Scalable MOSFET Short-channel Charge Model for All Regions

Guan Huei See[*], Siau Ben Chiah[*,**], Xing Zhou[*], Karthik Chandrasekaran[*], Wangzuo Shangguan[*], Zhaomin Zhu[*], Guan Hui Lim[*], Shesh Mani Pandey[**], Michael Cheng[**], Sanford Chu[**], and Liang-Choo Hsia[**]

[*]School of Electrical & Electronic Engineering, Nanyang Technological University
Nanyang Avenue, Singapore 639798, exzhou@ntu.edu.sg
[**]Chartered Semiconductor Manufacturing Ltd, 60 Woodlands Industrial Park D, St. 2, Singapore 738406

ABSTRACT

A scalable short-channel MOSFET charge model valid for all bias regions is presented. As the device length is reduced, long-channel intrinsic charge model cannot predict the short-channel dynamic behavior correctly. The extrinsic capacitances, such as overlap capacitance and bias-dependent fringing capacitance, must be included in the charge model as they are comparable to the intrinsic capacitance at short-channel dimensions. In order to ensure model scalability over geometry, short-channel effects must be included in the core charge model. This paper extends the short-channel models, such as bulk-charge sharing and potential-barrier lowering, for charge modeling that is valid for all regions. The model is verified by comparison with numerical simulations for three short-channel devices, 0.5, 0.25 and 0.09 μm. It is shown that the model accurately scales with the short-channel capacitances.

Keywords: bulk charge sharing, potential barrier lowering, quasi-2D Poisson solution, intrinsic/extrinsic capacitances, scalable compact model

1 INTRODUCTION

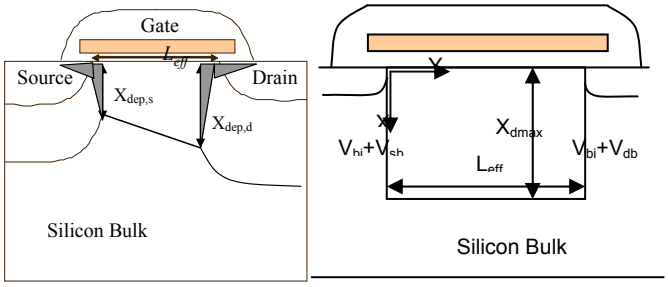
As MOSFET gate length reduces, short-channel effects (SCEs) become more apparent, resulting in deviation from the long-channel dynamic behavior. The main source of SCEs in the channel arises from the lateral-field effect of the source/drain terminal. Even at zero applied source-to-drain bias, the built-in potential of the source/drain junctions is affecting the intrinsic channel charges and charges in the overlap region. In addition, the extrinsic or parasitic capacitances, such as the fringing capacitance [1, 2], bias-dependent inner fringing capacitance [3], and lightly-doped drain (LDD) overlap region capacitance [4], are comparable to the intrinsic channel capacitance, which need to be included for short-channel charge modeling.

MOSFET parasitic capacitances are important for short-channel charge modeling. However, most of these capacitances are gate-length independent for a given technology. The key to charge-model scalability therefore lies in the physical modeling of the gate-length dependent intrinsic charges. Short-channel models such as bulk-charge sharing (BCS) [5] and potential-barrier lowering (PBL) [6] have been successfully implemented to model the SCEs in DC model [7]. This paper extends the short-channel models for charge modeling that is valid for all regions of operation. In section 2.1, scalable intrinsic charges are formulated. Section 2.2 covers the formulation of extrinsic charges/capacitances, which are added to the intrinsic charges/capacitances. Section 3 discusses the model verification in comparison with numerical data, and followed by the conclusions in section 4.

