ELSEVIER

Microelectronics Journal 37 (2006) 1148–1157

Microelectronics Journal

www.elsevier.com/locate/mejo

Thermal conduction in sub-100 nm transistors

S. Sinha*, K.E. Goodson

Thermosciences Division, Mechanical Engineering, Stanford University, Stanford, CA 94305-3030, USA

Received 27 December 2004; received in revised form 12 May 2005; accepted 22 July 2005
Available online 12 September 2005

Abstract

Heat conduction in integrated circuits spans length scales across several orders of magnitude: From the lattice spacing at a few Angstroms to the substrate thickness at hundreds of micrometers. The smaller length scale becomes increasingly important in devices with feature size well below 100 nm. This paper provides an overview of sub-continuum electro-thermal transport. We use the phonon Boltzmann transport equation to model heat conduction in the device and show that phonons emitted by hot electrons in the drain create a phonon hotspot. The resulting non-equilibrium leads to increased thermal resistance within the device. At the limits of scaling, the resistance is comparable to that due to the substrate and packaging.

PACS: 63.20.Kr; 66.70.+f; 44.10.+f

Keywords: Phonon; Heat; Conduction; Transistor

1. Introduction

The problem of increasing power density in logic circuits is a dominating concern today. Clock drivers and latches add up to approximately 70% of the total power dissipated in a microprocessor [1], and additionally have very high power densities. For example, a clock driver in a 0.18 μm technology can have a dynamic power consumption of about 49 mW over a 225 μm^2 area [2]. The usual definition of power density as the power dissipated per unit layout area ignores the spatial distribution of Joule heat within the transistor. In this paper, we discuss the problems associated with a 'second' power density: the power dissipated per unit volume inside the transistor itself. The peak power density inside a 90 nm bulk device [3] is estimated to be approximately 5 $W/\mu m^3$ from device simulations. This number can be as high as 65 $W/\mu m^3$ in an 18 nm ultrathinbody SOI device [4], as shown in Fig. 1. Even though the power dissipated per device decreases in future technology nodes, the volumetric power density continues to increase steadily. Consequently, the peak temperature inside a transistor is significantly higher than any average temperature [2] estimated by ignoring the granularity of heat generation.

The overall junction to ambient thermal resistance is approximately 0.5 K/W [5] for the current bulk-Si CMOS technology. However, the thermal pathway becomes increasingly resistive with the introduction of novel device geometries such as those shown in Fig. 2. The reason for this increase is the use of thermally resistive structures such as the silicon-on-insulator (SOI) or silicon–germanium on insulator. The improvement in electrostatics in these novel structures comes at the cost of reduced heat conduction from the device. Since, SOI substitutes a layer of bulk silicon by a layer of oxide whose thermal conductivity is a hundred times smaller, the thermal resistance of the device substrate increases by about 14% in an 18 nm technology node for example. The silicon thin-film constituting the device body also shows a significant reduction in thermal conductivity. This reduction is nearly 80% compared to the bulk value when the film thickness is 20 nm [6].

A more subtle contribution to the thermal resistance arises from sub-continuum electro-thermal phenomena inside the drain terminal of the transistor. The peak electric field in nanotransistors increases due to device scaling such that the heat source shrinks in size while

*Corresponding author.
E-mail address: sanjiv@stanfordalumni.org (S. Sinha).

0026-2692/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mejo.2005.07.015

increasing in intensity. Using a sub-continuum electro-thermal transport model, we show that this leads to an additional thermal resistance near the heat source in the drain. We find that this resistance scales in proportion to the peak power density and increases in future technology nodes. While the use of novel device geometries such as strained silicon on graded SiGe, ultra-thinbody SOI and FinFETs, is likely to lead to increased device temperatures compared to the bulk at the same power levels, the increase will not just be due to altered material properties but also due to sub-continuum conduction phenomena. The transport model discussed in this paper helps in understanding these phenomena and in predicting their impact on the performance of sub-100 nm transistors.

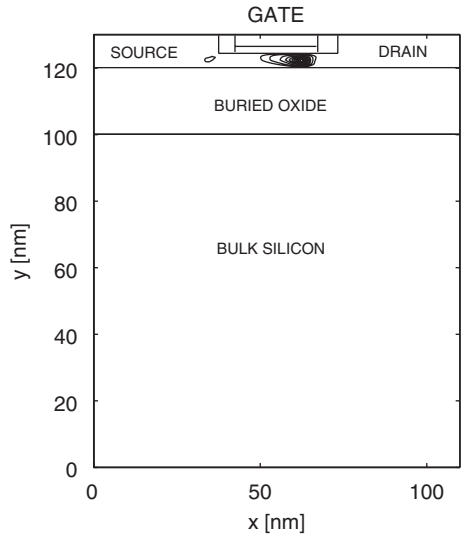

Fig. 1. The location of the heat source is shown in an ultra-thinbody SOI transistor operated at 1.2 V with a gate length of 18 nm [23]. The contours correspond to the spatial distribution of the heat generated inside the device and are equi-spaced at 0.5 W/m^3 with a peak value of 65 W/m^3 at the center.

2. Sub-continuum heat conduction

While a sub-continuum description of charge transport in terms of the Boltzmann transport equation (BTE) for electrons is quite common in device simulations, an analogous treatment of heat conduction is quite rare. However, current device gate lengths are well below the mean free path of room temperature energy carriers in silicon. The continuum heat diffusion equation is inadequate at describing transport at this length scale. A valid description at this scale is a semi-classical treatment of the quanta of lattice vibrations or phonons, which are the dominant energy carriers in silicon. Atoms located at the lattice sites of a dielectric crystal like silicon undergo small oscillations about their equilibrium positions at every temperature. The resulting atomic displacement field stores and transports energy in the crystal. Classically, this is described by a displacement vector obeying a linear, homogeneous wave equation of second order in space and time. Thus, the field may be described as superpositions of plane waves, with the frequency and wave vector obeying a non-linear dispersion relationship. These so-called normal modes of the crystal can have multiple polarizations depending on whether the displacement is perpendicular (transverse) or parallel (longitudinal) to the

