INVITED PAPER

Heat Generation and Transport in Nanometer-Scale Transistors

Heat problems in ever-smaller integrated circuits include hot-spots at transistor drain areas, reduced heat conduction in new devices and higher thermal resistance at material boundaries.

By Eric Pop, Sanjiv Sinha, and Kenneth E. Goodson

ABSTRACT | As transistor gate lengths are scaled towards the 10-nm range, thermal device design is becoming an important part of microprocessor engineering. Decreasing dimensions lead to nanometer-scale hot spots in the transistor drain region, which may increase the drain series and source injection electrical resistances. Such trends are accelerated by the introduction of novel materials and nontraditional transistor geometries, including ultrathin body, FinFET, or nanowire devices, which impede heat conduction. Thermal analysis is complicated by subcontinuum phenomena including ballistic electron transport, which reshapes the heat generation region compared with classical diffusion theory predictions. Ballistic phonon transport from the hot spot and between material boundaries impedes conduction cooling. The increased surface to volume ratio of novel transistor designs also leads to a larger contribution from material boundary thermal resistance. This paper surveys trends in transistor geometries and materials, from bulk silicon to carbon nanotubes, along with their implications for the thermal design of electronic systems.

KEYWORDS | Ballistic; carbon nanotubes; germanium-on-insulator (GOI); heat generation; Monte Carlo; MOSFET; nonequilibrium; phonon; power density; scaling; silicon-on-insulator (SOI); temperature

Manuscript received February 9, 2005; revised February 7, 2006. This work was supported by the Semiconductor Research Corporation (SRC) under Task 1043. The work of E. Pop was supported by a joint SRC/IBM graduate fellowship. The work of S. Sinha was supported by a Stanford Graduate Fellowship (SGF) and an Intel fellowship.

E. Pop was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with Intel Corp. Santa Clara, CA 95054 USA.
S. Sinha was with the Department of Mechanical Engineering, Stanford University, Stanford CA 94305 USA. He is now with Intel Corp, Hillsboro, OR 97124 USA.
K. E. Goodson is with the Department of Mechanical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: goodson@stanford.edu).

Digital Object Identifier: 10.1109/JPROC.2006.879794

I. INTRODUCTION

The continued miniaturization of integrated circuits and the current trend toward nanoscale electronics have led to tremendous integration levels, with hundreds of millions of transistors assembled on a chip area no larger than a few square centimeters. Circuit densities are projected to reach the gigascale as the smallest lateral device feature sizes approach 10 nm. An often cited technological roadblock of this scaling trend is the "power problem," i.e., power densities, heat generation, and chip temperatures reaching levels that will prevent the reliable operation of integrated circuits. Chip-level power densities are currently on the order of 100 W/cm^2. If the rates of integration and miniaturization continue to follow the International Technology Roadmap for Semiconductors (ITRS) guidelines [1], the chip-level power density is likely to increase even further [2], as illustrated in Fig. 1. Increasing power density levels will quickly drain batteries in portable devices and may render many electronic systems unusable without significant advances in cooling technology, or without fundamental shifts in design. The situation is compounded by millimeter-scale hot spots on the chip, i.e., localized regions of higher heat generation rate per unit area and hence higher temperatures (e.g., near the clock) [3].

While chip-level hotspots are troubling circuit designers [4], device designers are beginning to encounter thermal management problems at nanometer-length scales within individual transistors [5]. Novel and complex device geometries tend to make heat removal more difficult (Fig. 2) and most new materials being introduced in device processing have lower thermal conductivities than bulk silicon (Table 1). Self-heating of devices during operation occurs through the interaction of electrons (the current carriers, set in motion by the electric fields) with the lattice vibrations (phonons, whose net motion is governed by gradients in their density). Modern device technologies

0018-9219/$20.00 ©2006 IEEE

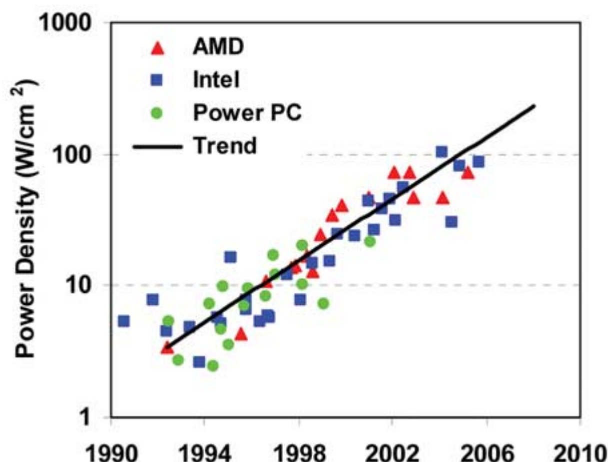

Fig. 1. ***Trends of on-chip power density over the past 10+ years. Note the vertical axis is logarithmic and the horizontal one (years) is linear. The solid line marks an exponential trend. By comparison, the power density of a hot plate is of the order 10 W/cm^2, that of a nuclear reactor around 100 W/cm^2, while the surface of the sun is about 7000 W/cm^2. Data compiled by F. Labonte, Stanford.***

operate at length scales comparable to both the electron and phonon mean free paths (approximately 5–10 and 200–300 nm in bulk silicon at room temperature, respectively [6], [7]), and future technologies are going to forge deeper into this subcontinuum regime. Ballistic conditions dominate both electron (current) and phonon (heat) transport at such length scales, leading to nonequilibrium conditions between the energy carriers. The electron–phonon interaction is neither energetically nor spatially uniform [8], [9] and the generated phonons have widely varying contributions to heat transport: optical phonons have small group velocity and contribute little to the thermal conductivity, which is dominated by acoustic phonon transport [7], [8]. In addition, the thermal conductivity of semiconductor films thinner than the phonon mean free path is significantly reduced by phonon confinement and boundary scattering. This increases the thermal resistance of a device, leading to higher operating temperatures in thin-film and novel, confined-geometry transistors [Fig. 2(b) and (c)] compared to bulk transistors operating at the same power input [Fig. 2(a)].

This paper provides an overview of heat generation and conduction issues within nanometer-scale transistors. The following section describes the mechanism of Joule heating in semiconductor devices and presents two methods for computing the heat (phonon) generation. The next section reviews the state-of-the-art in heat conduction at nanometer length scales. Section IV discusses the impact of various material choices on the heat transport characteristics of future-generation transistors. The remainder of the paper is dedicated to an overview of heat transfer issues in nanometer-scale devices from bulk silicon to carbon nanotubes, with particular emphasis on the subcontinuum aspects of their behavior.

Table 1 Approximate Thermal Conductivities of a Few Materials Used in Device Fabrication. Phonon-Boundary Scattering Significantly Reduces the Thermal Conductivity of a 10 nm Thin Si Film. Bulk Si and Ge are Assumed Undoped. Note That Lattice Vibrations (Phonons) Are Responsible for the Thermal Conductivity of all Dielectric Materials Listed, But Electrons are the Heat Carriers in Silicides (e.g., NiSi), Which are Metals

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

II. HEAT GENERATION

In the context of a field-effect transistor (FET), the applied voltage leads to a lateral electric field which peaks near the device drain. This field accelerates the charge carriers (e.g., conduction band electrons in a n-type FET)

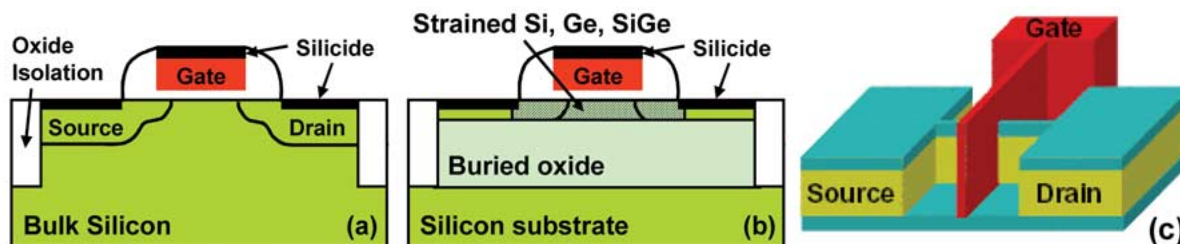

Fig. 2. ***Evolution of transistor designs from: (a) bulk silicon (b) to (strained) silicon or germanium on insulator to (c) multiple-gate or FinFET devices. Advanced device designs like (b) and (c) introduce more complicated geometries and lower thermal conductivity materials, making heat removal more difficult. The gate length is expected to be scaled to the 10-nm range, while the semiconductor film and "fin" thickness of (b) and (c), respectively, to about one-third to one-half the gate length. Fig. (c) is from the ITRS [1].***

which gain energy and heat up. Electrons can scatter with each other, with lattice vibrations (phonons), with material interfaces, imperfections or impurity atoms. Of these, the electron population only loses net energy by scattering with phonons, consequently heating up the lattice through the mechanism known as Joule heating. Other scattering mechanisms chiefly affect the electron momentum [6]. The lattice absorbs the extra electron energy, warms to a higher temperature (T), and in return affects the electronic transport properties of the material: The electron mobility in bulk (undoped) silicon decreases approximately as $T^{-2.4}$ around room temperature owing to higher phonon populations and increased scattering rates. When other scattering mechanisms come into play, the electron mobility is more weakly dependent on temperature: it decreases approximately as $T^{-1.7}$ in highly doped silicon and $T^{-1.4}$ in nanometer-thin silicon layers, where boundary scattering becomes of importance [6], [10].

