A General Approach to Compact Threshold Voltage Formulation Based on 2-D Numerical Simulation and Experimental Correlation for Deep-Submicron ULSI Technology Development

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

***Abstract*—A unified compact threshold voltage model is developed, which accounts for the normal and reverse short-channel effects with full range of body- and drain-bias conditions, and has been verified with experimental data down to 0.18 μm. The model only has five process-dependent fitting parameters with a simple one-iteration extraction procedure, and can be correlated to process variables for aiding new deep-submicron technology development. The approach to the model formulation is original and general, and can be extended to other key device performance parameters.**

***Index Terms*—Compact modeling, deep-submicron MOSFET, reverse short-channel effect, short-channel effect, TCAD, technology development, threshold voltage.**

I. Introduction

THERE exists a variety of definitions, models, and methods for measuring or extracting the threshold voltage (V_t) of a MOSFET. However, few analytic V_t model [1] exists that can accurately predict deep-submicron gate-length dependent threshold behavior, including V_t roll-off [due to short-channel effects, (SCE's)] and roll-up [due to reverse short-channel effects, (RSCE's)], with a full range of drain (V_{ds}) and substrate bias (V_{bs}) conditions. Although technology computer-aided design (TCAD) tools are becoming more popular in aiding new technology development, they require significant efforts in tool calibration and, very often, are difficult to extrapolate to regions where the process/device models are not well calibrated or invalid. Compact models for circuit simulation, such as the BSIM model [2], are empirically-based and efficient, but require extraction of a large number of parameters. Therefore, they are not meant for use by technology developers since process variables, such as implant dose or oxidation temperature, are not expressed explicitly.

This paper presents a general approach to formulating compact threshold voltage model based on a combined two-dimensional (2-D) numerical simulation, empirical formulation, and experimental correlation, with the main objective of aiding deep-submicron CMOS ULSI technology development. The approach is based on the following basic assumption (or belief): Analytically-derived equations incorporate device physics but usually cannot model complex devices for a wide range of variables because of the idealized assumptions that have to be made; whereas the 2-D numerical model (even noncalibrated) can provide useful information on the structural nonuniformity and carrier transport nonlinearity, which is supposed to be a good approximation of the real device. Whenever experimental data are unavailable or inaccessible, TCAD can be used to aid model formulation in finding the "global" functionality to cover a wide variable range, leaving "insensitive" parameters for empirical fitting. Nonphysics-based models that require curve fitting for parameter extraction are, in general, not useful, but empirical fitting is inevitable if the model is to be applied to real deep-submicron devices. If the model such developed (as opposed to analytical derivation or nonphysical fitting) can cover a wide range of variables with reasonable accuracy, the behavior of the real device can be predicted by *interpolation* (as opposed to TCAD calibration and extrapolation). Furthermore, if the model parameters can be correlated to process variables, a very efficient, empirically-based model will then be extremely useful in aiding new technology development.

Manuscript received August 25, 1998; revised July 1, 1999. The review of this paper was arranged by Editor C. Y. Yang.

X. Zhou and K. Y. Lim are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: exzhou@ntu.edu.sg).

D. Lim is with Chartered Semiconductor Manufacturing Ltd., Singapore 738406.

Publisher Item Identifier S 0018-9383(00)00167-2.

II. Model Formulation

This section presents the general approach (not just the equations) used to formulate the compact threshold voltage model for nMOSFET based on the experimental data of an 0.25-μm CMOS logic technology as well as the simulated data of the corresponding devices using MEDICI [3] within the drift-diffusion formalism. The experimental devices cover drawn gate lengths (L_g) from 10 μm down to 0.18 μm, with gate oxide thickness (t_{ox}) of 59 Å and lightly-doped drain (LDD) junction depth (x_j) of 75 nm. Various device structures are created with Gaussian-shaped channel and S/D doping profiles as well as lateral Gaussian profiles to simulate boron pile-up due to pocket-implant for the observed RSCE [4].

Both measured and simulated I_{ds}–V_{gs} data at various V_{ds} and V_{bs} conditions are used to extract the threshold voltages with the following definition. In linear mode (low V_{ds}), the threshold voltage (V_{t0}) is extracted from extrapolation of the linear I_{ds}–V_{gs} curve at maximum slope to zero current (maximum-g_m definition). In saturation mode (intermediate and high

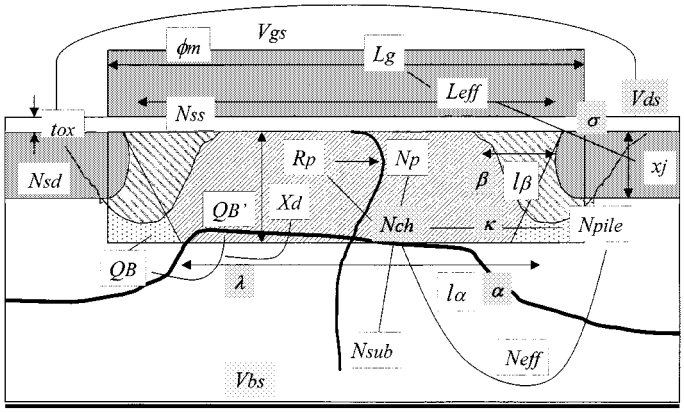


Fig. 1. Macromodel view of an nMOS transistor. All the major physical and empirical parameters of the proposed model are depicted and explained in the paper.

