Thermal Phenomena in Deeply Scaled MOSFETs

Jeremy Rowlette[1], Eric Pop[1], Sanjiv Sinha[2,3], Mathew Panzer[2], and Kenneth Goodson[2]

Depts. of [1]Electrical and [2]Mechanical Engineering, Stanford University, Stanford, CA and [3]Intel Corporation, Hillsboro, OR
Contact: rowlette@stanford.edu, Bldg 530 Room 101, Stanford CA 94305-3030, tel 650.725.3314

Abstract- Thermal phenomena are having an increasing influence on drive and leakage currents in modern transistors. This trend is accelerated for confined-geometry devices, which include thermally-resistive interfaces and materials with low thermal conductivity (e.g. SiO_2, $Si_{1-x}Ge_x$). This paper summarizes the nanotransistor thermal design challenges and reviews the latest advancements in electro-thermal modeling.

Introduction

Within the last decade, there has been growing concern with thermal phenomena occurring on the microprocessor, including interconnect and via self-heating. Self-heating of modern transistors has received increased attention as an important consideration for simulations at the device and circuit levels.

Dynamic power dissipation in nanotransistors will not be significantly reduced by device scaling and the volumetric heat generation density will therefore rise sharply each generation. The dynamic power of digital logic circuits based on nanotransistor technologies scales roughly as $C_{eff}V_{dd}^2f$, where the effective capacitance (C_{eff}) includes significant contributions from the interconnects. The square-law dependence on the operating voltage (V_{dd}) appears to offer the strongest lever for reducing dynamic power. However, a reduction in V_{dd} would require a comparable decrease in threshold voltage (V_t) in order to maintain drive current. The threshold voltage cannot be arbitrarily reduced since source-drain leakage generally grows exponentially with reductions in V_t [1]. A tradeoff between performance and power consumption naturally ensues with the threshold voltage acting as the central parameter in optimization. As a consequence of this tradeoff, V_{dd} and V_t have historically scaled slowly in comparison to the physical dimensions of devices. This long-term trend has led to stronger electric fields and rapidly increasing power densities [2] to the extent that major shifts in process design, performance targets, and architecture are now necessary [3]. Even with a slower projected trend in frequency scaling, only modest reductions in V_{dd} are expected. Therefore, volumetric power density of nanotransistors will continue to grow rapidly with each process generation and will be compounded by shifts towards confined geometry devices.

Leakage power and its relationship to thermal transport should receive more attention in future device design. A key characteristic difference between leakage and dynamic power consumption should be made clear at this point. Unlike dynamic power, leakage power tends to increase with junction temperature, which implies that an increase in either power or thermal resistance will cause further increases in leakage power consumption by means of positive feedback [4]. As frequencies and power densities increase, a greater fraction of the heat generated during device switching persists in the *off* state, increasing total leakage. Furthermore, the device performance, e.g. the *CV/I* metric, is a relatively weak function of device temperature compared to that for leakage [1].

Heat Generation and Conduction

Electrons accelerate in the electric field and heat the lattice through electron-phonon scattering (Fig. 1) at a total rate proportional to the familiar macroscopic power relation: P=I•V. Nanoscale transistors require greater attention to high-electric-field, ballistic, and quantum phenomena, both for the electron and phonon systems and for proper description of their interaction. As device channel lengths decrease below 50 nm, a simulation approach deliberately incorporating electron-phonon scattering and the complete dispersion characteristics of each particle type becomes more relevant. Furthermore, it becomes important to track the detailed populations of phonons generated by electron scattering, since the phonon distribution function strongly influences the ability of heat to conduct away from the device.

In deeply-scaled transistors, energy is transferred to the phonon system within about 50 nm of the highly localized (~10 nm) electric field region centered near the channel-drain boundary [5]. Energy is then carried out to the contacts and package boundaries primarily by means of acoustic phonon propagation. Resistance to acoustic phonon transport by scattering with other phonons, impurities, and boundaries leads to increases in the phonon population within the device, which can be described using a higher effective temperature (T_{eff}) [6]. Phonon occupation ($N=1/(\exp(\hbar\omega/k_BT_{eff})-1)$) directly affects the electron-phonon scattering rates which are proportional to ($N + ½ \pm ½$), where the signs correspond to phonon emission (+) and absorption (-) processes, respectively [7]. Phonon scattering influences the electron transport in two important ways. First, the effective mobility is reduced, leading to lower drive currents. Second, phonon absorption near the source-channel boundary can lead to increased electron injection efficiency. The second effect can enhance device performance in the *on* state by increasing drive current but can also increase subthreshold leakage in the *off* state.

