# COMPARISON OF QUANTUM CORRECTIONS FOR MONTE CARLO SIMULATION

Brian Winstead*, Hideaki Tsuchiya**, and Umberto Ravaioli*
*Beckman Institute and the Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana Illinois 61801, USA
**Department of Electrical and Electronics Engineering
Kobe University 1-1, Rokko-dai, Nada-ku, Kobe 657-8501, Japan

**Abstract**

As semiconductor devices are scaled down to nanometer scale dimensions, quantum mechanical effects can become important. For many device simulations at normal temperatures, an efficient quantum correction approach within a semi-classical framework is expected to be a practical way applicable to multi-dimensional simulation of ultrasmall integrated devices. In this paper, we present a comparative study on the three quantum correction methods proposed to operate within the Monte Carlo framework, which are based on Wigner transport equation, path integrals, and Schrödinger equation. Quantitative comparisons for the strengths and weaknesses of these methods are discussed by applying them to size quantization and tunneling effects.

*Index Terms*—ultrasmall MOSFET, Monte Carlo methods, nanotechnology, quantum correction, Schrödinger equation, quantum effects

This work was partially supported by the Semiconductor Research Corporation, contract 99-NJ-726. H. T also thanks for the support by the Ministry of Education, Science, Sports and Culture of Japan, Grand-in-Aid for Encouragement of Young Scientists, 13750061, 2001.

# 1 Introduction

As semiconductor devices are scaled down to nanometer scale dimensions, quantum mechanical effects can become significant, and a full quantum transport model is necessary if coherent effects dominate device behavior. However, for many practical devices, an efficient alternative is to include quantum corrections within a semi-classical framework. If a physically-based model such as Monte Carlo is used, it is easier to include important transport physics than in most available quantum transport approaches. For example, in MOSFETs scaled below 100 nm, bandstructure and scattering mechanisms must still be modeled to a certain degree of sophistication, while coherence effects should only play a secondary role because the potential profiles along the transport path are typically smooth, minimizing quantum mechanical reflections. Instead of coherent transport, the major quantum effects to be concerned about in this case are size quantization and tunneling. Size quantization can be captured with quantum corrections because in the direction perpendicular to the transport, the device is essentially in quasi-equilibrium conditions, and the major issue is to adjust the statistical occupation probabilities. Tunneling occurs in the direction of transport, but for sufficiently wide or high single barriers, the quantum region of action can be assumed to be strongly localized in the neighborhood of the barrier itself.

Quantum corrections can be incorporated into a semi-classical Monte Carlo simulator by introducing a quantum potential term which is superimposed onto the classical electrostatic potential seen by the simulated particles. The essence of the technique is illustrated pictorially for a single tunneling barrier in Fig. 1. Raising a particle's potential energy in a quantum well, or lowering it at the top of a barrier can modify the semi-classical transport, thus reproducing to first-order the average effects of quantization and tunneling on the carrier distribution.

Several quantum correction approaches are possible. These procedures in general entail the self-consistent calculation of a correction potential which is added to the semi-classical solution. Approximate quantum models are used to obtain the corrected potential from the semi-classical potential itself, to steer the transport toward a situation that mimics as much as possible the quantum behavior. The methods proposed to operate within the Monte Carlo framework include methods based on Wigner equation [1], path integrals [2], and Schrödinger equation [3]. The goal of this paper is to review comparatively these three main approaches,

underscoring the strengths and weaknesses of each of them. Quantitative comparisons are presented to help in understanding for which applications one method might be more efficient or appropriate over the others.

# 2 Description of Quantum Corrections

## 2.1 Wigner-based Correction

The Wigner-based quantum correction can be derived starting from a suitable form of the Wigner transport equation [4]

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f - \frac{1}{\hbar} \nabla_r V \cdot \nabla_k f + \sum_{\alpha=1}^{\infty} \frac{(-1)^{\alpha+1}}{\hbar 4^{\alpha} (2\alpha+1)!} (\nabla_r \cdot \nabla_k)^{2\alpha+1} V f = \left( \frac{\partial f}{\partial t} \right)_c \tag{1}$$

Here, $k$ is the crystal momentum, $V$ is the classical potential and the term on the right-hand side represents the effects of collisions. The non-local quantum mechanical effects are represented in the fourth term on the left-hand side of (1). In the limit of slow spatial variations, the non-local terms disappear and we recover the standard Boltzmann Transport equation (BTE). The simplest approach to quantum correction is to start by using only the lowest order term with $\alpha = 1$ in the summation. Following this approximation, one obtains an equation that closely resembles the structure of the BTE, with one additional term providing a quantum correction. This quantum corrected BTE takes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f - \frac{1}{\hbar} \nabla_r V_w \cdot \nabla_k f = \left( \frac{\partial f}{\partial t} \right)_c \tag{2}$$

where the term $V_w$ contains the quantum potential. $V_w$ depends on the distribution function, which in turn can be resolved numerically by Monte Carlo simulation, for equilibrium or non-equilibrium cases. We take here a simpler approach, which assumes a drifted maxwellian distribution function with parabolic dispersion relation. It allows us to represent $V_w$ with an analytical form. Limiting the derivation to one dimension for clarity, $V_w$ becomes [5]

$$V_w(k,n) = V + \frac{k_B T}{24} \left[ \gamma^2 (k - \overline{k})^2 - 3\gamma \right] \frac{\partial^2 \ln(n)}{\partial x^2} \quad , \quad \gamma = \frac{\hbar^2}{m^* k_B T} \tag{3}$$

where $n$ is the carrier concentration. In (3), the corrected potential, $V_w$, depends on both the location and the momentum of the individual particles. A simplified version of $V_w$ can also be derived by assuming in (3)

a thermal equilibrium average energy as [1][5]

$$V_w(n) = V - \frac{\hbar^2}{12m^*}\frac{\partial^2 \ln(n)}{\partial x^2} \tag{4}$$

This simplified, momentum-independent formulation has some advantage over the more complex momentum-dependent version (3) because in addition to use in Monte Carlo, it can be applied for quantum corrections in lower levels of the transport simulation hierarchy such as hydrodynamic [6] or drift-diffusion [7]. We have to add that for a multi-dimensional problem the Wigner-based correction should be represented in terms of a quantun force correction, not a quantum potential correction [5].

## 2.2 Effective Potential Correction

The effective potential approach to quantum correction was developed by Feynman [8]. To derive the effective potential, a variational method can be used to calculate to contribution to the path integral of a particle's quantum fluctuations around its classical path. Using a trial potential to first order in the average point on each path, the effective classical potential becomes

$$V_{eff}(x) = \frac{1}{\sqrt{2\pi}a}\int_{-\infty}^{\infty} V(x_0)e^{-\frac{(x_0-x)^2}{2a^2}}dx_0 \quad , \quad a^2 = \frac{\hbar^2}{12m^*k_BT} \tag{5}$$

Equation (5) represents a smearing of the electrostatic potential on a length scale of the parameter, $a$, which can also be interpreted as the effective quantum mechanical "size" of the particle [2].