V_{ds}), the threshold voltage is defined as the gate voltage when $I_{ds} = I_{crit}$, where the critical current is the drain current at $V_{gs} = V_{t0}$. This modified constant-current definition (known as "critical-current at linear-threshold") [5] not only avoids the arbitrary choice in the normal constant-current definition of V_t, but also provides a consistent and unique definition of V_t for all intermediate V_{ds}. More importantly, it gives a consistent value for linear and saturation threshold voltages for every transistor of the same gate length.

The macromodel view of an nMOS transistor (our "mental image of reality") is shown in Fig. 1, which illustrates all major physical parameters and variables used in our compact V_t model. The general idea is to retain the simple one-dimensional (1-D) form of the ideal V_t equation while building the 2-D SCE in the channel charge, surface potential, and effective channel doping. In essence, this is to "derive" (or "formulate") a V_t model based on the *theoretical* (surface potential) definition to match that based on the *measurement* [(terminal current–voltage (I–V)] definition. This is different from the approach [6] of "defining" a V_t with all SCE's included, and then, measuring (or extracting) the model parameters based on that definition. The following is a brief outline of the way our model is formulated.

- The theoretical V_t definition of long-channel, uniformly-doped transistors with substrate doping $N_{\rm sub}$ is used as the basis of the V_t model.
- $N_{\rm sub}$ is modified to account for the nonuniform channel doping profile (N_{ch}) for long-channel devices, characterized by two parameters, peak doping (N_p) and peak location (R_p) [7].
- The charge-sharing model [8] is included to account for SCE with a fitting parameter λ.
- The potential barrier lowering effect at short channel [9] is added with a fitting parameter α to model the characteristic length l_α.
- RSCE is modeled by adding a boron pile-up charge ($N_{\rm pile}$), which relates to the channel doping (N_{ch}) by a fitting parameter κ, to the effective channel doping ($N_{\rm eff}$) with a characteristic length l_β modeled by a fitting parameter β.

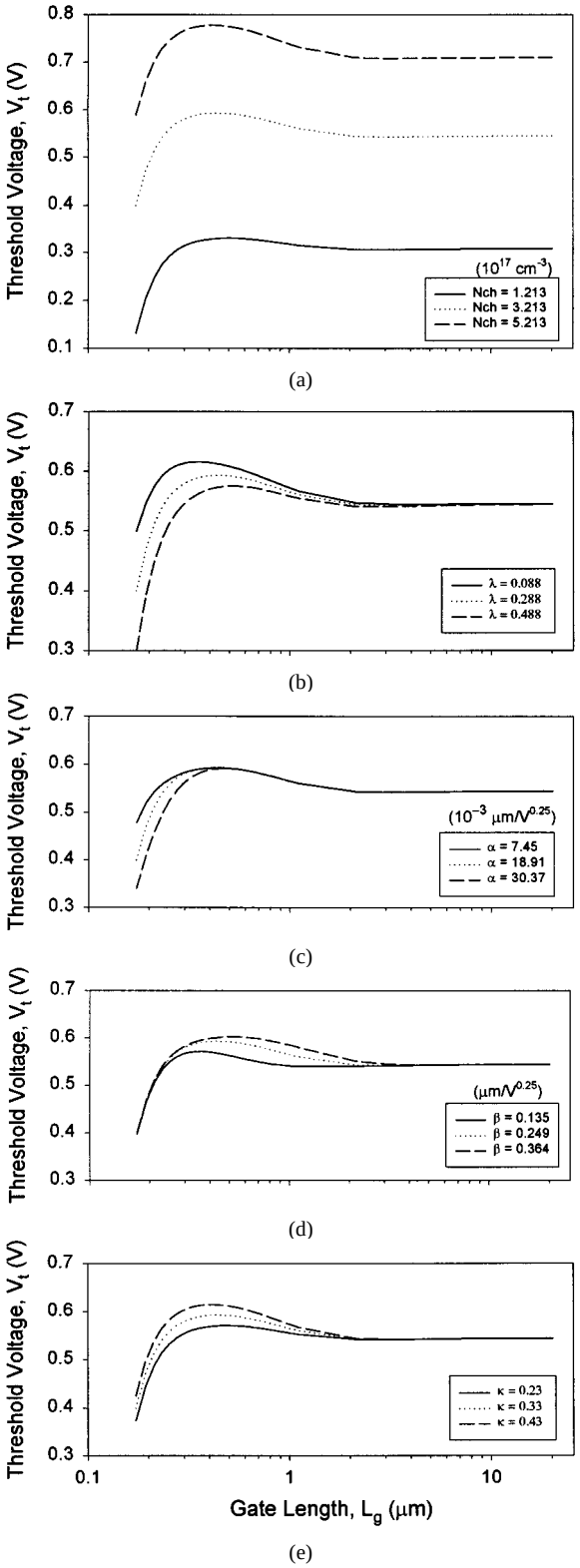


Fig. 2. The compact V_t model behavior. One fitting parameter is changed at a time with others taking their nominal values (dotted lines).

- V_t dependence on V_{bs} (body effect) and V_{ds} (drain-induced barrier lowering, DIBL) are modeled by first-order linear approximations of the fitting parameters κ and (α, β), respectively.

- Effective channel length ($L_{\rm eff}$) is related to the drawn gate length (L_g) by a simple relationship with a fitting parameter σ.

It should be emphasized that the fitting parameters in our model all have their respective physical meanings.

