

PARALLEL COMPUTATION OF THE BOLTZMANN TRANSPORT EQUATION FOR MICROSCALE HEAT TRANSFER IN MULTILAYERED THIN FILMS

S. Srinivasan and R. S. Miller

Department of Mechanical Engineering, Clemson University, Clemson, South Carolina, USA

E. Marotta

Department of Mechanical Engineering, Texas A&M University, College Station, Texas, USA

Results are presented from parallel computations of one-dimensional (1-D) and two-dimensional (2-D) microscale heat conduction in multilayered films involving the materials silicon (Si) and silicon dioxide (SiO₂). The equation of phonon radiative transport (ERPT), in its spectral as well as frequency-independent form, is considered for numerical modeling using finite-difference methods. Parallelization strategies based on Message Passing Interface (MPI) routines are explored in an effort to achieve computational efficiency. The numerical solution results address the effects of film thickness, grain boundary scattering, and interfacial boundary conditions on the time-dependent temperature distribution within the microscale films.

1. INTRODUCTION

Thin films at micrometer and submicrometer dimensions are finding increasing application in modern microelectronics devices. For example, dielectric and semiconductor thin films are widely used in microprocessors and electronic circuits. Accurate prediction and modeling of heat transfer within these devices can be crucial to efficient operation and design. Such thin films appear with dimensions ~ 0.1 – $100\ \mu\text{m}$, for which size effects are generally considered important for thermal modeling [1–3]. A second class of devices utilizing thin films are pyroelectric sensors, which operate based on measurable voltage generated within pyroelectric materials layered between various electrodes and insulators in response to incident heat flux [4]. Constant heat flux sources require that the incident radiation be modulated, and accurate prediction of the internal transient material temperature distribution is

Received 24 October 2003; accepted 1 December 2003.

This work was partially supported under contract DAAD19-00-1-0002 from the U.S. Army Research Office and the Defense Advanced Research Projects Agency through MINSA. The authors would also like to thank Drs. Daniel Stanzione and Walter Ligon for access to the PARL Beowulf cluster.

Address correspondence to R. S. Miller, Department of Mechanical Engineering, Clemson University, Clemson, SC 29634-0921, USA. E-mail: rm@clemson.edu

NOMENCLATURE			
C	specific heat, J/kg K	τ	relaxation time, s
$\mathcal{D}(\omega)$	density of states, $\text{m}^{-3} \text{s}$	ϕ	azimuthal angle, rad
e	internal energy, $\text{W}/\text{m}^{-3} \text{s}$	ω	angular frequency of phonons, s^{-1}
\hbar	Planck's constant divided by 2π , J s		
I	phonon intensity, $\text{W}/\text{m}^2 \text{ST}$	Subscripts	
k_B	Boltzmann constant, J/K	A	material film A
L	film thickness, m	B	material film B
\mathcal{L}	integral operator	BE	Bose-Einstein statistics
N	number of grid points	c	cutoff value
q	heat flux, W/m^2	d	diffuse
$R_{\alpha\beta}$	phonon reflectivity	D	Debye
t	time, s	f	thin film
T	temperature, K	i	phonon polarization mode
$T_{\alpha\beta}$	phonon transmissivity	L	longitudinal transverse acoustic mode
v	phonon group velocity, m/s	TU	high-frequency transverse acoustic mode
x, y, z	Cartesian coordinates, m	T0	low-frequency transverse acoustic mode
α	thermal diffusivity ($\kappa/\rho C$), m^2/s		
α_v	Courant number	Superscripts	
θ	zenith angle, rad	*	nonvariable
κ	thermal conductivity, $\text{W}/\text{m K}$	0	equilibrium
Λ	mean free path, m	+	positive y direction
ρ	density, kg/m^3	–	negative y direction
σ	Stefan-Boltzmann constant for phonons, $\text{W}/\text{m}^2 \text{K}^4$		

required in order to predict sensor performance and sensitivity [1, 4, 5]. Previous analyses of these devices have been limited nearly exclusively to an assumption of heat transfer governed by the macroscopic Fourier's heat conduction law (presented for constant properties) [5]:

$$\frac{\partial T}{\partial t} = \alpha \nabla^2 T \quad (1)$$

where T is the temperature, $\alpha = \kappa/\rho C$ is the thermal diffusivity, κ is the thermal conductivity, ρ is the material density, and C is the material heat capacity. Fourier's law is valid for length scales much larger than the mean free path between phonon collisions [6]; however, modern thin-film devices are being designed with material thicknesses at or below mean free paths. For example, a recently published sensor design [4] includes a pyroelectric layer having a thickness equal to $L = 1.5 \mu\text{m}$; the mean free phonon path (Λ) is $\Lambda \approx 1 \mu\text{m}$. At such length scales, significant discrepancies between the predictions of this conventional law and various versions of the relatively more fundamental Boltzmann transport equations have been reported [3, 7, 8]. This has consequently instigated a re-evaluation of approaches to heat transfer design in such situations.

The Boltzmann transport equations in its general form has no known analytic solutions and therefore must be solved numerically. The principal assumption that makes the Boltzmann transport equation tractable to numerical solution is the relaxation time approximation to the scattering term on the right-hand side of the equation (refer to the following section). Through the assumption, all phonons are

allowed to “relax” to their corresponding isotropic equilibrium distribution. Incorporating this assumption are the two distinct forms of the simplified transport equation predominantly encountered in the literature.

The first approach is the phonon-frequency integrated formulation [3, 9] (referred to hereinafter as the BTE), which is consistent with the assumptions of a gray-body medium. This formulation describes the time- and space-dependent variation in the phonon distribution as a function of polar phonon propagation angles, but assumes independence of the phonon frequency. The calculation of the heat flux and internal energy follows the standard procedure employed in typical radiative heat transport problems. A “pseudo-temperature,” based on the equilibrium distribution of the phonons, is deduced from the local internal energy and the bulk-material specific heat.

A second common approach to microscale heat condition is the equation of phonon radiative transport (EPRT) [7]. This formulation retains the frequency dependence of the distribution evolution and is applicable to materials for which the relaxation time based on grain boundary scattering is treated as frequency-dependent. Here as well, the heat flux and the internal energy are calculated in the same way as the previous approach, by integrating the distribution function. The temperature, however, is calculated by equating the phonon intensity distribution with an assumed equilibrium having Bose-Einstein statistics [6, 7, 10].

Previous solutions to these equation sets have typically been limited to simplified one-dimensional (1-D) problems [1, 3, 7, 9, 11, 12]. Only a relatively few references that involve two-dimensional (2-D) domains exist in the literature [13, 14]. One of the main reasons for this deficiency is related to the substantial computational demands posed by such multidimensional geometries (see below). The primary objective of the present investigation is therefore to determine efficient parallelization strategies for both the BTE and EPRT for multidimensional microscale heat transfer problems. The model problem considered is that of two or more thin-film materials in direct layered contact with rectangular geometry. A single surface is subject to an instantaneous temperature increase and the ensuing transient and steady-state temperature distributions are studied. Dimensions and material properties are taken to be relevant to modern microelectronics devices. Two materials of largely varying phonon mean free paths are chosen for this investigation (monocrystalline silicon, Si, and silicon dioxide, SiO₂), in order to address applications in which either a hybrid BTE/Fourier or EPRT/Fourier approach is warranted.

The specific objectives of the research are to present: (1) an efficient parallel finite-difference scheme for the solution of the transient 1-D and 2-D versions of the integrated BTE, the EPRT, and the two hybrid Fourier formulations, (2) an investigation of computational issues related to simulating heat transfer in materials of largely varying mean free paths, and (3) example 1-D and 2-D results from simulations of microscale heat transfer in multimaterial thin films comprised of silicon and silicon dioxide layers. The article is organized as follows. The mathematical formulation of the BTE, EPRT and the hybrid BTE/Fourier and EPRT/Fourier methods are described in Section 2. The numerical approach to the solution of the governing equations is described in Section 3. Code parallelization and scaling results, code validation, resolution requirements, simulation times, and example physical results to a typical 2-D microscale heat transfer configuration are provided in Section 4. Conclusions and discussions are given in Section 5.

2. MATHEMATICAL FORMULATION

2.1. The Equation for Phonon Radiative Transfer (EPRT)

Based on an analogy between photons and phonons as discrete wave packets of thermal energy, the equation for phonon radiative transfer (EPRT) for the phonon intensity (I) is derived directly from the Boltzmann transport equation [1]:

$$\frac{\partial I}{\partial t} + \mathbf{v} \cdot \nabla I = \frac{I^0 - I}{\tau} \quad (2)$$

where the characteristic relaxation time τ is in general a function of the phonon frequency, ω , and temperature, T . The acoustic velocity vector \mathbf{v} is hereinafter assumed to be isotropic with magnitude v . The Cartesian (x, y, z)-coordinate velocity components for phonons moving in polar directions θ, ϕ are therefore:

$$v_x = v \cos \Phi \sin \theta \quad (3)$$

$$v_y = v \sin \Phi \sin \theta \quad (4)$$

$$v_z = v \cos \theta \quad (5)$$

for $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$ (see the inset in Figure 1). The intensity $I(t, x, y, z, \theta, \phi, \omega)$ is defined as the flux of energy in the direction of phonon propagation per unit area, per unit solid angle, per unit frequency. A relaxation-time approximation has been used in the above to model the scattering term in the Boltzmann transport equation which drives the local intensity towards an equilibrium intensity, I^0 (described below).

