Adjoint-based deviational Monte Carlo methods for phonon transport calculations

Jean-Philippe M. Péraud and Nicolas G. Hadjiconstantinou

Phys. Rev. B **91**, 235321 — Published 30 June 2015

DOI: 10.1103/PhysRevB.91.235321

Adjoint-based deviational Monte Carlo methods for phonon transport calculations

Jean-Philippe M. Péraud and Nicolas G. Hadjiconstantinou,
Department of Mechanical Engineering, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

June 3, 2015

Abstract

In the field of linear transport, adjoint formulations exploit linearity to derive powerful reciprocity relations between a variety of quantities of interest. In this paper, we develop an adjoint formulation of the linearized Boltzmann transport equation for phonon transport. We use this formulation for accelerating deviational Monte Carlo simulations of complex, multiscale problems. Benefits include significant computational savings via direct variance reduction, or by enabling formulations which allow more efficient use of computational resources, such as formulations which provide high resolution in a particular phase-space dimension (e.g. spectral). We show that the proposed adjoint-based methods are particularly well suited to problems involving a wide range of lengthscales (e.g. nanometers to hundreds of microns) and lead to computational methods that can calculate quantities of interest with a cost that is independent of the system characteristic length scale, thus removing the traditional stiffness of kinetic descriptions. Applications to problems of current interest, such as simulation of transient thermoreflectance experiments or spectrally resolved calculation of the effective thermal conductivity of nanostructured materials are presented and discussed in detail.

1 Introduction

Nanoscale heat transport mediated by phonons has received considerable attention in recent years [1,2], both due to the scientific challenges arising from the failure of Fourier's law at small scales, as well as the potential applications to nanoscale engineering [3]. In this context, numerical methods for solving the Boltzmann transport equation are invaluable because they enable the solution of problems of practical interest but also provide insight into the physics of phonon transport. Due to the high dimensionality associated with the Boltzmann equation, important problems of practical interest remain computationally expensive, if not intractable, making the development of new and more efficient methods very desirable, especially for treating multiscale problems.

This paper focuses on the benefits that can be derived by exploiting the observation that many phonon transport problems of scientific and practical interest involve

relatively small driving forces (temperature differences) and thus may be treated by the linearized Boltzmann equation. Recent work [4] has shown that using the linearized Boltzmann equation in the presence of temperature differences of magnitude smaller than 10% of the reference temperature (e.g. temperature differences on the order of ±30K for a reference temperature of 300K) results in errors on the order of a few percent. Examples of problems featuring small temperature differences include calculations of the effective thermal conductivity of nanostructures (where, in fact, small temperature differences are *required* to prevent a nonlinear response), simulation of transient thermoreflectance experiments [5], and simulations of the thermal behavior of thin films, nanowires and superlattices. This observation was already used in a previous paper [5] in which a kinetic-Monte-Carlo-type scheme for simulating the Boltzmann equation was developed. In that work, linearity enabled the decoupling of deviational particle trajectories leading to an algorithm that was significantly faster and easier to code while exhibiting no timestep error [4, 5].

In the present paper, we present formulations which again exploit linearity of the governing equation to gain computational advantages (speedup, simplicity). The developments presented here, however, are completely distinct from the work presented in [5], but at the same time complementary–that is, the computational advantages of the two can be compounded (multiplicatively in the case of speedup). They are based on an adjoint formulation which exploits the duality between the linearized Boltzmann equation and its adjoint.

The adjoint formulation for the phonon Boltzmann equation in the frequency-dependent relaxation-time approximation is derived and discussed in section 3. We note that adjoint formulations have been developed in other domains of linear transport (e.g. radiation, neutron transport) [6,7] and have served as inspirations for this work. Acceleration techniques based on this new formulation are discussed in section 4 in the context of problems of practical interest. In sections 5 and 6 we discuss the use of adjoint formulations for developing schemes that are particularly suited to multiscale problems.