Feynman later improved this simple correction using a second-order trial potential [9] which yields the following effective potential, $W_1$

$$\begin{aligned} W_1(x_0) &= \min_{a^2(x_0),\Omega(x_0)}\{\tilde{W}_1(x_0, a^2(x_0), \Omega(x_0))\}, \text{ where} \\ V_{a^2}(x) &\equiv \frac{1}{\sqrt{2\pi}a}\int_{-\infty}^{\infty} V(x_0)e^{-\frac{(x_0-x)^2}{2a^2}}dx_0 \\ \tilde{W}_1(x_0, a^2, \Omega) &= \frac{1}{\beta}\ln\frac{\sinh(\beta\Omega/2)}{\beta\Omega/2} - \frac{\Omega^2}{2}a^2 + V_{a^2}(x_0) \end{aligned} \tag{6}$$

$V_{eff}$ in (5) corresponds to $W_1$ in (6) with the special non-optimal choice of $\Omega \equiv 0$. A typical solution of $W_1$ for MOS quantization effects is indicated in Figure 2. For this application, the benefits of using the

$W_1$ effective potential relative to the simpler $V_{eff}$ with $a$ allowed to vary as a tuning parameter appear to be marginal. For practical application, the primary focus in this work will be on the $V_{eff}$ version of the correction.

## 2.3 Schrödinger-based Correction

In the Schrödinger-based approach for quantum correction, the Schrödinger equation is solved periodically in a simulation using the self-consistent electrostatic potential as input. In contrast to the Wigner-based and effective potential corrections, the quantum potential in this method is calculated from the exact energy levels and wavefunctions corresponding to the electrostatic potential solution. The first step in the procedure is to calculate the overall shape of the quantum density by filling the energy levels according to an equilibrium Maxwell-Boltzmann distribution. This quantum density shape is mapped to a quantum potential through

$$V_{schr}(z) = -kT\log(n_q(z)) - V_p(z) + V_0 \tag{7}$$

Here, $V_{schr}$ is the quantum correction, $z$ is the direction normal to the interface, $n_q$ is quantum density from the Schrödinger equation or equivalently the converged Monte Carlo concentration, $V_p$ is the potential from the Poisson solution, and $V_0$ is an arbitrary reference potential determined by the knowledge that the correction should go to zero away from the quantum region, where the behavior is semi-classical. Only the shape of the quantum density is used, therefore, one does not need to invoke the exact Fermi level in the calculation. In this way the correction can be adapted to treat nonequilibrium device simulation [3].

The quantum-corrected potentials, $V_w$, $V_{eff}$, or $V_{schr}$, differ in their method of calculation and their underlying assumptions. However, they are all incorporated into a Monte Carlo simulation in a similar way. As a Monte Carlo simulation evolves in time, the corrections are recalculated along with the Poisson equation to maintain self-consistency. The quantum-corrected potential is then used instead of the electrostatic potential to calculate the forces on the Monte Carlo particles. Other than this modification of the classical forces applied to the particles, the quantum-corrected Monte Carlo simulation can be carried out in the same manner as the uncorrected case.

# 3 Quantization effects

To study quantization effects, the models described in the preceding section were implemented in the 2-D full-band Monte Carlo simulator, MOCA [10]. Because of its technological importance as a building block for devices, the MOS capacitor was used as a prototype structure for this comparative study. For verification, the quantum mechanical charge density and potential were also calculated using self-consistent Schrödinger/Poisson simulation.

Figure 2 illustrates the typical behavior of the quantum potential for the different methods. Here, the "ideal" quantum potential is the correction which would exactly reproduce the quantum density from the Schrödinger-Poisson solution. The results indicate that the Schrödinger-based correction provides the most accurate model, which closely matches the ideal value with no fitting parameters. This is expected because the approaches makes use of a complete solution of the Schrödinger equation instead of an approximate quantum solution. In addition, since there are no fitting parameters, the accuracy of the method is not sensitive to variations in the physical parameters of the MOS capacitor. Figures 3 and 4 compare the detailed solutions for concentration obtained from a full quantum calculation and from a Schrödinger-corrected Monte Carlo simulation, over a wide range of gate biases and for substrate dopings of $N_A = 2 \times 10^{17} cm^{-3}$ and $N_D = 2 \times 10^{17} cm^{-3}$.

The Wigner-based quantum potential is also found to be accurate for quantization effects in the MOS capacitor, if a fitting parameter is used for the density at the interface. Results obtained using the Wigner correction method are also shown in Figs. 3 and 4. The fitting parameter used here is an empirical charge layer of 1 x $10^{15}$ $cm^{-3}$ which is included in the oxide region for the calculation of the correction at the interface point. Beside this necessary adjustment at the interface, the quantum correction (4) is applied with no additional fitting parameters. This scheme allows for the proper adjustment of the interface density for a wide range of biases and doping while giving a reasonably accurate quantum density elsewhere.

For the Feynman effective potential given by (5), the "size" parameter, $a$ was treated as an empirical fitting parameter, as suggested by Ref. [2]. The best fit value for the size parameter in the MOS structures studied here was found to be $a = 4.5$ Å. The effective potential method is accurate in reproducing integrated quantities. Figure 5 shows the total sheet charge for a Monte Carlo simulation of the MOS capacitor with

the effective potential correction, and Fig. 6 shows, for the same simulation, the average displacement of the carriers from the $Si/SiO_2$ interface, which is indicative of the quantum repulsion. However, if the detailed spatial behavior of the effective potential correction is analyzed, one can see significant deviations from the quantum solution. Figure 7 shows the detailed concentration under the gate of the MOS capacitor. Typically, the correction is very large at the interface, leading to a layer of width $\sim a$ next to the oxide interface, where the concentration is significantly lower than that is expected by the quantum solution. Compensating for that, the correction becomes smaller than the expected one at the deeper location inside the substrate, leading to typically a larger peak concentration than the quantum solution.

It can be shown theoretically that the momentum-independent Wigner-based method (4) is an approximation to the effective potential [2]. However, the simulation results presented here indicate that the momentum-independent Wigner-based correction gives a solution which is substantially closer to the detailed quantum behavior. This is due to the fact that the Wigner correction is local, while the effective potential correction is non-local. Neither correction is strictly valid at a heterojunction. However, a single parameter can be used to fit the singularity at the interface for the Wigner correction, since it is local. The silicon region in which the transport actually occurs has a more slowly-varying potential than in the neighborhood of the interface, and thus no fitting is necessary. The application of a non-local effective potential act differently in the overall correction schemes. The adjustment of a fitting parameter to accommodate the strong influence of the interface on the overall solution requires a compensation in the silicon region where the solution has to deviate to maintain the averages.