Electrons with energies below 50 meV scatter mainly with acoustic phonons in silicon, while those with higher energy scatter strongly with the optical modes. The optical phonon modes have low group velocities (on the order of 1000 m/s) and their occupation is also relatively low, hence they contribute very little to heat transport [7]. The primary heat carriers in silicon are the faster acoustic phonon modes, which are significantly populated and have group velocities from 5000 m/s for transverse modes to 9000 m/s for longitudinal acoustic modes. Optical phonons decay into acoustic modes, but over relatively long time scales, i.e., picoseconds, compared to the electron-optical phonon scattering time which is on the order of tenths of picoseconds [11]. Under high field conditions, this can lead to the creation of a phonon energy bottleneck which can cause the density of optical phonon modes to build up over time, leading to more scattering events and impeding electron transport [12]. These processes are symbolically illustrated in Fig. 3.

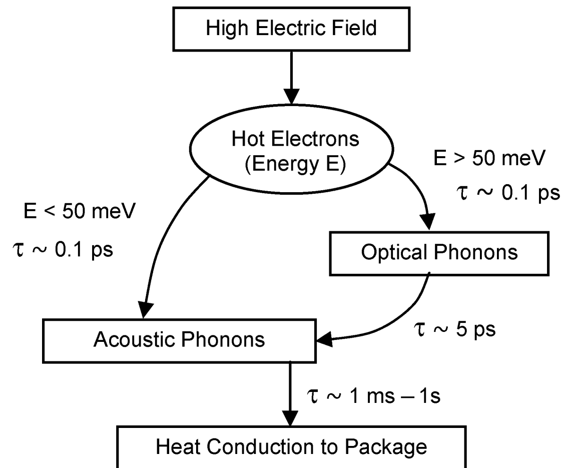

Fig. 3. ***Diagram and characteristic time scales of energy transfer processes in silicon. Scattering with slow optical phonons is the dominant energy relaxation mechanism for electron energies above 50 meV. This may create a phonon energy bottleneck until the optical phonons decay into the faster acoustic modes. A similar nonequilibrium heat dissipation bottleneck has also been recently observed in suspended carbon nanotubes [61], where the optical phonon energy is almost four times higher (see Section IX).***

In the context of traditional semiconductor device simulators, the Joule heating rate per unit volume is typically computed starting from the dot product of the electric field ($\mathbf{E}$) and current density ($\mathbf{J}$):

$$H = \mathbf{J} \cdot \mathbf{E} + (R - G)(E_G + 3k_B T) \tag{1}$$

where the second term represents the heating rate due to nonradiative generation (G) and recombination (R) of electron-hole pairs, E_G is the semiconductor band gap, k_B the Boltzmann constant and T the lattice temperature [13]. The equation is also applicable in degenerately doped semiconductors as long as bandgap narrowing is properly taken into account. However, this approach does not account for the microscopic nonlocality of the phonon emission near a strongly peaked electric field region, such as the drain of a transistor. Although electrons gain most of their energy at the location of the peak electric field, they must travel several inelastic mean free paths before releasing all of it to the lattice, in decrements of (at most) the optical phonon energy. In silicon transistors, for example, electrons can gain energies that are a significant fraction of an eV, while the optical phonon energy is only about 50–60 meV. Assuming an electron velocity of 10^7 cm/s (the saturation velocity in silicon) and an electron–phonon scattering time around 0.05–0.10 ps in the high-field region, the inelastic mean free path is then on the order of 5–10 nm [6]. The full electron energy relaxation length is therefore even longer, on the order of several mean free paths. While such a discrepancy may be neglected on length scales of micrometers, or even tenths of a micrometer, it must be taken into account when analyzing heat generation on length scales of the order 10 nm. The highly localized electric field in such devices leads to the formation of a nanometer-sized hot spot in the drain region, that is spatially displaced (by several mean free paths) from continuum theory predictions. In addition, the $\mathbf{J} \cdot \mathbf{E}$ formulation of the Joule heating also does not differentiate between electron energy exchange with the various phonon modes, and does not give any spectral information regarding the types of phonons emitted.

The mechanism through which Joule heating occurs is that of electron scattering with phonons, and consequently only a simulation approach which deliberately incorporates all such scattering events will capture the

complete microscopic, detailed picture of lattice heating. A Monte Carlo (MC) simulation method was recently introduced to compute subcontinuum and phonon frequency-specific heat generation rates, with applications at nanometer-length scales [8], [14]. Both the electron energy bands and the phonon dispersion can be analytically approximated as ellipsoids and quadratics, respectively. This is an efficient approximation for devices operating at voltages below the silicon bandgap ($E_G/q = 1.1$ V), such as those of future technologies, and it enables significantly faster code that is easier to implement and debug. Impact ionization and scattering with higher conduction band valleys are both strongly suppressed at voltages below E_G/q, so they can be safely ignored. Inelastic scattering with all intervalley as well as intravalley phonons is incorporated, as given by scattering selection rules and by the phonon dispersion. Particular care is taken to treat the acoustic phonon interaction inelastically, including the phonon dispersion. An isotropic, analytic fit to the dispersion relation [see Fig. 4(a)] is also used when computing the final state after scattering with both optical and acoustic phonons, satisfying momentum and energy conservation.

The phonon emission and absorption events during a simulation run are tallied and full heat generation statistics can be collected. The total heat generation rate per unit volume can be computed as

$$H = \frac{n}{N_{\text{sim}}\Delta t}\sum(\hbar\omega_{\text{ems}} - \hbar\omega_{\text{abs}}) \qquad (2)$$

where n is the electron density, N_{sim} is the number of simulated particles and Δt is the simulation time. Fig. 4(b) shows the net (emission minus absorption) numbers of generated phonons during a typical simulation of current flow and Joule heating in a silicon resistor with constant 50 kV/cm applied electric field. The shape of the generated phonon distribution approximately follows the density of states, while the specific phonons involved are given by the various scattering selection rules [15]: only phonons of specific energy and momentum may scatter electrons in silicon, in transitions between the six equivalent conduction band valleys (intervalley scattering), or within a single valley (intravalley scattering), as given by energy and momentum conservation. The area under each peak is proportional to the square of the coupling constant (deformation potential) of the respective electron–phonon interaction, and is a measure of the strength of each scattering type [14], [16]. The broadening of the phonon generation peaks occurs when energetic electrons, farther away from the conduction band minima in the Brillouin zone, may scatter with phonons outside those strictly dictated by the geometric selection rules. This is a direct consequence of using a continuous analytic representation of the phonon dispersion in selecting the final states. The dips between the phonon generation peaks in Fig. 4(b) are more pronounced at lower fields, and less evident at higher fields, when more phonon modes are allowed to participate [8].

Similar simulations carried out in *strained* silicon reveal that the phonon generation spectrum is essentially the same as in bulk silicon at high electric fields, when electrons have enough energy to emit across the entire phonon spectrum despite the strain-induced band splitting (in strained silicon, four of the six conduction band valleys are raised in energy, by an amount proportional to the strain). However, at low fields the heat generation spectrum in strained silicon is dominated by g-type longitudinal optical phonon emission since f-type (zone edge) phonon scattering is strongly reduced by the band splitting [8], [14]. In fact, it is only at low fields and in strained silicon that heat generation is dominated by longitudinal optical (LO) phonons, whereas in most practical situations the emission spectrum is significantly broadened, as shown in Fig. 4(b). In bulk silicon and at high fields, LO phonons account for just over half of the energy dissipation, while longitudinal acoustic (LA) modes account for approximately one third, and transverse optical (TO) for about one-tenth [8]. The resulting phonon generation spectra can be used as inputs for molecular

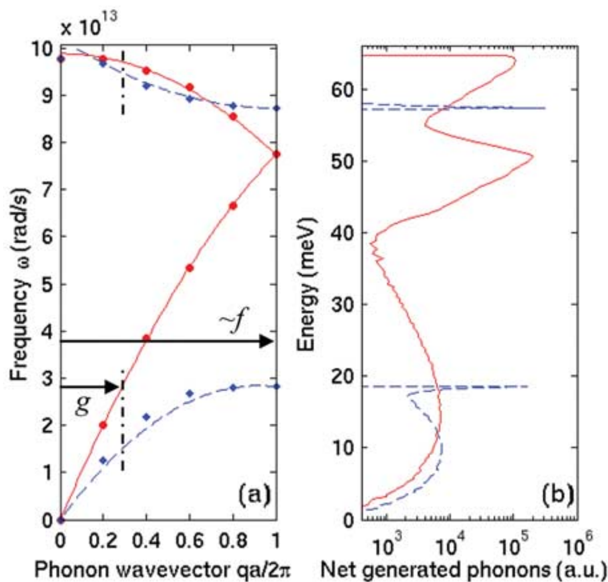

Fig. 4. ***(a) Phonon dispersion in the (100) direction of silicon and (b) computed net (emission minus absorption) phonons generated by Joule heating. The symbols in (a) are from neutron scattering data and the lines are a quadratic fit [8]. Solid lines are for longitudinal, and dashed lines are for transverse phonons. Optical modes are above and acoustic modes have energies below approximately 50 meV, respectively. The wave vectors of the f- and g-type phonons responsible for intervalley scattering are shown on the figure. Note the vertical axes are matched ($E = \hbar\omega$) to facilitate comparison of the dispersion and generated modes.***

dynamics [17] or microscopic phonon transport simulations [18], [19], and they offer a comprehensive look at the complexity of Joule heating in silicon.

