Ballistic MOS Model (BMM) Considering 2D Quantum Effects

Zhiping Yu, Dawei Zhang, and Lilin Tian
Institute of Microelectronics
Tsinghua University
Beijing, China
yuzhip@tsinghua.edu.cn
Outline

- Background
- Surface carrier density in ballistic transport
- 1D QM correction
- 2D QM correction
- Effects in poly-gate region
- Mobility modeling
- Results
Background

- Threshold-voltage based MOS modeling
- Ballistic transport in sub-100nm MOSFETs
- Extensive work on 1D QM correction
- QM effects along the channel
  - Tunneling through the S/D barrier
  - Threshold voltage lowering due to QM (neglected so far)
- Modeling of Poly-depletion and QM in poly
Ballistic MOS Modeling


\[ F^+ = (1 - r)F_S^+ + rF_D^- \]
\[ F^- = F_D^- \]
\[ F = F^+ - F^- = (1 - r)F_S^+ - (1 - r)F_D^- \]
\[ F_D^- = F_S^+ e^{-qV_{ds}/k_BT} \]
Flux Method

- At the peak of source-end barrier (0-):

\[
\frac{Q(0)}{q} v_T = (1 + r) F^+_S + (1 - r) F^-_D
\]

\[
I_{ds} = qWF = qW \left[ (1 + r) F^+_S + (1 - r) F^-_D \right] \frac{(1 - r) F^+_S - (1 - r) F^-_D}{(1 + r) F^+_S + (1 - r) F^-_D}
\]

\[
= WQ(0) v_T \, \frac{1 - r}{1 + r} \, \frac{1 - e^{-qV_{ds}/k_B T}}{1 + \frac{1 - r}{1 + r} e^{-qV_{ds}/k_B T}}
\]

- The reflection coef. \( r \) is function of \( V_{gs} \) and \( V_{ds} \).
Drain Current Expression


\[
\begin{align*}
    r &= \frac{l}{l + \lambda} \\
    l(V_{ds}) &= L \left( \beta \frac{V_T}{V_{ds}} \right)^\alpha, \\
    \lambda(V_{gs}, V_{ds}) &= 2V_T \frac{\mu(V_{gs})}{\nu_T} \frac{\Im^2(\eta)}{\Im_{-1}(\eta) \Im_{1/2}(\eta)} \\
    \frac{I_D}{W} &= Q(0) \frac{1 - r}{1 + r} \left[ \nu_T \frac{\Im_{1/2}(\eta)}{\Im_0(\eta)} \right] \\
        & \quad \cdot \left( 1 - \frac{\Im_{1/2}(\eta - U_D)}{\Im_{1/2}(\eta)} \right) \\
        & \quad \cdot \left( 1 + \frac{\Im_0(\eta - U_D)}{\Im_0(\eta)} \right)
\end{align*}
\]
Model Parameters

\( \alpha \) and \( \beta \) are fitting params, \( U_D = V_{ds}/V_T, \ \eta = \frac{E_{fS} - E_{\text{max}}}{k_B T} \)

\( \eta \) is determined from \( Q(0) \) and \( r \) by \( V_{gs} \) and \( V_{ds} \).

\[
Q(V_{gs}) = qN_{2D} \mathcal{Z}_0(\eta)(1+r) \left[ 1 + \frac{1-r}{1+r} \frac{\mathcal{Z}_0(\eta-U_D)}{\mathcal{Z}_0(\eta)} \right]
\]

– Applications: double gate (DG) MOSFET

\( L_g = 20\text{nm}, \ t_{si} = 1.5\text{nm}, \ t_{ox} = 1.5\text{nm} \)

– Compared to results from applying NEGF
Purdue’s Results of DG-MOS

Graph showing the relationship between $I_{DS}$ [μA/μm] and $V_{DS}$ [V] for different values of $V_{GS}$: 0.55V, 0.50V, 0.45V, and 0.40V.
Extension to Bulk CMOS and Full 2D QM Modeling

- Modeling of $Q(0)$ using $V_{th}$ approach

\[ Q(0) = C_{ox} V_{gst,\text{eff}}, \quad V_{gst} = V_{gs} - V_{th} \]

- Key expression (Yutao Ma, Zhiping Yu, et al., Tr.-ICCAD, 2001)

\[
V_{gst,\text{eff}} = \frac{2s V_T \ln \left[ 1 + e^{V_{gst} / 2s V_T} \right]}{1 + 2s C_{ox} \sqrt{2\phi_{s0} / q\varepsilon_0 \varepsilon_s N_{sub}}} \left[ V_{gst} - 2(V_{gs} - V_{th} - V_{off}) \right] / 2s V_T
\]

\[
V_{gst} = V_{gs} - V_{th}, \quad V_{th} = V_{FB} - \phi_{s0} - Q_{dep} / C_{ox}
\]
1D QM Charge-Based Model

• Features:
  – $V_{th}$ shift due to 1D QM
  – Considering finite inversion layer thickness
  – Widening of depletion region width in strong inversion due to the increased surface potential

Modeling of 2D QM Effects

- Separation of variables

\[
\Psi(x, y, z) = \frac{A}{\sqrt{p(x)}} \exp \left( \frac{i}{\hbar} \int^x p(\xi) d\xi \right) \left[ X(x) \sqrt{\frac{2}{W}} \sin \left( \frac{n_y \pi}{W} y \right) \cdot \varphi_{n_z}(z) \right] Y(y) Z(z)
\]

- WKB method for solving 1D Schroedinger eq. along the channel
2D QM Effects (cont’d)

\[ p(x) = \sqrt{2m_x \left[ E - U(x, y, z) \right]} \]

is imaginary if \( E < U(x, y, z) \)