The heat flux \mathbf{q} is related to the net phonon flux and its components are given by [15]

$$q_x = \mathcal{L}(I \cos \Phi \sin \theta) \quad (6)$$

$$q_y = \mathcal{L}(I \sin \Phi \sin \theta) \quad (7)$$

$$q_z = \mathcal{L}(I \cos \theta) \quad (8)$$

where the operator \mathcal{L} is defined by:

$$\mathcal{L}(\Psi) = \int_0^{\omega_c} \int_0^{2\pi} \int_0^\pi \Psi \sin \theta \, d\theta \, d\Phi \, d\omega \quad (9)$$

where ω_c is a maximum ‘‘cutoff frequency’’ (defined below). In a similar manner, the internal energy (e) is given by

$$e = \mathcal{L}\left(\frac{I}{v}\right) \quad (10)$$

Applying the operator \mathcal{L} to the EPRT [Eq. (2)] yields the first law of thermodynamics:

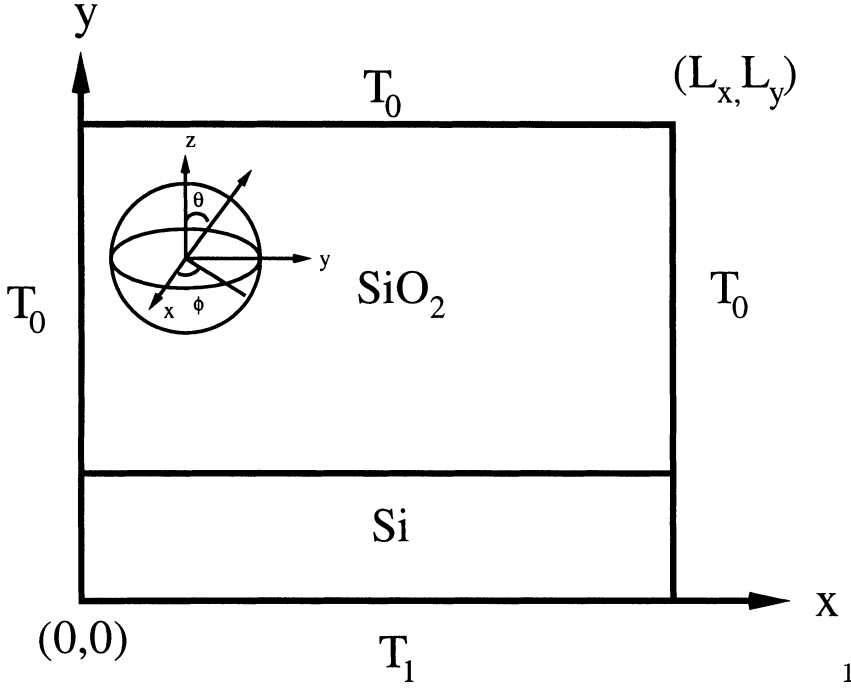


Figure 1. Schematic of the 2-D Si/SiO₂ domain. The Cartesian and polar-angle coordinate system is shown in the inset.

$$\frac{\partial e}{\partial t} + \nabla \cdot \mathbf{q} = \mathcal{L} \left(\frac{I^0 - I}{v\tau} \right) \quad (11)$$

which requires that the right-hand side be null:

$$\mathcal{L} \left(\frac{I^0 - I}{v\tau} \right) = 0 \quad (12)$$

A strict imposition of this equality is typically employed at each frequency [7]:

$$I^0(t, x, y, z, \omega) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi I \sin \theta \, d\theta \, d\Phi \quad (13)$$

where the equilibrium intensity I^0 is independent of the polar angle. Note that although the classical thermodynamic concept of temperature is poorly defined for dimensions of order or smaller than the phonon mean free path, a “pseudo-temperature” can be defined by assuming that an equivalent equilibrium intensity exists and is characterized by Bose-Einstein statistics:

$$I_{\text{BE}}^0(T) = \frac{1}{4\pi} \sum_i v_i \frac{\hbar \omega \mathcal{D}(\omega)}{\exp[\hbar \omega / k_B T] - 1} \quad (14)$$

where $\hbar = 6.625 \times 10^{-34} / 2\pi$ Js Planck’s constant, $k_B = 1.3806 \times 10^{-23}$ J/K is Boltzmann’s constant, $\mathcal{D}(\omega) = \omega^2 / 2\pi^2 v^3$ is the Debye density of states, and the

summation is over the three phonon polarizations [7]. The temperature appearing in Eq. (14) can be calculated using an iterative approach by considering the following equality [from Eqs. (12) and (14)]:

$$\mathcal{L}(I_{\text{BE}}^0) = \mathcal{L}(I) \quad (15)$$

which yields

$$\int_0^{\omega_c} I_{\text{BE}}^0 d\omega = \frac{1}{4\pi} \int_0^{\omega_c} \int_0^{2\pi} \int_0^{\pi} I \sin \theta d\theta d\Phi d\omega \quad (16)$$

2-D and 1-D EPRT formulation. The complete three-dimensional (3-D) formulation of the EPRT represents a transport equation for a six-dimensional array $I(t, x, y, z, \omega, \theta, \phi)$ (no array storage for the time coordinate), and its numerical solution is prohibitive even on modern parallel processing computers. We therefore consider a two-dimensional (x, y) reduction of the EPRT which not only reduces the spatial dimensionality [to $I(t, x, y, \omega, \theta, \phi)$] but also makes use of a symmetry in the intensity function implied by $q_z = 0$:

$$I\left(t, x, y, \frac{\pi}{2} - \theta, \Phi, \omega\right) = I\left(t, x, y, \frac{\pi}{2} + \theta, \Phi, \omega\right) \quad (17)$$

which reduces the polar coordinate interval to $0 \leq \theta \leq \pi/2$ and $0 \leq \phi \leq 2\pi$ (i.e., half of the original polar volume). In this case, the integration operator \mathcal{L} described above takes the form

$$\mathcal{L}(\Psi) = 2 \int_0^{\omega_c} \int_0^{2\pi} \int_0^{\pi/2} \Psi \sin \theta d\theta d\Phi d\omega \quad (18)$$

This change in operator, when substituted into the above 3-D formulation, defines the 2-D EPRT with the exceptions of Eqs. (13) and (16), which become

$$I^0(t, x, y, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi/2} I \sin \theta d\theta d\Phi \quad (19)$$

$$\int_0^{\omega_c} I_{\text{BE}}^0 d\omega = \frac{1}{2\pi} \int_0^{\omega_c} \int_0^{2\pi} \int_0^{\pi/2} I \sin \theta d\theta d\Phi d\omega \quad (20)$$

respectively.

The derivation of the 1-D form of the EPRT makes use of symmetries in the phonon intensity with azimuthal angle ϕ . Therefore, for the case of energy transport in only the z direction, the intensity reduces to a function of four variables, in this case $I(t, z, \omega, \theta)$. This is two dimensions less than the five-dimensional array problem size for the 2-D EPRT formulation and therefore is significantly more tractable. However, the primary emphasis of the present study is on multidimensional heat conduction. Hereinafter, all 1-D simulation results are solutions of the 2-D equations for energy flux in the y direction setting $q_x = q_z = 0$. This is implemented in the 2-D code by using three grid points in the x direction with periodic boundary conditions imposed. All 1-D parallel scaling results will therefore be reflective of the full 2-D code performance. This approach is also used for the BTE discussed below.

2.2. The Integrated BTE Formulation

For a gray medium, the methodology for solving the EPRT [Eq. (2)] can be simplified by considering the phonon intensity and relaxation time to be independent of frequency. In this case, modifications are made to the 2-D formulation in order to reflect its frequency independence:

$$\mathcal{L}(\Psi) = 2 \int_0^{2\pi} \int_0^{\pi/2} \Psi \sin \theta \, d\theta \, d\Phi \quad (21)$$

and

$$I^0(t, x, y) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi/2} I \sin \theta \, d\theta \, d\Phi \quad (22)$$

replace Eqs. (18)–(19), respectively. In addition, the temperature in this case is calculated directly from the definition of the internal energy:

$$e = CT \quad (23)$$

where C is the bulk heat capacity for the medium.

2.3. Hybrid EPRT/Fourier and BTE/Fourier Formulations

In considering multimaterial thin films with disparate ranges of phonon mean free paths and/or material dimensions, it may occur that one material is well described by the bulk Fourier's heat conduction law, whereas the other material requires a microscale heat transfer description via either the EPRT or the BTE. For example, the mean free paths of silicon and silicon dioxide considered in this study are 409 Å [9] (this value is used only in the BTE simulations, while a frequency-dependent relaxation time is used in the EPRT simulations) and 5.58 Å [3], respectively. As will be discussed below, resolution requirements for the silicon (due to its relatively very small mean free path) become prohibitive for relevant film thicknesses. We therefore propose a hybrid macro-/microheat transfer formulation based on the work of Sverdrup et al. [13], which allows for either a combined hybrid EPRT/Fourier or BTE/Fourier formulation. The approach within either material is identical to the pure Fourier and EPRT or BTE problems, respectively. Boundary conditions along material/approach interfaces are discussed below.

2.4. Calculation of Material Properties

The EPRT and the integrated BTE formulations require knowledge of the phonon relaxation time, the group velocity, and the specific heat. As a reasonable first approximation for many materials, these could either be borrowed or inferred from the corresponding published bulk values for the appropriate operating conditions (temperature, pressure, etc.) The analogy with the kinetic theory of gases is typically employed to calculate the phonon mean free path [9,16]. This is achieved by substituting the value for the bulk thermal conductivity (κ_{KT} is the conductivity

consistent with kinetic theory), specific heat (C), the phonon group velocity, v , and the phonon mean free path (Λ) in the following relation:

$$\kappa_{\text{KT}} = \frac{1}{3} C v \Lambda \quad (24)$$

The relaxation time (τ) is then $\tau = \Lambda/v$. A frequency-independent relaxation time is consistent with the gray-medium approximation.

Either the above approach or the macroscopic Fourier's law is used for the silicon dioxide modeled in this investigation. However, Ashegi and Goodson [17] observed that for silicon (and germanium) thin films, the adoption of the above approach may result in an overestimation of the thermal conductivity at room temperature. Therefore, to obtain more realistic result, the phonon scattering rate has to be determined by taking into account the film thickness, impurity and dopant composition, void concentrations, and grain structure (resulting in a frequency-dependent relaxation time). The procedure for determining the scattering rates for monocrystalline silicon (considered in this study) is summarized below.

