FinFET Design Considerations Based on 3-D Simulation and Analytical Modeling

Gen Pei, *Student Member, IEEE*, Jakub Kedzierski, *Member, IEEE*, Phil Oldiges, *Member, IEEE*, Meikei Ieong, *Senior Member, IEEE*, and Edwin Chih-Chuan Kan, *Member, IEEE*

***Abstract*—Design considerations of FinFET have been investigated by three-dimensional (3-D) simulation and analytical modeling in this paper. Short-channel effects (SCE) of FinFET can be reasonably controlled by reducing either silicon fin height or fin thickness. Analytical solution of 3-D Laplace's equation is employed to establish the design equations for the subthreshold behavior in the fully depleted silicon fins. Based on the 3-D analytical electrostatic potential in the subthreshold region, the threshold voltage (V_{th}) roll-off and the subthreshold swing (S) are estimated by considering the source barrier changes in the most leaky channel path. V_{th} roll-off is an exponential function of the ratio of effective channel length to drain potential decay length, which can then be expressed as a function of the fin thickness, the fin height and the gate oxide thickness. The drain-potential decay lengths of single-gate fully depleted SOI MOSFET (FDFET), double-gate MOSFET (DGFET), rectangular surrounding-gate MOSFET (SGFET), and FinFET are compared. The drain potential scaling length and V_{th} roll-off can be included into a universal relation for convenient comparison.**

***Index Terms*—Double-gate MOSFET, FinFET, short-channel effects, silicon-on-insulator (SOI).**

I. Introduction

THE relentless down scaling of CMOS device dimensions has been the major driving force of the growth in microelectronics during the past three decades. For lithography nodes from 20 μm to 0.1 μm, MOSFET gate oxide thickness has been aggressively scaled together with complex channel doping design to suppress short channel effects (SCE). According to the projection of the 1999 *International Technology Roadmap for Semiconductors*, devices with gate length down to 20 nm can be expected in 2014 [1].

CMOS designs below 0.1 μm are severely constrained by lateral SCE and vertical gate insulator tunneling [2]–[5]. One of the approaches to circumvent the gate tunneling restriction is to change the device structure in such a way that MOSFET gate length can be scaled further even with thicker oxide. Double-gate MOSFET (DGFET) has been considered as one of the most promising candidates for the channel length in the range of 10–30 nm [6]–[9]. The alignment of the top and the bottom gates to each other and to source/drain (S/D) doping is crucial to device performance, because misalignment may cause extra gate-to-S/D overlap capacitance as well as S/D series resistance [10]. In order to optimize the performance of double-gate devices, self-aligned processes and structures are proposed, with FinFET being one of the most promising [11]–[16].

FinFET is developed with special emphasis on process simplicity and compatibility with conventional planar CMOS technology [17]. Devices with gate length of 18 nm and gate oxide thickness of 2.5 nm have been experimentally demonstrated with acceptable short-channel characteristics [15]. Based on the subthreshold behavior of FinFET, an empirical scaling rule has been proposed. In order to suppress SCE, the fin thickness has to be less than one third of the channel length [15]. A simple capacitive model was proposed by [16] to predict the relationship between DIBL and S swing. Most of the results reported on FinFET are based on experiments or 2-D simulation [14]. Due to the asymmetrical three-dimensional (3-D) structure of FinFET, the coupling between gates that surround the channel cannot be neglected in aggressively scaled device dimensions. Hence, 3-D analyses are necessary to obtain reasonable prediction of device performance and structure optimization.

A basic structure of FinFET is shown in Fig. 1. The critical geometrical parameters of FinFET are defined below. Also shown together is the definition of these parameters in the MOS control box as used in the setup of analytical and numerical modeling.

- L_{gate}: physical gate length of FinFET defined by the spacer gap;
- H_{fin}: height of silicon fin, defined by the distance between top gate and buried oxides;
- T_{fin}: thickness of silicon fin, defined by the distance between front and back gate oxides;
- L_{eff}: effective channel length of FinFET are estimated by the metallurgical junction for abrupt junctions;
- W: geometrical channel width defined as $W = 2\times H_{fin} + T_{fin}$.

Intuitively, when T_{fin} is much larger than H_{fin} or when top gate oxide is much thinner than the front and back oxides, FinFET can be approximately treated as single-gate fully depleted SOI MOSFET (FDFET) as long as the silicon fin remains fully depleted. On the other hand, when H_{fin} is much larger than T_{fin} or top gate oxide is much thicker than the front and back oxides, FinFET can be approximately treated as DGFET. The two limits of FinFET, FDFET and DGFET, are widely studied and well understood [8], [9], [21], [22], but in

Manuscript received February 4, 2002; revised May 2, 2002. The work was supported in part by NSF Award ID 0120328 and SRC Task ID 935.001 under the Program of Advanced Mixed-Signal Device and Circuit Technologies. The review of this paper was arranged by Editor J. Vasi.

G. Pei and E. C.-C. Kan are with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, 14853 NY.

J. Kedzierski and M. Ieong are with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA.

P. Oldiges is with the IBM Microelectronics Divisions, East Fishkill, NY 12533 USA.

Publisher Item Identifier 10.1109/TED.2002.801263.

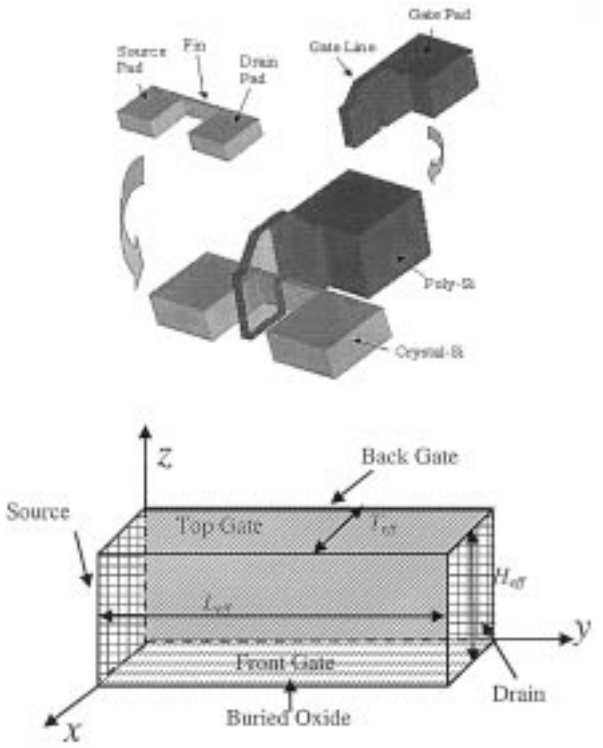

Fig. 1. Illustration of 3-D FinFET structure and the corresponding MOS control box modeling set-up.

