UTQUANT2.0

User's Guide

W.-K. Shih, S. Jallepalli[+], G. Chindalore
Microelectronics Research Center
Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, TX78758
[+] Now at STL, Motorola, Austin, TX 78721

I Introduction

UTQUANT is a quasi-static CV simulator for one-dimensional silicon MOS structures. It was originally written [1, 2] for the purpose of studying the quantization effects in MOS inversion layers. In UTQUANT2.0, simple models consistent with the quantum-mechanical treatment of the inversion layers have been added to study the gate leakage current due to electron tunneling through thin gate oxide [3].

For given channel doping profile, oxide thickness, gate material and/or polysilicon gate doping level, UTQUANT self-consistently calculates sheet carrier concentration at each gate bias step and extracts quasi-static CV characteristics in accumulation, depletion, and strong inversion regions based on both classical and quantum mechanical treatment of the inversion- and accumulation-layer carriers. By comparing the classically and quantum-mechanically predicted CV characteristics, the effective increase of electrically measured oxide thickness due to quantum mechanical effects is extracted. UT QUANT2.0 also takes into account the effects due to band-gap narrowing, depletion in the polysilicon gate, interface states as well as oxide fixed charge:

Besides CV simulations, UTQUANT has a set of switches that allow users to calculate the inversion (and accumulation) layer subband dispersions and wave functions at different levels of sophistication. Following features have been implemented in the 2.0 release:

(1) Effective-mass approximation for CPU-efficient simulations.

(2) More rigorous pseudopotential calculation for users with fundamental interest in the subband structure of the MOS inversion layers.

(3) Evaluation of the inversion- and accumulation-layer effective mobility from the the UT local-field mobility model.

(4) Extraction of threshold voltage using current-based definition.

In addition to the above features, post-processing routines have been included in UTQUANT2.0 for estimating the NMOS gate leakage current due to both direct tunneling and Fowler-Nordheim tunneling through the gate oxide [3]. In the 2.0 release, only the tunneling current from the nMOS inversion layer to the gate is calculated. The quantum treatment of the inversion layer carrier distribution has to be turned on in order to invoke this function.

Following is the arrangement of this user's guide. In Section II, instructions on how to set up an inputdeck are given. Section III briefly describes the two important models used in UTQUANT2.0, i.e., the models for treating channel quantization effets and estimating the gate leakage current in the direct tunneling regime. Examples on using the program are given in Section IV.

II Using UTQUANT2.0

To start a UTQUANT simulation under the UNIX system prompt, users should use the folowing command:

utquant <filename of the inputdeck>

The filename of the inputdeck is the only argument needed in the command line. The inputdeck contains most of the structural and material information of the 1D MOS capacitor to be simulated. It also contains keys to turn on the desired physical models in the simulation. Besides, information on the substrate and polysilicon gate doping levels can be specified if they are assumed uniform for the simulation. According to the setup of the inputdeck, other input files might be needed to initialize the simulation. These include files that specify the doping concentration along the depth direction of the substrate and/or poly gate and interface state density within the bandgap. The formats and usage of these input files are described as follows

II.1 The inputdeck

The inputdeck is needed (with user-defined filename) to specify the MOS material structure and the models to be turned on during the program execution. In the inputdeck, the MOS structure to be simulated is given by the substrate doping profile, the gate material and the oxide thickness, etc.. The models that users are allowed to turn on/off include the quantization treatment, the options in tunneling current calculation and the step sizes for both the CV extraction and the tunneling current estimation. Following is a list of keys in the inputdeck. Please refer to the last section to see sample files.

EPS_INS	Relative dielectric permmitivity of the gate dielectric material.
insulator_thickness	Thickness of the gate dielectric in angstroms.

gate_voltage The value given to gate_voltage defines the swing of the gate bias in both the accumulation and inversion regions. A UTQUANT simulation starts from the gate bias corresponding to the flatband condition, enters the depletion region and continues into the inversion region until the magnitude of the gate bias reaches the value specified for gate_voltage. After this, the gate bias goes back to the flatband condition and continues with the opposite swing. The simulations end when the (magnitude) gate bias reaches the value specified for gate_voltage for the second time.

dVg Gate voltage is swept by the code in steps of ‘dVg' V, in both the positive and negative directions. Recommended dVg is in the range of 0.50 - 0.1.

