Unified MOSFET Compact I–V Model Formulation through Physics-Based Effective Transformation

Xing Zhou, *Senior Member, IEEE,* and Khee Yong Lim, *Student Member, IEEE*

***Abstract*—A one-region compact I_{ds} model from subthreshold to saturation, which resembles the same form as the well-known long-channel model but includes all major short-channel effects (SCEs) in deep-submicron (DSM) MOSFETs , has been formulated through physics-based effective transformation. The model has 23 process-dependent fitting parameters, which requires an 11-step, one-iteration extraction procedure. The new approach to modeling channel-length modulation (CLM), subthreshold diffusion current, and edge-leakage current, all in a compact form, has been verified with the 0.25-μm experimental data. The model covers the full range of gate length (without "binning") and bias conditions, and can be correlated to true process variables for aiding technology development.**

***Index Terms*—Compact model (CM), effective transformation, MOSFET, parameter extraction, process correlation, technology development.**

I. Introduction

ADVANCES in technology scaling have been the driving forces for MOSFET miniaturization, which also lead to increasing challenges in compact model (CM) development. The current trend shows a rather disturbing version of "Moore's Law"—the number of CM parameters doubles every decade. There exists a large body of literature on MOSFET CMs [1]–[7], [27]. Among them, BSIM3v3 [1], [27] has been considered as the *de facto* industry standard in deep-submicron (DSM) MOSFET circuit simulation. However, the philosophy for BSIM parameter extraction is based on local/group-device optimization [1], [27], which may lead to unphysical parameter values. The requirement for "binning" effectively makes the model "piece-wise" although it is one-piece for all bias conditions. Moreover, it is difficult to relate the model directly to process variations, which becomes increasingly demanding for DSM technology development and early success in circuit design.

Another observation in DSM device modeling that departs from conventional long-channel models is the fact that fluctuation of device geometrical and structural parameters, which is unavoidable, has major influence on its electrical characteristics. Unlike for long-channel devices, "precise" determination of *individual* device critical dimension (CD), oxide thickness, channel doping, junction depth, etc., may not make sense for DSM CMs since these quantities are strongly subject to process fluctuations, and the CM only models the average values of these quantities. Models that are based on single-device extraction [6]–[7] may not be useful for DSM device/circuit modeling since the trend now is toward technology, rather than transistor, characterization. In other words, the CM for DSM devices should be developed for, and its parameters extracted from, a given technology of varying gate lengths, with short-channel effects (SCEs) "calibrated" to the length-independent long-channel devices of the same technology (wafer).

In this paper, the approach to formulating unified CMs for DSM MOSFETs is presented, which is based on physics-based "*effective transformation,*" a step-by-step process of adding higher order effects to the well-known long-channel MOSFET equations. The idea is similar to the "effective voltage transformation" [8], and has been demonstrated in our previous work on compact threshold-voltage formulation [9]. The approach is based on the belief that the SCE demonstrates itself as a *gradual* effect as the gate length alone is decreased and, thus, modeling of which must be separated from (and can be calibrated by) its long-channel counterpart. The general procedure is to incorporate physically-derived or empirically-based equations for each individual SCE with effective quantities that contain process-dependent fitting parameters, which also approach the values of their long-channel counterpart in the long-channel limit. All the fitting parameters have their physical meanings, and their extraction follows a unique, *one-iteration* procedure at the "boundary" condition at which their effect is most pronounced. The sequence of model parameter extraction should be such that when a parameter is to be extracted, those that have not been extracted should have little effect; and once extracted, its value should be fixed in subsequent extraction. In addition, third-order effect should only be considered after first- and second-order effects have been accounted for. The general guide in the tradeoff between detailed physics and compact form is what Albert Einstein described: "Everything should be made as simple as possible, but not any simpler."

Our unified compact I_{ds} model is developed with the help of a set of experimental data of a 0.25-μm CMOS shallow trench isolation (STI) wafer with drawn gate lengths (L_{drawn}) from 10 μm down to 0.2 μm. Section II describes the model-equation formulation with a step-by-step effective transformation that results in a one-region equation from subthreshold to saturation covering all gate length and bias conditions, which also resembles the form of the long-channel I_{ds} model. Section III presents the one-iteration parameter extraction procedure as well as the necessary measurement data. Section IV reports another unique feature of the approach to correlating the model to *true* process

Manuscript received May 31, 2000; revised October 4, 2000. The review of this paper was arranged by Editor J. M. Vasi.

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: exzhou@ntu.edu.sg).

Publisher Item Identifier S 0018-9383(01)02359-0.

variables. Results of the developed model are demonstrated and discussed in Section V. Finally, Section VI concludes the paper. The way our CM is formulated as well as the sequence the CM parameters are extracted also demonstrates the principles behind our novel approach to compact $I_{\rm ds}$ modeling.

II. Compact Model Formulation

The rationale on the sequence of our model formulation is the following. Accurate modeling of the threshold voltage (V_t) [9], including V_t roll-off SCE and roll-up reverse SCE (RSCE) as well as drain-induced barrier lowering (DIBL) and body effect, is first carried out. V_t represents a critical point on the $I_{\rm ds}$–$V_{\rm gs}$ curve and is relatively independent of the channel mobility and S/D series resistance. The (electrical) effective channel length ($L_{\rm eff}$), which also appears in the short-channel V_t expression, is modeled with a simple conceptual model for the bias-independent metallurgical channel length ($L_{\rm met}$) with CD correction and lateral lightly-doped drain (LDD) diffusion. Mobility due to the vertical field (μ_0) is then modeled semi-empirically [10], followed by a separate, semi-empirical, gate-bias-dependent S/D series resistance ($R_{\rm sd}$) modeling [11]. Next, the turn-on $I_{\rm ds}$ ("first-order") is formulated by the conventional approach with a two-region velocity–field model and a newly-derived saturation voltage ($V_{\rm dsat}$), followed by channel-length modulation (CLM) modeling [12] described by an effective Early voltage ($V_{A\rm eff}$). Subthreshold current ("second-order") is then formulated by a modified effective gate overdrive ($V_{\rm gg}$) for the correct diffusion current and a novel transformation to retain the compact form. Edge-leakage current ($I_{\rm edge}$) in STI structures [13] ("third-order"), together with diode-leakage current ($I_{\rm diode}$), is finally added to the main MOS current ($I_{\rm mos}$), which "borrows" the same MOS $I_{\rm ds}$ model.

A. Threshold Voltage and Effective Channel Length

The complete V_t model has been presented [9]. In this subsection, we present an enhanced version with a different, more accurate, DIBL modeling.

The ideal threshold-voltage equation is given by

$$V_t = V_{FB} + \phi_{s0} + \gamma\sqrt{\phi_{s0} - V_{\rm bs}} \tag{1a}$$

$$\phi_{s0} = 2\phi_B = 2(kT/q)\,\ln(N_{\rm sub}/n_i) \tag{1b}$$

$$\gamma = \sqrt{2q\varepsilon_{si}N_{\rm sub}}/C_{\rm ox} \tag{1c}$$

where symbols have their usual meanings. The basic idea in [9] is first to transform the uniform substrate doping $N_{\rm sub}$ to an effective vertical nonuniform doping $N_{\rm ch}$ [14], which also extends to short-channel devices; and then, transform $N_{\rm ch}$ to an effective lateral nonuniform doping $N_{\rm eff}$ [9].

To model the charge-sharing effect including the effect of $V_{\rm ds}$ [15], [16], the average source and drain depletion width (X_s and X_d) is modeled, with two fitting parameters, λ (major) and δ (fine-tune)

$$\gamma_{\rm eff} = \gamma\left(\frac{Q_{B'}}{Q_B}\right) = \gamma\left(1 - \frac{\lambda}{L_{\rm eff}}\,\frac{X_s + X_d}{2}\right)$$
$$= \gamma - \frac{\lambda}{L_{\rm eff}}\,\frac{2\varepsilon_{si}}{C_{\rm ox}}\left(\sqrt{\phi_s - V_{\rm bs}} + \frac{\delta V_{\rm ds}}{\sqrt{\phi_s - V_{\rm bs}}}\right) \tag{2a}$$

where ϕ_s is the (short-channel) surface potential at strong inversion. δ (for DIBL) is approximated by a linear function of $V_{\rm bs}$

$$\delta = \delta_0 + \delta_1 V_{\rm bs}. \tag{2b}$$

Replacing γ in (1) by the effective body factor $\gamma_{\rm eff}$ in (2a), V_t remains the same form as in the ideal V_t equation (1a)

$$V_t = V_{FB} + \phi_s + \gamma_{\rm eff}\sqrt{\phi_{s0} - V_{\rm bs}}. \tag{3}$$

For $\delta = 0$, (2a) becomes the one presented in [9].