In addition to accuracy, another important consideration in practical Monte Carlo device simulation is the execution time. For all three methods, the CPU time required to calculate the corrections is negligible relative to the overall Monte Carlo simulation time. However, there is an important difference, in the fact that the Schrödinger-based correction and the effective potential correction are calculated using the electrostatic potential as input, while the Wigner-based correction is calculated from concentration. The noise in the Monte Carlo concentration estimator is always higher than for the potential, and a Wigner-corrected Monte Carlo simulation can take significantly longer time to converge than an uncorrected semi-classical Monte Carlo simulation. In contrast, adding the Schrödinger-based or the effective potential correction to the Monte

Carlo procedure does not increase total cpu time in a very significant way. A self-consistent simulation with the full-band MOCA code using 30000 particles and a non-uniform grid of 300 x 200 nodes for the Poisson equation requires approximately 80 MB of RAM. On a standard 800 Mhz Intel processor, approximately 1000 iterations per hour are executed, where one iteration normally corresponds to a time step of 1 fs or less.

# 4 Tunneling effects

For the purpose of studying quantum corrections in the context of tunneling, the Wigner-based correction and the effective potential were implemented into a 1D GaAs/AlGaAs Monte Carlo simulator. For this case, the Schrödinger correction was not applied, since it is best suited for capturing quantum confinement effects. The tunneling test structure consists of a 4-nm wide GaAs/AlGaAs single barrier with a conduction band discontinuity of 0.22 eV and a temperature of 300 K.

As shown previously, the effective potential correction encounters difficulties in the neighborhood of the abrupt transition between oxide and silicon in the MOS system, since there is very large energy jump of about 3.1 eV and the underlying assumptions behind the theory tend to break down. The problem should not be as severe in the presence of smaller barriers, as is the case for the GaAs/AlGaAs system and the effective potential should be a very good candidate for practical inclusion of tunneling effects.

In applying the Wigner-based correction to MOS quantization, the difficulties near the large barrier were overcome by tuning the correction at the interface point. However, for tunneling it is necessary to model transport on both sides of the interface, and this scheme, that is based on assuming a concentration layer, cannot be used. Instead, for tunneling simulations we apply the theoretical value of the Wigner correction. In order to increase the accuracy, here we implement the momentum-dependent method (3) in addition to the momentum-independent method (4) used in the quantization simulations.

For the tunneling simulation the bias was varied from 0 to 0.3 V, and the GaAs effective mass of $m^* = 0.067m_0$ was used in all three corrections. From (5), this corresponds to a value of 1.9 nm for $a$ in the effective potential. To benchmark the results, the quantum tunneling current was also calculated using a transfer matrix method [11]. Figure 8 plots the resulting current from the transfer matrix and Monte Carlo methods.

All of the quantum corrected results improve significantly upon the classical simulation. The momentum-dependent Wigner correction and the effective potential are the more accurate methods. However the details of their results differ, which is expected because each method stems from a different set of assumptions. The momentum-independent Wigner method is less accurate, which is consistent with the fact that it can be considered an approximation to either the momentum-dependent method or to the effective potential. These same trends also hold for other small tunneling barriers, such as possible source-drain tunneling in highly scaled MOSFETs. If an improved accuracy is desired, the $a$ parameter in the effective potential, or equivalently, the $m^*$ parameter in the Wigner-based corrections can be adjusted for a best fit. However, in such a case, recalibration of the fitting parameters may be required when different barriers are considered.

# 5 Conclusions

Three methods for introducing quantum corrections in semi-classical Monte Carlo simulation have been studied and compared. For the size-quantization case in the MOS system, the Schrödinger-based correction has some intrinsic advantage, since this method does not require fitting parameters, it is accurate, and it adds only negligible CPU time to a Monte Carlo simulation. In contrast, while the Wigner-based method can be tuned to be as accurate, it is in general slower and it requires a fitting parameter. The effective potential method reproduces reasonably well integrated quantities related to size quantization, but it is spatially inaccurate even if the fitting parameter is optimized. For the tunneling case, the Schrödinger-based correction is not appropriate. Instead, the momentum-dependent Wigner correction or effective potential methods can be used with similar accuracy. In this case, the effective potential should have a computational advantage, since the Wigner correction is more affected by noise. On the other hand, the Wigner formulation can still be useful for detailed physical investigations, since it can be extended to include momentum-dependent distributions. One could introduce an analytical distribution function or even a numerical one evaluated directly with the semi-classical Monte Carlo procedure.

## References


[1] H. Tsuchiya and T. Miyoshi, "Quantum transport modeling of ultrasmall semiconductor devices," *IEICE Trans. Electron.*, vol. E82-C, no. 6, pp. 880-887, 1999.

[2] D. K. Ferry, "Effective potentials and the onset of quantization in ultrasmall MOSFETs," *Superlattices and Microstructures*, vol. 28, no. 5/6, pp. 419-423, 2000.

[3] B. Winstead, U. Ravaioli, "A coupled Schrödinger/Monte Carlo Technique for quantum-corrected device simulation" *Device Research Conference*, Notre Dame, Indiana, June 2001.

[4] E. Wigner, "On the quantum correction for thermodynamic equilibrium" *Phys. Rev.*, vol. 40, pp. 749-759, 1932.

[5] H. Tsuchiya and U. Ravaioli, "Particle Monte Carlo simulation of quantum phenomena in semiconductor nanostructures" *J. Appl. Phys.*, vol. 89, pp. 4023-4029, 2001.

[6] J. R. Zhou and D. K. Ferry, "Modeling of quantum effects in ultrasmall HEMT devices" *IEEE Trans. Electron Devices* vol. 40, pp.421-427, 1993.

[7] M. Ancona and H. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor" *Phys. Rev. B* vol. 39, pp. 9536-9539, 1989

[8] R. P. Feynman, A. R. Hibbs, *Quantum Mechanics and Path Integrals,* New York: McGraw-Hill, 1965.

[9] R. Feynman and H. Kleinert, "Effective Classical Partition Functions" *Phys. Rev. A*, vol. 34, no. 6, pp. 5080-5084, 1986.

[10] A. Duncan, U. Ravaioli, and J. Jakumeit, "Full-band Monte Carlo investigation of hot carrier trends in the scaling of metal-oxide-semiconductor field-effect transistors" *IEEE Trans. Electron Devices* vol. 45, pp.867-876, 1998.

[11] K. Brennan and C. Summers, "Theory of resonant tunneling in a variably spaced multiquantum well structure: An Airy function approach" *J. Appl. Phys.* vol. 61, pp. 614-623, 1987

## Figure Captions

Fig. 1 - Illustration of how quantum effects are treated by adding a "quantum potential" to the electrostatic potential.

Fig. 2 - Typical behavior of the quantum potential for an MOS capacitor using several different quantum correction approaches.

Fig. 3 - Electron concentration distributions in an inverted MOS capacitor from two different quantum-corrected Monte Carlo and self-consistent Schrödinger-Poisson methods over a range of gate bias.