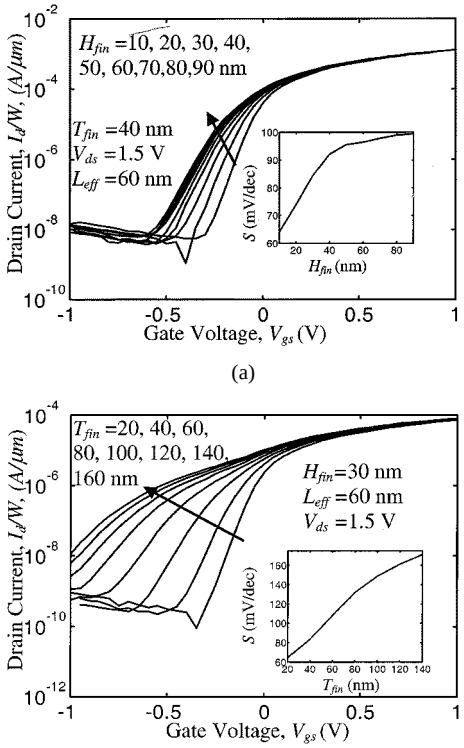

Fig. 2. I_d–V_{gs} characteristics for NMOS FinFET with 60 nm effective channel length from 3-D FIELDAY simulation. H_{fin} and T_{fin} change from less than $L_{eff}/3$ to larger than L_{eff} to show the parameter design space.

the regime where both fin height and fin thickness have control over SCE, the dependence of SCE on device dimensions is not well characterized. In order to establish the functional expressions of SCE, analytical solution of 3-D Laplace's equation is used to derive the design equations for the subthreshold behavior of FinFET. Based on the 3-D electrostatic potential distribution in the subthreshold region, the subthreshold swing S and the threshold voltage V_{th} roll-off are estimated by considering the source barrier changes in the most leaky channel path. The distributive channel doping effects cannot be easily incorporated into this model with tractable complexity [24], [29]. Many experimental DGFETs have undoped SOI body due to the difficult controllability on dopant and SOI body thickness fluctuation [19].

The paper is organized as follows: 3-D simulations results of FinFET are presented and discussed in Section II; 3-D analytical solution is derived and compared with the numerical solution in Section III; the methodology is extended to other variations of fullydepleted devices and design equations are presented in Section IV.

II. Short-Channel Effects by 3-D Distributive Simulation

3-D FinFET with undoped channel, n^+ polysilicon gate, uniform 1.5 nm gate oxide, abrupt source-to-channel and drain-to-channel junctions has been simulated in FIELDAY [18]. The poly gate is specified by work function and the effects of poly depletion are not included. The parasitic resistance and capacitance are treated in simulation by including 10-nm source-to-gate and drain-to-gate overlap. It is difficult to calibrate more accurate parasitic structures without complicating the resulting observation. The transport model is based on the improved drift-diffusion formalism with trajectory-based tunneling enabled [24]. The band-to-band tunneling model follows that in [35]. The mobility model in FIELDAY is the one for deep submicron bulk MOSFETs [34]. The definition of effective field in FinFET is supposed to be different from that in bulk CMOS because of the 3-D structure of FinFET [32]. The validity of using this mobility model calibrated with bulk CMOS for FinFET simulation is questionable. Long-range Coulomb scattering is important when gate oxide thickness is less than 2.5 nm [30], and it is not included in mobility model. Velocity overshoot may be important for the channel length under 60 nm, but only the velocity saturation is included here. Therefore, the above threshold current may not be accurate enough to give a quantitative discussion. Only qualitative discussion remains meaningful. The subthreshold current is supposed to be accurate enough for the study of subthreshold behavior.

The simulated I_d-V_{gs} characteristics are shown in Fig. 2, which compares reasonably well with experimental measurements on limited collections of device designs in consideration of physical measurement errors and fluctuations [14], [15], [31]. Direct experimental comparison will be presented in later figures after device parameter extraction [31].

The drain current normalized by the channel width W at the same V_{gs} is almost independent of H_{fin} while fixing T_{fin}. The small differences in the normalized drain current for devices

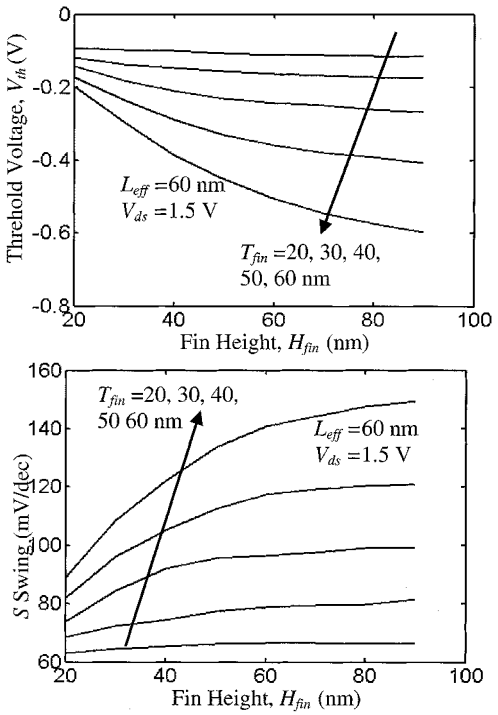

Fig. 3. Dependence of threshold voltage roll-off on H_{fin} and T_{fin} extracted from Fig. 2.

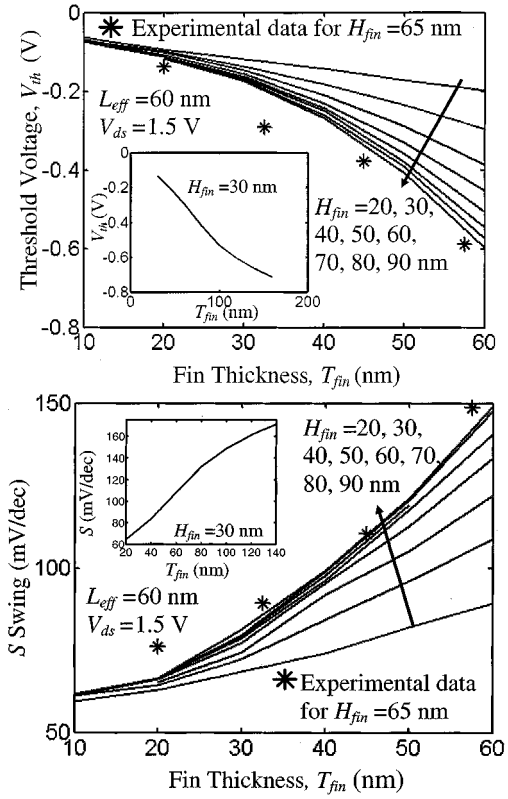

Fig. 4. Dependence of subthreshold swing on H_{fin} and T_{fin} extracted from Fig. 2. Details of experimental data can be found in [31].

with the same H_{fin} and different T_{fin} come from the threshold voltage roll-off due to the increase in T_{fin}. For comparison, if the drain current is normalized by $2H_{fin}$ as in [15], the normalized drain current of FinFET with 10 nm H_{fin} is almost twice higher than that of FinFET with 30 nm H_{fin}. This deviation in [15] is insignificant since the gate oxide on the top is much thicker than that on the side, where FinFET is similar to a self-aligned DGFET in operation. The gate-induced drain leakage (GIDL) is on the same order for all the devices simulated and is independent of H_{fin} and T_{fin}, because it is most sensitive to the drain junction overlap and doping concentration in the overlapped region, both of which are fixed throughout the simulation. To illustrate SCE of FinFET clearly, in Fig. 2(a), T_{fin} is fixed at 40 nm; as H_{fin} is increased from 10 to 90 nm, the subthreshold swing S changes from 74.0 to 99.4 mV/dec and the changes gradually saturate. In Fig. 2(b), H_{fin} is fixed at 30 nm. As T_{fin} is increased from 20 to 160 nm, the subthreshold swing S increases from 64.0 to 184.2 mV/dec in the way that S increases more rapidly first and the increase rate slows down gradually.