Holland [18] considered transverse and longitudinal phonons separately to determine their respective contributions to the bulk thermal conductivity. The regime (with respect to the lower and upper phonon frequency limits) for each mode is determined from its phonon spectrum. Transverse phonons are demarcated into high- and low-frequency modes and are denoted by subscripts TU and T0, respectively. Similarly, longitudinal phonons are also demarcated into LU and L0 branches. However, typically only the low-frequency longitudinal mode, L0 (denoted hereinafter with the subscript L) is considered and yields a reasonable accuracy [18]. The relaxation time for each mode contains contributions from boundary scattering, impurity scattering, and phonon interactions (the three terms on the right-hand sides of the equations below, respectively), which are summed in a Matthiessen sense. With the definition of the cutoff frequency, ω_k , as $k_B \theta_k / \hbar$ (where $k = 1, 2, 3$ and the corresponding θ_k values are taken from Table 1) for each mode i (=TU, T0, L), the relaxation times are

$$\tau_{\text{T0}}(\omega)^{-1} = \frac{v_b}{L_y F} + A \omega^4 + B_T \omega T^4 \quad 0 \leq \omega \leq \omega_1 \quad (25)$$

Table 1. Parameters used for Holland's model [18]

v_{T0}	5,860 m/s
v_{TU}	2,000 m/s
v_L	8,480 m/s
v_b	6,400 m/s
θ_1	180 K
θ_2	210 K
θ_3	570 K
θ_D	658 K
A	$1.145 \times 10^{-45} \text{ s}^3$
B_T	$9.3 \times 10^{-13} \text{ K}^{-4}$
B_{TU}	$5.5 \times 10^{-18} \text{ s}$
B_L	$2.0 \times 10^{-24} \text{ s/K}^3$

$$\tau_{\text{TU}}(\omega)^{-1} = \frac{v_b}{L_y F} + A\omega^4 + B_{\text{TU}}\omega^2 \sinh(\hbar\omega/R_B T) \quad \omega_1 < \omega \leq \omega_2 \quad (26)$$

$$\tau_L(\omega)^{-1} = \frac{v_b}{L_y F} + A\omega^4 + B_L\omega^2 T^3 \quad 0 < \omega \leq \omega_3 \quad (27)$$

Values for the constants appearing in the above equations are provided in Table 1 [18]. The above relations correspond to bulk material predictions, so a modification is necessary to account for size effects in thin films. Ashegi et al. [17] proposed that this could be achieved by retaining all the terms (with their respective constants unchanged) in the above equation except for the grain boundary scattering term, $v_b/L_y F$, and then applying Sondheimer's [19] method for estimating the reduction factor for the bulk-material mean free path. The reduction factor for totally diffuse surface scattering (the motivation for the use of which is explained below) is given as

$$\gamma_i(\omega) = 1 - \frac{3L_y}{2\Lambda_i(\omega)} \int_{t=1}^{\infty} \left(\frac{1}{t^3} - \frac{1}{t^5} \right) \left\{ 1 - \exp\left[\frac{-\Lambda_i(\omega)t}{L_y} \right] \right\} dt \quad (28)$$

with $\Lambda_i(\omega) = \tau_i(\omega)v_i$. The reduction factor is then multiplied with the original bulk-material relaxation time (excluding the grain boundary scattering term),

$$\tau_{f,i}(\omega) = \gamma_i(\omega)\tau_i(\omega) \quad (29)$$

The predicted thermal conductivity of monocrystalline silicon as a function of film thickness according to Holland's model [18], adapted for thin films by Ashegi and Goodson [17], is plotted in Figure 2a. The predicted conductivities are based on relaxation times calculated from Eqs. (26)–(29) substituted into ‘‘Holland's thermal conductivity integrals.’’

$$\kappa_{\text{T0}} = \frac{k_B}{3\pi^2 v_{\text{T0}}} \left(\frac{k_B T}{\hbar} \right)^3 \int_0^{\theta_1/T} \tau_{f,\text{T0}} \frac{x^4 \exp x}{(\exp x - 1)^2} dx \quad (30)$$

$$\kappa_{\text{TU}} = \frac{k_B}{3\pi^2 v_{\text{TU}}} \left(\frac{k_B T}{\hbar} \right)^3 \int_{\theta_1/T}^{\theta_2/T} \tau_{f,\text{TU}} \frac{x^4 \exp x}{(\exp x - 1)^2} dx \quad (31)$$

$$\kappa_L = \frac{k_B}{6\pi^2 v_L} \left(\frac{k_B T}{\hbar} \right)^3 \int_0^{\theta_3/T} \tau_{f,L} \frac{x^4 \exp x}{(\exp x - 1)^2} dx \quad (32)$$

Based on this figure, the high-frequency transverse phonon mode appears to be the dominant phonon mode in heat conduction. This is to be expected, since the trend (for thin films) mirrors the behavior of the phonon modes (with respect to thermal conductivity) for bulk materials at room temperature as per the original model, differing only by a reduction factor for thin films.

However, studies conducted on bulk materials with phonon spectra similar to monocrystalline silicon [20], following Holland's work, have revealed that the longitudinal acoustic phonon mode dominates the other two modes. For thin films, Ju and Goodson [21] proposed modeling the thermal conductivity contributed by longitudinal phonons as the sole heat carriers. They use Holland's thermal

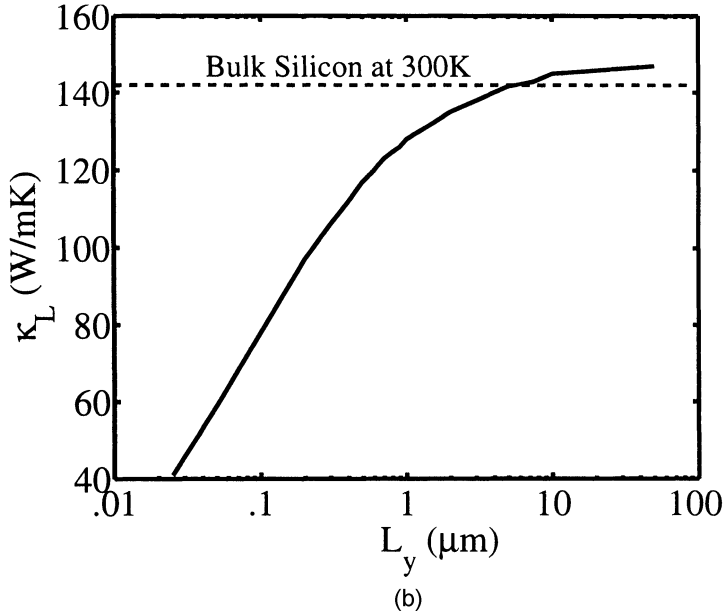
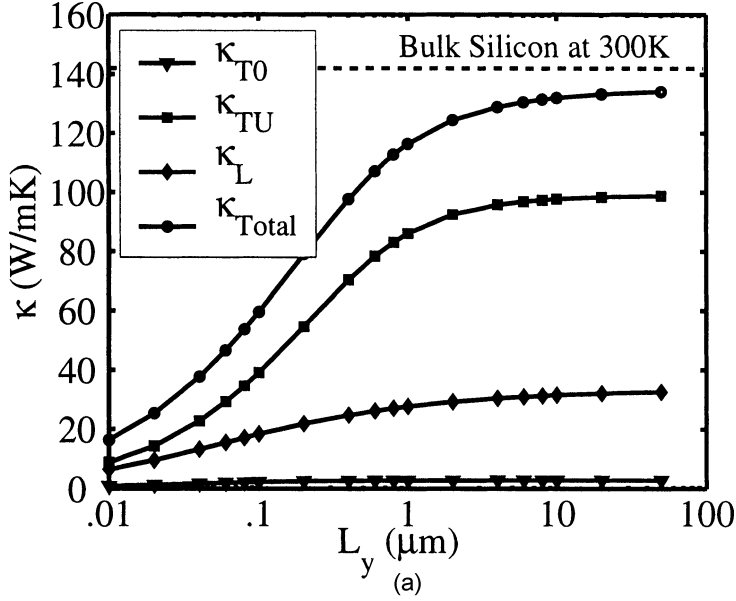


Figure 2. Thermal conductivity predictions for Si at $T = 300\text{K}$: (a) contributions of the three acoustic modes, and (b) assuming the longitudinal acoustic phonon mode to be the dominant mode.

conductivity integral for longitudinal phonons alone, and adjust the three-phonon interactions constant $B_L(T)$ in Eq. (27) such that the predicted thermal conductivity matches the total thermal conductivity for bulk silicon. They also suggest that the frequency, ω , in the phonon-interaction term be raised to the power 1.7 rather than 2

as in the original formulation. This leads to (neglecting the boundary scattering terms as before)

$$\tau_L(\omega)^{-1} = A\omega^4 + B_L\omega^{1.7}T^3 \quad 0 < \omega \leq \omega_3 \quad (33)$$

where $A = 1.145 \times 10^{-45} \text{ s}^3$ and $B_L(T = 300 \text{ K}) = 1.42 \times 10^{20} \text{ K}^{-3} \text{ s}^{-0.7}$. The thermal conductivity based on this approach is plotted in Figure 2b; the bulk silicon value is 142 W/mK [17], in good agreement with the model prediction. The relaxation time calculated by the same approach is incorporated in all simulations involving monocrystalline silicon in what follows.

2.5. Boundary and Initial Conditions

Both Fourier's heat conduction law [Eq. (1)] and the (pure and hybrid) EPRT and BTE formulations presented above are solved numerically for both 1-D and 2-D rectangular geometries (Figure 1). For 1-D cases the solution is on the interval $0 \leq y \leq L_y$, while for 2-D cases the rectangular geometry is bounded by $0 \leq x \leq L_x$ and $0 \leq y \leq L_y$. In both cases, the material is initially at uniform temperature T_0 . At time $t = 0$, a single edge of the material is raised instantaneously to temperature $T_0 + \Delta T$. This occurs at location $y = 0$ for 1-D cases and along the $y = 0$ surface in 2-D cases. For the EPRT the phonon intensity is initialized by assuming angular independent Bose-Einstein statistics as a function of the specified temperature [Eq. (14)]. For the frequency-integrated BTE, the initialization is [3]

$$I^0 = \frac{1}{4\pi} C v T \quad (34)$$

After time $t = 0$, boundary conditions for Fourier's law are fixed temperature, whereas for the EPRT the intensity function for *incoming* phonons remains fixed; the intensity of outgoing phonons evolves via the EPRT or the integrated BTE. All simulations use $T_0 = 300 \text{ K}$ and $\Delta T = 0.1 \text{ K}$; conditions relevant to several modern pyroelectric sensors and the same conditions as those employed in the 1-D simulations of Majumdar [1].

The rectangular geometries considered include both single- and multimaterial films. The contact surfaces between layers are normal to the primary direction of heat flux from the heated boundary; i.e., the thickness of material Ψ in the y direction is $L_y^{(\Psi)}$, such that $\sum L_y^{(\Psi)} = L_y$. For thin films, increased surface scattering significantly reduces the effective thermal conductivity. The nature of surface scattering can be purely diffuse, purely specular, or a combination of the two, depending on the material and temperature ranges encountered. The coefficient for specular reflection is given by [15]

$$p \approx \exp\left(\frac{-16\pi^3\eta^2}{\lambda_{\text{dom}}^2}\right) \quad (35)$$

with the dominant phonon wavelength, λ_{dom} , being of the order of $2\pi\hbar v/k_B T$. When an appropriate value for the surface roughness factor η (which ranges between 1 and 10\AA for the silicon wafers considered here [17]) is substituted, the maximum value of the specularly factor is less than 0.01 at room temperature. Therefore, for the

present investigation it is reasonable to assume purely diffuse surface scattering (such that all phonons, either reflected or transmitted, emerge isotropically distributed) and use Eq. (28).