2 Background

The non-linear Boltzmann equation for phonon transport in the relaxation time approximation can be written in the form

$$\frac{\partial f}{\partial t} + \mathbf{V}_g \cdot \nabla_{\mathbf{x}} f = \frac{f^{\mathrm{loc}} - f}{\tau} \tag{1}$$

where $f(\mathbf{x}, \omega, p, \mathbf{\Omega}, t)$ is the occupation number of phonon modes, $\mathbf{V}_g$ denotes the phonon group velocity (obtained from the dispersion relation) and the unit vector $\mathbf{\Omega}$ denotes the phonon traveling direction. Here, we use

$$f_T^{\mathrm{eq}} = \frac{1}{\exp\left(\frac{\hbar\omega}{k_b T}\right) - 1} \tag{2}$$

to denote the Bose-Einstein distribution with temperature parameter T; under this notation, $f^{\text{loc}} = f^{\text{eq}}_{T_{\text{loc}}}$ is a Bose-Einstein distribution with $T = T_{\text{loc}}$. The latter temperature, T_{loc}, is referred to as the"pseudo-" temperature and is defined such that the scattering processes remain strictly energy-conservative (see, for example, [8]). In the isotropic model considered here, the relaxation time $\tau = \tau(\omega, p, T)$ depends on the phonon frequency ω, the polarization p, and the temperature T.

It has been shown previously [9] that the deviational, energy-based Boltzmann equation,

$$\frac{\partial e^{\text{d}}}{\partial t} + \mathbf{V}_g \cdot \nabla_{\mathbf{x}} e^{\text{d}} = \frac{(e^{\text{loc}} - e^{\text{eq}}_{T_{\text{eq}}}) - e^{\text{d}}}{\tau(\omega, p, T)} \tag{3}$$

where $e^{\text{d}} = e - e^{\text{eq}}_{T_{\text{eq}}}$, $e = \hbar\omega f$ and $e^{\text{eq}}_{T_{\text{eq}}} = \hbar\omega f^{\text{eq}}_{T_{\text{eq}}}$, lends itself naturally to Monte Carlo solution of phonon transport problems, especially for problems involving small deviations from equilibrium. In such simulations, energy conserving deviational particles represent the distribution

$$\frac{D(\omega, p)}{4\pi}(e - e^{\text{eq}}_{T_{\text{eq}}}) \tag{4}$$

The "control" temperature T_{eq} is chosen by balancing simplicity (of the resulting algorithm) with computational efficiency (maximizing variance reduction); because of the small deviation from equilibrium, the "optimal" value of this parameter is close, if not equal, to the system reference temperature (note that T_{eq} can be spatially variable; this is discussed at length in [4] and section 5 of this paper).

When temperature deviations are sufficiently small, Eq. (3) can be linearized by approximating the non-linear scattering operator using the expansion

$$\frac{e^{\text{eq}}_{T_{\text{loc}}} - e^{\text{eq}}_{T_{\text{eq}}} - e^{\text{d}}}{\tau(\omega, p, T)} = \frac{\frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT}(T_{\text{loc}} - T_{\text{eq}}) - e^{\text{d}}}{\tau(\omega, p, T_{\text{eq}})} + \mathcal{O}\left[(T_{\text{loc}} - T_{\text{eq}})^2\right], \tag{5}$$

leading to

$$\frac{\partial e^{\text{d}}}{\partial t} + \mathbf{V}_g \cdot \nabla_{\mathbf{x}} e^{\text{d}} = \frac{\mathcal{L}(e^{\text{d}}) - e^{\text{d}}}{\tau}, \tag{6}$$

where the operator $\mathcal{L}$ is given by

$$\mathcal{L}(e^{\text{d}}) = \frac{\int \frac{D}{4\pi\tau} e^{\text{d}} d\omega d^2\mathbf{\Omega}}{\int \frac{D}{\tau} \frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT} d\omega} \frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT} \tag{7}$$

Here and in what follows, the sum over polarizations is implied by the integral over frequencies ω. Moreover, in the interest of simplicity, the integration range for variables $\omega, \mathbf{\Omega}, \mathbf{x}$ will be shown explicitly under a different integral sign only when different from the whole phase space associated with the problem of interest; in the case of time, the integration range will be shown if different from $(-\infty, \infty)$.