gate_material The material type for the gate is specified in this key. gate_material is +/- 1 for N+ poly, +/- 2 for P+ poly, or 3 for metal. If the `-' option is chosen, the poly work function is set to the value given by the key work_function. If the`+' options is used, the program will default the poly work function to the crystalline value.

variable_polydoping Set to 0 when constant poly doping is used. In this case, the doping level specified in the key poly_doping is taken. Set to -1 or 1 when the doping profile is provided in the file polydope.in. In case of -1, the program generates its own mesh in the poly gate for the simulation. If 1 is given, the mesh in polydope.in is used.

poly_doping Doping level for the polysilicon gate. This key is ignored by the simulator unless variable_polydoping is set to 0.

work_function Work function for the gate material. Its value is ignored if gate_material is set to +1 or +2.

substrate_bias Substrate bias in V.

variable_doping Set to 0 when constant substrate doping is used. In this case, the doping level specified in the key doping_conc is taken. Set to -1 or 1 when the doping profile is provided in the file dope.in. In case of -1, the program generates its own mesh in the substrate for the simulation. If 1 is given, the mesh in dope.in is used.

doping_conc Doping level for the substrate. Its value should be positive for n-type substrate and negative for p-substrate. This key is ignored by the simulator unless variable_doping is set to 0.

strain Set to 1 for simulating CV in strained silicon/SiGe substrate []. The mole-fraction of Ge in the SiGe substrate (mole_fraction) has to be specified if strain is set to 1.

mole_fraction Ge mole-fraction in the SiGe substrate. This key is ignored unless strain is set to 1.

CLASSICAL Set to 1 when only the semi-classical solution is desired. When CLASSICAL is 1, carrier distribution along the device depth direction is calculated in consistency with the pseudopotential bulk band structure, the Fermi-Dirac statistics and the Poisson equation. When CLASSICAL is set to 0, the CV calculation further includes the effects due to inversion- and accumulation-layer quantization. Note that the classical CV is still calculated in this case.

TUNNEL Set to 1 in order to turn on the tunneling current post-processor. Two distinct approaches in tunneling current calculation have been implemented in UTQUANT2.0: the WKB approximation and the fully numerical solution. For efficient simulation, the WKB approach is recommended. However, to ensure the accuracy of the WKB estimation, users should perform spot checks by comparing the WKB results with the fully-numerical solutions. Note that CLASSICAL must be set to 0 in order to estimate the tunneling current. Also, note that only electrons tunneling from substrate to the gate is treated in the present version.

WKB This key is set to 1 when WKB approximation is to be used in estimating the tunneling current and set to 0 when the fully numerical calculation is desired.

FNTNL Set to 0 if tunneling current is to be calculated in both the direct tunneling and the Fowler-Nordiem tunneling regimes. Set to 1 when only the tunneling current in the FN regime is of interest.

IMAGE This key allows users to specify if image potential is to be included in the tunneling current calculation.

dVg_tunnel The value given here (in volts), together with the value given for dVg, determines the gate-bias interval in which the tunneling current estimation is performed. When the gate bias is incremented by dVg, the new gate bias is compared with the gate bias at which the most recent tunneling current calculation is performed. The tunneling curent estimator is turned on only if this difference is greater than or equal to dVg_tunnel.

TL The lattice temperature in K.

NFIX Areal concentration of fixed interface charge.

const_Dit Set to 1 when constant interface density of states is used. In this case, the interface state density is treated as uniformly distributed within the bandgap. Otherwise, the distribution of the trap density is specified in the file Dit.in.

Dit Interface state density in cm^{-2}.

option set to 0 turns on the pseudopotential full-band based calculation of the subband energies and wave functions at a given bias point (equivalent to option = 3 + full band formalism). option is set to 1 for a quasi-static capacitance calculation, 2 for the extraction of data through post-processing and 3 for obtaining the solutions (potential, carrier concentration, etc.) at a given bias point.

II.2 Other input files

As mentioned in II.1 (see the descriptions for **variable_polydoping**, **variable_doping**, and **const_Dit**), some other input files will be needed if non-uniform doping profiles or interface state distribution are to be used in the simulation. In case that **variable_polydoping**=0 or **variable_doping**=0, the files **polydope.in** or **dope.in** have to be provided by users. These files contain information on the polysilicon doping concentration and the substrate doping concentration along the device depth direction, respectively, both of two-column format, with the distance along the depth direction in the first column and the net concentration of electrically active dopants on the second. Values of doping have to be specified in units of cm^{-3} while the distance should be given in μm. Besides, donor concentration has to be positive and acceptor negative.