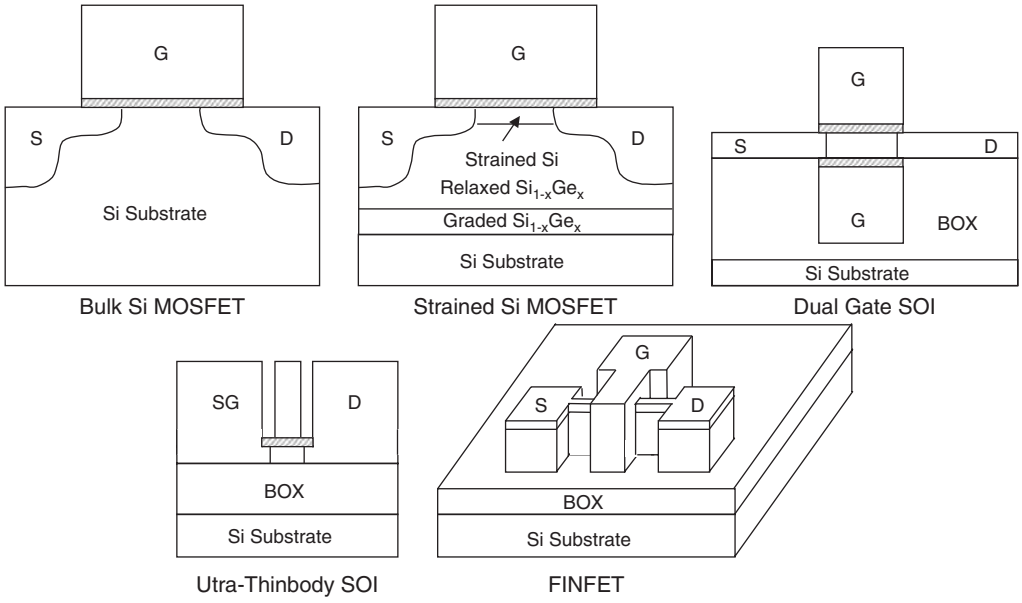

Fig. 2. The schematics show various FET device structures starting from the bulk silicon MOSFET on the top left. The locations of the source (S), drain (D) and gate (G) terminals are indicated as is location of the buried oxide (BOX) layer. The stained silicon FET uses a strained silicon layer as the channel for electron flow, which enhances the mobility of electrons. The use of dual gates and an ultra-thin film silicon channel enables better current control in short-channel devices. The FINFET is a three-dimensional structure with the gate surrounding a vertically oriented fin that serves as the channel.

wave vector. In the long wavelength limit, the longitudinal vibrations are identical to sound waves in the solid. Quantum mechanically, the atomic displacement field may be described either as an infinite number of distinguishable, quantized oscillators, or as a gas of indistinguishable particles called phonons. The particle description is particularly useful in treating interactions with other systems such as electrons. In thermal equilibrium, phonons are described by the Planck distribution, which applies to Bosons without rest mass and without conservation of particle number.

The dispersion relation that links the phonon frequency to the wave vector greatly affects phonon generation and transport in transistors. Since, a silicon crystal has two basis atoms per primitve cell, there are six degrees of freedom per primitive cell. This leads to six different branches of phonons, each describing a different relative motion of the atoms. Phonons in silicon are first classified as acoustic (A) or optical (O) depending on how the basis atoms within a primitive cell of silicon vibrate with respect to each other, whether in-phase or out-of-phase. The classification further separates optical and acoustic phonons into one longitudinal (L) and two transverse (T) polarizations per mode of vibration, leading to the six branches. We refer the reader to a standard text such as Ref. [7] for a more formal introduction to phonons.

The level of detail a heat conduction model must include depends on the length scale of the problem and the temperature. The phonon distribution obeys quantum statistics at low temperatures but approaches the classical limit as the lattice temperature approaches the Debye temperature, θ_D. Thus, quantum mechanical details that must be accounted for at very low temperatures become less important at higher temperatures. The length and time scale of the problem determines whether phonons may be modeled as particles or whether the wave nature of phonons is important. The important length scales for comparison are the phonon mean free path, Λ, the phonon wavelength, λ, and the lattice spacing, a_0. The important time scales are the phonon relaxation time, t, and the time period of a lattice wave, $1/v$, where v is the phonon frequency. The validity of various modeling regimes based on length scales is summarized in Fig. 3. The diffusion length is included for the sake of completeness. Typical silicon semiconductor device dimensions lie in between the phonon mean free path and the dominant phonon wavelength in room temperature silicon. As shown in the figure, semi-classical particle transport and/or classical atomistic models are applicable in this regime. When domain dimensions are comparable to the phonon wavelength and the temperature is small compared to the Debye temperature, a quantum mechanical model is necessary. Finally, the extent of the crystal can alter the phonon dispersion relation [8] and the density of phonon states. As dimensions of the crystal are reduced, surface effects become important and the assumption of an infinite crystal may not hold along one or more dimensions.

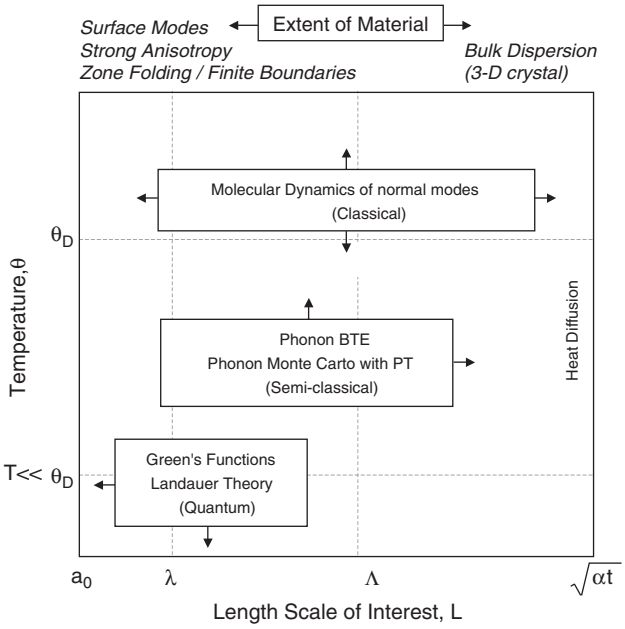

Fig. 3. Sub-continuum thermal modeling depends on the dimension of the system and the temperature. At dimensions comparable to the phonon wavelength and temperatures much smaller than the Debye temperature (bottom left), quantum mechanical nature is strongly manifest. At larger dimensions and room temperatures, a semi-classical approach is more pragmatic. PT in the figure stands for perturbation theory and a is the thermal diffusivity.