From the quasi-2D model [15], the surface potential (at strong inversion) in short-channel devices is lowered by $\Delta\phi_s$ from the long-channel one ($2\phi_B$)

$$\phi_s = \phi_{s0} - \Delta\phi_s = 2\phi_B - \Delta\phi_s. \tag{4a}$$

At high $V_{\rm ds}$, the channel surface potential becomes asymmetric and the minimum potential is no longer at $y = L_{\rm eff}/2$ [15]. It can be shown [16] that

$$\Delta\phi_s = \frac{1}{\cosh(L_{\rm eff}/2l_\alpha)} \times \left[(V_{\rm bi} - \phi_{s0})\,\cosh\left(\frac{z}{2}\right) + \frac{\varphi V_{\rm ds}}{2}\,\frac{\sinh\left(\frac{L_{\rm eff}}{2l_\alpha} - \frac{z}{2}\right)}{\sinh\left(\frac{L_{\rm eff}}{2l_\alpha}\right)}\right] \tag{4b}$$

where

$$z = \ln\left(\frac{V_{\rm bi} - \phi_{s0} + V_{\rm ds}}{V_{bi} - \phi_{s0}}\right) \tag{4c}$$

$$l_\alpha = \alpha(\phi_{s0} - V_{\rm bs})^{0.25} \tag{4d}$$

$$V_{\rm bi} = (kT/q)\,\ln(N_{\rm sd}N_{\rm ch}/n_i^2). \tag{4e}$$

In (4b), a fitting parameter φ (for DIBL) is added, which is approximated by a linear function of $V_{\rm bs}$:

$$\varphi = \varphi_0 + \varphi_1 V_{\rm bs}. \tag{4f}$$

When $\varphi = 1$ and z in (4c) is small, (4b) reduces to that presented in [9].

For completeness, the empirical RSCE model $N_{\rm eff}$ in [9] is shown below, which replaces $N_{\rm ch}$ in the previous equations

$$N_{\rm eff} = N_{\rm ch} + \frac{N_{\rm pile}}{\cosh(L_{\rm eff}/2l_\beta)} \tag{5a}$$

$$l_\beta = \beta(\phi_{s0} - V_{\rm bs})^{0.25} \tag{5b}$$

$$N_{\rm pile} = \kappa N_{\rm ch} \tag{5c}$$

$$\kappa = \kappa_0 + \kappa_1 V_{\rm bs}. \tag{5d}$$

The effective channel length is a critical parameter that influences the electrical behavior of a compact terminal $I_{\rm ds}$ model. For DSM devices, conventional approach to extracting $L_{\rm eff}$ starts to become invalid because of the nonscaling behavior of the total linear resistance [17], and the partition of the intrinsic channel resistance and S/D series resistance becomes strongly definition dependent. In our CM, we use a very simple model [18] for the actual poly-gate length with a constant CD

correction ($\Delta_{\rm CD}$) to account for poly-gate lithography and etching variations

$$L_g = L_{\rm drawn} - \Delta_{\rm CD} \tag{6a}$$

and a simple model for $L_{\rm met}$ with a fitting parameter σ to model the lateral LDD diffusion

$$L_{\rm met} = L_g - \Delta L = L_g - 2\sigma x_j \tag{6b}$$

where x_j is the LDD junction depth. $L_{\rm eff}$ is then assumed to be $L_{\rm met}$, with two physical parameters, $\Delta_{\rm CD}$ and x_j

$$L_{\rm eff} = L_{\rm met} = L_{\rm drawn} - \Delta_{\rm CD} - 2\sigma x_j \tag{6c}$$

and a new method of extraction together with V_t (see Section III). This bias-independent $L_{\rm eff}$ model makes subsequent $I_{\rm ds}$ modeling and parameter extraction much simpler. The bias-dependent two-dimensional (2-D) SCE is modeled by V_t based on the new "critical-current at linear-threshold" definition [19] as well as separate modeling of $R_{\rm sd}$ [11] (see Section II-B).

B. Effective Mobility and Series Resistance

Our CM adopts a separate and physical modeling of the effective mobility [10] and series resistance [11]. In [10], the vertical-field channel mobility is modeled semi-empirically, with a compact form to minimize correlation among the three fitting parameters, μ_1, μ_2, and μ_3

$$\mu_0 = \frac{\mu_1}{1 + (\mu_1/\mu_2)E_{\rm eff}^{1/3} + (\mu_1/\mu_3)E_{\rm eff}^2} \tag{7a}$$

where the effective (vertical) field is given by the well-known expression [20]

$$E_{\rm eff} = (Q_B + \eta Q_{\rm inv})/\varepsilon_{si} = (V_{\rm gs} + V_t)/6t_{\rm ox}. \tag{7b}$$

Each fitting parameter has its own physical meaning related to doping or temperature [10], and the μ_0 model will be extended to short-channel devices.

Likewise, source and drain series resistance is modeled physically by a bias-independent (extrinsic) part and a $V_{\rm gs}$-dependent (intrinsic) part as [11]

$$R_{\rm sd} = R_{\rm ext} + R_{\rm int} = \frac{2\rho S}{x_j W} + \frac{v}{V_{\rm gs} - V_t} \tag{8}$$

with two fitting parameters, ρ and v, to be fitted to the short-channel linear $I_{\rm ds}$ data. ρ represents some effective resistivity in the LDD region with a spacer thickness of S. A first-order x_j and W dependence is also built into the $R_{\rm sd}$ model. Since $L_{\rm eff}$ is modeled/extracted separately from $R_{\rm sd}$, what is modeled is actually the voltage drop across the LDD region [11], leaving the correct intrinsic voltage drop across the MOSFET effective channel.

C. Turn-on Current and Effective Saturation Voltage

Like V_t model formulation, our $I_{\rm ds}$ modeling also starts with the well-known long-channel equation. Assuming the two-region piecewise velocity-field relation, the drain current is given by [21]

$$I_{ds0} = G_{\rm eff0} V_{\rm ds} \tag{9a}$$

$$G_{\rm eff0} = \mu_{\rm eff0} C_{\rm ox}(W/L_{\rm eff})[(V_{\rm gs} - V_t) - A_b V_{\rm ds}/2] \tag{9b}$$

$$\mu_{\rm eff0} = \frac{\mu_0}{1 + V_{\rm ds}/(E_{\rm sat} L_{\rm eff})} \tag{9c}$$

$$E_{\rm sat} = 2v_{\rm sat}/\mu_0 \tag{9d}$$

where the subscript "0" denotes the condition for $R_{\rm sd} = 0$.

$$A_b = 1 + \zeta \frac{\gamma}{2\sqrt{\phi_s - V_{\rm bs}}} \tag{9e}$$

is the bulk-charge factor [22, p. 128] in which ζ is a fitting parameter. The effect of $R_{\rm sd}$ is modeled by the conceptual total resistance ($V_{\rm ds}/I_{\rm ds}$) partitioned into the channel resistance ($R_{\rm ch}$) and the S/D resistance ($R_{\rm sd}$)

$$I_{\rm ds} = \frac{V_{\rm ds}}{R_{\rm ch} + R_{\rm sd}} = \frac{V_{\rm ds}/R_{\rm ch}}{1 + R_{\rm sd}/R_{\rm ch}} = \frac{I_{\rm ds0}}{1 + (R_{\rm sd} I_{\rm ds0})/V_{\rm ds}} \tag{10}$$

where $I_{\rm ds0} = I_{\rm ds}(R_{\rm sd} = 0) = V_{\rm ds}/R_{\rm ch}$. Substituting (9a) into (10), with some algebra it can be shown that (10) becomes

$$I_{\rm ds} = G_{\rm eff} V_{\rm ds} \tag{11a}$$

$$G_{\rm eff} = \mu_{\rm eff} C_{\rm ox}(W/L_{\rm eff})[(V_{\rm gs} - V_t) - \tfrac{1}{2} A_b V_{\rm ds}] \tag{11b}$$

$$\mu_{\rm eff} = \frac{\mu_{\rm eff0}}{1 + R_{\rm sd} G_{\rm eff0}}. \tag{11c}$$

This formulation, however, will *not* be used as our final $I_{\rm ds}$ model but only for deriving a consistent saturation voltage ($V_{\rm dsat}$), as detailed below.