The theoretical definition of the threshold voltage

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

is based on the "strong-inversion" condition, i.e., the surface potential is twice of the bulk Fermi potential

$$\phi_{s0} = 2\phi_B \tag{1a}$$

where

$$\phi_B = \frac{kT}{q}\,\ln\left(\frac{N_{\rm sub}}{n_i}\right) \tag{1b}$$

is the bulk Fermi potential

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

is the body factor, and

$$V_{\rm FB} = \phi_m - \left(\chi + \frac{1}{2}E_g + \phi_B\right) - \frac{qN_{ss}}{C_{ox}} \tag{1d}$$

is the flat-band voltage. ϕ_m is the poly-gate workfunction, N_{ss} the fixed oxide charge density, χ the electron affinity, E_g the energy bandgap, n_i the intrinsic carrier concentration, and

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \tag{1e}$$

is the gate oxide capacitance. Verification of the relationship between (1) and the maximum-g_m definition of V_t has been established through numerical simulations [7].

To account for the nonuniform channel doping profiles, all $N_{\rm sub}$ in (1) are replaced by

$$N_{ch} = N_{tot}(1 - \eta R_p^2) \tag{2}$$

where

$$N_{tot} = N_{\rm sub} + N_p \tag{2a}$$

with two parameters, peak doping N_p and peak location R_p, to characterize the final long-channel doping profile. Equation (2) is developed through fitting V_t equation to a large number of MEDICI data with different N_p and R_p values. Through analytic derivation [7] and numerical fitting, the following "global" function for η has been found:

$$\eta = \frac{1}{3c}\left(1 - \frac{N_{\rm sub}}{N_{tot}}\right) \tag{2b}$$

where c is the first fitting parameter, which is "insensitive" to a large range of N_p and R_p values. Since no measured profiles available from the experimental devices, N_{ch} will be used as our first fitting parameter in this work. The validity of (2) has been verified with a set of MEDICI simulation data [7], which is more applicable to surface-doped devices.

Next, the charge-sharing model [8] is applied to model the SCE on V_t

$$V_t = V_{\rm FB} + \phi_s + \gamma\sqrt{\phi_{s0} - V_{bs}}\left(1 - \lambda\frac{X_d}{L_{\rm eff}}\right) \tag{3}$$

where

$$X_d = \varsigma\sqrt{\phi_{s0} - V_{bs}}, \quad \varsigma = \sqrt{\frac{2\varepsilon_{\rm Si}}{qN_{ch}}} \tag{3a}$$

is the channel depletion width, $L_{\rm eff}$ is the effective channel length, and λ is the second fitting parameter. ϕ_s and ϕ_{s0} are the surface potentials for short- and long-channel devices, respectively.

From the result of the quasi-2–D model [9], it is derived that the surface potential barrier in short-channel device is lowered by $\Delta\phi_s$ due to the S/D finger fields such that it is smaller than $2\phi_B$ at strong-inversion condition

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

where

$$\Delta\phi_s = \left(V_{bi} - \phi_{s0} + \frac{1}{2}V_{ds}\right)\frac{1}{\cosh(L_{\rm eff}/2l_\alpha)} \tag{4a}$$

with a characteristic length

$$l_\alpha = \alpha(\phi'_{s0} - V_{bs})^{0.25} \tag{4b}$$

and α is the third fitting parameter.

$$V_{bi} = \frac{kT}{q}\,\ln\left(\frac{N_{sd}N_{ch}}{n_i^2}\right) \tag{4c}$$

is the S/D built-in potential and N_{sd} is the doping concentration in the LDD region. In (4b),

$$\phi'_{s0} = 2\,\frac{kT}{q}\,\ln\left(\frac{N_{ch}}{n_i}\right) \tag{4d}$$

is the long-channel surface potential at strong inversion. It is found that potential-barrier lowering is needed in addition to the charge-sharing effect since they describe different contributions to the observed SCE (V_t roll-off).

To model the RSCE, we assume that it is due to the 2-D boron pile-up, which modifies the channel doping N_{ch} [4]. Although our ideal MEDICI devices with lateral Gaussian pile-up profiles can predict the observed RSCE quite well, there is obvious no simple analytical solutions [1]. To retain the simple 1-D V_t formula, it is assumed that a pile-up charge N_{pile} is added to the effective channel doping

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

which is modulated by a channel-length dependent function analogous to the quasi-2–D model solution (4a), and the characteristic length

$$l_\beta = \beta(\phi'_{s0} - V_{bs})^{0.25} \tag{5a}$$

for the pile-up charge is controlled by the forth fitting parameter β. This assumption is purely intuitive and empirical. Through a lot of fitting to the measured and simulated V_t data, it is found

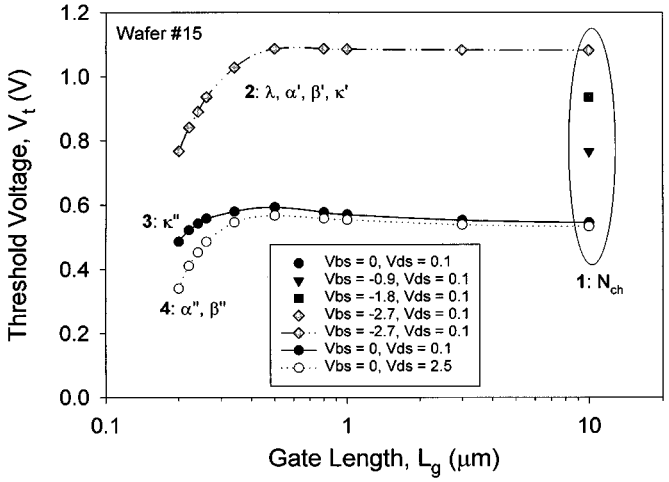


Fig. 3. Parameter-extraction procedure, with four measurements and four steps, as explained in the text.

that the ratio of N_{pile} to N_{ch} is always relatively constant. It is then decided to express

$$N_{pile} = \kappa N_{ch} \tag{5b}$$

and use κ as the fifth fitting parameter. N_{eff} of (5) should replace all N_{ch} in the previous equations [except ϕ'_{s0} of (4d) to avoid convergence problem during nonlinear regression]. This RSCE-model development demonstrates a general approach to empirical formulation in combining analytical and numerical solutions.

