A Charge Conserving Non-Quasi-Static (NQS) MOSFET Model for SPICE Transient Analysis

Hong-June Park, Ping Keung Ko, Member, IEEE, and Chenming Hu, Fellow, IEEE

Abstract—An analytic charge conserving non-quasi-static (NQS) model has been derived for long channel MOSFET's and has been implemented in SPICE3. The model is based on approximate solutions to the transient current continuity equation, and analytic equations have been derived for node charges using the charge-sheet formulation. The NQS effects in several test circuits which include a pass transistor, a CMOS inverter chain, and a differential sample-hold circuit, are simulated. Excellent agreements have been observed among this work, PISCES (2-D device simulation), the 1-D numerical simulation, the multiple lump model, and CODECS (a mixed device and circuit simulation). However, large differences have been observed between this work and conventional quasi-static (QS) models. The model computation time of this work implemented in SPICE3 is about 2-3 times larger than those of QS models (BSIM, Level-2 Meyer) in SPICE3.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>Lateral dimension along the channel (y = 0 at source, y = L at drain) (centimeters).</td>
</tr>
<tr>
<td>L</td>
<td>Effective channel length (centimeters).</td>
</tr>
<tr>
<td>W</td>
<td>Effective channel width (centimeters).</td>
</tr>
<tr>
<td>t</td>
<td>Time (seconds).</td>
</tr>
<tr>
<td>(Q'_c(y, t))</td>
<td>Inversion charge density normalized by (W C_{OX}) (volts).</td>
</tr>
<tr>
<td>(C_{OX})</td>
<td>Gate oxide capacitance per unit area (farads per square centimeter).</td>
</tr>
<tr>
<td>(I_D(t))</td>
<td>Drain current (1) (amperes).</td>
</tr>
<tr>
<td>(I_S(t))</td>
<td>Source current (2) (amperes).</td>
</tr>
<tr>
<td>(I_G(t))</td>
<td>Gate current (amperes).</td>
</tr>
<tr>
<td>(I_B(t))</td>
<td>Bulk current (amperes).</td>
</tr>
<tr>
<td>(I_{DC}(t))</td>
<td>DC transport current (amperes).</td>
</tr>
<tr>
<td>(Q_D(t))</td>
<td>Drain charge (3) (coulombs).</td>
</tr>
<tr>
<td>(Q_S(t))</td>
<td>Source charge (4) (coulombs).</td>
</tr>
<tr>
<td>(Q_G(t))</td>
<td>Gate charge (32) (coulombs).</td>
</tr>
<tr>
<td>(\psi_L(y, t))</td>
<td>Surface potential at a channel point (volts).</td>
</tr>
<tr>
<td>(\psi_{SD}(t))</td>
<td>Surface potential at source end of channel (volts).</td>
</tr>
<tr>
<td>(\psi_{DS}(t))</td>
<td>Surface potential at drain end of channel (volts).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_{GS}(t))</td>
<td>Equivalent to ((V_{GS} - V_{TH})), but is non-negative (12) (volts).</td>
</tr>
<tr>
<td>(F_B)</td>
<td>Correction factor due to bulk (substrate) charge effect.</td>
</tr>
<tr>
<td>(P_S(t))</td>
<td>((Q'_S(0, t) + F_B V_T)^2) (square volts).</td>
</tr>
<tr>
<td>(P_B(t))</td>
<td>((Q'_B(L, t) + F_B V_T)^2) (square volts).</td>
</tr>
<tr>
<td>(a(t))</td>
<td>(1 - (P_B(t)/P_S(t))).</td>
</tr>
<tr>
<td>(\mu)</td>
<td>Mobility of channel carriers ((cm^2/V \cdot s)).</td>
</tr>
<tr>
<td>(D_p)</td>
<td>Diffusion constant of channel carriers (square centimeter per second).</td>
</tr>
<tr>
<td>(q)</td>
<td>An electronic charge (coulombs).</td>
</tr>
<tr>
<td>(k)</td>
<td>Boltzmann constant ((V \cdot C/\text{deg})).</td>
</tr>
<tr>
<td>(T)</td>
<td>Absolute temperature (degrees).</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>Permittivity of silicon (farads per centimeter).</td>
</tr>
<tr>
<td>(\epsilon_{ox})</td>
<td>Permittivity of silicon dioxide (farads per centimeter).</td>
</tr>
<tr>
<td>(N_{SUB})</td>
<td>Bulk (substrate) doping concentration ((cm^{-3})).</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Bulk effect coefficient ((\sqrt{2\epsilon_{ox} q N_{SUB}/C_{OX}})/V).</td>
</tr>
<tr>
<td>(V_{FB})</td>
<td>Flat band voltage (volts).</td>
</tr>
<tr>
<td>(V_T)</td>
<td>Thermal voltage ((kT/q)) (volts).</td>
</tr>
<tr>
<td>(V_{GS})</td>
<td>Applied voltage between gate and bulk nodes (volts).</td>
</tr>
<tr>
<td>(V_{DS})</td>
<td>Applied voltage between gate and source nodes (volts).</td>
</tr>
<tr>
<td>(V_{BS})</td>
<td>Applied voltage between bulk and source nodes (volts).</td>
</tr>
<tr>
<td>(V_{TH})</td>
<td>Threshold voltage at source for QS models (volts).</td>
</tr>
<tr>
<td>(N_{SUB}/n_i)</td>
<td>Intrinsic carrier concentration of silicon ((cm^{-3})).</td>
</tr>
<tr>
<td>(\Phi_F)</td>
<td>(V_T \cdot \log((N_{SUB}/n_i)) (volts).</td>
</tr>
<tr>
<td>(n_i)</td>
<td>Intrinsic carrier concentration of silicon ((cm^{-3})).</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>Empirical channel length modulation factor ((V^{-1})).</td>
</tr>
</tbody>
</table>

1. INTRODUCTION

In simulating charge storage circuits such as switched analog circuits [1] and DRAM’s, it is important that the device model used [2]-[4] conserves charge because charge nonconserving models produce erroneous simulation results. Charge-based models which give ex-
plicit expressions for node charges are guaranteed to con-
serve charge [2], [4].

During the turn-off transient of MOSFET switches, the
channel charge is injected into either the source or the
drain node. The injection of the channel charge into the
substrate (charge pumping) can be neglected under nor-
mal circuit operations [5]. This channel charge injection
is known to be one of the major distortion sources in the
low distortion switched capacitor filters [6] and is also one
of the major obstacles in designing high-speed and high-
resolution MOS A/D converters [7]. Since the conven-
tional quasi-static (QS) SPICE models cannot predict
channel charge injection accurately, a non-quasi-static
(NQS) MOSFET model is required.

Turchetti et al. [8] reported a NQS transient MOSFET
model. They assumed a linear channel charge density pro-
file with the lateral dimension for the QS component and
a symmetrical profile with respect to source and drain for
the NQS component. Since the profiles are over-simpli-
fied, the usefulness of the model [8] is limited. We
reported a NQS MOSFET model [9] which predicts the
NQS behavior accurately both for the transient [10] and ac
[11] analyses but does not conserve charge for the tran-
sient analysis. The charge nonconservation was found to be
due to the fact that the model in [10] is not based on
charges. Chai et al. [12] reported a NQS MOSFET
model, but it uses an iterative method which is ex-
pected to take much longer computation time than ana-
lytic methods. This work describes an analytic, charge-
based, and hence, charge conserving NQS MOSFET
model for the transient analysis. A preliminary version of
this paper was reported in [13] and the ac counterpart of
this paper can be found in [9], [11], [14].

A recently published paper [15] shows that the Meyer
capacitance model, which is not a charge-based model,
can be made to conserve charge by adding terms in the
model implementation. However, it is not clear whether
the approach in [15] guarantees the charge conservation,
although the simulation results suggest it conserves
charge. On the contrary, the charge-based models which
use charges as the state variables clearly guarantee the
charge conservation as shown in [2] and [4]. Further-
more, the model in [15] is almost as complex as charge-
based models (see [15, fig. 1], [14, fig. A2.5]) but is not
as accurate. Hence, it is still believed that the charge-
based approach is the right direction.

