AN ACCURATE MOSFET INTRINSIC CAPACITANCE MODEL CONSIDERING QUANTUM MECHANIC EFFECT FOR BSIM3v3.2

by

Weidong Liu, Xiaodong Jin, Ya-Chin King
And Chenming Hu

Memorandum No. UCB/ERL M98/47

21 July 1998

cover

AN ACCURATE MOSFET INTRINSIC CAPACITANCE MODEL CONSIDERING QUANTUM MECHANIC EFFECT FOR BSIM3v3.2

by

Weidong Liu, Xiaodong Jin, Ya-Chin King and Chenming Hu

Memorandum No. UCB/ERL M98/47

21 July 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

An Accurate MOSFET Intrinsic Capacitance Model Considering Quantum Mechanic Effect For BSIM3v3.2

Weidong Liu, Xiaodong Jin, Ya-Chin King, and Chenming Hu

Electronics Research Laboratory
Department of Electrical Engineering and Computer Sciences
University of California at Berkeley, Berkeley, CA 94720-1772

Abstract

As the gate oxide thickness is scaled down, charge thickness due to quantization effects must be considered in MOSFET intrinsic capacitance modeling for accurate circuit simulation. We present a novel analytical charge thickness model from numerical self-consistent solution of Schrodinger, Poisson and Fermi-Dirac equations. Based on the charge layer thickness model, we develop a compact MOSFET intrinsic capacitance model which accounts for the charge thickness throughout all operating regions of MOSFET's. Its universality and accuracy are demonstrated with many technologies including TiN gate and $SiON/Ta_2O_5$ dielectric with equivalent SiO_2 thickness of 18Å. Extensive tests demonstrate significantly improved computational efficiency and convergence properties over state-of-the-art compact MOSFET capacitance models due to the inherent continuity of the model. This model has been implemented as a new intrinsic capacitance model (**Capmod=3**) in the industry standard **BSIM3v3.2**.

1. Introduction

Physically accurate and computationally efficient modeling of MOSFET intrinsic capacitance is becoming important for circuit design. This is because of the desire for accuracy in circuit performance simulation, the need for a physical and predictive model for statistical process modeling, and better convergence behavior.

Almost all the MOSFET capacitance models available in the circuit simulators today are piecewise ones [1]-[6]. In these models, different sets of charge-voltage equations are used for different regions of device operation, i.e. accumulation, depletion and inversion regions. Unfortunately, this approach usually results in non-smoothness in capacitance-voltage (C-V) characteristics near the threshold and flat-band voltages, and therefore poor fitting of capacitance in the moderate inversion, where analog circuits are usually biased. The non-smoothness is also believed to be a reason for non-convergence in transient circuit simulations. While charge sheet models [7]-[9] using one single equation for all operating regions guarantee continuity and smoothness of the capacitance, it adds computational overhead because of the numerical iterations needed for the solution of the surface potential. Similarly, models based on the surface potential formulations as proposed in [10]-[13] are also continuous, but they are also computationally expensive. Furthermore and more important, charge sheet models assume Boltzmann, instead of Fermi-Dirac, statistics and ignore the non-zero thickness of the accumulation and inversion layers — an unacceptable omission for thin-T_{ox} CMOS technologies.

In other words, all of the MOSFET capacitance models reported so far [1]-[17] suffer the inability to model all the significant physical phenomena for $T_{ox} \leq 7$nm. These physical effects include the polysilicon gate depletion and inversion layer charge quantization effects, both of which significantly affect the C-V characteristics of CMOS devices [18]-[25]. It has already been widely accepted that the polysilicon depletion is caused by the partial activation of the dopant and/or insufficient doping in the polysilicon gate and consumes some of the applied gate voltage, and the quantization effect causes the spatial charge distribution to spread deep (20A to 100A) from the interface (therefore the *charge sheet* assumption fails). Based on the numerical surface potential approach, the quantization effect on the surface potential has been modeled through a correction for the intrinsic carrier density [26], where the spatial charge distribution was not considered.

We have demonstrated in the previous work [27], [28] that this finite charge thickness (or average charge centroid) can not be ignored any more for the accurate modeling of charge-voltage characteristics in scaled MOSFET's as illustrated in Figure 1. In this paper, we report a complete, computationally efficient and accurate compact model (Charge-Thickness Model (**CTM**)) for thin-oxide MOSFET intrinsic capacitance, including the finite charge thickness from the accumulation through depletion to inversion regions as well as the polysilicon depletion effects. The results of comparisons with measured data and BSIM3v3.1 [29] clearly show the higher accuracy of this model in all regions of operation. Since no numerical iterations are involved, application of this model in the circuit simulation reveals high efficiency and robustness. This model has been implemented as capMod=3 in BSIM3v3.2, and is inherently continuous and smooth in all regions of operation. It should be noted that since all the critical physical parameters, such as the threshold voltage and bulk charge coefficient, are consistent with the DC model, this new model preserves the high scalability and accurate modeling of the non-uniform doping, mobility degradation, velocity saturation and drain induced barrier lowering (DIBL) effects that are characteristic of BSIM3.

In the following, formulations for the polydepletion, finite charge thickness from the accumulation through depletion to inversion regions, and bias-dependent surface potential due to the bulk charge in the inversion region are presented and discussed in section 2, followed by the charge equations and channel charge partitioning in section 3. In section 4, this new model is compared with the measured data and BSIM3v3.1. Section 5 gives the evaluation of the simulation performance and some discussions, followed by the conclusion of this work.

2. Physics and Modeling

2.1 Polysilicon gate depletion effect

The polysilicon gate depletion (or polydepletion) effect results in a voltage drop V_{poly} across the polysilicon gate, thus reducing the gate capacitance. There are in general two approaches to model V_p in compact MOSFET modeling: one is to include it in the threshold voltage and the bulk charge factor [20], while the other is to replace the gate voltage V_{gs} with an effective gate voltage V_{gse} calculated by subtracting the polysilicon gate band bending from V_{gs}, i.e. $V_{gse} = V_{gs} - V_p$ [6]. V_p is dependent on the gate oxide thickness T_{ox}, the voltage across the oxide and the polysilicon gate doping concentration, denoted by N_p. By solving the

Poisson equation in the polysilicon under the depletion approximation [6], [21], V_{gse} can be formulated as

$$V_{gse} = V_{fb} + \varphi_{s0} + \frac{q\varepsilon_{si}N_p T_{ox}^2}{\varepsilon_{ox}^2}\left(\sqrt{1+\frac{2\varepsilon_{ox}^2\left(V_{gs}-V_{fb}-\varphi_{s0}\right)}{q\varepsilon_{si}N_p T_{ox}^2}}-1\right) \tag{1}$$

where V_{fb} is the flat band voltage, ε_{si} and ε_{ox} are the dielectric constants of the silicon and SiO_2, respectively, and φ_{s0} is the surface potential equaling twice the Fermi potential ϕ_f.