Also important at the material layer boundaries is interface scattering. Among the simplest and most popularly employed approaches to modeling interface scattering are the acoustic mismatch model (AMM) [9, 3] and the diffuse mismatch model (DMM) [9, 3]. The former considers elastic phonon collisions such that the reflected and transmitted phonons are at the same frequency as they were before the event. In addition, across the interface of two dissimilar materials, only the phonons with frequencies lesser than the minimum of either medium's cutoff frequency are transmitted. Another restriction is imposed on the phonons with angles of incidence greater than the critical angle that are also confined within the medium of origin. Therefore, the transmissivity and reflectivity of the phonons predicted by the AMM have angular and frequency dependence. This model is typically employed when operating temperatures are much lower than the Debye temperature, or when the Debye temperatures of the two materials are substantially different. In contrast to this approach is the DMM. Inelastic collisions are treated within this model such that phonons of any frequency can be transmitted across the interface with the transmitted and reflected phonons emerging with an isotropic distribution. The probability for transmission or reflection of the incident phonons is calculated according to a procedure discussed in the paragraph below. The DMM is typically applied at temperatures of order and larger than the Debye temperature. The Debye temperatures for the materials considered in this work are 570 K and 492 K for Si [18] and SiO₂ [13], respectively. We therefore assume the DMM to be applicable for both materials and employ it hereinafter. Note that although the Debye temperature of the dominant heat carrying phonon mode for Si is slightly larger than the room-temperature conditions considered in this work, heat flux and temperature predictions by the DMM are deemed sufficiently accurate for the purposes of this study. The DMM has been used for Si under similar conditions by Sverdrup et al. [13].

The coefficient of transmissivity of phonons across a material interface is first calculated by performing an energy balance at the interface using the relevant initial conditions [Eq. (14) or (34)]. For the EPRT formulation the transmissivity coefficient, $T_{\alpha\beta}$ (α and β indicate the particular material and take values A or B), is a function of the material phonon velocities [22],

$$T_{\alpha\beta} = \frac{1/v_{\beta}^2}{1/v_{\alpha}^2 + 1/v_{\beta}^2} \quad (36)$$

whereas for the integrated BTE form it is the material heat capacities which determine the transmissivity,

$$T_{\alpha\beta} = \frac{C_{\beta}v_{\beta}}{C_{\alpha}v_{\alpha} + C_{\beta}v_{\beta}} \quad (37)$$

The ‘‘coefficient of reflectivity’’ is then given by $R_{\alpha\beta} = 1 - T_{\alpha\beta}$. The isotropically distributed diffuse intensity emerging from the surface and entering material A ($I_{A,D}$) is then given in terms of the intensities of phonons exiting material A and entering material B ,

$$I_{A,D} = \frac{2}{\pi} T_{BA} \int_0^\pi \int_0^{\pi/2} I_B \sin^2 \theta \sin \phi \, d\theta \, d\phi - \frac{2}{\pi} R_{AB} \int_\pi^{2\pi} \int_0^{\pi/2} I_A \sin^2 \theta \sin \phi \, d\theta \, d\phi \quad (38)$$

while the corresponding intensity emerging from the surface and entering material B is

$$I_{B,D} = \frac{2}{\pi} T_{AB} \int_\pi^{2\pi} \int_0^{\pi/2} I_A \sin^2 \theta \sin \phi \, d\theta \, d\phi - \frac{2}{\pi} R_{BA} \int_0^\pi \int_0^{\pi/2} I_B \sin^2 \theta \sin \phi \, d\theta \, d\phi \quad (39)$$

For additional details the reader is referred to [3, 9].

Hybrid EPRT/Fourier and BTE/Fourier Boundary Conditions. For materials such as those considered here, which have large diversities in the phonon mean free path, one material may be well predicted by the macroscopic Fourier's law while the other requires a microscale heat transfer description. For such cases either a hybrid EPRT/Fourier or BTE/Fourier algorithm is employed as follows. For the present simulations, SiO_2 occupies the nonheated portion of the domain in the region $L_{y,A} \leq y \leq L_y$ (Figure 1). Therefore, the temperatures along the peripheries of the silicon dioxide domain are constant ($T = T_0$):

$$T_B(0, y = L_{y,A}) = T_B(L_x, y = L_{y,A}) = T_B(x, y = L_y) = T_0 \quad (40)$$

At the interface of Si/SiO_2 , the intensity distribution is calculated by equating the y -direction heat fluxes from both sides of the interface. Here, it is assumed that the phonons leaving the oxide side of the interface are at equilibrium and a hypothetical equilibrium intensity is calculated [Eq. (14) or (34)] with the oxide-interface temperature obtained from the solution to Fourier's law. These phonons are then transmitted into the silicon layer according to the transmissivity calculated according to the DMM as explained in the previous subsection. For the silicon side of the interface, a fraction of phonons is reflected back (reflectivity again calculated according to the DMM) to the silicon domain. Therefore, for the silicon side of the interface we have the following isotropic intensity:

$$I_{A,D} = T_{BA} I_B^0(T_B) - \frac{2}{\pi} R_{AB} \int_\pi^{2\pi} \int_0^{\pi/2} I_A \sin^2 \theta \sin \phi \, d\theta \, d\phi \quad (41)$$

As before, the outgoing phonon intensities from the silicon layer interface evolve according to the EPRT or the integrated BTE and the heat flux, $q_{y,\text{Si}}$, is calculated according to Eq. (7) or its integrated BTE equivalent.

For the oxide interface the y -direction heat flux is calculated iteratively (with respect to Eq. (41) and the equation below) by guessing the interface temperature, T_B , at each grid point successively such that

$$q_{y,\text{Si}} \approx q_{y,\text{SiO}_2} = -\kappa_{\text{SiO}_2} \frac{dT}{dy} \quad (42)$$

Convergence is established with the aid of an imposed tolerance when the two heat fluxes fall within this limit (see below).

3. NUMERICAL APPROACH

A finite-difference-based approach is employed for the numerical solution of the governing equations. This approach is investigated as an alternative to statistical Monte Carlo-based methods [23] due to: (1) the availability of inexpensive parallel Beowulf clusters, making large multidimensional simulations feasible; (2) the ease of programming and parallelizing finite-difference methods; and (3) the relatively simple manner in which unsteady simulations can be performed (as opposed to Monte Carlo methods, which require many realizations to be conducted for statistical averaging of the results). Nondiffusive second-order central differencing is used for Fourier's heat conduction law, whereas a first-order upwinding scheme is used for the EPRT because of its hyperbolic nature [1]. The coordinates are discretized on the intervals $0 \leq x \leq L_x$, $0 \leq y \leq L_y$, $0 \leq \omega \leq \omega_D$, $0 \leq \theta \leq \pi/2$, and $0 \leq \phi \leq 2\pi$ using N_x , N_y , N_ω , N_θ , and N_ϕ grid points with uniformly spaced intervals Δx , Δy , $\Delta\omega$, $\Delta\theta$, and $\Delta\phi$, respectively. For the sake of simplicity we make the further restrictions that $\Delta x = \Delta y$ and $\Delta\theta = \Delta\phi$. The maximum frequency (ω_D) is determined from the Debye temperature of the material using the relation $\omega_D = k_B\theta_D/h$. The Courant number based on velocity, $\alpha_v = v\Delta t/\Delta x$, is typically $\alpha_v = 1.0$ for the smaller-material-dimension cases; however, it must be decreased for larger film thicknesses due to restrictive temporal resolution requirements. All integrations are performed using the trapezoidal rule. The temperature is calculated from Eq. (14) using the bisection method with an imposed tolerance of 1×10^{-12} K. The code has provisions for performing 1-D simulations by imposing laterally periodic boundary conditions. In this case, the domain width is equal to the grid size along the direction of heat flow (Δy). Also in this section, it is noted that the values " $2/\pi$ " appearing on the right-hand sides of Eqs. (38) and (39) originate from the integral $\int_0^\pi \int_0^{\pi/2} \sin^2\theta \sin\phi \, d\theta \, d\phi$. If the analytic solution is used, sensitivity of the interface conditions will result in substantial errors in matching phonon intensities across the material boundaries. This is overcome in practice by using the value of the integral achieved using the same trapezoidal integration as is used in evaluating integrals on the right-hand sides of the equations. The same is true for the value 4π in the denominator of Eq. (13), which originates from the integral $2 \int_0^{2\pi} \int_0^{\pi/2} \sin\theta \, d\theta \, d\phi$.

For the hybrid approach (EPRT/Fourier or integrated BTE/Fourier), an analysis of characteristic thermal diffusion time scales shows that unsteady solutions to Fourier's law must be considered: The characteristic time for thermal diffusion in the SiO₂ is $L_{y,\text{SiO}_2}^2/\alpha \approx 10^{-8}$ s, while that for the silicon material is $L_{y,\text{Si}}/v \approx 10^{-12}$ s. Therefore, Eq. (1) is solved numerically in the substrate using the forward time and central space finite-difference scheme. The iterative procedure associated with Eq. (42) is handled easily because of the explicit nature of the numerical approach in the substrate. As mentioned in the previous section, at each new time level, the temperature at a point along the interface (with the new interior temperature distribution already available in the substrate) is iterated such that the EPRT heat flux corresponding to the phonon-intensity distribution on the monocrystalline silicon side of the interface at that point is the same as the Fourier heat flux in the oxide side

of the interface. The iteration is carried out in a grid point-by-grid point manner using the bisection method, with the speed of convergence being related to the imposed tolerance for acceptable heat flux mismatch. Also, with the time step for the substrate being two orders of magnitude larger than the silicon region (for the aforementioned grid configuration), the EPRT-related time step is chosen to ensure time stability.

4. RESULTS

Results of the present study are presented in the following subsections relating to code parallelization, code validation, resolution requirements, total simulation times, and an example physical microscale heat transfer simulation.

4.1. Parallelization Scaling Results

Solution of the 2-D EPRT represents the time advancement of a five-dimensional array, $I(x, y, \omega, \theta, \phi)$, and would be prohibitively expensive on typical serial computers of the type commonly available. For example, consider a moderate discretization on a square domain having $N_x = N_y = N_\omega = 4N_\theta = N_\phi = 40$. The total number of grid points for this problem is $40^5/4 = 25,600,000$. On the other hand, parallel computers are becoming highly prevalent and extremely cost-effective options for large problem sizes. Pure parallelization of the above problem using 16 processors would then result in 16 separate serial jobs having a tractable 1,600,000 grid points each. With the increased availability of parallel Beowulf clusters, we therefore choose to explore the use of “direct” finite-difference schemes for the 2-D EPRT as an alternative to stochastic-based approaches such as the Monte Carlo method.