II. FORMULATION OF MODEL EQUATIONS

2.1. Formulation of the State Equation

From the current continuity equation, one can obtain the
drain and source currents, \( I_D(t) \) and \( I_S(t) \), as follows
[5], [8], [14]:

\[
I_D(t) = I_{DC}(t) + \frac{dQ_D(t)}{dt} \tag{1}
\]

\[
I_S(t) = I_{DC}(t) + \frac{dQ_S(t)}{dt} \tag{2}
\]

where

\[
Q_D(t) = -WC_{OX} \cdot \int_0^L y \cdot Q_S'(y, t) \, dy \tag{3}
\]

\[
Q_S(t) = -WC_{OX} \cdot \int_0^L \left( 1 - \frac{y}{L} \right) \cdot Q_D'(y, t) \, dy \tag{4}
\]

where \( I_{DC}(t) \) is the dc transport current [2], \( Q_D(t) \) and
\( Q_S(t) \) are the drain and source charges, respectively. \( y \)
is the lateral dimension from source \((y = 0)\) to drain \((y = L)\), \( W \) is the effective channel width, \( C_{OX} \) is the gate oxide
capacitance per unit area, and \( Q' \) is the inversion charge
density normalized by \( WC_{OX} \) (\( Q' \) has a unit of volts).

From the current relation and the current continuity
equation, one can derive [10]

\[
\frac{\partial Q_S'(y, t)}{\partial t} = \frac{\mu_n}{2F_B} \cdot \frac{\partial^2}{\partial y^2} \left[ (Q_D'(y, t) + F_B V_t)^2 \right] \tag{5}
\]

where

\[
F_B(t) = 1 + \frac{0.5\gamma}{\sqrt{\Psi_S(t) - V_t}} \cdot \frac{1}{1.744 + 0.8364 \cdot (\Psi_S(t) - V_t)} \tag{6}
\]

where \( F_B \) is the correction factor due to the bulk (sub-
strate) charge effect [17], [18]. \( \Psi_S(t) \) is the surface
potential at the source end of the channel, and \( V_t \) is the
thermal voltage \((kT/q)\).

In the charge nonconserving NQS model [10], \((Q_D'(y, t) + F_B V_t)^2\) is multiplied to both sides of (5) and the variable \((Q_D'(y, t) + F_B V_t)^2\) is solved and analytic equations have been derived for node currents. However, it turns out that the model in [10] does not conserve charges. In this work, instead of \((Q_D'(y, t) + F_B V_t)^2\), \( Q_D'(y, t) \) is solved directly from (5) so that analytic equations can be derived for node charges (charge-based) (see (3) and (4)) and it guarantees the charge conservation.

To solve \( Q_S'(y, t) \) from the partial differential equation
(5), we decompose \( Q_S'(y, t) \) into a QS component and a
NQS component. The QS component can be derived from
(5) by setting the time derivative to zero. We assume that
the QS approximation holds for inversion charge densities
at both the source and drain ends, i.e., we assume that
the inversion charge densities at source and drain ends
respond instantaneously to applied biases. Based on this
approximation, the NQS component of \( Q_S'(y, t) \) is zero at
source \((y = 0)\) and drain \((y = L)\) ends, respectively. Since
any profile which vanishes at both ends can be represented
by a Fourier sine series expansion, we’ll express the NQS
component of \( Q_S'(y, t) \) in terms of a Fourier sine series:

\[
Q_S'(y, t) = \sqrt{P_S(t) - P_D(t)} \cdot \frac{y}{L} - F_B V_t + \sum_{n=1}^{10} A_n(t) \cdot \sin \left( \frac{\pi n \cdot y}{L} \right) \tag{7}
\]
\[ P_s(t) = (Q_s(0, t) + F_b V_i)^2 \]
\[ P_d(t) = (Q_s(L, t) + F_b V_i)^2 \]
\[ Q_s(0, t) = V_{GS}(t) \]
\[ Q_s(L, t) = V_{GS}(t) - F_b \cdot (\Psi_{SL}(t) - \Psi_{SO}(t)) \]
\[ V_{GST}(t) = V_{GS}(t) - V_{FB} - \Psi_{SO}(t) - \gamma \sqrt{\Psi_{SO}(t)} - V_i \]

where \( \Psi_{SL} \) and \( \Psi_{SO} \) are surface potentials at the drain and source ends of the channel, respectively, and are computed from the charge sheet formulation using the long channel drain saturation voltage, i.e., \( V_{DSAT} = V_{GST}/F_b + V_i \). \( V_{GST}(t) \) is equivalent to \( (V_{GS} - V_{TH}) \) in the conventional notation, but has a non-negative value due to the charge sheet formulation [14].

The square root term and \( F_b V_i \) term in (7) are QS components of \( Q_s(y, t) \). In this work, 10 terms are used for the Fourier sine series in (7).

Substituting (7) into (5), one can obtain
\[
\frac{\partial G(y, t)}{\partial t} + \sum_{n=1}^{10} \frac{dA_n(t)}{dt} \cdot \sin \left( \frac{n\pi y}{L} \right)
= \frac{\mu_n}{2F_b} \frac{\partial^2}{\partial y^2} \left[ D(y, t) \cdot \sum_{n=1}^{10} A_n(t) \cdot \sin \left( \frac{n\pi y}{L} \right) \right]
\]

where
\[ G(y, t) = \sqrt{P_s(t)} \cdot \sqrt{1 - a(t) \cdot \frac{y}{L} - F_b V_i} \]
\[ D(y, t) = 2\sqrt{P_s(t)} \cdot \sqrt{1 - a(t) \cdot \frac{y}{L}} \\
+ \sum_{n=1}^{10} A_n(t) \cdot \sin \left( \frac{n\pi y}{L} \right) \]
\[ a(t) = 1 - \frac{P_d(t)}{P_s(t)}. \]

\( a(t) \) has a value between 0 and 1; \( a(t) = 0 \) in the cutoff region or when \( V_{GS} = 0 \), and \( a(t) \) is close to 1 in the saturation region. Approximating \( \frac{\partial G(y, t)}{\partial t} \) and \( D(y, t) \) in (13) by a Fourier sine series in \( y \) and a Fourier cosine series in \( y \), respectively, one can obtain

\[
\frac{dG(y, t)}{dt} = \frac{dG(y, t)}{\partial P_s} \cdot \frac{dP_s}{dt} + \frac{dG(y, t)}{\partial P_d} \cdot \frac{dP_d}{dt}
= -\sum_{n=1}^{10} G_{S,n}(t) \cdot \sin \left( \frac{n\pi y}{L} \right) \cdot \frac{dP_s}{dt} \\
- \sum_{n=1}^{10} G_{D,n}(t) \cdot \sin \left( \frac{n\pi y}{L} \right) \cdot \frac{dP_d}{dt}
\]

\[ D(y, t) \approx D_0(t) + 2D_1(t) \cdot \cos \left( \frac{\pi y}{L} \right) \]

Expressions for \( G_{S,n}(t) \), \( G_{D,n}(t) \), \( D_0(t) \), and \( D_1(t) \), and their derivations are shown in Appendix 1. Only two terms are used for the Fourier cosine series expansion in (18) for computational efficiency, which will be explained. Substituting (17) and (18) into (13), and matching the coefficients of each Fourier sine series term on both sides of the equation, one can obtain

\[
\frac{dA_n(t)}{dt} = G_{S,n}(t) \cdot \frac{dP_s(t)}{dt} + G_{D,n}(t) \cdot \frac{dP_d(t)}{dt} \\
- \frac{\mu_n}{2F_b} \cdot \frac{(n\pi)^2}{L^2} \cdot (A_{n-1}(t) \cdot D_1(t) + A_{n+1}(t) \cdot D_1(t)) \\
+ A_n(t) \cdot D_0(t) + A_{n+1}(t) \cdot D_1(t)
\]

for \( n = 1, 2, \cdots, 10 \)

where \( A_0(t) \) and \( A_{11}(t) \) are set to 0 in (19).