2.2 *Finite charge layer thickness model*

The present model is a charge-based model and therefore starts with the DC charge thickness, X_{DC}. X_{DC} is defined as $\int \rho(x)xdx \Big/ \int \rho(x)dx$. Qualitatively, the charge thickness introduces a capacitance in series with C_{ox} as illustrated in Figure 2, resulting in an effective C_{ox}, C_{oxeff}.

$$C_{oxeff} = \frac{C_{ox}C_{cen}}{C_{ox}+C_{cen}} \tag{2}$$

where $C_{cen} = \varepsilon_{si} \Big/ X_{DC}$.

Based on self-consistent solution of the Schrodinger, Poisson and Fermi-Dirac equations [27], [28], we have developed a universal model for the finite charge thickness from the accumulation to depletion region:

$$X_{ad} = \frac{1}{3}L_{debye}\exp\left[acde\cdot\left(\frac{N_{sub}}{2\times10^{16}}\right)^{-0.25}\cdot\frac{V_{gs}-V_{bs}-V_{fb}}{T_{ox}}\right] \quad \text{[cm]} \tag{3}$$

where L_{debye} is the Debye length, $\left(V_{gs}-V_{bs}-V_{fb}\right)\Big/T_{ox}$ has the unit of MV/cm, V_{bs} is the body bias and *acde* is a fitting parameter with the default value of 1. Figure 3 shows that (3) agrees with the numerical quantum simulations for different T_{ox} and channel doping concentrations. However, note that (3) can not be implemented directly in the code, since the very large accumulation layer thickness is physically correct but numerically unacceptable due to numerical overflow. An upper bound, X_{max}, is therefore imposed for the accumulation and depletion charge layer thickness:

$$X_{\max} = X_{ad}\Big|_{V_{gs}=V_{bs}+V_{fb}} = \frac{1}{3}L_{debye} \tag{4}$$

To smooth the transition from (3) to (4), we propose

$$X_{DC} = X_{\max} - \frac{1}{2}\left(X_0 + \sqrt{X_0^2 + 4\delta_t X_{\max}}\right) \tag{5}$$

where $X_0 = X_{\max} - X_{ad} - \delta_t$ and $\delta_t = 10^{-3} \cdot T_{ox}$. (5) reduces to (3) for $V_{gs} < (V_{bs} + V_{fb})$ and (4) for $V_{gs} > (V_{bs} + V_{fb})$.

For the inversion region, the inversion charge layer thickness is proposed as

$$X_{DC} = \frac{1.9\times10^{-7}}{1+\left(\frac{V_{gtx} + 4(V_{th} - V_{fb} - 2\phi_f)}{2T_{ox}}\right)^{0.7}} \quad \text{[cm]} \tag{6}$$

where the second term in the denominator has the unit of MV/cm. Through V_{fb} in the above equation, the model is applicable to N^+ or P^+ poly-Si gates as well as any other future gate materials. V_{gtx} has the form

$$V_{gtx} = noff \cdot nv_t \ln\left(1 + \exp\left(\frac{V_{gse} - V_{th} - voffcv}{noff \cdot nv_t}\right)\right) \tag{7}$$

where n is the geometry and bias dependent subthreshold swing parameter [29], v_t is the thermal voltage and *noff* and *voffcv* are model parameters which can be extracted from measured data. V_{gtx} becomes (V_{gse} - V_{th} - *voffcv*) when (V_{gse} - V_{th} - *voffcv*) > $3noff \cdot nv_t$ and rapidly drops to zero when (V_{gse} – V_{th} - *voffcv*) < $-3noff \cdot nv_t$. Figure 4 shows that (6) can match numerical quantum simulations very well over wide ranges of gate oxide thicknesses and channel doping concentrations.

2.3 Bulk charge induced bias dependent surface potential in inversion region

The classical condition for strong inversion is defined as the surface potential $\varphi_s = \varphi_{s0} = 2\phi_f$ [30], and pinned at that value even when V_{gs} exceeds the threshold voltage V_{th}. In reality, φ_s varies with the gate bias even in strong inversion. This approximation is one cause for the sharp turn in C-V around V_{th} in the modeled capacitance, which can give rise to significant inaccuracies for analog circuit and low voltage/power designs, for which the moderate inversion region (V_{th} ~ (V_{th} + $6v_t$) [31]) is of great importance.

Considering both the inversion charge (q_{inv}) and bulk charge (q_B) layer thickness in the inversion region, the surface potential φ_s can be written as

$$\varphi_s = -\frac{X_{DC} \cdot q_{inv}}{\varepsilon_{si}} - \frac{X_{bulk} \cdot q_B}{\varepsilon_{si}} \tag{8}$$

where X_{DC} is expressed by (6) and X_{bulk} is the bulk charge layer thickness. q_{inv} can now be formulated as

$$q_{inv} = -C_{oxeff} \cdot (V_{gtx} - \varphi_\delta) \tag{9}$$

where $\varphi_\delta = -\frac{X_{bulk} \cdot q_B}{\varepsilon_{si}} - \varphi_{s0}$. By solving the Poisson equation and assuming zero inversion charge layer thickness, we propose an analytical formulation for φ_δ as

$$\varphi_\delta = v_t \, ln\left(\frac{V_{gtx} \cdot (V_{gtx} + 2\gamma\sqrt{\varphi_{s0}})}{moin \cdot \gamma v_t^2} + 1\right) \tag{10}$$

where γ is the body bias coefficient and *moin* is a fitting parameter with a typical value of 15. Note that (10) rapidly drops to zero for $(V_{gse} - V_{th}) < -3noff \cdot n v_t$ as inversion disappears. Figure 5 shows that equation (10) agrees with the numerical quantum simulation results with *moin*=15.

Figure 6 shows the comparison of the measured and modeled channel charge density q_{inv} as a function of V_{gs} for a large NMOSFET with T_{ox}=40A and N_p=5.5×10^{19}cm^{-3}. The measured q_{inv} was obtained by numerical integration of the gate-to-channel capacitance C_{gc}. The excellent agreement between CTM *qinv* and data implies that CTM can potentially be used in DC IV model.

3. Charge Formulation and Partitioning

In this section, we will introduce charge equations and inversion charge partitioning which are essential to derive the total 16 capacitances. The physical effects described in Section 2 will be considered. The capacitance will be shown to be inherently continuous and smooth from the accumulation to inversion region in Section 4.

