Heat Generation and Transport in SOI and GOI Devices

Eric Pop

Micro and Nanotechnology Laboratory, Dept. of Electrical and Computer Engineering
University of Illinois Urbana-Champaign, Urbana IL 61801, USA

This paper surveys recent progress in our understanding of heat generation and transport in nanoscale transistors. Monte Carlo simulations show that under quasi-ballistic transport conditions, most Joule heat is generated in the device drain. Measurements and modeling find the device thermal resistance scales inversely with device size, reaching well over 100 K/mW for sub-100 nm transistors. This trend is partly also driven by decreased thermal conductivity of ultra-thin films, as well as the increased role of boundary thermal resistance. Somewhat surprisingly, the analysis shows that well-behaved GOI transistors ought to outperform SOI from a *thermal* viewpoint as well, highlighting the importance of electro-thermal co-design at device length scales approaching 10 nm.

Introduction

Modern CMOS device designs favor confined geometry transistors, such as Silicon-on-Insulator (SOI), FinFET, tri- or surround-gate, and ultimately nanowire (or nanotube) designs (1, 2). This trend is driven by the need for better electrostatic control of the device active region, and the need to lower parasitic capacitance. Another trend is to incorporate higher mobility materials (like germanium) into the active device channel.

The drive towards such device geometries has had several negative implications from a thermal point of view. The most commonly used electrical insulator (SiO_2) is also a very good thermal insulator (100x less thermally conducting than silicon). This has led to well known observations of severe steady-state self-heating in SOI transistors (3-6). Germanium itself has a bulk thermal conductivity only 40% that of silicon, while the thermal conductivity in very thin layers or nanowires is reduced even further, for both materials (1, 7, 8). Thin active layers and narrow device bodies confine the heat generation and dissipation regions, leading to higher local power densities, while a larger surface-to-volume ratio increases the role of interface thermal resistance between materials, impeding heat flow.

From a scaling point of view, it is also interesting to note that device operating voltages have been scaling linearly (or sometimes not at all) with the device dimensions (9), whereas the device volume and surface area scale cubically and quadratically with the linear dimension. As mentioned above, this leads to higher power densities *per device*, whose effects are compounded by the higher thermal resistances involved. This may lead, at the very least, to devices whose steady-state (I-V) characteristics suffer from significant self-heating and cannot be used to understand their dynamic operation. Or, at worst, to devices whose reliability, performance and leakage is compromised both in dynamic (digital) and steady-state operation, and whose thermal drawbacks may negate most advantages obtained from the advanced electro-static design.

Heat Generation in Nanoscale Devices

Heat generation takes an interesting shape in nanoscale devices, under non-equilibrium transport conditions. Monte Carlo (MC) analysis shows that due to quasi-ballistic transport in the active device region (channel < 50 nm), most Joule heat is generated in the drain region and near the contact (1, 10). This asymmetric heat generation is in contrast with lumped analyses of this problem which assumed uniform heat generation in the channel (5), or drift-diffusion (DD) based finite element modeling which places the heat generation region in sync with the local electric field (11).

The standard drift-diffusion approach calculates the heat generation as a dot product of the local current density and electric field ($H = \boldsymbol{J}\cdot\boldsymbol{E}$), whereas the Monte Carlo method implicitly accounts for all phonon generation and absorption events as electrons drift and scatter inside the device (10, 12). The former tends to overestimate the peak heat generation rate and generally predicts a narrower heat generation region, which follows the shape of the electric field. Figure 1 illustrates the heat generation profile computed along the 20 nm channel of a quasi-ballistic silicon device. The differences between the drift-diffusion and Monte Carlo computations are clearly evident at such short length scales (10). The Monte Carlo method yields a broader heat generation domain extending inside the device drain, and chiefly limited by the electron-phonon scattering rate there.

Non-Equilibrium Heat Generation

The Monte Carlo approach also shows that heat generation in silicon is not evenly divided among phonon modes, but that, rather, the acoustic phonon modes receive approximately one third and optical phonons the remaining two thirds of the Joule power. More specifically, the longitudinal optical (LO) *g*-type phonon receives approximately 60% of the total energy dissipation (10). The optical phonons have group velocities of 1000 m/s or less, and are thus much slower than the (zone center) acoustic phonons typically responsible for heat transport in silicon (group velocity 5000-9000 m/s). This non-equilibrium phonon generation has energy transfer bottleneck implications (10, 13). In other words, a significant non-equilibrium OP population may build up, particularly for