To model the body effect, it is observed from the measured and simulated $V_t(V_{bs}) - L_g$ data that the amount of V_t roll-up (RSCE) reduces and diminishes as V_{bs} becomes more negative. (This is contrary to what was observed in [1].) This can be explained by our simple model in which the "effective" channel depletion charge (N_{ch}) increases at increasing $|V_{bs}|$ while N_{pile} is supposed to be fixed. By using a linear approximation for the ratio κ dependence on V_{bs}

$$\kappa = \kappa_0 + \kappa_1 V_{bs} \tag{6a}$$

and extracting two values at low and high V_{bs}, it is found that the RSCE at different V_{bs} can be predicted reasonably well. In the same way, DIBL is modeled by modifying α and β to include a linear dependence on V_{ds}

$$\alpha = \alpha_0 + \alpha_1 V_{ds} \tag{6b}$$

$$\beta = \beta_0 + \beta_1 V_{ds} \tag{6c}$$

which is based on the assumption that the characteristic lengths l_α and l_β, should be affected by the lateral potential V_{ds}.

Finally, although the effective channel length (L_{eff}) can be extracted experimentally or numerically, they are not consistently and unambiguously defined. The following simple model is adopted:

$$L_{\text{eff}} = L_g - 2\sigma x_j \tag{7}$$

such that the V_t model is expressed in terms of the well-defined drawn gate length, L_g. σ is considered as a fitting parameter to model the lateral diffusion of LDD junctions, which can usually be fixed at 0.7–0.75. Complete modeling of L_{eff} has been presented elsewhere [10].

The final compact V_t model, including SCE and RSCE as well as V_{bs} and V_{ds} dependencies, is given by

$$\begin{aligned} V_t = {} & \phi_m - \left(\chi + \frac{1}{2}E_g\right) - \frac{qN_{ss}}{C_{ox}} + \frac{1}{2}\phi_{s0} \\ & - \left(V_{bi} - \phi_{s0} + \frac{1}{2}V_{ds}\right)\frac{1}{\cosh(L_{\text{eff}}/2l_\alpha)} \\ & + \frac{\sqrt{2q\varepsilon_{\text{Si}}N_{\text{eff}}(\phi_{s0} - V_{bs})}}{C_{ox}} \\ & \cdot \left(1 - \frac{\lambda}{L_{\text{eff}}}\sqrt{\frac{2\varepsilon_{\text{Si}}}{qN_{\text{eff}}}(\phi_s - V_{bs})}\right) \end{aligned} \tag{8}$$

with only five fitting parameters: N_{ch}, λ, α, β, and κ. The major physical parameters in this model are t_{ox}, x_j, ϕ_m, N_{ss}, N_{sd}, N_p, and R_p. The principal variables are L_g, V_{bs}, and V_{ds}.

The behavior of the developed V_t model is shown in Fig. 2 as one of the five fitting parameters is varied with others taking their nominal values. It is observed that only N_{ch} influences the long-channel V_t (parallel shift of the V_t–L_g curve) while others only affect the short-channel behavior. λ models the amount of V_t roll-off (due to charge sharing). α influences the onset of V_t roll-off (due to lateral spread of surface potential change). β controls the onset of V_t roll-up (due to lateral spread of the pile-up charge) with no effects on long- and short-channel V_t. κ determines the amount of V_t roll-up (due to the magnitude of the pile-up charge) with little effects on long- and short-channel V_t. These behaviors can be and are all well explained by the physics the respective parameters represent. This is the first model that combines all of the above effects into one single equation with physics-based parameters. The success of the model is attributed to the separate, yet unified, modeling of the individual SCE's.

III. Parameter Extraction

An empirically-based model will not be useful unless a consistent and simple parameter-extraction procedure can be adopted, nor will it be of any interest if experimental data are needed to extract the parameters every time the model is used.

Our compact V_t model uses a very simple, four-step, one-iteration extraction procedure, which requires the measured (or simulated) V_t–L_g data from the same process (or devices) with only four sets of measurements: V_t (long L_g) for all V_{bs}, V_t (low V_{ds}, high V_{bs}), V_t (low V_{ds}, low V_{bs}), and V_t (high V_{ds}, low V_{bs}) for all L_g. The four-step extraction procedure is illustrated in Fig. 3 and described as follows.

Step 1) A simple nonlinear regression on (1) (with N_{sub} replaced by N_{ch}), fitting the long-channel V_t–V_{bs} data to extract N_{ch}. Full equation (8) can also be used by setting the parameters (λ, α, β, κ) to zero since they will not affect the long-channel V_t. The extracted N_{ch} value will be fixed in the subsequent steps.