3.1 Equations for Accumulation Charge

In this region, the inversion charge Q_{inv} is zero, and the gate charge Q_G is mirrored in the substrate as the accumulation charge Q_{acc} near the silicon surface. Q_{acc} is computed by

$$Q_{acc} = WLC_{oxeff} \cdot V_{gbacc} \tag{11}$$

where V_{gbacc} is the effective gate-to-body voltage and given by

$$V_{gbacc} = \frac{1}{2} \cdot \left[V_0 + \sqrt{V_0^2 + 4\delta_v V_{fb}} \right] \tag{12}$$

where $V_0 = V_{fb} + V_{bs} - V_{gse} - \delta_v$ and δ_v = 0.02V so that V_{gbacc} reduces to $(V_{fb} + V_{bs} - V_{gse})$ in the accumulation region $((V_{gse} - V_{fb} - V_{bs}) << -\delta_v)$ and zero in other operating regions $((V_{gse} - V_{fb} - V_{bs}) >> \delta_v)$.

3.2 *Equations for Depletion Charge*

Under the depletion approximation, the bulk charge in the depletion region [31], [32] can be obtained by solving the Poisson equation as

$$Q_{B0} = -WLC_{oxeff} \cdot \frac{\gamma^2}{2} \cdot \left[-1 + \sqrt{1 + \frac{4(V_{gse} - V_{fbx} - V_{bs} - V_{gtx})}{\gamma^2}} \right] \tag{13}$$

where $V_{fbx} = V_{fb} - V_{gbacc}$ and γ is the body effect factor. V_{gtx}, approximately equal to zero in the depletion region, is introduced in (13) so that Q_{B0} can account for the V_{ds}-independent portion of the bulk charge in the inversion region. Note that (13) becomes zero in the accumulation region.

3.3 *Charge Equations for Inversion Region*

Classical MOSFET modeling uses the assumptions of constant surface potential and zero thickness of the inversion layer for the inversion regime. Instead, the inversion charge in the linear region is proposed as

$$Q_{inv} = -WC_{oxeff} \cdot \int_0^L (V_{gtx} - \varphi_\delta - \alpha V_y) dy \tag{14}$$

where α is the bulk charge coefficient considering geometry and body bias dependencies of the bulk charge [29]. To express the integration upper limit in (14) in terms of V_{ds}, dy has to be replaced by dV_y. dy is related to dV_y by

$$dy = \frac{dV_y}{E_y} \tag{15}$$

The source-drain current I_{ds}, considering the finite charge thickness effect, has the form of

$$I_{ds} = W\mu_{eff} C_{oxeff} \cdot (V_{gtx} - \varphi_\delta - \alpha V_y) E_y \tag{16}$$

where μ_{eff} is the bias-dependent effective mobility [29]. Substituting (15) into (16) and integrating from source to drain will permit solution to I_{ds}.

By combining (14)-(16) and performing integration, we can obtain the expression for Q_{inv} in the linear region as

$$Q_{inv} = -WLC_{oxeff} \cdot \left[V_{gtx} - \varphi_\delta - \frac{1}{2}\alpha V_{ds} + \frac{\alpha^2 V_{ds}^2}{12 \cdot \left(V_{gtx} - \varphi_\delta - \alpha V_{ds}/2 \right)} \right] \tag{17}$$

Similarly, the bulk charge in the linear region can be written as

$$Q_B = -WC_{oxeff} \cdot \int_0^L \left(\gamma \sqrt{\varphi_{s0} - V_{bs}} - (1-\alpha) V_y \right) dy \tag{18}$$

Q_B can be divided into two parts: one is V_{ds} independent denoted by Q_{B0} and the other component is a function of V_{ds} called $Q_{B\delta}$. Q_{B0} is given by (13) where V_{gtx} reduces to (V_{gse}-V_{th} - *voffcv*) in the strong inversion region, while $Q_{B\delta}$ is formulated as

$$Q_{B\delta} = WLC_{oxeff} \cdot \left[\frac{1-\alpha}{2} V_{ds} - \frac{(1-\alpha) \cdot \alpha V_{ds}^2}{12 \cdot \left(V_{gtx} - \varphi_\delta - \alpha V_{ds}/2 \right)} \right] \tag{19}$$

In order for (17) and (19) to be applicable in the saturation region, an effective V_{ds}, namely V_{dsx}, is used to replace V_{ds} in both (17) and (19). V_{dsx} has the form of

$$V_{dsx} = V_{dsat} - \frac{1}{2} \cdot \left(V_1 + \sqrt{V_1^2 + 4\delta_v V_{dsat}} \right) \tag{20}$$

where $V_1 = V_{dsat} - V_{ds} - \delta_v$. V_{dsx} is equal to V_{ds} in the linear region and V_{dsat} in the saturation region. The saturation voltage V_{dsat} is proposed as

$$V_{dsat} = \frac{V_{gtx} - \varphi_\delta}{\alpha} \tag{21}$$

3.4 Channel Charge Partitioning

Channel charge Q_{inv} (Eq. (17)) must be seperated into drain and source charge components. There are three popular channel charge partitioning schemes, i.e. 50/50, 40/60 and 0/100 which are defined as the ratios of Q_d (the drain terminal charges) to Q_s (the source terminal charges) in the saturation region. 50/50 partitioning is the simplest which assumes the inversion charges are contributed equally from the source and drain nodes, and Q_d and Q_s are written as

$$Q_S = Q_D = \frac{1}{2} Q_{inv} = -\frac{WLC_{oxeff}}{2} \left[V_{gtx} - \varphi_\delta - \frac{1}{2}\alpha V_{dsx} + \frac{\alpha^2 V_{dsx}^2}{12 \cdot \left(V_{gtx} - \varphi_\delta - \alpha V_{dsx}/2 \right)} \right] \tag{22}$$

40/60 partitioning represents the most physical, under the quasi-static assumption, of the three partitioning schemes in which the channel charges are allocated to the source and drain electrodes by assuming a linear dependence as proposed by Ward and Dutton [2]. Q_d and Q_s are obtained as the following

$$Q_S = -\frac{WLC_{oxeff}}{2\left(V_{gtx}-\varphi_\delta-\alpha V_{dsx}/2\right)^2}\left[(V_{gtx}-\varphi_\delta)^3-\frac{4}{3}(V_{gtx}-\varphi_\delta)^2\alpha V_{dsx}+\frac{2}{3}(V_{gtx}-\varphi_\delta)(\alpha V_{dsx})^2-\frac{2}{15}(\alpha V_{dsx})^3\right] \tag{23}$$

$$Q_D = -\frac{WLC_{oxeff}}{2\left(V_{gtx}-\varphi_\delta-\alpha V_{dsx}/2\right)^2}\left[(V_{gtx}-\varphi_\delta)^3-\frac{5}{3}(V_{gtx}-\varphi_\delta)^2\alpha V_{dsx}+(V_{gtx}-\varphi_\delta)(\alpha V_{dsx})^2-\frac{1}{5}(\alpha V_{dsx})^3\right] \tag{24}$$

