Electro-Thermal Comparison and Performance Optimization of Thin-Body SOI and GOI MOSFETs

Eric Pop, Chi On Chui, Sanjiv Sinha[†], Robert Dutton and Kenneth Goodson[†]

Department of Electrical and [†]Mechanical Engineering, Stanford University
Contact: epop@alum.mit.edu, Bldg 500 Room 501S, Stanford CA 94305, Tel 650.723.8482, Fax 650.723.7657

ABSTRACT

This paper examines self-heating trends in ultra-scaled fully depleted SOI and GOI devices. We introduce a self-consistent model for calculating device temperature, saturation current and intrinsic gate delay. We show that the raised device source/drain can be designed to simultaneously lower device temperature and parasitic capacitance, such that the intrinsic gate delay (CV/I) is optimal. We find that a raised source/drain height approximately 3 times the channel thickness would be desirable both from an electrical and thermal point of view. Optimized GOI devices could provide at least 30 percent performance advantage over similar SOI devices, despite the lower thermal conductivity of the germanium layer.

INTRODUCTION

Ultra-thin body, fully depleted silicon-on-insulator (SOI) devices offer great promise for scaling near the end of the roadmap [1] due to better control of short channel effects and lower parasitic capacitance [2][3]. Very recently germanium-on-insulator (GOI) structures and devices have been reported [4][5], that could be even more attractive because germanium offers a mobility enhancement up to $2\times$ compared to silicon, for both electrons and holes. However, the thermal conductivity of bulk germanium is only 40 percent as large as that of silicon, which combined with the poor thermal conductivity of the buried oxide may lead to worse thermal problems for GOI than those already well documented for SOI [6][7]. In this work we analyze self-heating trends in GOI and SOI devices and show that despite the lower thermal conductivity of Ge, the temperature rise in GOI may be comparable to that in similar SOI devices, owing mainly to reduced power dissipation. We also show that ultra-thin body GOI and SOI devices can be designed to provide optimal performance, taking self-heating into account self-consistently.

SOI AND GOI MODEL ASSUMPTIONS

Thin-body SOI and GOI devices are analyzed with the lumped thermal resistance model shown in Fig. 1(a) and described in [7]. This model correctly reproduces the experimentally observed steady-state temperature rise in 100 nm channel length SOI devices [8]. In this work, the model is applied to end-of-roadmap SOI and GOI devices. The gate length (L_g), saturation current (I_d), nominal voltage (V_{dd}) and gate oxide thickness (t_{ox}) used in this study follow the most recent ITRS guidelines [1]. Other assumptions made regarding
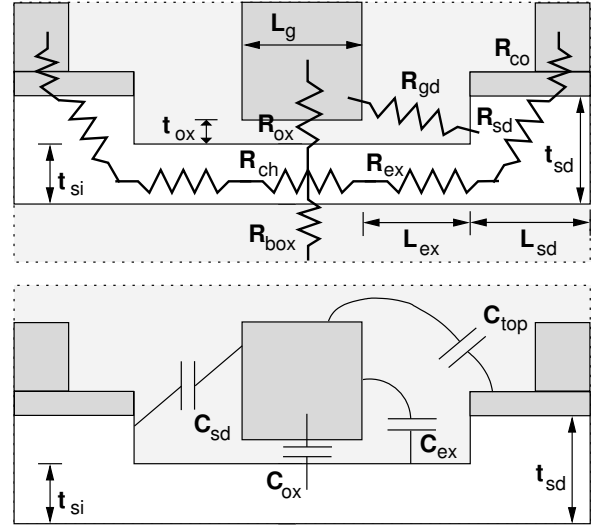

Fig. 1. Ultra-thin body MOSFET and the thermal resistances (a) and parasitic capacitances (b) used in our model. The dark gray represents the metalized gate and contacts, and the light gray is the surrounding oxide insulator. The image is not drawn to scale.

the device geometry are as follows. The SOI body thickness needed to ensure good electrostatics scales as $t_{si} = L_g/4$ [9]. The GOI body thickness should then scale by a factor of the material permittivity ratio, as $t_{ge} = \epsilon_{si} t_{si}/\epsilon_{ge} = 3t_{si}/4$. The buried oxide thickness scales as $t_{BOX} = 2L_g$ [1]. The thin body is assumed to be essentially undoped to prevent dopant fluctuation effects on the threshold voltage. The threshold voltage is then mainly determined by the choice of gate metal workfunction, which in this work is taken to be a metallic alloy with a thermal conductivity of 40 W/m/K, typical of silicides.