3. Electro-thermal transport physics

Charge transport in a semiconductor device is strongly coupled with thermal transport through electron–phonon interactions. The electrons in the channel gain energy from the applied electric field and subsequently lose it to the lattice while restoring thermodynamic equilibrium through scattering. Electrons may also absorb phonons, thereby gaining energy from the lattice. This occurs near the device source where phonon absorption enables the less energetic electrons to overcome the potential barrier. At the drain terminal 'hot' electrons shed energy gained from the field by emitting phonons. Electron–phonon scattering rules determine the spectral distribution of the emitted phonons. The evolution of the emitted phonons depends on phonon–phonon interactions, but affects electron transport through phonon reabsorption processes. We consider electron–phonon and phonon–phonon interactions below in greater detail.

3.1. Electron–phonon scattering

In n-type silicon, the electrons involved in transport are located at the bottom of the conduction band. The constant energy surface at the bottom of the conduction bands in silicon consists of six equivalent ellipsoids along the $\langle 1\,0\,0 \rangle$ directions. The conduction band minima occur at about 0.85 of the distance from the zone center to the zone edge along these directions [9]. Electron–phonon

scattering can be categorized as inter-valley and intra-valley depending on whether the scattering moves the electrons within a valley or from one valley to another in wave vector space. Intra-valley scattering in silicon is entirely due to acoustic phonons since, optical phonons are forbidden from energy and symmetry considerations [10]. Inter-valley scattering, however, involves mainly optical and zone-edge acoustic phonons. Energetic electrons tend to scatter more with optical phonons since such processes result in a higher energy loss. An electron must have energy in excess of about 51 meV, the minimum energy of optical phonons in silicon, in order to emit an optical phonon. Phonon modes that conserve crystal momentum in inter-valley processes have been calculated by Long [11]. Inter-valley scattering requires a large change in the wave vector of the electron. The two types of electron transitions, called g- and f-scattering, involve scattering between valleys along the same axis and along different axes, respectively. In the reduced zone scheme, both are umklapp processes as shown in Fig. 4.

The g-process requires a phonon with wave vector $0.3 \times (2\pi/a)\langle 0\,0\,1\rangle$. The f-process requires a phonon with wave vector that is 11° off $\langle 1\,0\,0\rangle$. Further, symmetry considerations based on a zeroth order expansion dictate that only LO phonons are allowed in a g-process and LA and TO are allowed in an f-process [12]. However, it has been shown by a first order expansion that low frequency TA and LA phonons are also involved in the g-process [9]. In fact, scattering with these modes must be considered in order to match the mobility data in silicon [10]. The f- and g-transition picture holds up to moderately high electric fields ($<\sim 10^5$ V/cm).

3.2. Phonon–phonon scattering

The phonon distribution emitted by electrons thermalizes over a time scale of 10–100 ps. The thermalization occurs through phonon–phonon scattering in which the emitted non-equilibrium population scatters with phonons in the lattice heat bath. The time and length scale of this thermalization essentially determines the importance of sub-continuum thermal phenomena in the device. If the mean free path for the thermalization is much larger than the length scale for emission, then the heat source region remains perpetually out of equilibrium with the lattice. Chen [13] used the equation for phonon radiative transfer under a gray-body approximation to show that the thermal conductance in the region surrounding such a heat source effectively decreases as the ratio of the source size to the mean free path. The magnitude of this size effect thus depends on the degree to which thermal transport is truly ballistic in the device drain.

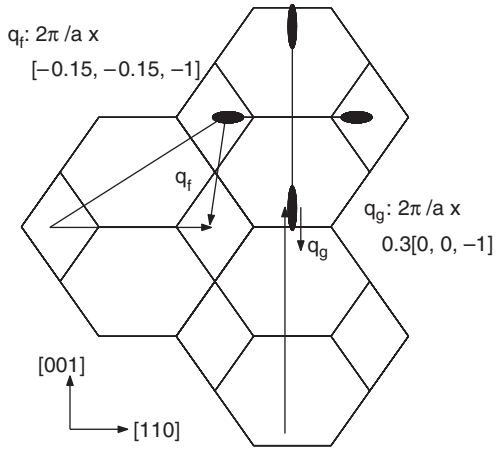

Fig. 4. The f- and g-processes are shown in a (110)-plane to identify the wave vectors of the involved phonons. The f-process phonon has the wave vector $\mathbf{q}_f$ and the g-process phonon has the wave vector $\mathbf{q}_g$ purely from geometry considerations.

If the time scale for thermalization is large compared to the time scale for emission, and the mean free path of the emitted phonons is small relative to the source size, then the emitted population can be driven severely out of thermodynamic equilibrium. Such phonons are referred to in literature as 'hot' phonons and such a heat source is called a phonon hotspot [14]. The question of whether emitted phonons do in fact become hot in a typical silicon MOSFET has not been investigated in great detail. Artaki and Price [15] estimated the electric field required for the onset of hot phonon related mobility degradation. Despite their conclusion that the field required was well within the range of typical device fields, this issue has not received much attention. Most of the work on hot phonons has focused on polar semiconductors such as GaAs where carriers are strongly coupled to zone center longitudinal optical phonons (see Refs. [16,17] for example).

The thermalization length scale mentioned above is essentially a diffusion length for phonons, and depends on the phonon group velocity, and the phonon–phonon and phonon-impurity scattering rates. The thermalization time scale, however, depends on the strength of phonon–phonon interactions only. The presence of a diffusely scattering material boundary such as the gate oxide interface affects both scales. The relaxation time for phonon–phonon interactions is difficult to compute directly. In particular, the relaxation times for optical phonon modes that dominate the emission spectrum in a silicon device are not well known, either through theory or experiments. In an earlier work, we have used atomistic simulations to investigate the anharmonic scattering of optical phonons at device-like hotspots and computed their scattering rates in the absence of a heat bath [4]. The hotspot is modeled as a phonon wave packet in this study. All emission is assumed to be in the g-type longitudinal optical phonon mode. The decay of energy in the initial high frequency optical mode and the creation of lower frequency acoustic modes are shown in Fig. 5. An important inference from this work is that the phonon scattering time is strongly dependent on local energy density under non-equilibrium conditions when a thermodynamic temperature cannot be defined.