The dependences of V_{th} roll-off and subthreshold swing S on H_{fin} and T_{fin} for $V_{ds} = 1.5$ V are further shown in Figs. 3 and 4. Here V_{th} is defined as the gate voltage when $I_d = 300\,\text{nA} \times W/L_{eff}$. In Fig. 3, as H_{fin} is increased from 20 to 90 nm, the saturation of V_{th} roll-off and S is observed. The critical H_{fin} needed for saturation is different for devices with different T_{fin}. For devices with larger T_{fin}, the critical H_{fin} is correspondingly larger. In Fig. 4, V_{th} roll-off and S change more and more rapidly as T_{fin} changing from 10 to 60 nm; as T_{fin} changing from 60 to 140 nm, the change rate of V_{th} roll-off and S increases at first and then slows down. We can expect V_{th} roll-off and S to saturate when T_{fin} goes to even higher value. Experimental data [31] are only available for $H_{fin} = 65$ nm with four T_{fin} variations. Since the simulation predicts saturation of V_{th} and S with increasing H_{fin}, the simulation results actually compare reasonably well with the experimental measurement, given the device parameter fluctuations.

Depending on the relative ratio of T_{fin} to H_{fin}, either T_{fin} or H_{fin} can be the dominant parameter that controls SCE of the device. For a given L_{eff}, changing the dominant parameter will lead to big changes in V_{th} roll-off and S, while changing the nondominant parameter will lead to small changes or almost no changes in V_{th} roll-off and S. In the regime where both H_{fin} and T_{fin} have noticeable control over SCE, the same changes in T_{fin} and H_{fin} lead to different changes in V_{th} roll-off and S. Intuitively, this is because the channel is controlled by two gates on the sides while only one gate on the top and no gate at the bottom.

It is also very important to study the device characteristics in the weak to strong inversion region for analog circuit applications [25]. A representative figure of merit of the device is g_m/I_d versus $\log(I_d)$ as shown in Fig. 5. The bump in the curve is due to corner effects (Fig. 6). Because of the high electric field in the corner, the corner transistors turn on earlier than the main transistor. The width of the corner transistor is smaller than that of main transistor and its current contribution is only important before the main transistor turns on. In the above threshold regime, parasitic resistance is an important factor and decreases g_m/I_d because of the small fin size.

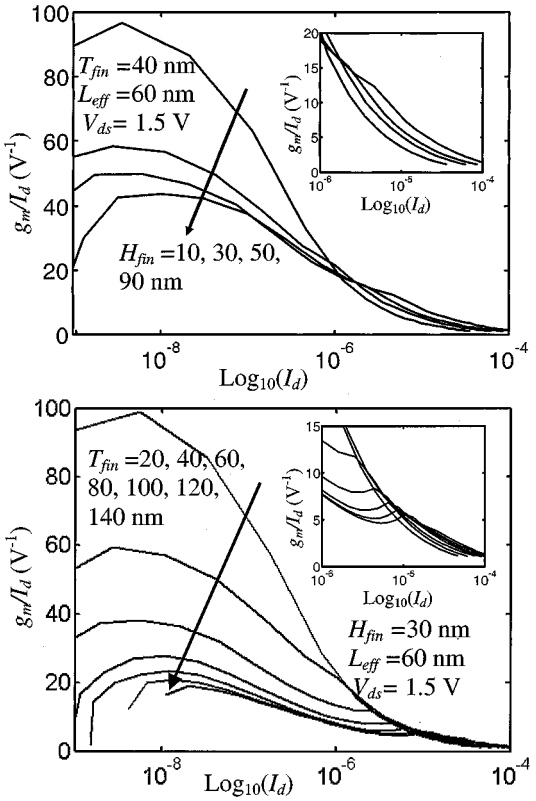

Fig. 5. g_m/I_d as a function of $\log(I_d)$ from deep subthreshold to weak inversion to strong inversion. Effects from corner transistors can be observed.

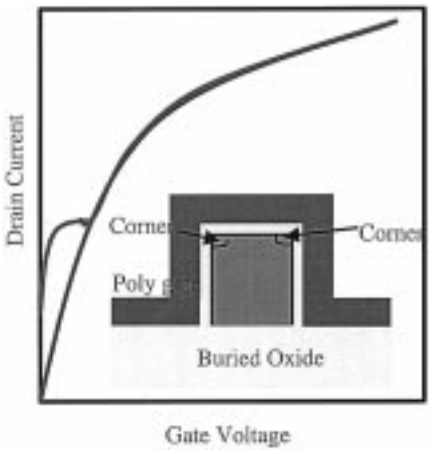

Fig. 6. Illustration of corner effects and its influences on the I_d-V_{gs} characteristics of devices.

III. Short-Channel Effects by 3-D Analytical Modeling

Subthreshold conduction is governed by the potential distribution in the entire SOI body. The parabolic potential assumption is widely used in DGFET modeling [8], [9], [20], [21]. It is hard to assume a simple potential distribution in FinFET because of the asymmetric 3-D structure. Analyses based on evanescent nodes are more applicable [26]. The electrostatic potential in the subthreshold region can be described by 3-D Laplace's equation if the potential perturbation due to the carrier and doping concentrations can be ignored, which is the case for undoped fin.

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=0. \tag{1}$$

Approximations have to be made to reduce the formulation to a manageable level. The six boundary conditions are set by the top gate, front gate, back gate, source, and drain and buried oxide (Fig. 1). The buried oxide is assumed to be thick enough that any finite potential across the buried oxide leads to a negligible electric field. The boundaries between gate oxide and silicon fin are eliminated by replacing the physical dimensions with effective dimensions. The whole region is treated as homogeneous silicon with effective thickness (T_{eff}), effective channel length (L_{eff}) and effective height (H_{eff}) [22]. Two ways to define H_{eff} and T_{eff} are examined here. The first definition in (2) is to replace the oxide with an equivalent region with thickness of the oxide region modified by $\epsilon_{si}/\epsilon_{ox}$[22]. The simplification is based on the fact that the normal component of the electric field changes by a factor of $\epsilon_{si}/\epsilon_{ox}$ across the silicon-oxide boundary. The drawback is that it may cause errors in the tangential field whose magnitude does not change across the silicon-oxide boundary. The errors are expected to be small when the gate oxide field is dominated by its normal component. However, for very short-channel device when the lateral electric field becomes comparable to the normal electric field, the errors cannot be neglected.