0/100 partitioning was useful in fast transient and high frequency small signal simulations to suppress the drain current spikes by assigning all inversion charges in the saturation region to the source electrode. With V_{dsat} in the form of (21) and following the procedure proposed by Yang *et al.* [3], Q_S and Q_D can be formulated as

$$Q_S = -\frac{WLC_{oxeff}}{2}\cdot\left[V_{gtx}-\varphi_\delta+\frac{1}{2}\alpha V_{ds}-\frac{\alpha^2V_{ds}^2}{12\cdot\left(V_{gtx}-\varphi_\delta-\alpha V_{ds}/2\right)}\right] \tag{25}$$

$$Q_D = -\frac{WLC_{oxeff}}{2}\cdot\left[V_{gtx}-\varphi_\delta-\frac{3}{2}\alpha V_{ds}+\frac{\alpha^2V_{ds}^2}{4\cdot\left(V_{gtx}-\varphi_\delta-\alpha V_{ds}/2\right)}\right] \tag{26}$$

Q_G can be obtained directly from the charge conservation principle. The MOSFET intrinsic capacitance can be obtained by differentiating the terminal charges described above with respect to the terminal voltages, which are defined as

$$C_{ij} = \Delta_{ij}\frac{\partial Q_i}{\partial V_j}, \quad \left(\Delta_{ij} = 1\, for\, i = j; \Delta_{ij} = -1\, for\, i \neq j\right) \tag{27}$$

where i and j represent MOSFET's four terminals.

4. Experimental Verification and Discussion

In this section, model validity and accuracy are first investigated using measured C-V characteristics from many technologies and compared with a BSIM3v3.1 capacitance model, Capmod2, which is a continuous single-equation model but does not consider the finite charge layer thickness and variable surface potential effects. Model overall performance is then evaluated by simulating various devices under different bias conditions and at different

temperatures. The measured data were obtained at 100kHz, which is high enough to provide good signal to noise ratio and low enough to suppress distributed RC effects. The model parameters were extracted by *BSIMpro* [33]. Unless otherwise stated, those newly introduced model parameters (i.e. *acde*, *moin*, *noff* and *voffcv*) in the present model use their default values in what follows.

Figure 7 shows that CTM can model very accurately the measured NMOSFET gate capacitance *Cgg* for arbitrary *Tox* and polysilicon and substrate doping concentrations.

Although the model equations were developed based on NMOSFET's, we found they are equally applicable to the PMOSFET case. A buried channel PMOSFET is employed to demonstrate the model validity and capability. Figure 8 shows C_{gg} as a function of V_{gs} but for an n-type polysilicon gate (buried channel) PMOSNFET with T_{ox}=75A and N_p=1.5×10^{19}cm^{-3}. The model indeed fits data smoothly and with high accuracy throughout the operating regions. For p-type polysilicon gate (i.e. surface channel) PMOSFET's, when the channel surface is inverted, the gate gets depleted. This is identical to and can be modeled similarly to the NMOSFET case. But for the n-type gate (buried channel) PMOSFET's, the opposite is true: the polysilicon is accumulated at the interface in the inversion regime and becomes depleted when the silicon surface is accmulated. From the figure, the charge layer thickness and polysilicon depletion effects are seen to be modeled correctly and accurately in buried channel PMOSFET's as well.

The new model was also verified extensively for other trans-capacitance components of both NMOS and PMOS transistors fabricated by several different technologies. For brevity, we show in Figure 9 (a) and (b) examples of the gate-to-drain (C_{gd}) and gate-to-source (C_{gs}) capacitance versus the source drain voltage V_{ds} under different V_{gs} for an NMOSFET with T_{ox}=45A. The model again shows relatively high accuracy, smoothness and, in particular, accurate prediction of the transition from the linear to saturation region, demonstrating the effectiveness of V_{dsx} and V_{dsat} formulations, where the parameter *moin* can be extracted from CV data for even better fit of the slopes of the transition region.

The contribution of each physical effect described above is depicted in Figure 10, where the measured and modeled gate-to-channel capacitance C_{gc} is plotted against V_{gs}. Only when all those effects are accounted for, a good fit can be achieved as shown in the figure. Table 1 quantifies the simulation errors relative to the measured data extracted from Figure 10 for two

typical V_{gs} values, which correspond to the moderate and strong inversion regimes. It is apparent that ignoring any specific effect will result in appreciable discrepancies in either the moderate inversion region or the strong inversion region.

Table 1

Quantification of C_{gc} errors in the moderate and strong inversion regions when different physical mechanisms are considered as shown in Figure 10

Physical mechanism considered	Error percentage (%)	
	Moderate inversion (V_{gs}=0.7V)	Strong inversion (V_{gs}=2.0V)
CapMod=2	36.8	20.7
CTM without φ_δ	28.0	3.9
φ_δ only	13.5	19.2
φ_δ and polydepletion	7.8	7.3
CTM	1.9	-0.3

A more dramatic example of the universality of CTM is shown in Figure 11, where the C_{gg} of a SiON/Ta_2O_5/TiN NMOSFET with an equivalent T_{ox} of 18Å is accurately modeled with CTM. Figure 12 shows the comparison of inverter delays for different technology generations between CTM and classical non-CTM (capMod=2 of BSIM3v3.1) simulations. As technology shrinks, the significance of CTM increases.

The model overall performance has been quantitatively evaluated at the device level for various device dimensions and under extensive testing conditions as listed in Table 2. In the evaluation, all the 16 C_{ij}'s except for C_{dd}, C_{ds}, C_{ss} and C_{sd} were investigated. The test results are shown in Table 3, where the results for BSI3v3.1 Capmod2 are also given for comparison. In Table 3 "total signals" means the total number of tests made for the 12 C_{ij}'s, "error signals" stands for the number of tests where certain trans-capacitance has very small negative values, "error points" is the total number of bias points where C_{ij} shows small negative values, and the error rate is the ratio of the error signal number to the number of total signals. The present model performance is substantially improved over Capmod2.