In addition to the above, we will also be using the following notation:

- The *deviational* temperature will be denoted by T (instead of $T - T_{\text{eq}}$).

- The mode-dependent free path $\Lambda_{\omega,p}$ is defined as the product $V_g(\omega,p)\tau(\omega,p,T_{\text{eq}})$.

- We define the mean free path as

$$\langle\Lambda\rangle = \frac{\int D\frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT}\Lambda_{\omega,p}d\omega}{\int D\frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT}d\omega} \tag{8}$$

- The frequency and polarization-dependent Knudsen number $\text{Kn}_{\omega,p}$ is defined as $\Lambda_{\omega,p}/L$, where L is the smallest characteristic lengthscale in the problem. The Knudsen number based on the mean free path is defined as $\langle\text{Kn}\rangle = \langle\Lambda\rangle/L$.

- The specific heat capacity C is given by

$$C = 4\pi\int \Xi d\omega \tag{9}$$

where

$$\Xi(\omega,p) \equiv \frac{D(\omega,p)}{4\pi}\frac{de^{\text{eq}}_{T_{\text{eq}}}(\omega)}{dT} \tag{10}$$

2.1 Kinetic Monte Carlo for linearized problems

The linearized BTE (6) lends itself to very efficient simulation methods. Here we summarize the Kinetic Monte Carlo (KMC) algorithm described in [5,10] that is particularly efficient for the types of problems considered here and will be referred to extensively in this work.

One of the key features of the KMC method is that computational particles are treated independently and thus sequentially. Let N denote the total number of particles; what follows describes the calculation of the trajectory of each particle as a sequence of linear (straight-line) segments separated by scattering events or collisions with boundaries:

- Randomly draw the particle initial properties from the source distribution [4] which includes contributions from the initial condition, boundary conditions, heat sources etc. Sources in linear phonon transport are discussed in detail in [4]. Each particle is assigned a (constant) weight called the "effective energy", $\mathcal{E}_{\text{eff}}$, which corresponds to the total energy emitted by the source divided by the total number of particles to be used by the simulation. In steady problems, the "effective energy" has the unit of an energy rate.

- Calculate the particle trajectory until the time the particle exits the computational domain (via absorbing boundaries, or when the particle leaves the time domain of interest in time-dependent cases) by repeating the following steps:

- Calculate the time between the latest scattering event i and the next scattering event, $i+1$, using $\Delta t_i = -\tau(\omega_i, T_{\text{eq}}, p_i)\ln(R)$, where R is a uniform random variate in the interval $(0,1)$. The next scattering time is $t_i + \Delta t_i$ at location $\mathbf{x}_i + \mathbf{V}_{g,i}\Delta t_i$.
- If no boundary is encountered between $\mathbf{x}_i$ and $\mathbf{x}_i + \mathbf{V}_{g,i}\Delta t_i$, then the particle's updated position is $\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{V}_{g,i}\Delta t_i$. If on the other hand one or several boundaries are encountered along this trajectory, the next position is the intersection point between the segment $[\mathbf{x}_i, \mathbf{x}_i + \mathbf{V}_{g,i}\Delta t_i]$ and the first boundary. The time of scattering event $i+1$, t_{i+1}, is set to the time of encounter with the boundary.
- If $\mathbf{x}_{i+1}$ corresponds to a scattering event, then the frequency, polarization and traveling direction of the particle are reset. The new properties are drawn from the linearized post-scattering distribution

$$\frac{D(\omega,p)}{4\pi\tau(\omega,p,T_{\text{eq}})}\mathcal{L}(e^{\text{d}})(\omega,p) \tag{11}$$

- If $\mathbf{x}_{i+1}$ corresponds to an encounter with a boundary, properties will be updated depending on the type of boundary (e.g. diffusely reflective, partially transmissive, etc). An absorbing boundary simply terminates the current particle trajectory.

- Accumulate the contribution of the calculated trajectories to the quantities of interest. For instance, if the quantity of interest is the average temperature in a given region of space at time t_{measure}, then a given particle contributes to the estimate if it is located within that volume when the time is equal to t_{measure}. In that case, the contribution $\mathcal{E}_{\text{eff}}/(CN)$ is added to the estimate.