Rewriting (19) in matrix notation, one can get the state equation:

\[
\frac{dA(t)}{dt} = D(t) \cdot A(t) + G_s(t) \cdot \frac{dP_s(t)}{dt} \\
+ G_d(t) \cdot \frac{dP_d(t)}{dt}
\]

where \( A(t) \) is a column matrix for the coefficients \( \{A_n(t)\} \), \( D(t) \) is a tridiagonal square matrix, \( G_s(t) \) and \( G_d(t) \) are column matrices for the coefficients \( \{G_{S,n}(t)\} \) and \( \{G_{D,n}(t)\} \) which account for excitations at source and drain ends, respectively. The coefficients of the square matrix \( D(t) \) are

\[
D_{i,i-1}(t) = -\frac{\mu_n}{2F_b} \cdot \frac{\pi^2}{L^2} \cdot i^2 \cdot D_1(t) \]
\[
D_{i,i}(t) = -\frac{\mu_n}{2F_b} \cdot \frac{\pi^2}{L^2} \cdot i^2 \cdot D_0(t) \]
\[
D_{i,i+1}(t) = -\frac{\mu_n}{2F_b} \cdot \frac{\pi^2}{L^2} \cdot (i^2 + 1) \cdot D_1(t)
\]

for \( i = 1, 2, 3, \cdots, 10 \),

but \( D_{0,0}(t) = D_{10,10}(t) = 0 \).

All the other coefficients of the matrix \( D(t) \) except the diagonal term in (22) and the two off-diagonal terms in (21) and (23) are 0.

The number of terms in the Fourier cosine series expansion of \( D(y, t) \) in (18) is important for computational efficiency. If more than 2 terms are used in (18), the matrix \( D \) is no longer a tridiagonal matrix and the total computation time is about twice longer.

Although several integration methods can be used to solve the state equation (20), the trapezoidal integration method is used in this work.

At \( t = 0 \) (the initial time point for the transient analysis), all the coefficients \( \{A_n(t)\} \) for \( n = 1, 2, 3, \cdots, 10 \) are set to 0, since NQS components are 0 at the initial
dc operating point. Old values (values at the previous time point) are used for matrices $D$, $G_y$, and $G_2$ in (20) to make the computation of derivatives of node currents with respect to node voltages easier. So $D$, $G_y$, and $G_2$ are treated as constant matrices during the small time interval considered.

Applying the trapezoidal integration scheme to the state equation (20), one can get

$$
\left( I - \frac{k}{2} D(t_0) \right) \cdot \left( A(t) - A(t_0) \right) = k D(t_0) \cdot A(t_0) + G_3(t_0) \cdot (P_3(t) - P_3(t_0)) + G_2(t_0) \cdot (P_2(t) - P_2(t_0))
$$

(24)

where $I$ is a unit matrix, $t$ is the present time point, $t_0$ is the previous time point, and $k$ is the time step which is equal to $(t - t_0)$. By solving (24), one can find the coefficients $\{A_i(t)\}$ at the new time point $t$. A new time step control scheme is used to keep the local truncation error in approximating (20) into (24) under reasonable range [14].

2.2. Equations for Node Currents and Node Charges

Including the empirical channel length modulation effect, the dc transport current $I_{DC}(t)$ can be computed as [10], [14]

$$
I_{DC}(t) = \frac{W}{L} \cdot C_{OX} \cdot \frac{\mu_s}{2F_B} \cdot (P_3(t) - P_2(t)) \cdot \left( 1 + \lambda \cdot V_{DS} \right)
$$

(25)

where $\lambda$ is the empirical channel length modulation factor and $V_{DS}$ is the applied drain to source voltage. In the derivation of (25), the mobility $\mu_s$ is assumed to be constant for all $y$ [10], [14].

$I_{DC}(t)$ in (1), (2), and (25) can be replaced by any dc current equations. In fact, the SPICE Level-2 dc model with all the short channel effects included has been combined with this NQS charge model and has been implemented in SPICE3, as shown in [19].

Substituting the equation for $Q_d(y, t)$ (7) into (3) and (4), one can find

$$
Q_d(t) = -WL_{OX} \cdot \left[ \sqrt{P_3(t)} \cdot \left( \frac{2}{3} - \frac{1 - (1 - a)^{1.5}}{a^2} \right) \right.
- \frac{2}{5} \cdot \frac{1 - (1 - a)^{2.5}}{a^2} \cdot F_B V'_I
$$

$$
\left. + \sum_{n=1}^{10} (-1)^{(n-1)} \cdot \frac{A_n(t)}{n \pi} \right]
$$

(26)

$$
Q_y(t) = -WL_{OX} \cdot \left[ \sqrt{P_3(t)} \cdot \frac{2}{3} \cdot \frac{1 - (1 - a)^{1.5}}{a^2} \right]
- \frac{2}{5} \cdot \frac{1 - (1 - a)^{2.5}}{a^2} \cdot F_B V'_I
$$

$$
\left. + \sum_{n=1}^{10} (-1)^{(n-1)} \cdot \frac{A_n(t)}{n \pi} \right]
$$

(27)

where $a(t)$ is shown in (16). The argument $(t)$ of $a(t)$ is not shown in (26) and (27) for clarity. Since $a(t)$ appears in the denominator of (26) and (27), (26) and (27) are converted into asymptotic forms for small $a(t)$. Hence, for $a(t) \approx 0$:

$$
Q_d(t) = -WL_{OX} \cdot \left[ \frac{0.5}{\sqrt{P_3(t)}} \cdot \left( 1 - a \cdot \frac{1}{3} \cdot \frac{a}{16} \right) \right.
- \frac{F_B V'_I}{2} + \sum_{n=1}^{10} (-1)^{(n-1)} \cdot \frac{A_n(t)}{n \pi} \right]
$$

(28)

$$
Q_y(t) = -WL_{OX} \cdot \left[ \frac{0.5}{\sqrt{P_3(t)}} \cdot \left( 1 - a \cdot \frac{1}{6} \cdot \frac{a}{48} \right) \right.
- \frac{F_B V'_I}{2} + \sum_{n=1}^{10} \frac{A_n(t)}{n \pi} \right]
$$

(29)

The gate current $I_G(t)$ can be expressed as the time derivative of the gate charge $Q_G(t)$:

$$
I_G(t) = \frac{dQ_G(t)}{dt}
$$

(30)

where

$$
Q_G(t) = WL_{OX} \cdot \int_0^y V_{GB}(t) - V_{FB} - \Psi_S(t) \, dy.
$$

(31)

$Q_G(t)$ can be represented in terms of $Q_d(t)$ and $Q_y(t)$ [14]:

$$
Q_G(t) = WL_{OX} \cdot \left[ V_{GB}(t) - V_{FB} - \Psi_S(t) - \frac{V_{GB}(t)}{F_B} \right]
$$

$$
- Q_d(t) + Q_y(t)
$$

(32)

Derivatives of node currents and node charges with respect to node voltages can be computed from (25)–(29) and (32). These derivatives are used to form the Jacobian matrix for the nonlinear Newton–Raphson iterations in solving circuit equations. The bulk (substrate) current $I_B(t)$ can be computed from the other three current components, $I_D(t)$, $I_S(t)$, and $I_C(t)$, shown in (1), (2), and (30), respectively.

2.3. Moving Boundary Condition

When the transistor is suddenly turned on, the channel charge density profile is nonzero only near the source [10]. During this time period, $I_D(t)$ is zero. We call this time period the "moving-boundary period" and introduce a test condition to ensure that $I_D(t) = 0$ until the end of this time period. Without such a test condition, there will be small but troublesome nonzero $I_D(t)$’s originating from the errors introduced by using the Fourier series with finite terms to represent a two-section profile [10].