Figure 2 plots the thermal conductivity of undoped ultra-thin silicon and germanium films based on a Matthiessen's rule estimate [7] for the phonon mean free path. The silicon film thermal conductivity is consistent with recent measurements on 20 nm thin films ($k_{si} = 22$ W/m/K) [10], however no thermal conductivity data on thin germanium films is yet available. The ratio of the thermal conductivities, k_{ge}/k_{si}, is closer to unity (higher) in ultra-thin films than in bulk, where germanium (60 W/m/K) is only 40 percent as thermally conductive as silicon (148 W/m/K). It should be noted that ultra-thin film thermal conductivity is largely independent of temperature, because heat transport is limited by phonon boundary scattering with the film thickness [10].
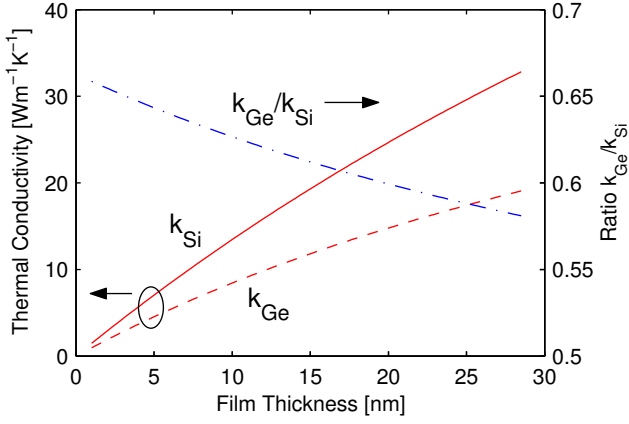
Fig. 2. Estimated thermal conductivity of thin Si and Ge layers. As the film is thinned, the thermal conductivity decreases due to phonon boundary scattering, but it decreases less (vs. bulk) for Ge films due to the shorter phonon mean free path of this material. In bulk form, the thermal conductivity ratio is $k_{ge}/k_{si} = 60/148 \simeq 0.40$, but this fraction is closer to unity for ultra-thin films.

Electron mobility in thin germanium layers is about $2\times$ higher than in thin silicon layers near room temperature [5][11]. Recent devices built by Yu *et al* [5] indicate this mobility enhancement means GOI devices can carry the same on-current (I_d) at 40 percent lower voltage (V_{dd}) than comparable SOI transistors. This is the assumption we use in the current work when comparing otherwise similar SOI and GOI transistors (except in Fig. 8 where this assumption is relaxed). However, since the FETs in Ref. [5] are large ($L_g = 10~\mu$m), this may be a conservative estimate for very small devices, where velocity saturation is less important and the $2\times$ mobility advantage of germanium could play a stronger role [12][13]. With our assumption, a GOI device dissipates 40 percent less power ($P = I_d \times V_{dd}$) than an equivalent SOI device, while generating the same ITRS-specified [1] drive current.

TEMPERATURE DEPENDENCE OF ON-CURRENT

To estimate the temperature dependence of the saturation current (per unit width) for devices near the limit of scaling, we use the following simple model [12]:

$$I_d = v_T \frac{\lambda}{2l + \lambda} C_{ox}(V_{gs} - V_t) \qquad (1)$$

where v_T is the unidirectional thermal velocity, λ the electron mean free path (both at the source), l is the distance of the first $k_B T/q$ potential drop in the channel, C_{ox} the gate oxide capacitance per unit area and V_t is the threshold voltage. The various temperature dependencies are [12][14]:

$$v_T = v_{T_o}(T/T_o)^{1/2} \qquad (2)$$
$$\lambda = \lambda_o(T/T_o)^{1/2+\alpha} \qquad (3)$$
$$l = l_o(T/T_o) \qquad (4)$$
$$V_t = V_{to} + \eta(T - T_o) \qquad (5)$$
$$\mu = \mu_o(T/T_o)^{\alpha} \qquad (6)$$