Boundary scattering will be a major source of thermal resistance in future low-dimensional devices such as FinFETs [8,9]. Although bulk silicon has a high intrinsic thermal conductivity, the effective thermal conductivities of ultra-thin silicon films are severely reduced (~10x) by phonon boundary scattering [10,11]. Furthermore, such devices routinely use SiO_2 or other low intrinsic thermal conductivity materials to electrically isolate the active regions of the device. An electro-thermal optimization, which incorporated boundary

0-7803-9269-8/05/$20.00 (c) 2005 IEEE

scattering and thermal resistances of the surrounding medium, was performed by Pop *et al*. for a raised S/D thin body FET with sub-35 nm channel lengths [11]. It was shown that the benefits of increasing the S/D layer thickness in order to improve heat conduction was countered by the associated increases in capacitive coupling to the gate (Fig. 2). An optimal S/D layer thickness was found to be about 3-4x the body thickness. As the spatial density of interfaces increases with the reduction of device dimensions, heat transport across boundaries will play an increasing role in determining the overall thermal properties of the device [6]. For example, the thermal interface resistance associated with a metal-oxide-silicon stack was shown to be equivalent to the thermal resistance of ~20 nm of oxide [12]. The development of accurate boundary resistance models has proven to be extremely difficult and is still the subject of active research today [13-15].

Within 10's of nanometers of the heat source, neither charge nor heat energy transport is diffusive in nature. Rather, electrons and phonons propagate under near-ballistic conditions leading to highly non-equilibrium distributions for each carrier type. The Boltzmann transport equation (BTE) can be solved to obtain accurate distribution profiles for each particle type provided that certain quantum phenomena can be neglected [7]. Significant progress has been made in the detailed modeling of independent electron and phonon transport processes in silicon devices by solving various limiting forms of the BTE (e.g., Refs. [6, 16, 17]). A rigorous solution to the electron BTE (e-BTE) was performed in Ref. [18] using an electron Monte Carlo (e-MC) technique optimized for calculating phonon generation rates (spectra) in silicon devices, which included optical and acoustic phonon dispersion. Detailed simulations [19] showed that ~2/3 of the heat generation occurred in the optical phonon branches for electric field strengths typical of modern transistors. However, despite receiving a majority of the energy during hot carrier relaxation in the drain, optical phonons have essentially no contribution to heat transport because their group velocities are small compared to those of the long wavelength acoustic modes. Understanding the decay processes of optical phonons into the heat conducting acoustic modes is therefore an essential aspect of electro-thermal device modeling at nanoscales.

A recent molecular dynamics study by Sinha *et al*. [20] offered crucial insight into the decay processes of the intervalley *g*-type longitudinal optical phonons (*g*-LO). Phonons centered around the *g*-LO mode in k-space (~0.3 of the distance to the edge of the Brillouin zone) are of special interest because they scatter strongly with hot electrons and have a relatively low density-of-states, characteristics which lead to *hot* phonon behavior [5, 21]. Determining the dependence of optical phonon lifetimes on power density is also of great interest to electro-thermal modeling of realistic devices. Because of a lack of experimental data, most optical phonon lifetime parameters used in phonon transport models rely exclusively on optical Raman data and calculations where there *is* a wealth of knowledge spanning decades of research [22-24]. For equilibrium conditions near 300 K, Raman mode lifetimes have been reported to be around 4 ps [22] and are expected to decrease with increasing local power density [20].