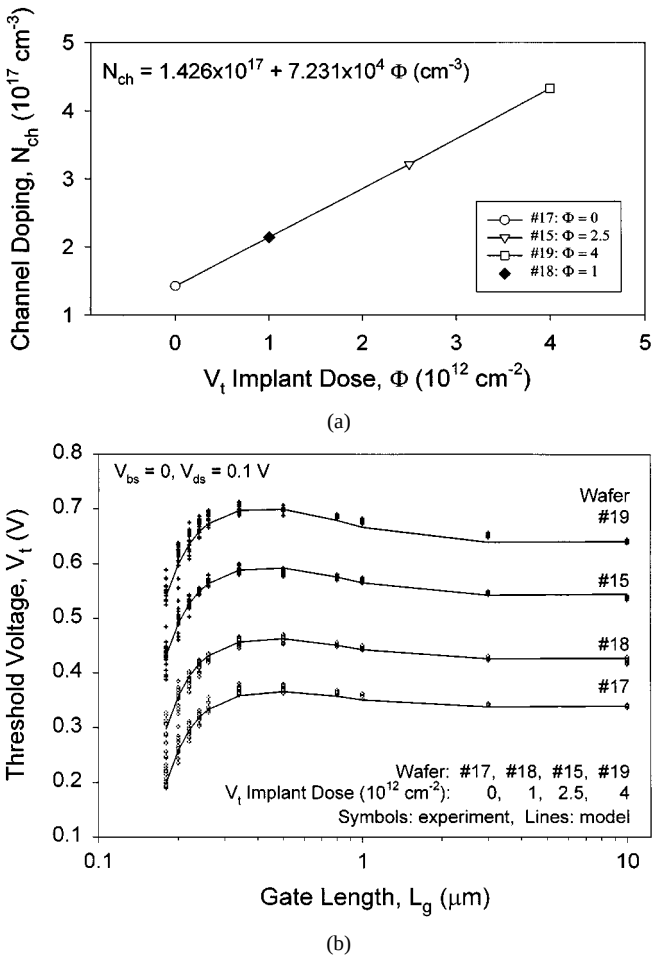


Fig. 4. Experimental correlation of the model parameter N_{ch} to V_t implant dose Φ. The full set of fitting parameters is extracted from wafer #15 data. The empirical correlation $N_{ch} - \Phi$ in (a) is found from long-channel data of wafers #17 and #19. Prediction of wafer #18 in (b) is made by the compact model with Φ as input.

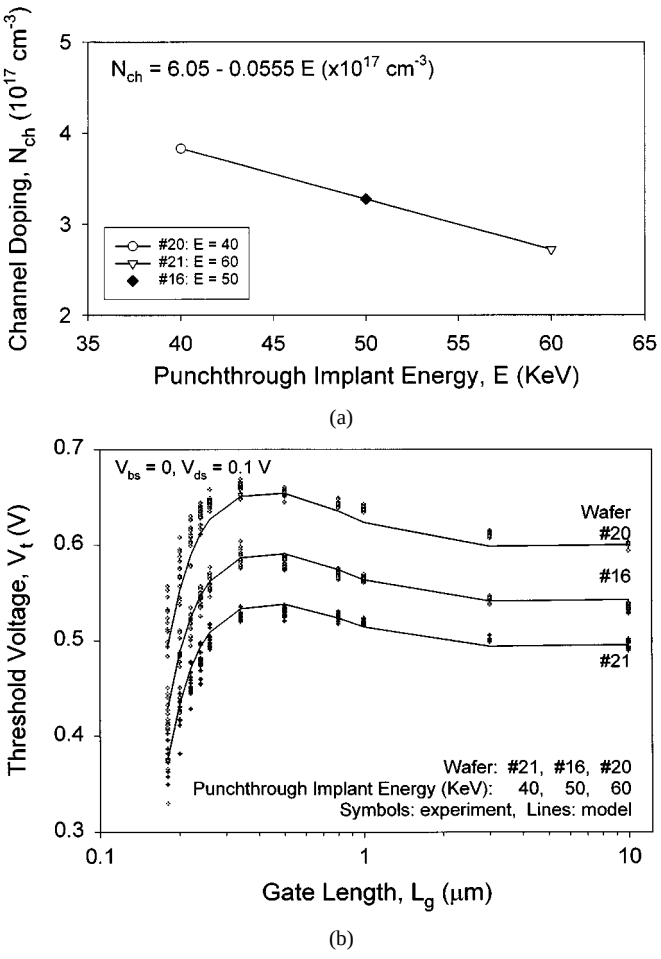


Fig. 5. Experimental correlation of the model parameter N_{ch} to punchthrough implant energy E. The same set of fitting parameters from wafer #15 data is used. The empirical correlation $N_{ch} - E$ in (a) is found from long-channel data of wafers #20 and #21. Prediction of wafer #16 in (b) is made by the compact model with E as input.

Step 2) Nonlinear regression on (8), fitting the V_t (low V_{ds}, high V_{bs}) versus L_g data to extract λ, α', β', and κ'.

Step 3) Nonlinear regression on (8) with fixed values of (λ, α', β') from Step 2, fitting the V_t (low V_{ds}, low V_{bs}) versus L_g data to extract a second value of κ''. With κ' extracted from Step 2, κ_0 and κ_1 can be calculated from (6a).

Step 4) Nonlinear regression on (8) with fixed κ from Step 3, fitting the V_t (high V_{ds}, low V_{bs}) versus L_g data to extract a second set of α'' and β''. With the two sets of (α', α'') and (β', β''), α and β can be calculated from (6b) and (6c), respectively.

