

Joule Heating under Quasi-Ballistic Transport Conditions in Bulk and Strained Silicon Devices

Eric Pop[†], Jeremy A. Rowlette[†], Robert W. Dutton[†] and Kenneth E. Goodson^{*}

Dept. of [†]Electrical and ^{*}Mechanical Engineering, Stanford University, U.S.A

Contact: epop@stanford.edu, Bldg 530 Room 224, Stanford CA 94305-3030, U.S.A., Tel. 650-387-3501

Abstract – We use Monte Carlo simulations to examine self-heating in ultra-short silicon devices when quasi-ballistic transport conditions dominate. The generated phonon spectrum in strained silicon is found to be different from bulk silicon at low electric fields, but essentially the same under high fields. Joule heat dissipation in ultra-short devices occurs almost entirely in their drain region, since transport across the channel is quasi-ballistic. The results of this work can be used to gauge the electro-thermal performance of ultra-scaled device geometries.

I. INTRODUCTION

As the dimensions and voltage of semiconductor devices are down-scaled linearly, the volume available for heat dissipation decreases cubically. Such trends can lead to significant power density issues within nanoscale semiconductor devices. The resulting consequences are relevant both within circuit operation and in the context of device ESD (electro-static discharge) behavior. In addition, incorporating materials with lower thermal conductivities than silicon (Table I) is only serving to accelerate such trends, while the thermal boundary resistance between materials also begins to play a role in device designs with large surface-area-to-volume ratio [1]. Confined device geometries (i.e. FinFET, ultra-thin body SOI, surround gate FETs and nanowire devices) will further limit the heat dissipation volume, leading to increased temperatures and the associated consequences on performance (lowered drive current, decreased reliability, weakened ESD immunity).

Electro-thermal simulations for semiconductor devices have typically taken one of two approaches. At the circuit level the thermal pathways of a device are reduced to an equivalent thermal resistance which is used to compute an average temperature rise given the dissipated power [2,3]. At the device level the heat generation rate is typically simulated using the dot product of the current density and the electric field across the simulation mesh, and the temperature profile is computed through a solution of the heat diffusion equation (Fourier Law) [4]. In reality, the physical mechanism through which self-heating occurs is that of electron scattering with phonons, and thus only a simulation approach which deliberately incorporates all such scattering events will capture the full microscopic, detailed picture. Consequently, in this work we study heat generation using Monte Carlo simulations. The heat generation rate is computed as a sum of all phonon emission events minus all phonon absorption events. Heat generation in ultra-short devices

is found to occur almost entirely in the drain, beyond the peak electric field region when transport across the channel is quasi-ballistic. The heat generation region (“hot spot”) extends deep into the drain, because the energetic electrons take a relatively long time and space to fully relax their energy to the lattice. We extract a compact analytic model for the size and intensity of the drain hot spot, and the results of this work can be used to gauge the electro-thermal performance of ultra-scaled device geometries.

II. MONTE CARLO IMPLEMENTATION

The extensive details of our Monte Carlo (MC) implementation have been described elsewhere [5]. The electron energy bands are modeled with the analytic nonparabolic band approximation, including the six ellipsoidal conduction X valleys of silicon:

$$E(1 + \alpha E) = \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z} \right) \quad (1)$$

where E is the electron energy, $\alpha = 0.5 \text{ eV}^{-1}$ is the nonparabolicity parameter and (k_i, m_i) represent the electron wave vector and effective mass along the i -th crystal direction. This is a good approximation for device voltages near or below the silicon band gap (1.1 eV), such as those of future nano-technologies, and is significantly faster than the full-band MC method [6]. In addition, the low electron energy range means that both impact ionization (which is controlled by the band gap) and interband scattering with higher energy bands can be safely neglected (the X - L band separation in silicon is about 1.2 eV). Inelastic scattering with six types of

Table I: Thermal conductivities of a few materials used in device fabrication. Phonons are thermal carriers in all materials listed except silicides (metals) whose heat conduction is by electrons. Phonon boundary scattering significantly reduces the thermal conductivity of 10 nm thin silicon films, while alloy scattering reduces that of $\text{Si}_{1-x}\text{Ge}_x$ layers.