Coupled Electrical and Thermal Simulations

Despite significant advances in independent transport simulations, self-consistent electro-thermal simulations in the sub-continuum regime (< 100 nm) are very limited in number and in scope. Nearly a decade ago, Lai and Majumdar [25] developed a coupled hydrodynamic solution separating the optical and acoustic phonon branches. Recently, Sinha *et al*. [26] introduced a computationally efficient split-flux solution to the phonon BTE (p-BTE), which captures ballistic phonon conduction near the nanometer-sized heat source and yields a convenient interface to continuum calculations far from the heat source. Using the detailed phonon generation rates calculated by e-MC in Ref. [19] the split-flux p-BTE model was used to simulate 2-D steady-state and 1-D transient heating inside of an 18 nm depleted substrate SOI device (Figs. 3 and 4) [26]. Most recently, we have performed steady-state, self-consistent, coupled, electron-phonon transport simulations of an isolated, bulk, 20 nm, n+/n/n+ silicon device (Figs. 5-7) [27] by combining the e-MC presented in Ref. [18] with the split-flux p-BTE method presented in Ref. [26]. The impact of self-heating for the n+/n/n+ device was studied for three different voltages, the results of which are summarized in Figs. 8-10. Results showed an effective lattice temperature rise as high as ~50 K with respect to the reference temperature (300 K) at the device contacts for the 1.0 V bias condition. For such a temperature rise, it was found that current was reduced by only 2.2%. In a more-realistic thin-body device structure, phonon boundary scattering would act to increase the effective thermal resistance of the device by about an order of magnitude. Neglecting the strong negative feedback on electron transport that would occur, such an increase in thermal resistance would lead to projected temperature rises of ~10x that which was calculated for the bulk device.

Conclusion

We have reviewed the recent progress in electro-thermal simulation of advanced silicon devices. We have examined the effects of boundary and interface scattering as well as non-equilibrium ballistic transport on charge and heat conduction inside of deeply scaled MOSFETs. Recent work has culminated in the demonstration of self-consistent, electron-phonon simulations using an electron Monte Carlo technique and a split-flux form of the phonon BTE. It is apparent that the small, restricted dimensions of future devices will affect both current and heat transport, while thermal cross-talk is expected to be of significance between closely packed devices, device fins or nanowire/tube segments.

Acknowledgments

This work was supported by the Semiconductor Research Corporation (SRC) Task 1043.

References

[1] Y. Taur and T. Ning. Fundamentals of Modern VLSI Devices *Cambridge. U. Press* (1998)
[2] S. Borkar. *Design Automation Conference*, 293–296 (2001)
[3] D. Geer. IEEE *Computer* **38**, 8-13 (2005)
[4] K. Banerjee, S. Lin, A. Keshavarzi, S. Narendra, V. De. *IEDM Tech. Digest*, 887-890 (2004)
[5] E. Pop, J. Rowlette, R. Dutton, K. Goodson. *Proc. of IEEE SISPAD* (2005)
[6] D. Cahill, W. Ford, K. Goodson, G. Mahan, A. Majumdar, H. Maris, R. Merlin, S. Phillpot. *J. Appl. Phys*. **93**, 793 (2003)
[7] M. Lundstrom. Fundamentals of Carrier Transport. *Cambridge U. Press* (2000)
[8] Y. Choi, K. Asano, N. Lindert, V. Subramanian, T. King, J. Bokor, C. Hu. *IEEE Electron. Dev. Lett.* **21**, 254-255 (2000)
[9] B. Doris *et al. IEDM Tech. Digest*, 267-270 (2002)
[10] Y. S. Ju and K. E. Goodson. *Appl. Phys. Lett*. **74**, 3005 (1999)
[11] E. Pop, C. Chui, R. Dutton, S. Sinha, K. Goodson. *IEDM Tech. Digest*, 411-414 (2004)
[12] T. Yamane, N. Nagai, S. Katayama, M. Todoki. *J. Applied Physics* **91**, 9772 (2002)
[13] A. Smith, J. Hostetler, P. Norris. *Microscale Thermophysical Engineering* 4:51-60 (2000)
[14] R. Stevens, P. Norris, L. Zhigilei. *Mech. Eng. Cong. and Expo. (IMECE)* Nov. (2004)
[15] A. Majumdar and P. Reddy. *Appl. Phys. Lett*. **84**, 4768 (2004)
[16] G. Chen. *Phys. Rev. Lett*. **86**, 2297 (2001)
[17] K. Etessam. et al. *Proc. of IEEE SEMI-THERM Symp*. (2005)
[18] E. Pop, R. Dutton, K. Goodson. *J. Appl. Phys*. **96**, 4998 (2004)
[19] E. Pop, R. Dutton, K. Goodson. *Appl. Phys. Lett*. **86**, 082101 (2005)
[20] S. Sinha, P. Schelling, S. Phillpot, K. Goodson *J. Appl. Physics* **97**, 023702 (2005)
[21] M. Artaki and P. Price. *J. Appl. Phys*. **65**, 1317 (1989)
[22] J. Menendez and M. Cardona. *Phys. Rev. B* **29**, 2051 (1984)
[23] C. Ulrich, E. Anastassakis, K. Syassen, A. Debernardi, M. Cardona. *Phys. Rev. Lett*. **78**, 1283 (1997)
[24] A. Debernardi, S. Baroni, E. Molinari. *Phys. Rev. Lett*. **75**, 1819 (1995)
[25] J. Lai and A. Majumdar. *J. Appl. Phys*. **79**, 7353 (1996)
[26] S. Sinha, E. Pop, K. Goodson. *Mech. Eng. Cong. and Expo. (IMECE)* Nov. (2004)
[27] J. Rowlette, E. Pop, S. Sinha, M. Panzer, K. Goodson. *Proc. of IEEE ICCAD* (2005)