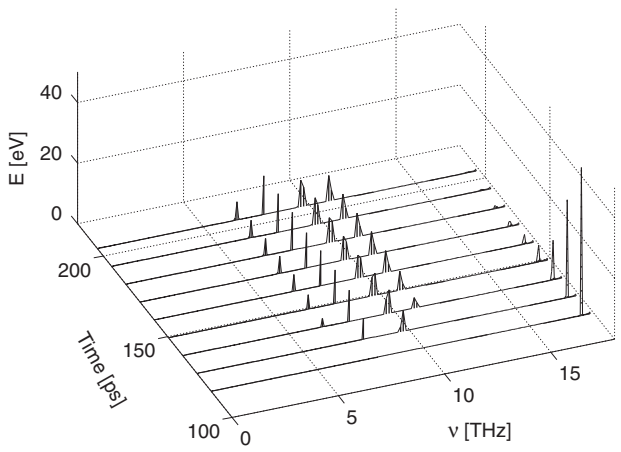

Fig. 5. Anharmonic phonon decay can be studied using atomistic simulations. The distribution of energy in the phonon spectrum is shown at various times during the decay of g-type longitudinal optical phonons at a hotspot with 50 eV energy.

The scattering rate is proportional to the square root of the energy density at the hotspot under such conditions. Fig. 5 compares the scattering times for phonons at two different energy densities.

4. The split-flux model for device hotspot

The distribution function for phonons evolves in a sixdimensional phase space due to their motion in real space, and due to scattering with other phonons, impurities and boundaries. Peierls [18] described this evolution in terms of a Boltzmann transport equation for phonons. Under the relaxation time approximation, the phonon BTE at steady state is

$$\mathbf{v}_{\mathbf{k},s} \cdot \nabla N_{\mathbf{k},s} = -\frac{N_{\mathbf{k},s} - \bar{N}}{\tau_{\mathbf{k},s}} \tag{1}$$

where ($\mathbf{k}$, s) refers to the phonon mode with wave vector $\mathbf{k}$ and polarization s, $\mathbf{v}$ is the group velocity, N is the phonon occupation number, $\bar{N}$ is the Bose–Einstein distribution function, τ is the net relaxation time from all scattering events. The distribution function can further be written in terms of a departure from equilibrium as

$$N = \bar{N}(T) + \eta_{\mathbf{k},s}. \tag{2}$$

where η is the departure function. The BTE is usually solved for this departure function.

In the case of a device, the BTE must also include a source term due to electron–phonon scattering. Thus, the steady state evolution can be written as

$$\mathbf{v}_{\mathbf{k},s} \nabla N_{\mathbf{k},s} = -\frac{N_{\mathbf{k},s} - \bar{N}}{\tau_{\mathbf{k},s}} + \dot{n}_{\mathbf{k},s}, \tag{3}$$

where $\dot{n}$ is the net phonon emission. In our previous work, we have proposed a departure function for the phonon source in a transistor. In our development, we assume that the phonon flux at any point in space is due to a nearequilibrium part that obeys the Fourier law; this is superposed on a secondary far-from-equilibrium contribution that does not obey the Fourier law. As we move away from the hotspot, we expect the emitted phonons to get scattered and the secondary contribution to diminish rapidly. The Fourier law contribution must increase proportionately to maintain energy continuity. We write the phonon departure from equilibrium in this 'split-flux' model [19] as

$$\eta_{\mathbf{k},s} = -\tau_{\mathbf{k},s}\mathbf{v}_{\mathbf{k},s} \cdot \nabla T \frac{\partial \bar{N}}{\partial T} + n_{\mathbf{k},s}, \tag{4}$$

where $n_{\mathbf{k},\mathrm{s}}$ is the far-from-equilibrium departure function. The BTE thus becomes

$$\cdot\left(\nabla n_{\mathbf{k},s} - \nabla\left(\tau_{\mathbf{k},s}\mathbf{v}_{\mathbf{k},s} \cdot \nabla T \frac{\partial \bar{N}}{\partial T}\right)\right) = \frac{n_{\mathbf{k},s}}{\tau_{\mathbf{k},s}} + \dot{n}_{\mathbf{k},s}. \tag{5}$$

Additionally, macroscopic energy continuity must be satisfied and this is expressed at steady state as

$$\nabla \cdot \mathbf{J} = \dot{Q}, \tag{6}$$

where $\mathbf{J}$ is the heat flux vector and $\dot{Q}$ is the heat generation rate. These quantities are expressed in terms of the nonequilibrium phonon distribution function as follow.

$$\dot{Q} = \sum_s \frac{1}{8\pi^3} \int \dot{n}_{\mathbf{k},s} \hbar\omega_{\mathbf{k},s} \mathrm{d}\mathbf{k} \tag{7}$$

$$\mathbf{J} = \sum_s \frac{1}{8\pi^3} \left[\int \mathbf{v}_{\mathbf{k},s} n_{\mathbf{k},s} \hbar\omega_{\mathbf{k},s}\, \mathrm{d}\mathbf{k} - \int \mathbf{v}_{\mathbf{k},s}\tau_{\mathbf{k},s}\mathbf{v}_{\mathbf{k},s} \cdot \nabla T \frac{\partial \bar{N}}{\partial T} \hbar\omega_{\mathbf{k},s}\, \mathrm{d}\mathbf{k} \right] \tag{8}$$

where the limits of integration span the phonon wavevector space. The second term in the heat flux vector can be written in terms of the thermal conductivity tensor, $\mathbf{K}$, so that Eq. (8) reduces to

$$\mathbf{J} = \sum_s \frac{1}{8\pi^3} \int \mathbf{v}_{\mathbf{k},s} n_{\mathbf{k},s} \hbar\omega_{\mathbf{k},s} \mathrm{d}\mathbf{k} - \mathbf{K} \cdot \nabla T. \tag{9}$$

With the above expression for the heat flux vector, Eq. (6) can be integrated to give

$$\sum_s \frac{1}{8\pi^3} \int \mathbf{v}_{\mathbf{k},s} n_{\mathbf{k},s} \hbar\omega_{\mathbf{k},s} \mathrm{d}\mathbf{k} - \mathbf{K} \cdot \nabla T = \int \dot{Q} \mathrm{d}\mathbf{r}. \tag{10}$$

Eqs. (5) and (10) form a closed system with the unknowns being $n_{\mathbf{k},s}$ and T. The above equations can also be rewritten to retain the time dependence if needed. Analytical methods to solve the above system are discussed in Ref. [19]. Here we discuss some of the important results obtained through applying the split-flux model to a device hotspot.