The simulation can also be run in the context of a realistic device design, collecting heat generation information across the entire domain (Fig. 5). In the limit of an ultrashort MOSFET when transport across the channel is quasi-ballistic and the lateral field peaks strongly at the channel–drain junction, it is found that electrons relax their energy far into the drain, as qualitatively described above. The spatial distribution of heat generation in a quasi-ballistic device is shown in Fig. 6, both as computed with the classical $\mathbf{J} \cdot \mathbf{E}$ approach (1) and with the Monte Carlo method (2). We note the heating of the lattice in the latter calculation is entirely due to electron–phonon scattering, since the electron–electron interaction only serves to redistribute the energy and momentum within the electron population, and electron-impurity scattering is elastic (affecting carrier momentum, but not energy) [6]. Electron–plasmon scattering *may* provide an indirect channel for lattice heating in polar semiconductors (like GaAs) through the stronger plasmon–phonon coupling, but this is unlikely the case here. In silicon, electron-plasmon interactions simply serve to slightly reshape the electron distribution at very high energies [20], but the precise shape of the heat generation spectrum at high fields has been found to be largely independent of the electron distribution (skewed as it may be), and rather to scale fairly linearly with the field itself [21]. When non-equilibrium transport is simulated with the Monte Carlo approach described here, we have found it is a very good assumption that *the entire* Joule heat is dissipated in the *drain* of quasi-ballistic devices [9]. This implies that drain-end transistor design ought to play an important role for future device thermal behavior, just as source-end design is generally thought to be of most importance for the electrostatic behavior (e.g., to control injection velocity and threshold voltage). Furthermore, the proximity of the drain contact to the heat generation region also implies that drain contact design may affect the heat dissipation and spreading from the active device region. These results are further utilized below, in the discussion regarding scaling of ultrathin body transistors.

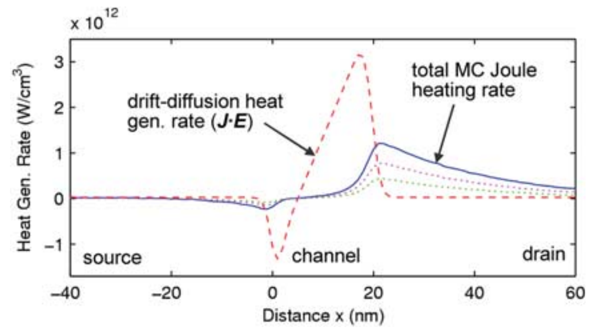

Fig. 6. ***Heat generation in an $n+/n/n+$ quasi-ballistic device with channel length 20 nm [9]. The source and drain are doped to 10^{20} cm^{-3}, the applied voltage is 0.6 V. Unlike the classical (drift-diffusion) result, the Monte Carlo (MC) simulation shows that heat is dissipated far into the device drain. The dotted lines represent the optical phonon (upper) and acoustic phonon (lower) heat generation rates from the MC result.***

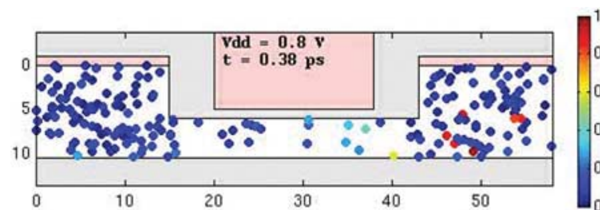

Fig. 5. ***Snapshot of electron transport in a future-generation thin-body silicon-on-insulator (SOI) device simulated with the Monte Carlo code MONET. The color scale is the electron energy in eV and the device dimensions are in nm. The contacts are on top, and the device is surrounded by oxide. The gate and drain are both tied to $V_{dd} = 0.8$ V, the source is grounded, and the location of all terminals from left to right is consistent with Figs. 8 and 11.***

At this point, it is worthwhile to revisit the conditions which may lead to the existence of hot phonons in nanoscale silicon devices [12]. Our observations in Fig. 4 show that phonon generation is more evenly spread across the phonon spectrum than the long-held (simplifying) assumption of heat dissipation predominantly with the optical phonon modes [22]. Nevertheless, we find that significant non-equilibrium optical phonon populations may still build up, particularly for the *g*-type LO mode. The generation rates for 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 nonequilibrium effects are less significant. Assuming a 5 ps phonon lifetime, we find the occupation number of the *g*-type LO phonon may exceed $N_{LO} > 0.1$ and become comparable to unity for power densities greater than 10^{12} W/cm^3 [9]. Such power densities are attainable in the drain of 20 nm (or shorter) channel length devices at operating voltages from the current ITRS guidelines [1], as shown in Fig. 6. Nonequilibrium phonon populations would increase electron scattering in the drain, leading (at the very least) to a magnification of the drain series resistance. The full ramifications of such behavior are still under investigation.

Similar heat generation statistics could be evaluated for different transistor materials (e.g., germanium) as well as for strained or confined nanostructures. Restricted dimensions may lead to confined phonon modes [23], but the electron–phonon scattering rate can be recomputed taking into account the modified phonon dispersion. This can be done efficiently as long as the phonon dispersion is recast

as a set of analytic functions (e.g., polynomials). This Monte Carlo tool (named MONET) will be made available through the NSF/NNI Computational Nanohub [24].

III. HEAT CONDUCTION

The energy gained by charge carriers moving along the electric field is given up to the lattice by means of the electron–phonon scattering, as described above. The conduction of heat from the device layer to the heat sink is impeded by an overall thermal resistance of approximately 0.6 K/W [4] in the 90-nm bulk silicon technology (for the entire chip). A simple thermal resistance circuit for heat flow from a MOSFET is shown in Fig. 7. The fundamental objective of a cooling solution is to increase heat transfer by increasing the surface area and thereby spreading heat. This necessitates the use of integrated heat spreaders and external heats sinks or heat pipes. The dominant contribution to the thermal resistance typically lies within the package. However, it must be noted that the contribution from outside the package (due to the heat pipe, the heat sink, and convection to the ambient) is equally important in space-constrained mobile applications. The contribution from the silicon die R_{DIE} is approximately 0.2 K/W. The resistance to one-dimensional heat flow across the bulk silicon is only 0.06 K/W. However, the nonuniformity of heat generation inherent in microprocessors increases this resistance by nearly three times. The transistor component R_{INT} arises in silicon-on-insulator devices due to the presence of the buried oxide. The magnitude depends on the dimensions of an individual device. For a 22-nm gate length ultrathin body SOI device, the device thermal resistance will be between 0.1 and 1 K/μW per transistor. If we assume a duty factor of 10% for 100 million transistors, the cumulative SOI resistance R_{INT}, is between 0.01 and 0.1 K/W. The exact number will depend on the geometry, material properties, and interfacial resistances. The higher end figure is of the same order of magnitude as the total thermal resistance today and is worrisome. We believe that an understanding of the heat conduction processes in the vicinity of transistors will help in minimizing this contribution.

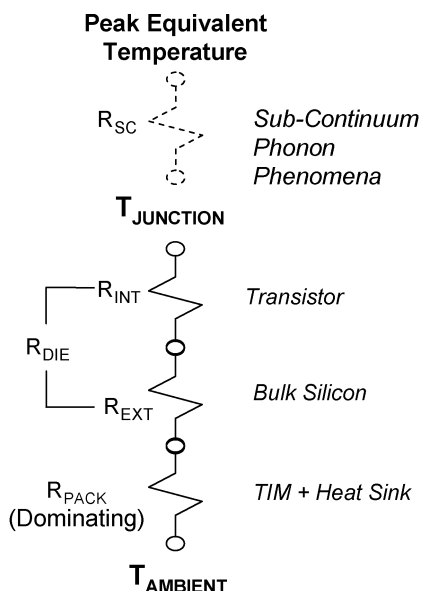

Fig. 7. ***Thermal resistance circuit for heat flow from a transistor. The package side resistance R_{PACK} dominates heat conduction in modern transistors (TIM represents the thermal interface materials). The subcontinuum component R_{SC}, will become increasingly larger in future transistors.***

The diffusion time constant for a heat pulse initiated at the drain of a 90-nm gate length SOI device to reach the bulk silicon is about 100 ns [25]. However, heat dissipation in digital CMOS transistors primarily occurs over a period of several tens of picoseconds, during a switching event [26]. Thus, the temperature field within a transistor is quasi-steady during the on-state of the device. The dynamic on-state peak temperature is not predicted by the resistance circuit of Fig. 7, which only provides an estimate of the average value over several million switching cycles. The silicon lattice comprising the body of the transistor is driven out of equilibrium during each switching period. The length and time scales are too small to permit a continuum assumption for heat flow. In the subcontinuum regime, we view heat conduction as the transport of phonons inside the transistor. The concept of a temperature is meaningless under such nonequilibrium conditions. It is more useful to consider the distribution of phonons within the transistor, that is, a "phonon field" instead of the usual temperature field. The distribution of phonons can be converted to an "equivalent temperature" field by equating the local energy density due to the nonequilibrium distribution to that based on a Bose–Einstein distribution. The temperature used in the Bose–Einstein distribution to provide the same local energy density is then the equivalent temperature. We note here that the quasi-steady continuum temperature field is still useful while describing global chip conditions, such as in circuit simulations. However, it must be treated with caution when predicting or understanding the behavior of an individual device.

The phenomenological phonon Boltzmann transport equation (BTE) predicts the evolution of phonon distributions. Under the relaxation time approximation, the equation for the phonon mode with wave vector $\mathbf{q}$ and polarization s is written as

$$\frac{\partial N_{\mathbf{q},s}}{\partial t} + v \cdot \nabla N_{\mathbf{q},s} = -\frac{N_{\mathbf{q},s} - \overline{N}_{\mathbf{q},s}}{\tau_{\mathbf{q},s}} + N_{\mathbf{q},s}^{e-p} \tag{3}$$

where N is the distribution function, v is the group velocity and τ is the single-mode relaxation time. The Bose–Einstein equilibrium distribution function $\overline{N}$ is evaluated

at the thermodynamic temperature of the lattice. However, the relaxation time must be evaluated at the local equivalent temperature for the approximation to work. The last term on the right is due to electron–phonon scattering and represents the net generation of phonons. The BTE under the relaxation time approximation does not contain enough information in itself to yield solutions that conserve energy implicitly. Thus, an additional *macroscopic* energy conservation equation must be constructed and solved along with the above system of BTE equations for each of the phonon modes. We clarify here that the macroscopic moment equation does not conserve energy for each phonon scattering event but only ensures a global energy conservation. An efficient treatment that accounts for microscopic energy conservation is yet to be developed. Instead, the conservation equation is constructed from an energy moment of the BTE and has the following form:

$$C\frac{\partial T}{\partial t}+\frac{\partial E_{SC}}{\partial t}=H-\nabla\cdot(-\overleftrightarrow{K}\,\nabla T+\vec{J}_{SC}) \qquad (4)$$

where C is the volumetric heat capacity, H is the volumetric heat generation rate as computed in the previous section, and $\overleftrightarrow{K}$ is the thermal conductivity tensor. The subscript "SC" marks the subcontinuum terms in the equation, with E being the energy density and $\vec{J}$ being the heat current due to unrelaxed phonons. In the absence of these subcontinuum terms, (4) is the usual heat conduction equation.