2 MODEL FORMULATION

2.1 Intrinsic Charges

BCS and quasi-two-dimensional (quasi-2D) PBL are the two models used to handle the SCEs in threshold-voltage (V_t) modeling [7]. This paper extends the BCS and quasi-2D PBL models so that these models are valid for all regions in the unified regional charge model [8]. The BCS model is derived based on charge sharing in the depletion layer of source/drain regions. The gate electric field overlapping with the source/drain junction fields, as shown in the shaded area in Fig. 1(a), causes the reduction of the effectiveness of the gate charge control when depletion layer is formed in the channel. BCS is modeled as the average of the depletion layer thickness near the source and drain, $X_{dep,s}$ and $X_{dep,d}$, with respect to the long-channel (without BCS) bulk charge, $Q_{B,L}$ [7]:



(a) (b)

Figure 1: (a) Schematic for BCS model for calculating the charge reduction. (b) Gaussian box and boundary conditions in quasi-2D analysis for PBL.

$$\Delta Q_{B,BCS} = \frac{\lambda}{L_{eff}}\left(\frac{X_{dep,s}+X_{dep,d}}{2}\right)Q_{B,L} \tag{1}$$

where λ is a fitting parameter. Effectively, this model is quite similar to the lateral-field gradient (LFG) model [9], both of which are only modifying the surface potential and charge above the flat-band voltage. If only BCS or LFG model alone is used, it will miss the SCE in accumulation. Therefore, PBL due to lateral field needs to be included for improvement in accumulation in the charge model.

The quasi-2D PBL is formulated from 2D Poisson equation but reduced to solving 1D Poisson equation (or the so-called "quasi-2D") using the boundary conditions of a Gaussian box, as shown in Fig. 1(b). We changed the depletion charge (used in our V_t model) on the right-hand side of the equation to the total gate charge to derive the quasi-2D PBL that is valid for all regions:

$$V_{gb}-V_{FB}-\psi(y)+\frac{\varepsilon_{Si}X_{d\max}}{\eta C_{ox}}\frac{dE_s(y)}{dy}=\frac{Q_g+Q_{ox}}{C_{ox}} \tag{2}$$

where $E_s(y)$ is the lateral electric field, $X_{d\max}$ the maximum depletion depth in the channel, V_{gb} the gate–bulk voltage, V_{FB} the flat-band voltage, ε_{Si} the silicon permittivity, C_{ox} the oxide capacitance (per unit area), Q_{ox} the fixed oxide charge density, and η is a fitting parameter for PBL. $\psi(y)$ is the position-dependent potential in the channel which, from quasi-2D solution [10], is given as

$$\begin{aligned}\psi(y) = {} & \phi_s+(V_{db}+V_{bi}-\phi_s)\frac{\sinh(y/l_\alpha)}{\sinh(L_{eff}/l_\alpha)} \\ & +(V_{sb}+V_{bi}-\phi_s)\frac{\sinh\left[(L_{eff}-y)/l_\alpha\right]}{\sinh(L_{eff}/l_\alpha)}\end{aligned} \tag{3}$$

where $l_\alpha=(\varepsilon_{Si}X_{d\max})/(\eta C_{ox})$, ϕ_s is the long-channel surface potential (without PBL), V_{bi} the built-in potential of source/drain junctions, V_{db} the drain–bulk voltage, V_{sb} the source–bulk voltage. For $V_{ds}\neq 0$ and $L_{eff}>>l_\alpha$, $\psi(y)$ is approximated to

$$\begin{aligned}\psi(y) = {} & \phi_s+(V_{bi}+V_{db}-\phi_s)e^{(y-L_{eff})/l_\alpha} \\ & +(V_{bi}+V_{sb}-\phi_s)e^{-y/l_\alpha}+(V_{bi}+V_{db}-\phi_s)e^{-L_{eff}/l_\alpha}\end{aligned} \tag{4}$$

The surface-potential solution is y-position dependent. For simplicity and model compactness, only the potential difference at the minimum potential in the channel is used to derive the change in surface potential, $\Delta\phi_s$, due to PBL. This can be done by equating the derivative of (4) to zero. Therefore, a general solution of the change in surface potential due to PBL is derived as:

$$\Delta\phi_s=\frac{1}{\cosh\left(\frac{L_{eff}}{2l_\alpha}\right)}\left[(V_{bi}+V_{sb}-\phi_s)\cosh\left(\frac{z}{2}\right)+\frac{V_{ds}}{2}\frac{\sinh\left(\frac{L_{eff}}{2l_\alpha}-\frac{z}{2}\right)}{\sinh\left(\frac{L_{eff}}{2l_\alpha}\right)}\right] \tag{5}$$