5. Steady state distributions

The integro-differential nature of the phonon BTE makes it difficult to solve for realistic device geometries.

The splitflux model discussed above reduces some of the complexity by using the BTE to only solve for the ballistic component of phonon transport. In this section we use the split-flux model to describe heat conduction from a device-like hotspot in a 90 nm gate length bulk device geometry. Fig. 6 shows a schematic of the device along with the thermal boundary conditions. The hotspot is assumed to be located at the origin to exploit symmetry and reduce computations. In reality the hotspot is slightly skewed toward the drain but the displacement is much smaller compared to the lateral extent of the domain. The top boundary is assumed to be adiabatic due to the insulating gate oxide. The small heat flux across this interface is ignored in order to obtain an analytical solution. The side boundaries are both adiabatic due to thick isolation oxides and the presence of neighboring devices. Some heat is lost through the metallic contacts at the top as shown. Finally, heat flows out to the heat sink through the bulk silicon at the bottom. We estimate the heat transfer coefficient at the bottom by assuming a junction to ambient thermal resistance of 0.5 K/W and the heat transfer coefficient at the top by treating the contact as an infinite fin. We assume the power density at the source to be 5 W/μm^3 and the radius to be 20 nm, which closely approximate the hotspot in a 90 nm bulk silicon NMOS operating at a supply voltage of 1.2 V. Finally, the device hotspot is assumed to be a step function in space to approximate the profile in a real device.

We obtain the net phonon emission term on the right hand side of Eq. (3) from the emission spectrum for electrons in silicon computed at a field of 4 MV/m using Monte Carlo simulations [20]. Another input required in the model are the phonon relaxation times. For acoustic modes, we use the expressions developed by Holland [21] and fit to thermal conductivity data for bulk silicon. We use a single value of 10 ps for all optical modes, which is on the order of the room temperature lifetime of zone-center optical modes in silicon measured by Raman spectroscopy

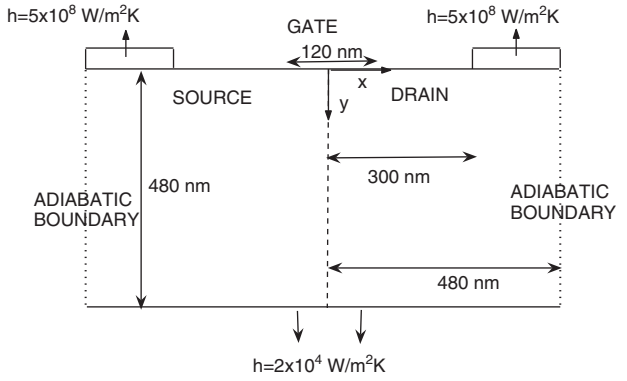

Fig. 6. The boundary conditions used in the device calculations are shown. The dashed line at the center is taken to be a line of symmetry. The hotspot is assumed to be semi-circular with a power density of 5 W/μm^3. Heat flows out through the bottom boundary with the heat transfer coefficient, h, as indicated. There is a second path for heat conduction, through the contacts shown at the top.

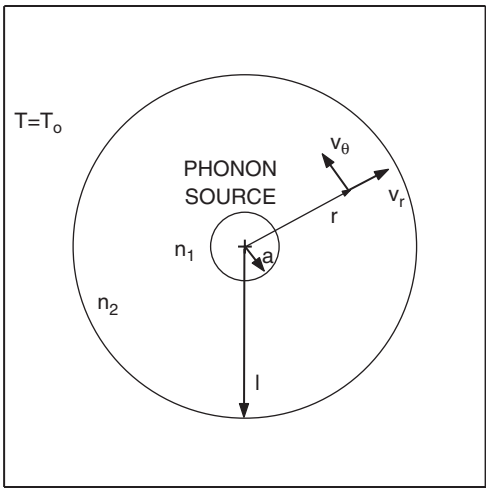

Fig. 7. The BTE is solved in a radial geometry with the source assumed to be confined within a radius equal to a. The solution is obtained in the two domains indexed as 1 and 2.

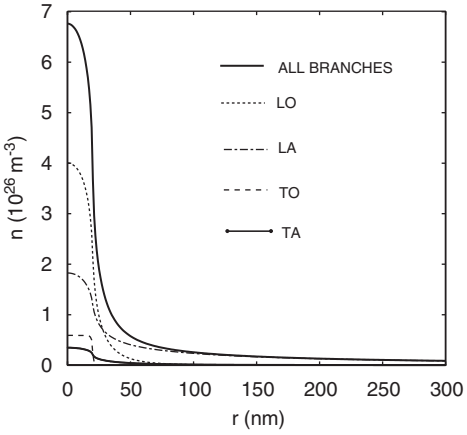

Fig. 8. The phonon number density for different branches is shown close to the hotspot which is 20 nm in extent. The LO contribution dominates in terms of number density and consequently, in terms of energy density.

[22]. We note that the actual scattering rates at a device hotspot are likely to be higher than these near-equilibrium rates and ongoing work may provide more accurate rates in future.

We exploit the radial symmetry in the problem to rewrite the steady-state phonon BTE in radial co-ordinates as:

$$v_r \frac{\partial n_{\mathbf{k},s}}{\partial r} + \frac{v_\theta^2}{r} \frac{\partial n_{\mathbf{k},s}}{\partial v_r} - \frac{v_\theta v_r}{r} \frac{\partial n_{\mathbf{k},s}}{\partial v_\theta} + \frac{n_{\mathbf{k},s}}{\tau_{\mathbf{k},s}} = \dot{n}_{\mathbf{k},s} \tag{11}$$

where v_r and v_θ are the radial and tangential velocities as depicted in Fig. 7. Due to axial symmetry, there is no dependence on the azimuthal angle, θ. The solution to the above equation is obtained by employing the method of characteristics. We describe this in Ref. [19].

In comparing the non-equilibrium contribution of the different phonon branches to the number density and the non-Fourier heat flux, we find the longitudinal optical contribution to be more than 50% of the total. Fig. 8

shows the contribution from all four branches near the hotspot. In the figure, L refers to longitudinal polarization, T to transverse polarization, A to an acoustic branch and O to an optical branch. The energy density is also proportionately higher for the LO branches since, LO frequencies are higher compared to other branches. The LO and the LA branches are dominant contributors to the heat flux. This is shown in Fig. 9 where the contribution of individual branches to the heat flux is plotted. We note that the flux at the center is zero for all branches due to the imposed symmetry condition. The contributions of TA and TO are insignificant. The non- Fourier heat flux diminishes rapidly outside the hotspot.