While several subcontinuum models based on the phonon BTE under the relaxation time approximation are available in the literature, most of them suffer from several deficiencies when it comes to device simulations. We refer the reader to [27] for an in-depth discussion of the strengths and limitations of several models. The models of particular interest to device simulation are the equation for phonon radiative transport (EPRT) [28], the ballistic-diffusive equations [29], energy moment formulations [30], and the split-flux model [31], [32]. Recent work has shown that it is necessary to include a phonon dispersion model accounting for all phonon branches. Fig. 8 shows the equivalent temperature field in an 18-nm gate length silicon-on-insulator (SOI) device computed using the split-flux model. More details of this computation are provided in [33].

Past work shows that nonequilibrium phonon populations serve to increase the thermal resistance locally in the drain of a transistor, compared to predictions based on heat diffusion theory. The subcontinuum resistance is approximately proportional to the peak value of the volumetric power density in the device. The lower bound on the peak power density can be taken as the product of the duty factor, typically 0.1, and the saturation peak power

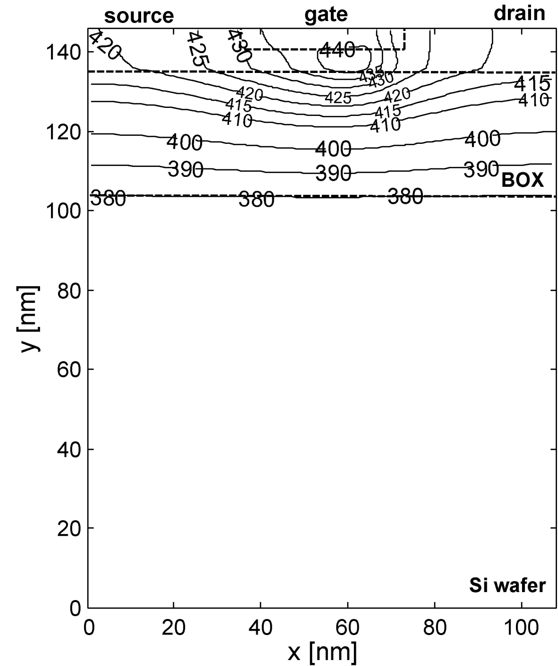

Fig. 8. ***Contours of equivalent temperature in an 18-nm gate length SOI device computed using the split-flux model [31]–[33]. The raised source is on the left and the raised drain is on the right-hand side at the top of the figure. The rectangle at the top enclosed by the three dashed lines is the gate region. The buried oxide (BOX) layer extends approximately from 100 to 136 nm on the y axis as indicated by the dashed lines, and below is the rest of the silicon wafer.***

density. The lower bound on the increase in resistance is about 13% in a 90-nm gate length device and about 30% in a 18-nm gate length device [33]. Since the volumetric power density increases gradually in future technologies following ITRS guidelines [1], we expect the resistance to result in higher peak "temperatures" in the drain than predicted by the diffusion theory. The increase in the volumetric power density in future nodes is shown in Fig. 9(a). We have assumed a 17% clock frequency increase per year to estimate the dynamic power. The rest of the numbers follow from stock ITRS projections.

Fig. 9(b) shows the estimated temperature rise at future device technology nodes based on the power density numbers in Fig. 9(a). We include both bulk silicon and SOI MOSFETs. Unless otherwise indicated, the graphs represent the average temperature rise in the device. We assume a normalized junction to ambient thermal resistance of 0.6 Kcm^2/W. In order to estimate the actual resistance we have to account for the spatial nonuniformity in the heat source distribution. This is achieved by means of a density factor which must be multiplied to the normalized resistance. To obtain the temperature in an SOI device, we add an extra thermal resistance to that of the corresponding bulk MOSFET. This additional resistance depends on the geometry of the SOI device, in particular

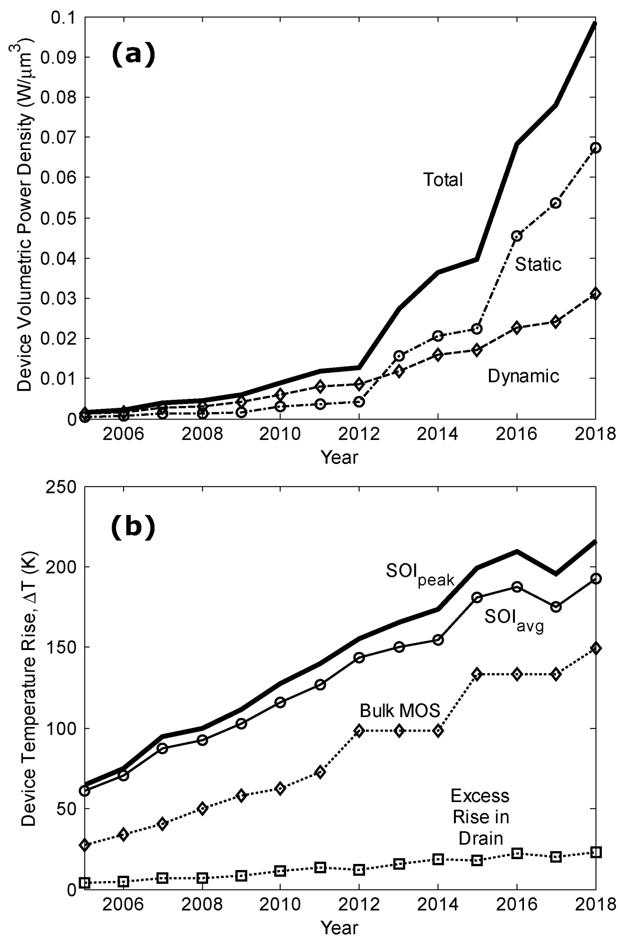

Fig. 9. ***(a) The power dissipated per unit volume in a device. This increases by an order of magnitude within a decade as per ITRS figures. (b) The average temperature rise in a device under saturation conditions is shown for generic bulk and silicon-on-insulator MOSFETs. The temperature of the drain contact is consistently higher in devices, especially SOI, due to prevailing thermal nonequilibrium in the lattice. This contribution is shown toward the bottom. The peak temperature is obtained by adding this to the average. This is shown for SOI devices.***

the dimensions of the contacts and the channel. We assume the channel thickness to be half the printed gate length. Further, we only consider one-dimensional heat flow along the contacts and employ results from the theory of heat conduction for extended surfaces [34], [35]. Finally, we plot the excess temperature rise in the drain contact of a device due to thermal nonequilibrium. The phonon bath in the drain is driven out of equilibrium when a current is flowing through the transistor. This results in a peak "temperature" that exceeds the temperature predicted assuming heat diffusion. In order to calculate the excess temperature of the phonon bath in the drain, we assume a volumetric power density which is half of the saturation power density. This is based on the fact that devices in digital CMOS circuits are typically operated up until half of the saturation point [26]. Thus, the peak point in the power curve during a switching transient corresponds to a drive current that is approximately half the saturation current. The temperature rise is then evaluated using the phonon Boltzmann transport equation. Although the excess heating in the drain is insignificant at present, we find that this increases at future technology nodes. The peak temperature in SOI devices shown in Fig. 9(b) includes this excess contribution. However, the principal contribution to the increase in thermal resistance is due to the reduction in the layout area of the device and the switch to SOI structures.

We note here that all the above calculations are based on the assumption of a bulk phonon dispersion relationship. While we do expect the vibrational modes in an ultrathin silicon film to be different from those of the bulk, we note that such a film does not exist in isolation in a device but is always structurally coupled to the substrate. Hence, the phonon modes of the silicon channel will be coupled to those of the source, the drain and the oxide underneath. This should lead to a complex dispersion relation with even long wavelength modes exhibiting nonlinear dispersion. It is possible to obtain some idea of the dispersion relationship for long wavelength modes under a continuum assumption using the elasticity equation. We refer the reader to the literature on nanoelectromechanical systems (NEMS) for existing calculations on phonon modes in thin-film structures [36]. Atomistic lattice dynamics of a thin silicon film is much more expensive to solve than the bulk since the simplifying assumptions based on lattice periodicity cannot be applied to a silicon unit cell along the thickness of the film. Hepplestone and Srivastava [37] have used plane wave pseudopotentials under the local density approximation to calculate the phonon modes. However, these calculations do not account for coupling with the substrate.

IV. MATERIAL THERMAL PROPERTIES

Traditional, bulk transistor designs [Fig. 2(a)] have typically incorporated only a few materials, most notably silicon, silicon dioxide insulators and silicided (e.g., NiSi) contacts. The high thermal conductivity of bulk silicon (148 $\mathrm{Wm^{-1}K^{-1}}$) facilitates heat transport from the transistor channel down to the backside of the chip, where it has been traditionally removed with a heat sink, as described above.