When $R_{\rm sd}$ is ignored, $I_{\rm ds}$ in saturation can be obtained from the well-known gradual-channel approximation

$$I_{\rm ds0} = v_{\rm sat} W C_{\rm ox}(V_{\rm gs} - V_t - A_b V_{\rm ds}) \tag{12a}$$

and when $V_{\rm ds} = V_{\rm dsat}$

$$I_{\rm dsat0} = v_{\rm sat} W C_{\rm ox}(V_{\rm gs} - V_t - A_b V_{\rm dsat}). \tag{12b}$$

When $R_{\rm sd}$ is considered, $V_{\rm gs}$ in (12a) should be replaced by $V_{\rm gs} - I_{\rm ds} R_s$ and $V_{\rm ds}$ replaced by $V_{\rm ds} - I_{\rm ds}(R_s + R_d)$. Although we have assumed symmetrical R_s and $R_d (R_s = R_d)$ based on their physical interpretation, their electrical effect (voltage drop) is different for source and drain, which will be included in the $V_{\rm dsat}$ expression. Denoting $I_{\rm dsat}$ for $I_{\rm ds}$ when $V_{\rm ds} = V_{\rm dsat}$ in saturation, with the above replacement for $V_{\rm gs}$ and $V_{\rm ds}$, it is found from (12a)

$$I_{\rm dsat} = v_{\rm sat} W C_{\rm ox}\{(V_{\rm gs} - I_{\rm dsat} R_s) - V_t - A_b[V_{\rm dsat} - I_{\rm dsat}(R_s + R_d)]\}. \tag{13a}$$

After some algebra,

$$I_{\rm dsat} = \frac{I_{\rm dsat0}}{1 - v_{\rm sat} W C_{\rm ox}[(A_b - 1)R_s + A_b R_d]} \tag{13b}$$

where $I_{\rm dsat0}$ is given by (12b). This newly-derived saturation current should "join" the one in the linear region. Equating (13b)

to $I_{\rm ds}(V_{\rm ds} = V_{\rm dsat})$ in (11a), and extracting $V_{\rm dsat}$ out of the equation, it can be shown that

$$V_{\rm dsat} = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{13c}$$

$$\begin{aligned} a &= v_{\rm sat} W C_{\rm ox} A_b R_s \\ b &= -[V_{\rm gs} - V_t + v_{\rm sat} W C_{\rm ox}(V_{\rm gs} - V_t)(2R_s + A_b R_{\rm sd}) \\ &\quad + A_b E_{\rm sat} L_{\rm eff}] \\ c &= (V_{\rm gs} - V_t) E_{\rm sat} L_{\rm eff} + 2 v_{\rm sat} W C_{\rm ox} R_{\rm sd} (V_{\rm gs} - V_t)^2. \end{aligned} \tag{13d}$$

This formulation (as well as result) is different from, and simpler than, the BSIM3v3 expression [1], [27].

To achieve a smooth transition from linear to saturation region, the smoothing function in BSIM3v3 [1], [27]

$$\begin{aligned} V_{d\rm eff} = V_{\rm dsat} - \tfrac{1}{2}\Big[& V_{\rm dsat} - V_{\rm ds} - \delta_s \\ & + \sqrt{(V_{\rm dsat} - V_{\rm ds} - \delta_s)^2 + 4\delta_s V_{\rm dsat}}\Big] \end{aligned} \tag{13e}$$

is used to replace $V_{\rm ds}$ in (9), where δ_s is chosen as a fixed parameter. $V_{d\rm eff}$ approaches $V_{\rm ds}$ when $V_{\rm ds} < V_{\rm dsat}$ and $V_{\rm dsat}$ when $V_{\rm ds} > V_{\rm dsat}$. Then, (9) becomes a unified one-region equation

$$\begin{aligned} I_{ds0} = \mu_0 C_{\rm ox} \frac{W}{L_{\rm eff}} & \frac{1}{1 + V_{d\rm eff}/(E_{\rm sat} L_{\rm eff})} \\ & \times [(V_{\rm gs} - V_t) V_{d\rm eff} - \tfrac{1}{2} A_b V_{d\rm eff}^2] \end{aligned} \tag{14a}$$

which joins (13b) for $V_{\rm ds} > V_{\rm dsat}$. Similarly, (10) becomes

$$I_{\rm ds} = \frac{I_{\rm ds0}}{1 + (R_{\rm sd} I_{ds0})/V_{d\rm eff}}. \tag{14b}$$

Strictly speaking, $I_{\rm ds0} \neq V_{\rm ds}/R_{\rm h}$ when $R_{\rm sd}$ is not ignored and, hence, (14b) is not accurate since $V_{d\rm eff}$ contains $R_{\rm sd}$. This is the real case where $R_{\rm sd} \neq 0$, even for long-channel devices. However, as will be shown in Section III, our separate μ_0 and $R_{\rm sd}$ extraction makes the error involved to be minimal.

D. Channel-Length Modulation and Effective Early Voltage

So far, (14b) has included the effects of vertical- and parallel-field mobility, bulk charge, velocity saturation, and series resistance, but no CLM. A new approach to modeling CLM including high-field effect based on the quasi-2D formulation [21] has been developed [12].

The piecewise velocity–field relation assumes that when the electric field $E > E_{\rm sat}$, electron velocity saturates, $v = v_{\rm sat}$. However, the quasi-2D solution [21] reveals that the electric field in the velocity saturation region (VSR) of length Δl increases exponentially as $E_{\rm sat}\cosh(y/l)$. Since it is not practical to include "local" quantities in a CM, an "effective average field" is introduced [12], defined as

$$E_{\rm av} \equiv \frac{1}{L_{\rm eff}} \left[\int_0^{L_{\rm eff} - \Delta l} E_{\rm sat}\, dy + \int_{L_{\rm eff} - \Delta l}^{L_{\rm eff}} E_{\rm sat}\, \cosh(y/l)\, dy \right]. \tag{15}$$

We assume that the saturation field $E_{\rm sat}$ in (14a) (without CLM) is replaced by $E_{\rm av}$ (with CLM based on the quasi-2D solution). The physics behind this assumption is to model the voltage drop across the VSR such that the voltage across the intrinsic channel can be modeled correctly, with length and bias dependencies (see [12, Figs. 9 and 10]). With this replacement, it can be shown that (14a) becomes

$$I_{d\rm eff} = \left(1 + \frac{V_{\rm ds} - V_{d\rm eff}}{V_{A\rm eff}}\right) I_{\rm ds0} \tag{16a}$$

where I_{ds0} is given by (14a), which takes the familiar form of the "pinch-off" model. The effect of CLM due to increased $E_{\rm av}$ at decreasing $L_{\rm eff}$ is included in an *effective Early voltage* given by [12]

$$V_{A\rm eff} = \frac{E_{\rm sat} L_{\rm eff} (E_{\rm sat} L_{\rm eff} + V_{\rm ds})}{\xi V_{d\rm eff}} \tag{16b}$$

in which ξ is a fitting parameter. As (14a) changed to (16a) with the inclusion of CLM, (14b) is then changed to

$$I_{\rm ds} = \frac{I_{d\rm eff}}{1 + (R_{\rm sd} I_{d\rm eff})/V_{d\rm eff}} \tag{16c}$$

which will not affect the characteristics in the linear region.

E. Subthreshold Current and Effective Gate Overdrive

Accurate subthreshold modeling (second-order) is only meaningful after threshold voltage, mobility, series resistance, CLM, etc., have been modeled accurately from turn-on current (first-order). In order to obtain a unified equation from subthreshold to strong inversion, the smoothing function in BSIM3v3 [23], [1], [27], for an "effective gate overdrive" ($V_{g\rm eff}$) is adopted to replace all $V_{\rm gs}$–V_t in the previous equations such that $I_{\rm ds0}$ in (14a) becomes

$$I_{\rm ds0} = \mu_{\rm eff} C_{\rm ox} \frac{W}{L_{\rm eff}} [V_{g\rm eff} V_{d\rm eff} - \tfrac{1}{2} A_b V_{d\rm eff}^2] \tag{17a}$$

where $V_{g\rm eff}$ is given by [23]

$$V_{g\rm eff} = \frac{2 n v_{th} \ln\left(1 + e^{(V_{\rm gs} - V_t)/(2 n v_{th})}\right)}{1 + 2n(C_{\rm ox}/C_d) e^{-(V_{\rm gs} - V_t - 2V_{\rm off})/(2 n v_{th})}} \tag{17b}$$

with $V_{\rm off}$ as a fitting parameter. $v_{\rm th} = kT/q$ is the thermal voltage, and

$$n = 1 + C_d/C_{\rm ox} \tag{17c}$$

$$C_d = \frac{\gamma C_{\rm ox}}{2\sqrt{0.75\phi_s - V_{\rm bs}}}. \tag{17d}$$

C_d employs a modified expression ([22, p. 139]) for the surface potential at weak inversion ($\phi_{s0} = 1.5\phi_B$ instead of $2\phi_B$). The length-dependent SCE in γ (or $N_{\rm eff}$) and ϕ_s, respectively, given by (1c) and (4a), are supposed to be included in C_d. Similar to (9c), $\mu_{\rm eff}$ in (17a) is given by

$$\mu_{\rm eff} = \frac{\mu_0}{1 + V_{d\rm eff}/(E_{\rm sat} L_{\rm eff})}. \tag{17e}$$

$V_{g\rm eff}$ in (17b) is based on *charge smoothing* in linear region (small $V_{\rm ds}$) [1], [27], [23], i.e., $C_{\rm ox} V_{g\rm eff}$ gives the correct inversion charge in both strong and weak inversion regimes, but only

in linear region without considering the effects of bulk charge at high $V_{\rm ds}$. $V_{g\rm eff}$ approaches $V_{\rm gs}-V_t$ when $V_{\rm gs} \gg V_t$, which gives the correct drift current. In the subthreshold regime ($V_{\rm gs} < V_t$), however, it leads to an incorrect expression for the diffusion current. According to the BSIM3v3 manual [27, eq. (2.7.1)] the diffusion current (including the effect of high $V_{\rm ds}$) is given by

$$\mu_{\rm eff} C_{\rm ox} \frac{W}{L_{\rm eff}} v_{\rm th}^2 \frac{C_d}{C_{\rm ox}} e^{(V_{\rm gs}-V_t-V_{\rm off})/(nv_{\rm th})} \left(1-e^{-V_{\rm ds}/v_{\rm th}}\right). \tag{18a}$$