The departure from equilibrium is evident from the sharp temperature gradients near the hotspot in Fig. 10. The splitflux model predicts an excess temperature rise of 5 K above the peak temperature rise of about 38 K from heat diffusion. The temperature obtained from the split-flux model is not a thermodynamic temperature in the strict sense but an equivalent temperature at which a Planck population would have the same energy density. The difference of 13% in the temperature rise is significant considering that the boundary conditions account for realistic heat flow to the ambient. Thus, these boundary conditions provide the best case scenario for heat conduction.

The fact that the split-flux temperature is not greater by a higher margin indicates that the size effect described in Section 3.2 though present is not currently dominant. The macroscopic contributions of the thermal interface material and the heat sink dominate the overall thermal resistance. We find further explanation for this by computing the mean free path of the emitted phonons. The spread in phonon free paths at room temperature in silicon is shown in Fig. 11. The curves show the free path as a function of phonon frequency and polarization. The mean free paths for device hotspot phonons in each phonon branch as computed through the split-flux model are also provided in the figure. Since the mean free path of the dominant LO branch is comparable but not significantly larger than the size of the hotspot, the size effect is limited. We recall that the size effect increases as the ratio of the mean free path to the source size. Next we consider the impact of hot phonons. Longitudinal optical phonons, which are strongly coupled to electrons, have an equivalent branch temperature of approximately 358 K. The occupation of the g-type LO phonon is about 15% higher than its equilibrium occupation at 338 K, the peak temperature estimated from the heat diffusion equation. However, an increase in the peak electric field, and consequently peak power density can increase this substantially in future technology nodes.

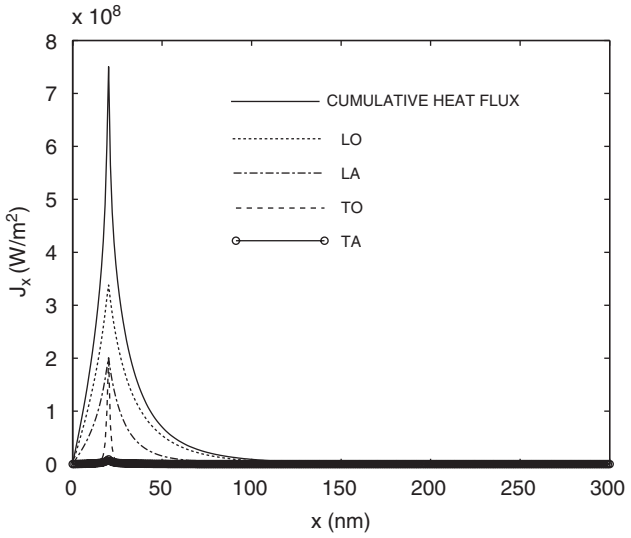

Fig. 9. A branch-wise break up of the far-from-equilibrium heat flux near a hotspot in a bulk device [19] shows the dominance of LO phonons.

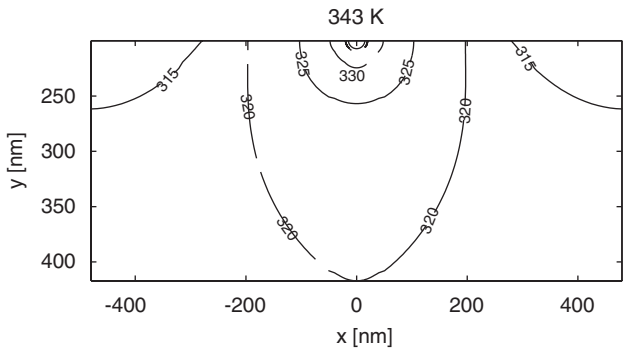

Fig. 10. The equivalent temperature field for phonon heat conduction from a device-like hotspot in bulk 90 nm geometry.

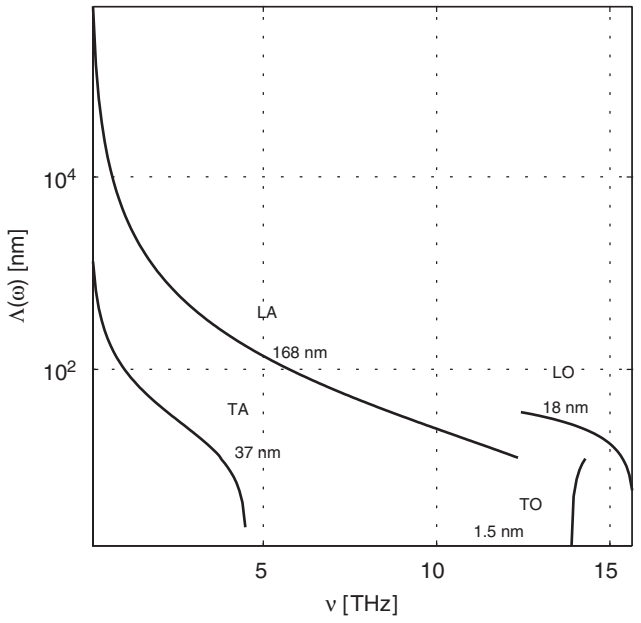

Fig. 11. The distribution of free paths in room temperature silicon as a function of phonon frequency and polarization is shown. The mean free path of phonons emitted by hot electrons in a device are also given for comparison. The use of a gray-body approximation for the heat source leads to large errors in predicting the magnitude of the hotspot effect.