Having **const_Dit**=0 in the inputdeck implies that the interface trap density is non-uniform in energy in the band-gap. In this case, the file **Dit.in** is needed. It has to be a two-column ascii file with the energy in the first column specified (in eV) every 10meV (referenced to the mid-gap) and the 2-dimensional density of states (in $eV^{-1}cm^{-2}$) in the second column.

The change in the band-gap (in strained Si/SiGe sbstrate) has to be provided in the file **strain.in**. The format is: depth (in μm), dEc (in eV), dEv (in eV). The conduction band is assumed to be moved up and the valence band moved down. It should be noted that the strained Si layer has to be thick enough to include all of the inversion layer. Electrostatics/ transport in SiGe are not simulated (it is assumed to be a Si layer). Also, note that, in this version, only mobility in (100) Si-SiO2 system is modelled correctly. A beta version of an effective mobility model is available for electrons in strained silicon. We would not recommend it highly, though. It should only be better than using the silicon model.

All related model parameters chosen for the tunneling current calculation can be found in the file **gatus.h**. Among them, the values for the electron effecitve mass in the oxide bandgap and the oxide barrier height are the most uncertain ones among the published literature. In the current release, the former is set to 0.55 (in the units of free-electron mass) while the latter is set to 3.1 (in eV). Changes in these values might be needed in order to calibrate the model for explaining the experimentally measured data (see AI.3 for more details).

II.3 Simulation output

All filenames of the output files from a UTQUANT simulation end with **.out** and are in ASCII format. The content and format of each file are given below.

Clascap.out A two-column file with the gate bias (in V) in column 1 and the classical quasi-static capacitance (in F/cm^2) in column 2.

QMcap.out A two-column file with the the gate bias (in V) in column 1 and quantum-mechanical quasi-static capacitance (in F/cm^2) in column 2.

Claschg.out A five-column file with the the gate bias (in V) (column 1), the electron sheet concentration (column 2), the hole sheet concentration (column 3), the integrated dopant concentration (column 4), and the density of occupied interface states (column 5). All sheet concentration and density of states are obtained within the classical picture and are in the units of cm^{-2}.

QMchg.out In the same format as **Claschg.out**, except that all the quantities are obtained with the channel quantization taken into account.

igvg.out A five-column file arranged in the following way: gate bias (V), total I_g, $I_g(1)$, $I_g(2)$, $I_g(3)$, where $I_g(i)$ is the contribution to I_g from the *i*-th subband. All currents are in A/cm^2.

rate.out A four-column file arranged in the following way: oxide field (V/cm), $R_t(1)$, $R_t(2)$, $R_t(3)$, where $R_t(i)$ is the tunneling rate (in s^{-1}) of an electron in the *i*-th subband.

Clasdelta_tox.out Δt_{ox}, the difference between the electrically extracted and the physical oxide thickness in nm, versus E_{ox}, the oxide electric field in kV/cm, both obtained classically. This is a two-column file with E_{ox} in column 1 and Δt_{ox} in column 2.

QMdelta_tox.out Same as **Clasdelta_tox.out** , except that both E_{ox} and Δt_{ox} are obtained with channel quantization taken into account.

delta_tox.out Same as **Clasdelta_tox.out**, except that the Δt_{ox} is the difference between those in **QMdelta_tox.out** and **Clasdelta_tox.out**.

gm.out Transconductance in Simens/μm (column 2) as a function of V_g in V (column 1). The transconductance is extracted by takeing the derivative of the current value given in the file **cur.out** with respect to $V_{g.}$

polydrop.out Potential drop in V (column 2) across the gate as a function of the gate bias (column 1).

grid.out A two-column file that contains grid index (column 1) and the coordinate of the corresponding grid (column 2). The grid coordinate is referenced to the Si/SiO$_2$ interface and is in the units of Å.

threshold.out Threshold voltage (in V) for the MOS structure under simulation.

cur.out Linear region drain current in A/μm (column 2) versus V_g (column 1). The drain current is calculated by assuming a drain-source bias of 0.1V across a 1μm channel length, based on the classical inversion charge concentration if **CLASSICAL**=1 and the quantum charge concentration if **CLASSICAL**=0.

mobeff_fld.out The inversion or accumulation layer effective mobility (in cm^2/Vs) (column 2) versus the transverse effective field (in V/cm)(column 1). Note: only the effective-field mobility model is used in the calculation.

mobility.out This file is of the same format as **mobeff_fld.out** except that the local mobility model is used in the calculation.