(17a) can be rewritten as

$$\begin{aligned} I_{\rm ds0} &= \mu_{\rm eff} C_{\rm ox} \frac{W}{L_{\rm eff}} V_{g\rm eff} V_{d\rm eff} \left(1-\frac{A_b}{2}\frac{V_{d\rm eff}}{V_{g\rm eff}}\right) \\ &= \mu_{\rm eff} C_{\rm ox} \frac{W}{L_{\rm eff}} V_{g\rm eff} V_{\rm de} \end{aligned} \tag{18b}$$

in which $V_{\rm de}$ is defined as

$$V_{\rm de} = V_{d\rm eff}\left(1-\frac{A_b}{2}\frac{V_{d\rm eff}}{V_{g\rm eff}}\right). \tag{18c}$$

Through a novel idea for a modified effective gate overdrive

$$V_{\rm gg} = \frac{2nv_{\rm th}\ln\left(1+e^{(V_{\rm gs}-V_t)/(2nv_{\rm th})}\right)V_{de}}{1+V_{\rm de}\dfrac{2n(C_{\rm ox}/C_d)e^{-(V_{\rm gs}-V_t-2V_{\rm off})/(2nv_{\rm th})}}{v_{\rm th}\left(1-e^{-V_{\rm ds}/v_{\rm th}}\right)}} \tag{18d}$$

for *current smoothing*, it can be seen that when $V_{\rm gs} \gg V_t$, $V_{\rm gg}$ approaches

$$(V_{\rm gs}-V_t)V_{\rm de} = [(V_{\rm gs}-V_t)V_{d\rm eff} - \tfrac{1}{2}A_b V_{d\rm eff}^2] \tag{18e}$$

as in (14a), and when $V_{\rm gs} \ll V_t$, $V_{\rm gg}$ approaches

$$v_{\rm th}^2 \frac{C_d}{C_{\rm ox}} e^{(V_{\rm gs}-V_t-V_{\rm off})/(nv_{\rm th})}\left(1-e^{-V_{\rm ds}/v_{\rm th}}\right) \tag{18f}$$

as in (18a). By replacing $V_{g\rm eff}V_{\rm de}$ in (18b) with $V_{\rm gg}$, the following unified one-region expression

$$I_{\rm ds0} = \mu_{\rm eff} C_{\rm ox} \frac{W}{L_{\rm eff}} V_{\rm gg} \tag{18g}$$

leads to correct currents for both strong-inversion and subthreshold regions. The final complete $I_{\rm ds}$ model including the effect of CLM and $R_{\rm sd}$ is then given by (16c).

F. Edge- and Diode-Leakage Current

Since the experimental data of our STI wafer exhibit significant edge-leakage current, a novel approach to modeling and extracting $I_{\rm edge}$ is developed. We first rename our developed $I_{\rm ds}$ model as $I_{\rm mos}$ for the main MOS current. Knowing the fact that edge-leakage current in STI structures is due to the parallel parasitic MOSFET along the edge of the channel [13], it is assumed that $I_{\rm edge}$ should have the same length and bias dependencies as the main MOSFET but with a different channel width (W_{e0}) and a scaled threshold voltage ($\varsigma_{v0}V_t$), as modeled by the same complete $I_{\rm mos}$ model

$$I_{\rm edge} = I_{\rm mos}(W_{e0},\, \varsigma_{v0}V_t) \tag{19a}$$

where W_{e0} and ς_{v0} are two fitting parameters to be extracted from the $I_{\rm ds}$ data when $I_{\rm edge}$ is most pronounced (Section III). Together with a simple model for the diode-leakage current

$$I_{\rm diode} = I_{s0}\left[e^{(V_{\rm bs}-V_{\rm ds})/v_{\rm th}}-1\right] \tag{19b}$$

where I_{s0} is a fitting parameter, the final compact drain current is given by

$$I_{\rm ds} = I_{\rm mos} + I_{\rm edge} - I_{\rm diode}. \tag{19c}$$

III. Model Parameter Extraction

The philosophy behind our model parameter extraction is based on three principles: 1) minimum measurement data requirement; 2) separate fitting and physical parameters; and 3) one-iteration extraction. Process-dependent *fitting* parameters ("unknown") should be extracted at the average values of the process-variable *physical* parameters ("known" or estimated), and the former should be fixed in subsequent application of the model with the latter varied for statistical analysis of process fluctuations.

The unified $I_{\rm ds}$ model requires 11 steps to extract its 23 fitting parameters, which will be detailed in this section. The idea behind our one-iteration extraction is to "calibrate" (or fit) the model at "extreme" (length and bias) conditions—the model already had the correct physics built in but it needs to fit to the "particular process" at hand, and this needs to be done only at the "boundary" cases. Of course, the technology data must include a full range of gate lengths (down to the V_t roll-off region) from the same wafer. In this paper, I–V data are based on ten devices of $L_{\rm drawn} = 10, 3, 1, 0.8, 0.5, 0.34, 0.26, 0.24, 0.22, 0.2$ μm ($W = 20$ μm), and the following "extreme" conditions are used: the longest gate $L_\infty = 10$ μm, the shortest gate $L_0 = 0.2$ μm, the medium gate (with maximum V_t) $L_m = 0.8$ μm, low $V_{\rm ds} = V_{d0} = 0.1$ V, high $V_{\rm ds} = V_{\rm dd} = 2.5$ V, low $V_{\rm bs} = V_{b0} = 0$, and high $V_{\rm bs} = V_{\rm bb} = -2.7$ V. We had altogether 200 sets of measured I–V data.

There are four independent variables (inputs): $L_{\rm drawn}$, $V_{\rm gs}$, $V_{\rm ds}$, and $V_{\rm bs}$. The process fitting parameters will be extracted with the assumed (measured or estimated) "average" values of the physical parameters, mainly $t_{\rm ox} = 59$ Å, $x_j = 75$ nm, and secondarily: $\Delta_{\rm CD} = 0$, $N_{\rm sd} = 10^{21}$ cm^{-3}, $S = 120$ nm, $v_{\rm sat} = 8\times10^6$ cm/s, and $N_{\rm ss} = 3\times10^{10}$ cm^{-2}.

The principal I–V sweep that is required is the linear $I_{\rm ds}$–$V_{\rm gs}$ curve (at V_{d0} and V_{b0}) for each device. From these curves, *linear threshold voltage* (V_{t0}) is extracted based on the maximum-g_m definition, and the corresponding *critical drain current* ($I_{\rm crit} = I_{\rm ds}$) at $V_{\rm gs} = V_{t0}$ is interpolated (i.e., "measured") for each device. The threshold voltage based on the "$I_{\rm crit}@V_{t0}$" definition [19] includes the effects of mobility and series resistance at $I_{\rm ds} = I_{\rm crit}$ [18], which is a key to the success of the model. In principle, all other required V_t data are "point measurements," i.e., one pair of (I, V) data, similar to constant–current method. After $I_{\rm crit}@V_{t0}$ determination, the required measurement data and the sequence of parameter extraction are described as follows.

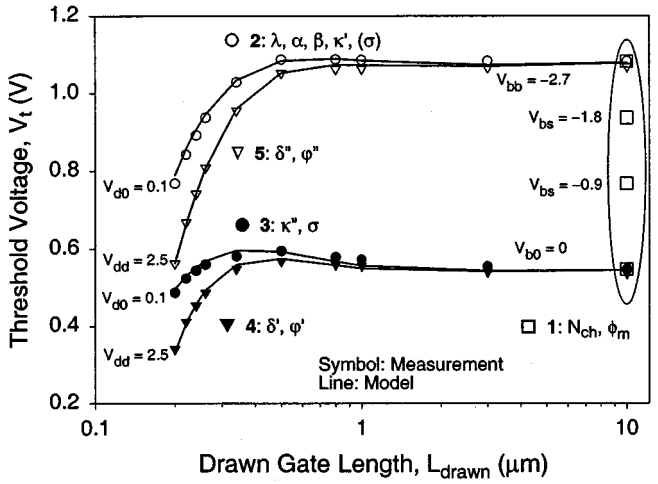

Fig. 1. Measured (symbols) and modeled (lines) V_t–$L_{\rm drawn}$ curves at the indicated bias conditions showing the 5-step V_t and $L_{\rm eff}$ parameter-extraction procedure.

A. Threshold Voltage and Effective Channel Length

Threshold voltage is the most sensitive parameter on an $I_{\rm ds}$ model, whose formulation and extraction are independent of mobility and series resistance based on our V_t definition. Our V_t model attempts to build in all major SCEs and RSCEs with its 11 fitting parameters: $N_{\rm ch}$, ϕ_m, λ, α, β, κ_0, κ_1, δ_0, δ_1, φ_0, and φ_1; together with the $L_{\rm eff}$ parameter, σ. The extraction follows a five-step procedure, as illustrated (with real data) in Fig. 1 and detailed below.