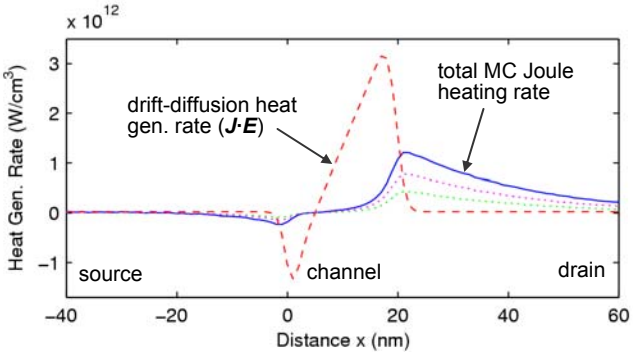

Figure 1: Heat generation in a n+/n/n+ quasi-ballistic silicon device with channel length L = 20 nm. The source and drain are doped to 10^{20} cm^{-3}, the applied voltage is 0.6 V. Unlike the classical (drift-diffusion) result, the Monte Carlo (MC) simulation shows that heat is dissipated far into the device drain. The dotted lines represent the optical phonon (upper) and acoustic phonon (lower) heat generation profiles from the MC result.

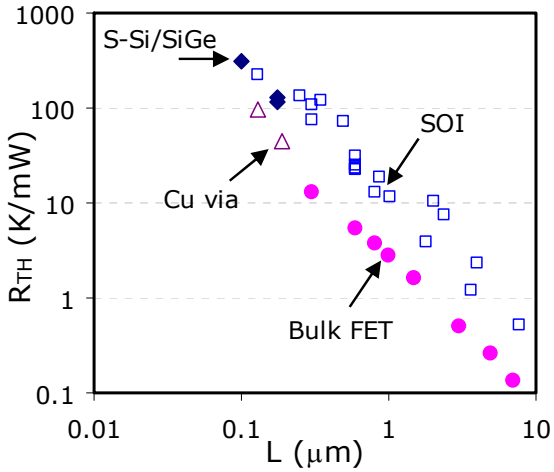

Figure 2. Thermal resistance data reported in the literature over nearly two decades of research across many bulk silicon FET (5, 14) and SOI device technologies (3-5, 15-18). The thermal resistances of single Cu vias and strained Si (on SiGe buffer) transistors (19, 20) are included for comparison. The *x*-axis represents the FET gate lengths, and the via diameters. The raw numbers have been normalized to a $W/L = 4$ device aspect ratio.

the *g*-type longitudinal optical (LO) mode. The generation rates for the other phonon modes are either smaller or their density of states (DOS) is larger (the DOS is proportional to the square of the phonon wave vector, which is largest at the edge of the Brillouin zone) and non-equilibrium effects are less significant. Assuming a 10 ps phonon lifetime, the occupation number of the *g*-type LO phonon would exceed $N_{LO} > 0.1$ and become comparable to unity for power densities greater than 10^{12} W/cm^3 (10). Such power densities are attainable in the drain of 20 nm (or shorter) channel length devices at operating voltages from the current ITRS guidelines (Fig. 1). Non-equilibrium phonon populations will increase electron scattering in the drain, leading (at the very least) to a magnification of the drain series resistance, and decreased device reliability.

Heat Transport in Nanoscale Devices

Device Thermal Resistance

Heat transport from a *lumped* electronic device can be quantified by measuring its thermal resistance (R_{th}) to the environment. This approach yields an average temperature rise of the lumped element as $\Delta T = PR_{th}$ where P is the dissipated power. In practice, the thermal resistance has been measured through noise thermometry (15), gate electrode electrical resistance thermometry (5, 14), pulsed I-V measurements (16, 19) or an AC conductance method (3, 4, 17, 18). Figure 2 presents a summary of this experimental data produced in the literature over the past sixteen years, covering a wide range of device dimensions and technologies. A clear trend emerges, showing that device thermal resistance increases as a power law of the (reduction in) device dimensions, and is reaching values well above 100 W/mK for sub-100 nm gate length devices. The simplest model illustrating the inverse proportionality of the device thermal resistance with its dimensions is $R_{th} = 1/(2kD)$ for the thermal spreading resistance of a heated disk (diameter D) on a semi-infinite plane with thermal conductivity k (21). This can be extended to the case of a

