Thermal Analysis of Ultra-Thin Body Device Scaling

Eric Pop, Robert Dutton and Kenneth Goodson[†]

Dept. of Electrical and [†]Mechanical Engineering, Stanford University, Stanford, CA 94305
Tel: 650.723.8482, Fax: 650.723.7657, Email: epop@alum.mit.edu

Abstract

This paper explores the effect of confined dimensions and complicated geometries on the self-heating of ultra-thin body SOI and FinFET devices. A compact thermal model is introduced, incorporating the most advanced understanding of nanoscale heat conduction available. Novel device scaling is analyzed from a thermal point of view. We show device temperatures are very sensitive to the choice of drain and channel extension dimensions, and suggest a parameter design space which can help alleviate thermal problems. ITRS power guidelines below the 25 nm technology node should be revised if isothermal scaling of thin-body devices is desired.

Introduction

Ultra-thin body (UTB) silicon-on-insulator (SOI) devices have shown great potential for scaling toward channel lengths of 10 nm and below [1][2]. Since the thermal conductivity of the oxide insulator is two orders of magnitude less than that of silicon, SOI devices have been associated with self-heating problems [3]. This is somewhat alleviated if the buried oxide becomes thinner. However, confined dimensions and more complicated device geometries can still lead to significant self-heating. This paper estimates such trends in single-gate (SG) UTB-SOI and FinFET devices near the limits of scaling. The impact of the reduced thermal conductivity of doped ultra-thin layers on steady-state device temperatures is investigated. While other studies have looked at the electrostatic scaling properties of such novel devices [4]–[6], this is the first paper to analyze their scaling from a thermal point of view.

Previous work [7][8] has suggested that sub-continuum heat conduction effects play a role in bulk nanotransistors because the area of most intense heat generation (i.e. the phonon "hot spot") is much smaller than the phonon mean free path ($\Lambda_b \sim 100$ nm in bulk silicon). In thin body transistors the phonon mean free path is limited by the body thickness (t_{si}), hence the sub-continuum effect predicted for bulk devices is of less importance. However, the thermal conductivity of doped ultra-thin films is strongly reduced from the bulk silicon value (148 $\mathrm{Wm^{-1}K^{-1}}$) by phonon boundary and impurity scattering. The thermal conductivity can be written as [9]

$$k = Cv\Lambda/3 \tag{1}$$

where C is the volumetric heat capacity, v the average phonon velocity and Λ is an appropriately averaged phonon mean free path. Ju [10] showed that the thermal conductivity of undoped crystalline silicon is halved in films of thickness $t_{si} \sim 100$ nm. Other recent experimental work [11] found k decreasing about 30 percent in highly doped bulk silicon characteristic of modern devices. The combined thin film and impurity effects can be approximated by writing the phonon mean free path via Mathiessen's rule as

$$1/\Lambda = 1/\Lambda_b + 1/t_{si} + 1/\Lambda_{imp} \tag{2}$$

where the last term is the mean phonon-impurity scattering length. This can be extrapolated to ultra-thin films that have not been experimentally measured. For example, the thermal conductivity of a highly doped 10 nm silicon film is estimated at 13 $\mathrm{Wm^{-1}K^{-1}}$, less than 9 percent of the bulk value.

The thermal conductivity of bulk silicon dioxide ($k_{ox} = 1.38\ \mathrm{Wm^{-1}K^{-1}}$) is about two orders of magnitude less than that of silicon. Measurements on metal-oxide-silicon (MOS) structures have shown that for very thin films the interface resistance of the materials plays a significant role. The data of [12] and [13] can be interpreted if the MOS oxide thermal resistance is written as

$$R_{ox} = (R_i + t_{ox}/k_{ox})/A \tag{3}$$

where t_{ox} is the thickness, A is the area of the oxide, and the total interface resistance R_i is about $2 \times 10^{-8}\ \mathrm{m^2KW^{-1}}$. The origin of this value is not well understood, since it is about an order of magnitude larger than theoretical predictions based on phonon dispersion mismatch at the boundaries [14]. This theoretical lower bound is slightly enhanced by heat carrier (electron-phonon) conversion at the metal/oxide interface. More subtle boundary effects may be playing a significant role, e.g. near-interfacial disorder in the metal, such as porosity or small grain size. The value of R_i was reported to be relatively independent of processing conditions, and is approximately equivalent to the thermal resistance of a 20 nm oxide film.