where the subscript o denotes the value at room temperature. Electron mobility in ultra-thin silicon layers has been recently

reported to vary as $T^{-1.4}$ ($\alpha = -1.4$) near room temperature, and to be largely independent of the layer thickness [15][16]. The temperature coefficient of mobility enters the mean free path from $\lambda = 2\mu(k_B T/q)/v_T$ [14]. No data is yet available on the temperature dependence of electron mobility in ultra-thin germanium layers. However, it is well known that electron mobility in bulk germanium is less temperature sensitive ($T^{-1.7}$) than in bulk silicon ($T^{-2.4}$), due to the lower optical phonon energy in germanium. By extension, in this work we assume the thin layer germanium mobility to have a T^{-1} dependence.

Finally, the threshold voltage of ultra-thin body fully depleted SOI devices varies linearly with temperature, with a coefficient η, which can be approximated as [17][18]:

$$\eta \simeq \frac{\partial \phi_F}{\partial T} = \frac{k_B}{q}\left[\ln\left(\frac{N_a}{\sqrt{N_c N_v}}\right) + \frac{1}{2k_B}\frac{\partial E_g}{\partial T} - \frac{3}{2}\right] \qquad (7)$$

where all quantities have their usual meanings (see, e.g., [17]). Recent experimental work [18] has found $\eta \simeq -0.7$ mV/K for fully depleted thin-body SOI devices. Although such data is not yet available for similar GOI devices, a quick estimate (accounting for the smaller germanium band gap, different conduction and valence band effective density of states) yields a value of η in the same range.

Taking the above into account, we obtain an expression for the temperature dependence of the saturation current for devices near the limit of scaling:

$$\frac{\Delta I_d}{I_{do}} = \left[\frac{1}{T_o}\left(\frac{1}{2} + \frac{2\alpha - 1}{2 + \lambda_o/l_o}\right) - \frac{\eta}{V_{gs} - V_{to}}\right]\Delta T, \qquad (8)$$

which is a generalization of the expression in [14]. All values with subscript o are taken to be at room temperature ($T_o = 300$ K), and in the rest of this work we assume I_{do} and V_{to} to be the values of saturation current and threshold voltage, respectively, targeted by the ITRS guidelines [1]. The temperature rise due to self-heating, ΔT, is assumed to be that at the source end of the channel, since it is this region which affects the injection velocity, mean free path and threshold voltage in Eq. 1 and in the rest of our model.

SELF-CONSISTENT CURRENT ESTIMATE

We have implemented a self-consistent iterative solution of the device temperature and current based on the model in Fig. 1 and the discussion above. The total amount of heat ($I_d \times V_{dd}$) is assumed to be entirely generated in the device drain, based on previous Monte Carlo simulation results [7]. This power is input to the thermal resistance model assuming (at first) the current to be the room-temperature value (I_{do}) targeted by the ITRS. The model yields a temperature rise at the source end of the channel (ΔT) which is used to adjust the current based on Eq. 8. The new device power is used again to solve for the device temperature, and this loop is repeated until the temperature and current are obtained self-consistently.