Step 1) At the longest gate L_∞: Biased at $V_{\rm ds} = V_{d0}$, measure $V_t(V_{\rm bs}) \equiv V_{\rm gs}$ @ $I_{\rm ds} = I_{\rm crit}$ for a few values of $V_{\rm bs}$. Fit the long-channel $V_t = f(V_{\rm bs})$ to the V_t–$V_{\rm bs}$ data to extract the channel-doping parameter ($N_{\rm ch}$) and workfunction (ϕ_m), which are then fixed. Values of all other parameters are irrelevant and set to zero.

Step 2) At each $L_{\rm drawn}$: Biased at $V_{\rm ds} = V_{d0}$ and $V_{\rm bs} = V_{\rm bb}$, measure $V_t(L_{\rm drawn}) \equiv V_{\rm gs}$ @ $I_{\rm ds} = I_{\rm crit}$. With a few trial values of σ (ranging, for example, from 0.5 to 1), fit $V_t = f(L_{\rm drawn}; N_{\rm ch})$ to the $V_t(V_{d0}, V_{\rm bb}) - L_{\rm drawn}$ data to extract the parameters for charge-sharing (λ), barrier-lowering (α), and RSCE (β, κ'). Values of the DIBL parameters δ and φ are set to zero and to be fine-tuned in step 4. The V_t equation has good properties in nonlinear regression for any practical $V_t - L_{\rm drawn}$ data [9].

Step 3) From the extracted $V_{t0}(V_{d0}, V_{b0}) - L_{\rm drawn}$ data: For each value of σ as well as the extracted (λ, α, β) in step 2, fit $V_t = f(L_{\rm drawn}; N_{\rm ch}, \sigma, \lambda, \alpha, \beta)$ to extract κ''. The best parameter set (σ, λ, α, β, κ') is selected with minimum error in all $V_t(L_{\rm drawn})$ values in the extraction of κ'', and fix the extracted κ in (5d)

$$\kappa_0 = \kappa''; \quad \kappa_1 = (\kappa' - \kappa'')/V_{\rm bb}. \tag{5e}$$

Steps 2 and 3 are the only steps involving the concept of optimization, but it is done with a few trial values and one iteration. This novel $L_{\rm eff}$ extraction has been proven to be simple and efficient, which gives a bias-independent $L_{\rm eff}$ that is supposed to be close to $L_{\rm met}$ since it is extracted at $V_{\rm gs} = V_{t0}$ (i.e., zero gate overdrive).

Step 4) At each $L_{\rm drawn}$: Biased at $V_{\rm ds} = V_{\rm dd}$ and $V_{\rm bs} = V_{b0}$, measure $V_t(L_{\rm drawn}) \equiv V_{\rm gs}$ @ $I_{\rm ds} = I_{\rm crit}$. Fit $V_t = f(L_{\rm drawn}; N_{\rm ch}, \sigma, \lambda, \alpha, \beta, \kappa)$ to the $V_t(V_{\rm dd}, V_{b0}) - L_{\rm drawn}$ data to extract the first set of the DIBL parameters (δ', φ').

Step 5) At each $L_{\rm drawn}$: Biased at $V_{\rm ds} = V_{\rm dd}$ and $V_{\rm bs} = V_{\rm bb}$, measure $V_t(L_{\rm drawn}) \equiv V_{\rm gs}$ @ $I_{\rm ds} = I_{\rm crit}$. Fit $V_t = f(L_{\rm drawn}; N_{\rm ch}, \sigma, \lambda, \alpha, \beta, \kappa)$ to the $V_t(V_{\rm dd}, V_{\rm bb}) - L_{\rm drawn}$ data to extract the second set of the DIBL parameters (δ'', φ''), which completes extraction of the $V_{\rm bs}$-dependent δ in (2c):

$$\delta_0 = \delta'; \quad \delta_1 = (\delta'' - \delta')/V_{\rm bb} \tag{2d}$$

and φ in (4f):

$$\varphi_0 = \varphi'; \quad \varphi_1 = (\varphi'' - \varphi')/V_{\rm bb} \tag{4g}$$

and, hence, the complete V_t model.

B. Mobility

Vertical-field mobility (μ_0) should be determined at long channel in linear mode after V_t characterization. This is shown in Fig. 2 by the solid circles.

Step 6) At the longest gate L_∞: Use the measured $I_{\rm ds} - V_{\rm gs}$ (@V_{d0}, V_{b0}) data (for $V_{\rm gs} > V_t$). Fit $I_{\rm ds} = f(V_{\rm gs}; V_t)$ to the $I_{\rm ds} - V_{\rm gs}$ data to extract the mobility parameters (μ_1, μ_2, μ_3), which are then fixed. At long channel and low $V_{\rm ds}$, bulk charge, series resistance, and CLM are negligible, hence, the long-channel I_{ds0} (with $\zeta = 1$) can be used. This step has the most parameter dependency among all the steps. The formulation as a ratio (7a) helps to reduce correlation in nonlinear regression [10].

C. Bulk Charge

With the fixed μ_0 from step 6, bulk-charge effect is then characterized from long-channel device at high $V_{\rm ds}$, where the effect is most pronounced, as shown in Fig. 2 by the solid squares.

Step 7) At the longest gate L_∞: Biased at $V_{\rm ds} = V_{\rm dd}$ and $V_{\rm bs} = V_{b0}$, measure $I_{\rm ds} - V_{\rm gs}$ (for $V_{\rm gs} > V_t$). Fit $I_{\rm ds} = f(V_{\rm gs}; V_t, \mu_0)$ to the $I_{\rm ds} - V_{\rm gs}$ data to extract the bulk-charge parameter (ζ), which is then fixed.

D. Series Resistance

Once the mobility (μ_0) and bulk-charge factor (A_b) are characterized, they are extended to short-channel devices. Series resistance ($R_{\rm sd}$) is then extracted from short channel in linear mode, as shown in Fig. 2 by the solid triangles.

Step 8) At the shortest gate L_0: Use the measured $I_{\rm ds} - V_{\rm gs}$ (@V_{d0}, V_{b0}) data (for $V_{\rm gs} > V_t$). Fit $I_{\rm ds} = f(V_{\rm gs}; V_t, \mu_0, A_b)$ to the $I_{\rm ds} - V_{\rm gs}$ data to extract the series resistance parameters (ρ, v), which are then fixed. In linear mode, CLM is unimportant, so $\xi = 0$.

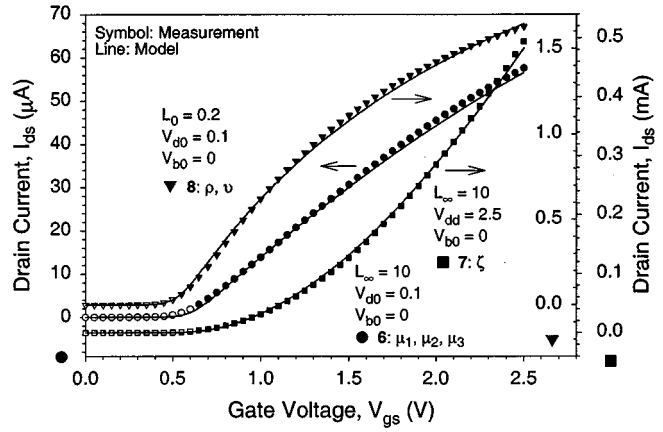

Fig. 2. Measured (symbols) and modeled (lines) $I_{\rm ds}$–$V_{\rm gs}$ characteristics at the indicated length and bias conditions showing the μ_0, A_b, and $R_{\rm sd}$ parameter-extraction procedure.

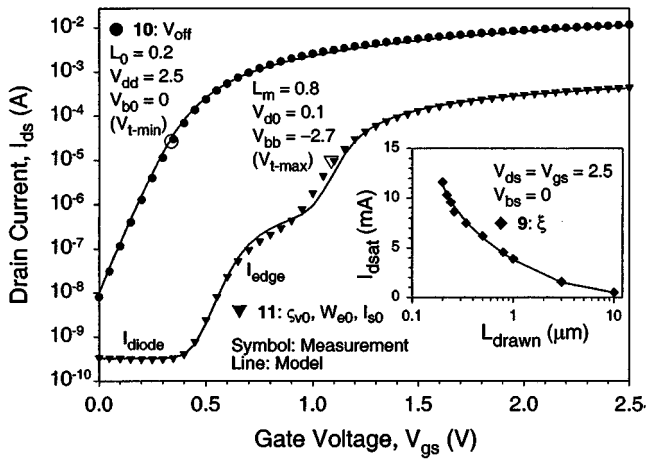

Fig. 3. Measured (symbols) and modeled (lines) $I_{\rm ds} - V_{\rm gs}$ characteristics at the indicated length and bias conditions (inset shows $I_{\rm dsat} - L_{\rm drawn}$) showing the $V_{A\rm eff}$, $V_{\rm off}$, and $I_{\rm edge}$ parameter-extraction procedure.