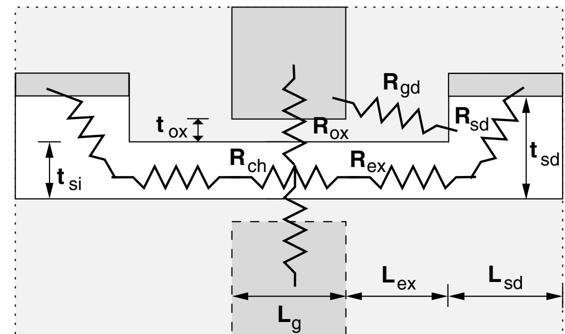

Fig. 1. Schematic of the ultra-thin body FET, thermal resistances and dimensions of interest. The dark gray represents the metalized gate and contacts. The second gate of a dual-gate device is drawn with a dashed line. For a FinFET this becomes the top view and the gate is wrapped over the channel. The device is otherwise surrounded by oxide insulator.

0-7803-7873-3/03/$17.00 (c) 2003 IEEE

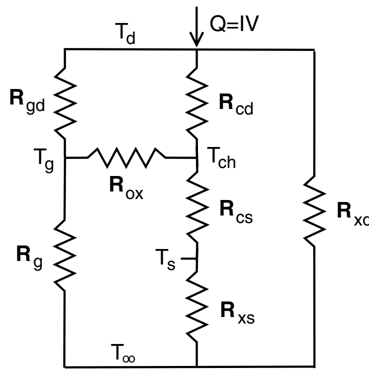

Fig. 2. Equivalent thermal circuit of the thin-body FET. R_g is the gate thermal resistance, R_{cd} and R_{cs} are the drain- and source-side thermal resistance of the thin body channel. R_{xd} and R_{xs} contain R_{sd} from Fig. 1 in series with the drain- and source-side component of the thin channel extension, respectively. Other thermal resistances are defined in Fig. 1.

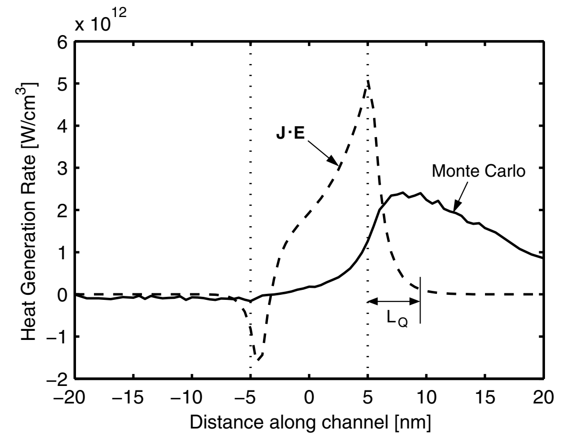

Fig. 3. Monte Carlo (MC) simulation of heat generation along the channel of a 10 nm DG-SOI with $V_{dd} = 0.6$ V [17]. The source and drain are to the left and right of the vertical dotted lines, respectively. Unlike the profile of the $\mathbf{J} \cdot \mathbf{E}$ product (dashed line), the MC result (solid line) shows heat being generated past the electric field peak and well inside the drain. The centroid of the heat generation region is a distance L_Q from the edge of the gate.

This means that the thermal resistance of the nanometer-thin silicon dioxide films found in practical MOSFET devices is essentially independent of the oxide thickness. In other words smaller, thinner FETs with metal gates do not have an advantage of increased cooling through the gate oxide, despite a thinner insulator.

THERMAL MODEL

A typical UTB-SOI device is modeled using the thermal circuit in Figs. 1 and 2. All dimensions, thermal resistances and temperatures of interest are labeled on the two diagrams. The FinFET thermal circuit is similar, except the gate and gate oxide thermal resistances (R_g and R_{ox} respectively) are lower due to a larger surface area. The FinFET body ("fin") resistance is also different due to additional boundary scattering from the limited fin height, itself less than the bulk silicon phonon mean path. The thermal resistance of each device portion is written as

$$R = L/(kA) \tag{4}$$

where the material thermal conductivity k is adjusted for boundary and impurity scattering, L is the dimension along the heat conduction direction and A is the cross-sectional area of heat flux perpendicular to it.