Fig. 4 - Electron concentration distributions in an accumulated MOS capacitor from two different quantum-corrected Monte Carlo and self-consistent Schrödinger-Poisson methods over a range of gate bias.

Fig. 5 - Sheet charge density in a MOS capacitor as a function of gate bias calculated with effective potential corrected Monte Carlo and self-consistent Schrödinger-Poisson methods.

Fig. 6 - Location of charge centroid in a MOS capacitor as a function of gate bias calculated with effective potential corrected Monte Carlo and self-consistent Schrödinger-Poisson methods.

Fig. 7 - Electron concentration distributions in an inverted MOS capacitor calculated with effective potential Monte Carlo and self-consistent Schrödinger-Poisson methods.

Fig. 8 - Tunneling current in a 4-nm GaAs-AlGaAs tunneling barrier over a range of bias calculated with three quantum-corrected Monte Carlo and transfer matrix methods.

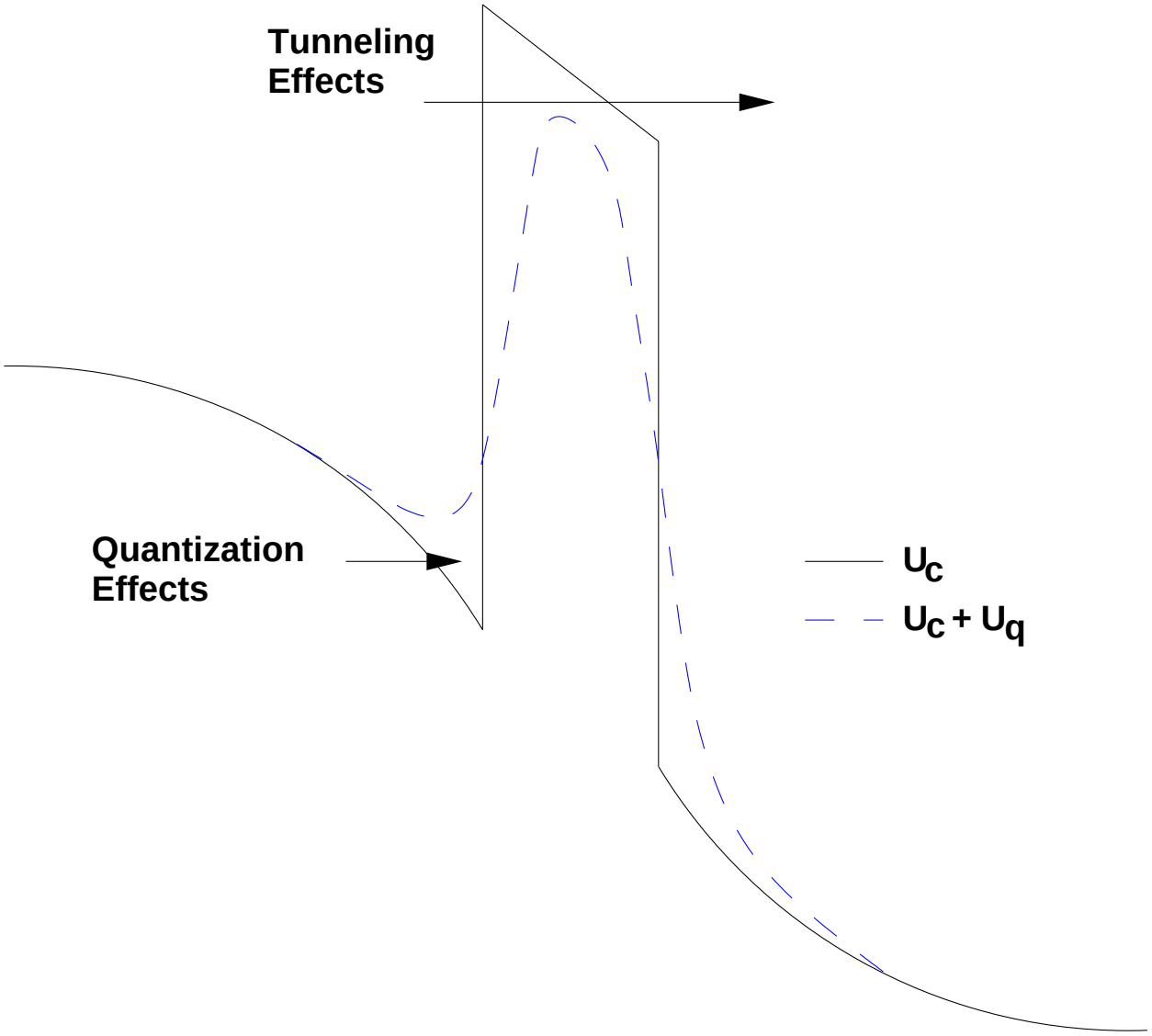


Figure 1.