When the linear V_{t0} at zero body bias is to be modeled, only a two-step extraction is needed: Step 1 for N_{ch} and Step 3 for (λ, α, β, κ).

As mentioned in the Introduction, the essence of the proposed compact-modeling approach is to predict physical behavior by interpolation. This means that one needs to design the experiment to cover a wide range of interest (e.g., full range of L_g, V_{bs}, and V_{ds}) such that process-dependent empirical parameters can be extracted at the "extreme" conditions. Thereby, the model can be employed for prediction by interpolation at intermediate values of the variables.

IV. Experimental Correlation

Another unique concept and application of the proposed compact-modeling approach, which is different from the BSIM modeling, is to correlate the compact-model parameters to process variables such that the compact model can be used by technology developers to efficiently evaluate process windows or process fluctuations on device performance. This idea is demonstrated with the following examples.

Fig. 4(b) (symbols) shows the measured V_{t0}–L_g data from the same wafer split with only V_t-adjustment implant dose (Φ) changed. One set of V_t data (wafer #15) is used to extract all the necessary model parameters (N_{ch}, λ, α, β, κ). With another two measurements (wafers #17 and #19) for the long-channel $V_{t0}(V_{bs})$, the respective N_{ch} can be extracted. Since there is a one-to-one correlation of the implant dose Φ and the parameter N_{ch}, an empirical equation can be found

$$N_{ch} = 1.426 \times 10^{17} + 7.231 \times 10^{4} \Phi \ (\text{cm}^{-3}) \qquad (9a)$$

where Φ is in cm^{-2}, which turns out to be linear, as shown in Fig. 4(a). Then, a full compact V_t equation with L_g and Φ as input is obtained. Excellent match of the model [lines in Fig. 4(b), wafers #17 and #19] to the experimental data is observed. Furthermore, the V_t behavior of a new wafer (#18) with a different implant dose can be predicted (by interpolation) by the compact model, as demonstrated.

This correlation method has been verified with another set of experimental data in which only punchthrough implant energy (E) has been varied. Wafer #16 has exactly the same process conditions as wafer #15, whereas wafers #20 and #21 are different only in the punchthrough implant energy. By correlating N_{ch} to the measured long-channel $V_{t0}(V_{bs})$ for wafers #20 and #21, the following linear relationship (since only two points) is obtained: R_p [as in (2)], and correlate them to implant dose and energy. Similar correlation can be formulated, e.g., by modifying the model to include an empirical fitting parameter for

$$N_{ch} = 6.05 - 0.0555E \ (\times 10^{17}\ \text{cm}^{-3}) \tag{9b}$$

where E is in KeV, as shown in Fig. 5(a). Then, the compact V_t model (with L_g and E as input), using the same set of parameters (N_{ch}, λ, α, β, κ) extracted from wafer #15, can predict wafer #16 quite well, as shown in Fig. 5(b).

If SIMS data on the channel profile of the final device is available, the compact V_t model should be able to incorporate two variables, peak doping N_p and peak location t_{ox} and correlate it to the gate oxidation temperature or time. Since this approach relates the behavior of the final experimental device to the macromodel, it can be very useful and efficient in reducing wafer splits if they are carefully designed.

V. Results and Discussion

Results of the compact model $V_t = f(L_g, V_{bs}, V_{ds})$ are shown in Fig. 6 (lines) and compared with the experimental data (symbols), from which the model parameters are first extracted as outlined in Section III. Fig. 6(c) shows the model result at an intermediate $V_{ds} = 1$ V at which model prediction is made by "interpolation." The excellent match over a wide range of gate length and bias conditions demonstrates the validity and efficiency of the developed model. Similar results are also obtained for the simulated TCAD data (only the "interpolated" results at $V_{ds} = 1$ V is shown in Fig. 7), in which the $V_t = f(L_g, V_{bs}, V_{ds})$ data has a little bit different behavior from the experimental data. This shows that as long as the data is obtained consistently from the same set of devices, following the parameter-extraction procedure, our compact V_t equation can model the behavior of the real or virtual devices quite well. The model presented here can be further improved by more accurate modeling of the surface potential for a better characterization of the DIBL effect [11].

The approach demonstrated for the compact-model formulation can be generalized to other device-performance figures of merit, such as "on" and "off" state currents, I_{on} and I_{off}, which would be very useful in device design and optimization. Once a good V_t model is available, developing a saturation-current model is mainly a matter of mobility (μ) and series resistance (R_{sd}) modeling. The approach, again, is to formulate an I_{on} compact equation as a function of L_g, V_{bs}, and V_{ds}, with extraction of empirical fitting parameters for process-dependent mobility and resistance. One such preliminary result [11] (based on [12]) is shown in Fig. 8, in which one set of the measured