Material	Thermal Conductivity ($\text{Wm}^{-1}\text{K}^{-1}$)
Si (bulk)	148
Ge (bulk)	60
Silicides	40
Si (10 nm)	13
$\text{Si}_{0.7}\text{Ge}_{0.3}$	8
SiO_2	1.4

intervalley phonons is incorporated. Three of these are of g -type, assisting electron scattering between valleys on the same k -axis, and three are of f -type, for scattering between valleys on orthogonal axes [5]. Care is also taken to treat intravalley scattering (within the same conduction valley) with low-energy acoustic phonons inelastically, to properly account for the energy dissipation. All phonon dispersion modes are modeled with quadratic, analytic approximations, which properly capture the phonon speed and density of states across the Brillouin zone [7]:

$$\omega(q) = \omega_0 + v_s q + c q^2 \quad (2)$$

where $\omega(q)$ is the phonon frequency, q is the phonon wave vector and the other parameters are adjusted to mimic the experimental dispersion relationship [7]. Ionized impurity scattering is included with an efficient model which reduces the number of time-consuming small angle scattering events [8]. In the context of device simulation, the electric fields are self-consistently updated along with the charge transport through solutions of the Poisson equation.

Electron transport in strained silicon is incorporated by accounting for the split in the conduction band introduced by the biaxial strain in silicon grown on a $\text{Si}_{1-x}\text{Ge}_x$ substrate. The strain lifts four of the six conduction valleys by $\Delta E \approx 0.67x$. The in-plane conductivity effective mass of the two lower valleys is the lighter transverse mass $m_t = 0.19m_0$ of silicon (m_0 being the mass of the free electron). The strain-induced energy splitting reduces f -type intervalley scattering, since the maximum available f -type phonon energy in silicon is only about 50 meV. For $x > 0.15$ the energy splitting is large enough (> 0.1 eV) to essentially suppress f -type scattering between the lower and upper valleys, and the strained silicon mobility enhancement is dominated by conduction via the two lower valleys with the lighter transverse mass.

III. JOULE HEATING IN BULK AND STRAINED SI

Fig. 1 shows the net energy dissipation rates in bulk and strained ($x = 0.3$) silicon with each branch of the phonon spectrum for various steady-state electric fields. Note the enhancement of g -type LO (longitudinal optical) phonon emission and the reduction in energy relaxation through f -type LA (longitudinal acoustic) and TO (transverse optical) phonons in strained silicon at low electric fields. At fields larger than 30-40 kV/cm the average electron energy (in excess of 0.25 eV) becomes comparable to or larger than the 0.2 eV strain-induced band splitting (for Ge fraction $x = 0.3$), and the phonon generation rates are the same in bulk and strained silicon. This is the case when electrons have enough energy to emit phonons across the entire spectrum despite the band splitting. For both strained and bulk silicon the total Joule heat dissipated to the lattice, over all four phonon branches, conveniently sums up to

$$Q''' = \mathbf{J} \cdot \mathbf{E} = env|\mathbf{E}| \quad (3)$$

as anticipated, where J is the current density, E the electric field, e the elementary charge, n the electron density and v the average drift velocity (as calculated, e.g., in Ref. [5]).

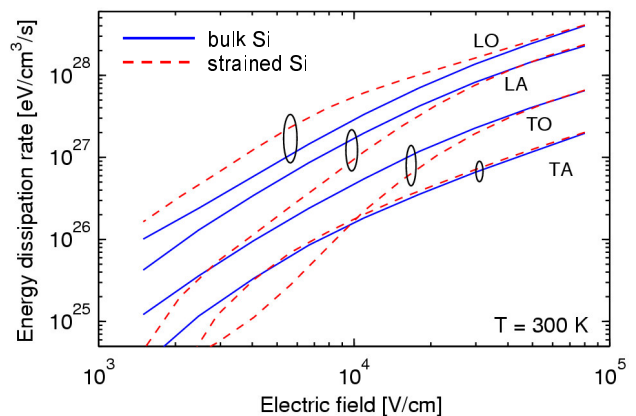


Fig. 1. Computed heat generation rates vs. electric field with each phonon mode in bulk and strained ($x = 0.3$) Si.

Even a quick examination of Fig. 1 reveals that unlike it is usually assumed [9], *not* all electron energy is dissipated into the optical phonon modes. Table II summarizes the relative proportion of energy dissipated into each phonon mode in bulk and strained silicon, at low and high electric fields. This is of importance because the generated phonons have widely varying contribution to heat transport: optical phonons make virtually no contribution to the thermal conductivity of silicon, which is dominated by acoustic phonon transport [10]. Emission of LO modes makes up nearly 90 percent of the energy dissipated in strained silicon at low fields, but only about half the heat generation in bulk silicon at most practical electric fields. Intervalley f -type LA phonon emission accounts for about one third, and TO emission for one tenth of the heat generation rate in bulk silicon. Finally, transverse acoustic (TA) phonons have relatively low energy and low generation rates, and they are consequently the weakest energy relaxation (heat generation) mechanism.