Figure 3 shows the calculated average temperature rise at 20 percent duty factor for SOI and GOI devices along the
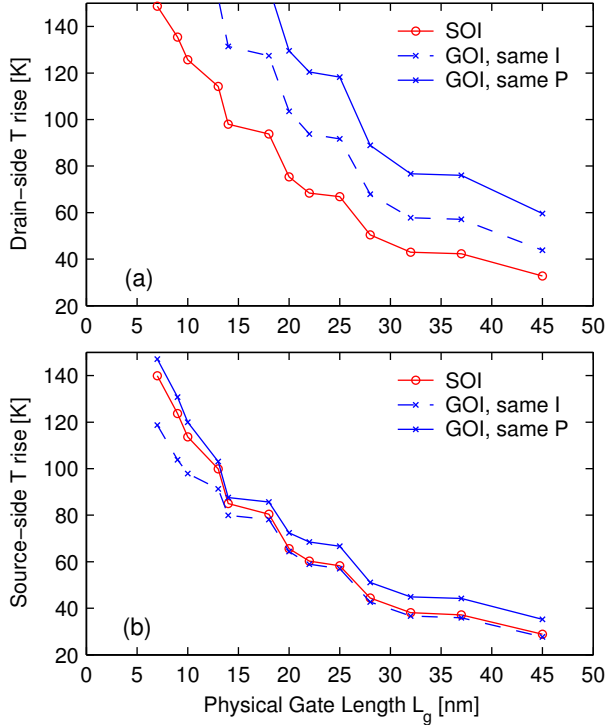
Fig. 3. Self-consistently computed average drain- (a) and source-side (b) temperature rise in SOI and GOI devices operated with a duty factor of 20 percent. Two GOI cases are shown, one with the same current (but 40 percent lower V_{dd}) as the SOI, and one with the same power as the SOI. The raised SD thickness scales as $t_{sd} = 3t_{si}$ and the channel extension as $L_{ex} = L_g/2$.
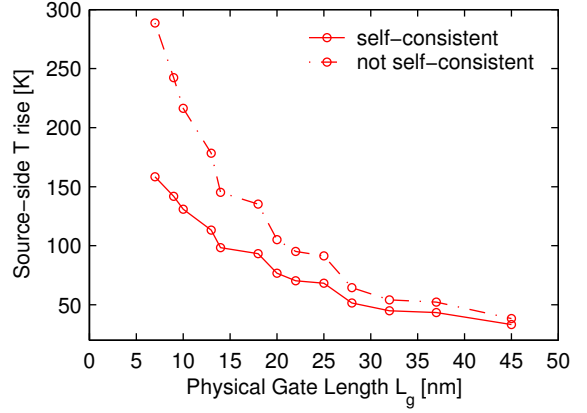

Fig. 4. Comparison of SOI source-side temperature estimate obtained from the self-consistent temperature-current calculation (solid line) and a calculation where the current is not iteratively adjusted for changes in temperature (dash-dotted line). The temperature-current consistency is important, especially for the smallest devices where the error is near 100 percent.

roadmap. The relationship between maximum (DC) temperature and the average temperature for a given duty factor f can be written as $T_{avg} = fT_{dc}$, since device thermal time constants (tens of nanoseconds) are much longer than device switching times (tens of picoseconds) [8]. Both same-current (but lower V_{dd}, hence lower power) and same-power scenarios are compared for GOI and SOI in Fig. 3. The drain temperature rise of GOI is expected to be higher in either case, due to the lower overall thermal conductivities. However, the source temperature rise is generally comparable, and even slightly lower for the same-current GOI vs. SOI case. This is due to the larger GOI channel thermal resistance, along with the lower dissipated power. Self-consistency is important in these calculations, since without it the temperature may be overestimated by close to 100 percent for the smallest devices, as shown in Fig. 4. Owing to their less temperature-sensitive mobility, GOI devices show less current degradation due to self-heating, as shown in Fig. 5.

DESIGN CONSIDERATIONS

It has been previously suggested [7] that raised source/drain (SD) and shorter extension L_{ex} are essential not only to reduce electrical series resistance, but also to reduce the thermal resistance of a device, and therefore lower its operating temperature. However, a raised SD and shorter L_{ex} can increase the gate fringing capacitance. We quantify the performance impact of the modified SD by estimating the intrinsic gate

delay, $C_g V_{dd}/I_d$. The gate capacitance components are shown in Fig. 1(b), and modeled as in Ref. [19]. For example, the fringing component C_{ex} can be written as:

$$C_{ex} = \frac{2\beta\epsilon_{sw}}{\pi} \ln\left(1 + \frac{L_{ex}}{t_{ox}}\right) \qquad (9)$$