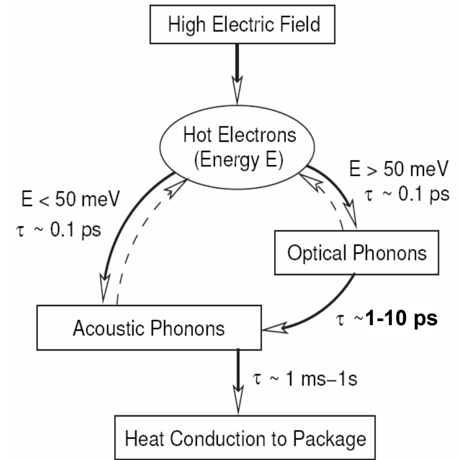
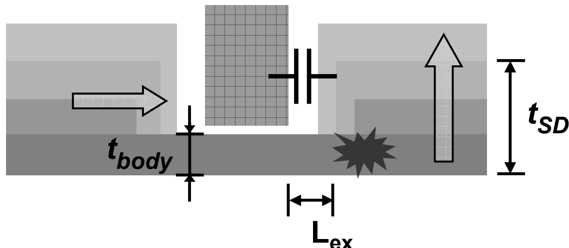


Figure 1. Time scales for various energy transfer processes between the electrons and lattice phonons in silicon. The most appropriate optical phonon lifetime values to use in simulating nanometer scale devices under typical operating conditions remains to be fully answered. Ref. 20 offers insight into this question in the framework of molecular dynamics simulations.



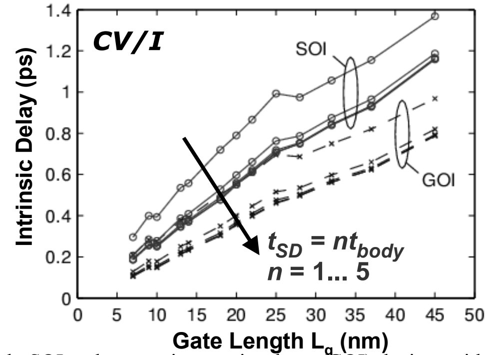


Figure 2. Electro-thermally self-consistent comparative study of "well-behaved" ultra-thin body SOI and germanium-on-insulator (GOI) devices with 18 nm channel lengths [11]. A larger (e.g. elevated) source and drain design will alleviate S/D series resistance and heat dissipation problems, but eventually lead to an increase in parasitic capacitance which reduces the intrinsic speed gain.

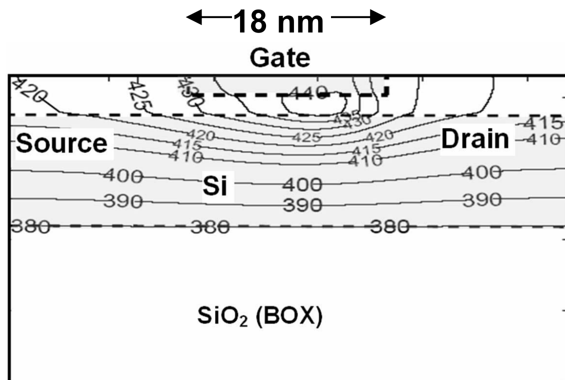


Figure 3. Effective temperature contours of an 18 nm fully depleted SOI FET calculated using a split-flux solution to the phonon BTE [26].