We note that both the f -type TA and TO phonons are generated at the edge of the Brillouin zone (BZ), where their group velocity is nearly zero, making them essentially stationary [7]. However, their density of states is large at the edge of the BZ, implying that their occupation is likely to be very small. Hence, an overpopulation of zone-edge TA and TO modes in a silicon device is unlikely. The most likely hot phonon effects in silicon devices would arise within modes which have strong generation rates and low density of states. Given our findings above, the g -type LO modes (generated

Table II: Approximate percentages of the total Joule heating rate dissipated with every branch of the phonon spectrum, based on Fig. 1. Note each column adds up to unity.

	Bulk (all fields) and high-field strained Si	Low-field strained Si
TA	< 0.03	0.02
LA	0.32	0.08
TO	0.09	< 0.01
LO	0.56	0.89

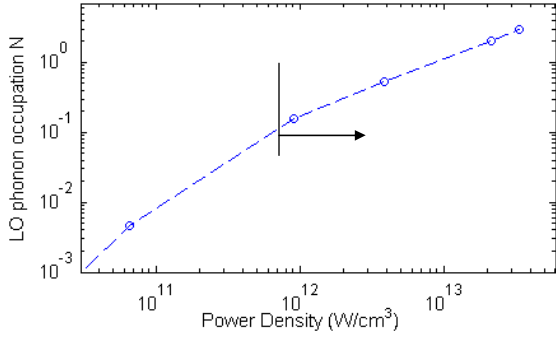


Fig. 2: Estimated occupation of the g -type LO phonon as a function of total dissipated power density. Hot phonon effects are expected when the occupation becomes comparable to unity, for total power densities $> 10^{12}$ W/cm³.

around 0.3 of the distance to the BZ edge [7]) are most likely to have high occupation numbers for large enough power dissipation densities. Given the LO phonon generation rates computed here, we can estimate the LO occupation as

$$N \approx (Q_{LO}\tau_{LO})/[\hbar\omega_{LO}g(\omega)\Delta\omega] \quad (4)$$

where $g(\omega)=[q^2/(2\pi^2)](dq/d\omega)$ is the phonon density of states and $\tau_{LO} \sim 10$ ps is the LO phonon lifetime [14]. We plot this estimate in Fig. 3 and find that hot LO phonon effects may be of importance for power dissipation densities greater than 10^{12} W/cm³. Such power densities may be reached within devices of 20 nm channel length and below (see Figs. 3c and 4, and discussion below). The resulting hot phonons will likely affect the device drain-side series resistance and reliability, but further work (currently in progress) must be done to properly investigate such concerns.

IV. JOULE HEATING IN QUASI-BALLISTIC DEVICES

We have also applied our MC method to evaluate heat generation in one-dimensional devices. Three $n^+/n/n^+$ devices are considered, with channel lengths $L = 500, 100$ and 20 nm (Fig. 3). The source and drain regions are assumed doped to $10^{18}, 10^{19}$ and 10^{20} cm⁻³, and the applied voltages are 2.5, 1.2 and 0.6 V respectively. The latter are roughly equivalent to operating voltages recommended by Technology Roadmap (ITRS) guidelines [11] for CMOS devices of similar channel lengths. The channel is assumed lightly doped at 10^{16} cm⁻³. Heat generation simulations using our MC code are compared to those computed using the commercial drift-diffusion simulator Medici, as shown in Fig. 3. Both approaches self-consistently solve the transport problem along with Poisson's equation.

As expected, the MC results are very similar to those of the drift-diffusion calculations for the longest simulated device with channel length (500 nm) much longer than the electron-phonon scattering length (5-10 nm), i.e. in the continuum limit (Fig. 3a). This provides a check on the accuracy of our MC simulation and enables a study of the conditions under which the drift-diffusion heat generation calculations break down. For the two shorter devices we find the MC heat generation region is significantly “displaced” from the

drift-diffusion predictions (Figs 3b and 3c). The error incurred in computing the geometrical center of the heat generation region with the drift-diffusion vs. the physically accurate MC approach is $\Delta L/L \approx 0.10, 0.38$ and 0.82 for the three device lengths. This can be qualitatively understood by noting that electrons gain most of their energy as they pass through the high field region, then they travel several scattering lengths until releasing this energy back to the lattice. Consequently, the hot spot extends remarkably far into the drain. Electrons can only exchange energy with phonons in discrete amounts of the phonon energy ($\hbar\omega$) on average every 5-10 nm, the inelastic scattering length. The total energy relaxation length in the drain is approximately

$$l \approx v\tau(E - 3k_B T / 2)/(\hbar\omega) \quad (5)$$

where τ is the inelastic scattering time (0.05-1 ps) and E is the average energy of electrons injected across the channel. By calculating the heat generation profile as a function of the applied voltage V_d (Fig. 4) we find $E \approx 0.4V_d$ (assuming