Parallelization of the EPRT proves to be highly efficient, resulting in near-linear scaling with added processors *if* an appropriate parallelization scheme is chosen. The code described hereafter is written in Fortran 77 and utilizes the Message Passing Interface (MPI) parallel communication routines. Two Beowulf clusters are used for testing. The first cluster, operated by the Department of Mechanical Engineering (ME) at Clemson University, is comprised of 20 dual-processor nodes with varying processors based on additions made over three years of upgrades. The first eight nodes contain 1.28-GHz Pentium III processors. The next eight nodes consist of 933-MHz Pentium III processors. The final four nodes contain 700-MHz Pentium III processors. All nodes contain 1 GB of memory, and interprocessor communication is via a dedicated Myrinet network operating at 1.28 gigabits per second (gbps). A single processor is allocated from each machine before any single machine has both processors active and sharing the memory bus. Therefore, linear scaling can only be expected up to 20 processors, even if all nodes contained identical hardware. The second cluster is operated by the Parallel Architecture Research Laboratory (PARL) at Clemson University. This cluster (at the time the simulations were conducted) has 90 nodes, each containing dual 1-GHz Pentium III processors. Communication is via “fast” 100-megabit per second (mbps) ethernet. Linear scaling could therefore be expected up to 90 processors. Testing is hereinafter limited to no more than 20 and 90 processors on the two clusters, respectively.

The options for parallelizing the problem tested in this study involve domain decomposition in combinations of spatial or spectral coordinates. Two versions of the code were developed with the sets of parallelized variables, θ , ϕ , ω and x , θ , ω , respectively. Both codes solve either the Fourier, BTE, or EPRT formulation or the hybrid approaches. Individual simulations for BTE or EPRT cases can be run using 1-D, 2-D, or 3-D parallel decompositions by specifying the number of processors along each of the three parallel-variable directions. Parallelization for Fourier domains is in either x or y or both. We also note here that costly iterative temperature calculations for the EPRT code [Eq. (16)] are only performed at select intervals used for output. In general, the temperature must be calculated at every time step in order to evaluate the temperature-dependent relaxation times. However, the temperature rise considered in what follows is 0.1 K and temperature changes are therefore negligible to accurate relaxation-time evaluations. Additional testing not reported here shows that temperature evaluations can be neglected even for much larger rises, up to approximately 50 K. All results below therefore correspond to cases for which the temperature is not updated every time step (except along the material interface boundary for hybrid EPRT/Fourier cases). No iterations are required for BTE-based temperature calculations. Parameters for scaling and performance evaluations discussed below are provided in Table 2.

Parallelization in the spatial domain. The most obvious method of parallelization is to decompose the spatial domain into uniform intervals, each handled by its own processor. This approach is taken for the Fourier approach to heat conduction and is simply applied in this case with nearly perfect linear scaling of the solution with added processors. Unfortunately, the same approach is ill-suited to the solution of the EPRT and BTE equations because of extreme communication requirements. Consider a rectangular domain based on Fourier's heat conduction law, decomposed into two identically sized subdomains in the y direction at position $y=L'$. At each time step, interprocessor communication is required in order to calculate derivatives across the processor boundary at $y=L'$. Second-order central finite differencing requires that the single set of temperature from the edge cells be communicated from each side of the boundary (i.e., $y=L'^+$ temperatures are sent to the $y=L'^-$ side and vice versa). Higher-order finite differences require a larger stencil size and therefore extra rows of temperatures are communicated; however, even for an eighth-order central scheme, only four rows of temperature data are communicated from each side of the processor boundary. This type of parallelization is highly efficient, as only relatively small amounts of data require communication.

Table 2. Details of the base test cases used in the parallelization study

Case	Methodology	Domain size	$N_x \times N_y \times N_\theta \times N_\phi (\times N_\omega)$
A1	1-D EPRT/EPRT	300 nm (50-nm-thick Si)	$3 \times 48 \times 16 \times 61 \times 180$
A2	1-D EPRT/EPRT	300 nm (50-nm-thick Si)	$3 \times 24 \times 64 \times 253 \times 36$
B1	2-D EPRT/EPRT	$75 \times 25 \text{ nm}^2$ (50-nm-thick Si)	$40 \times 30 \times 24 \times 93 \times 36$
B2	2-D BTE/Fourier	$250 \times 325 \text{ nm}^2$ (75-nm-thick Si)	$36 \times 48 \times 10 \times 37$

In contrast, for the EPRT and the BTE solutions the phonon distribution functions are multidimensional, and much larger communication requirements result. In these cases, two sets of phonon distribution arrays corresponding to the same boundary values are either four-dimensional ($N_x \times N_\omega \times N_\theta \times N_\phi$) or three-dimensional ($N_x \times N_\theta \times N_\phi$) for the EPRT and the BTE, respectively. These relatively very large data sets must be communicated every time step (for first-order explicit time stepping; more for higher-order multistep methods such as Runge-Kutta schemes) and along every interprocessor boundary. These stringent communication requirements quickly overload even a high-speed network such as the Myrinet network on our primary test cluster, even when only a 1-D heat conduction problem is simulated. For example, a 1-D test case scaling relative to the time on a single processor is presented in Figure 3a as a function of the number of processors on each of the two clusters used for this investigation. The “speed-up” is defined as the average real time per iteration on a single processor relative to that on N processors (ideal scaling would be linear, as denoted by the dashed line in the figure). The test case of this figure is Case B1 (Table 2), involving an EPRT formulation for 2-D heat transfer in a Si/SiO₂ domain of size $75 \times 25 \text{ nm}^2$, for which the silicon domain width is $L_{y,A} = 50 \text{ nm}$. The grid size used is $40 \times 30 \times 24 \times 93 \times 36$ ($N_x \times N_y \times N_\theta \times N_\phi \times N_\omega$), with parallelization in the x coordinate only. Even on this relatively large mesh, communication quickly overcomes any advantages from added processors. Linear scaling is not observed even for two processors, and a saturation of performance increases occurs after only a few processors. Performance is nevertheless better with the high-speed network on the ME cluster compared to the 100-mbps ethernet network on the PARL cluster.

Parallelization in the frequency domain. The advantages of parallelization of the EPRT in the ω domain lie in the fact that in solving for the EPRT, partial derivatives with respect to ω are not encountered, precluding the type of communications found in spatial parallelizations. Only integrals over frequency are contained in the governing equations [Eqs. (6), (7), (8), (10), and (16)]. The benefit of not having partial derivatives over ω is that data from neighbouring processors need not be communicated to evaluate the partial derivative at the processor boundary. Interprocessor communications are required in this case only at select times for which the temperature is calculated in the EPRT formulation (discussed above). Therefore, the EPRT with respect to each ω interval can be evaluated in the processors *independently* of the others until the need arises to calculate the temperature for output. Even in this case, the communication load is easily minimized by communicating only the single scalar value representing the local integral over each processor’s portion of the frequency domain, which are then summed by an appropriate MPI call.

Figure 3b presents the code speed-up for Case A1 with 1-D domain decomposition in the frequency domain as a function of the number of processors on the PARL cluster. The 1-D, two-material EPRT simulation has a resolution of $3 \times 48 \times 16 \times 61 \times 180$ grid points (Case A1). Near-linear scaling is observed up to 90 processors. At this point only two ω grid points are contained within each processor’s portion of the problem. This illustrates one of the limitations of pure frequency-domain decomposition, as the number of processors cannot exceed the

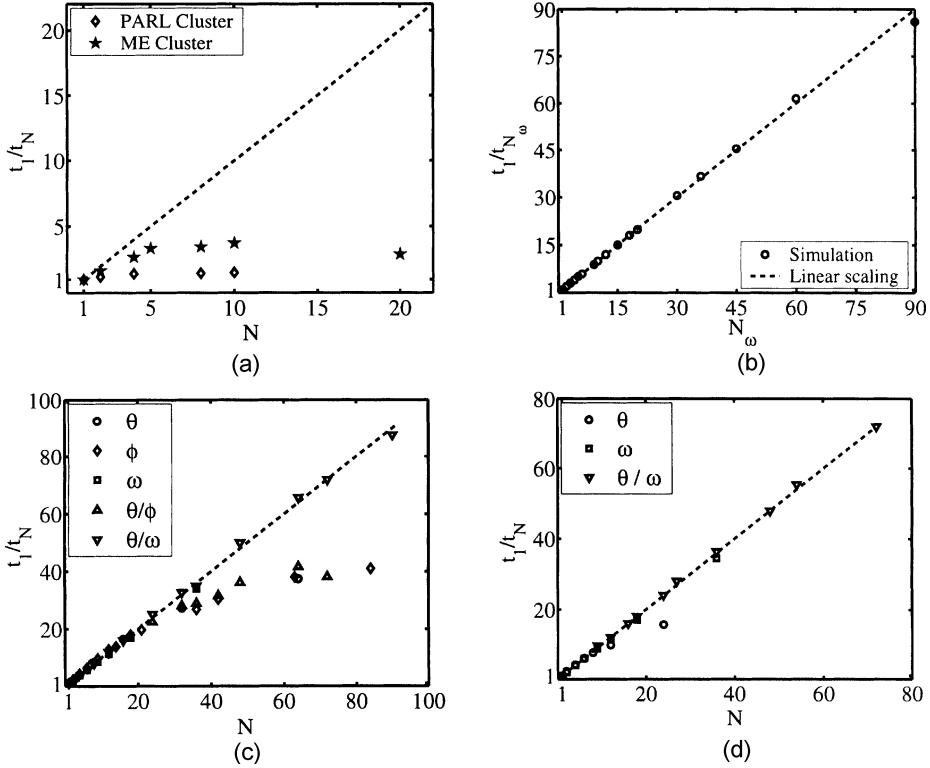


Figure 3. Speed-up for (a) Case B1 with 1-D parallelization in x , (b) Case A1 with 1-D parallelization in ω , (c) Case A2 and (d) Case B1 with 1-D and 2-D parallelization.

number of grid points used to discretize the frequency domain (which is in general fewer than the $N_\omega = 180$ considered for this test case). The observed linear scaling is further confirmed in Figure 3c, corresponding to a 1-D test case with grid resolution of $3 \times 24 \times 64 \times 253 \times 36$ grid points (Case A2).