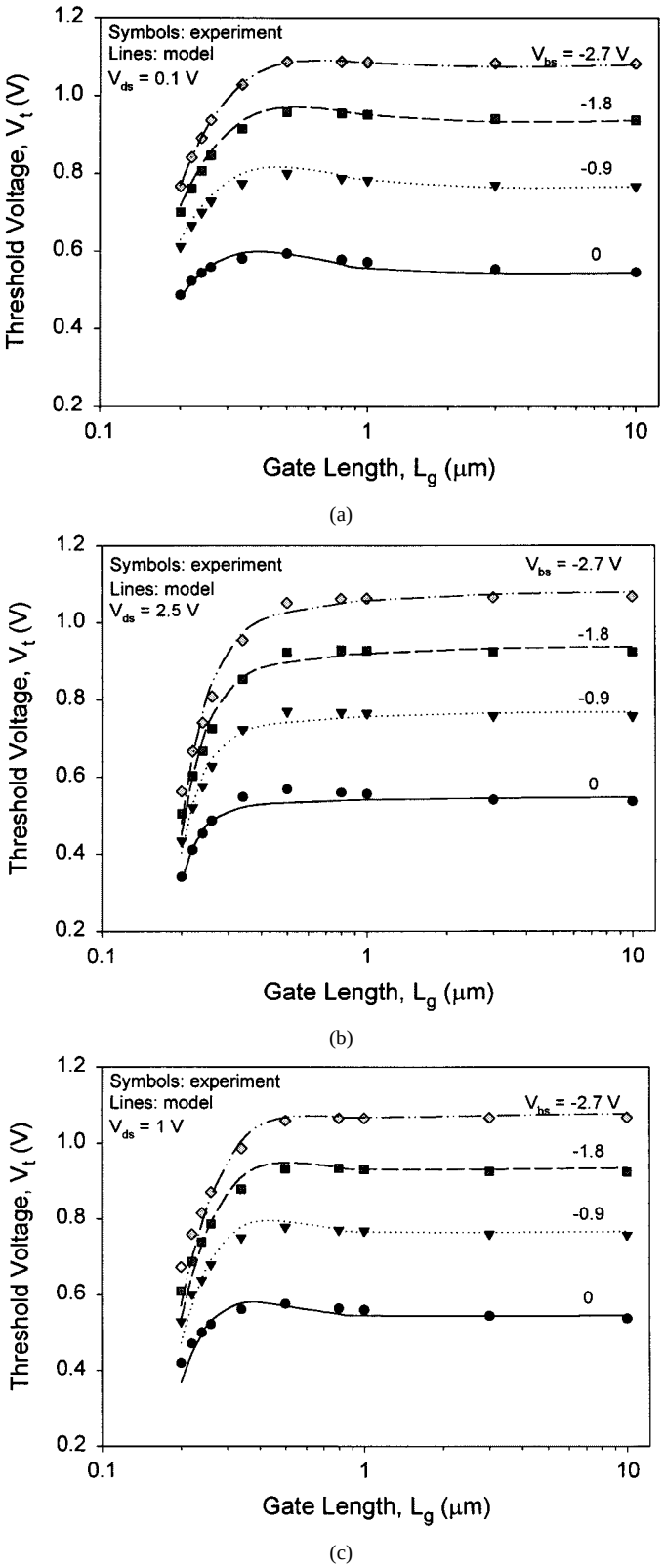


Fig. 6. Results of the compact V_t model (lines) compared with the experimental data (symbols), with full range of gate length, body, and drain biases.

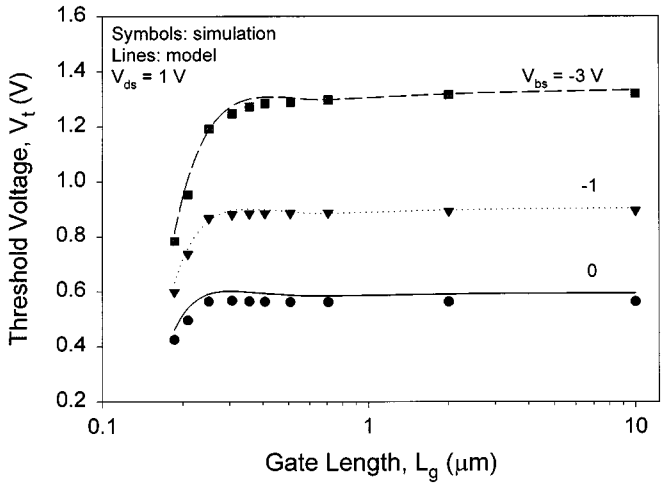


Fig. 7. Prediction of the compact V_t model (lines) on the simulated data (symbols) at an intermediate drain bias. The fitting parameters are extracted from the simulated data similar to Fig. 6(a) and (b) (not shown).

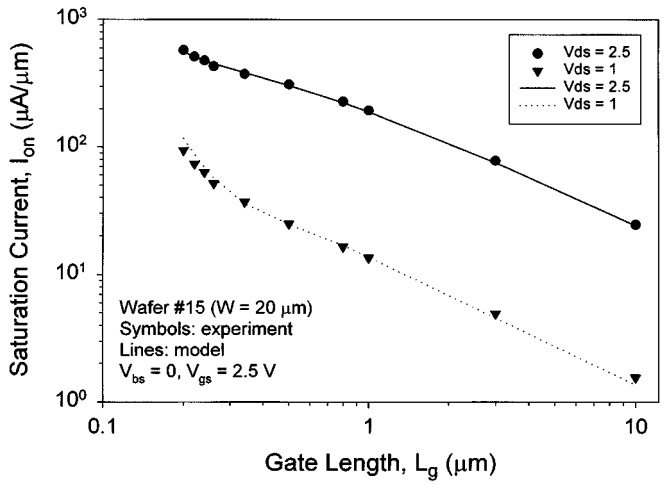


Fig. 8. Preliminary results of the saturation-current model using the developed V_t model. The measured $I_{\rm on}$ at $V_{gs} = V_{ds} = 2.5$ V (circles) is used to extract the parameters for mobility and series resistance. The dotted line is predicted by the model at $V_{gs} = V_{ds} = 1$ V compared to the experimental data (triangles).