Another definition of T_{eff} and H_{eff} in (3) is inspired from the scaling length of DGFETs with the parabolic potential approximation [9], [27], [28]. The scaling length of regular DGFETs is $\lambda \propto \sqrt{T_{fin}\times(T_{fin}+(4\varepsilon_{si}/\varepsilon_{ox})T_{ox})}$, which is the geometrical average of $(4\varepsilon_{si}/\varepsilon_{ox})T_{ox}$ and T_{fin}, while in the first definition the algebraic average is used. The simplification of taking T_{eff} and H_{eff} by the geometrical average is based on the reason that when one of them is much larger than the other, FinFET can be reduced to a compatible DGFET or FDFET.

$$\begin{aligned} T_{eff} &= T_{fin}+\frac{2\varepsilon_{si}}{\varepsilon_{ox}}T_{ox} \\ H_{eff} &= H_{fin}+\frac{\varepsilon_{si}}{\varepsilon_{ox}}T_{ox} \end{aligned} \tag{2}$$

$$\begin{aligned} T_{eff} &= \sqrt{T_{fin}\times\left(T_{fin}+\frac{4\varepsilon_{si}}{\varepsilon_{ox}}T_{ox}\right)} \\ H_{eff} &= \sqrt{H_{fin}\times\left(H_{fin}+\frac{2\varepsilon_{si}}{\varepsilon_{ox}}T_{ox}\right)}. \end{aligned} \tag{3}$$

With all the assumptions and approximations above, the boundary conditions are defined as:

Top gate:

$$\psi\,|_{z=H_{eff}} = V_g - V_{fb}. \tag{4}$$

Front gate:

$$\psi\,|_{x=T_{eff}} = V_g - V_{fb}. \tag{5}$$

Back gate:

$$\psi\,|_{x=0} = V_g - V_{fb}. \tag{6}$$

Source:

$$\psi\,|_{y=0} = -\phi_{ms}. \tag{7}$$

Drain:

$$\psi\,|_{y=L_{eff}} = -\phi_{ms} + V_{ds}. \tag{8}$$

Buried oxide:

$$\frac{\partial \psi}{\partial z}\,|_{z=0} = 0 \tag{9}$$

where ϕ_{ms} is the work function difference of S/D to the channel. For undoped fin body and n^+S/D, ϕ_{ms} is about half of the Si bandgap. Superposition principle can be applied to solve the 3-D linear Laplace's equation and the solutions are as follows. See (10)–(19), shown at the bottom of the page.

We have validated the previous equations by observing device symmetry. The symmetry of FinFET is reflected in the height direction similar to the rectangular surrounding gate MOSFET. The only change is the bottom buried-oxide boundary condition is replaced with bottom-gate boundary condition in (20). Due to the symmetrical structure in the z direction, the buried oxide boundary condition is implicit in the model and the same solutions can be reached.

Bottom gate

$$\psi\,|_{z=-H_{eff}} = V_g - V_{fb}. \tag{20}$$

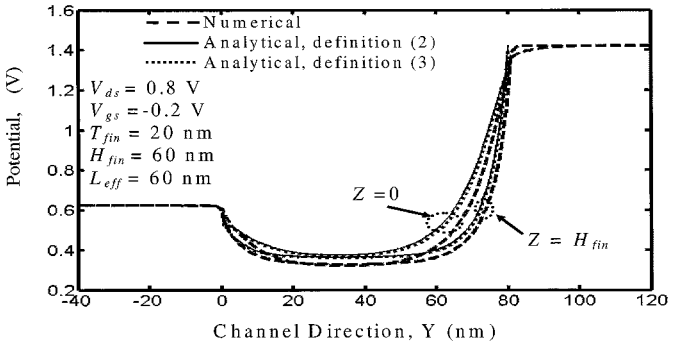

Fig. 7. Comparison of potential distribution at $x = T_{eff}/2$, $z = 0$ and $x = T_{eff}/2$, $z = H_{fin}$ from numerical and analytical solutions around the threshold voltage at high V_{ds}.

For further verification, the analytical solutions of the potential in the two cut lines of $z = 0$, $x = T_{eff}/2$ and $z = H_{fin}$, $x = T_{eff}/2$ are compared with numerical simulation in Fig. 7. The first cut line is the channel closest to top gate and the other is the channel furthest from top gate. The simulation and analytical solution have shown reasonable match and the potential distribution based on two different definitions of T_{eff} and H_{eff} is very close for the device of interest here. The source barrier of the current channel at $z = 0$ is pulled down more by the drain potential and hence more sensitive to device scaling. The potential difference between these two z cut lines is small because

$$\psi = \psi_{tg} + \psi_{fg} + \psi_{bg} + \psi_s + \psi_d + \psi_{BO} \tag{10}$$

$$\psi_{tg} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{16(V_g - V_{fb})}{(2n-1)(2m-1)\pi^2} \sin\left(\frac{(2m-1)\pi}{L_{eff}}y\right) \sin\left(\frac{(2n-1)\pi}{T_{eff}}x\right) \frac{\cos h\left(\frac{k_z\pi}{H_{eff}}z\right)}{\cos h(k_z\pi)} \tag{11}$$

$$\psi_{fg} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{16(V_g - V_{fb})(-1)^{n+1}}{(2n-1)(2m-1)\pi^2} \sin\left(\frac{(2m-1)\pi}{L_{eff}}y\right) \cos\left(\frac{(2n-1)\pi}{2H_{eff}}z\right) \times \frac{\sin h\left(\frac{k_x\pi}{T_{eff}}x\right)}{\sin h(k_x\pi)} \tag{12}$$

$$\psi_{bg} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{16(V_g - V_{fb})(-1)^{n+1}}{(2n-1)(2m-1)\pi^2} \sin\left(\frac{(2m-1)\pi}{L_{eff}}y\right) \cos\left(\frac{(2n-1)\pi}{2H_{eff}}z\right) \times \frac{\sin h\left(\frac{k_x\pi}{T_{eff}}(T_{eff}-x)\right)}{\sin h(k_x\pi)} \tag{13}$$

$$\psi_s = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{-16\phi_{ms}(-1)^{n+1}}{(2n-1)(2m-1)\pi^2} \sin\left(\frac{(2m-1)\pi}{T_{eff}}x\right) \cos\left(\frac{(2n-1)\pi}{2H_{eff}}z\right) \times \frac{\sin h\left(\frac{k_y\pi}{L_{eff}}(L_{eff}-y)\right)}{\sin h(k_y\pi)} \tag{14}$$

$$\psi_d = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{-16(\phi_{ms} - V_{ds})(-1)^{n+1}}{(2n-1)(2m-1)\pi^2} \sin\left(\frac{(2m-1)\pi}{T_{eff}}x\right) \cos\left(\frac{(2n-1)\pi}{2H_{eff}}z\right) \times \frac{\sin h\left(\frac{k_y\pi}{L_{eff}}y\right)}{\sin h(k_y\pi)} \tag{15}$$

$$\psi_{BO} = 0 \tag{16}$$

$$k_x = \sqrt{\left(\frac{2m-1}{L_{eff}}\right)^2 + \left(\frac{2n-1}{2H_{eff}}\right)^2} \times T_{eff} \tag{17}$$

$$k_y = \sqrt{\left(\frac{2m-1}{T_{eff}}\right)^2 + \left(\frac{2n-1}{2H_{eff}}\right)^2} \times L_{eff} \tag{18}$$

$$k_z = \sqrt{\left(\frac{2n-1}{T_{eff}}\right)^2 + \left(\frac{2m-1}{L_{eff}}\right)^2} \times H_{eff}. \tag{19}$$

this is a relatively long channel device with $T_{fin} = 20$ nm. Although the two cut lines are different in distance from the top gate, but they have the same distance to the side gates, which is around 10 nm and is much smaller than H_{fin} at 60 nm. Hence, the sides have a good control over the channel at $z = 0$ and overwhelm the influence of the top gate. It can be predicted that larger barrier difference will occur if T_{fin} is increased or shorter channel device is considered [8].