Advanced device fabrication introduces a number of new materials with lower thermal conductivities. The thermal properties of these materials are therefore expected to play a more significant role in device design and thermal behavior. Table 1 lists the thermal conductivities of a few materials used in semiconductor device fabrication. Bulk germanium transistors, for example, may suffer from increased operating temperatures due to a substrate thermal conductivity approximately 60% lower than bulk

silicon transistors. Strained silicon channel devices [Fig. 2(b)] grown on a graded $Si_{1-x}Ge_x$ buffer layer benefit from an increased mobility due to band degeneracy splitting and a lighter effective mass in the strained film. However, their thermal behavior is adversely affected by the lower thermal conductivity of the $Si_{1-x}Ge_x$ alloy layer. The situation is worse for thin-body devices grown on a silicon dioxide film. From an electrical point of view, SOI devices benefit from lower capacitive coupling with the substrate, and hence increased switching speeds. Thermally, however, they are significantly affected by the very low thermal conductivity of the buried oxide layer, which is about two orders of magnitude less than that of silicon [25]. The thermal conductivity of thin semiconductor films (thinner than the phonon mean free path) is also significantly reduced by phonon boundary scattering. The thermal conductivity of a 10-nm-thin silicon film is expected to be reduced by an order of magnitude from that of bulk silicon [5]. Although experimental data for such thin films does not yet exist, an estimate can be based on extrapolations to the latest available data [38], and also supported by recently measured reduced thermal conductivities in silicon nanowires [39]. For film or nanowire thicknesses on the order of a few nanometers, and therefore comparable to the phonon wavelength, phonon confinement effects may become important as well, further contributing to a degradation in thermal conductivity [23].

The small dimensions of future device designs also imply a large surface-to-volume ratio (Fig. 2), and hence a stronger effect of material boundary resistance. Few estimates exist on the magnitude of the *thermal* boundary resistance between dissimilar materials (e.g., a dielectric and a metal) [40], [41]. The origins of this resistance are not well understood, but it is likely caused by phonon dispersion mismatch, heat carrier (electron–phonon) energy conversion, and atomic-scale surface roughness or porosity at the interface between the two materials [42]. Some measurements indicate it is on the order of the thermal resistance of a 20-nm-thick silicon dioxide film [43], and fairly independent of processing conditions or the specific type of metal and dielectric involved. This is a significant value for nanometer scale devices, and very important to understand. As more materials (e.g., high-k dielectrics, germanium, various silicides) are introduced in semiconductor processing, there is a growing need to understand the magnitude of boundary thermal resistance and its significance in nanoscale device behavior. The boundary thermal resistance may play a significant role, for example, when a metal electrode is placed on top of the high-k gate dielectric (as expected for threshold voltage control in future technologies) as well as for device metal contacts and interconnects. More measurements are necessary in this area, while more available data on thermal boundary resistance would also help toward a better understanding and modeling of the atomic scale interaction at the interface between two materials.

V. BULK SILICON TRANSISTORS

The bulk CMOS silicon transistor [Fig. 2(a)] has been a building block of the semiconductor industry over the past 40 years. The high thermal conductivity of bulk silicon facilitates heat transport from the transistor channel down to the backside of the chip, where it is usually removed with a heat sink. Thermal transport in bulk transistors has traditionally been modeled in the classical limit, as subcontinuum effects can be neglected for device dimensions larger than the phonon mean free path.

As devices are scaled in the deca-nanometer range, subcontinuum heat conduction results in nonequilibrium phonon populations as discussed above. The equivalent temperatures computed using subcontinuum transport models exceed the temperatures predicted by the heat diffusion equation. It is important to note that in devices where both charge and heat transport fall well below the continuum regime, these two transport phenomena are inherently coupled. Thus, investigating only one of them is unlikely to yield accurate answers. It will be difficult to predict the exact effect of subcontinuum heat conduction on device characteristics without a coupled simulation of electrons and phonons. The effort thus far has been to develop subcontinuum heat conduction models, and only the most recent efforts have begun targeting coupled transport models [21], [44].

For the present, it is not yet clear how the elevated equivalent temperature in the drain may affect the device current. Some indications exist that in the limit of the smallest achievable bulk silicon FETs (10 nm) the optical phonons generated in the drain may reach the source before decaying into acoustic phonons, directly affecting the source injection velocity and consequently the ultimately achievable current drive of the device [45]. On another look, any expression for the drive current based on the assumption of a thermodynamic equilibrium temperature will clearly underestimate the effect. However, one result that emerges from our investigations on subcontinuum heat conduction in devices is that the perturbation in the phonon population cannot be ignored. The implications of elevated equivalent temperatures may also be more significant for bulk transistors due to the greater temperature sensitivity of the leakage currents in these devices. Furthermore, since the bulk transistor subcontinuum effects are expected to take place in the drain region, their effect is also likely to be pronounced on device reliability.

VI. THIN-FILM TRANSISTORS

Thin-film transistors are partially or fully depleted single-gate SOI devices, planar (buried gate) dual-gate devices, or vertical FinFETs, as in Fig. 2(b) and 2(c). Multigate transistors also fall essentially in the same category as FinFETs from the thermal point of view. Such devices

benefit from lower capacitive coupling with the substrate, hence higher switching speeds, and more gate control over the channel, hence better turn-on characteristics and threshold voltage control. The subcontinuum phenomena present in bulk nanotransistors are expected to play a much lesser role in thin-film transistors. This occurs because phonon-boundary scattering reduces the average phonon mean free path in the thin film to a scale comparable to the device dimensions and the size of the drain-side phonon hot spot.

However, unlike for bulk devices the thermal conductivity of the active device region, i.e., the thin-film body, is much lower (see Fig. 10). This thermal conductivity is directly proportional to the average phonon mean free path [46], which is strongly affected by boundary scattering. In addition, a larger surface area to volume ratio for ultra-scaled SOI devices also implies a stronger contribution from the thermal boundary resistance of the various material interfaces. Finally, the very low thermal conductivity of the buried oxide significantly impedes heat transfer toward the substrate. Heat dissipation through the device contacts may play an important role, especially for the smallest, scaled devices. These effects have been captured in the thermal resistance model introduced in [47], showing that device thermal properties are strongly dependent on the transistor geometry, particularly on the dimensions of the channel extension length (L_{ex}) and the source/drain height (t_{SD}). A shorter extension and elevated source/drain are desirable from a thermal point of view, since this choice would minimize the thermal conduction path toward the device contacts and increase the volume available for heat dissipation. These device geometry tradeoffs are schematically shown in the Fig. 11(b) image. While a raised source/drain is also desirable from an electrical point of view to minimize electrical series resistance, a shorter extension length can lead to significant parasitic gate-to-drain capacitance, adversely affecting the switching speed of such a device [5]. The impact can be quantified by considering the intrinsic gate delay (CV/I) for several technology nodes towards the end of the Semiconductor Roadmap [1], as illustrated in Fig. 11(a). Through self-consistent electrothermal simulations we have found that a raised source/drain thicker than approximately three times the device body does not yield any additional speed gain. This is due to an increase in the parasitic capacitance C that cannot be offset by any more improvements in drive current $I(T)$, even as the device temperature is lowered by a larger source/drain. The same studies have also suggested that the optimal extension length L_{ex} ought to scale approximately as half the physical device gate length for such well-behaved electrothermally optimized thin-body devices [5].

Another electrical versus thermal tradeoff exists for the thickness of the buried oxide, t_{BOX}: A thinner oxide is desirable from a thermal point of view, although electrically this would lead to higher parasitic coupling capacitance with the substrate. Similar arguments can be made about isolated, vertical FinFET devices to show that they generally have better thermal properties owing to their thicker (hence less thermally resistive) body, larger oxide area, and overlapped gate [47]. However, a FinFET device required to deliver a larger current will generally contain several parallel fins. In this case, the fins in the middle will be at a more elevated temperature than the outside ones, and this additional temperature rise will depend on the fin spacing [48].

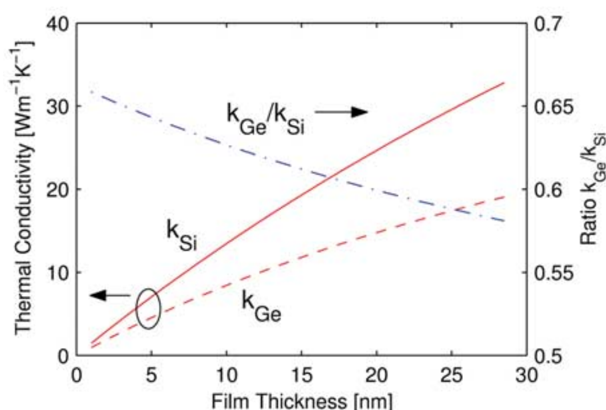

Fig. 10. ***Estimated reduction in the thermal conductivity of silicon and germanium thin films due to phonon boundary scattering alone (i.e., no confinement). Thin silicon films suffer from a proportionally stronger reduction in thermal conductivity (versus germanium) due to the longer bulk silicon phonon mean free path [5].***

VII. GERMANIUM TRANSISTORS

Germanium field effect transistors (FETs) were largely abandoned 40 years ago because germanium lacks a reliable, native insulator like SiO_2 for silicon. However, with the recent introduction of compatible high-k gate dielectrics, the existence of a viable germanium transistor has become a renewed possibility [49]. Such devices are attractive because germanium has higher low field mobility than silicon (both for electrons and holes) which is thought to determine the ultimate current drive of a device [50]. From a thermal point of view, however, large-scale integration with bulk germanium transistors is more problematic because germanium has a thermal conductivity only about 40 percent that of silicon.