III Models

III.1 Quantization models for inversion and accumulation layers

Two distinct approaches have been implemented in UTQUANT2.0 for treating the inversion-layer quantization of holes and electrons. With the effective-mass approximation (EMA), the effective-mass Hamiltonian in k-space is diagonalized in order to obtain the wavefunctions and subband dispersions. With the pseudopotential calculation, plane waves in the dimension parallel to the Si/SiO2 interface and sinusoidal functions in the quantization direction are used as the basis to construct the Hamiltonian. By default, the inversion layer sheet charge concentration is then calculated based on Fermi statistics and the density of states obtained from the subband dispersion. Five subbands have been included. Regardless the different means in obtaining the subband energies and wavefunctions, both EMA and pseudopotential calculation reaches self-consistent solution with the Poisson equation. For the detailed procedure of the pesudopotential calculation, users are referred to the work by Jallepalli et al. .

The treatment of accmulation layer carriers is different from that of the inversion carriers since the unbound states have a significant contribution to the total carrier sheet concentration, i.e., a large portion of the accumulation carriers have to be considered as classical particles, although the rest of the carriers are confined by the band bending near the interface. UTQUANT2.0 partitions the entire carrier population into the quantum and classical domains according to the carrier total energy. For energy greater than the band edge inside the bulk, classical (3D) density of states is used in computing the local carrier concentration. Otherwise, quantum (2D) density of states is used. In addition, since it becomes more computationally expensive and less accurate to treat the higher subbands, a cutoff energy is set at the subband edge of the 6th subband if the number of bound states exceeds five. This is expected to be a valid approximation since the subband spacing decreases as energy increases and the 2D density of states approaches its 3D counterpart.

III.2 Models for electron tunneling through the gate oxide

The tunneling current is calculated in a fashion consistent with the quantum-mechanical treatment of the inversion layers. Either the WKB approximation or the more rigorous solution to the quasi-bound state Schrodinger equation can be used in the tunneling current post processor. Both approaches calculate the lifetime of the inversion-layer quasi-bound states. Within the effective mass approximation, carriers in the same inversion-layer subband have the same tunneling rate and the total tunneling current density is simply

$$J = \sum_i n_i \Gamma_i$$

where n_i. and.Γ_i. are respectively the carrier population and the tunneling rate of the *i*-th subband. When the WKB treatment is desired, the WKB tunneling rate

$$\exp(-2\int \sqrt{2m(E-V(x))}/\hbar \ \ dx)$$

is used in estimating the lifetime of each subband. The WKB is, in general, more valid with a gradually varying potential relative to the carrier kinetic energy. A more rigorous treatment requires solving the scattering-matrix problem associated with the 1D tunneling structure. The procedure of solving such a quasi-bound state problem has been detailed in [3].

It should be noted that in the tunneling current calculation, the carrier energy distribution has been assumed to be in equilibrium with the lattice temperature. Physically, an implicit assumption has been made that either the carrier thermal generation rate in the depletion layer is much higher comparred to the tunneling rate, or a source/drain region exists to supply the minority carriers in the inversion layer when the tunneling rate dominates the thermal generation rate. It is uncertain whether the present tool applys when optical excitation is used to generate the inversion layer electrons.

IV Examples

The use of UTQUANT2.0 is demonstrated by going through some of the electrical analysis on MOS capacitors. This section contains three examples of performing CV simulations and tunneling current estimations on both n- and p-MOS capacitors. In Example 1, we perform CV simulation on a pMOS structure with uniformly doped substrate and p+ polysilicon gate. In Example 2, both CV and Ig-Vg are simulated on a nMOS structure with 2.5 nm gate oxide. In Example 3, CV simulation is performed on an experimental nMOS structure and comparison between theoretical and experimental results are made.