Parallelization in the polar angles and multidimensional parallelization. As in the case with frequency parallelization, the formulations of the governing EPRT and BTE do not contain derivatives with respect to polar angles, making for efficient parallelization in these variables. However, whereas the frequency integral requires calculation only at times for which the temperature is required, calculation of the equilibrium intensities for the relaxation terms mandates polar angle integrations at every time step for both the EPRT and the BTE [Eqs. (13) and (22), respectively]. As a result, optimal parallelization is achieved with domain decompositions solely in the frequency domain for the EPRT. However, if the number of grid points used to discretize ω is insufficient for the desired number of processors (or if the BTE is being solved), additional levels of parallelization in polar angles can be added. Figures 3c and 3d present the code scaling performance for Case A2 and Case B1, testing performance in both single- and multidimensional domain decompositions. Multidimensional parallelization in ω and θ domains exhibits optimal performance, scaling nearly

linearly over the entire range of the number of processors available. Pure parallelization in θ is limited by communication overhead and begins to flatten after approximately 20 processors for the conditions chosen (Figure 3c). Parallelizations involving ϕ and ω are also nonideal and generally perform less favorably than those involving θ , because of the order in which the nested polar angle integrations are performed.

Parallelization of the hybrid interface temperature calculation. For the hybrid BTE/Fourier model, the code performance is additionally limited by the iterative nature of the temperature calculation at the interface. While the calculation of heat flux on the substrate side of the interface is trivial (involving Fourier's Law), the calculation of heat flux on the silicon-layer side involves evaluation of integrals (Eqs. (7) and (41)) and therefore multiple calls to the MPI_ALLREDUCE routine *per* grid point. This puts strict limitations on the scaling efficiency of the hybrid EPRT/Fourier and BTE/Fourier codes. For such cases we choose to parallelize the temperature calculation alone. That is, the general solution to both the Fourier and the EPRT or BTE portions of the domain is treated as in a serial code. However, at each time step when the interface temperature requires updating, the interface boundary is decomposed evenly among a specified number of processors. Each processor then performs the iterations necessary to calculate the interface temperatures and then sends these values back to the master processor. In support of the above arguments, it was observed that running the code with interface temperature and the polar angles as the parallel variables was inefficient compared to 1-D parallelization in interface temperature calculation alone. Also, for 1-D cases involving this approach, the serial code worked better than any of the parallel versions.

Code speed-up as a function of the number of processors was investigated for the hybrid BTE/Fourier solver. The configuration considered is a 2-D Si/SiO₂ test case of dimension $250 \times 325 \text{ nm}^2$ with a 75-nm-thick silicon layer. The grid resolution is $36 \times 48 \times 10 \times 37$ ($N_x \times N_y \times N_\theta \times N_\phi$) grid points (Case B2). Code performance is enhanced by added processors but only up to six processors, after which performance is degraded by added processors, because of enhanced communication overhead. Even with less than six processors, the observed scaling is less than linear. Additional results for the same case but using θ parallelization reveal that communication overhead actually increases the simulation time using even only two processors (not shown). However, for the EPRT/Fourier hybrid model, 1-D ω parallelization worked significantly better than 1-D parallelization in temperature calculation of θ space. This is due to the fact that the number of integrations related to the MPI call is *halved*, since communication is required only during the evaluation of the integral over phonon frequency [Eq. (7)]. Using the same spatial domain and grid resolution as in the test case discussed above, but additionally with 32 grid points in frequency, a speed-up with as many as 16 processors was observed, although it was not linear (not shown).

Although the code speed-up is extremely limited for the hybrid approaches, it may become necessary to use the hybrid solver for material dimensions much larger than the mean free path of the material (or the minimum mean free path for multimaterial simulations). This is because associated spatial resolution requirements for

the pure EPRT or BTE solver can become overly taxing for large film thickness, thus negating the scaling performance of the pure solvers (see below).

Problem scaling efficiency. Finally, the code was tested for performance with respect to the size of the problem (represented by the total mesh size). This is done by studying the scaling of the average time per iteration with the problem size for a fixed number of processors. For this study, 2-D EPRT simulations of Si/Si were performed with different mesh sizes. Parallelization is provided by 32 processors decomposing the frequency domain only. The spatial domain size for this case was chosen to be $10\ \mu\text{m} \times 10\ \mu\text{m}$ with $L_{y,A} = 5\ \mu\text{m}$. For the base case, a mesh with 100 million grid points was considered having $N_x = 50$, $N_y = 50$, $N_\theta = 18$, $N_\phi = 69$, and $N_\omega = 32$. The corresponding speed-up is then defined as the ratio of the average time per iteration relative to the associated time for the large base case. The speed-up of the frequency parallelized code was found to be nearly ideal over the range of problem sizes considered (not shown).

4.2. Code Validation

The code is checked for possible errors by performing simulations corresponding to cases from published results [1, 24]. The heat flux predicted by the EPRT code is easily verified by the radiative heat transport theory [25, 26]. According to this theory, under gray-body assumptions, the predicted steady-state 1-D heat flux corresponding to the boundary conditions used in this work is

$$q_y = \pi \int_0^{\omega_D} I_{\omega,y=0}^0 d\omega - \pi \int_0^{\omega_D} I_{\omega,y=L_y}^0 d\omega \quad (43)$$

where the equilibrium intensity is calculated according to Eq. (14) or (34). Simulations were then performed for a diamond domain with a range of film thicknesses between 0.01 and $10\ \mu\text{m}$. The properties used for the simulations are taken from the NIST Web database and are provided in Table 3. The grid resolution used for these simulations is $3 \times 30 \times 10 \times 37 \times 32$ ($N_x \times N_y \times N_\theta \times N_\phi \times N_\omega$). The maximum deviation between the theory and the code performance was approximately 3.6%, thereby validating the basic functionality of the code.

The implementation of the spectral relaxation-time model and frequency-resolution requirements for the EPRT code are examined next. For EPRT cases involving a spectral relaxation time, the bulk thermal conductivity predicted by Fourier's law ($k = |q_y| \Delta y / \Delta T$) is compared with the values calculated directly from Holland's [18] thermal conductivity integrals modified for thin films [21, 17] (presented previously in Figure 2a). The values obtained from the complete EPRT code for three independent cases involving a $50\ \mu\text{m}$ -thick silicon domain for each of the three phonon branches considered in isolation—T0, TU, and L—are 2.86, 100.4, and 31.24 W/mK, respectively. The grid resolution used for these simulations is $3 \times 30 \times 10 \times 37 \times 32$. Corresponding values were also obtained from a separate code that exclusively evaluates the thermal conductivity integrals (without solving the EPRT) with 500 grid points in frequency space. The values obtained from this code are 2.87, 98.8, and 32.52 W/mK, respectively. The EPRT code was run for varying numbers of grid points in the frequency domain until this near-match in con-

Table 3. Property values at $T = 300 \text{ K}^a$

Property	Diamond	SiO ₂
λ [J/ms K]	3,320	1.38
ρ [kg/m ³]	3,510	2,278
C [J/kg K]	517.1	740
v [m/s]	1.239×10^4	6.4×10^3
Λ [m]	4.431×10^{-7}	5.58×10^{-10}
τ [s]	8.949×10^{-12}	6.39×10^{-12}

^aMacroscopic properties for diamond from the NIST Web database and from Zeng and Chen [3] for SiO₂.

ductivities was obtained. The near-match in conductivities between the approaches both validates the implementation of the spectral relaxation time as well as determines that a 32-point discretization in frequency is adequate for the present conditions. Additional simulations from Majumdar [1] were reproduced using the present code as further validation (not shown). In addition, all cases were tested to ensure energy conservation within the steady-state solutions by integrating the net heat transfer into and out of the material domain.

4.3. Resolution Requirements

Further studies of grid dependence show that minimal resolution requirements needed to obtain grid independence of the predicted temperature profiles may be insufficient for obtaining grid independence of the heat flux rate (for steady-state conditions, the heat flux should be constant across a 1-D material domain). As shown below, these effects are due primarily to the large discrepancy between the phonon mean free paths in Si and SiO₂. In order to completely resolve a simulation the grid spacing, Δy must be of the order of the minimum mean free path within the specific configuration. For simulations involving SiO₂ this restricts the grid spacing to small dimensions, typically of order 10^{-10} m (the mean free path of SiO₂ is $5.58 \times 10^{-10} \text{ m}$). Therefore, use of the pure EPRT or BTE formulation can become computationally prohibitive for large film thicknesses, because of spatial resolution restrictions. In order to quantify these effects, 1-D and 2-D simulations were conducted by considering both the EPRT and BTE formulations as well as hybrid formulations for varying film thicknesses. The values for velocity and the Debye temperature for the EPRT simulations are those corresponding to the lower longitudinal branch of phonons in silicon given in Table 1. For the BTE simulations the aforementioned values have been taken to be 6,400 m/s and 658 K, respectively.

We first address the simulation of relatively very thin films (of relevant dimensions) on bulk substrates. Using the modified Holland's model and assuming a spectral relaxation time for monocrystalline silicon governed by Eq. (33), 1-D simulations of monocrystalline silicon on a bulk SiO₂ substrate were conducted. The DMM was incorporated in imposing the interface boundary conditions as described by Eqs. (38) and (39). The chosen dimensions of the two materials were typical of silicon-on-oxide (SOI) microdevices [2, 13]: 50 nm for silicon and 250 nm for the oxide substrate. The relaxation time for SiO₂ is calculated from the bulk thermal

conductivity [using Eq. (24) and the properties for SiO_2 listed in Table 3], since the dimension chosen for the substrate material is within the regime of continuum heat transfer [17].

The steady-state temperature profile is presented in Figure 4a for varying grid resolutions. A corresponding examination of the spatial and frequency dependence of the relaxation-time distribution shows that the mean free path of the dioxide is two orders of magnitude smaller than the smallest spectral component of the silicon relaxation time. Therefore, grid resolution in space becomes an important issue to ensure continuity of heat flux at the interface. However, simulations performed with four different grid configurations result in the almost identical temperature profiles shown in Figure 4a, where the nondimensional temperature is $T^* = (T - T_{Ly}) / (T_0 - T_{Ly})$. For this 1-D simulation the heat flux is also well resolved and grid-independent, with the error related to heat flux continuity at the interface being less than 3%. Unfortunately, for larger film thicknesses this minimal spatial resolution must still be maintained, thereby linearly increasing the total computational demands as a function of the (1-D) film thickness. Computational requirements will therefore quickly become prohibitive for large material dimensions; this is particularly true for 2-D films because of the added spatial coordinate resolution requirements. However, the very fact that the substrate dimension is much large than its mean free path implies that heat conduction within it is in the continuum regime. Therefore, this problem can be circumvented by incorporating Fourier's law (for the substrate) coupled with EPRT (or the BTE) for the microscale silicon layer for predicting the microscale heat transfer.