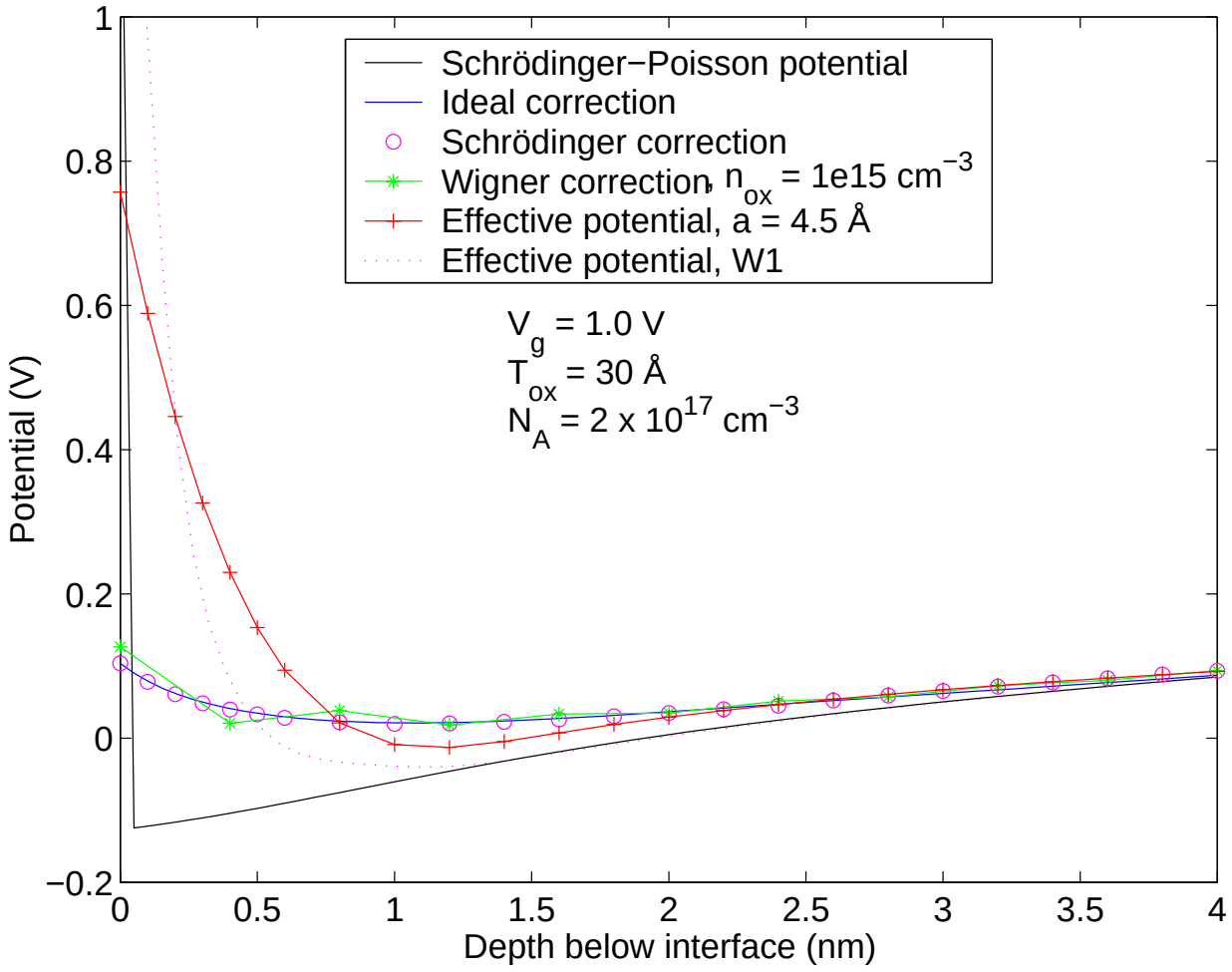
Schrödinger−Poisson potential
Ideal correction
Schrödinger correction
Wigner correction, n_ox = 1e15 cm^−3
Effective potential, a = 4.5 Å
Effective potential, W1
V_g = 1.0 V
T_ox = 30 Å
N_A = 2 x 10^17 cm^−3
Potential (V)
Depth below interface (nm)
1
0.8
0.6
0.4
0.2
0
−0.2
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 2.

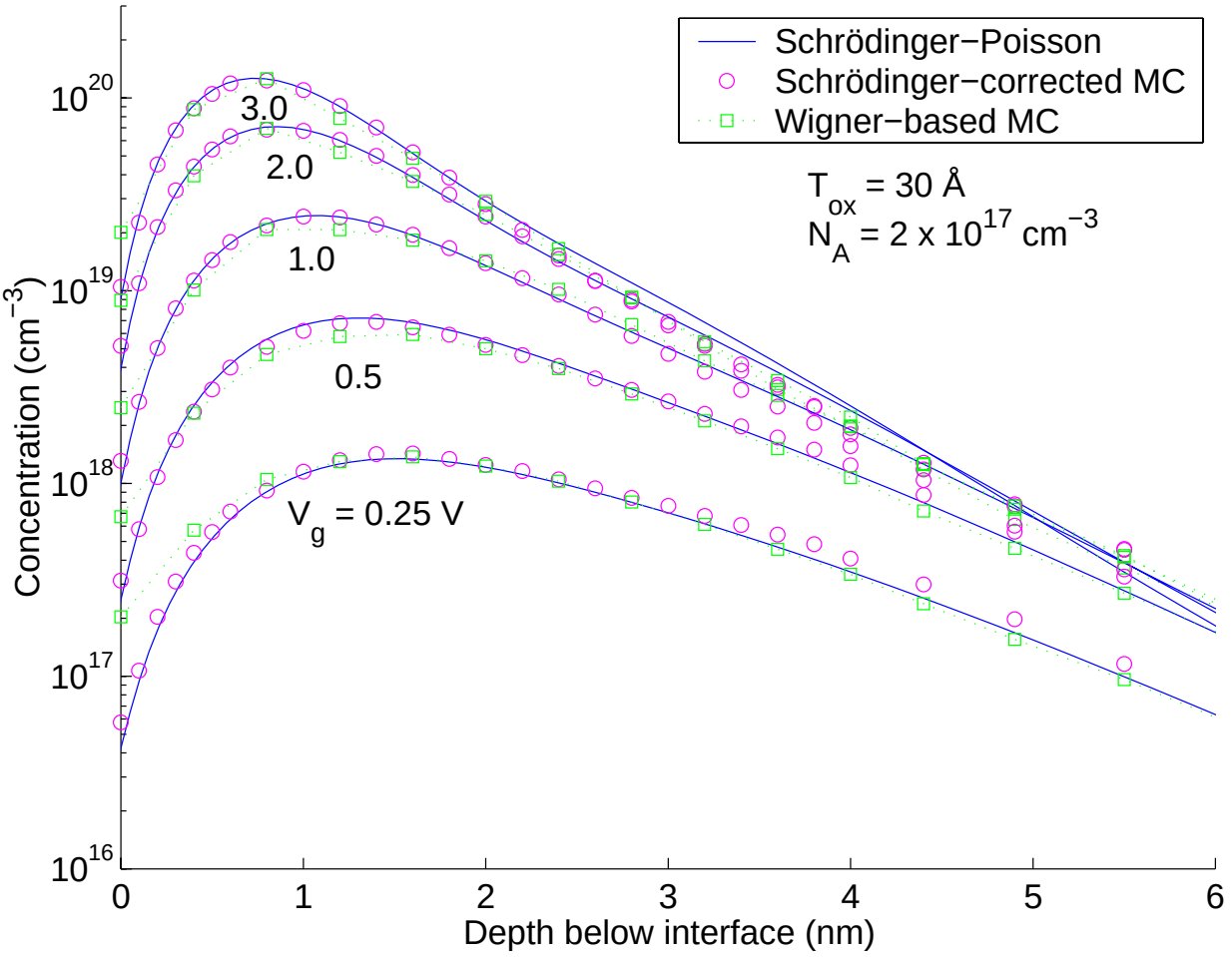
Schrödinger−Poisson
Schrödinger−corrected MC
Wigner−based MC
T_ox = 30 Å
N_A = 2 x 10^17 cm^−3
3.0
2.0
1.0
0.5
V_g = 0.25 V
Concentration (cm^−3)
Depth below interface (nm)
10^20
10^19
10^18
10^17
10^16
0
1
2
3
4
5
6

Figure 3.