where ϵ_{sw} is the dielectric constant of the sidewall material (here assumed to be oxide) and $\beta \simeq 0.8$ is a geometrical shape factor [19]. Figure 6 shows the computed intrinsic delay for SOI and GOI devices with the same drive current. An elevated source/drain lowers the device temperature [7] and thus improves I_d, but at the same time increases the fringing capacitance. For this reason, it appears that raising the source/drain thickness t_{sd} much beyond $3\times t_{si}$ does not result in significant additional speed gain. In Fig. 7 we optimize both the SD height as well as the extension length for an 18 nm gate device. The delay contours again suggest an optimal SD height around 3-4$\times t_{si}$ and an extension length approximately $L_g/3$ for GOI and closer to $L_g/2$ for SOI devices. Finally, in Fig. 8 we use these "optimized" device geometries and show the impact of V_{dd} scaling on GOI intrinsic gate delay. The figure explores various scenarios of $V_{dd}(\mathrm{Ge})/V_{dd}(\mathrm{Si})$, since it is not yet known what voltage well-behaved GOI devices may operate at (or, rather, at what fraction of the SOI V_{dd}).

SUMMARY

This study compares the electro-thermal behavior of GOI and SOI devices near the limits of scaling. We develop a self-consistent model for calculating device temperature, current and intrinsic gate delay. We show how the device source/drain can be designed to help simultaneously minimize device temperature and parasitic capacitance, such that the intrinsic gate delay is optimal. Finally, we show that optimized GOI devices could provide at least 30 percent performance advantage over similar SOI devices, even when self-heating is taken into account.
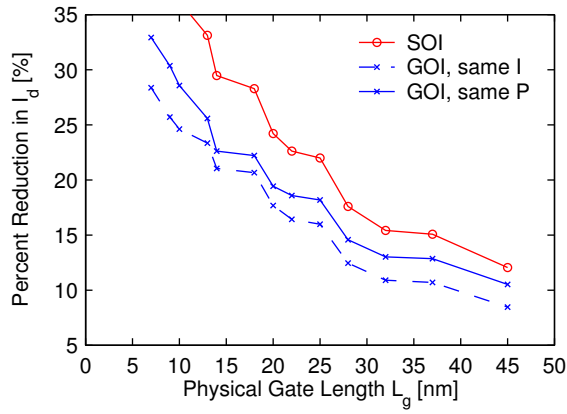
Fig. 5. Self-consistently computed percentage decrease in drain current due to self-heating (vs. the ITRS-targeted current), for the same cases as in Fig. 3
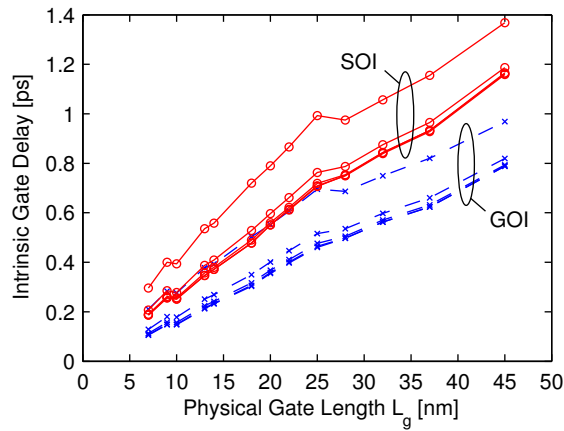

Fig. 6. Self-consistently computed intrinsic delay for SOI and GOI devices in the same-current scenario. The SD height t_{sd} is varied as a parameter, from t_{si} (no raised SD, top line in each set of curves) to $5t_{si}$. The extension length is assumed constant at each node, $L_{ex} = L_g/2$. The intrinsic delay is not reduced significantly for $t_{sd} > 3t_{si}$.