The test condition for the moving boundary period is the same as the shown in [10], but different implementation schemes are required since this work is a charge-based model while the work described in [10] is not. The coefficients $\{A_i(t)\}$ computed from (24) are adjusted to ensure that $I_D(t)$ does not change with time and also to
guarantee the smooth transition between the moving-boundary period and the nonmoving-boundary period. This adjustment of \( A_s(t) \) guarantees the continuity of \( Q_s(t) \) between the moving-boundary and the nonmoving-boundary periods without affecting \( Q_2(t) \) and \( I_2(t) \). However, a small discontinuity in \( I_2(t) \) results between those two time periods, since \( I_2(t) \) involves the time derivatives of \( A_s(t) \) as can be seen from (1) and (26), and it could not be made to be continuous without affecting \( Q_s(t) \) and \( I_3(t) \). More details are described in Appendix 2.

### III. Simulation Results and Comparison with Other Models

This work has been implemented in SPICE3 [14], [19], [21], and many example circuits are simulated using this work and the results are compared with those from other models.

#### 3.1. Charge Conservation

Since this work is a charge-based model, it guarantees charge conservation [2], [4]. To demonstrate its charge conservation property, the turn-off transient of a pass transistor has been simulated using this work. Results obtained with a QS charge-based model (BSIM [23]), and the Meyer capacitance model [22] are included for comparison. A train of pulses is applied to the gate. As shown in Fig. 1, the charge-based models mentioned in this paper and the BSIM 0/100 model correctly predict the output voltage which returns to 0 after each clock cycle, while the Meyer capacitance model shows incorrect results due to charge nonconservation.

#### 3.2. Comparison with the Multiple Lump Model

This work has been compared with the QS multiple lump approach for emulating QNS effects, which decomposes each MOSFET into N-MOSFET's in series, with channel length of \( 1/N \) times the original channel length and simulates the resulting circuit using QS models [25].

Fig. 2(a) shows the output voltage waveforms at the turn-off transient of a NMOS pass transistor, predicted by this work and the QS multiple lump model. As \( N \) (number of lumped elements) increases, the behaviors of this QS multiple lump model approach this work. The QS 8-lump model gives almost the same waveform as the QS 4-lump model. To concentrate on the intrinsic phenomenon, overlap and junction capacitances are not included in the simulation. Table I shows the comparison of run statistics for this example. This work takes about the same time as the QS 4-lump model. However, it must be considered that the charge sheet formulation used in the QS part of this work takes about twice longer in model-computation time than the conventional QS SPICE models for the same circuit. It must be noted that, as \( N \) (the number of lump elements) increases, it becomes more difficult to converge as partially reflected in the number of (rejected time points) in Table I.

Fig. 2(b) shows the output voltage waveform of a NQS multiple lump approach where each MOSFET is decomposed into N-MOSFET's in series and the resulting circuit is simulated using this work (a QNS model). The output voltage waveform of the NQS 2-lump model is almost the same as that of the NQS 1-lump model (this work). This clearly demonstrates that this work is QNS.

#### 3.3. Comparison with CODECS

The turn-off transient of a NMOS pass transistor has been simulated using this work, the SPICE Level-2
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>This Work</th>
<th>1-Lump</th>
<th>2-Lump</th>
<th>4-Lump</th>
<th>8-Lump</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (second)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Run</td>
<td>3.1</td>
<td>0.7</td>
<td>1.1</td>
<td>3.1</td>
<td>837</td>
</tr>
<tr>
<td>Number of</td>
<td>591</td>
<td>145</td>
<td>176</td>
<td>422</td>
<td>85 546</td>
</tr>
<tr>
<td>Iterations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transient Time</td>
<td>277</td>
<td>70</td>
<td>65</td>
<td>100</td>
<td>14 617</td>
</tr>
<tr>
<td>Points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rejected Time</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>3645</td>
</tr>
<tr>
<td>Points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Load Time (sec)</td>
<td>1.9</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>613</td>
</tr>
</tbody>
</table>

Fig. 3. Comparison of this work (SPICE Level-2 dc model + NQS charge model) with CODECS [26] and the SPICE Level-2 charge-based model [24], for the output error voltages versus the fall time of the gate pulse, at the turn-off transient of a NMOS pass transistor. W/L of the NMOSFET is 20 μm / 3 μm (drawn dimension). Model parameters used for this work and the SPICE Level-2 charge-based model are extracted from the dc current characteristics of the CODECS simulation and are as follows. (LEVEL = 2, VDD = 1.13 V, VDS = 0.050 μm, NGS = 2 × 10^6 cm^-2, UT = 800, LD = 0.4 μm, GAMMA = 1.34 × V, NPS = 5 × 10^15 cm^-3, VDD = 85 K-m/s, NDD = 2, CGG = 276 pF, CG = 276 pF/m, XJ = 0.2 μm).

For the slow turn-off (large T_F), the node S is connected to ground via a high impedance path (2-pF capacitor), and the node D is connected to ground via a relatively low impedance path (10K resistor). Hence, most of the channel charge is injected to the node D and very little channel charge is injected to the node S. In this case, all models show the small output error voltage for the slow turn-off (large T_F) in Fig. 3.

For the fast turn-off (T_F is less than 20 ns (RC time constant of 2-pF capacitor and 10K resistor)), the impedance between the node D and ground becomes comparable to that between the node S and ground. Hence, half of the channel charge is injected to the node D, as shown by this work and CODECS in Fig. 3. However, the SPICE Level-2 charge-based model with XQC = 0.0 gives an output error voltage much larger than that of this work because it assigns most of the channel charge to the source node S. For the SPICE Level-2 charge-based model with XQC = 0.4999, about half of the channel charge is assigned to each of the drain or source nodes. Hence, this is physically correct for the fast turn-off and it gives good agreement with both this paper and CODECS. Therefore, XQC = 0.4999 is the good partitioning scheme for this particular example.

For the extremely fast turn-off (T_F < 40 ps), this work deviates slightly from CODECS, which is believed to be correct. Since the channel transit time τ_F of the MOSFET in Fig. 3 with VGS = 5 V is around 30 ps, T_F = 40 ps approximately corresponds to |dVGS/dt| = (VGS - VTH)/τ_F, where VGS = 5 V in this example. Hence, from this example, we can observe that this work is valid for signals of |dVGS/dt| ≤ (VGS - VTH)/τ_F.

3.4 Turn-On Transient of a NMOSFET

Fig. 4(a) shows the circuit diagram to evaluate the node current and the node charges for the turn-on transient of a NMOSFET. A rising ramp voltage is applied at gate and V_{D} = 2 V, V_{B} = V_{S} = 0. Initially the MOSFET is in the cutoff region and enters the saturation region and then the linear region as time goes on. This turn-on transient has been simulated using this work, QS charge-based models (BSIM [23]), PISCES (2-D device simulation) [27], and the numerical solution of the 1-D current continuity equation [20]. The ratios 40/60, 0/100, and 50/50 represent the ratios between drain and source charges in the saturation region. Conceptually, the BSIM 40/60 model is the same as the Ward-Dutton model [2] and the BSIM 0/100 model is the same as the Yang-Chatterjee model [4].

Fig. 4(b) shows the drain current waveform for the turn-on transient. At t = 0, all models predict the same drain current which is slightly negative due to the feedthrough current through the gate–drain overlap capacitance. At t = 0.35 ns when V_{GS} reaches V_{TH}, there are sudden negative jumps in drain currents for the BSIM 50/50 and the BSIM 40/60 models. This sudden negative jump is caused by the assignment of fixed ratio (50% for the BSIM 50/50 and 40% for the BSIM 40/60 model) of channel charge to the drain node without considering the channel
transit time effect. However, in this paper, PISCES and the 1-D numerical simulation all show that the drain current does not change with time until $t = 1.17$ ns when the inversion carriers which started from the source node at the turn-on time ($t = 0.35$ ns) reach the drain node. The channel transit time for this example is $(1.17 - 0.35) = 0.82$ ns. The drain current component of the BSIM 0/100 model in the saturation region is purely dc (transport) since the drain charge is 0 and so the displacement current component is 0 in the saturation region. This paper, PISCES, and the 1-D numerical solution all give similar results, except for a small discontinuity in drain current of this paper at $t = 1.17$ ns. This small discontinuity of drain current between the moving boundary period and the nonmoving-boundary period is caused by the adjustment of coefficients $\{A_n\}$ as explained in Section II-2.3.
Although this discontinuity in drain current degrades the convergence property of circuit simulation, one must note that there are big discontinuities in drain current at the turn-on time $t = 0.35$ ns for the BSIM 40/60 and the BSIM 50/50 models as shown in Fig. 4(b), and yet both BSIM 40/60 and BSIM 50/50 models work fine in real circuit simulations [2], [23], [24].