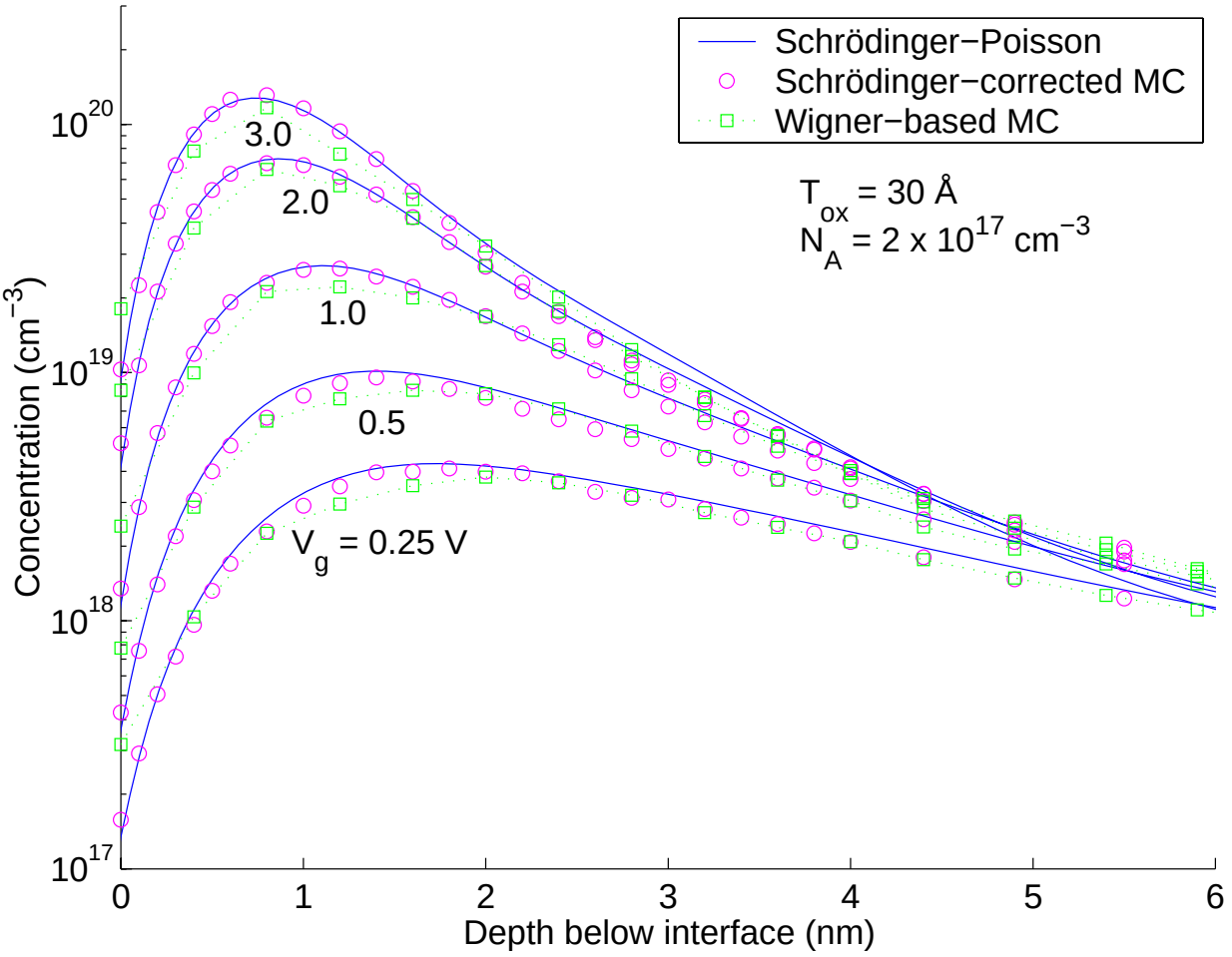
Schrödinger−Poisson
Schrödinger−corrected MC
Wigner−based MC
T_ox = 30 Å
N_A = 2 x 10^17 cm^−3
3.0
2.0
1.0
0.5
V_g = 0.25 V
Concentration (cm^−3)
Depth below interface (nm)
10^20
10^19
10^18
10^17
0
1
2
3
4
5
6

Figure 4.

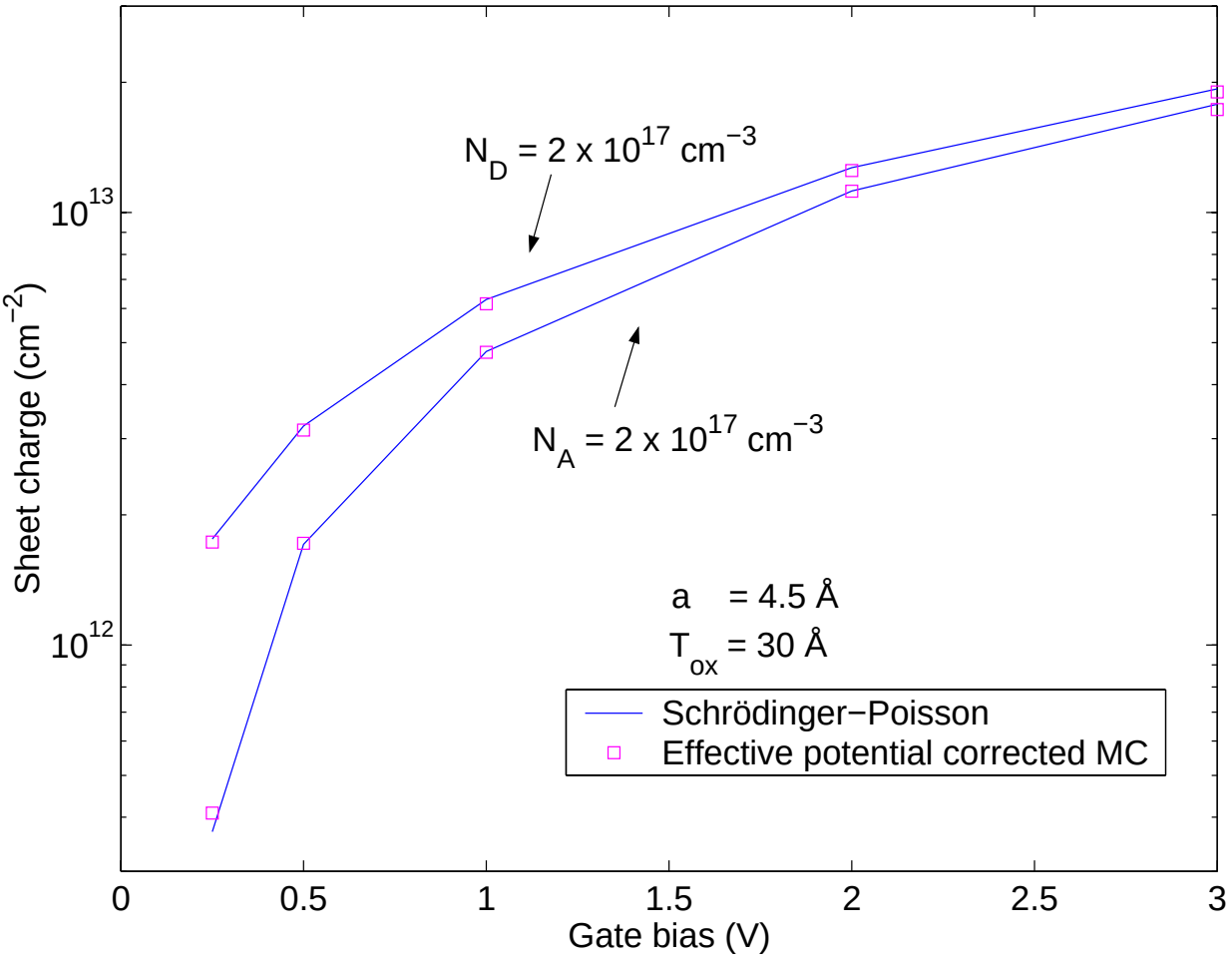
N_D = 2 x 10^17 cm^−3
N_A = 2 x 10^17 cm^−3
a = 4.5 Å
T_ox = 30 Å
Schrödinger−Poisson
Effective potential corrected MC
Sheet charge (cm^−2)
Gate bias (V)
10^13
10^12
0
0.5
1
1.5
2
2.5
3