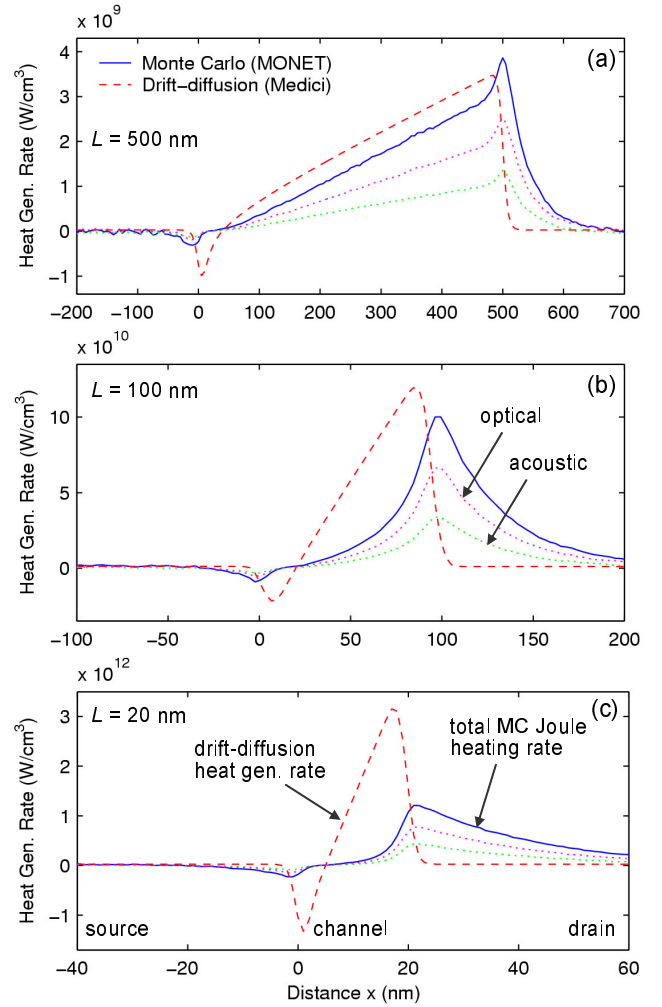


Fig. 3. Heat generation in $n^+/n/n^+$ devices with channel lengths $L = 500, 100$ and 20 nm. Solid lines are results of our MC simulations, dashed lines are from Medici. Dotted lines represent the optical (upper) and acoustic (lower) phonon generation rates given by the MC simulation.

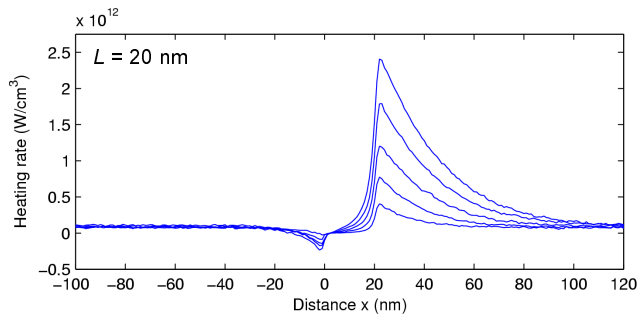


Fig. 4. Monte Carlo simulation results of the heat generation profile in a short channel (20 nm) ballistic diode with applied voltages $V_d = 0.2, 0.4, 0.6, 0.8$ and 1 V from bottom to top. The edges of the channel are at 0 and 20 nm.

the electron charge is $e = 1$), and we use $\hbar\omega \approx 50$ meV. The characteristic (exponential) decay length of the heat generation region in the drain is always approximately ~ 20 nm regardless of V_d (see Fig. 4). This can be understood since the velocity v scales as the square root of the kinetic energy E , while the inelastic scattering time τ scales as $1/\sqrt{E}$ because the phonon scattering rate ($1/\tau$) scales with \sqrt{E} from the density of states [5]. We note that the drain is a heterogeneous mixture of two populations, one being the “hot” electrons injected across the channel, the other made up of the many “cold” electrons already there due to the high doping. The cold electrons have an average energy of $3k_B T/2$ (thermal average) and they do not contribute to the net heat generation. The net heat generation in Fig. 4 is entirely caused by hot electrons injected almost ballistically across the channel. Finally, since most Joule heat is dissipated in the drain and *not* along the channel of quasi-ballistic devices, this suggests the design of the drain region and contacts will play a significant role for heat dissipation in ultra-short devices, especially those on insulating substrates (SOI, FinFET). Our findings are consistent with Ref. [12], implying that heat dissipation in mesoscopic devices occurs in or near the contacts rather than in the active device region.