IV.1 Example 1: CV simulation on pMOS

The inputdeck shown in Figure 1 is used to simulate the CV characteristics on a pMOS test structure. In this example, the n-type substrate doping level is fixed at 1x10^{17} cm^{-3} and oxide of 3 nm is used. P-poly doped at 6x10^{19} cm^{-3} is used . **gate_material** = +2 in the inputdeck specifies that the p-poly is used for the gate material and the gate work functioin is set to that of the crystaline silicon at equivalent doping level. and quantum treatment is desired. Zero interface state density is assumed and the interface fixed charge density is set to 3x10^{10}.cm^{-2} The simulated CV results curves (saved in QMcap.out and Clascap.out) are shown in Figure 2. Effects due to poly depletion has been captured in the CV. Difference between classical and quantum mechanical CV in both inversion and accumulation regions is obvious.

```
EPS_INS = 3.9
insulator_thickness = 30.0
gate_voltage = 3.0
dVg = 0.05
gate_material : +2
variable_polydoping ? 0
poly_doping = -6.0e19
work_function = 4.45
substrate_bias = 0.0
variable_doping ? 0
doping_conc = 5e17
strain ? 0
mole_fraction = 0.0
CLASSICAL ? 0
TUNNEL ? 0
dVg_tunnel = 0.05
TL = 300.0
NFIX = 3.0e10
const_Dit ? 1
Dit = 0.0e12
option = 1
```

Figure 1: Inputdeck for Example 1

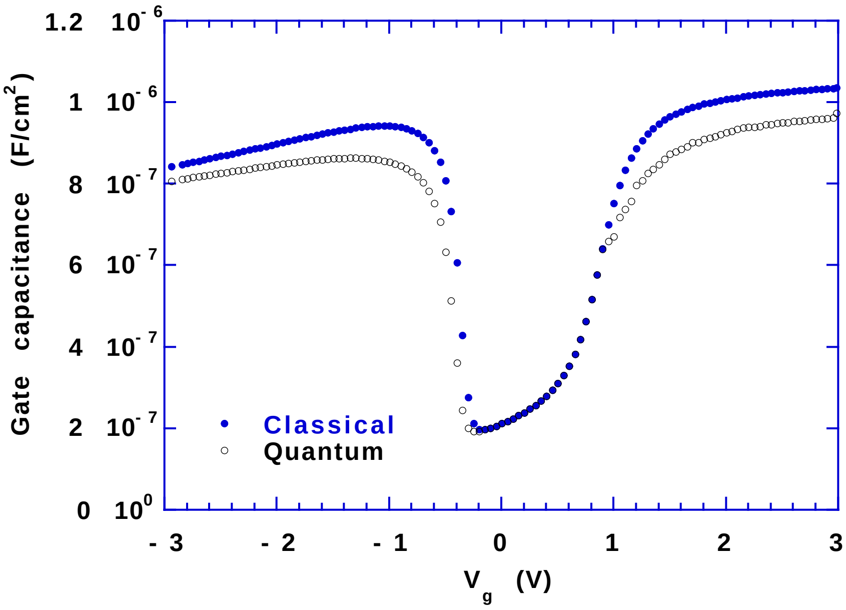

Figure 2: CV characteeristics of Example 1.

IV.2 Example 2: CV and Ig-Vg simulation on nMOS with uniform substrate doping

WKB approximation is used in this example to calculate the direct tunneling current. In the inputdeck shown in Figure 3, both **WKB** and **TUNNEL** are set to 1 and **CLASSICAL** is set to 0. **FNTNL** is set to 0 to enable the tunneling current estimation in the direct tunneling regime. Figure 4 shows the resulting I_G-V_G characteristics.

```
EPS_INS = 3.9
insulator_thickness = 25.0
gate_voltage = 4.0
dVg = 0.05
gate_material : 3
variable_polydoping ? 0
poly_doping = 6.0e19
work_function = 4.45
substrate_bias = 0.0
variable_doping ? 0
doping_conc = -1e17
strain ? 0
mole_fraction = 0.0
CLASSICAL ? 0
TUNNEL ? 1
WKB ? 1
FNTNL ? 0
IMAGE ? 0
dVg_tunnel = 0.10
TL = 300.0
```

```
NFIX = 3.0e10
const_Dit ? 1
Dit = 0.0e12
option = 1
```

Figure 3: Inputdeck for Example 2.