Figure 5.

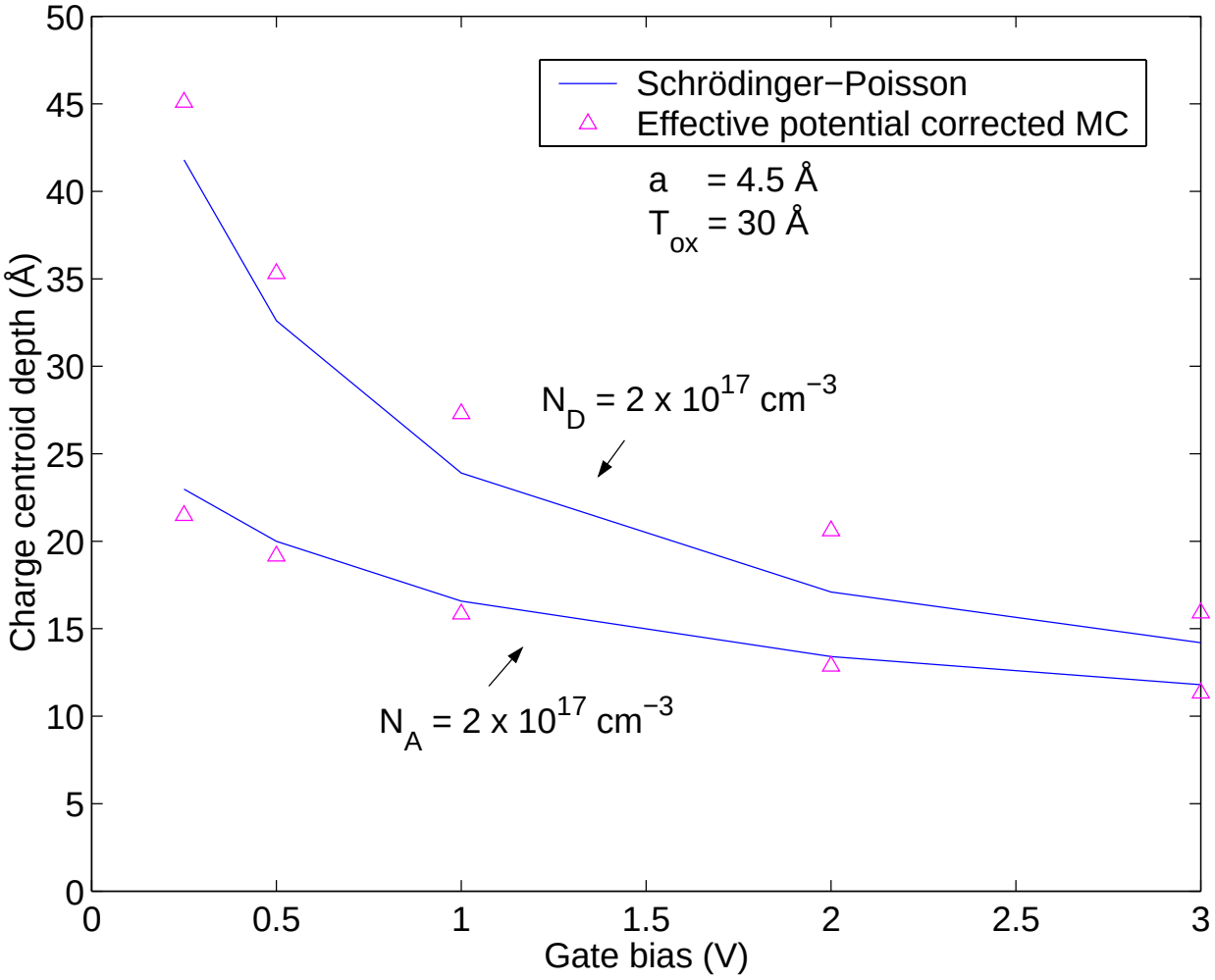
Schrödinger−Poisson
Effective potential corrected MC
a = 4.5 Å
T_ox = 30 Å
N_D = 2 x 10^17 cm^−3
N_A = 2 x 10^17 cm^−3
Charge centroid depth (Å)
Gate bias (V)
0
5
10
15
20
25
30
35
40
45
50
0
0.5
1
1.5
2
2.5
3

Figure 6.