For $t > 2$ ns, the applied bias remains constant. This work shows that the drain current should still be changing, which is verified by PISCES and the 1-D numerical model. All BSIM models, however, predict that the drain current does not change with time.

Fig. 5(b) shows the intrinsic drain charge waveform normalized by $W/L_{	ext{ox}}$. It does not include any charge components due to the overlap or junction capacitance. This work and the 1-D numerical simulation agrees with each other for all time points. This work and the BSIM 40/60 model gives the same drain charge at the steady state ($t = 3$ ns) but this work shows some delay in charge buildup during the time interval before $t = 2$ ns. At the steady state ($t = 3$ ns), the BSIM 0/100 model gives smaller drain charge than this work and the BSIM 50/50 model gives larger drain charge than this work.

Fig. 4(d) shows the source current waveform. This work, PISCES, and the 1-D numerical simulation give good agreements for all time points. All BSIM models give a sudden negative jump at the turn-on time ($t = 0.35$ ns). The BSIM 0/100 model gives the largest jump among the three BSIM models considered, because 100% of the channel charge is assigned to the source node in the saturation region for the BSIM 0/100 model.

Fig. 4(e) shows the intrinsic source charge waveform. Again, this work and the 1-D numerical simulation give good agreement with each other. The steady-state source charge at $t = 3$ ns is the same for this paper, the 1-D numerical simulation, and the BSIM 40/60 model. However, this work and the 1-D numerical simulation show a delay in the charge buildup compared to the BSIM 40/60 model. Among the models considered, the BSIM 0/100 model gives the largest magnitude of source charge while the BSIM 50/50 model gives the smallest.

Although the comparisons with other NQS models are not shown in this paper, as far as the current waveforms are concerned, all the other models in [8]–[12] and [19] showed qualitative agreements with this work.

3.5. Simulation of CMOS Inverter Chain

Delay times of a 10-stage CMOS inverter chain have been simulated using this work and other Q5 models and the results have been compared. Model parameters have been adjusted to guarantee the same dc characteristics among all the models considered. Fig. 5(a) shows the dc transfer curve and the dc supply current of a single CMOS inverter simulated using this paper, BSIM [23] (40/60, 0/100), and the Meyer model. The Meyer model refers to the SPICE Level-2 dc model [24] with the Meyer capacitance model [22]. Good agreements among models have been obtained in dc characteristics.

Fig. 5(b) shows the output voltage waveforms after 10 stages of CMOS inverter chain for the rising pulse input and the falling pulse input, respectively, as shown in the inset. Fig. 5(c) and (d) show the output voltage waveforms with a load capacitor (0.5 pF) in Fig. 5(c) and 3.0 pF in Fig. 5(d) connected at each inverter output node.

When the delay is limited by gate capacitances as in Fig. 5(b), this paper gives the smallest delay time and the BSIM 40/60 model gives the largest delay time. The reason why this paper gives the smallest delay time is probably due to the fact that the effective loading by the gate capacitance is much smaller in this work than in the Q5 BSIM models for the fast transient as shown in [10] and [14].

When the delay time is limited by stray capacitances as in Fig. 5(d) the maximum gate capacitance of an inverter (0.23 pF) << the load capacitance (3.0 pF), this work gives the largest delay time due to the NQS inertia in the driving current of this work as can be seen in Fig. 4(b).

In the intermediate case in Fig. 5(d), this work gives an intermediate delay time which is almost the same as that of BSIM 0/100 model.

3.6. Simulation of a Differential Sample-Hold Circuit

Fig. 6(a) shows the circuit diagram of a differential sample-hold circuit commonly used in MOS A/D converter circuits [7]. Input voltage waveforms are shown in Fig. 6(b) and (c).

When $V_{R1}$ and $V_{R2}$ are 5 V and $V_{R3}$ is -5 V, input voltages $V_{in.p}$ and $V_{in.m}$ are stored (sampled) in capacitors C1 and C3, respectively, and the output voltage $V_{OUT}$ is zero. We call this time period the ‘‘sampling time.’’ When $V_{R1}$ and $V_{R2}$ become -5 V and $V_{R3}$ becomes 5 V, the charges stored in capacitors C1 and C3 are dumped into capacitors C2 and C4, respectively. $V_{OUT}$ becomes equal to the sampled input voltage $(V_{in.p} - V_{in.m})$ in an ideal manner. During this time period, the output voltage does not change with time even when the input voltage is changing. We call this time period the ‘‘hold time.’’ However, due to charge injection, there are some errors between the output voltage at hold time and the sampled input voltage. Fig. 6(c) shows three kinds of differential input voltage waveforms: rising, falling, and dc inputs. Rising and falling inputs are a half-period of sine wave with frequency of 25 MHz and amplitude of 2 V. $V_{dc}$ is the superimposed dc voltage. In all cases, $V_{in.p}$ and $V_{in.m}$ have the same magnitude but with the opposite polarity. Ideal operational amplifiers are used to concentrate on the charge injection problem. The voltage gain of the operational amplifier is 80 dB throughout the frequency range considered and the input impedance is infinitely large and the output impedance is zero.

As shown in Fig. 6(b), M2 and M4 are turned off before $M1$ and $M3$ to eliminate the effect of charge injection of $M1$ and $M3$ on the output voltage [7].

Three time phases (1), (2), and (3) are shown in Fig. 6(b). The time phase (1) is the time period between $t = 10$ ns and $t = 11$ ns, during which $V_{R1} = 5$ V and $V_{R3} =$
Fig. 5. (a) Comparison of the dc transfer curve and the supply current of a CMOS inverter simulated using this work and other QS models. BSIM 23 is used for QS models. Model parameters for the NMOSFET are $V_{th} = -0.77 \ V$, $T_{ox} = 18 \ \text{nm}$, $N_{SUA} = 2 \cdot 10^{12} \ \text{cm}^{-2}$, $U_{d} = 500 \ \text{cm}^2/(\text{V} \cdot \text{s})$, $\lambda = 0.03 \ \text{V}^{-1}$, $C_{GSO} = 150 \ \text{pF/M}$, $C_{GSO} = 150 \ \text{pF/M}$, $C_{J} = 3 \cdot 10^{-4} \ \text{F/M}$, $C_{J} = 8 \cdot 10^{-10} \ \text{F/M}$. Model parameters for the PMOSFET are $V_{th} = 0.22 \ \text{V}$, $T_{ox} = 18 \ \text{nm}$, $N_{DUG} = 6 \cdot 10^{15} \ \text{cm}^{-2}$, $U_{d} = 180 \ \text{cm}^2/(\text{V} \cdot \text{s})$, $\lambda = 0.05 \ \text{V}^{-1}$, $C_{GSO} = 150 \ \text{pF/M}$, $C_{GSO} = 150 \ \text{pF/M}$, $C_{J} = 2 \cdot 10^{-4} \ \text{F/M}$, $C_{J} = 5 \cdot 10^{-10} \ \text{F/M}$. (b) Comparison of the output voltage waveform after 10 stages of CMOS inverter chain for the rising pulse input and the falling pulse input, respectively, with no load capacitors. (c) Comparison of the output voltage waveform after 10 stages of CMOS inverter chain for the rising pulse input and the falling pulse input, respectively, with a 0.5-pF load capacitor connected at the output node of each inverter. (d) Comparison of the output voltage waveform after 10 stages of CMOS inverter chain for the rising pulse input and the falling pulse input, respectively, with a 3.0-pF load capacitor connected at the output node of each inverter.