On the other hand, recent calculations suggest that ultrathin-film germanium films take on a lesser penalty in their thermal conductivity reduction than thin silicon films, as in Fig. 10. The longer bulk phonon mean free path of silicon also implies a stronger effect of boundary scattering on the thin-film thermal conductivity. As the film thickness is reduced, thin crystalline germanium films become more attractive (thermally) compared to silicon

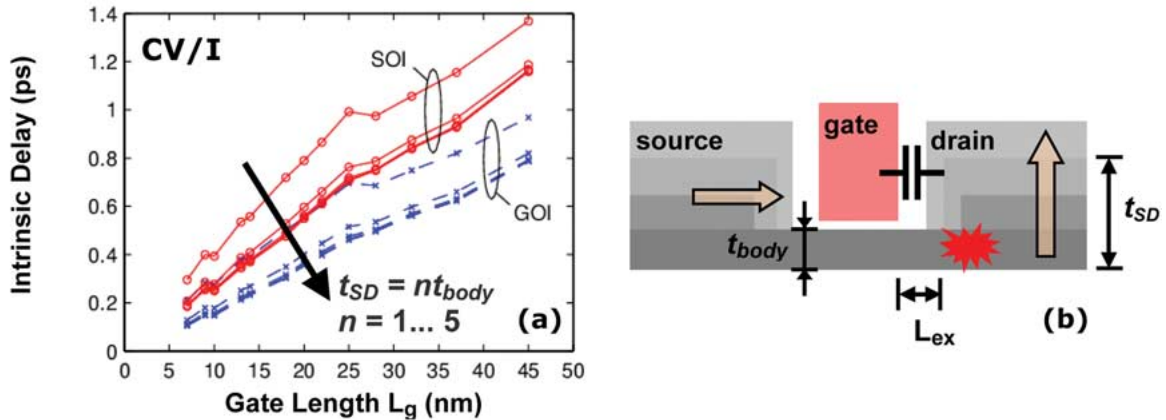

Fig. 11. *Electrothermally self-consistent comparative study of "well-behaved" ultrathin body SOI and GOI devices near the limits of scaling [5]. A larger (e.g., elevated) source and drain design will alleviate heat dissipation problems, but eventually lead to an increase in parasitic capacitance which reduces the intrinsic speed gain. The heat dissipation is assumed to be entirely in the drain, as suggested by Monte Carlo simulations.*

films versus the bulk case. The thermal conductivity ratio between the two material films is also plotted in Fig. 10. The thermal conductivity ratio between bulk germanium and silicon is approximately 0.4 (see Table 1), which increases to 0.65 in favor of germanium, for the thinnest films of practical interest. On the other hand, germanium devices also have a higher mobility, which is less sensitive to temperature compared to silicon [5]. The higher mobility implies that between otherwise similar, well-behaved GOI and SOI devices, the GOI devices can support the same amount of current (I) at a lower operating voltage (V). Moreover, the total dissipated power ($P = IV$) is expected to be lower for such GOI devices. Owing to their less temperature-sensitive mobility, and lower power dissipation, well-designed GOI devices are expected to maintain their performance advantage over similar SOI devices, despite the lower thermal conductivity of the thin germanium layer [5]. Fig. 11 quantifies and compares such trends for devices in the deca-nanometer range, and shows that electrothermally optimized GOI devices should maintain a 30% speed advantage over comparably well-behaved SOI. This thermal behavior is an additional reason making GOI technology an interesting alternative to SOI for the smallest thin-film devices at the limits of the technology roadmap. The lower fabrication temperatures of GOI devices also make them an attractive choice for three-dimensional integrated circuits.

VIII. SEMICONDUCTOR NANOWIRES

Transistors made from semiconductor nanowires (NWs) have recently received a lot of attention for their current-carrying capabilities, low synthesis temperature (as low as 275 °C for CVD-grown germanium NWs [51]), and relative ease of fabrication and integration with currently existing technologies [52], [53]. Nanowires also present an ideal vehicle for studying (low-dimensional) electronic and thermal transport at nanometer length scales, as well as coupled electrothermal transport. There are few data available on the mobility and thermal conductivity of nanowires, but it is strongly believed that confined electron *and* phonon conduction plays an important role in these devices. Recently measured silicon NW thermal conductivity suggests it is on the order of 5 $Wm^{-1}K^{-1}$ for wires of 20 nm diameter [39], which is predictably lower than the thermal conductivity of comparably thin silicon films [38], owing to additional phonon-boundary scattering with the surrounding wire perimeter.

Suspended semiconductor nanowires also represent the extreme limit of the electrothermally confined FinFET or surround-gate devices [54]. Their low thermal conductivity combined with their good current-carrying ability imply tight and possibly limiting coupling between electrons and the lattice, especially at high current levels. In other words, although such devices are currently known to be adversely affected by poor contact resistance, in practice their ultimate performance may be intrinsically limited by self-heating. This heat may be effectively dissipated through the surrounding gate or its contacts, but the thermal interface resistance between the wire, dielectric and gate has not yet been measured. Future work will also need to determine the thermal resistance of the contacts, along with the thermal crosstalk between closely packed parallel wires, as would be the case in a realistic device. However, it is possible that with controlled growth on a large scale, and a solid theoretical understanding from an electrothermal point of view, nanowires could possibly complement (although not necessarily displace) currently existing CMOS technology.

IX. CARBON NANOTUBES

Unlike nanowires, carbon nanotubes (CNTs) cannot be synthesized at relatively low temperatures, but rather only in the 500–1000 °C range, depending on the method and catalyst used [55]. However, both their electrical and thermal conductivity are outstanding, such that CNT-based devices have been thought to represent the ultimate limit of nanotransistors. Single-walled carbon nanotubes (SWNTs) grown on insulating substrates exhibit quantized, ballistic transport at low temperature and low bias, and extremely high current-carrying ability at high bias, up to 100 μA through very short tubes approximately 15 nm in length and 2 nm diameter [56]. This represents a current density greater than 10^{10} A/cm^2, more than two orders of magnitude higher than current densities achievable in other good conductors like Cu or Al.

With the introduction of ohmic Pd contacts, CNTs have been brought one step closer to realizing their potential device applications [57]. The existence of good quality contacts has also enabled studies which have shown that for nanotubes much longer than the electron-optical phonon (OP) mean free path (MFP), high-bias electron transport is not truly ballistic [58]. Despite the ballistic nature of CNTs at low-bias (where the intrinsic resistance is close to $h/4e^2 \approx 6.5$ kΩ), strong electron-OP scattering dominates high-field transport, when electrons can attain energies more than 0.16 eV, the zone-boundary optical phonon energy [59]. In general, the electron–acoustic phonon interaction in CNTs is assumed to be elastic (electrons are simply backscattered, but their energy is not changed) while the OP interaction is assumed to be strongly inelastic, such that electrons lose $\hbar\omega_{OP} \approx 0.16 - 0.20$ eV energy during an emission process [59]. Recent work has shown that despite their high thermal conductivity [60], the thermal *conductance* of SWNTs is relatively low, owing to their small diameter [61], [62]. The consequences of this low conductance are most evident at high-bias in *suspended* SWNTs [61], as shown in Fig. 12.

While even several micrometer-long SWNTs on insulating substrates exhibit very high (> 20 μA) current-carrying capability [59], we have found that freely suspended nanotubes carry much lower currents due to significant self-heating. The suspended nanotube resistance at high bias in Fig. 12 is greater than expected near $T \sim 600°$ C (the burning temperature of SWNTs in air), suggesting a lower lattice temperature but a higher non-equilibrium, hot OP population [61]. This observation is consistent with recent studies indicating much longer phonon lifetimes in suspended SWNTs [63]. This is attributed to the lack of intimate coupling with a substrate, which would otherwise provide additional phonon relaxation channels. Mann *et al.* have also shown that even gaseous "substrates" in contact with suspended SWNTs can serve to reduce hot OP phonon lifetimes and therefore deliver high-bias currents larger than in vacuum, but lower than for SWNTs on solid substrates [64]. Freely suspended, unperturbed SWNTs in vacuum seem to represent the extreme case of a thermally isolated and long phonon lifetime scenario for nanotubes. All such suspended SWNTs exhibit negative differential conductance (NDC), but the longer ones (10 μm) show it at much lower electric fields (~200 V/cm) than predicted by previous theoretical models which assume isothermal conditions (~5 kV/cm) [65]. This also indicates that the observed NDC is a thermal and not electrical (e.g., contact or field-related) effect.

A simple two-temperature (corresponding to acoustic and optical phonons) model can be used to calculate the theoretical *I-V* characteristics (solid lines) in Fig. 12 [61]. The optical phonons are assumed to be stationary, and the acoustic phonons are solely responsible for heat dissipation. The approach self-consistently computes the nanotube resistance, Joule heating and temperature along its

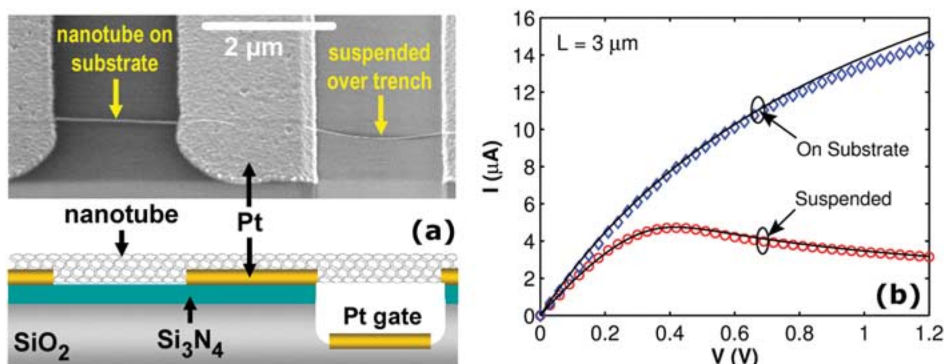

Fig. 12. *SEM (top left) and diagram (bottom left) of a metallic SWNT with one portion grown across a substrate, the other suspended across a trench. The measured (symbols) and calculated (lines) I-V characteristics of the two nanotube segments are also plotted for a similar device with length 3 μm and diameter 2.4 nm (right) [61].*

length. A key feature is the temperature dependence of the SWNT thermal conductivity, which is found to be essential for reproducing the high bias behavior of the electrical characteristics. The approximately $1/V$ shape of the suspended SWNT *I-V* characteristics at high bias is found to be a reflection of the $1/T$ dependence of the thermal conductivity at high temperature from Umklapp-limited phonon scattering. This provides an indirect way to measure the thermal conductivity of individual suspended SWNTs in the high temperature regime, indicating it to be approximately described by $3600(300/T)$ $\mathrm{Wm^{-1}K^{-1}}$ from $400 < T < 700$ K. A more detailed study of the thermal conductance of individual SWNTs is described in [60].