3 The adjoint Boltzmann equation

3.1 Background

The adjoint formulation is best introduced in a framework where boundary and initial conditions are incorporated into the governing equation as (special) sources of deviational particles. We remind the reader that in the deviational and linearized formulations, sources can emit positive or negative particles [4, 5, 9].

In what follows, we will use q to denote the generalized source term, namely the sum of all particle sources in a given problem. From this definition, it follows that energy-based deviational particles are emitted from $(4\pi)^{-1}D(de^{\text{eq}}_{T_{\text{eq}}}/dT)q$. We also recall that, here, T denotes the deviational temperature. With these definitions in mind, the deviational Boltzmann equation reads

$$\frac{\partial\psi}{\partial t} + \mathbf{V}_g\cdot\nabla\psi = \frac{\mathcal{L}(\psi)-\psi}{\tau} + q \tag{12}$$

where $\psi = e^{\mathrm{d}}(de^{\mathrm{eq}}_{T_{\mathrm{eq}}}/dT)^{-1}$ and the linearized operator $\mathcal{L}$ can now be written as

$$\mathcal{L}(\psi) = \frac{\int \frac{\Xi}{\tau}\psi d\omega d^2\mathbf{\Omega}}{4\pi \int \frac{\Xi}{\tau} d\omega} \tag{13}$$

We also define the scalar product

$$\langle \phi, \psi \rangle = \int \phi \Xi \psi d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt \tag{14}$$

with respect to which $\mathcal{L}/\tau$ is self-adjoint; namely,

$$\langle \phi, \mathcal{L}(\psi)/\tau \rangle = \langle \mathcal{L}(\phi)/\tau, \psi \rangle \tag{15}$$

In addition to sources, Monte Carlo simulations and experimental setups are also characterized by detectors, which sample phonons as a means of returning "measurements" of quantities of interest. Mathematically, a detector is defined by its characteristic function h; the quantity of interest, $\mathcal{I}$, is then written as

$$\mathcal{I} = \int h \frac{D}{4\pi} e^{\mathrm{d}} d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt = \int h \Xi \psi d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt \tag{16}$$

The function h prescribes both the type of quantity that is estimated (temperature, heat flux...) and the location (in phase space, including time) over which the quantity is averaged. For example, for the average deviational temperature within a volume V at time t such that $t_1 < t < t_2$, h is given by

$$h = \frac{1}{CV(t_2 - t_1)} \mathbb{1}_V \mathbb{1}_{[t_1, t_2]} \tag{17}$$

where $\mathbb{1}_V$ refers to the indicator function of V, i.e. the function that takes the value 1 inside the volume V and 0 otherwise. For the temperature at a given time t_0, h would instead be given by

$$h = \frac{1}{CV} \mathbb{1}_V \delta(t - t_0) \tag{18}$$

where $\delta(t - t_0)$ refers to the Dirac delta function centered in time on t_0. Although these expressions might not always seem intuitive, they can be verified by considering an equilibrium system at (deviational) temperature T: in the linearized framework, $e^{\mathrm{d}} = T de^{\mathrm{eq}}_{T_{\mathrm{eq}}}/dT$; substituting in equation (16) leads to $\mathcal{I} = T$.

3.2 The fundamental relation

We now introduce the adjoint Boltzmann equation

$$-\frac{\partial \psi^*}{\partial t} - \mathbf{V}_g \cdot \nabla_{\mathbf{x}} \psi^* = \frac{\mathcal{L}(\psi^*) - \psi^*}{\tau} + h \tag{19}$$

In this equation, particles simulating the adjoint distribution, ψ^*, evolve backwards in time and are emitted by the adjoint source, h, which is the function characterizing the detector in the original problem. The specification of the adjoint problem is completed by using the source q as the adjoint detector, in the sense

$$\mathcal{I}^* = \int q \Xi \psi^* d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt \tag{20}$$

The importance of the adjoint formulation can be summarized by the relation

$$\mathcal{I}^* = \mathcal{I} \tag{21}$$

which we will refer to as the fundamental relation. In words, this relation implies that any quantity of interest (of the form (16)) can be obtained by solving the adjoint problem which uses the detector (of the original problem), h, as a source and the source (of the original problem), q, as detector. Based on the observation that the adjoint equation describes particles that move backwards in time, we will frequently use the term "backward problem" to describe the adjoint problem defined by equations (19) and (20); in analogy, we will use the term "forward" to describe the original problem defined by equations (12) and (16).