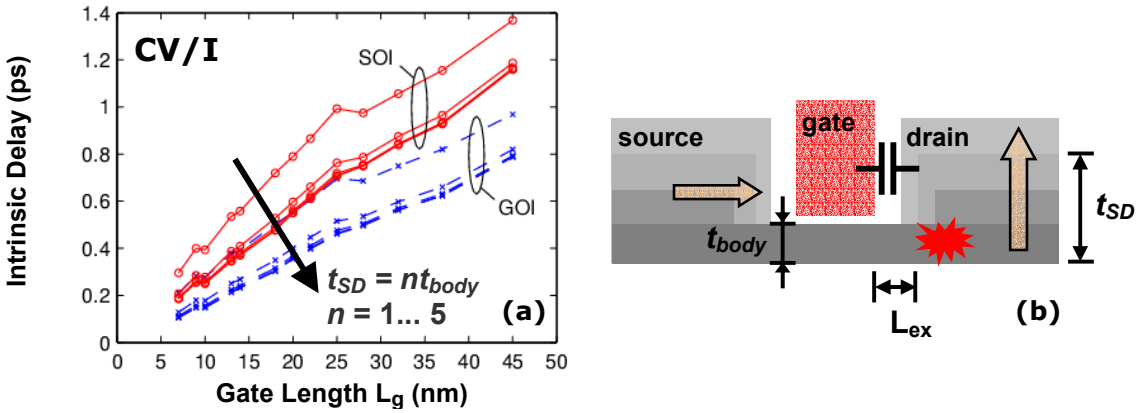

Figure 3. Electro-thermally self-consistent modeling of "well-behaved" SOI and GOI devices near the limits of scaling (26). A larger (e.g. elevated) source and drain design will alleviate heat dissipation problems, but eventually lead to an increase in parasitic capacitance which reduces the intrinsic speed gain. The heat dissipation was assumed to be entirely in the drain, as suggested by Monte Carlo simulations.

rectangular heat source (width W, length L) by replacing $D \approx (LW)^{1/2}$ and including additional 3-D heat spreading shape factors (14, 22). Several online tools are also available for quick, web-based spreading resistance calculations for various shapes and substrates (23). Many other models of varying sophistication have been published, all of which incorporate various inverse length and width dependencies (5, 14, 24-26). Naturally, the choice of such a model in practice depends on its complexity, and on the specific geometry of the device (for instance FinFET vs. bulk FET, to choose two fairly disparate cases).

Devices with Thermally Insulating Substrates

As expected, the thermal resistance of transistors built on thermally insulating substrates, like SOI and strained-Si on SiGe, can be particularly high, as shown in Fig. 2. For these devices, steady-state *I-V* measurements suffer from significant self-heating, and proper characterization must be done with the pulsed (16) or AC conductance methods (3). The trends in Fig. 2 suggest such characterization may need to be employed for near-10 nm *bulk* silicon FETs as well, and most certainly for surround-gate or FinFET type devices, although specific data are currently lacking.

Germanium-on-insulator (GOI) devices with high-κ dielectrics are very promising from an electrical point of view, but may also exhibit self-heating effects because the thermal conductivity of germanium is 60% lower than that of silicon. This would add to the thermal challenge already introduced by the buried insulator. Since data on well-behaved GOI devices is not yet available, this is an area where theoretical investigations are key at the moment. Such preliminary estimates (Fig. 3) suggest the high thermal resistance disadvantages of GOI may be mitigated by lower power dissipation (same current obtained at lower voltage), and also in part by a weaker temperature dependence of the germanium mobility (26). In other words, well-behaved ultra-thin body GOI devices are expected to maintain a performance advantage over similar SOI devices (27), assuming an adequate fabrication technique could be perfected.

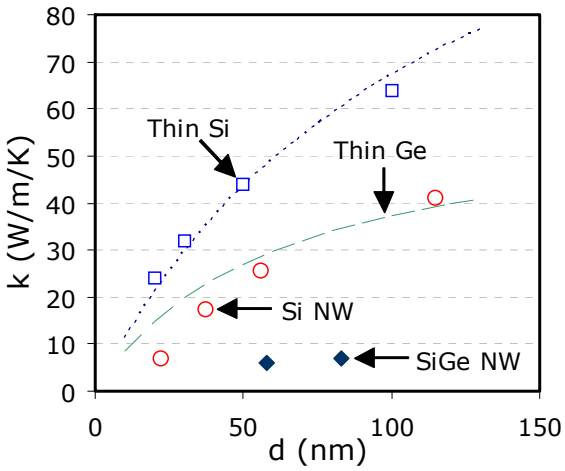

Figure 4. Thermal conductivity data along thin SOI layers, Si and SiGe superlattice nanowires (7, 8). The *x*-axis is the layer thickness or wire diameter, respectively. The dashed lines are simple models for Si and Ge (26). A significant decrease from the bulk thermal conductivity of Si (~140 W/m/K) and Ge (~60 W/m/K) is noted.