The channel region of the thin-body devices is assumed undoped, to prevent discrete dopant fluctuation effects on the threshold voltage. All other silicon regions (channel extension L_{ex}, source and drain L_{sd}) are highly doped to reduce series resistance. The gate is a metallic alloy whose work function sets the device threshold voltage as needed. The metalized contacts are assumed to be at the temperature T_∞ which all other temperature rise estimates are referenced to. This background temperature is a function of on-chip device density, layout, surrounding circuit activity [15] and cooling technology in the packaging (in a modern chip, T_∞ can reach 360-380 K near the clock [16] and this figure is expected to increase unless significant improvements are made in thermal packaging).

Monte Carlo simulations suggest the heat generation region extends several tens of nanometers into the drain [17] as in Fig. 3. For the purpose of this lumped thermal resistance model, the heating $Q = I_{on} \cdot V_{dd}$ is assumed to occur only at the centroid of the heat generation profile. The distance from this point to the drain edge of the gate can be written approximately as $L_Q = 5V_{dd}/0.6$ nm, where V_{dd} is the drain voltage. Depending on the values of V_{dd} and of the channel extension length L_{ex}, the heat generation centroid may be located in the channel extension (if $L_Q < L_{ex}$) or it may be pushed into the drain (if $L_Q > L_{ex}$).

For this scaling study, all standard device parameters (gate length L_g, oxide thickness t_{ox}, nominal voltage V_{dd} and current I_{on}) are taken from the ITRS roadmap [18]. The body thickness needed to ensure good electrostatics is assumed to vary as $L_g/4$ for SG-SOI devices, and as $L_g/2$ for dual-gate (DG) SOI and FinFETs [4]. Fig. 4 illustrates how the body thermal conductivity decreases for the thin SG-SOI device layers along the ITRS roadmap. The source/drain (SD) fan-out region must be epitaxially raised to reduce the parasitic series resistance [19]. This also brings a double thermal benefit: it improves the region's thermal conductivity by allowing a larger phonon mean free path (Eq. 2) and it lowers its thermal resistance by offering a larger area for lateral heat flow.

The thermal circuit model is first applied to estimate the steady-state temperature rise of scaled SG-SOI devices, as in Fig. 5. Having written all thermal resistances as a function of their dimensions and material properties, the thermal circuit of Fig. 2 can be solved by applying the thermal equivalent of Ohm's law, $T = Q \cdot R$. We define T_d at the point of maximum heating rate, making it the highest temperature, i.e. the worst-case scenario for a given device design. The temperature T_s is defined at the point directly under the source-side edge of the gate. This is the temperature at the region of electron injection into the channel, and it affects the channel injection velocity, backscattering coefficient and the ultimate current drive of the device [20]. Both T_d and T_s affect the temperature-dependent

0-7803-7873-3/03/$17.00 (c) 2003 IEEE

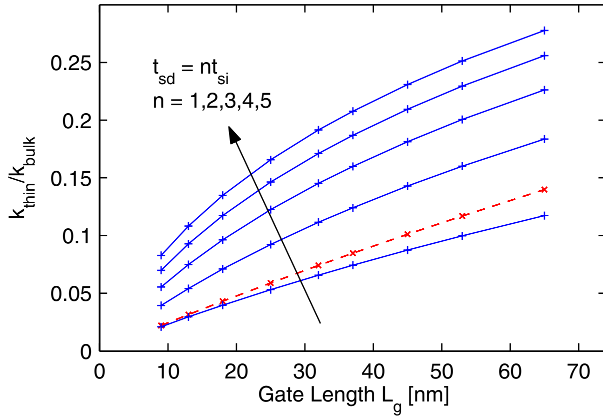

Fig. 4. Thermal conductivity scaling of ultra-thin device layers, as a fraction of the bulk silicon thermal conductivity. The dashed line represents the thermal conductivity of the undoped channel (assumed to scale as $t_{si} = L_g/4$). The five solid lines are the thermal conductivity of the highly doped source/drain regions, with varying thicknesses from $t_{sd} = t_{si}$ (bottom) to $t_{sd} = 5t_{si}$ (top). For the shortest devices, the thermal conductivity of the thin layers can drop well below 10 percent of its value in bulk silicon (148 $\mathrm{Wm^{-1}K^{-1}}$).