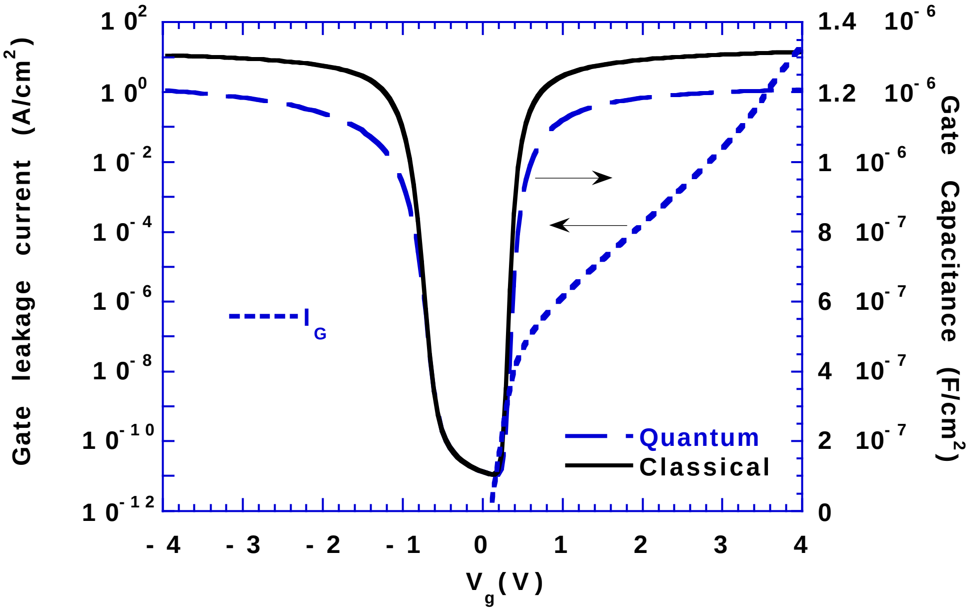

Figure 4: CV and I_g-V_g for Example 2.

IV.3 Example 3: CV simulation on an experimental nMOS capacitor

What we have demonstrated so far are simplified cases where test structures have been used to obtain the CV and I-V relations. A more realistic example is provided in this example, where experimental MOS device is simulated and the simulation result is compared to the measured data [4]. The goal of the experiment is to observe the shift in MOS threshold voltage due to quantization. The MOS capacitor with Al gate (work function set to 4.42 eV) and an oxide thickness of 22.1 nm is fabricated on p-type substrate with doping profile shown in Figure 7. Note that in the inputdeck (see Figure 5), the key **variable_doping** is set to -1, indicating that the external doping profile (saved in file **dope.in**) is to be used and, instead of using the mesh in dope.in, UTQUANT is to generate its own mesh for the simulation. If the mesh in dope.in were desired to be used, **variable_doping** should be set to 1. Note also that, in this example, we are not only interested in the CV characteristics but also other physical quantities such as the carrier concentration and electrostatic potential along the depth direction at the final V_g. For this purpose, **option** is set to 3.

```
EPS_INS = 3.9
insulator_thickness = 221.0
gate_voltage = 4.0
dVg = 0.05
gate_material : 3
variable_polydoping ? 0
poly_doping = 6.0e19
work_function = 4.42
substrate_bias = 0.0
variable_doping ? -1
doping_conc = -5e17
strain ? 0
mole_fraction = 0.0
CLASSICAL ? 0
TUNNEL ? 0
dVg_tunnel = 0.05
TL = 300.0
NFIX = 3.0e10
const_Dit ? 1
Dit = 0.0e12
option = 3
```