Although the 3-D analytical solution has clearly demonstrated the functional dependence of subthreshold characteristics on geometrical features, it needs further reduction to obtain design equations for subthreshold swing S and V_{th} roll-off. It will be further assumed that the leakage current is dominated by current in the channel with the minimum source barrier, the most leaky path. The assumption is based on the exponential dependence of the drain current on source barrier in subthreshold region. From the previous distributed device simulation, it can be seen that the source barrier difference between the two channels is very small for -0.2 V gate bias. The simplification is inevitable, since if the whole fin is used for integration to obtain the emission over the source barrier, the analytical form becomes intractable. Following the assumptions above, the subthreshold swing S can be derived. The general formula of S for the most leaky path at (x_c, z_c) is given in (21)

$$\begin{aligned}\frac{1}{S} =& \frac{\partial \log(I_d)}{\partial V_g} = \frac{q}{2.3KT}\frac{\partial \psi_s}{\partial V_g} \\ =& \frac{q}{2.3kT}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{16}{(2n-1)(2m-1)\pi^2}\sin\left(\frac{(2m-1)\pi}{L_{eff}}y_c\right) \\ & \times\left\{\sin\left(\frac{(2n-1)\pi}{T_{eff}}x_c\right)\frac{\cos h(\frac{k_z\pi}{H_{eff}}z_c)}{\cos h(k_z\pi)}\right. \\ & +\cos\left(\frac{(2n-1)\pi}{2H_{eff}}z_c\right)(-1)^{n+1} \\ & \left.\times\frac{\sinh(\frac{k_x\pi}{T_{eff}}(T_{eff}-x_c))+\sinh(\frac{k_x\pi}{T_{eff}}x_c)}{\sinh(k_x\pi)}\right\} \end{aligned} \tag{21}$$

$$y_c = \frac{L_{eff}}{2} + \frac{L_d}{2\pi}\ln\left(\frac{-\phi_{ms}}{-\phi_{ms}+V_{ds}}\right) \tag{22}$$

$$L_d = \frac{1}{\sqrt{\left(\frac{1}{T_{eff}}\right)^2+\left(\frac{0.5}{H_{eff}}\right)^2}} \tag{23}$$

k_x and k_z are defined in (17) and (18) respectively. L_d is the drain potential decay length. The position of the most leaky path (x_c, z_c) can be found algebraically. Since the source barriers of all channel paths are referred to the same source, finding the minimum source barrier is equivalent to finding the maximum potential at the source barrier plane. If the most leaky path is not at boundaries but in the interior, its position (x_c, z_c) can be found by solving two coupled equations: $\partial\psi(x, y_c, z)/\partial x = 0$ and $\partial\psi(x, y_c, z)/\partial z = 0$. And the second derivatives have to be checked to make sure the potential at (x_c, y_c, z_c) is maximum instead of minimum on the source barrier plane, $y = y_c$, where $\partial\psi^2(x_c, y_c, z_c)/\partial x^2 < 0$, $\partial\psi^2(x_c, y_c, z_c)/\partial x\partial z < 0$ and

$$\frac{\partial\psi^2(x_c,y_c,z_c)}{\partial x^2}\times\frac{\partial\psi^2(x_c,y_c,z_c)}{\partial z^2} - \frac{\partial\psi^2(x_c,y_c,z_c)}{\partial x\partial z} > 0.$$

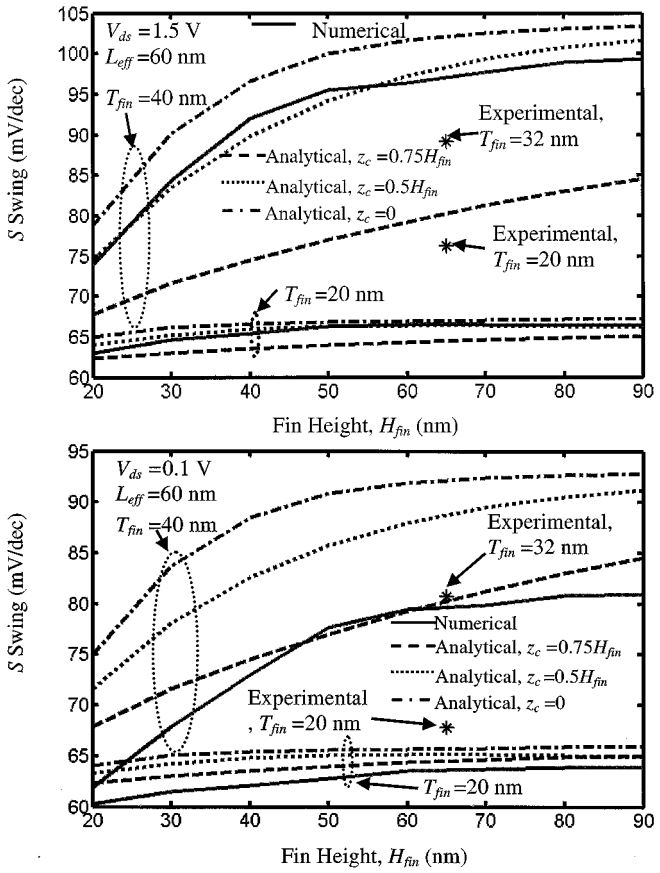

Fig. 8. Comparison of subthreshold swing from numerical and analytical solutions for both linear ($V_{ds} = 0.1$ V) and saturation ($V_{ds} = 1.5$ V) regions. Experimental measurements are from [31]. T_{eff} and H_{eff} are defined by (3).

However, the equation will be become unmanageable if the position of most leaky path is found this way.

The position of the most leaky path is highly dependent on the channel doping, gate bias and device geometry. For the case of undoped channel and deep subthreshold operation, the most important variable is the device geometry. The closer the channel path to the gate, the more source barrier is reduced by the gate, while the further the channel away from the gate, the more source barrier is reduced by the source/drain potential. If there is no potential penetration from the source and drain, then the most leaky path will be on the interface of silicon and gate oxide without considering quantum correction. If there is potential penetration from the source and drain, the most leaky path may move away from the interface into silicon fin. If the source/drain potential penetration is the dominant factor in determining the source barrier, the most leaky path will be the channel path furthest away from the gate in distance. The source/drain potential here include the build-in potential, ϕ_{ms} and applied drain bias, V_{ds}.