SWNTs on insulating substrates have also been recently shown to exhibit self-heating at high bias, although the presence of nonequilibrium hot phonons in this case is less certain. Researchers have been able to reproduce on-substrate *I-V* characteristics by including a simple Joule heating model, without the assumption of phonon nonequilibrium [62], [66], although some arguments to the contrary also exist [67]. It is, however, experimentally evident that self-heating at high-bias is indeed sufficient to burn SWNTs in air, the breakdown voltage scaling approximately as 5 V/μm for tubes longer than 1 μm, and suggesting a SWNT-to-substrate thermal conductance of about 0.15 $\mathrm{WK^{-1}m^{-1}}$ per nanotube length [62]. For much shorter tubes, it appears that high-bias current enhancement is aided by increased heat sinking through the contacts [66] rather than along the length of the tube itself. With these recent advances in understanding electrical *and* thermal transport in SWNTs, it remains yet to be seen if engineering advances can be made to optimize both types of transport, and consequently yield more reliable and efficient carbon nanotube device applications.

X. CONCLUSION

Heat transfer research for integrated electronics has traditionally focused on transport off the semiconductor chip, including air convection from the heat sink and conduction in the package. However, ITRS trends and the introduction of novel device technologies are projected to lead to significant thermal bottlenecks at the transistor and circuit level. The average device temperature will continue to be dominated by circuit architecture and packaging design choices. On the other hand, the temperature distribution in the active device region will be governed by subcontinuum electrothermal transport. The peak temperature in future bulk devices is likely to increase significantly due to subcontinuum effects driven by increasing volumetric power densities. Innovative experimental research is still required to fully understand the impact of such effects on electronic device transport.

The shift towards novel, confined device geometries (ultrathin body, nanowires) and lower thermal conductivity materials (e.g., germanium) will bring additional complexity to thermal management within individual devices and circuits. This situation motivates a new approach to device design, focused on nanoscale conduction physics, coupled electron and phonon transport modeling, and electrothermal transistor design optimization. Both *power* and *contact* issues are going to dominate novel device and circuit behavior. Material interfaces and device contact thermal resistance are going to play a more significant role as device volumes diminish and the surface-to-volume ratio increases. Heat will be generated almost entirely in the drain and near the drain contact of quasi-ballistic device technologies. Consequently, larger drain and contact designs may be used to provide an additional thermal cooling path at the device level, while large interlayer vias may be used for thermal management in 3-D circuits. It is yet unclear whether surround-gate devices will benefit from the additional cooling path through the gate, depending on the gate material and the interface thermal resistance. It is however apparent that the small, restricted dimensions of future devices will affect both current and heat transport, while thermal crosstalk is expected to be of significance between closely packed devices, device fins, or nanowire/tube segments. ■

REFERENCES

[1] International Technology Roadmap for Semiconductors (ITRS). [Online]. Available: http://www.public.itrs.net

[2] S. Borkar, "Low power design challenges for the decade," in *Proc. Design Automation Conf.* 2001, pp. 293–296.

[3] C. C. Liu, J. Zhang, A. K. Datta, and S. Tiwari, "Heating effects of clock drivers in bulk, SOI, and 3-D CMOS," *Electron Device Lett.*, vol. 23, pp. 716–718, 2002.

[4] R. Mahajan, R. Nair, V. Wakharkar, J. Swan, J. Tang, and G. Vandentop, "Emerging directions for packaging technologies," *Intel Technol. J.*, vol. 6, pp. 62–75, 2002.

[5] E. Pop, C. O. Chui, S. Sinha, K. E. Goodson, and R. W. Dutton, "Electro-thermal comparison and performance optimization of thin-body SOI and GOI MOSFETs," in *Proc. IEEE Int. Electron Devices Mtg. (IEDM)*, 2004, pp. 411–414.

[6] M. Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[7] Y. S. Ju and K. E. Goodson, "Phonon scattering in silicon thin films with thickness of order 100 nm," *Appl. Phys. Lett.*, vol. 74, pp. 3005–3007, 1999.

[8] E. Pop, R. W. Dutton, and K. E. Goodson, "Monte Carlo simulation of Joule heating in bulk and strained silicon," *Appl. Phys. Lett.*, vol. 86, p. 082101, 2005.

[9] E. Pop, J. Rowlette, R. W. Dutton, and K. E. Goodson, "Joule heating under quasi-ballistic transport conditions in bulk and strained silicon devices," in *Proc. Int. Conf. Simulation of Semiconductor Processes and Devices (SISPAD)*, 2005, pp. 307–310.

[10] D. Esseni, M. Mastrapasqua, G. K. Celler, C. Fiegna, L. Selmi, and E. Sangiorgi, "Low field electron and hole mobility of SOI transistors fabricated on ultrathin silicon films for deep submicrometer technology application," *IEEE Trans. Electron Devices,* vol. 48, no. 12, pp. 2842–2850, Dec. 2001.

[11] D. K. Ferry, *Semiconductor Transport*. New York: Taylor & Francis, 2000.

[12] M. Artaki and P. J. Price, "Hot phonon effects in silicon field-effect transistors," *J. Appl. Phys.*, vol. 65, pp. 1317–1320, 1989.

[13] G. K. Wachutka, "Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling," *IEEE*

Trans. Comput-Aided Design Integr. Circuits Syst., vol. 9, no. 11, pp. 1141–1149, Nov. 1990.

[14] 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.*, vol. 96, p. 4998, 2004.

[15] D. Long, "Scattering of conduction electrons by lattice vibrations in silicon," *Phys. Rev.*, vol. 48, p. 2024, 1960.

[16] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys.*, vol. 55, pp. 645–705, 1983.

[17] S. Sinha, P. K. Schelling, S. R. Phillpot, and K. E. Goodson, "Scattering of g-process LO phonons at hotspots in silicon," *J. Appl. Phys.*, vol. 97, p. 237002, 2005.

[18] S. Mazumder and A. Majumdar, "Monte Carlo study of phonon transport in solid thin films including dispersion and polarization," *J. Heat Transf.*, vol. 123, pp. 749–759, 2001.

[19] S. V. J. Narumanchi, J. Y. Murthy, and C. H. Amon, "Submicron heat transport model in silicon accounting for phonon dispersion and polarization," *J. Heat Transf.*, vol. 126, pp. 946–955, 2004.

[20] T. Mietzner, J. Jakumeit, and U. Ravaioli, "Local iterative Monte Carlo analysis of electron–electron interaction in short-channel Si-MOSFETs," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 2323–2330, Apr. 2001.

[21] J. Rowlette, E. Pop, S. Sinha, M. Panzer, and K. E. Goodson, "Thermal simulation techniques for nanoscale transistors," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, 2005, pp. 225–228.

[22] J. Lai and A. Majumdar, "Concurrent thermal and electrical modeling of sub-micrometer silicon devices," *J. Appl. Phys.*, vol. 79, pp. 7353–7361, 1996.

[23] A. Balandin and K. L. Wang, "Significant decrease of the lattice thermal conductivity due to phonon confinement in a free standing semiconductor quantum well," *Phys. Rev. B*, vol. 58, pp. 1544–1549, 1998.

[24] NSF/NCN Nanotechnology Simulation Hub. [Online]. Available: http://www.nanohub.org

[25] S. Polonsky and K. A. Jenkins, "Time-resolved measurements of self-heating in SOI and strained-silicon MOSFETs using photon emission microscopy," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 208–210, Apr. 2004.

[26] K. A. Jenkins and R. L. Franch, "Impact of self-heating on digital SOI and strained-silicon CMOS circuits," in *Proc. IEEE Int. SOI Conf.* 2003, pp. 161–163.

[27] S. Sinha and K. E. Goodson, "Review: multiscale thermal modeling in nanoelectronics," *Int. J. Multiscale Comput. Eng.*, vol. 3, pp. 107–133, 2005.

[28] A. Majumdar, "Microscale heat conduction in dielectric thin films," *J. Heat Transf.*, vol. 115, p. 7, 1993.

[29] G. Chen, "Ballistic-diffusive heat conduction equations," *Phys. Rev. Lett.*, vol. 86, p. 2297, 2001.

[30] P. G. Sverdrup, Y. S. Ju, and K. E. Goodson, "Sub-continuum simulations of heat conduction in silicon-on-insulator transistors," *J. Heat Transf.*, vol. 123, pp. 130–137, 2001.

[31] S. Sinha, E. Pop, and K. E. Goodson, "A split-flux model for phonon transport near hotspots," presented at the Int. Mech. Eng. Congress and Expo (IMECE), Anaheim, CA, 2004.

[32] S. Sinha, E. Pop, R. W. Dutton, and K. E. Goodson, "Non-equilibrium phonon distributions in sub-100 nm silicon transistors," *J. Heat Transf.*, vol. 128, pp. 638–647, 2006.

[33] S. Sinha, "Non-equilibrium phonons in silicon field-effect transistors," Ph. D. dissertation in mechanical engineering, Stanford Univ., Stanford, CA, 2005.

[34] K. E. Goodson and M. I. Flik, "Effect of microscale thermal conduction on the packing limit of silicon-on-insulator electronic devices," *IEEE Trans. Compon., Hybrids, Manuf. Technol.*, vol. 15, no. 5, pp. 715–722, Oct. 1992.

[35] L. T. Su, J. E. Chung, D. A. Antoniadis, K. E. Goodson, and M. I. Flik, "Measurement and modeling of self-heating in SOI NMOSFETs," *IEEE Trans. Electron Devices*, vol. 41, no. 1, pp. 69–75, Jan. 1994.