Unlike in the drain, electrons in the source are near thermal equilibrium with the lattice temperature. However, Figs. 3 and 4 reveal a small negative heat generation region at the beginning of the channel. This is a thermoelectric (Peltier) effect due electron injection across the potential barrier from the source into the channel. Electrons diffusing against the energy barrier extract the energy required to move up the conduction band slope from the lattice, through net phonon absorption. This phenomenon has been exploited in the design of heterojunction laser diodes, where energy barriers introduced by band structure offsets can be optimized to provide internal thermoelectric cooling near the active laser region [13]. However, similar techniques may be difficult to implement in CMOS devices, where control over the barrier height is essential for their electrical behavior.

V. CONCLUSION

In this work we have analyzed Joule heating in bulk and strained silicon, and under quasi-ballistic transport conditions in ultra-short devices down to dimensions comparable

to the electron-phonon scattering length (5-10 nm). Our analysis reveals that LO phonons make up approximately half the power dissipated in silicon at high electric fields and that hot phonon effects may occur in devices with channel lengths < 20 nm. The physically accurate Monte Carlo (MC) method also finds that power is almost entirely dissipated in the drain of such short devices, when transport across the channel is quasi-ballistic. This suggests that electro-thermal optimization of the drain and contact region may play an important role for such advanced technologies. The results of this work can also be used for electro-thermal analysis of ultra-short device scaling [1], and as inputs for complex phonon transport studies [14].

ACKNOWLEDGMENT

We wish to thank Cristoph Jungemann, Umberto Ravaioli and Sanjiv Sinha for stimulating discussions. This work was supported by the Semiconductor Research Corporation (SRC) Task 1043.

REFERENCES

- [1] E. Pop, C. O. Chui, S. Sinha, R. W. Dutton and K. E. Goodson, “Electro-thermal comparison and performance optimization of thin-body SOI and GOI MOSFETs,” *Proc. IEDM*, p. 411, San Francisco CA (2004).
- [2] L. Su, J. Chung, D. Antoniadis, K. E. Goodson and M. I. Flik, “Measurement and modeling of self-heating in SOI nMOSFETs,” *IEEE TED* **41**, 69 (1994).
- [3] G. O. Workman, J. G. Fossum, S. Krishnan and M. M. Pelella, “Physical modeling of temperature dependences of SOI CMOS devices and circuits including self-heating,” *IEEE Trans. Electron Devices* **45**, 125 (1998).
- [4] G. K. Wachutka, “Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling,” *IEEE Trans. Electron Devices* **9**, 1141 (1990).
- [5] E. Pop, R. W. Dutton and K. E. Goodson, “Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion,” *J. Appl. Phys.* **96**, 4998 (2004).
- [6] H. Mizuno, K. Taniguchi and C. Hamaguchi, “Electron transport simulation in silicon including anisotropic phonon scattering rate,” *Phys. Rev. B* **48**, 1512 (1993).
- [7] E. Pop, R. W. Dutton and K. E. Goodson, “Monte Carlo simulation of Joule heating in bulk and strained silicon,” *Appl. Phys. Lett.* **86**, 82101 (2005).
- [8] H. Kosina, “A method to reduce small-angle scattering in Monte Carlo device analysis,” *IEEE Trans. Electron Devices*, **46**, 1196 (1999).
- [9] J. Lai and A. Majumdar, “Concurrent thermal and electrical modeling of sub-micrometer silicon devices,” *J. Appl. Phys.* **79**, 7353 (1996).
- [10] Y. S. Ju and K. E. Goodson, “Phonon scattering in silicon films with thickness of order 100 nm,” *Appl. Phys. Lett.* **74**, 3005 (1999).
- [11] International Technology Roadmap for Semiconductors (ITRS). Online at <http://public.itrs.net>
- [12] R. Lake and S. Datta, “Energy balance and heat exchange in mesoscopic systems,” *Phys. Rev. B* **46**, 4757 (1992).
- [13] K. P. Pipe, R. J. Ram and A. Shakouri, “Internal cooling in a semiconductor laser diode,” *IEEE Phot. Tech. Lett.* **14**, 453 (2002).
- [14] S. Sinha, E. Pop and K. E. Goodson, “A split-flux model for phonon transport near hotspots,” *Proc. IMECE*, Anaheim CA (2004) and *J. Heat Transfer* in press (2005).