To prove the fundamental relation we write

$$\mathcal{I}^* = \int \left[\frac{\partial \psi}{\partial t} + \mathbf{V}_g \cdot \nabla \psi - \frac{\mathcal{L}(\psi) - \psi}{\tau} \right] \Xi \psi^* d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt \tag{22}$$

$$= \int \psi \Xi \left[-\frac{\partial \psi^*}{\partial t} - \mathbf{V}_g \cdot \nabla \psi^* - \frac{\mathcal{L}(\psi^*) - \psi^*}{\tau} \right] d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt \tag{23}$$

$$= \int \psi \Xi h d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt = \mathcal{I} \tag{24}$$

Obtaining expression (23) from (22) requires integration by parts and, depending on the problem of interest, some manipulation.

We now discuss this integration for the term involving the time derivative. The use of sources for imposing initial conditions allows us to extend the integration over time from $-\infty$ to ∞ by taking $\psi(t < 0) = 0$ and $\psi^*(t > t_{\text{final}}) = 0$ where t_{final} denotes the last detector instance. As a result:

$$\int_{t=-\infty}^{\infty} \frac{\partial \psi}{\partial t} \Xi \psi^* dt = [\psi \Xi \psi^*]_{-\infty}^{\infty} - \int_{t=-\infty}^{\infty} \psi \Xi \frac{\partial \psi^*}{\partial t} dt \tag{25}$$

$$= - \int_{t=-\infty}^{\infty} \psi \Xi \frac{\partial \psi^*}{\partial t} dt \tag{26}$$

We now consider the term

$$\int_{\mathbf{x} \in X} \int \mathbf{V}_g \cdot \nabla \psi \Xi \psi^* d\omega d^2\mathbf{\Omega} d^3\mathbf{x} \tag{27}$$

which can be written in the form

$$-\int_{\partial X}\int \mathbf{V}_g \cdot \mathbf{n}\psi\Xi\psi^* d\omega d^2\mathbf{\Omega} d^2\mathbf{x} - \int_{\mathbf{x}\in X}\int \psi\Xi\mathbf{V}_g \cdot \nabla\psi^* d\omega d^2\mathbf{\Omega} d^3\mathbf{x} \tag{28}$$

where $\mathbf{n}$ is the inward-pointing normal vector to the boundary ∂X. The above proof requires the first term in (28) to vanish. This will be established for various boundary conditions of interest below. In the case where the spatial domain is unbounded, one may proceed by assuming (as was done in this work) that the integral over the boundary ∂X tends to zero when the latter is made infinitely large. A sufficient condition for this is that $\psi\psi^* \to 0$ sufficiently fast as $\mathbf{x} \to \infty$; this is expected to be satisfied by problems that can be simulated by the deviational Monte Carlo method.

In addition to periodic boundary conditions, the most commonly encountered boundary conditions in phonon transport literature are diffusely/specularly reflective and prescribed temperature boundaries. The case of diffusely/specularly reflective boundaries is treated in section 3.4; prescribed temperature boundaries are treated in Appendix A. Periodic boundary conditions are discussed in section 4.2.

3.3 Adjoint particle dynamics and simulation

Comparison of the adjoint BTE (19) and the original linearized BTE (12) reveals strong similarities, suggesting that algorithms for performing forward simulations could also be used for backward simulations with small modifications. As expected, one difference between the two lies in the source term. In the forward case the energy-based particles are emitted from the distribution Ξq. By analogy, the adjoint particles must be emitted from the distribution Ξh. In contrast to the forward case where $\int \Xi q d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt$ has the unit of energy, $\int \Xi h d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt$ will not, in general, have the unit of energy. Nonetheless, energy will be conserved provided the number of computational particles is conserved during scattering events, since this guarantees

$$\int \frac{\psi^*\Xi}{\tau} d\omega d^2\mathbf{\Omega} = \int \frac{\mathcal{L}(\psi^*)\Xi}{\tau} d\omega d^2\mathbf{\Omega} \tag{29}$$

Note that although the quantity $\mathcal{E}^*_{\text{eff}} = \int \Xi h d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt/N$ does not always represent an energy, we will still refer to it as "adjoint effective energy".