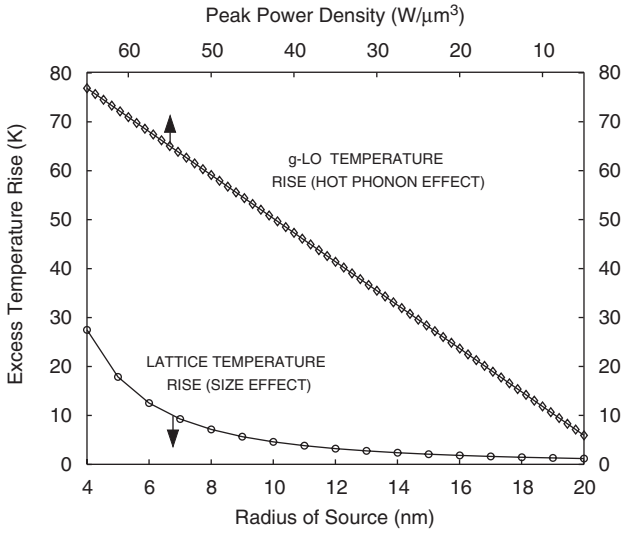

Fig. 12. The lower curve show the excess equivalent temperature rise in the lattice as a function of the heat source size. The source size is around 20 nm at the 90 nm node and is estimated to be around 4 nm for an 18 nm gate length device. The upper curve shows the excess equivalent temperature of the g-type longitudinal optical phonon as a function of the peak power density.

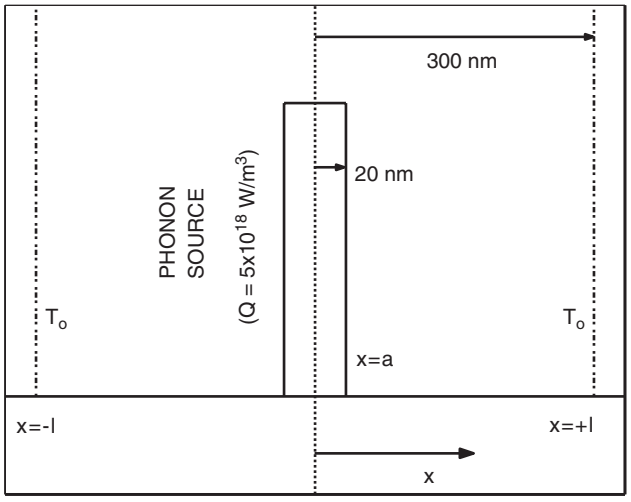

Fig. 13. A step-like phonon source symmetric about $x = 0$ with a uniform power density of $5 \times 10^{18}\,\mathrm{W/m^3}$ is considered as a sample calculation. The extent of the source, a, is taken to be 20 nm, consistent with device hotspots. A sink at 300 K is assumed to be at $x = \pm l$ where $l = 300$ nm.

Fig. 12 compares the magnitudes of the two subcontinuum effects in future technology nodes. At the right end is the 90 nm bulk Si NMOS ($V_{dd} = 1.2$ V) with a hotspot dimension of around 20 nm and a peak power density of $5\,\mathrm{W/\mu m^3}$, and at the left end is a model 18 nm ultrathin-body Si device ($V_{dd} = 1.1$ V) with a hotspot dimension of about 4 nm and a peak power density of $65\,\mathrm{W/\mu m^3}$. Both these devices correspond to 'well-tempered' devices provided in Ref. [3]. The ordinate shows the temperature rise in excess of the temperature rise predicted by diffusion theory. The lower curve shows the averaged lattice temperature in the hotspot, and is a measure of the size effect. As the hotspot shrinks in size compared to the mean free path of phonons, it remains increasingly out of equilibrium with its vicinity. The upper curve shows an equivalent temperature rise in g-type LO phonon modes that strongly couple with electrons in silicon. In our model this has a linear dependence on the peak power density, which is given by the abscissa at the top. Thus, optical phonons will become increasingly 'hotter' in future technologies. We are currently investigating how this may impact carrier mobility and threshold voltage.

6. Transients during switching

In a CMOS circuit, energy is mainly dissipated during a switching event, assuming leakage currents are low. However, with scaling this is increasingly not the case. As a first step toward investigating the impact of power transients on phonon populations in the device, we present a transient one-dimensional transport model. The domain is shown schematically in Fig. 13. We consider a onedimensional phonon source rather than the full device geometry in order to keep the problem tractable. The phonon source in this study is time dependent. The transistor typically switches on a time scale of about 100 ps. The device is on over only a fraction of this period, referred to as the duty cycle, which is typically less than 0.3. The switching time is, however, comparable to the relaxation times of some of the phonon modes. In this section, we report preliminary studies on phonon distributions during a switching event, neglecting leakage power.

In the domain considered, the phonon source is symmetric about $x = 0$, where x is the co-ordinate direction. A phonon sink is located at $x = \pm l$. The time-dependent BTE is of the form:

$$\frac{\partial n_{\mathbf{k},s}}{\partial t} + v_x \frac{\partial n_{\mathbf{k},s}}{\partial x} = -\frac{n_{\mathbf{k},s}}{\tau_{\mathbf{k},s}} + \dot{n}_{\mathbf{k},s} f(t) \tag{12}$$

where $f(t)$ is the switching function,

$$f(t) = 1 \; t \bmod t_0 < \alpha = 0 \; t \bmod t_0 > \alpha \tag{13}$$

with t_0 being the switching period and α being the duty cycle or the fraction of the time period that the device is on. As mentioned above, we have assumed negligible leakage power to construct the base case for comparison. Further, the boundary conditions are expressed as

$$n^{+}_{\mathbf{k},s}(x = 0, |v_x|, t) = n^{-}_{\mathbf{k},s}(x = 0, -|v_x|, t)$$
$$n^{-}_{\mathbf{k},s}(x = l, -|v_x|, t) = 0 \quad T(x = l, t) = T_0 \tag{14}$$

where $n^{+}_{\mathbf{k},s}$ is the departure function for phonons traveling to the right in Fig. 13 and $n^{-}_{\mathbf{k},s}$ is the function for phonons traveling to the left. Eq. (12) is linearized by computing the relaxation time at the temperature field obtained from the heat diffusion equation. The solution to the transient

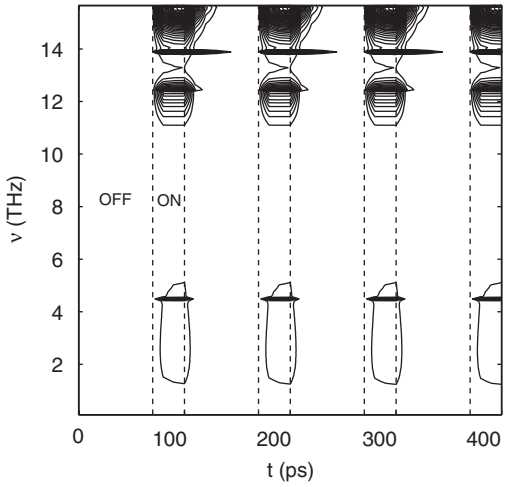

Fig. 14. Contours of the normalized phonon number density, spaced by 0.01, are shown as a function of frequency and time at $x = 0$. There is no phonon accumulation for a switching period of 100 ps with a duty cycle of 30%.