REFERENCES

[1] ITRS 2003 [online] http://public.itrs.net
[2] B. Doris et al, Proc. IEDM, p. 631, 2003
[3] T. Ernst, IEEE TED, v. 50, p. 830, 2003
[4] N. A. Bojarczuk et al, Appl. Phys. Lett. v. 83, p. 5443, 2003
[5] D. S. Yu et al, IEEE EDL, vol. 25, p. 138, 2004
[6] L. Su et al, IEEE TED, vol. 41, p. 69, 1994
[7] E. Pop et al, Proc. IEDM, p. 883, 2003
[8] S. Polonsky et al, IEEE EDL, v. 25, p. 208, 2004
[9] H.-S.P. Wong et al, Proc. IEDM, p. 407, 1998
[10] W. Liu et al, J. Heat Transfer, in press, 2004
[11] A. Khakifirooz et al, IEEE EDL, vol. 25, p. 80, 2004
[12] M. Lundstrom, IEEE EDL, vol. 22, p. 293, 2001
[13] S. Takagi, Symp. VLSI, p. 115, 2003
[14] M.-J. Chen et al, Proc. IEDM, p. 39, 2002
[15] D. Esseni et al, Proc. IEDM, p. 671, 2000
[16] F. Gamiz, Semic. Sci. Tech., v. 19, p. 113, 2004
[17] Y. Taur and T. Ning, Cambridge Univ. Press, 1998
[18] L. Vancaillie, IEEE Int. SOI Conf., p. 78, 2003
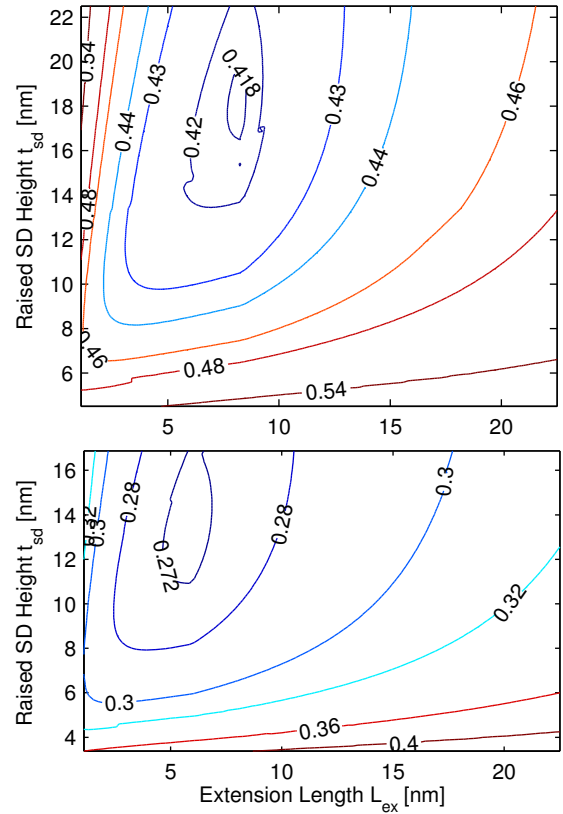[19] N. R. Mohapatra et al, IEEE TED, v. 50, p. 959, 2003

Fig. 7. Geometry optimization to minimize intrinsic delay for a SOI (top) and GOI device (bottom) with $L_g = 18$ nm and $t_{si} = 4.5$ nm, assuming the GOI device provides the same current at 40 percent less V_{dd}. The results are expressed as contour plots of the delay (in picoseconds) with the extension length (L_{ex}) and SD thickness (t_{sd}) as parameters.
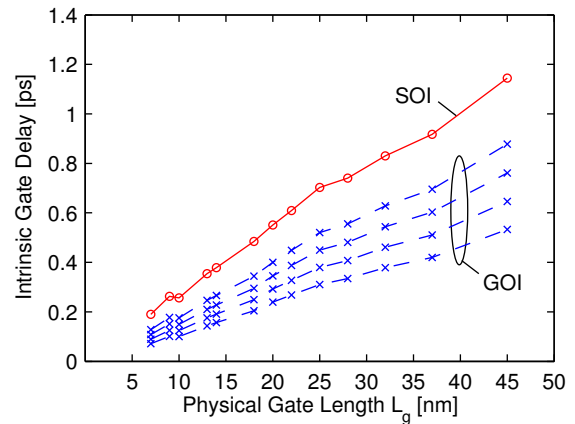

Fig. 8. Intrinsic gate delay for SOI and GOI devices with optimized $L_{ex} = L_g/3$ and $t_{sd} = 3t_{si}$. The GOI voltage is varied as a parameter from $0.5V_{dd}$ (bottom dashed line) to $0.8V_{dd}$ (top dashed line) in increments of $0.1V_{dd}$, where V_{dd} is the nominal SOI voltage from the ITRS guidelines [1].