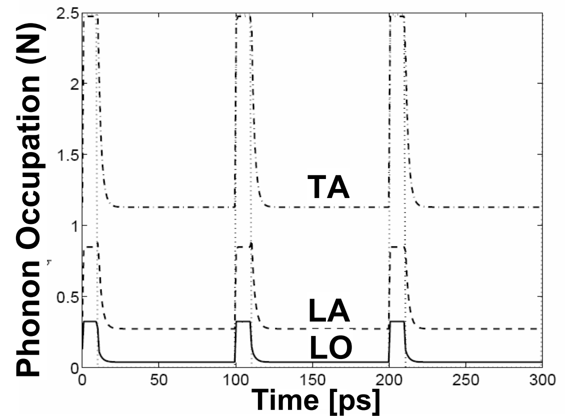


Figure 4. Transient phonon transport simulations with periodic 10 ps current pulses. The transverse acoustic (TA), longitudinal acoustic (LA) and longitudinal optical (LO) phonon branches are tracked [26].

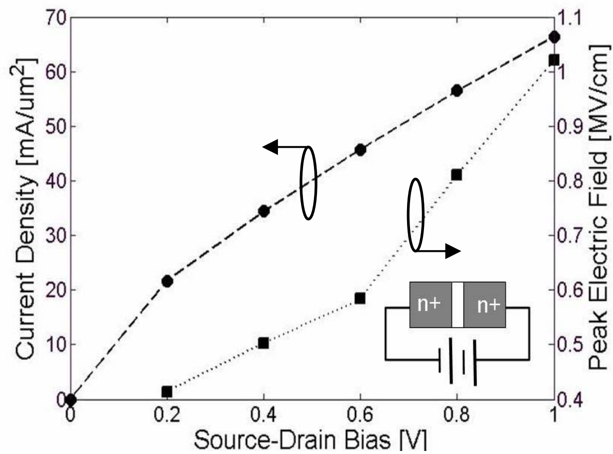


Figure 5. Current density (left axis) and peak electric field (right axis) vs. source-drain bias voltage (V) for a 1-D n+/n/n+ silicon diode modeled using self-consistent e-MC/split-flux p-BTE technique discussed in the text and in Ref. [27].

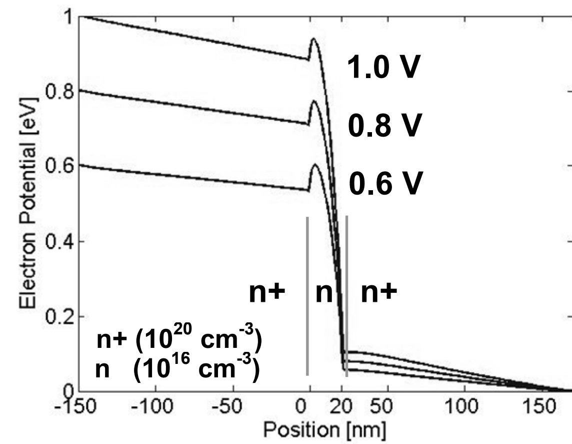


Figure 6. Electron potential energy vs. position within the n+/n/n+ device with electrical characteristics shown in Fig. 5 for three bias conditions. The device has three regions: two 150 nm n+ (10^{20} cm^{-3}) source and drain regions separated by a 20 nm n-type (10^{16} cm^{-3}) "channel". The doping is uniform in each region with a 1.25 nm/decade roll-off between regions.