$-5 \ \text{V}$, and $V_{P3}$ changes from $5 \ \text{V}$ to $5 \ \text{V}$. The time phase (2) is the time period between $t = 12 \ \text{ns}$ and $t = 13 \ \text{ns}$ during which $V_{P1}$ changes from $5 \ \text{V}$ to $5 \ \text{V}$ and $V_{P2} = -5 \ \text{V}$ and $V_{P3} = -5 \ \text{V}$. The time phase (3) is the time period between $t = 14 \ \text{ns}$ and $t = 15 \ \text{ns}$, during which $V_{P1} = V_{P2} = -5 \ \text{V}$ and $V_{P3}$ changes from $-5 \ \text{V}$ to $5 \ \text{V}$.

Fig. 6(d) shows the differential output voltage waveform predicted by this work, the BSIM 0/100 and 50/50 models, for the dc input ($V_{in, p} = 1 \ \text{V}$ and $V_{in, n} = -1 \ \text{V}$). The BSIM 40/60 model is not shown for clarity. The output voltage is reset to 0 during the sampling time ($t < 10 \ \text{ns}$) and moves toward 2 \ \text{V} during the hold time ($t > 15 \ \text{ns}$).

Fig. 6(e) shows the output error voltage at $t = 20 \ \text{ns}$ for three kinds of input voltage waveforms shown in Fig. 6(c). $V_{p}$ is the superimposed dc differential input voltage. An exact symmetry with respect to the center point can be observed, due to the differential symmetry of the circuit configuration. No overlap or junction capacitance is included in the simulation. Inclusion of those stray capacitances ruins the exact symmetry. This work gives results which are close to those of BSIM 40/60 model, but are quite different from those of BSIM 0/100 model.

For the dc input voltages ($V_{in}$: dc), all models show that the output error voltage is a nonlinear function of dc input voltage $V_{o}$. The BSIM 0/100 model gives the smallest output error voltage.

For ($V_{in}$: rising), all models predict that the output error voltage is shifted in the positive direction. This is partially due to the fact that the input voltage increases with time during the sampling period ($10 \ \text{ns} < t < 10.45 \ \text{ns}$).
and the real sampled voltage is slightly larger than $V_o$.
However, the difference in output error voltages between $V_{in}$: dc and and $V_{in}$: rising is not constant for different $V_o$'s.

3.7. Computational Costs

Table II shows the comparison of single model computation time which has been calculated by dividing the (load time) from SPICE3 [21] output listings by ((total number of iterations) * (number of MOSFET's in the circuit)) from the simulations of 11-stage CMOS inverter chain (Fig. 5) and ring oscillator. This assumes that the model computation time of extra devices in the circuit is negligible. (Note that there are only two extra devices (independent voltage source) in comparison with 22 MOSFET's in the 11-stage CMOS inverter chain and ring oscillator.) The (load time) includes both the model computation time and the computation time of the circuit matrix coefficients, but does not include either the matrix solution time or the input/output processing time. This work implemented in SPICE3 takes about 2.8 times longer than QS models (BSIM and Level-2 Meyer) in SPICE3.

Table III shows the run statistics comparison for the simulation of an 11-stage CMOS inverter chain (Rising $V_{in}$ (Fig. 5(b)). The simulation interval is 10 ns and the maximum time step is set to 0.01 ns for all models con-
TABLE III
COMPARISON OF RUN STATISTICS FOR THE SIMULATION OF THE 11-STAGE CMOS INVERTER CHAIN (RISING VIN) SHOWN IN FIG. 5(b)

<table>
<thead>
<tr>
<th>Model</th>
<th>This Work</th>
<th>BSIM 40/60</th>
<th>BSIM 0/100</th>
<th>Level-2 Meyer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Run Time (seconds)</td>
<td>176</td>
<td>66</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>Number of Iterations</td>
<td>2293</td>
<td>2356</td>
<td>2247</td>
<td>2262</td>
</tr>
<tr>
<td>Transient Time Points</td>
<td>1024</td>
<td>1013</td>
<td>1011</td>
<td>1015</td>
</tr>
<tr>
<td>Rejected Time Points</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Load Time (seconds)</td>
<td>141</td>
<td>55</td>
<td>51</td>
<td>59</td>
</tr>
</tbody>
</table>

TABLE IV
COMPARISON OF RUN STATISTICS FOR THE SIMULATION OF AN 11-STAGE CMOS RING OSCILLATOR

<table>
<thead>
<tr>
<th>Model</th>
<th>This Work</th>
<th>BSIM 40/60</th>
<th>BSIM 0/100</th>
<th>Level-2 Meyer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Run Time (seconds)</td>
<td>1249</td>
<td>145 (470)</td>
<td>119 (538)</td>
<td>152 (436)</td>
</tr>
<tr>
<td>Number of Iterations</td>
<td>17 953</td>
<td>5250 (18 715)</td>
<td>4673 (20 480)</td>
<td>5793 (16 976)</td>
</tr>
<tr>
<td>Transient Time Points</td>
<td>5703</td>
<td>1265 (5668)</td>
<td>1306 (6950)</td>
<td>1101 (4100)</td>
</tr>
<tr>
<td>Rejected Time Points</td>
<td>522</td>
<td>403 (1845)</td>
<td>416 (2190)</td>
<td>356 (1358)</td>
</tr>
<tr>
<td>Load Time (seconds)</td>
<td>1055</td>
<td>128 (404)</td>
<td>102 (460)</td>
<td>135 (385)</td>
</tr>
</tbody>
</table>

TABLE V
COMPARISON OF RUN STATISTICS FOR THE SIMULATION OF THE DIFFERENTIAL SAMPLE-HOLD CIRCUIT SHOWN IN FIG. 6(e)

<table>
<thead>
<tr>
<th>Model</th>
<th>This Work</th>
<th>QS 40/60</th>
<th>QS 0/100</th>
<th>QS 50/50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Run Time (seconds)</td>
<td>26.8</td>
<td>8.8</td>
<td>8.4</td>
<td>11.1</td>
</tr>
<tr>
<td>Number of Iterations</td>
<td>1390</td>
<td>854</td>
<td>855</td>
<td>1021</td>
</tr>
<tr>
<td>Transient Time Points</td>
<td>533</td>
<td>339</td>
<td>346</td>
<td>364</td>
</tr>
<tr>
<td>Rejected Time Points</td>
<td>38</td>
<td>70</td>
<td>74</td>
<td>89</td>
</tr>
<tr>
<td>Load Time (seconds)</td>
<td>20.4</td>
<td>5.5</td>
<td>5.7</td>
<td>7.0</td>
</tr>
</tbody>
</table>

considered. The total runtime of this work is about 2.5–3 times longer than QS models.

Table IV shows the run statistics comparison for the 11-stage CMOS ring oscillator simulation, where the time step is not forced but is determined by the time step control scheme. For this work the time step control parameters, (minton) and (qtrtol), are set to 1 and 80, respectively [14]. For other models, the time step control parameter (reitol) [28] is set to $10^{-3}$ and $10^{-4}$, respectively. Values for (reitol) = $10^{-4}$ are shown inside the parentheses in Table IV. With (reitol) = $10^{-3}$ for QS models, (transient time points) of this paper is much larger than those of QS models and this work takes about 8–10 times longer than QS models. However, with (reitol) = $10^{-4}$ for QS models, (transient time points) are comparable in all models and this work takes about 2–3 times longer than QS models.

Table V shows the run statistics comparison of the differential sample–hold circuit (Fig. 6) simulation where time steps are not forced but are determined by time step control schemes. With (minton) = 1, (qtrtol) = 30, and (reitol) = $10^{-4}$, this work takes about 2–3 times longer than QS models.