E. Channel-Length Modulation

CLM is next characterized from saturation currents for all gate-length devices, which is shown in the inset of Fig. 3.

Step 9) At each $L_{\rm drawn}$: Biased at $V_{\rm ds} = V_{\rm gs} = V_{\rm dd}$ and $V_{\rm bs} = V_{b0}$, measure $I_{\rm dsat}(L_{\rm drawn})$. Fit $I_{\rm ds} = f(L_{\rm drawn};\, V_t,\, \mu_0,\, A_b,\, R_{\rm sd})$ to the $I_{\rm dsat} - L_{\rm drawn}$ data to extract the CLM parameter (ξ) for the effective Early voltage, which is then fixed.

F. Subthreshold Current

After characterizing the turn-on current, subthreshold current is then extracted, which is second order compared to the turn-on current and, hence, it will not influence what has been characterized. This is shown in Fig. 3 by the solid circles.

Step 10) At the condition when V_t is minimum (i.e., L_0, $V_{\rm dd}$, and V_{b0}): measure $I_{\rm ds} - V_{\rm gs}$. Fit $I_{\rm ds} = f(V_{\rm gs};\, V_t,\, \mu_0,\, A_b,\, R_{\rm sd},\, V_{A\rm eff})$ to the $\log(I_{\rm ds}) - V_{\rm gs}$ data to extract the subthreshold-current parameter $(V_{\rm off})$, which is then fixed. At this condition, edge- and diode-leakage currents are negligible (third order), thus, W_{e0} and I_{s0} can be set to zero.

G. Edge- and Diode-Leakage Current

After the complete MOS current has been characterized, edge-leakage current in STI structures is extracted at the condition when it is most pronounced, i.e., when the main MOS $I_{\rm ds}$ is much smaller than $I_{\rm edge}$. The real data for this excellent example of extraction (i.e., extract $I_{\rm edge}$ which is embedded in $I_{\rm ds}$) is shown in Fig. 3 by the solid triangles.

Step 11) At the condition when V_t is maximum (i.e., L_m, V_{d0}, and $V_{\rm bb}$): measure $I_{\rm ds} - V_{\rm gs}$. One simple nonlinear regression is used to fit $I_{\rm ds} = I_{\rm mos}(V_{\rm gs};\, V_t,\, \mu_0,\, A_b,\, R_{\rm sd},\, V_{A\rm eff},\, V_{\rm off}) + I_{\rm mos}(V_{\rm gs};\, \varsigma_{v0} V_t,\, W_{e0}) + I_{\rm diode}(I_{s0})$ to the $\log(I_{\rm ds}) - V_{\rm gs}$ data to extract the edge-leakage $(\varsigma_{v0},\, W_{e0})$ and diode-leakage (I_{s0}) parameters.

In summary, to extract the 23 parameters used in the model, assuming N devices of varying gate lengths on the same wafer, only $(N+3)$ I–V sweeps plus $(4N+2)$ point (I, V) measurements for the various V_ts and $I_{\rm dsat}$ are needed. This compares favorably with BSIM3v3 [1], [27], which requires a minimum of 18 I–V sweeps (three devices, each at six different bias conditions).

IV. Empirical Process Correlation

One major objective and application of the developed model is for empirical process correlation, which has been demonstrated previously [9], [24]. In Section III, our model is extracted based on data from wafer #15, which has a V_t-implant dose of $\Phi = 2.5 \times 10^{12}$ cm^{-2}. We have measurement data from a split-lot with only Φ varied as 0, 1, and 4×10^{12} cm^{-2} for wafers #17, #18, and #19, respectively. There are 17 sites on each wafer from which E-test data have been measured. In this section, we present the prediction of our model with Φ as input on threshold voltage, on-state saturation current ($I_{\rm on}$), and off-state leakage current ($I_{\rm off}$), compared to the averaged values from those 17 sites. The original measured data have been presented in [9], [24].

By extracting the V_t model parameter, $N_{\rm ch}$, of wafers #17 and #19 from its long-channel (10-μm) $V_t(V_{\rm bs})$ data, a one-to-one correlation of $N_{\rm ch}$ to Φ has been found [9]:

$$N_{\rm ch} = 1.426 \times 10^{17} + 7.231 \times 10^4 \Phi \ ({\rm cm}^{-3}) \qquad (20)$$

where Φ is in cm^{-2}, and it is plotted in the inset of Fig. 4. Without using any other measurement data from wafers #17 and #19 (and none from wafer #18), excellent prediction of our model (with $L_{\rm drawn}$ and Φ as inputs) to the (average) measured V_t and $I_{\rm on}$ are shown in Figs. 4 and 5, respectively.

For $I_{\rm off}$ prediction, increase in $I_{\rm off}$ due to V_t roll-off at decreasing $L_{\rm drawn}$ should be equally-well modeled by the V_t model correlated to Φ by (20). However, since our diode-leakage modeling is too simple, long-channel $I_{\rm off}$ data from the four wafers are taken to formulate an empirical relation $I_{s0} - \Phi$

$$I_{s0} = 1/(0.07 + 0.3526\Phi) \ ({\rm pA}/\mu{\rm m}) \qquad (21)$$

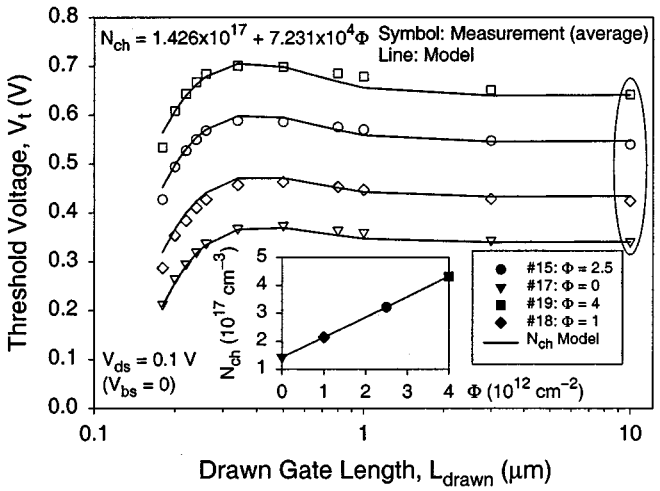

Fig. 4. Threshold voltage prediction of the model (lines) on the four wafers with different V_t-implant dose Φ shown by the average from 17 sites (symbols) through a simple correlation of the long-channel parameter $N_{\rm ch}$ to Φ (inset).

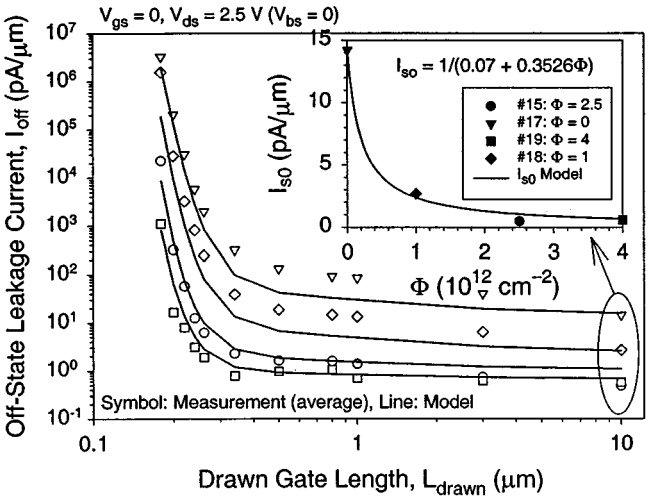

Fig. 6. Leakage current prediction of the model (lines) on the four wafers (symbols) in Fig. 4 using the same $V_t - \Phi$ empirical model together with a simple $I_{s0} - \Phi$ correlation (inset).

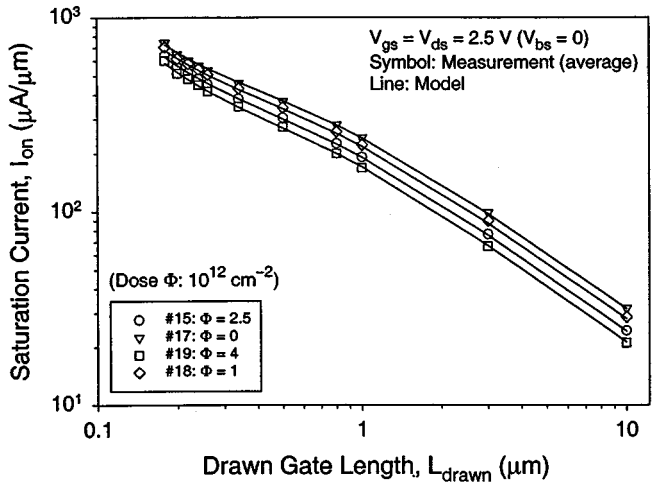

Fig. 5. Saturation current prediction of the model (lines) on the four wafers (symbols) in Fig. 4 using the same $V_t - \Phi$ correlation.