where

$$z=\ln\frac{V_{bi}+V_{db}-\phi_s}{V_{bi}+V_{sb}-\phi_s}. \tag{6}$$

In our unified regional approach, ϕ_s in (5) and (6) is modeled by the (long-channel) unified regional surface potential in accumulation, ϕ_{acc}, and from depletion to strong inversion, ϕ_{ds}, such that asymptotically

$$\phi_s=\begin{cases}\phi_{acc} & (V_{gb}<V_{FB})\\ \phi_{ds} & (V_{gb}>V_{FB})\end{cases}. \tag{7}$$

In accumulation region, the bulk charge can be approximated by [11]

$$\begin{aligned}Q_{B,acc} &\approx -Q_{G,acc}=-C_{ox}WL_{eff}\left[V_{gb}-V_{FB}-(\phi_{acc}+\Delta\phi_{acc})\right] \\ &=\underbrace{-C_{ox}WL_{eff}(V_{gb}-V_{FB}-\phi_{acc})}_{Q_{BL,acc}}+\underbrace{C_{ox}WL_{eff}\Delta\phi_{acc}}_{-\Delta Q_{B,PBL}}\end{aligned} \tag{8}$$

where the first term is the long-channel bulk charge, and the second term defines the charge reduction due to PBL in accumulation. The influence of PBL from depletion to strong inversion is in the potential changes above the flat-band voltage, $\Delta\phi_{ds}$, which subsequently can be partitioned into source and drain charges [8].

2.2 Extrinsic Capacitances

There are a few extrinsic capacitances need to be modeled, namely, the overlap capacitance [4] and fringing capacitance [3, 12]. The overlap capacitances are derived based on the unified regional surface potential of the opposite type from the channel. The overlap region is usually highly doped as compared to the channel. In the normal operating range of a MOSFET, it is safe to ignore carrier inversion in the overlap region. Only charges due to (accumulated) electrons and (depleted) donors are included in the Pao–Sah voltage solution [13] (for nMOSFETs) in the overlap region (ignoring hole inversion), given by

$$(V_{gc}-V_{FB,ov}-\phi_{ov})=\gamma_{ov}\sqrt{v_{th}\left(\exp\left(\frac{\phi_{ov}}{v_{th}}\right)-1\right)-\phi_{ov}} \tag{9}$$

where V_{gc} is the gate–channel (or source/drain) voltage, $V_{FB,ov}$ the flat-band voltage in overlap region, v_{th} the thermal voltage, and ϕ_{ov} is the surface potential in the overlap region with γ_{ov} being the body factor in that region.

Following the unified regional formulation for charges, the source-side overlap region bulk charge in accumulation and depletion are modeled through the unified regional surface potentials in accumulation ($\phi_{ov,acc}$) and depletion ($\phi_{ov,sub}$):

$$\begin{aligned}Q_{Sov,acc} &= -WL_{ov}C_{ox}(V_{gsf}-\phi_{ov,acc}) \\ Q_{Sov,sub} &= -WL_{ov}C_{ox}(V_{gsr}-\phi_{ov,sub})\end{aligned} \tag{10}$$

where $Q_{Sov}=Q_{Sov,acc}+Q_{Sov,sub}$ is the total bulk charge at the source side of the overlap region of length L_{ov}, V_{gsf} and V_{gsr} are the forward and reverse interpolation functions, respectively [8]. Drain side has similar expressions.

Similar to the intrinsic channel region, the overlap region also has charge sharing. Therefore, the BCS model is extended to the overlap region modeling. The source-side overlap region BCS is modeled by

$$\Delta Q_{Sov,BCS} = 0.5\lambda_{ov}\left(X_{dov,s}/L_{ov}\right)Q_{Sov} \tag{11}$$

where $X_{dov,s}$ is the source-side overlap region depletion depth. For the drain-side overlap charge, similar analysis to the source side is applied by replacing V_{sb} by V_{db}.