Figure 5: Inputdeck for Example 3

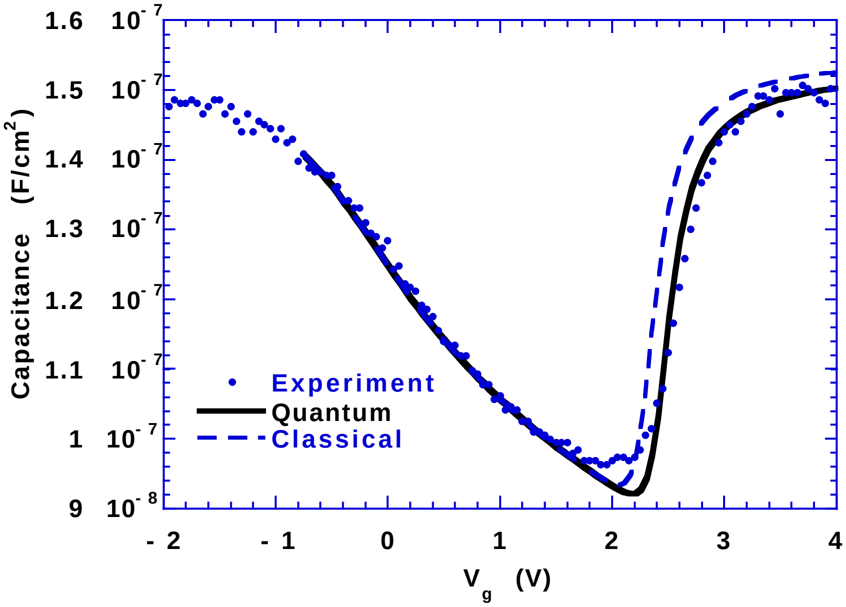

Figure 6: Experimantal vs. simulated CV characteristics.

The comparison between experimental and simulated CV characteristics is shown in Figure 6. The simulation is completed at V_g=4V, at which the solution variables are desired. The carrier concentrations at this bias point are saved in files **QMelec_conc.d** and **Claselec_conc.d** (see Figure 7).

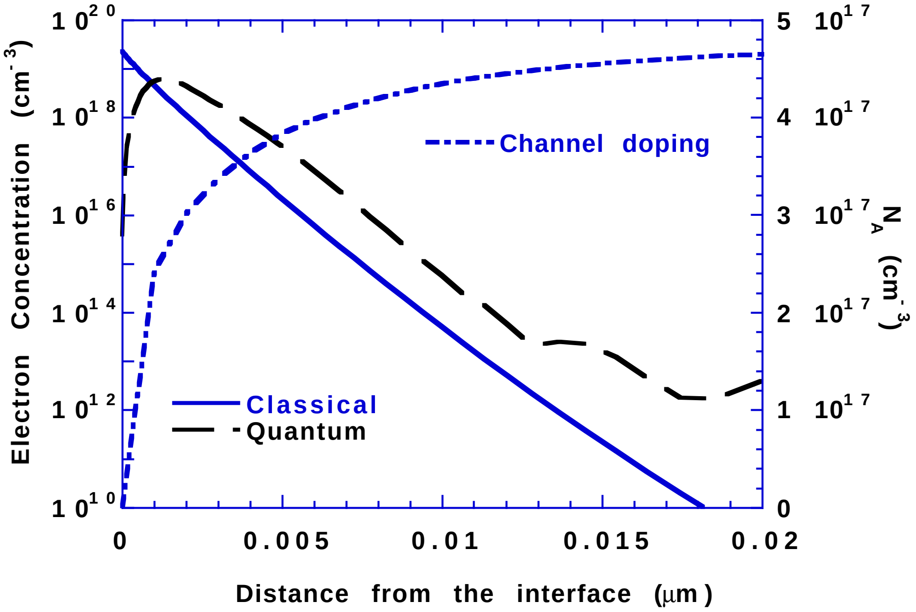

Figure 7: Electron concentration along the MOS depth direction obtained at V_g=4.0 V with and without the quantization effects. Channel doping of the experimental MOS capacitor is also shown.

V Technical Contacts

For technical assistance and additional information on the models and code implementation, please contact:

Christine Maziar
Phone: (512) 471-4456
Fax: (512) 471-0577
email: cmaziar@mail.utexas.edu

Al F. Tasch
Phone: (512) 471-1640
Fax: (512) 471-5625
email: tasch@roz.mer.utexas.edu

VI References

[1] S. Jallepalli, J. Bude, W.-K. Shih, M. R. Pinto, C. M. Maziar, and A. F. Tasch, “Electron and hole quantization and their impact on deep submicron silicon p- and n-MOSFET characteristics,” *IEEE Trans. Electron Devices*, vol. 44, pp. 297, 1997.

[2] S. Jallepalli, Ph.D. dissertatioin, The University of Texas at Austin, 1996.

[3] W.-K. Shih, X. Wang, S. Jallepalli, C. M. Maziar, and A. F. Tasch, “Modeling gate leakage current in nMOS structures due to direct tunneling through ultra-thin oxide,” , 1997.

[4] G. Chindlore, S. A. Hareland, S. Jallepalli, A. F. Tasch, C. M. Maziar, V. K. F. Chia, and S. Smith, “Experimental determination of threshold voltage shifts due to quantum mechanical effects in MOS electron and hole inversion layers,” *Electron Device Letters*, vol. 18, 1997.