In the case of no source/drain potential penetration, the most leaky path can be illustrated by making a cutline at $y = y_c$ and $z = 0$. For simplicity, $H_{fin} \gg T_{fin}$ is assumed and the potential penetration from top gate to the cutline is also negligible. If the partial derivative is taken in the x direction, $\partial\psi(x, y_c, 0)/\partial x = 0$ gives $x = T_{eff}/2$. However, $\partial\psi^2(T_{eff}/2, y_c, 0)/\partial x^2 > 0$, which means the channel path $x = T_{eff}/2$ and $z = 0$ has the maximum source barrier. The channel path with the minimum barrier then has to happen

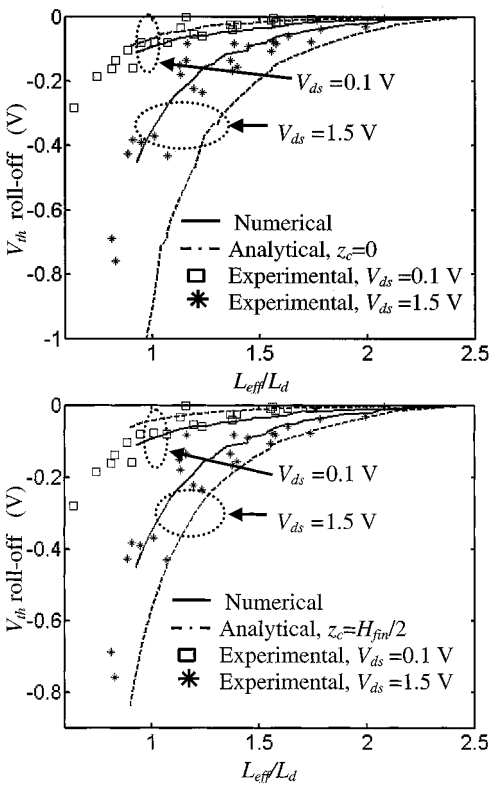

Fig. 9. Comparison of threshold voltage roll-off from numerical and analytical solutions for both linear ($V_{ds} = 0.1$ V) and saturation ($V_{ds} = 1.5$ V) regions. Experimental measurements are from [31]. T_{eff} and H_{eff} are defined by (3).

on the boundary, which means the most leaky path is on gate oxide-silicon interface.

The subthreshold slope S extracted from simulation data is compared with that from the analytical solution in Fig. 8. Three different choices of most leaky path are presented: $x_c = 0.5T_{eff}$, $z_c = 0$; $x_c = 0.5T_{eff}$, $z_c = 0.5H_{fin}$; $x_c = 0.5T_{eff}$, $z_c = 0.75H_{fin}$. The analytical solution with $x_c = 0.5T_{eff}$ and $z_c = 0.5H_{fin}$ matches the simulation results best at high V_{ds}, while at low V_{ds} channel at $x_c = 0.5T_{eff}$ and $z_c = 0.75H_{fin}$ matches the simulation results best. Because of less DIBL at low V_{ds}, the most leaky path moves closer to the top gate. The build-in source/channel and drain/channel potential ϕ_{ms} contributes to potential penetration at low V_{ds}. Available experimental points [31] are also included in Fig. 8 for $H_{fin} = 65$ nm and $T_{fin} = 20$ and 32 nm. This is mainly for validation purposes due to the scattered nature of the experimental parameter extraction, which is more evident in Fig. 9.

V_{th} roll-off caused by DIBL is estimated by the necessary gate voltage change to compensate the DIBL in the most leaky path (x_c, z_c), (24). In the derivation, only the first term of the series summation is taken because of the fast decay of the magnitude of higher order summation terms

$$\begin{aligned}\Delta V_{th} =& \frac{\partial V_{gs}}{\partial \psi_c}\Delta\psi_c(V_{ds}) = \frac{Sq}{2.3KT}\Delta\psi_c(V_{ds}) \\ =& -\frac{Sq}{2.3kT}\frac{32}{\pi^2}\sin\left(\frac{\pi}{T_{eff}}x_c\right)\cos\left(\frac{\pi}{2H_{eff}}z_c\right) \\ & \times \exp\left(-\frac{L_{eff}}{2L_d}\pi\right)\left[\sqrt{\phi_{ms}(\phi_{ms}-V_{ds})} - |\phi_{ms}|\right].\end{aligned} \tag{24}$$

V_{gs} is the gate bias and ψ_c is the source potential barrier in the most leaky path. The formula shows that V_{th} roll-off is an exponential function of L_{eff}/L_d. The position of the most leaky path will not change the exponential term but only change the prefactor. If the channel furthest away from the gates, i.e., $x_c = T_{eff}/2$, $z_c = 0$, is chosen as the most leaky path, the sin and cos terms will be reduced to 1. ΔV_{th} extracted from distributive simulation at different V_{ds} is compared with the analytical solution in Fig. 9. The analytical model overestimates the V_{th} roll-off at high V_{ds} and underestimates the threshold voltage roll-off at low V_{ds}. The threshold voltage from numerical simulation is extracted with constant current definition, which is different from the ΔV_{th} definition in (24). This is the first error sources that cause overestimation at high V_{ds}. The underestimation of V_{th} roll-off at low V_{ds} may partly come from the built-in diode charge sharing that is not captured in the analytical model. Another possible reason for underestimation at low V_{ds} is the energy quantization effect that is accounted for in FIELDAY by the Poisson-Schrödinger solution in the 1-D MOSC slices, but not considered in the analytical modeling. The quantization effect is larger for smaller fins at lower V_{ds} since the drain potential penetration will help reduce the inversion carrier confinement. Normalized experimental measurement results [31] are also included in Fig. 9 for comparison, where our modeling and 3-D FIELDAY simulations are reasonably accurate for these ΔV_{th} versus L_{eff}/L_d curves.

IV. Insights Into Device Design

The six boundary conditions for 3-D Laplace's equation can be used to construct the analytical model of various SOI structures including DGFET, FDFET and SGFET. The only modifications are on the boundary conditions. Dielectric Isolation (DI) is assumed, which assumes that normal component of electric field across all isolation boundaries is zero. For DGFET, only the top boundary condition needs to be replaced with

$$\frac{\partial \psi}{\partial z}\,|_{z=H_{fin}} = 0. \tag{25}$$

For FDFET, front and back boundary conditions need to be replaced with

$$\frac{\partial \psi}{\partial z}\,|_{x=T_{fin}} = 0 \tag{26}$$

$$\frac{\partial \psi}{\partial z}\,|_{x=0} = 0. \tag{27}$$

For SGFET, only the bottom boundary condition needs to be replaced with

$$\psi\,|_{z=0} = V_g - V_{fb}. \tag{28}$$

T_{eff} and H_{eff} have to be redefined depending whether the effective dimension includes one gate oxide and two gate oxides. The formula is in the same format of (2), (3). For DGFET and FDFET with dielectric isolation, the potential solution of 3-D Laplace's equation is exactly the same as that of 2-D Laplace's equation, and the 2-D solution of DGFET can be found in [33].