Extrinsic and intrinsic fringing capacitances are formulated based on charges for simple addition. The extrinsic fringing charge is partitioned into bulk and inversion charge, respectively, as follows:

$$\begin{aligned} Q_{B,FR} &= \nu_b Q_{B,L}\left(C_{side}+C_{bottom}\right)\Big/\left(L_g C_{ox}\right) \\ Q_{I,FR} &= \nu_i Q_{I,L}\left(C_{side}+C_{bottom}\right)\Big/\left(L_g C_{ox}\right) \end{aligned} \tag{12}$$

where $Q_{B,L}$ and $Q_{I,L}$ are the long-channel bulk charge and inversion charge, respectively, C_{side} and C_{bottom} are the side-wall and bottom capacitances given in [12], L_g is the gate length, ν_b and ν_i are two fitting parameters for extrinsic fringing capacitances. The intrinsic fringing charge follows the semi-empirical approach of [3]:

$$Q_{I,IF} = -\kappa_{if} C_{IF,\max}\,\mathrm{erf}\left(\frac{V_{gb}-V_{FB}-\alpha_{if}}{\beta_{if}}\right) \tag{13}$$

where κ_{if}, α_{if}, and β_{if} are fitting parameters that can be tuned for the magnitude, mean location, and spread of the Gaussian profile, respectively, for modeling the inner fringing capacitance.

The terminal charges are therefore defined as follows:

$$\begin{aligned} Q_D &= Q_{D,L} + Q_{D,OV} - \Delta Q_{Dov,BCS} + 0.5Q_{I,FR} \\ Q_S &= Q_{S,L} + Q_{S,OV} - \Delta Q_{Sov,BCS} + 0.5Q_{I,FR} \\ Q_B &= Q_{B,L} - \Delta Q_{B,PBL} - \Delta Q_{B,BCS} + Q_{B,FR} + Q_{I,IF} \\ Q_G &= -\left(Q_D + Q_S + Q_B + Q_{OX}\right) \end{aligned} \tag{14}$$

where $Q_{D,L}$, $Q_{S,L}$, $Q_{B,L}$, and Q_{OX} are the long-channel drain charge, source charge, bulk charge, and oxide trapped charge, respectively [8]. The terminal capacitances are numerically calculated from

$$C_{ij} = \delta_{ij}\frac{\partial Q_i}{\partial V_j}, \qquad \delta_{ij} = \begin{cases} 1 & (i=j) \\ -1 & (i\neq j) \end{cases} \tag{15}$$

which are used for comparison with the Medici numerical capacitance data for model verification.

3 RESULTS AND DISCUSSION

Numerical simulations are used to aid the development of the short-channel charge model. Fig. 2 shows the normalized gate capacitance (C_{gg}) of three short-channel devices, L_{eff} = 0.5, 0.25, 0.09 µm (each normalized to its own $C_{ox}WL_g$), generated from the Medici numerical devices with uniformly doped channel and LDD regions, as shown in the inset. SCE is seen from the changing capacitances in all regions as the gate length decreases.

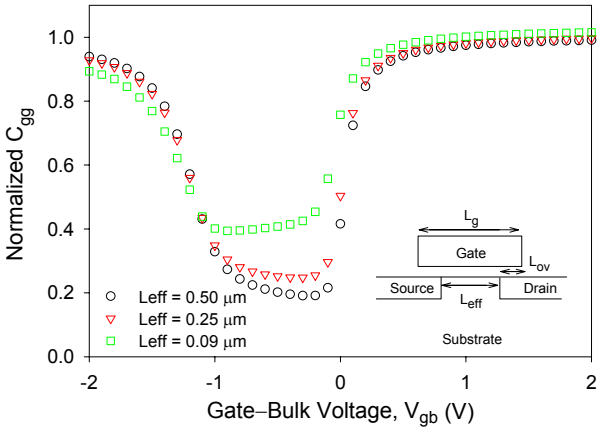


Figure 2: Numerical simulation of C_{gg} (normalized to each $C_{ox}WL_g$) for L_{eff} = 0.5 µm, 0.25 µm, and 0.09 µm. *Inset*: Medici structure to generate the C_{gg} data.