[36] M. C. Cross and R. Lifshitz, "Elastic wave transmission at an abrupt junction in a thin plate with application to heat transport and vibrations in mesoscopic systems," *Phys. Rev. B*, vol. 64, p. 085324, 2001.

[37] S. P. Hepplestone and G. P. Srivastava, "Lattice dynamics of ultrasmall silicon nanostructures," *Appl. Phys. Lett.*, vol. 87, p. 231906, 2005.

[38] W. Liu and M. Asheghi, "Thermal conduction in ultrathin pure and doped single-crystal silicon layers at high temperature," *J. Appl. Phys.*, vol. 98, p. 123523, 2005.

[39] D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar, "Thermal conductivity of individual silicon nanowires," *Appl. Phys. Lett.*, vol. 83, pp. 2934–2936, 2003.

[40] R. J. Stoner and H. J. Maris, "Kapitza conductance and heat flow between solids at temperatures from 50 to 300 K," *Phys. Rev. B*, vol. 48, pp. 16373–16387, 1993.

[41] D. G. Cahill, W. K. Ford, K. E. Goodson, G. D. Mahan, A. Majumdar, H. J. Maris, R. Merlin, and S. R. Phillpot, "Nanoscale thermal transport," *J. Appl. Phys.*, vol. 93, pp. 793–818, 2003.

[42] A. Majumdar and P. Reddy, "Role of electron-phonon coupling in thermal conductance of metal-nonmetal interfaces," *Appl. Phys. Lett.*, vol. 84, pp. 4768–4770, 2004.

[43] T. Yamane, N. Nagai, S. Katayama, and M. Todoki, "Measurement of the thermal conductivity of silicon dioxide thin films using a 3-omega method," *J. Appl. Phys.*, vol. 91, pp. 9772–9776, 2002.

[44] J. Rowlette, E. Pop, S. Sinha, M. Panzer, and K. E. Goodson, "Thermal phenomena in deeply scaled MOSFETs," in *Tech. Dig. IEEE Int. Electron Devices Meeting*, 2005, pp. 984–987.

[45] E. Pop, K. Banerjee, P. G. Sverdrup, R. W. Dutton, and K. E. Goodson, "Localized heating effects and scaling of sub-0.18 micron CMOS devices," in *Tech. Dig. IEEE Int. Electron Devices Meeting (IEDM)*, 2001, pp. 31.1.1–31.1.4.

[46] J. M. Ziman, *Electrons and Phonons*. Oxford, U.K.: Oxford Univ. Press, 1960.

[47] E. Pop, R. W. Dutton, and K. E. Goodson, "Thermal analysis of ultra-thin body device scaling," in *Tech. Dig. IEEE Int. Electron Devices Meeting (IEDM)*, 2003, pp. 883–886.

[48] R. V. Joshi, J. A. Pasual-Gutiérez, and C. T. Chuang, "3-D thermal modeling of FinFET," presented at the Eur. Solid-State Device Research Conf. (ESSDERC), Grenoble, France, 2005.

[49] C. O. Chui, H. Kim, P. C. McIntyre, and K. C. Saraswat, "A germanium NMOSFET process integrating metal gate and improved high-k dielectric," in *Tech. Dig. IEEE Int. Electron Devices Meeting (IEDM)*, 2003, pp. 18.3.1–18.3.4.

[50] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 133–141, Jan. 2002.

[51] D. Wang, Q. Wang, A. Javey, R. Tu, H. Dai, H. Kim, P. C. McIntyre, T. Krishnamohan, and K. C. Saraswat, "Germanium nanowire field-effect transistors with SiO2 and high-k HfO2 gate dielectrics," *Appl. Phys. Lett.*, vol. 83, p. 2432, 2003.

[52] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, "High performance silicon nanowire field effect transistors," *Nano Lett.* vol. 3, pp. 149–152, 2003.

[53] S.-M. Koo, M. D. Edelstein, Q. Li, C. A. Richter, and E. M. Vogel, "Silicon nanowires as enhancement-mode Schottky barrier field-effect transistors," *Nanotechnology*, vol. 16, pp. 1482–1485, 2005.

[54] A. Tilke, L. Pescini, A. Erbe, H. Lorenz, and R. H. Blick, "Electron-phonon interaction in suspended highly doped silicon nanowires," *Nanotechnology*, vol. 13, p. 491, 2002.

[55] H. Dai, "Growth and characterization of carbon nanotubes," in *Carbon Nanotubes*, vol. 80, *Topics in Applied Physics*, M. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds. Berlin, Germany: Springer-Verlag, 2000.

[56] A. Javey, P. Qi, Q. Wang, and H. Dai, "Ten- to 50-nm-long quasi-ballistic carbon nanotube devices obtained without complex lithography," *Proc. Nat. Acad. Sci.*, vol. 101, p. 13408, 2004.

[57] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, "Ballistic carbon nanotube field-effect transistors," *Nature*, vol. 424, pp. 654–657, 2003.

[58] A. Javey, J. Guo, M. Paulsson, Q. Wang, D. Mann, M. Lundstrom, and H. J. Dai, "High-field quasi-ballistic transport in short carbon nanotubes," *Phys. Rev. Lett.*, vol. 92, p. 106804, 2004.

[59] Z. Yao, C. L. Kane, and C. Dekker, "High-field electrical transport in single-wall carbon nanotubes," *Phys. Rev. Lett.*, vol. 84, pp. 2941–2944, 2000.

[60] E. Pop, D. Mann, Q. Wang, K. E. Goodson, and H. J. Dai, "Thermal conductance of an individual single-wall carbon nanotube above room temperature," *Nano Lett.*, vol. 6, pp. 96–100, 2006.

[61] E. Pop, D. Mann, J. Cao, Q. Wang, K. E. Goodson, and H. J. Dai, "Negative differential conductance and hot phonons in suspended nanotube molecular wires," *Phys. Rev. Lett.*, vol. 95, p. 155505, 2005.

[62] E. Pop, D. Mann, J. Reifenberg, K. E. Goodson, and H. J. Dai, "Electro-thermal transport in metallic single-wall carbon nanotubes for interconnect applications," in *Tech. Dig. IEEE Int. Electron Devices Meeting (IEDM)*, 2005, pp. 253–256.

[63] B. J. LeRoy, S. G. Lemay, J. Kong, and C. Dekker, "Electrical generation and absorption of phonons in carbon nanotubes," *Nature*, vol. 432, p. 371, 2004.

[64] D. Mann, E. Pop, J. Cao, Q. Wang, K. E. Goodson, and H. J. Dai, "Thermally and molecularly stimulated relaxation of hot phonons in suspended carbon nanotubes," *J. Phys. Chem. B*, vol. 110, pp. 1502–1505, 2006.

[65] V. Perebeinos, J. Tersoff, and P. Avouris, "Electron–phonon interaction and transport in semiconducting carbon nanotubes," *Phys. Rev. Lett.*, vol. 94, p. 086802, 2005.

[66] M. A. Kuroda, A. Cangellaris, and J.-P. Leburton, "Nonlinear transport and heat dissipation in metallic carbon nanotubes," *Phys. Rev. Lett.*, vol. 95, pp. 266803–266806, 2005.

[67] M. Lazzeri, S. Piscanec, F. Mauri, A. C. Ferrari, and J. Robertson, "Electron transport and hot phonons in carbon nanotubes," *Phys. Rev. Lett.*, vol. 95, p. 236802, 2005.

ABOUT THE AUTHORS

Eric Pop received the M.Eng. and S.B., degrees in electrical engineering and the S.B. degree in physics from the Massachusetts Institute of Technology, Cambridge, in 1999 and the Ph.D. degree from Stanford University, Stanford, CA, in electrical engineering in 2005, supported by a joint SRC/IBM fellowship.

In 2005 he did postdoctoral research on the electrical and thermal properties of carbon nanotubes at Stanford. He is currently an Intel Researcher in Residence in the Stanford Center for Integrated Systems. He is interested in charge and heat transport in nanoscale and molecular devices, power issues in nanoelectronics, phase change memory, and device and materials characterization.

Sanjiv Sinha received the B.Tech degree in mechanical engineering from the Indian Institute of Technology, Delhi, in 1999 and the M.S. and Ph. D. degrees in mechanical engineering from Stanford University, Stanford, CA, in 2001 and 2005, respectively.

He currently works at the Systems Technology Lab, Intel Corp., Portland, OR. He is broadly interested in solid-state physics as well as transport phenomena and surface science.

Kenneth E. Goodson received the S.B. degrees in mechanical engineering and humanities in 1989 and the M.S. and Ph.D. degrees in mechanical engineering from the Massachusetts Institute of Technology in 1991 and 1993, respectively.

After graduation, he spent 18 months with the materials group at DaimlerBenz AG working on diamond heat sinks. He joined the Mechanical Engineering Faculty at Stanford University, Stanford, CA, in 1994, and his research group is a leader in the area of electronics cooling. He cofounded Cooligy in 2001 during a leave of absence from Stanford, and the startup develops and supplies pumped liquid cooling systems for computers. From a fundamental perspective, his group studies heat transfer at very small length and time scales, in particular those relevant to the design of transistors, semiconductor lasers, and microelectromechanical systems (MEMS). His students have contributed to the development of two-phase microchannel and micro-jet impingement cooling technology for semiconductor chips. He has authored or coauthored more than 150 journal and conference papers and five book chapters.

Prof. Goodson has received a number of awards in recognition of his work, including the ONR Young Investigator Award, the NSF CAREER Award, the Journal of Heat Transfer Outstanding Reviewer Award (1999), a JSPS Visiting Professorship at the Tokyo Institute of Technology (1996), as well as Best Paper Awards at SEMI-THERM (2001), the Multilevel Interconnect Symposium (1998), and SRC TECHCON (1998).