Table 2

Model performance testing conditions

NMOSFET and PMOSFET sizes				Biases (Opposite signs for PMOS)	Temperatures (°C)		
W (μm)		L(μm)		V_{ds}: 0 ~ 5.5V / 0.1V step	Min	Nominal	Max
Small	Large	Small	Large	V_{gs}: -1.0 ~ 5.5V / 0.05V step			
1.008	8.0	0.672	9.6	V_{bs}: 0, -5.5V	-55	27	150

Table 3

Summary of model performance test results with testing conditions listed in Table 2

	Capmod2	New model
Total signals	384	384
Error points	3323	86
Error signals	68	3
Error rate (%)	17.7083	0.7813

5. Simulation Results and Comparisons

Not only does this model fit the measured capacitance accurately and show good continuities, but also it gives the simulation results of sub-half CMOS circuits as expected and demonstrates desirable simulation performance as well [34]. This received intensive and extensive testing at many Compact Model Council member companies and no non-convergence problems have ever been reported. As an example, for a logic book simulation, the delay is decreased by about 2 percent and operational amplifier (Op Amp) roll off shows about 0.1 decade sooner while its gain is maintained, as compared with BSIM3v3.1 capmod2 [34]. In the following, we will show some typical simulation results and make comparisons of the simulation performance between this model and the capacitance models available in BSIM3v3.1. All the newly introduced model parameters used their default values and 40/60 charge partitioning was used in all cases.

Transient simulation performance was evaluated with a typical 101-stage 0.18μm CMOS ring oscillator (RO). simulated using this model and BSIM3v3.1 Capmod2. The gate oxide thickness T_{ox} was 42Å and the polysilicon gate doping concentration N_p was $6.0\times10^{19}cm^{-3}$. The overlap capacitance [29] were also included in the simulation. The total

CPU time and total iteration number are 693.02 seconds and 19673 for CTM, while the counterparts for capmod2 are 637.27 seconds and 19758, respectively.

The simulation performance has been further investigated in both AC and transient analyses of some CMOS circuits between this model and all BSIM3v3.1 capacitance models. Table IV compares the run statistics for the AC analysis of a CMOS Op Amp. This model takes less CPU time than capmod2. Capmod0 is a long-channel model and its equations are algebraically much simpler, resulting in the least CPU time. Table 5 shows the comparisons of the run statistics of the transient analysis of a CMOS one-bit comparator. Again as in the case of RO, this model requires the least number of total iterations due to the continuity in charges and capacitance. Although this model contains more logarithmic and exponential functions, the CPU time penalty, however, is only 9 percent when compared to BSIM3v3.1 capmod2.

Tabel 4
Run statistics comparison for the AC analysis of a CMOS Op Amp
between this model and BSIM3v3.1 capacitance models

Model	Capmod0	Capmod1	Capmod2	New model
Total CPU time (s)	0.28	0.31	0.30	0.29
Load time (s)	0.02	0.02	0.02	0.02
Total iterations	16	16	16	16

Tabel 5
Run statistics comparison for the transient analysis of a CMOS one-bit comparator
between this model and BSIM3v3.1 capacitance models

Model	Capmod0	Capmod1	Capmod2	New model
Total CPU time (s)	2.03	2.05	2.23	2.42
Load time (s)	1.42	1.42	1.64	1.80
Total iterations	676	656	657	625
Transient iterations	657	637	638	606
Transient time points	172	167	167	159
Rejected time points	21	19	19	16

Figure 13 (a) and (b) show respectively the gate currents for turn-on and turn-off transients of a 0.18μm NMOSFET with T_{ox}=42A simulated by this model and BSIM3v3.1

capmod2. Capmod2 suffers oscillations in the gate current across the threshold voltage. However, the new model exhibits stable gate currents in both turn-on and turn-off periods because of the continuous capacitance, which should lead to better convergence behavior in large circuit simulation.

6. Conclusion

An accurate compact model for MOSFET intrinsic capacitance has been presented for ultra-thin T_{ox} CMOS circuit simulation in both analog and digital applications. This model considers the finite charge layer thickness due to quantization effects in all operating regions and variable surface potential in the inversion region as well as polysilicon depletion. No numerical iterations are used in the model. Extensive and intensive model tests reveals high accuracy, good universality and high performance of the model with excellent convergence properties. This model is inherently continuous and physics based. This model has been implemented as a new capacitance model in BSIM3v3.2. It is anticipated that this model will be the preferred capacitance/charge model in future CMOS circuit simulation.

7. Acknowledgement

The authors are thankful to Drs. Keith Green, Tom Vrontsos, Britt Brooks, and Vinod Gupa at Texas Instruments, Drs. David Newmark and B. Gadepally at Motorola, Drs. A. Yang and Bob Daniel at Avant!, Dr. Ming-Chie Jeng at Cadence Systems, Drs. Sally Liu and Mishel Matloubian at Rockwell Technology, Shiuh-Wuu Lee and Ling-Chu Chien at Intel, Drs. Ping-Chin Yeh and Dick Dowell at Hewlett Packart, Dr. Josef Watts at IBM, and T. Saito at NEC for valuable discussions, code testing and providing test wafers and some of the measured data. The authors would especially acknowledge Compact Model Council for coordinating BSIM3v3.2 code testing and release.

This work is partially supported by Compact Model Council and SRC SJ-417-97.

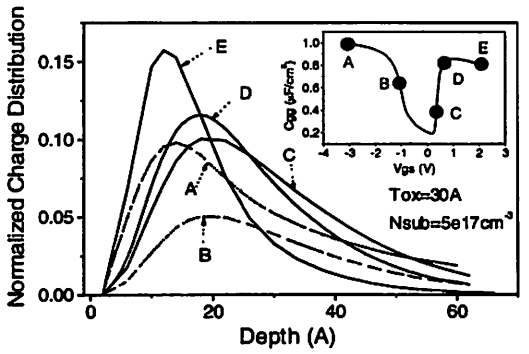


Figure 1. Charge distributions from numerical quantum simulations show significant charge thickness at various bias conditions shown in the inset.

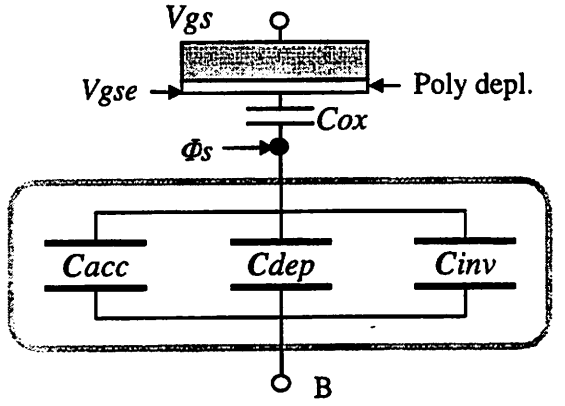


Figure 2. Charge-thickness capacitance concept in this model. *Vgse* accounts for poly-depletion.