Reduced Thermal Conductivity of Thin Films and Nanowires

In addition to the thermally insulating substrate and boundary thermal resistance (1), thin-body or nanowire (NW) devices also suffer from reduced thermal conductivity owed to increased phonon-surface scattering in the active device region. Recently available data on such thin films and nanowires at room temperature is summarized in Fig. 4. Such data does not yet exist for thin Ge layers, thus a theoretical estimate (dash-dot line) is provided instead. Nevertheless, the most notable feature is the significant decrease in thermal conductivity (up to an order of magnitude) from bulk values for both Si and Ge. However, the decrease in Ge thin film thermal conductivity is expected to be proportionally less severe owing to the shorter phonon mean free path of this material (26). The lowest thermal conductivity is found along Si/Ge superlattice nanowires, where phonon conduction is particularly suppressed by scattering off the additional interfaces.

Transient Heat Conduction

The thermal resistance models mentioned above are sufficient for evaluating the *steady-state* behavior of semiconductor devices, i.e. relevant during *I-V* characterization, analog operation, or to estimate the temperature rise owed to device leakage. However, an understanding of *transient* heat conduction is necessary for short duration pulsed operation, such as during digital switching or electro-static discharge (ESD) events.

To first order, the temperature rise of a pulse-heated volume can be obtained from the energy of the heating pulse (E) and the heat capacity of the volume being heated (C): $E = Pt \approx C\Delta T$ where t is the pulse duration and P is its power. During digital operation, the duration of an inverter switching event is approximately $t \approx$ 50-100 ps, which is significantly faster than the thermal time constant of modern devices, $\tau \approx$ 50-100 ns (28). This is the "adiabatic limit," where the device is essentially thermally decoupled from its environment. In other words, while the device is ON there is little "smearing" of the heated volume outside the area where the actual heating takes place, i.e. the drain of the transistor. Any additional temperature spreading extends approximately $d \approx (\alpha t)^{1/2}$ outside the

immediately heated volume (29), where α is the heat diffusion coefficient. This distance is of the order 10 nm into silicon or germanium, and 1 nm into immediately adjacent SiO_2 layers (like the top passivation layer, or the buried oxide below SOI). Typical estimates of the temperature rise during digital switching have shown this dynamic value does not exceed a few degrees (e.g., 5 K) for sub-micron device technologies (28, 30). However, the exact value for devices in the 10 nm range is unknown, and will be highly dependent on the ultimate choice of device geometry and materials.

For longer time scales, comparable to or larger than the device thermal time constants, several models have been proposed (22, 25, 31, 32). These bridge the time range from the adiabatic limit to the steady-state operation of a device, and are typically based on a Green's functions solution of the heat diffusion equation. This approach is faster and offers more physical insight than solutions based on finite-element (FE) solvers. The disadvantage of such methods vs. the FE approach is their applicability to only a limited range of geometries, like heated sphere, infinite cylinder or rectangular parallelepiped, not taking into account the full geometry and diverse materials making up a modern semiconductor device.

Conclusions

This paper summarizes recent advances in our understanding of heat generation and transport in sub-100 nm transistors. Particular attention is given to non-equilibrium effects, of importance in small thin-film silicon- (or germanium-) on-insulator (SOI/GOI) transistors. Heat generation predominantly occurs in the drain of such quasi-ballistic devices. The lumped thermal resistance of nanoscale transistors can easily surpass 100 K/mW for SOI, FinFET, as well as (smaller) bulk transistors. The thermal conductivity of thin films and nanowires is reduced by up to an order of magnitude from the bulk values in silicon and germanium, partly contributing to the increased thermal resistance of small devices. The electro-thermal co-design of nanoscale transistors is seen as an increasingly important area of research.

Acknowledgments

I acknowledge valuable discussions with Sanjiv Sinha and Kenneth Goodson. Support for some of this work was provided by the Semiconductor Research Corporation (SRC).