V. CONCLUSION

An analytic charge conserving NQS model for long channel MOSFET’s has been derived and implemented in SPICE3. It is based on an approximate solution to the current continuity equation. Analytic equations are derived for node charges to guarantee charge conservation. Simulation results of some test circuits using this paper and other models have been shown. Excellent agreements have been observed among this paper, PISCES (2-D device simulation) [27], the numerical solution of the 1-D current continuity equation [20], the multiple lump model [25], and CODECS (a mixed device and circuit simula-

tion) [26]. However, large differences have been observed between this work and conventional QS models (BSIM [23], Level-2 Meyer [24]).

Although the NQS effect on overall circuit performance is not large in digital circuits as shown in the inverter chain simulation (Fig. 5), this model is expected to be crucial for the simulation of switched analog circuits (Figs. 3 and 6) and other analog circuits where what kind of channel charge partitioning scheme to use is not clear or the channel transit time effect is important.

The model computation time of this work implemented in SPICE3 is about 2–3 times longer than those of other QS MOSFET models in SPICE3 (BSIM, Level-2 Meyer).

Also, this NQS charge model has been combined with the SPICE Level-2 [24] dc model and has been implemented in SPICE3 [19]. In this way, all the short channel effects considered by the SPICE Level-2 dc model have been included in the implementation.

APPENDIX 1

DERIVATION OF $\{G_{Na}(t)\}, \{G_{Nn}(t)\}, D_{c}(t),$ and $D_{n}(t)$

$\frac{\partial G(y, t)}{\partial t}$ in (13) is approximated into a Fourier sine series in (17). For this purpose, $G(y, t)$ in (14) is approximated into (A1.1) by expanding the square root term in
(14) as a Taylor series with three terms. Hence:

\[
G(y, t) = \sqrt{P(t)} \cdot \left[ 1 - \frac{a(t)}{2} \cdot \frac{y}{L} - b(t) \cdot \left( \frac{y}{L} \right)^2 \right]
- F_y V_t
\]

(A.1.1)

where

\[
b(t) = 3 - \frac{3}{4} \cdot a(t) - 2 \cdot \frac{1 - (1 - a(t))^{1.5}}{a(t)}
\]

(A.1.2)

where \(a(t)\) is shown in (16), and \(b(t)\) has been chosen such that it guarantees the integration of the right-hand side of (14) with \(y\) from \(y = 0\) to \(y = L\) is equal to the integration of the right-hand side of (A.1). \(b(t)\) is reduced to \((a(t))^2 / 8\) for \(a(t) = 0\). \(D(y, t)\) in (15) is approximated as a Fourier cosine series so that the multiplication of a cosine series and a sine series in the right-hand side of (13) gives another sine series and it enables the matching of the sine series coefficients in both sides of (13). Hence, \(D(y, t)\) can be written as (18). \(D(t)\) in (18) is the average value of \(D(y, t)\) between \(y = 0\) and \(y = L\). Hence:

\[
D(t) = \frac{1}{L} \int_0^L D(y, t) \, dy.
\]

(A.1.3)

Substituting (15) into (A.1.3), one can get

\[
D(t) = \frac{4}{3} \cdot \sqrt{P(t)} \cdot \frac{1 - (1 - a(t))^{1.5}}{a(t)}
+ \frac{2}{\pi} \cdot \sum_{n=1}^{\infty} \frac{A_{2m-1}(t)}{2m - 1}
\]

\[
= 2 \sqrt{P(t)} \cdot \left( 1 - \frac{a(t)}{4} \right) + \frac{2}{\pi} \cdot \sum_{n=1}^{\infty} \frac{A_{2m-1}(t)}{2m - 1}
\]

for \(a(t) = 0\).

(A.1.4)

\[\text{where } H(a) = a^{0.25}.\]

(A.1.6)

The function \(H(a)\) is added to set \(D_t = 0\) when \(a = 0\), that is, when the MOSFET is in the cutoff region or when \(V_{DS} = 0\). The exponent 0.25 in (A.1.6) is a heuristic factor. \(D_t\) is limited to have values within the range \([-0.5 \cdot D(t), +0.5 \cdot D(t)]\) to guarantee \(D(y, t) \geq 0\) for all \(y(0 \leq y \leq L)\), and hence, the numerical stability of the solution method.

Substituting (18) into (13), one can find

\[
\frac{\partial G(y, t)}{\partial P} \cdot \frac{dP}(t) + \frac{\partial G(y, t)}{\partial D} \cdot \frac{dD}(t)
- \frac{dF_y}{dt} \cdot V_t + \sum_{s=1}^{10} \frac{dA_s}{dt} \cdot \sin \left( \frac{n \pi \cdot y}{L} \right)
\]

\[
= - \frac{\mu_p}{2 F_y} \cdot \sum_{n=1}^{10} A_n(t) \cdot \left[ \left( n - 1 \right) \pi \frac{D(t)}{L^2} \cdot \sin \left( \frac{n \pi \cdot y}{L} + \left( n + 1 \right) \pi \right) \right]
\]

\[
\text{Expanding } \left( \frac{y}{L} \right)^2 \text{ in (A.1.3) as Fourier sine series, one can represent } \frac{\partial G(y, t)}{\partial P} \text{ and } \frac{\partial G(y, t)}{\partial D}
\]

in (A.1.7) as Fourier sine series. Hence:

\[
\frac{\partial G(y, t)}{\partial P} = - \sum_{n=1}^{10} G_{S_n}(t) \cdot \sin \left( n \pi \cdot \frac{y}{L} \right)
\]

(A.1.8)

\[
\frac{\partial G(y, t)}{\partial D} = - \sum_{n=1}^{10} G_{D_n}(t) \cdot \sin \left( n \pi \cdot \frac{y}{L} \right)
\]

(A.1.9)

where

\[
G_{S_n}(t) = \frac{1}{\sqrt{P(t)}} \cdot \left[ \frac{1}{n \pi} - \left( 1 - \frac{a}{2} \right) \cdot \frac{(-1)^{n+1}}{n \pi} - \frac{b + 2(1 - a) \cdot \frac{db}{da}}{n \pi} \cdot \frac{(-1)^n}{(n \pi)^3} \cdot \left( 1 - (-1)^n \right) \right]
\]

\[
G_{D_n}(t) = - \frac{1}{\sqrt{P(t)}} \cdot \frac{(-1)^{n+1}}{n \pi} \cdot \left( 1 + \frac{db}{da} \right)
\]

(A.1.10)

\[
\frac{2}{(n \pi)^3} \cdot (1 - (-1)^n)
\]

(A.1.11)

In (A.1.10) and (A.1.11), the argument \(n\) is dropped from \(a(t)\) and \(b(t)\) for clarity, and \(db/da\) is the derivative of \(b(t)\) in (A.1.2) with respect to \(a(t)\). Using (A.1.8) and (A.1.9), one can derive (17).

APPENDIX 2

\[Q_D(t) \text{ and } \{A_0(t)\} \text{ DURING THE MOVING-BOUNDARY PERIOD}\]

During the moving-boundary period, the drain current does not change with time. Hence:

\[I_D(t) = I_{DC}(t) + I_{D}(t) = I_D(t_0)\]

(A2.1)

where

\[I_D(t) = \frac{dQ_D(t)}{dt}\]

(A2.2)
where \( t \) is the present time point, and \( t_0 \) is the previous time point. The dc transport current \( I_{DC}(t) \) can be determined by (25) as a function of applied biases (node voltages) only. To compute the displacement current \( I_{DR}(t) \) from (A2.2), different orders of integration schemes can be used. Two most commonly used integration schemes, backward Euler and trapezoidal integration schemes, are considered in the following.