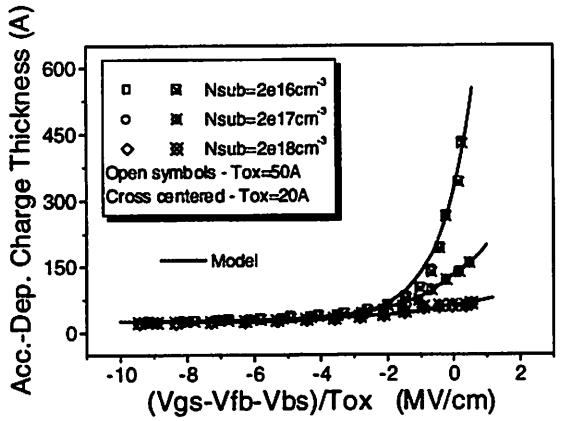


Figure 3. For all *Tox* and *Nsub*, modeled charge thickness agrees with numerical quantum simulations throughout accumulation and depletion regions.

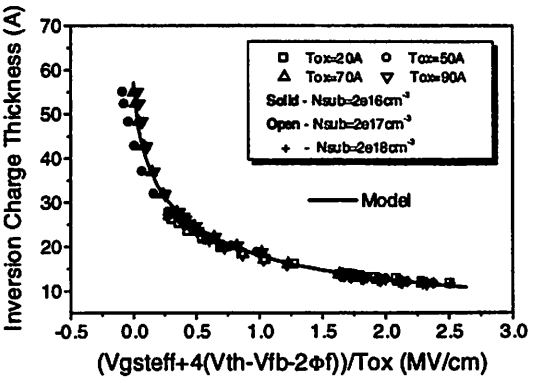


Figure 4. For all *Tox* and *Nsub*, modeled inversion charge thickness agrees with numerical quantum simulations.

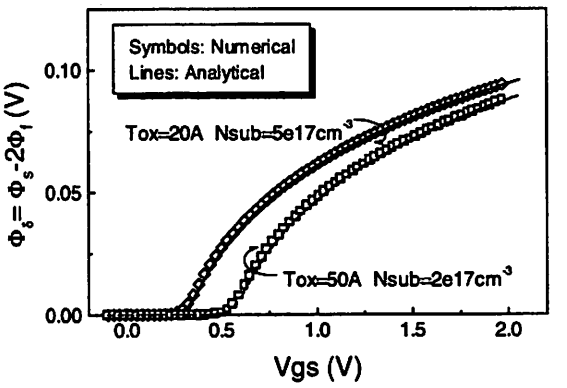


Figure 5. $\Phi\delta$ agrees with numerical quantum simulations.

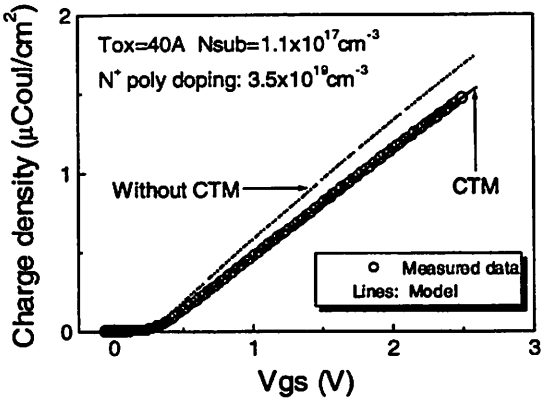


Figure 6. Present model (CTM) agrees with measured *qinv* (obtained from integrating *Cgc* data).

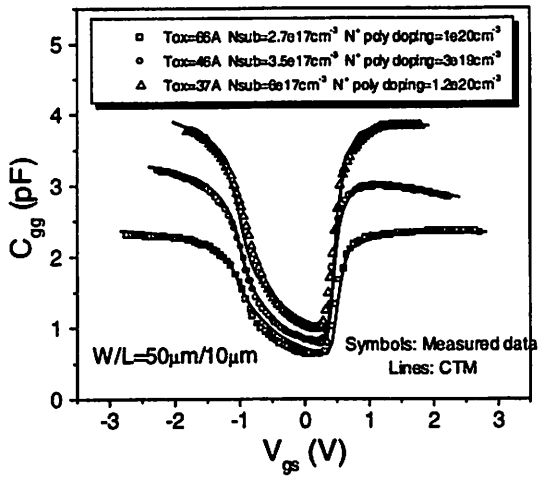


Figure 7. CTM verified with measured NMOSFET *Cgg* for various oxide thickness, substrate and polysilicon gate doping concentrations.

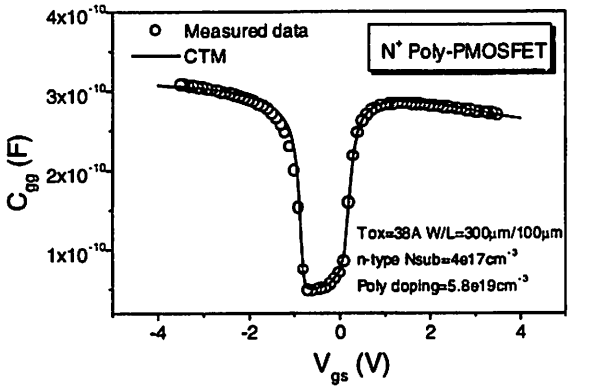
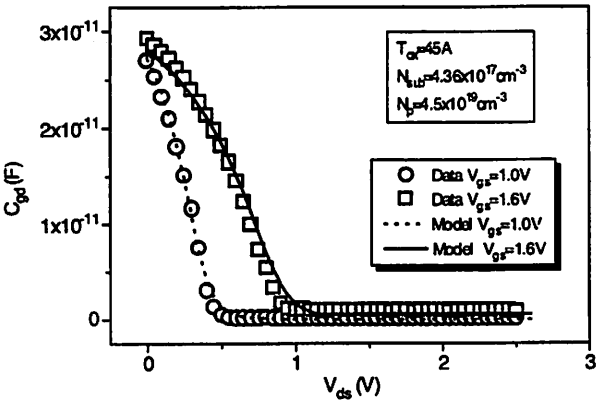
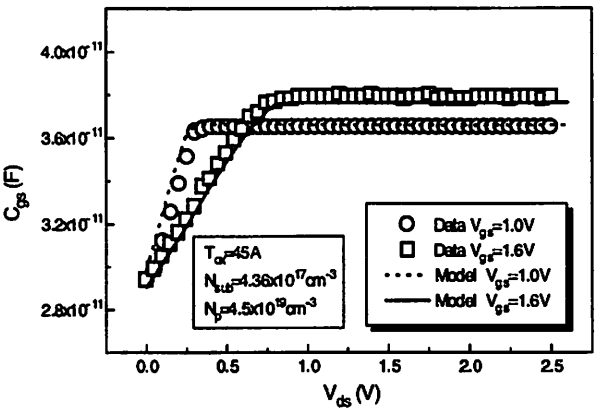


Figure 8. Measured and CTM modeled *Cgg* of N^+-poly (buried-channel) PMOSFET. Gate depletion occuring in the accumulation region is correctly modeled.