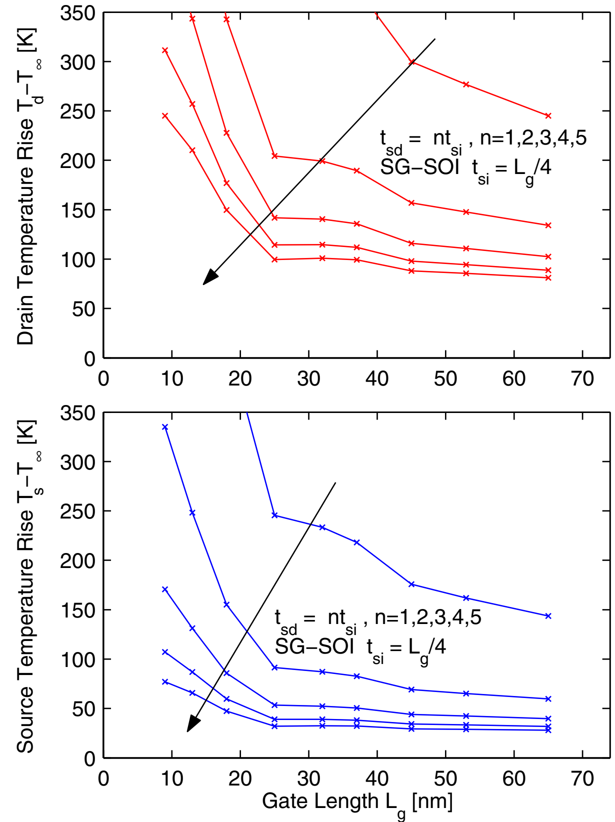

Fig. 5. Estimated drain-side (top) and source-side (bottom) steady-state temperature rise for single-gate UTB-SOI devices along the ITRS roadmap. The extension length is assumed to scale as $L_{ex} = L_g/2$. The five sets of lines correspond to raised source/drain heights between $t_{sd} = t_{si}$ and $t_{sd} = 5t_{si}$. The taller source/drain reduces electrical series resistance and also helps decrease device temperature, owing to its higher thermal conductivity and larger conduction area.

series resistance. The "kinks" in the temperature trends occur due to non-uniform scaling of the discrete V_{dd} and I_{on} values adopted from the ITRS roadmap [18]. At the smallest device dimensions the reduced thermal conductivity of the ultra-thin layers is expected to significantly impede heat transport out of the device, and hence raise its temperature.

Since a thin body is required by electrostatics and because heat transfer through the oxide is limited by interface thermal resistance, another way to ensure heat is more easily transferred out is to lower the thermal resistance of the source and drain regions. From the trends in Fig. 5, isothermal device scaling requires $t_{sd} \sim 5t_{si}$ if the channel extension length scales as $L_{ex} = L_g/2$, except at technology nodes below 25 nm. The temperature of the shortest devices increases despite the gradually planned power reduction along the ITRS roadmap from 1100 (at the 65 nm node) to 600 $\mu\mathrm{W}/\mu\mathrm{m}$ (at the 9 nm node). All else being equal, the power should be decreased more to achieve isothermal scaling of thin-body devices below the 25 nm node.