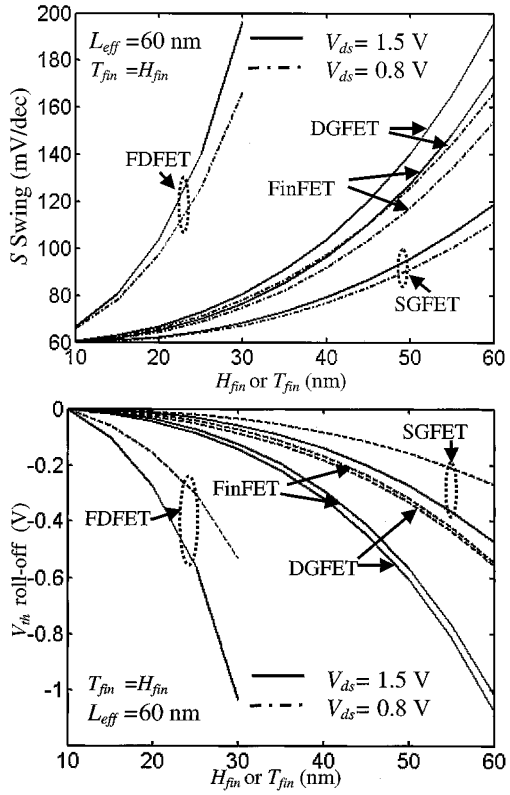

Fig. 10. Comparison of analytical S and V_{th} roll-off for FinFET, SGFET, DGFET, FDFET at $V_{ds} = 0.8$ V and 1.5 V.

For an easy comparison, the most leaky path is taken at the channel furthest away from the gates. The V_{th} roll-off of the four types of devices can be expressed in one formula

$$\Delta V_{th} = -\alpha \frac{Sq}{2.3KT} \exp\left(-\frac{L_{eff}}{2L_d}\pi\right) \cdot \left[\sqrt{\phi_{ms}(\phi_{ms} - V_{ds})} - |\phi_{ms}|\right] \quad (29)$$

where $\alpha = 32/\pi^2$ for FinFET and SGFET, $\alpha = 8/\pi$ for DGFET and FDFET. α is different because of the truncation error of the series. L_d is the drain potential decay length defined as

$$L_d = \frac{1}{\sqrt{\left(\frac{a}{T_{eff}}\right)^2 + \left(\frac{b}{H_{eff}}\right)^2}}. \quad (30)$$

$a = 1$, $b = 0.5$ for FinFET; $a = 1$, $b = 0$ for DGFET [33]; $a = 1$, $b = 1$ for SGFET; $a = 0.5$, $b = 0$ for FDFET. S for SGFET can be expressed as

$$\frac{1}{S} = \frac{q}{2.3kT} \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{16}{(2n-1)(2m-1)\pi^2} \sin\left(\frac{(2m-1)\pi}{L_{eff}} y_c\right) \times \left[\frac{1}{\cos h(0.5k_z\pi)} + \frac{1}{\cos h(0.5k_x\pi)}\right] \quad (31)$$

$$k_x = \sqrt{\left(\frac{2m-1}{L_{eff}}\right)^2 + \left(\frac{2n-1}{H_{eff}}\right)^2} \times T_{eff}. \quad (32)$$

k_z and y_c are defined the same as that of FinFET, (19), (22). S for DGFET and FDFET can be expressed by one formula

$$\frac{1}{S} = \frac{1}{2.3}\frac{q}{kT} \sum_{n=1}^{\infty} \frac{4}{(2n-1)\pi} \quad \frac{\sin\left[\frac{(2n-1)\pi}{L_{eff}} y_c\right]}{\cos h\left[\frac{(2n-1)\pi}{2L_{eff}} L_d\right]}. \quad (33)$$

S and V_{th} roll-off are plotted in Fig. 10. Since there are two variables from SGFET and FinFET, but only one variable for DGFET and FDFET, we take the case of $T_{fin} = H_{fin}$.

V. Conclusions

Using 3-D numerical simulation and analytical modeling, the scaling effects from SCE in FinFET device design are presented. The studies in this paper also provide a tractable theoretical base for design space and technology selection among FDFET, DGFET, FinFET, and SGFET.

Acknowledgment

The authors would like to thank Y. Taur and P. Soloman of IBM T. J. Watson Laboratories for many helpful discussions.

References

[1] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association (SIA), San Jose, CA, 1999.
[2] Y. Taur, L. H. Wann, and D. J. Frank, "25 nm CMOS design considerations," in *IEDM Tech. Dig.*, 1998, pp. 789–792.
[3] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S. P. Wong, "CMOS scaling into the nanometer regime," *Proc. IEEE*, vol. 85, no. 4, pp. 486–504, 1997.
[4] D. J. Frank, R. H. Dennard, and E. Nowak, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, 2001.
[5] Y. Taur, "CMOS scaling beyond 0.1 μm: how far can it go?," in *Proc. Symp. VLSI Technology*, 1999, pp. 6–9.
[6] D. J. Frank, S. E. Laux, and M. V. Fischetti, "Monte Carlo simulation of 30 nm dual-gate MOSFET: how short can Si go?," in *IEDM Tech. Dig.*, 1992, pp. 553–556.
[7] ——, "Monte Carlo simulation of p-and n-channel dual-gate Si MOSFETs at the limits of scaling," *IEEE Trans. Electron Devices*, vol. 40, p. 2103, Nov. 1993.
[8] R.-H. Yan, A. Ourmazd, and K. F. Lee, "Scaling the Si MOSFET: from bulk to SOI to bulk," *IEEE Trans. Electron Devices*, vol. 39, pp. 1704–1710, July 1992.
[9] K. Suzuki, T. Tanaka, Y. Tasaka, H. Horie, and Y. Arimoto, "Scaling theory for double-gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, pp. 2326–2329, Dec. 1993.
[10] H.-S. P. Wong, D. J. Frank, Y. Taur, and J. M. C. Stork, "Design and performance considerations for sub-0.1 μm double-gate SOI MOSFETs," in *IEDM Tech. Dig.*, 1994, pp. 747–750.
[11] H.-S. P. Wong, K. K. Chan, and Y. Taur, "Self-aligned (top and bottom) double-gate MOSFET with a 25 nm thick silicon channel," in *IEDM Tech. Dig.*, 1997, pp. 427–430.
[12] J.-H. Lee, G. Taraschi, A. Wei, T. A. Langdo, E. A. Fitzgerald, and D. A. Antoniadis, "Super self-aligned double-gate (SSDG) MOSFETs utilizing oxidation rate difference and selective epitaxy," in *IEDM Tech. Dig.*, 1999, pp. 71–74.
[13] T. Su, J. P. Denton, and G. W. Neudeck, "New planar self-aligned double-gate fully depleted P-MOSFETs using epitaxial lateral overgrowth (ELO) and selectively grown source/drain (S/D)," in *IEEE Int. SOI Conf.*, Oct. 2000, pp. 110–111.
[14] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, R. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Trans. Electron Devices*, vol. 47, pp. 2320–2325, Dec. 2000.