(a)



(b)

Figure 9. Comparison of measured and modeled (CTM) gate-to-drain C_{gd} (a) and gate-to-source C_{gs} (b) capacitance as a function of the source drain voltage V_{ds}. W/L = 100μm/100μm, T_{ox}=45A, N_p= 4.5× 10^{19} cm^{-3} and N_{sub}=4.36x$10^{17}$$cm^{-3}$.

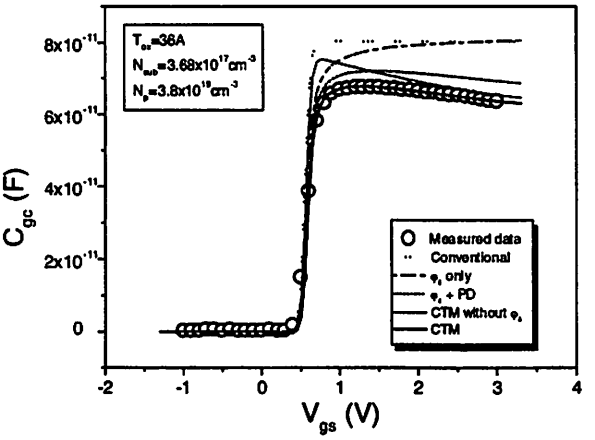


Figure 10. Measured and modeled gate-to-channel capacitance C_{gc} of an NMOSFET with different physical effects considered in simulation. T_{ox} =36A, W/L=100μm/100μm, N_p=3.8x$10^{19}$$cm^{-3}$ and N_{sub}=3.68x$10^{17}$$cm^{-3}$. *PD* stands for Polysilicon Depletion.

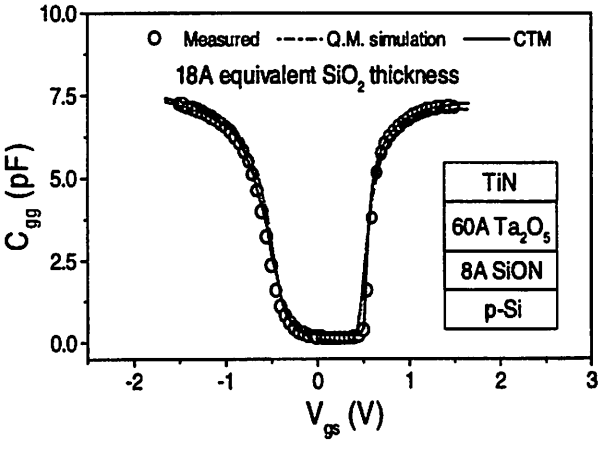


Figure 11. Universality of CTM is demonstrated by modeling the *Cgg* of NMOSFET with SiON/Ta_2O_5/TiN gate stack.

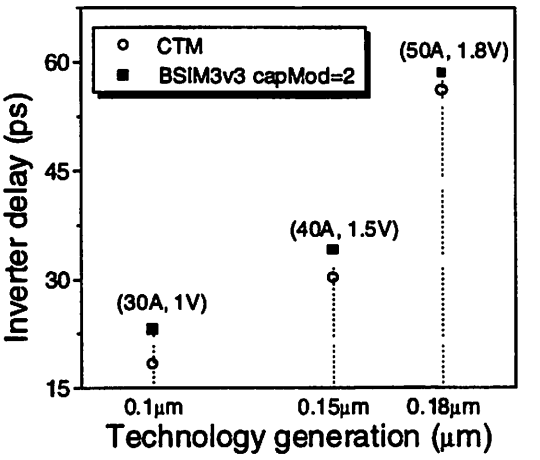
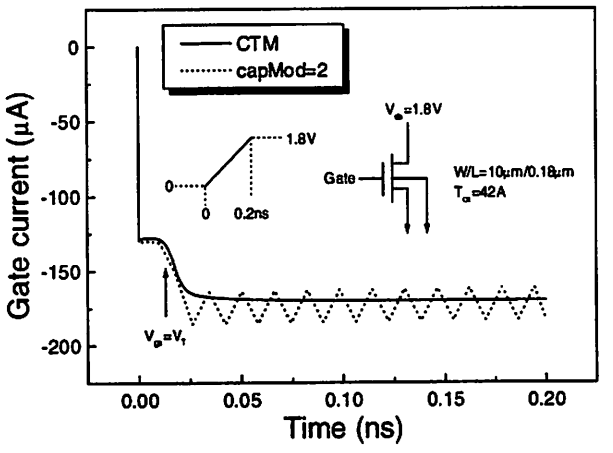


Figure 12. Simulated unloaded CMOS inverter delay is shorter with CTM than BSIM3v3 capMod=2. *Wn/Wp*=10μm/15μm.



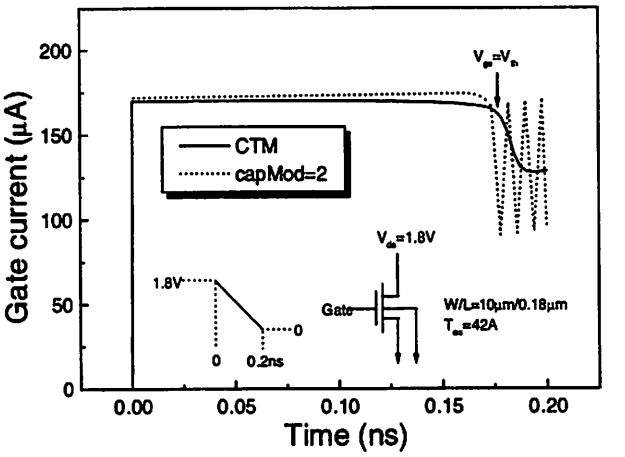


Figure 13. Comparison of gate currents in turn-on and turn-off transients between CTM and BSIM3v3.1 capmod=2 for an NMOSFET with *W/L*=10μm/0.18μm and T_{ox}=42A. The absence of the gate current oscillation in CTM helps convergence.

8. Reference

[1] J. E. Meyer, "MOS models and circuit simulation," *RCA Review*, vol. 32, pp. 42-63, 1971.

[2] D. E. Ward and R. W. Dutton, "A charge-orient model for MOS transistor capacitances," *IEEE J. Solid-State Circuits*, vol. 13, pp. 703-708, 1978.