$I_{\rm on}$–L_g data at $V_{gs} = V_{ds} = 2.5$ V is used to extract the model parameters, and the model is verified with the prediction at $V_{gs} = V_{ds} = 1$ V. If a compact $I_{\rm off}$ model is developed, $I_{\rm on}/I_{\rm off}$ tradeoff can be made easily, from which design optimization and process windows can be obtained if correlation to process variables can be found.

There are certain limitations, though, in the proposed compact-modeling approach. The model requires fairly accurate values of the device physical parameters, such as L_g, x_j, t_{ox}, ϕ_m, N_{ss}, N_{sd}, etc.; otherwise, the model may be fitted to the "wrong" device during parameter extraction for N_{ch}, σ, λ, α, β, and κ. However, this may not be a fatal problem, and it may turn out to be an advantage since the uncertainties in $(\Delta L_g, \Delta x_j; \Delta t_{ox}, \Delta\phi_m, \Delta N_{ss}; \Delta N_{sd})$, which are unavoidable in deep-submicron devices, may be "absorbed" in the fitting parameters (σ; N_{ch}; α). In this work, with the measured t_{ox} and x_j, we have used nonlinear regression on (8) fitting the long-channel $V_t(V_{bs})$ data to obtain N_{ch} as well as ϕ_m. Optimum value of σ is also predetermined by best fit of the short-channel V_t in the roll-off region. Another disadvantage compared to the TCAD approach is in capturing the details of process and device variables. This is obvious because compact models are at a higher level of abstraction. There is a tradeoff between the effects to be built into the model and the complexity of the model. However, combined with a full-loop calibrated TCAD, the demonstrated compact-modeling approach (together with experimental correlation) will prove to be extremely powerful in providing an efficient and accurate guide to technology developers.

VI. Conclusion

A general approach to formulating physics-based compact model equations for deep-submicron technology development has been described and demonstrated with the threshold voltage model, which can predict the behavior of devices with gate lengths down to 0.18 μm and full range of body- and drain-bias conditions. There are three basic ingredients in the approach: 1) multi-level modeling, which combines low-level accuracy and high-level efficiency; 2) empirical formulation, which combines analytic physics and realistic behavior, and 3) experimental correlation, which combines theoretical definitions of macromodel parameters and practical measurable variables. The developed V_t model takes into account nonuniform channel profile, charge-sharing, and barrier-lowering effects, normal and reverse SCE's with combined body- and drain-bias dependencies. Such a unified compact model will not be feasible with rigorous analytical derivations, nor will it be a simple task to get the remarkably accurate prediction on real device performance by TCAD-calibration approach. Combined with the efforts in new process/device model development and TCAD calibration, the proposed approach will be a first step toward the implementation of a multi-level TCAD-synthesis methodology in aiding new technology development and deep-submicron transistor design.

References

[1] B. Yu, C. H. J. Wann, E. D. Nowak, K. Noda, and C. Hu, "Short-channel effect improved by lateral channel-engineering in deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 44, pp. 627–633, Apr. 1997.

[2] *BSIM3v3 Manual*, Univ. of California, Berkeley, 1996.

[3] *MEDICI (ver. 4.0)*, Technol. Model. Assoc., Inc., Palo Alto, CA, 1997.

[4] H. I. Hanafi, W. P. Noble, R. S. Bass, K. Varahramyan, Y. Lii, and A. J. Dally, "A model for anomalous short-channel behavior in submicron MOSFET's," *IEEE Electron Device Lett.*, vol. 14, pp. 575–577, Dec. 1993.

[5] X. Zhou, K. Y. Lim, and D. Lim, "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.

[6] Z. X. Yan and M. J. Deen, "Physically-based method for measuring the threshold voltage of MOSFET's," *Proc. Inst. Elect. Eng.*, vol. 138–G, pp. 351–357, Mar. 1991.

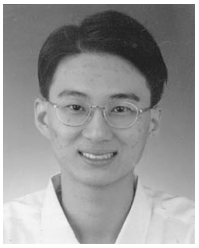
[7] K. Y. Lim and X. Zhou, "Modeling of threshold voltage with non-uniform substrate doping," in *Proc. 1998 IEEE Int. Conf. Semicond. Electron. (ICSE'98)*, Malaysia, Nov. 1998, pp. 27–31.

[8] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*. New York: McGraw-Hill, 1987, p. 186.

[9] 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.

[10] 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.
[11] K. Y. Lim, X. Zhou, and D. Lim, "A predictive length-dependent saturation current model based on accurate threshold voltage modeling," in *Proc. 2nd Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators (MSM99)*, Puerto Rico, Apr. 1999, pp. 423–426.
[12] K. Chen, H. C. Wann, J. Duster, D. Pramanik, S. Nariani, P. K. Ko, and C. Hu, "An accurate semi-empirical saturation drain current model for LDD N-MOSFET," *IEEE Electron Device Lett.*, vol. 17, pp. 145–147, Mar. 1996.

Xing Zhou (S'88–M'91–SM'99), for a photograph and biography, see this issue, p. 120.



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.

His present research interests focus on physical modeling, simulation and characterization of deep sub-micrometer MOSFET as well as the improved structures for transistor scaling.

David Lim, photograph and biography not available at the time of publication.