[15] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, R. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Sub-50 nm P-channel FinFET," *IEEE Trans. Electron Devices*, vol. 48, pp. 880–886, May 2001.
[16] D. M. Fried, A. P. Johnson, E. J. Nowak, J. H. Rankin, and C. R. Willets, "A sub-40 nm body thickness N-type FinFET," in *Proc. Device Res. Conf.*, 2001, pp. 24–25.
[17] S. H. Tang, L. Chang, N. Lindert, Y.-K. Choi, W.-C. Lee, X. Huang, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "FinFET-a quasiplanar double-gate MOSFET," in *Proc. Int. Solid-State Circuits Conf. (ISSCC)*, 2001, pp. 118–119.
[18] E. Buturla, J. Johnson, S. Furkay, and P. Cottrell, "A new 3D device simulation formation," in *NASECODE VI: Proc. 6th Int. Conf. Numerical Analysis Semiconductor Devices and Integrated Circuits*, 1989, pp. 291–296.
[19] D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *Proc. Symp. VLSI Circuits*, 1999, pp. 171–172.
[20] K. K. Young, "Short-channel effects in fully depleted SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 36, pp. 399–402, Feb. 1989.
[21] ——, "Analysis of conduction in fully depleted SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 36, pp. 504–506, Mar. 1989.
[22] T. N. Nguyen, "Small-geometry MOS transistors: physics and modeling of surface- and buried-channel MOSFETs," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1984.
[23] L. Chang, S. Tang, T.-J. King, J. Bokor, and C. Hu, "Gate length scaling and threshold voltage control of double-gate MOSFETs," in *IEDM Tech. Dig*, 2000, pp. 719–722.
[24] M. Ieong, P. M. Soloman, S. E. Laux, H.-S. P. Wong, and D. Chidambarrao, "Comparison of raised Schottky source/drain MOSFETs using a novel tunneling contact model," in *IEDM Tech. Dig.*, 1998, pp. 1080–1083.
[25] Y. Tsividis, *Mixed Analog-Digital VLSI Devices and Technology*. New York: McGraw-Hill, 1996.
[26] D. J. Frank, Y. Taur, and H.-S. P. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Device Lett.*, vol. 19, pp. 385–388, 1998.
[27] K. Suzuki, Y. Tosaka, and T. Sugii, "Analytical threshold voltage model for short channel n^+-p^+ double-gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 43, pp. 732–738, 1996.
[28] K. Suzuki and T. Sugii, "Analytical models for n^+-p^+ double gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 42, pp. 1940–1948, 1995.
[29] H. van Meer and K. De Meyer, "A 2-D analytical threshold voltage model for fullydepleted SOI MOSFETs with halos or pockets," *IEEE Trans. Electron Devices*, vol. 48, pp. 2292–2302, Oct. 2001.
[30] M. V. Fischetti and S. E. Laux, "Long-range Coulomb interactions in small Si devices. Part I: Performance and reliability," *J. Appl. Phys.*, vol. 89, no. 2, pp. 1205–1231, Jan. 2001.
[31] J. Kedzierski, D. M. Fried, E. J. Edward, T. Kanarsky, J. H. Rankin, H. Hanafi, W. Natzle, D. Boyd, Y. Zhang, R. A. Roy, J. Newbury, C. Yu, Q. Yang, P. Saunders, C. P. Willets, A. Johnson, S. P. Cole, H. E. Young, N. Carpenter, D. Rakowski, B. A. Rainey, P. E. Cottrell, M. Ieong, and H. -. P. Wong, "High-performance symmetric-gate and CMOS-compatible V_t asymmetric-gate FinFET devices," in *IEDM Tech. Dig.*, 2001, pp. 437–450.
[32] M. J. Sherony, L. T. Su, J. E. Chung, and D. A. Antoniadis, "SOI MOSFET effective channel mobility," *IEEE Trans. Electron Devices*, vol. 41, pp. 276–278, Feb. 1994.
[33] G. Pei, V. Narayannan, Z. Liu, and E. C. Kan, "3-D analytical subthreshold and quantum mechanical analyses of double-gate MOSFET," in *IEDM Tech. Dig.*, 2001, pp. 103–106.
[34] S. A. Mujtaba, "Advanced mobility models for design and simulation of deep submicron MOSFETs," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1995.
[35] E. O. Kane, "Theory of tunneling," *J. Appl. Phys.*, vol. 32, no. 1, pp. 83–91, 1961.

Gen Pei (S'01) received the B.E. degree from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1999. She is now pursuing the Ph.D. degree in electrical and computer engineering at Cornell University, Ithaca, NY.

In summer 2001, she was a Research Intern with IBM. T. J. Watson Research Center, Yorktown Heights, NY. Her current research interests include CMOS technology scaling, optimal double-gate device design for mixed-signal circuits, and integrated RF systems.

Jakub Kedzierski (M'02), photograph and biography not available at the time of publication.

Phil Oldiges (M'00) received the B.S. degree in physics from Thomas More College, Fort Mitchell, KY, in 1981, and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY, in 1984 and 1988, respectively.

From 1984 to 1986, he was a Visiting Research Scientist with Toshiba Corporation, Kawasaki, Japan, where he investigated the dynamics alpha particle induced soft errors in dynamic memories. From 1988 to 1993, he worked at Sony Corporation, Atsugi, Japan, in the area of TCAD applications of advanced silicon device technologies. From 1993 to 1998, he was a member of the TCAD Group, Digital Semiconductor, Hudson, MA, working in the area of TCAD semiconductor process simulation and the modeling of alpha-particle effects in semiconductor devices and circuits. Since 1998, he has been advancing the capabilities of the FIELDAY device simulator and investigating soft error physics in emerging technologies at IBM Microelectronics Division, East Fishkill, NY.

Meikei Ieong (SM'01), photograph and biography not available at the time of publication.

Edwin Chih-Chuan Kan (M'91) received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1984, and the M.S. and Ph.D. degrees from University of Illinois, Urbana, in 1988 and 1992, respectively, all in electrical engineering.

From 1984 to 1986, he served as a second lieutenant in the Taiwanese Air Force. In January 1992, he joined Dawn Technologies as a Principal CAD Engineer of advanced electronic and optical device simulators and technology CAD framework. He was then with Stanford University, Stanford, CA, as a Research Associate from 1994 to 1997, leading projects such as TCAD 1–2-3D tool development, software architecture definition, model hierarchy, and MEMS modeling. From 1997 to 2002, he was an Assistant Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, where he is now an Associate Professor. He has spent the summers of 2000 and 2001 at IBM Microelectronics, Yorktown Heights, NY, and Fishkill, NY, as part of the IBM Faculty Partner Program. His main research areas include CMOS technology, semiconductor device physics, system-on-a-chip, composite CAD development, and numerical methods for PDE and ODE.

Dr. Kan received the Presidential Early Career Award for Scientists and Engineers (PECASE) from the White House in October 2001.