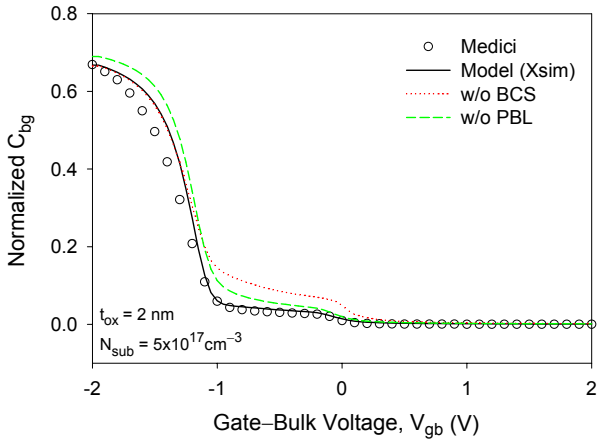


Figure 3: Modeled C_{bg} from complete model (—), without BCS (····), and without PBL (– –); compared with the Medici data for L_{eff} = 0.09 µm.

Fig. 3 shows the comparison of bulk–gate capacitance (C_{bg}) between the Medici and model data at the shortest channel, L_{eff} = 0.09 µm. The BCS and quasi-2D PBL models are turned off one at a time to show the effective region of influence of each model. The PBL model is affecting all regions, from accumulation to strong inversion, where as BCS only has effect above the flat-band voltage.

Fig. 4 shows the comparison of gate–channel capacitance (C_{gc}) between the Medici and model data at L_{eff} = 0.09 µm. Overlap charge model, bulk charge-sharing model in the overlap region, extrinsic fringing model, and intrinsic fringing model are turned off one at a time to observe the model behavior. Overlap charges have influence across all regions, while overlap-region BCS affecting mostly in accumulation, depletion, and a small portion of strong-inversion region. The extrinsic fringing has control over the level in C_{gg}, while inner fringing gives rise to the "bell" shape effect in depletion region.

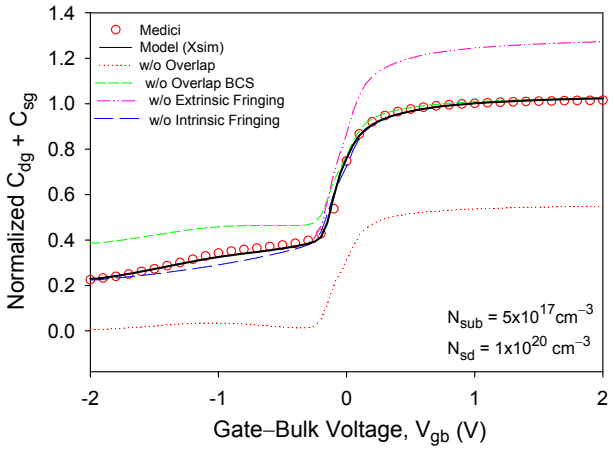


Figure 4: Modeled $|C_{dg} + C_{sg}|$ from complete model (—), without overlap charge (····), without overlap BCS (- - -), without extrinsic fringing charge (-···-), and without intrinsic fringing charge (– –); compared with the Medici data for L_{eff} = 0.09 µm.

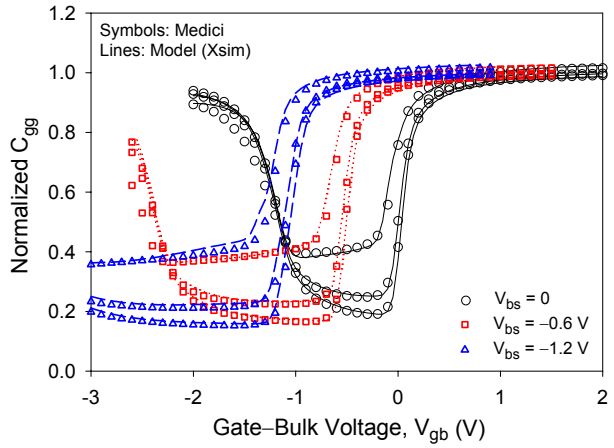


Figure 5: Modeled (lines) and simulated (symbols) C_{gg} for short-channel devices L_{eff} = 0.5 µm, 0.25 µm and 0.09 µm in all regions of operation for V_{bs} = 0, –0.6 and –1.2 V.