A. Backward Euler Integration Scheme

If the backward Euler integration (order 1) is used, (A2.2) can be rewritten as

\[
I_{DR}(t) = \frac{Q_D(t) - Q_D(t_0)}{k} \tag{A2.3}
\]

where \( k \) is the time step which is equal to \((t - t_0)\). Substituting (A2.3) into (A2.1), one can find the new drain charge \( Q_D(t) \) during the moving boundary period, as

\[
Q_D(t) = Q_D(t_0) + k \cdot (-I_{DC}(t) + I_D(t_0)) \tag{A2.4}
\]

New drain capacitances can be computed from (A2.4) as

\[
C_{DX}(t) = \frac{\partial Q_D(t)}{\partial V_X} = -k \cdot \frac{\partial I_{DC}(t)}{\partial V_X},
\]

where \( X = G, D, S, B. \) \tag{A2.5}

\( \partial I_{DC}(t)/\partial V_X \) in (A2.5) can be computed from (25).

B. Trapezoidal Integration

If the trapezoidal integration (order 2) is used, (A2.2) can be rewritten as

\[
I_{DR}(t) = -I_{DR}(t_0) + \frac{k}{2} \cdot (Q_D(t) - Q_D(t_0)) \tag{A2.6}
\]

Substituting (A2.6) into (A2.1), one can find

\[
Q_D(t) = Q_D(t_0) + \frac{k}{2} \cdot (-I_{DC}(t) + I_D(t_0) + I_D(t_0)) \tag{A2.7}
\]

New drain capacitances can be derived from

\[
C_{DX}(t) = \frac{\partial Q_D(t)}{\partial V_X} = -\frac{k}{2} \cdot \frac{\partial I_{DC}(t)}{\partial V_X},
\]

where \( X = G, D, S, B. \) \tag{A2.8}

C. Adjustment of Coefficients \( A_i(t) \)

Coefficients \( \{A_i(t)\} \) in (7), and (26)–(29) are adjusted to obtain the smooth transition between the moving-boundary period and the nonmoving-boundary period. This adjustment of \( \{A_i(t)\} \) guarantees the continuity of \( Q_D(t) \) between the moving-boundary period and the nonmoving-boundary period without changing \( Q_S(t) \):

\[
A_{2m-1}(t) = A_{2m-1}(t) - 0.5 \cdot \pi \cdot \frac{(Q_D(t) - Q_D(t_0))}{(2m - 1) \cdot WLC_{OX} \cdot \sum_{i=1}^{5} \frac{1}{(2i - 1)^2}}
\]

\[\text{for } m = 1, 2, 3, 4, 5. \tag{A2.9}\]

\[
A_{2m}(t) = A_{2m}(t) + 0.5 \cdot \pi \cdot \frac{(Q_D(t) - Q_D(t_0))}{(2m-1) \cdot WLC_{OX} \cdot \sum_{i=1}^{5} \frac{1}{(2i)^2}}
\]

\[\text{for } m = 1, 2, 3, 4, 5. \tag{A2.10}\]

\( A_{2m}^{(t)} \) and \( A_{2m-1}^{(t)} \) are computed from the solution of (24), \( Q_D(t) \) is computed from (26) or (28) and \( Q_D(t) \) is computed from (A2.4) or (A2.7). Although this scheme guarantees the continuity of \( Q_D(t) \) between the moving boundary period and the nonmoving boundary period, the displacement current \( dQ_D/dt \) is discontinuous between those two time periods, as shown by the drain current discontinuity of this work in Fig. 4(b).

Acknowledgment

H. J. Park would like to express his special thanks to Prof. P. R. Gray and D. A. Hodges, University of California at Berkeley, for their encouragement in this paper, and to a graduate student B. S. Kim for discussions.

References


[15] M. A. Critt, "The Meyer model revisited: Why is charge not con-
nonquasistatic model for the four-terminal MOSFET valid in all re-
(19) H. J. Park, P. K. Ko, and C. Hu, "SPICE3 implementation of a non-
UCB/ERL M89/70, Electron. Res. Lab., Univ. of California, Berke-
ley, June 8, 1989.
(21) T. L. Quartes, A. R. Newton, D. O. Pederson, and A. Sangiovanni-
Univ. of California, Berkeley, May 1985.
[24] A. Vladimirescu and S. Liu, "The simulation of MOS integrated cir-
cuits using SPICE2G.6," Memo. No. UCB/ERL M80/7, Electron.
circular geometry MOS pass transistors," IEEE J. Solid-State Cir-
[26] K. Magee, "CODECS: A mixed-level circuit and device simula-
EECS, Univ. of California, Berkeley, Nov. 21, 1988.
Manual, Stanford Electronics Lab., Stanford Univ., Palo Alto, CA,
1984.
[28] L. W. Nagel, "SPICE2: A computer program to simulate semi-
of California, Berkeley, CA, May 9, 1975.

Hong-June Park received the B.S. degree in electronics engineering from Seoul National Uni-
versity in 1979, the M.S. degree in electrical en-
ingineering from K.A.I.S.T., Seoul, Korea, in
1981, and the Ph.D. degree in electrical engi-
neering and computer sciences from University of California Berkeley in 1989.
From 1981 to 1984, he worked as a research
staff member in the Design Automation Labora-
tory, Electronics Telecommunication Research Institute, Daejon, Korea, where he worked on
SPICE MOSFET model parameter extraction system. Since 1989, he has been with Technology CAD, Intel, Santa Clara, CA, where he is working
on the next-generation BIT and MOSFET models for circuit simulation. His research interests include semiconductor device modeling, intercon-
nect and package modeling, circuit simulation, electrical, and computer-
aided designs of integrated circuits.

Ping Keung Ko (M’89) received the B.S. degree in physics with special honors from Hong Kong University, and the M.S. and Ph.D. degrees in
electrical engineering from the University of Cal-
ifornia at Berkeley, in 1978 and 1982, respec-
tively.
In 1982 and 1983, he was a Member of Tech-
nical Staff at Bell Labs, Holmdel, NJ, where he
was responsible for the development of high-speed
MOS technologies for communication circuits. He
joined the Berkeley faculty in 1984, where he is
now a Professor of electrical engineering and computer sciences. His pre-
sent research interests include high-speed VLSI technologies and devices,
device modeling for circuit simulation, and electronic neural networks. He
has authored or co-authored one book and over 100 research papers. Cur-
rently, he is the director of the Berkeley Microlaboratory.

Dr. Ko has served on the program committees of the International VLSI
He has also been Associate Editor of IEEE TRANSACTIONS ON ELECTRON
DEVICES since March 1988.

Chenming Hu (S’71-M’76-SM’83-F’90) received the B.S. degree from the National Taiwan
University and the M.S. and Ph.D. degrees in
electrical engineering from University of Cal-
ifornia, Berkeley, in 1979 and 1983, respectively.
From 1973 to 1976 he was an Assistant Profes-
sor at Massachusetts Institute of Technology.
In 1976 he joined the University of California,
Berkeley, a Professor of Electrical Engineering
and Computer Sciences. While on industrial leave from the University in 1980–81 he was manager of
nonvolatile memory development at National Semiconductor. Since 1973 he
has served as a consultant to the electronics industry. He has also been
an adviser to many government and educational institutions. His present
research areas include VLSI devices, hot electron effects, thin dielectrics,
device reliability, nonvolatile semiconductor memories, power semicon-
ductor devices, and GaAs devices. He has also been awarded several
patents on semiconductor devices and technology. He has authored or co-
authored two books and over 200 research papers. He is also Honorary
Professor of Beijing University, China and Director of the Joint Services
Electronics Program at Berkeley.

Dr. Hu was an Associate Editor of IEEE TRANSACTIONS ON ELECTRON
DEVICES from 1986 to 1988, and Vice Chairman of IEEE Electron Devices
Society, Santa Clara Valley Chapter, 1980–1982. He was appointed the
first National Science Council Invited Chair Lecturer, Republic of China,
in 1987. He is a Board Chairman of East San Francisco Bay Chinese School.
He has delivered a dozen keynote addresses and invited papers at scientific
conferences, and has received four best paper awards. He has been listed in
American Men and Women of Science, California Men and Women of
Science and Technology, and Who’s Who in Technology.