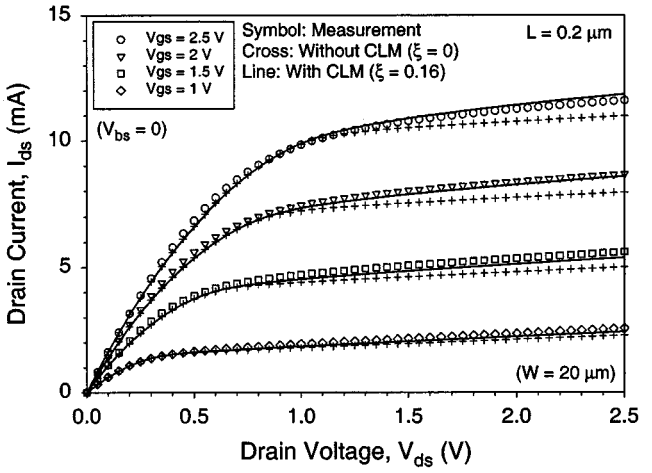

Fig. 7. Modeled $I_{\rm ds} - V_{\rm ds}$ characteristics with (solid lines) or without (crosses) CLM for the 0.2-μm device compared to the measurement data (symbols).

where Φ is in 10^{12} cm^{-2}, which is shown in the inset of Fig. 6. Using (20) and (21) with Φ as input, our model prediction to the measured $I_{\rm off} - L_{\rm drawn}$ is quite well, as shown in Fig. 6.

The excellent prediction of our model with a very simple process correlation is the result of the correct physics that has been built into the model. This approach, combined with a carefully designed wafer split, can be very efficient and useful in reducing experimental wafer split-lot.

V. Results and Discussion

Besides the presented unique approach to CM formulation through effective transformation, we proposed three new models: CLM (with $V_{A\rm eff}$), subthreshold modeling (with $V_{\rm gg}$), and edge-leakage prediction. These results will be presented in this section. Other sample results based on the same model have been presented in [24].

Fig. 7 shows the modeled $I_{\rm ds} - V_{\rm ds}$ curves with two values of $\xi = 0$ (crosses) and $\xi = 0.16$ (solid lines) for the 0.2-μm device compared to the measurement data (symbols). When $\xi = 0$ (no CLM), the small finite drain conductance in saturation is due to inclusion of the bulk-charge current. It can be seen that the $V_{\rm gs}$-dependent $V_{A\rm eff}$ model (see [12, Fig. 5]) can predict CLM very well.

Fig. 8 compares the new $V_{\rm gg}$ model (18d) before $I_{\rm edge}$ extraction (dotted lines), which is fitted to the $\log(I_{\rm ds}) - V_{\rm gs}$ (@ $V_{\rm dd}$) data (triangles), with the BSIM expression (17a) (crosses), which is fitted to the $\log(I_{\rm ds}) - V_{\rm gs}$ (@ V_{d0}) data (circles), with separate extraction of $V_{\rm off}$ (after the same turn-on current extraction). For the BSIM expression, a thermal voltage (v_{th}) has to be added to $V_{g\rm eff}$ [1], [27] in order to get the correct slope, but the extracted $V_{\rm off} = -0.1744$ V fails to predict the data well at high $V_{\rm ds}$, as it is only valid at low $V_{\rm ds}$ [1], [27]. Our new model, however, can easily fit the data with the correct slope at high $V_{\rm ds}$ (in Step 10) when edge leakage is negligible, and further extract

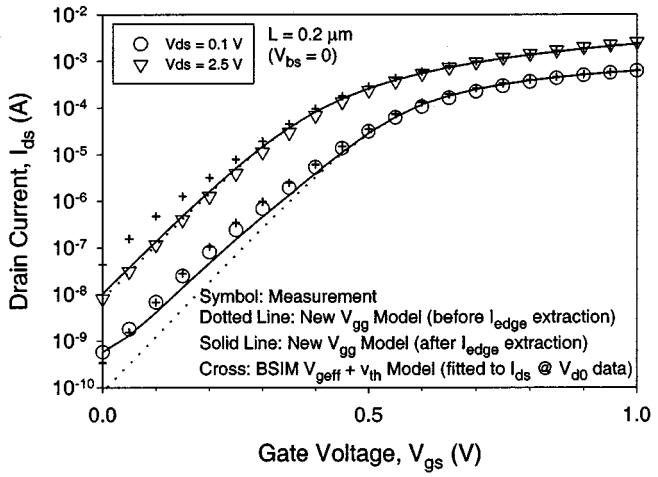

Fig. 8. Subthreshold $\log(I_{\rm ds}) - V_{\rm gs}$ characteristics of the new $V_{\rm gg}$ model before (dotted lines) and after (solid lines) $I_{\rm edge}$ extraction, and compared with the BSIM $V_{\rm geff}$ expression (crosses).

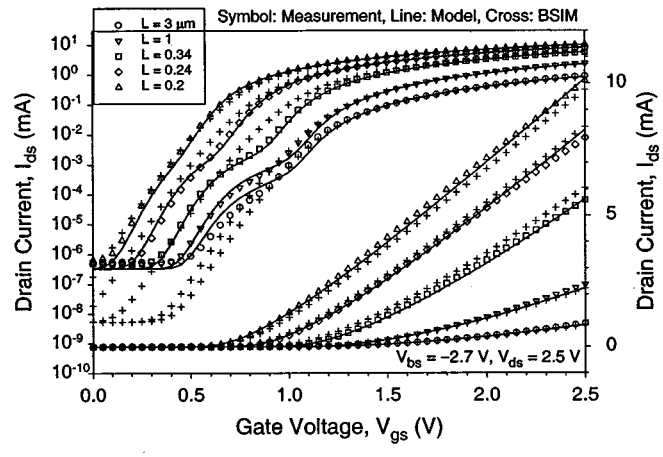

Fig. 9. Prediction of the length-dependent subthreshold and edge-leakage current (left axis) and turn-on current (right axis) of our one-region $I_{\rm ds}$ model (lines) on the measurement data (symbols) for five devices, and compared to the BSIM3v3 prediction (crosses).

$I_{\rm edge}$ at high V_t (in Step 11), which has little influence on the extracted subthreshold slope at low V_t, as shown by the solid lines in Fig. 8.

The best proof of validity and accuracy of our model is the excellent prediction of the edge-leakage currents at various gate lengths and biases. A sample result is shown in Fig. 9 (solid lines) in which *none* of the data (symbols) has been used in parameter extraction. Also shown (crosses) as a comparison are the BSIM3v3 results whose parameters are extracted automatically by BSIMPro using *all* (200 sets) of the available I–V data. Excellent predictions of the $V_{\rm bs}$-dependent subthreshold current ("hump") at fixed gate length have also been obtained, one of which was reported in [24], which further validates our unified model and the approach to STI current modeling. This is believed to be the first one-region CM for STI edge-leakage current, which has only two fitting parameters and one extraction.

Due to length limitation, we show in Fig. 10 one sample result of the $g_m/I_{\rm ds}$ versus $V_{\rm gs}$ curves for two devices in linear and

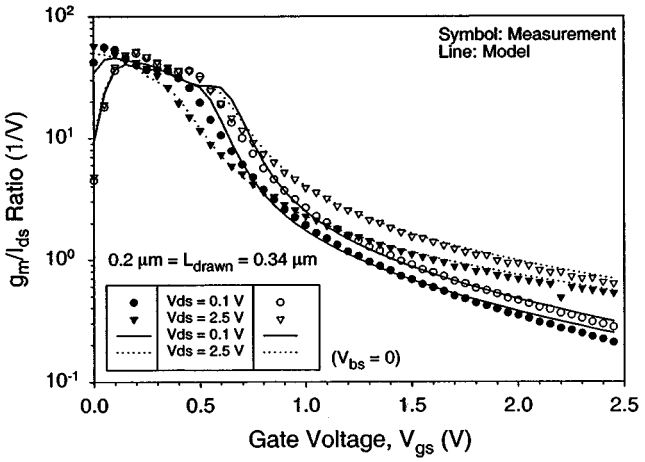

Fig. 10. Measured (symbols) and modeled (lines) $g_m/I_{\rm ds}$ – $V_{\rm gs}$ characteristics in linear and saturation mode for two devices.

saturation regions, which demonstrates smoothness of our one-region model.

So far, our model does not include many effects, such as narrow channel, substrate and gate leakage, poly depletion, quantum effects, etc., which can and will be formulated following the same approach.

With the separation of process and physical parameters, device performance fluctuations due to statistical process variations can be studied following the approach in [25]. For example, after process-dependent fitting parameters have been extracted and fixed, V_t, $I_{\rm on}$,, and $I_{\rm off}$ fluctuations can be related to variations in $t_{\rm ox}$, x_j, $\Delta_{\rm CD}$, S, etc., which, in turn, are due to process variations. This further research will be carried out as a novel application of the developed model.