In order to investigate the hybrid approach, a new set of 1-D simulations was performed using both the pure BTE (or EPRT) and the hybrid BTE/ (or EPRT/) Fourier codes on a 1-D domain of thickness 325 nm, with the silicon layer being 75 nm thick. Using a relatively coarse grid of 48 spatial grid points (36 in the substrate), the temperature distributions obtained from the four approaches are almost identical, as observed in Figure 4b. However, a problem with heat flux continuity at the interface (absent in the hybrid approaches) is observed for the BTE/BTE and EPRT/EPRT simulations. In the case of the BTE-BTE simulation a small and erroneous heat flux "jump" is observed at the material interface due to the overly restrictive spatial resolution requirements of the SiO_2 material. However, the mean flux (within the Si) compares well with the hybrid simulation results. In contrast, the pure EPRT/EPRT simulations experience a larger heat flux discontinuity which cannot be eliminated (tested using up to 372 spatial grid points). The corresponding thermal conductivity calculated from the heat flux distribution and the temperature gradient in the bulk substrate is 2.54 W/m K. In contrast, the thermal conductivity of amorphous silicon dioxide is 1.38 W/m K [27]. As mentioned earlier, the procedure for the calculation of the spectrally independent relaxation time from Eq. (24) may be rather simplistic, and some of the assumptions associated with the Deybe theory are not entirely accurate in the case of amorphous solids [26]. The fact that the BTE/BTE simulation yields a heat flux consistent with the bulk thermal conductivity of SiO_2 is a direct consequence of fixing the specific heat capacity, phonon velocity, and mean free path (from their respective published values) via the kinetic theory and *a priori* knowledge of the thermal conductivity. In contrast, for the EPRT/EPRT simulation, only the velocity and the relaxation time are fixed, with no

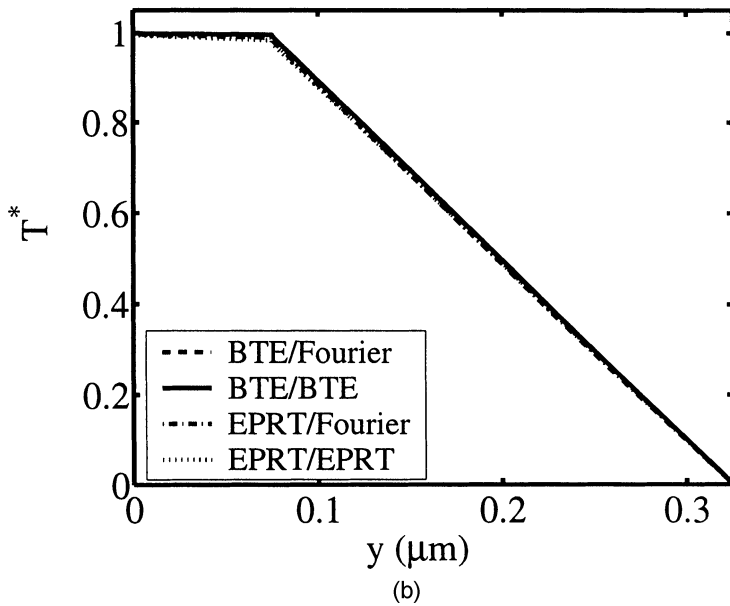
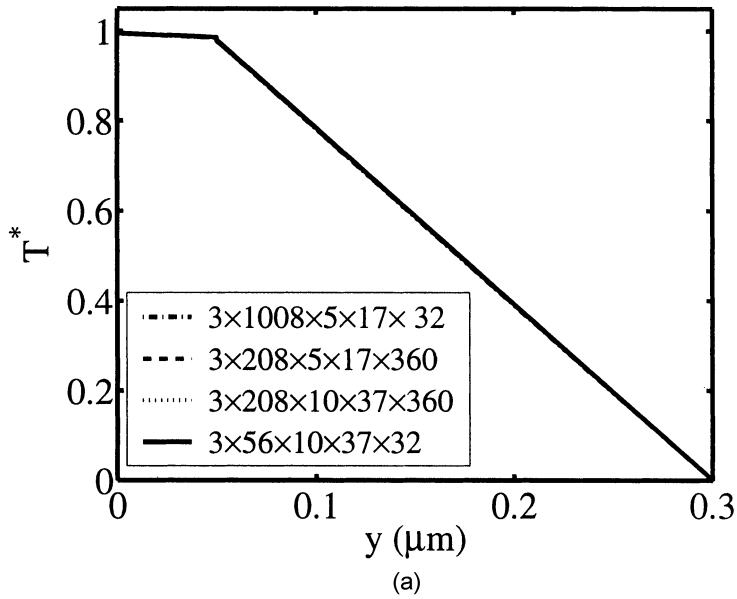


Figure 4. Steady temperature distributions as a function of y for 1-D Si/SiO₂ as a function of (a) grid resolution and (b) solution approach.

direct linkage to bulk conductivities. This is added to the fact that the procedure for initialization according to Eq. (14) may be technically inaccurate for amorphous silica, even though the predicted temperature distributions may not change significantly.

4.4. Total Simulation Times

Table 4 gives the performance of codes based on the four different formulations discussed in this work. All the simulations presented in this table have been performed using the ME Department's Beowulf cluster at Clemson University. Total simulation times for arbitrary numbers of processors and domain dimensions can be estimated from the base values reported in the table and the corresponding scaling plots presented previously. The 1-D Si/SiO₂ case involves a spatial domain of length 325 nm, with the silicon occupying 75 nm. The column "Simulation time" contains the approximate total clock time required to attain steady state. Note also that the 1-D simulations are overestimates of actual required mesh sizes and simulation times, due to the fact that the 2-D code is being solved using three grid points in the x direction. A true 1-D code would therefore require one-third fewer spatial grid points, and would not require discretization in ϕ , thereby greatly improving its performance. As noted above, the 2-D code was chosen for these simulations in order to more directly address scaling issues for general multidimensional heat conduction.

A measure of the accuracy of the simulations is also provided in Table 4 in the column "% $q_{y,error}$." For 1-D cases the steady heat flux should be constant; in these cases the error is therefore defined as $q_{y,error} = (q_{y,max} - q_{y,min})/q_{y,min}$. For 2-D simulations we alternatively define the error as $q_{y,error} = (q_{in} - q_{out})/q_{in}$, where q_{in} refers to the total heat flux entering the domain, obtained by integrating the component of heat flux perpendicular to the heated edge. Similarly, q_{out} is obtained by integrating the component of heat flux normal to the three nonheated edges. The angular resolution is $N_{\theta} = 10$ and $N_{\phi} = 37$ for all cases, while the spectral resolution for test cases involving the EPRT is $N_{\omega} = 32$. The spatial resolution, however, varies and has been informally optimized between numerical accuracy and computational expense (all cases maintain $\Delta x = \Delta y$). For the 1-D EPRT/EPRT simulation, 372 spatial grid points (y direction) were employed and the test case was run with 32 processors decomposing the frequency domain (as discussed above, the scaling with ω processors is nearly linear). For the 1-D BTE/BTE case also, the same spatial resolution was employed and the test case was run with the number of processors equal to the number of θ grid points (parallelization in θ). For both the 1-D EPRT/Fourier and the 1-D BTE/Fourier simulations, only 36 grid points in the

Table 4. Code performance for various test cases (simulation times correspond to the clock time to steady state)^a

Methodology	Mesh size	Number of CPUs	% $q_{y,error}$	Simulation time (min)
1-D EPRT/EPRT	13.2×10^6	32	6.4	750
1-D BTE/BTE	4.14×10^5	10	6.6	60
1-D EPRT/Fourier	1.704×10^6	16	4.1	75
1-D BTE/Fourier	5.34×10^4	1	1.7	15
2-D EPRT/Fourier	23×10^6	16	3.4	600
2-D BTE/Fourier	7.2×10^5	6	3.4	50

^aAll test cases involve a domain of dimension 325 nm (L_y) with a monocrystalline silicon film layer of 75 nm thickness. The dimension for 2-D test cases, in addition, also includes a width (L_x) of 250 nm.

y direction were employed and yet the error is smaller than in the previous two test cases. This is a direct consequence of iterating temperature at the interface so as to ensure heat flux continuity. As observed from the table, time-dependent microscale heat transfer simulations can be employed using a parallelized finite-difference solver for relevant material dimensions with reasonable accuracies and with reasonable simulation times.

4.5. Example Simulation: 2-D Si/SiO₂ Thin Film

As a final illustration of the capabilities of the developed code, results of a 2-D simulation are shown in Figure 5. A rectangular geometry with two materials (thin monocrystalline Si film on bulk SiO₂ substrate) is considered. In light of the facts mentioned in the previous section, the modeling approach involves the two hybrid approaches. Spatial grid resolution studies were conducted with the configurations of 36×48 , 72×148 , and 72×374 grid points, respectively, with almost identical resulting temperature and heat flux distributions. The attainment of steady state was gauged by performing a heat flux balance over the four edges of the two-material domain and evaluating the ratio $(q_{\text{out}} - q_{\text{in}})/q_{\text{in}}$. The 2-D EPRT/Fourier simulation was run on 16 ω processors, since increasing the number of processors beyond this only slowed down the simulations. However, for the reasons discussed in the parallelization section, the 2-D BTE/Fourier simulations were run on 6 processors by dividing the task of temperature calculation along the interface among 6 different processors.