problem is [19]:

$$n^{+}_{\mathbf{k},s} = \mathrm{e}^{-(x'|v_x|\tau)}\left[\int_0^x \frac{\dot{n}}{|v_x|} f\left(t + \frac{x'+x}{|v_x|}\right)\mathrm{e}^{(x''|v_x|\tau)}\,\mathrm{d}x' + \int_0^l \frac{\dot{n}}{|v_x|} f\left(t - \frac{x'+x}{|v_x|}\right)\mathrm{e}^{-(x''|v_x|\tau)}\,\mathrm{d}x'\right] \quad (15)$$

$$n^{-}_{\mathbf{k},s} = \mathrm{e}^{(x'|v_x|\tau)}\int_x^1 \frac{\dot{n}}{|v_x|} f\left(t - \frac{x'-x}{|v_x|}\right)\mathrm{e}^{-(x''|v_x|\tau)}\,\mathrm{d}x'$$

We use the above solution to compute phonon number density as a function of time, position and the phonon frequency. Fig. 14 shows the phonon number density contours at the hotspot. The plot provides a snapshot of phonons being created and annihilated at the hotspot as the device is switched. We find that the emitted population is thermalized over an effective time period of about 60 ps. We do not find any evidence of phonon retardation (or memory effects) in our investigation. With increasing leakage power, there will be significant phonon emission during the offstate, leading to perpetual phonon non-equilibrium in the drain. Ongoing work aims to study the impact of this nonequilibrium on device performance.

7. Concluding remarks

Sub-continuum conduction physics plays an important role in determining the thermal response of nanotransistors. The presence of a phonon hotspot in the drain leads to sharp gradients in the energy density at the drain, and gives rise to a peak temperature in excess of that predicted by the heat diffusion equation. This non-equilibrium effect may be characterized in terms of a sub-continuum thermal resistance. The resistance will scale upwards in future technologies due to the gradual increase in the peak electric field and an increasing peak volumetric power density. This will give rise to significant hot phonon populations. The influence of these on mobility and the threshold voltage is currently under investigation. It is illustrative to compare the relative magnitudes of various thermal resistances toward the limits of scaling. We choose the 18 nm gate length technology node as an example and assume the voltage to be 1.1 V and the saturation current to be a conservative 1000 μA/μm. We estimate the thermal resistance due to the packaging and the bulk silicon substrate to be about 6.5 K/μW for a single transistor. The use of SOI adds another 2.5 K/μW to the resistance. The sub-continuum contribution to the localized thermal resistance in the drain is a very significant 2.2 K/μW. The impact of leakage power on phonon departure from equilibrium and its influence on device performance remains an open question. Our preliminary results indicate that there may be significant non-equilibrium due to leakage and this needs further investigation. Finally, fully coupled electron–phonon transport modeling is necessary in order to better understand the impact of 'hot' phonons. Our work on phonon transport modeling is but the first step in this direction, and is intended to ultimately bring this aspect of device transport into parity with its electronic counterpart.

Acknowledgements

S. Sinha acknowledges the Intel PhD Fellowship, the Stanford Graduate Fellowship, and the Semiconductor Research Corporation (Task 1043) for support. The authors thank Eric Pop at Stanford (EE), Dr Ravi Prasher at Intel (Chandler) and Dr Siva Narendra at Intel (Hillsboro) for helpful discussions.

References

[1] P.E. Gronowski, W.J. Bowhill, R.P. Preston, M.K. Gowan, R.L. Allmon, IEEE J. Solid-State Circuits 33 (1998) 676.
[2] C.C. Liu, J. Zhang, A.K. Dutta, S. Tiwari, IEEE Electron. Device Lett. 23 (2002) 716.
[3] Well-Tempered Bulk-Si NMOSFET Device, Microsystems Technology Laboratory, MIT, http://www-mtl.mit.edu/Well/.
[4] S. Sinha, P.K. Schelling, S.R. Phillpot, K.E. Goodson, J. Appl. Phys. 97 (2005) 023702.
[5] R. Mahajan, R. Nair, V. Wakharkar, J. Swan, J. Tang, G. Vandentop, Intel Technol. J. 6 (2002) 62.
[6] Y.S. Ju, K.E. Goodson, Appl. Phys. Lett. 74 (1999) 3005.
[7] N.W. Ashcroft, N.D. Mermin, Solid State Physics, Brooks/Cole, USA, 1976.
[8] A. Balandin, K.L. Wang, Phys. Rev. B 58 (1998) 1544.
[9] D.K. Ferry, Semiconductor Transport, Taylor & Francis Inc., New York, 2000.
[10] D.K. Ferry, Phys. Rev. B 14 (1976) 1605.
[11] D. Long, Phys. Rev. 120 (1960) 2024.
[12] B.K. Ridley, Quantum Processes in Semiconductors, Clarendon Press, Oxford, 1999.
[13] G. Chen, J. Heat Transfer 118 (1996) 539.
[14] D.V. Kazakovtsev, I.B. Levinson, Sov. Phys. JETP 61 (1985) 1318.
[15] M. Artaki, P.J. Price, J. Appl. Phys. 65 (1989) 1317.
[16] P. Kocevar, Adv. Solid State Phys. 27 (1987) 197.

[17] P. Lugli, P. Bordone, L. Reggiani, M. Rieger, P. Kocevar, S.M. Goodnick, Phys. Rev. B 39 (1989) 7852.
[18] R.E. Peierls, Quantum Theory of Solids, Clarendon Press, Oxford, 1955.
[19] S. Sinha, E. Pop, K.E. Goodson, Imece-2004 in: International Mechanical Engineering Congress and Exposition, Anaheim, CA, 2004.
[20] E. Pop, R.W. Dutton, K.E. Goodson, Sispad (ieee) in: SISPAD, Boston, 2003.
[21] M.G. Holland, Phys. Rev. 132 (1963) 2461.
[22] J. Menèdez, M. Cardona, Phys. Rev. B 29 (1984) 2051.
[23] International Technology Roadmap for Semiconductors (ITRS), http://www.public.itrs.net.