[3] P. Yang, "Capacitance modeling for MOSFET," in *Advances in CAD for VLSI*, vol. 3, pt. 1, A. E. Ruehli, Ed. Amsterdam, The Netherlands: North Holland, pp. 107-130, 1986.

[4] B. J. Sheu, D. L. Scharfetter, P. K. Ko, and M. C. Jeng, "BSIM: Berkeley short channel IGFET model for MOS transistors," *IEEE J. Solid-State Circuits*, vol. 22, pp. 558-566, 1987.

[5] M. C. Jeng, "Design and modeling of deep-submicrometer MOSFET's," Ph.D. Thesis, U. C. Berkeley, 1989.

[6] J. H. Huang, Z. H. Liu, M. C. Jeng, K. Hui, M. Chan, P. K. Ko, and C. Hu, *BSIM3 Version 2.0 User's Manual*, UC Berkeley, March 1994.

[7] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, pp. 345-355, 1978.

[8] F. Van de Wiele, " A long-channel MOSFET model," *Solid-State Electron.*, vol. 22, pp. 991-997, 1979

[9] C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behavior of the MOS transistor in quasistatic operation," *Solid-State Electron.*, vol. 26, pp. 941-949, 1983.

[10] H. J. Park, P. K. Ko, and C. Hu, "A charge sheet capacitance model of short channel MOSFET's for SPICE," *IEEE Trans. Computer-Aided Design*, vol. 10, pp. 376-389, 1991.

[11] A. R. Boothroyd, S. W. Tarasewicz, and C. Slaby, "MISNAN – a physically based continuous MOSFET model for CAD applications," *IEEE Trans. Computer-Aided Design*, vol. 10, pp. 1512-1529, 1991.

[12] D. H. Cho, S. M. Kang, K. H. Kim, and S. H. Lee, "An accurate intrinsic capacitance modeling for deep submicrometer MOSFET's," ," *IEEE Trans. Electron Devices*, vol. 42, pp. 540-548, 1995.

[13] M. Miura-Mattausch, U. Feldmann, A. Rahm, M. Bollu, and D. Savignac, "Unified complete MOSFET model for analysis of digital and analog circuits," *IEEE Trans. CAD of Integrated Cir. and Sys.* vol. 15, pp. 1-7, 1996.

[14] H. K. Lim, and J. G. Fossum, "A charge-based large-signal model for thin-film SOI MOSFET's," *IEEE J. Solid-State Circuits*, vol. 20, pp. 366-377, 1985.

[15] B. Gharabagi, and M A. El-Nokali, "A charge-based model for short-channel MOS transistor capacitances," *IEEE Trans. Electron Devices*, vol. 37, pp. 1064-1072, 1990.

[16] K. M. Rho, K. Lee, M. Shur, and T. A. Fjeldly, "Unified quasi-static MOSFET capacitance model," *IEEE Trans. Electron Devices*, vol. 40, pp. 131-136, 1990.

[17] N. D. Arora, R. Rios, C. L. Huang, and K. Raol, "PCIM: a physically based continuous short-channel IGFET model for circuit simulation," *IEEE Trans. Electron Devices*, vol. 41, pp. 988-997, 1994.

[18] G. Yaron, and D. Frohman-Bentchkowsky, "Capacitance voltage characterization of poly Si-SiO2-Si structures," *Solid-State Electron.*, 23, pp. 433-439, 1980.

[19] C.-L. Huang and N. D. Arora, "Measurements and modeling of MOSFET I-V characteristics with polysilicon depletion effect," *IEEE Trans. Electron Devices*, vol. 40, pp. 2330-2337, 1993.

[20] N. D. Arora, R. Rios, and C.-L. Huang, "Modeling the polysilicon depletion effect and its impact on submicron CMOS circuit performance," *IEEE Trans. Electron Devices*, vol. 42, pp. 935-943, 1995.

[21] K. F. Schuegraf, C. C. King, and Chenming Hu, "Impact of polysilicon depletion in thin oxide MOS technology," *Proc. 1993 Int. Symp. VLSI Tech., Sys. and Appl.* (VLSI-TSA), Taiwan, pp. 86-90, 1993.

[22] C. Y. Lu, J. M. Sung, H. C. Kirsch, S. J. Hollenius, T. E. Smith, and L. Manchanda, "Anomalous C-V characteristics of implanted poly MOS structure in n^+/p^+ dual gate CMOS technology," *IEEE Electron Devices Lett., EDL*-10, pp. 192-194, 1989.

[23] R. Rios, and N. D. Arora, "Determination of ultra-thin gate oxide thickness for CMOS structures using quantum effects," *IEDM Tech. Dig.*, pp. 613-616, 1994.

[24] S. A. Hareland, S. Krishnamurthy, S. Jallepalli, C.-F. Yeap, K. Hasnat, A. F. Tasch, Jr., and C. M. Maziar, "A computationally efficient model for inversion layer quantization effects in deep submicron n-channel MOSFET's," *IEDM Tech. Dig.*, pp. 933-936, 1995.

[25] S. A. Hareland, S. Krishnamurthy, S. Jallepalli, C.-F. Yeap, K. Hasnat, Al F. Tasch, Jr., and C. M. Maziar, "A computationally efficient model for inversion layer quantization effects in deep submicron n-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, pp. 90-96, 1996.

[26] R. Rios, N. D. Arora, C.-L. Huang, N. Khalil, J. Faricelli, and L. Gruber, "A physical compact MOSFET model, including quantum mechanical effects, for statistical circuit design applications," *IEDM Tech. Dig.*, pp. 937-940, 1995.

[27] Ya-Chin King, H. Fujioka, S. Kamohara, W.-C. Lee, and Chenming Hu, "AC charge centroid model for quantization of inversion layer in NMOSFET," *Int. Symp. VLSI*

Technology, Systems and Applications, Proc. of Tech. Papers, Taipei, Taiwan, pp. 245-249, June 1997.

[28] Ya-Chin King, H. Fujioka, S. Kamohara, Kai Chen, and Chenming Hu, "DC charge centroid model for quantization of inversion layer in MOSFET for cirrent prediction," Submitted to *IEEE EDL*, 1997.

[29] *BSIM3v3 User's Manual*, UC Berkeley, 1995-1996.

[30] S. M. Sze, *Physics of Semconductor Devices*, 2nd Edition, John Wiley & Sons, Inc., 1981.

[31] N. D. Arora, *MOSFET Modeling for VLSI Circuit Simulation: Theory and Practice.* Wein, New York: Springer-Verlag, 1993.

[32] T. P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, New York, 1987.

[33] *BSIMpro User's Manual*, BTA Technology, Inc., Feb. 1996.

[34] Compact Model Council Meeting Reports, September 1997, Dallas.