Fig. 5 shows the comparison of the full model including the intrinsic and extrinsic short-channel effects with the simulation data for all three short-channel devices. The model is tuned to match the shortest-length C_{gg} at V_{bs} = 0 and played back to predict the intermediate lengths and different body bias conditions. The model agrees well with the Medici data. The model also converges to the simple long-channel model when the gate length is very large.

4 CONCLUSION

In conclusion, both the intrinsic and extrinsic short-channel models for low-field condition are implemented included in the unified regional charge model. The quasi-2D PBL model in accumulation is a unique feature due to the unified regional surface-potential approach, which is nontrivial for LFG or inversion-charge based approaches. The model shows good scalability from long to short gate lengths and all body biases due to the physically derived geometry-dependent charge model, which is essential for AC and transient circuit simulation.

Acknowledgment: This work was supported in part by Semiconductor Research Corporation under Contract 2004-VJ-1166 and in part by Nanyang Technological University under Grant RGM30/03.

REFERENCES

[1] K. Suzuki, "Parasitic capacitance of submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 46, pp. 1895–1900, 1999.

[2] D. Pattanayak, J. Poksheva, R. Downing, and L. Akers, "Fringing field effect in MOS devices," *IEEE Trans. Components, Hybrids, and Manufacturing Technology,* [see also: *IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part A, B, C*], vol. 5, pp. 127–131, 1982.

[3] F. Pregaldiny, C. Lallement, and D. Mathiot, "A simple efficient model of parasitic capacitances of deep-submicron LDD MOSFETs," *Solid-State Electron.*, vol. 46, pp. 2191–2198, 2002.

[4] P. Klein, K. Hoffmann, and B. Lemaitre, "Description of the bias dependent overlap capacitance at LDD MOSFETs for circuit applications," *IEDM Tech. Dig.*, 1993, pp. 493–496.

[5] L. D. Yau, "A simple theory to predict the threshold voltage of short-channel IGFET's," *Solid-State Electron.*, vol. 17, pp. 1059–1063, 1974.

[6] T. Toyabe and S. Asai, "Analytical models of threshold voltage and breakdown voltage of short-channel MOSFET's derived from two-dimensional analysis," *IEEE Trans. Electron Devices*, vol. 26, pp. 453–461, 1979.

[7] X. Zhou and K. Y. Lim, "Unified MOSFET compact *I–V* model formulation through physics-based effective transformation," *IEEE Trans. Electron Devices*, vol. 48, pp. 887–896, 2001.

[8] X. Zhou, S. B. Chiah, K. Chandrasekaran, G. H. See, W. Shangguan, S. M. Pandey, M. Cheng, S. Chu, and L.-C. Hsia, "Unified regional charge-based versus surface-potential-based compact modeling approaches," *Proc. NSTI Nanotech 2005*, Anaheim, CA, May 2005, WCM, pp. 25–30.

[9] T. L. Chen and G. Gildenblat, "An extended analytical approximation for the MOSFET surface potential," *Solid-State Electron.*, vol. 49, pp. 267–270, 2005.

[10] Z.-H. Liu, C. Hu, J.-H. Huang, T.-Y. Chan, M.-C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, pp. 86–95, 1993.

[11] G. H. See, S. B. Chiah, X. Zhou, K. Chandrasekaran, W. Shangguan, S. M. Pandey, M. Cheng, S. Chu, and L-C Hsia, "Unified regional charge-based MOSFET model calibration," *Proc. NSTI Nanotech 2005*, Anaheim, CA, May 2005, WCM, pp. 143–146.

[12] M. J. Kumar, V. Venkataraman, and S. K. Gupta, "On the parasitic gate capacitance of small-geometry MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, pp. 1676–1677, 2005.

[13] C.-T. Sah, "A history of MOS transistor compact modeling," *Proc. NSTI Nanotech 2005*, Anaheim, CA, May 2005, WCM, pp. 347–390.