Discussion

Since most of the heat is generated in the drain-side channel extension (or just inside the drain if $L_Q > L_{ex}$), the two device parameters with most impact on temperature are the extension length L_{ex}, and source/drain height t_{sd}. Figs. 6 and 7 illustrate their effect on the source and drain temperatures of 25 nm SG-SOI and FinFET devices, respectively. Note the sensitivity of temperature on these parameters, as they control the drain (and extension) thermal resistance. The contours suggest a range of parameter values for the thermal optimization of these devices, i.e. a short L_{ex} and tall t_{sd}. While preferable both from an electrical and thermal point of view, such a choice would likely increase the gate-to-source/drain parasitic capacitance. The smallest L_{ex} (assumed to be as thin as t_{ox} in Figs. 6 and 7) may in practice be limited by technological control over the spacer width. The choice of parameters must also account for the geometry effect on dopant diffusion into the channel extension and possibly under the gate. Fig. 8 compares the drain-side temperature rise for SG-SOI and FinFET devices along the ITRS roadmap. The thicker body and larger wrap-around gate/oxide surface of the FinFET help lower its temperature, despite increased phonon scattering with the limited fin height.

Conclusions

This study explores the thermal impact of several parameters on the scaling of UTB single-gate and FinFET devices. It is shown that device temperatures are very sensitive to the drain and extension dimensions. Lowering the drain region's thermal resistance (e.g. by epitaxially raising it) can aid heat dissipation. Steady-state temperatures are used throughout because they represent the worst-case scenario for a given device design. They also must be accounted for during testing and parameter extraction, and are of particular concern for devices that are ON most of the time (biasing elements, current mirrors) or have high duty cycles (clock, I/O driver). The ultimate parameter design choices will need to involve thermal

0-7803-7873-3/03/$17.00 (c) 2003 IEEE

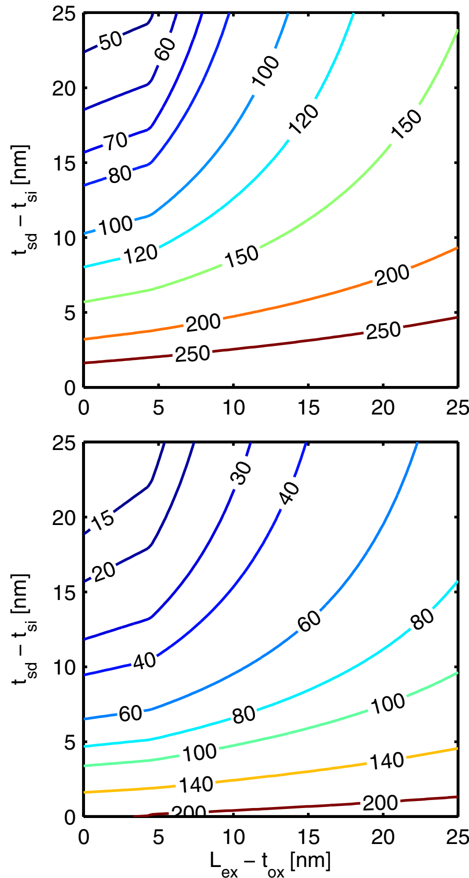

Fig. 6. Drain-side (top) and source-side (bottom) temperature rise of a SG-SOI device with $L_g = 25$ nm and $t_{si} = 7$ nm. The results are expressed as contour plots of the temperature rise (in Kelvins) above T_∞, with the extension length L_{ex} and raised source/drain height t_{sd} as parameters. The temperatures are minimized when t_{sd} is maximized and $L_{ex} < L_Q$, such that part of the heat generated near the drain can escape directly through the gate oxide and sidewall.