The model has been formulated and demonstrated with nMOSFETs. A direct application of the model following the described extraction has been applied to the pMOSFETs on the same wafer (#15), which demonstrated equally-well accuracy. The model has been automated and can be applied to automatic wafer test systems as a quick and reliable aid to technology developers, at least in the 0.25-μm regime. This unified compact $I_{\rm ds}$ model has been named as **Xsim**, which will be implemented in the mixed-mode circuit simulator (Xsim) [26].

VI. Conclusion

In conclusion, a unified one-region $I_{\rm ds}$ equation for DSM MOSFETs has been developed and verified through physics-based effective transformation. The novelty lies behind the philosophy of one-iteration parameter extraction, which follows a prioritized sequence for extracting the parameters being modeled at the condition when their effect is most pronounced, with process-dependent parameters fitted to the measured terminal $I_{\rm ds}$ data with assumed average physical parameters. The simple form of the formulated equations is a result of building SCEs into long-channel models, resulting in a true single-piece $I_{\rm ds}$ model for all gate lengths (no binning). Development of technology-dependent CMs for circuit-level simulation becomes one of the grand challenges for the 0.1-μm

technology node. The demonstrated approach in model formulation and process correlation will prove to be extremely useful for DSM technology modeling, process monitoring, as well as in bridging technology developers to circuit designers.

ACKNOWLEDGMENT

The authors would like to thank Chartered Semiconductor Manufacturing Ltd. for providing the experimental data for this work. K. Y. Lim would like to acknowledge the scholarship awarded by Nanyang Technological University and National Science and Technology Board, Singapore.

REFERENCES

[1] Y. Cheng, M.-C. Jeng, Z. Liu, J. Huang, M. Chan, K. Chen, P. K. Ko, and C. Hu, "A physical and scalable I–V model in BSIM3v3 for analog/digital circuit simulation," *IEEE Trans. Electron Devices*, vol. 44, pp. 277–287, Feb. 1997.

[2] 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, June 1994.

[3] S.-L. Jang and M.-C. Hu, "An analytical drain current model for submicrometer and deep submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 44, pp. 1896–1902, Nov. 1997.

[4] S.-L. Jang, S.-S. Liu, and C.-J. Sheu, "A compact LDD MOSFET I–V model based on nonpinned surface potential," *IEEE Trans. Electron Devices*, vol. 45, pp. 2489–2498, Dec. 1998.

[5] G. Baccarani and S. Reggiani, "A compact double-gate MOSFET model comprising quantum-mechanical and nonstatic effects," *IEEE Trans. Electron Devices*, vol. 46, pp. 1656–1666, Aug. 1999.

[6] C.-L. Lou, W.-K. Chim, D. S.-H. Chan, and Y. Pan, "A novel single-device DC method for extraction of the effective mobility and source-drain resistances of fresh and hot-carrier degraded drain-engineered MOSFET's," *IEEE Trans. Electron Devices*, vol. 45, pp. 1317–1323, June 1998.

[7] J. A. M. Otten and F. M. Klaassen, "A novel technique to determine the gate and drain bias dependent series resistance in drain engineered MOSFET's using one single device," *IEEE Trans. Electron Devices*, vol. 43, pp. 1478–1488, Sept. 1996.

[8] T. Skotnicki, G. Merckel, and T. Pedron, "The voltage- doping transformation: A new approach to the modeling of MOSFET short-channel effects," *IEEE Electron Device Lett.*, vol. 9, pp. 109–111, Mar. 1988.

[9] X. Zhou, K. Y. Lim, and D. Lim, "A general approach to compact threshold voltage formulation based on 2-D numerical simulation and experimental correlation for deep-submicron ULSI technology development," *IEEE Trans. Electron Devices*, vol. 47, pp. 214–221, Jan. 2000.

[10] K. Y. Lim and X. Zhou, "A physically-based semi-empirical effective mobility model for MOSFET compact I–V modeling," *Solid-State Electron*, vol. 45, no. 1, pp. 193–197, Jan. 2001.

[11] ——, "A physically-based semi-empirical series resistance model for deep-submicron MOSFET I–V modeling," *IEEE Trans. Electron Devices*, vol. 47, pp. 1300–1302, June 2000.

[12] X. Zhou and K. Y. Lim, "A compact MOSFET I_{ds} model for channel-length modulation including velocity overshoot," in *Proc. ISDRS-99*, Charlottesville, VA, Dec. 1999, pp. 423–426.

[13] A. Bryant, W. Haensch, S. Geissler, J. Mandelman, D. Poindexter, and M. Steger, "The current-carrying corner inherent to trench isolation," *IEEE Electron Device Lett.*, vol. 14, pp. 412–414, Aug. 1993.

[14] K. Y. Lim and X. Zhou, "Modeling of threshold voltage with nonuniform substrate doping," in *Proc. IEEE ICSE'98*, Malaysia, Nov. 1998, pp. 27–31.

[15] 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 MOSFET's," *IEEE Trans. Electron Devices*, vol. 40, pp. 86–94, Jan. 1993.

[16] K. Y. Lim, "Design, modeling, and characterization of submicron MOSFETs," Nanyang Technical University, First-Year Conversion Report, 1997/98.

[17] D. Esseni, H. Iwai, M. Saito, and B. Ricco, "Nonscaling of MOSFET's linear resistance in the deep submicrometer regime," *IEEE Electron Device Lett.*, vol. 19, pp. 131–133, Apr. 1998.

[18] X. Zhou, K. Y. Lim, and D. Lim, "A new 'critical-current at linear-threshold' method for direct extraction of deep-submicron MOSFET effective channel length," *IEEE Trans. Electron Devices*, vol. 46, pp. 1492–1494, July 1999.

[19] ——, "A simple and unambiguous definition of threshold voltage and its implications in deep-submicron MOS device modeling," *IEEE Trans. Electron Devices*, vol. 46, pp. 807–809, Apr. 1999.

[20] A. G. Sabnis and J. T. Clemens, "Characterization of the electron mobility in the inverted < 100 > Si surface," in *IEDM Tech. Dig.*, 1979, pp. 18–21.

[21] P. K. Ko, *VLSI Electronics: Microstructure Science*. New York: Academic, 1988, vol. 18, ch. 1.

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

[23] Y. Cheng, K. Chen, K. Imai, and C. Hu, "A unified MOSFET channel charge model for device modeling in circuit simulation," *IEEE Trans. Computer-Aided Design*, vol. 17, pp. 641–644, Aug. 1998.

[24] X. Zhou and K. Y. Lim, "A novel approach to compact I–V modeling for deep-submicron MOSFET's technology development with process correlation," in *Proc. MSM'2000*, San Diego, CA, Mar. 2000, pp. 333–336.

[25] R. Sitte, S. Dimitrijev, and H. B. Harrison, "Device parameter changes caused by manufacturing fluctuations of deep submicron MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, pp. 2210–2215, Nov. 1994.

[26] T. Tang and X. Zhou, "Multi-level digital/mixed-signal simulation with automatic circuit partition and dynamic delay calculation," *J. Modeling and Simulation of Microsystems*, vol. 1, no. 2, pp. 83–89, 1999.

[27] Y. Cheng *et al.*, "BSIM3v3 Manual," Univ. California, Berkeley, 1997–1998.

Xing Zhou (S'88–M'91–SM'99) received the B.E. degree from Tsinghua University, Beijing, China, in 1983, and the M.S. and Ph.D. degrees in electrical engineering from the University of Rochester, Rochester, NY, in 1987 and 1990, respectively.

From 1990 to 1991, he was a Research Associate in the Department of Electrical Engineering, the University of Rochester, where he worked on hot-carrier injection phenomena in MOS devices, as well as development of CAD tools for mixed-signal circuit simulation. From 1992 to 1995, he was a Research Fellow in the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore, where he worked on Monte Carlo and numerical modeling of semiconductor and optoelectronic devices as well as mixed-signal circuit modeling and simulation. He is currently an Associate Professor in the same school at NTU, teaching and researching deep-submicron CMOS technology and device modeling. In November and December of 1997 as well as in February and March 2001, he was a Visiting Fellow at the Center for Integrated Systems, Stanford University, Stanford, CA. His main interests are in the area of semiconductor device physics and modeling, novel device structures, CM development for advanced devices, technology modeling and simulation, mixed-signal CAD tools, hot-carrier transport, and ultrafast phenomena.

Dr. Zhou is a member of the IEEE EDS VLSI Technology and Cirsuits as well as Compact Modeling Technical Committees. He is listed in the *Marquis Who's Who in the World* and *Who's Who in Science and Engineering.*

Khee Yong Lim (S'98) was born in Malacca, Malaysia, in 1975. He received the B.Eng. (Hons.) degree in electrical engineering in 1997 from Nanyang Technological University, Singapore, where he is currently pursuing the Ph.D. degree in microelectronic engineering.

He is currently with Chartered Semiconductor Manufacturing Ltd., Singapore. His present research interests focus on physical modeling, simulation, and characterization of deep-submicrometer MOSFETs as well as the improved structures for transistor scaling.