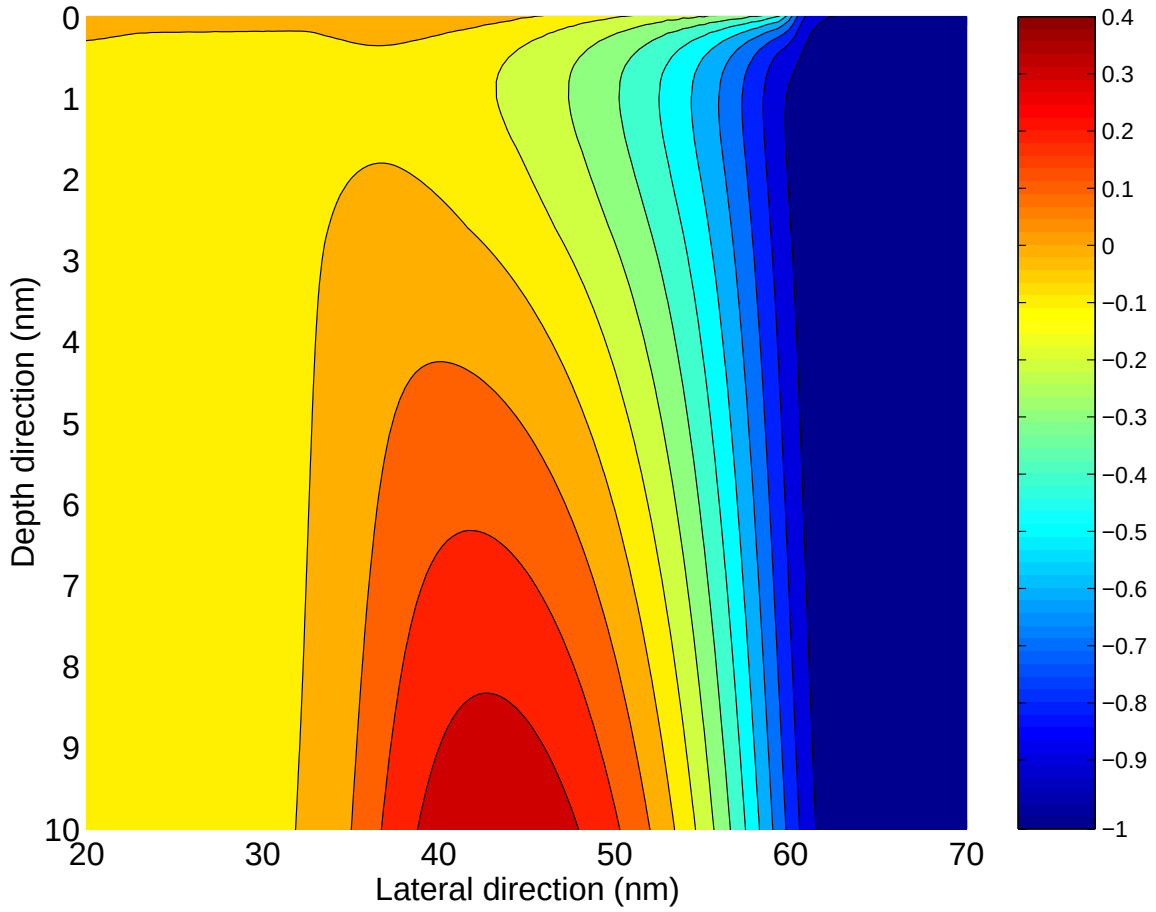


25 nm MOSFET – concentration



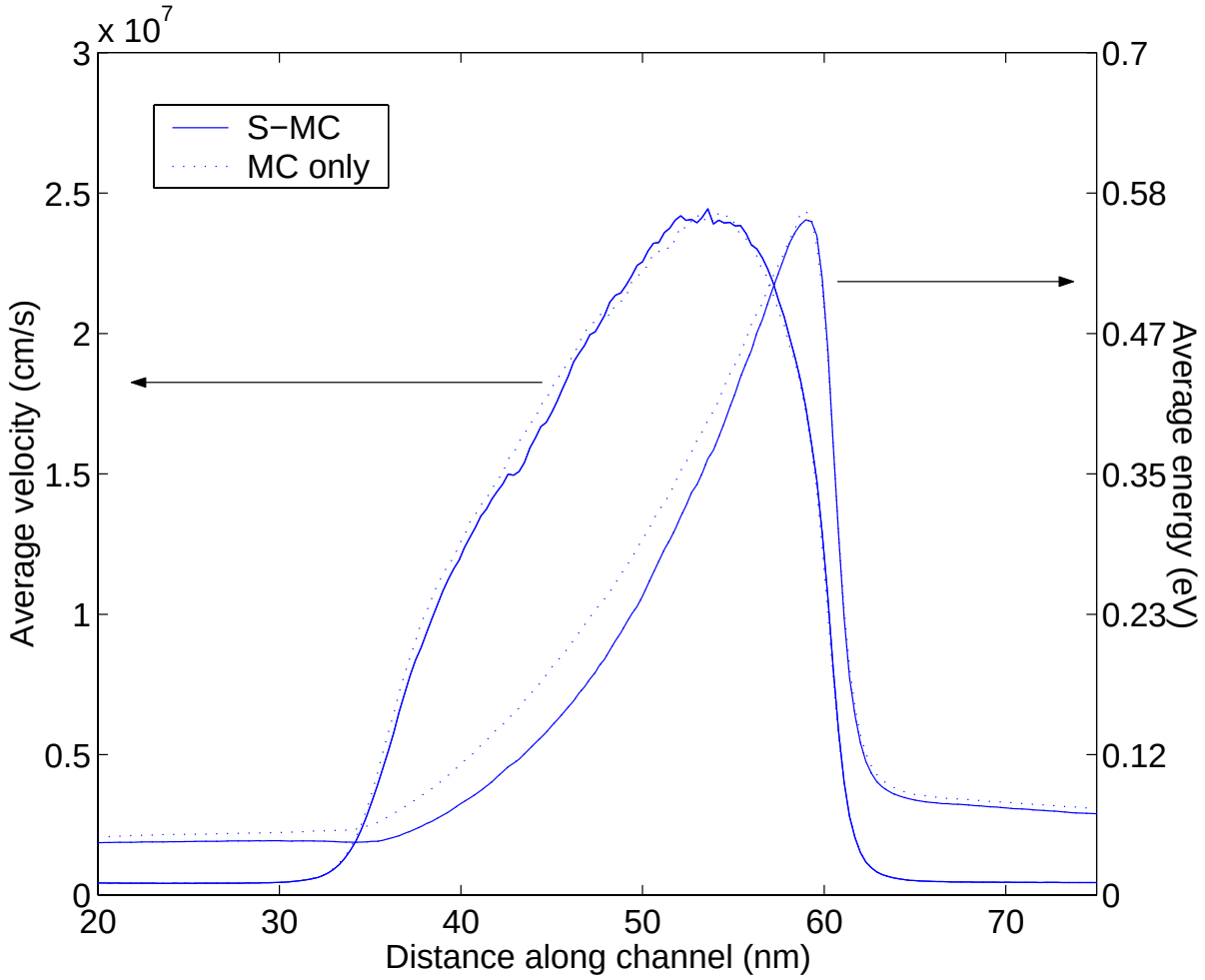
$L_{eff} = 25$ nm
$T_{ox} = 15$ Å
$V_g = 1$ V
$V_d = 1$ V

25 nm MOSFET – corrected potential



$L_{eff} = 25$ nm
$T_{ox} = 15$ Å
$V_g = 1$ V
$V_d = 1$ V

Properties along transport path



Conclusions

- Quantum corrections can be used to extend the validity of Monte Carlo device simulation to the 10-nm regime
- Wigner-based corrections
  - accurate, and momentum-dependent is interesting
  - somewhat impractical due to noise in Monte Carlo
- Effective potential
  - simple and fast
  - accurate for small heterojunctions
- Schrödinger-based correction
  - accurate and efficient with no fitting parameters
  - best choice for size quantization effects