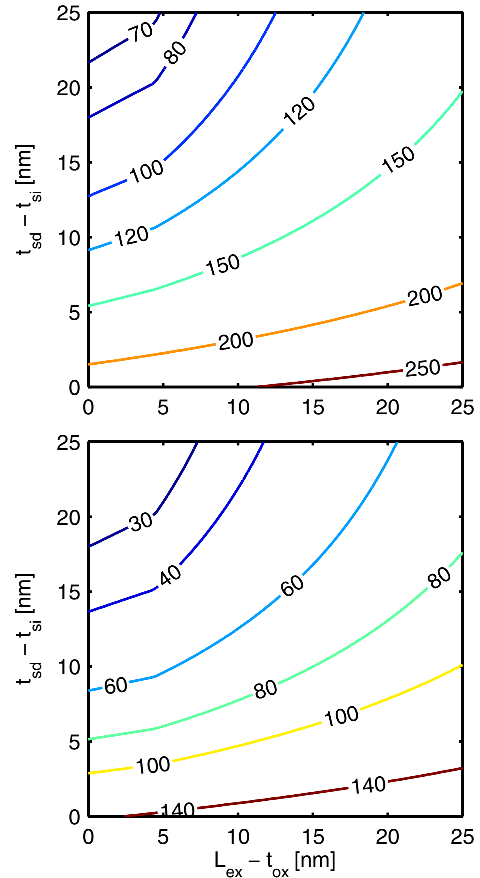

Fig. 7. Drain-side (top) and source-side (bottom) temperature rise of a FinFET device with $L_g = 25$ nm and $t_{si} = 13$ nm. The results are expressed as contour plots with the extension length and raised source/drain height as parameters. The FinFET channel width is 2 × fin height. Compare with the SG-SOI temperature contours from Fig. 6.

as well as electrical and technological considerations.

We thank C. Chui, A. McConnell and S. Sinha for helpful comments. E. Pop is supported by the IBM/SRC Fellowship.

References

[1] B. Yu et al, Proc. IEDM, p. 251, 2002
[2] M. Ieong et al, Proc. ISQED, p. 492, 2002
[3] L. Su et al, IEEE TED, vol. 41, p. 69, 1994
[4] H.-S.P. Wong et al, Proc. IEDM, p. 407, 1998
[5] G. Pei et al, IEEE TED, vol. 49, p. 1411, 2002
[6] L. Chang et al, IEEE TED, vol. 49, p. 2288, 2002
[7] E. Pop et al, Proc. IEDM, p. 677, 2001
[8] P. Sverdrup et al, J. Heat Transfer, vol. 123, p. 130, 2001
[9] J.M. Ziman, *Electrons and Phonons*, Oxford Press, 1960
[10] Y.S. Ju and K.E. Goodson, Appl. Phys. Let., vol. 74, p. 3005, 1999
[11] M. Asheghi et al, J. Appl. Phys., vol. 91, p. 5079, 2002
[12] S.-M. Lee et al, J. Appl. Phys., vol. 81, p. 2590, 1997
[13] T. Yamane et al, J. Appl. Phys., vol. 91, p. 9772, 2002
[14] E.T. Swartz and R.O. Pohl, Rev. Mod. Phys., vol. 61, p. 605
[15] K.E. Goodson et al, IEEE TCPMT, vol. 15, p. 715, 1992
[16] A. Akturk et al, Proc. SISPAD, p. 311, 2003
[17] E. Pop et al, Proc. SISPAD, p. 121, 2003
[18] ITRS Roadmap [Online] http://public.itrs.net
[19] J. Kedzierski et al, IEEE TED vol. 50, p. 952, 2003
[20] M. Lundstrom and Z. Ren, IEEE TED, vol. 49, p. 133, 2002

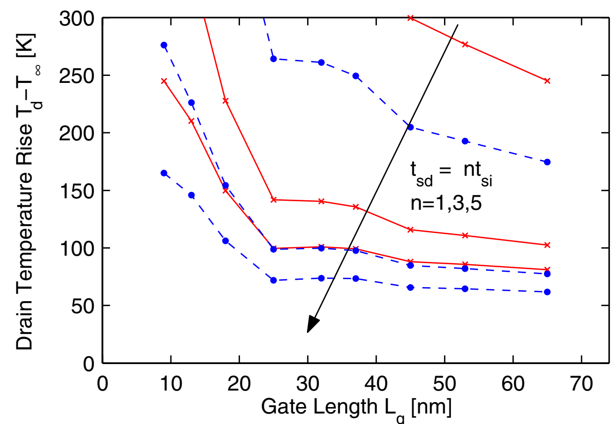

Fig. 8. Comparison of drain-side temperature rise for SG-SOI (solid lines) and FinFET (dashed lines) devices along ITRS roadmap. The FinFET channel width is 2 × fin height and the fin height is 4 × t_{si}. Extension lengths are assumed to scale as $L_{ex} = L_g/2$. The thicker body and larger wrap-around gate/oxide surface of the FinFET help lower its temperature.

0-7803-7873-3/03/$17.00 (c) 2003 IEEE