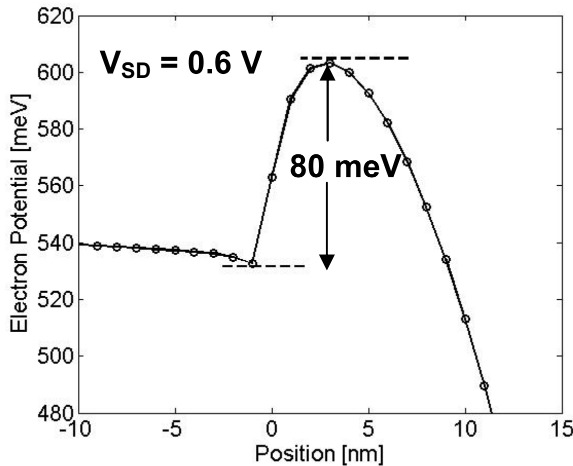


Figure 7. Zoomed-in view of the potential energy barrier seen by source electrons for the 0.6 V bias condition. Quantum transport effects (e.g. tunneling) were not included in this work, but will become more important as the barrier reduces in height and width.

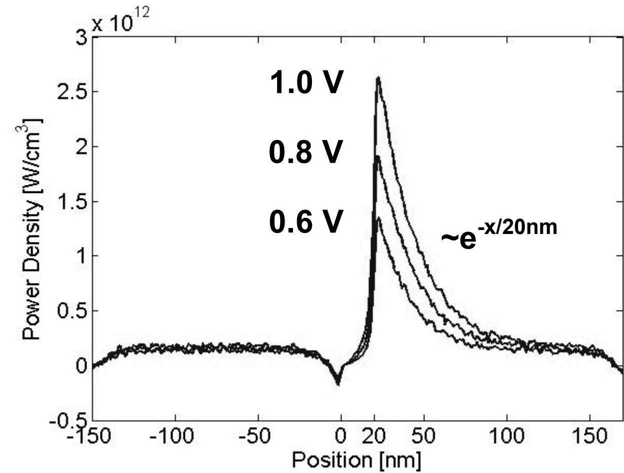


Figure 8. Volumetric power generation rate vs. position [5] within the 1-D n+/n/n+ device. The power generation at each grid point was computed by adding the appropriately weighted phonon emission events and subtracting the phonon absorption events and dividing by the simulation time interval.

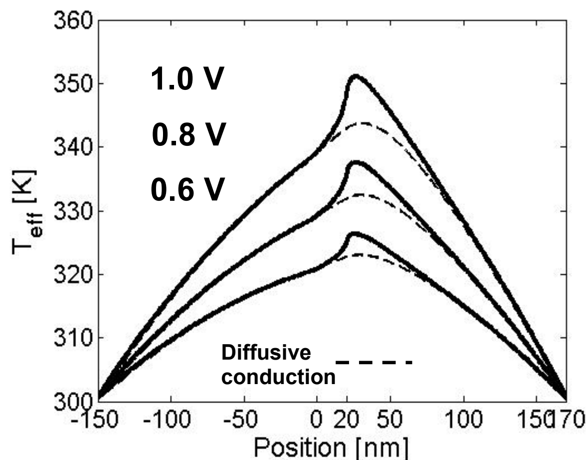


Figure 9. Effective lattice temperature (T_{eff}) computed using the self-consistent e-MC/split-flux p-BTE technique described in the text [27].

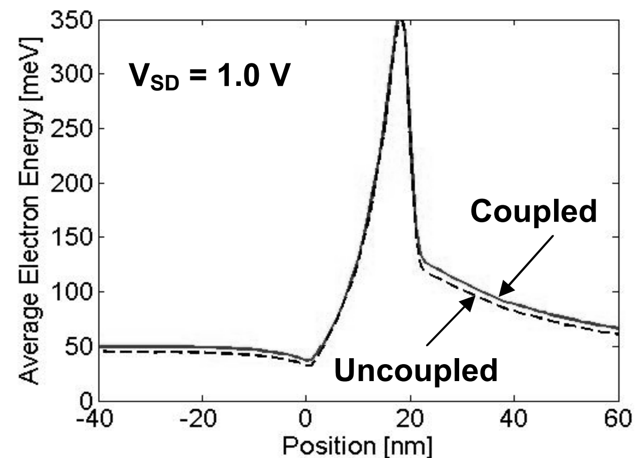


Figure 10. Effect on average source injected electron energy using the self-consistent e-MC/split-flux p-BTE technique described in the text [27].