Define the lowest eigenlevel with nonlocal state

\[
\frac{i}{\hbar} \int_a^b \sqrt{2m_x \left[ E_d - U(x) \right]} dx = -\alpha, \text{ s.t. } e^{-\alpha} \ll 1, \text{ say } \alpha = 4
\]
2D QM Effects (cont’d)

• Approximate the potential barrier using parabola

\[ E(x) = E_{\text{max}} - \sigma (x - x_{\text{max}})^2 \]

• Define the correction of 2D QM to \( V_{th} \) as

\[ \Delta V_{th,2\text{DQM}} = E_d - E_{\text{max}} \]

\[ \Delta V_{th,2\text{DQM}} = -\frac{8\hbar}{q\pi} \sqrt{\frac{\sigma}{2m_x}} \left( 1 + \frac{t_{ox}}{2\varepsilon_0\varepsilon_{ox}} \sqrt{\frac{q^2\varepsilon_0\varepsilon_{si}N_{\text{sub}}}{k_B T \ln(N_{\text{sub}}/n_i)}} \right) \]
Results (2D QM)

- $\sigma = 3 \times 10^{-2} e^{-L/16} \left(\text{eV/nm}^2\right)$ (\textit{L} units: nm) is insensitive to $t_{ox}$ and $N_{sub}$.
2D QM Modeling

Surface Charge Density (cm$^{-2}$) vs. Gate Voltage (V)

- Numerical $L=20$nm
- Numerical $L=80$nm
- 2D QM Analytical
- 1D QM Analytical

$T_{ox}=1.2$nm
$N_{sub}=2\times10^{18}$cm$^{-3}$

$T_{ox}=1.6$nm
$N_{sub}=2\times10^{18}$cm$^{-3}$

$T_{ox}=1.2$nm
$N_{sub}=4\times10^{18}$cm$^{-3}$
Poly-gate Modeling

- Poly depletion
- QM effects: the formation of dipole layer
  - Negative shift in $V_{th}$
  - Amount not sensitive to bias $V_g$

$$N_p = 5 \times 10^{20} \text{ cm}^{-3}$$
Poly-gate Modeling

• Poly-depletion effect
  – Moderately strong inversion, modeled as $V_{th}$ shift

  \[
  \Delta V_{th,\,pd} = \left( \frac{N_{\text{sub}}}{N_p} \right) (2\phi_B + V_{sb})
  \]

  – Strong inversion, has to be modeled as $V_g$ shift

  \[
  V_{gs,\,eff} = V_{FB} + \psi_S + \frac{1}{2k} \left[ \sqrt{1 + 4k(V_{gs} - \psi_S - V_{FB})} - 1 \right]
  \]

  \[
  k = \frac{\varepsilon_{ox}^2}{2q\varepsilon_{st}t_{ox}^2N_{\text{poly}}}
  \]
Modeling of Combined PD/QM Effects in Poly-Gate

\[ Q_{\text{strong\_inversion}} = C'_{ox} \left( V_g - V_{TH}' \right) / q \]
Threshold Voltage Shift

- Due to QM dipole in the poly

\[ -\Delta V_{th, QMpoly} = 0.0387 + 0.026 \times \log\left(10^{-20} \times N_p\right) \]

- Threshold voltage after all corrections:

\[ V_{th} = V_{th, 1Dqm} + \Delta V_{th, 2Dqm} + \Delta V_{th, QMpoly} + \Delta V_{th, SCE} + \Delta V_{th, DIBL} \]
Other Corrections to $V_{th}$

- **SCE**

$$
\Delta V_{th,SCE} = -\frac{0.5 \times DVT0}{\cosh\left(DVT1 \times \frac{L}{l_t}\right) - 1} (V_{bi} - \phi_{sqm})
$$

- **DIBL**

$$
\Delta V_{th,DIBL} = -\frac{0.5}{\cosh\left(D_{sub} \times \frac{L}{l_t}\right) - 1} \times ETA0 \times V_{ds}
$$
Mobility Modeling

- Dependence on both $V_{gs}$ and $L$

$$\mu = \frac{U_0}{1 + U_1 \times V_{gs} + \left[U_2 + U_3 \times \exp\left(-L/U_4\right)\right] \times V_{gs}^2}$$
Model Parameter Extraction

- Eleven parameters to extract:
  \[ \alpha, \beta, U0, U1, U2, U3, U4, DVT0, DVT1, D_{sub}, ETA0 \]
- Two parameters to SCE: \[ DVT0, DVT1 \]
- Five parameters to mobility: \[ U0-4 \]
- Two parameters to critical length, \( l \), in ballistic transport:
  \[ \alpha : 0.2 - 1, \quad (1/\alpha: 5-1), \quad \beta: 0.4-1.5 \]
- Two parameters to DIBL: \[ D_{sub}, ETA0 \]
BMM Validation

• Six sub-50 nm (14 - 44nm) bulk CMOS FETs from published literature ($t_{ox}$ and $N_{sub}$ are extracted and fixed from physical structure.)

• Model scalability is tested through three simulated groups of structure.
14 nm CMOS Test

  - EOT=1nm
  - $I_{on}/I_{off}$ (nMOS): 564µA/µm, 264nA/µm
  - $I_{on}/I_{off}$ (pMOS): 251µA/µm, 185nA/µm
  - $V_{gs}=V_{ds}=0.75V$
Test of Model Scalability

Solid lines are from BMM and data points from numerical simulation.
Conclusions

• BMM can provide smooth and accurate $I-V$ characteristics for sub-50nm MOSFETs for all regions of operation.

• 2D QM effects are modeled using 1D correction to the threshold voltage.

• In view of the QM/Poly-depletion effects in the gate, the effective gate dielectric thickness should be modeled as dependent upon the gate bias.

• More physics-based and sub-threshold work in progress