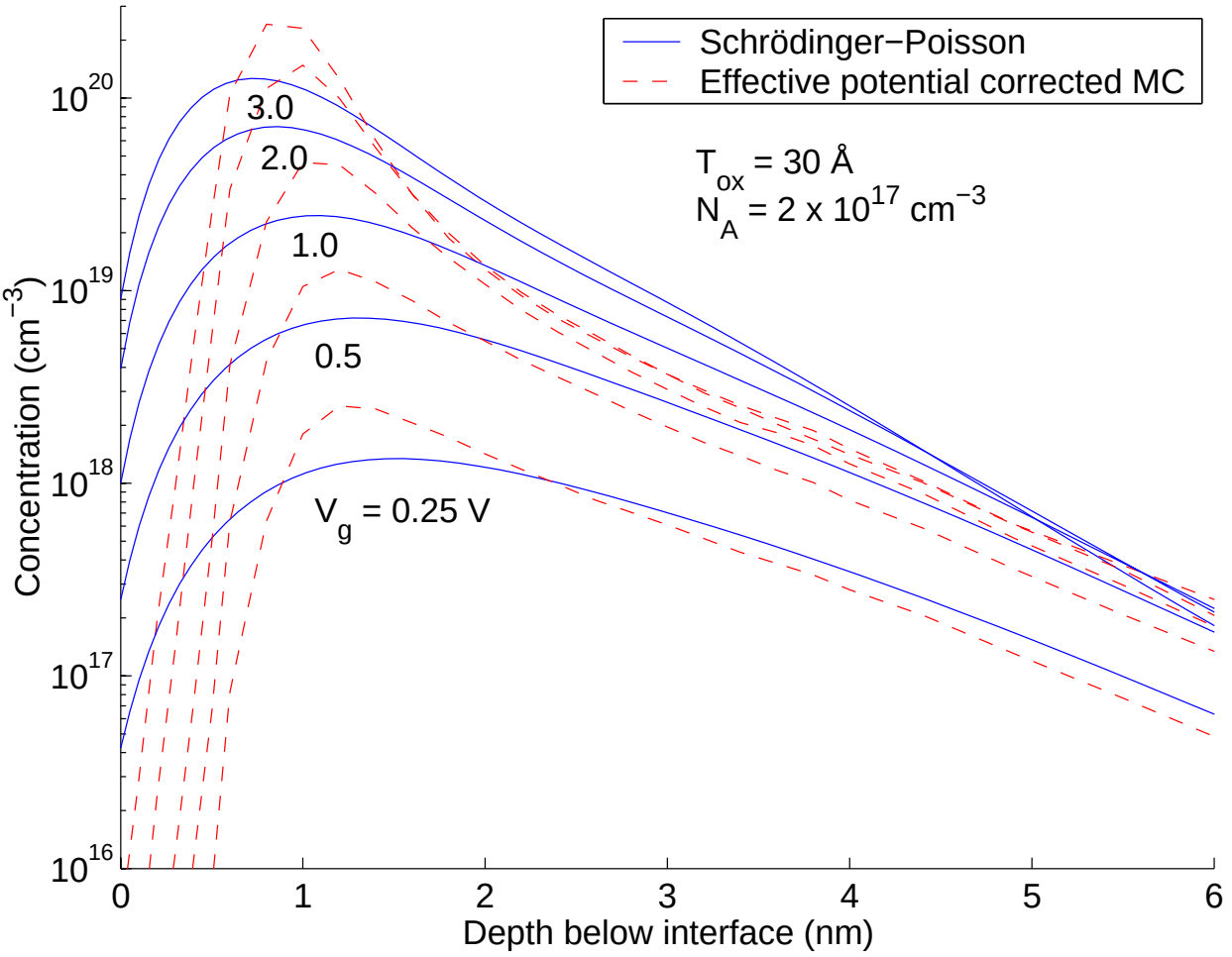
Schrödinger−Poisson
Effective potential corrected MC
T_ox = 30 Å
N_A = 2 x 10^17 cm^−3
3.0
2.0
1.0
0.5
V_g = 0.25 V
Concentration (cm^−3)
10^20
10^19
10^18
10^17
10^16
0
1
2
3
4
5
6
Depth below interface (nm)

Figure 7.

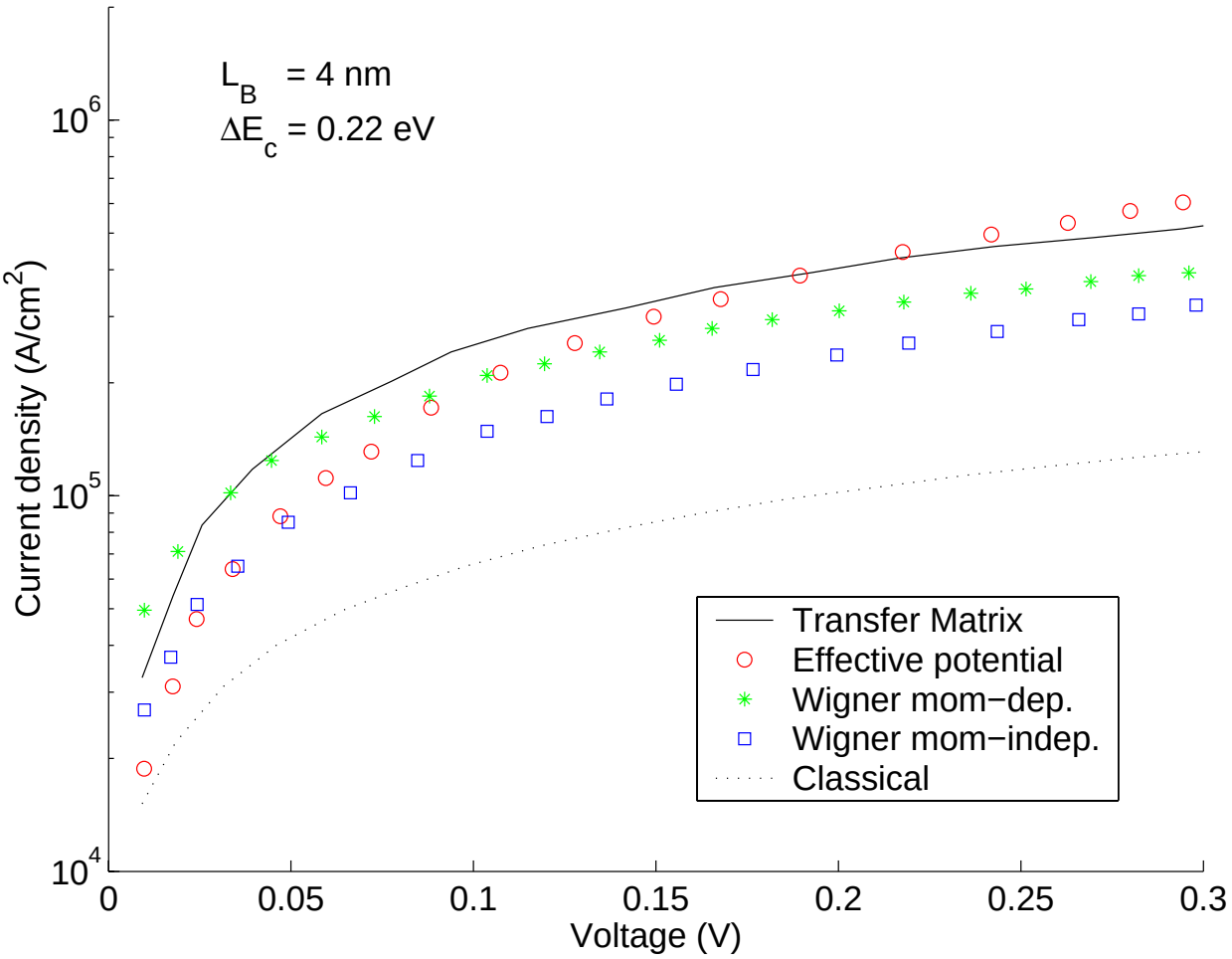
L_B = 4 nm
ΔE_c = 0.22 eV
Current density (A/cm^2)
10^6
10^5
10^4
0
0.05
0.1
0.15
0.2
0.25
0.3
Voltage (V)
Transfer Matrix
Effective potential
Wigner mom−dep.
Wigner mom−indep.
Classical

Figure 8.