The temperature distributions predicted by the EPRT/Fourier simulation and the BTE/Fourier simulation are shown in Figures 5a and 5b, respectively. The agreement between the hybrid EPRT and BTE predictions is within 2%. The reason for this might be attributed to the fact that the relaxation time used in the BTE/Fourier simulation is 6.4×10^{-12} s, while the spectral relaxation time at 300 K and 75 nm for the EPRT/Fourier simulation varies between approximately 6×10^{-11} s and 6×10^{-12} s, corresponding to a higher density of states. Hence, for the 2-D test case considered in this work, the BTE/Fourier simulation can be used with the gray approximation of constant relaxation time with relatively good accuracy. However, when the problem involves dissimilar boundary temperature values (with the difference being at least 100 K), the EPRT/Fourier method may be the appropriate technique, since the relaxation time is temperature-dependent. Figure 5c illustrates the deviation of the predicted temperature from the conventional Fourier's law applied to both the silicon thin layer and the bulk oxide substrate. The deviation has been calculated as $(T_{F,F} - T_{E,F})/T_{E,F}$, where the subscripts correspond to the EPRT and the Fourier's law formulations. Fourier's law underestimates the temperature prediction near and along the silicon film edges. This is because the temperature at the boundaries according to Fourier's law is fixed, whereas for the EPRT simulation it is only the intensity of the incoming phonons that is fixed by the prescribed boundary temperature. The intensity of outgoing (leaving the domain) phonons, however, evolves transiently, so that the temperature at the boundaries is not the same as the prescribed temperature. The same comments apply to the BTE/Fourier simulation as well. In contrast, at the bulk limit ($d_s > 100 \mu\text{m}$), the temperature along the outer edges of the silicon film layer tends to approach the

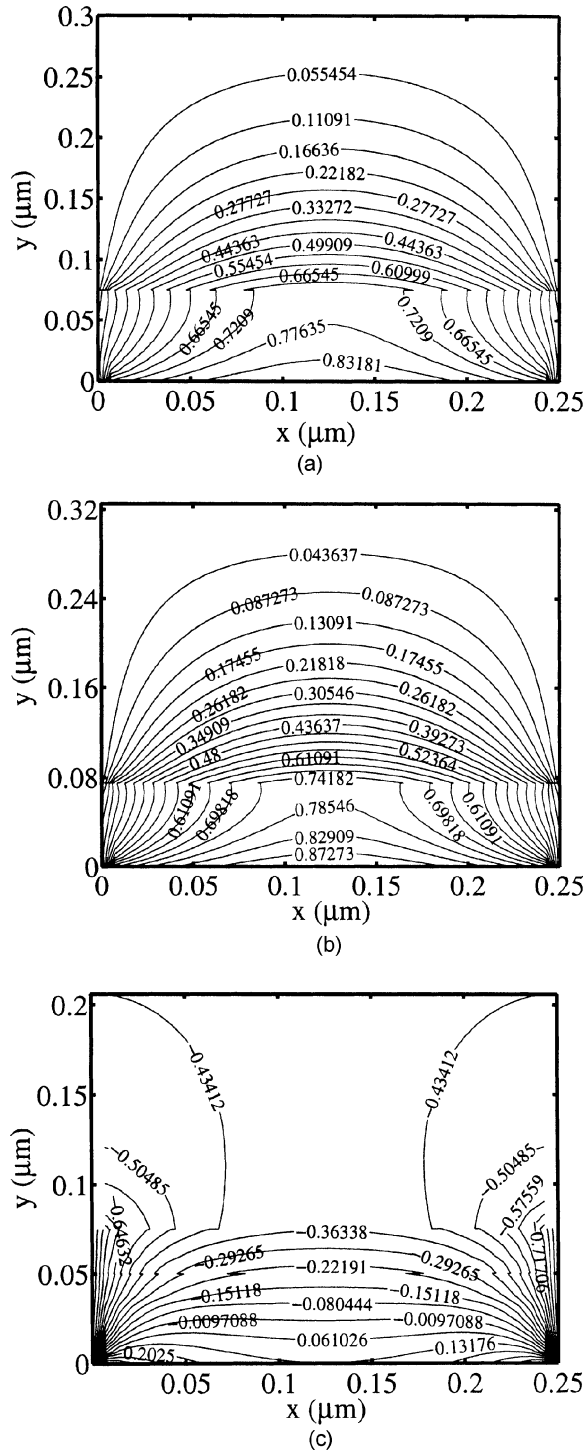


Figure 5. Steady nondimensional temperature contours in a 2-D Si/SiO₂ domain: (a) EPRT/Fourier solution, (b) BTE/Fourier solution, and (c) the relative difference between the EPRT/Fourier and a Fourier/Fourier solution.

imposed boundary value and the temperature distribution itself becomes consistent with that predicted by Fourier's law (not shown).

5. CONCLUSIONS

Numerical simulations of heat conduction in microscale-dimension thin layered films have been conducted for both the equation for phonon radiative transfer and a frequency-independent form of the Boltzmann transport equation. For cases in which the phonon mean free paths of the various film materials vary substantially, allowing one material to be described by the macroscopic Fourier's law, hybrid EPRT/Fourier, and BTE/Fourier approaches have been introduced. Two materials satisfying these criteria have been included for study: monocrystalline silicon, Si ($\Lambda = 409 \text{ \AA}$), and silicon dioxide, SiO₂ ($\Lambda = 5.58 \text{ \AA}$). A complete set of formulations has been presented for all sets of governing equations, including both interface scattering based on the DMM and a temperature and frequency-dependent relaxation time in the case of the EPRT. Simulations in both 1-D and 2-D spatial rectangular geometries have been investigated.

Efficient parallelization strategies have been presented for each of the equation sets described above based on direct finite differencing of the governing equations in spatial, polar angle, and frequency domains. Finite differencing has been studied as an alternative to statistical Monte Carlo-based solution algorithms, which are more difficult to parallelize and for unsteady results require many realizations to be conducted. Two Beowulf parallel computer clusters were used for the investigation, one based on a high-speed network and one based on a megabit-per-second "fast" ethernet. Code speed-up and parallelization efficiency have been studied as a function of parallel domain decomposition in each of the independent variables. Spatial domain decomposition is observed to be highly inefficient because of very large communication overhead. In contrast, 1-D domain decomposition in frequency space for the EPRT equation results in near-linear code speed-up for large numbers of processors (up to 90 were used in this study). For the BTE, or for the EPRT, when the number of processors desired is larger than the number of frequency grid points, decompositions in polar angle spaces are also found to be efficient. Other issues concerning parallelization of the various equations and interface boundary conditions have been presented, as well as example heat conduction results for both 1-D and 2-D layered thin films of Si and SiO₂.

REFERENCES

1. A. Majumdar, Microscale Heat Conduction in Dielectric Thin Films, *J. Heat Transfer*, vol. 115, pp. 7–16, 1993.
2. C. L. Tien, A. Majumdar, and F. Gerner, *Microscale Energy Transport*, Taylor & Francis, New York, 1998.
3. T. Zeng and G. Chen, Phonon Heat Conduction in Thin Films: Impacts of Thermal Boundary Resistance and Internal Heat Generation, *J. Heat Transfer*, vol. 123, pp. 340–347, 2001.
4. N. Neumann, R. Kohler, and G. Hofmann, Pyroelectric Thin Film Sensors and Arrays Based on P(VDF/TrFE), *Integrated Ferroelectrics*, vol. 6, pp. 213–230, 1995.

5. G. Teowee and D. R. Uhlmann, Modeling of a Thin Film Pyroelectric Pixel: Steady State Results, *Integrated Ferroelectrics*, vol. 22, pp. 411–419, 1998.
6. M. I. Flik, B. I. Choi, and K. E. Goodson, Heat Transfer Regimes in Microstructures, *J. Heat Transfer*, vol. 114, pp. 666–674, 1992.
7. A. A. Joshi and A. Majumdar, Transient Ballistic and Diffusive Phonon Heat Transport in Thin Films, *J. Appl. Phys.*, vol. 74, pp. 31–39, 1993.
8. G. Chen, Ballistic-Diffusive Equations for Transient Heat Conduction from Nano to Microscales, *J. Heat Transfer*, vol. 124, pp. 320–328, 2002.
9. G. Chen, Thermal Conductivity and Ballistic-Phonon Transport in the Crossplane Direction of Super-lattices, *Phys. Rev. B*, vol. 57, pp. 14958–14973, 1998.
10. S. Kumar, A. Majumdar, and C. L. Tien, The Differential-Discrete-Ordinate Method for Solutions of the Equation of Radiative Transfer, *J. Heat Transfer*, vol. 112, pp. 424–429, 1990.
11. C. Jen and C. Chieng, Microscale Thermal Characterization for Two Adjacent Dielectric Films, *J. Thermophys. Heat Transfer*, vol. 12, pp. 146–152, 1998.
12. A. Raisi and A. A. Rostami, Phonon Heat Conduction in Thin Films: Impacts of Thermal Boundary Resistance and Internal Heat Generation, *Numeric Heat Transfer A*, vol. 41, pp. 373–391, 2002.
13. P. G. Sverdrup, J. Y. Sungtaek, and K. E. Goodson, Sub-continuum Simulations of Heat Conduction in Silicon-on-Insulator Transistors, *J. Heat Transfer*, vol. 123, pp. 130–137, 2001.
14. T. Klitsner, J. E. Van Cleve, H. E. Fischer, and R. O. Pohl, Phonon Radiative Heat Transfer and Surface Scattering, *Phys. Rev. B*, vol. 38, pp. 7576–7594, 1988.
15. J. M. Ziman, *Electrons and Phonons*, Oxford University Press, London, UK, 1960.
16. G. Chen, Nonlocal and Nonequilibrium Heat Conduction in the Vicinity of Nanoparticles, *J. Heat Transfer*, vol. 118, pp. 539–545, 1996.
17. M. Ashegi, M. N. Touzelbaev, K. E. Goodson, Y. K. Leung, and S. S. Wong, Temperature-Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates, *J. Heat Transfer*, vol. 120, pp. 30–36, 1998.
18. M. G. Holland, Analysis of Lattice Thermal Conductivity, *Phys. Rev.*, vol. 132, pp. 2461–2471, 1963.
19. E. H. Sondheimer, The Mean Free Path of Electrons in Metals, *Adv. Phys.*, vol. 50, pp. 505–515, 2001.
20. K. C. Sood and M. K. Roy, Longitudinal Phonons and High-Temperature Heat Conduction in Germanium, *J. Phys.: Condensed Matter*, vol. 74, pp. 301–312, 1992.
21. Y. S. Ju and K. E. Goodson, Phonon Scattering in Silicon Films with Thickness of Order 100 nm, *Appl. Phys. Lett.*, vol. 74, pp. 3005–3007, 1999.
22. E. T. Swartz and R. O. Pohl, Thermal Boundary Resistance, *Rev. Mod. Phys.*, vol. 61, pp. 605–668, 1989.
23. S. Mazumder and A. Majumdar, Monte Carlo Study of Phonon Transport in Solid Thin Films Including Dispersion and Polarization, *J. Heat Transfer*, vol. 123, pp. 749–759, 2001.
24. G. Chen, Ballistic-Diffusive Equations for Transient Heat Conduction from Nano to Macroscales, *J. Heat Transfer*, vol. 124, pp. 320–328, 2002.
25. M. A. Heaslet and R. F. Warming, Radiative Transport and Wall Temperature Slip in an Absorbing Planar Medium, *Int. J. Heat Mass Transfer*, vol. 8, pp. 979–994, 1965.
26. G. Chen and C. L. Tien, Thermal Conductivities of Quantum Well Structures, *J. Thermophys. Heat Transfer*, vol. 7, pp. 311–318, 1993.
27. Y. S. Touloukian, R. W. Powell, C. Y. Ho, and P. G. Klemens, *Thermophysical Properties of Matter*, IFI/Plenum, New York, 1970.