References

1. E. Pop, S. Sinha and K. E. Goodson, *Proc. IEEE*, **94**, 1587 (2006).
2. H.-S. P. Wong, *IBM J. Res. Dev.*, **46**, 133 (2002).
3. W. Jin, W. Liu, S. K. H. Fung, P. C. H. Chan and C. Hu, *IEEE Trans. Electron Devices*, **48**, 730 (2001).
4. B. Tenbroek, M. S. L. Lee, W. Redman-White, R. J. T. Bunyan and M. J. Uren, *IEEE Trans. Electron Devices*, **43**, 2240 (1996).
5. L. T. Su, J. E. Chung, A. D. A., K. E. Goodson and M. I. Flik, *IEEE Trans. Electron Devices*, **41**, 69 (1994).
6. K. E. Goodson and M. I. Flik, *IEEE Trans. Components, Hybrids, Manufacturing Technol.*, **15**, 715 (1992).
7. D. Li, Y. Wu, P. Kim, L. Shi, P. Yang and A. Majumdar, *Appl. Phys. Lett.*, **83**, 2934 (2003).

8. W. Liu and M. Asheghi, *J. Appl. Phys.*, **98**, 123523 (2005).
9. International Technology Roadmap for Semiconductors (ITRS), in http://public.itrs.net.
10. E. Pop, J. Rowlette, R. W. Dutton and K. E. Goodson, in *Intl. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, p. 307, Tokyo, Japan (2005).
11. G. K. Wachutka, *IEEE Transactions on Electron Devices*, **9**, 1141 (1990).
12. T. Sadi, R. Kelsall and N. Pilgrim, *IEEE Trans. Electron Devices*, **53**, 1768 (2006).
13. M. Artaki and P. J. Price, *J. Appl. Phys.*, **65**, 1317 (1989).
14. P. G. Mautry and J. Trager, in *Intl. Conf. on Microelectronic Test Struct.*, p. 221 (1990).
15. R. J. T. Bunyan, M. J. Uren, J. C. Alderman and W. Eccleston, *IEEE Electron Device Letters*, **13**, 279 (1992).
16. K. A. Jenkins and J. Y.-C. Sun, *IEEE Electron Device Letters*, **16**, 145 (1995).
17. T.-Y. Lee and R. M. Fox, in *IEEE Intl. SOI Conference*, p. 78 (1995).
18. M. Reyboz, R. Daviot, O. Rozeau, P. Martin and M. Paccaud, in *IEEE Intl. SOI Conference*, p. 159 (2004).
19. K. A. Jenkins and K. Rim, *IEEE Electron Device Letters*, **23**, 360 (2002).
20. G. Nicholas, T. J. Grasby, E. H. C. Parker, T. E. Whall and T. Skotnicki, *IEEE Electron Device Letters*, **26**, 684 (2005).
21. M. M. Yovanovich, J. R. Culham and P. Teertstra, *IEEE Trans. Components, Packaging and Manuf. Tech.*, **21**, 168 (1998).
22. R. C. Joy and E. S. Schlig, *IEEE Trans. Electron Devices*, **17**, 386 (1970).
23. Micro Heat Transfer Lab (U. Waterloo), in http://www.mhtl.uwaterloo.ca/RScalculators.html.
24. A. M. Darwish, A. J. Bayba and H. A. Hung, *IEEE Trans. Microwave Theory and Tech.*, **53**, 306 (2005).
25. N. Rinaldi, *IEEE Trans. Electron Devices*, **48**, 2796 (2001).
26. E. Pop, C. O. Chui, S. Sinha, K. E. Goodson and R. W. Dutton, in *IEEE Intl. Electron Devices Mtg. (IEDM)*, p. 411, San Francisco, CA (2004).
27. X. An, R. Huang, X. Zhang and W. Yangyuan, *Semicond. Sci. Technol.*, **20**, 1034 (2005).
28. K. A. Jenkins and R. L. Franch, in *IEEE Intl. SOI Conference*, p. 161 (2003).
29. K. Banerjee, A. Amerasekera, N. Cheung and C. Hu, *IEEE Electron Device Letters*, **18**, 405 (1997).
30. B. Tenbroek, M. S. L. Lee, W. Redman-White, C. F. Edwards, R. J. T. Bunyan and M. J. Uren, in *Proc. IEEE Intl. SOI Conf.*, p. 156 (1997).
31. V. M. Dwyer, A. J. Franklin and D. S. Campbell, *Solid-State Electronics*, **33**, 553 (1990).
32. Y. J. Min, A. L. Palisoc and C. C. Lee, *IEEE Trans. Components, Hybrids, Manufacturing Technol.*, **13**, 980 (1990).