The second difference can be found in the rules for calculating a particle trajectory. The minus signs in the left-hand side of (19) means that:

- the time parameter of a particle monotonically decreases
- a particle with parameter $\mathbf{\Omega}$, moves in the $-\mathbf{\Omega}$ direction.

In practice, the isotropy of the collision operator and typical boundary conditions (e.g. diffuse reflection, prescribed temperature boundary) means that the "backward" algorithm differs very little from the "forward" algorithm.

3.4 Reflecting boundaries

Let us consider a point $\mathbf{x}_\mathrm{b}$ on a reflective boundary, whose inward (towards the material) pointing normal is denoted by $\mathbf{n}$. When a particle encounters a diffusely reflective boundary, it is reflected back, and its traveling direction is randomized, such that the outgoing distribution is isotropic. As a result, the distribution ψ, for a given frequency and polarization, obeys the following relation at the boundary ($\mathbf{x} = \mathbf{x}_\mathrm{b}$):

$$\psi(\mathbf{\Omega} \cdot \mathbf{n} > 0) = -\frac{1}{\pi} \int_{\mathbf{\Omega}\cdot\mathbf{n}<0} \psi \mathbf{\Omega} \cdot \mathbf{n} d^2\mathbf{\Omega} \tag{30}$$

Since particles subject to the adjoint Boltzmann equation travel backward in time, the diffusely reflective boundary conditions for the adjoint distribution ψ^* reads:

$$\psi^*(\mathbf{\Omega} \cdot \mathbf{n} < 0) = \frac{1}{\pi} \int_{\mathbf{\Omega}\cdot\mathbf{n}>0} \psi^* \mathbf{\Omega} \cdot \mathbf{n} d^2\mathbf{\Omega} \tag{31}$$

We may now use (30) and (31) to write

$$\int \mathbf{\Omega} \cdot \mathbf{n} \psi \psi^* d^2\mathbf{\Omega} = \int_{\mathbf{\Omega}\cdot\mathbf{n}<0} \mathbf{\Omega} \cdot \mathbf{n} \psi \psi^* d^2\mathbf{\Omega} + \int_{\mathbf{\Omega}\cdot\mathbf{n}>0} \mathbf{\Omega} \cdot \mathbf{n} \psi \psi^* d^2\mathbf{\Omega} \tag{32}$$

$$= \psi^*(\mathbf{\Omega} \cdot \mathbf{n} < 0) \int_{\mathbf{\Omega}\cdot\mathbf{n}<0} \mathbf{\Omega} \cdot \mathbf{n} \psi d^2\mathbf{\Omega} + \psi(\mathbf{\Omega} \cdot \mathbf{n} > 0) \int_{\mathbf{\Omega}\cdot\mathbf{n}>0} \mathbf{\Omega} \cdot \mathbf{n} \psi^* d^2\mathbf{\Omega} \tag{33}$$

$$= -\pi \psi^*(\mathbf{\Omega} \cdot \mathbf{n} < 0) \psi(\mathbf{\Omega} \cdot \mathbf{n} > 0) + \pi \psi(\mathbf{\Omega} \cdot \mathbf{n} > 0) \psi^*(\mathbf{\Omega} \cdot \mathbf{n} < 0) = 0 \tag{34}$$

which proves that the surface integral over the diffusively reflective boundary is zero.

For specular reflective walls, proving that

$$\int \mathbf{\Omega} \cdot \mathbf{n} \psi \psi^* d^2\mathbf{\Omega} = 0 \tag{35}$$

follows by noticing that if both ψ and ψ^* satisfy the specular reflection condition, then so does their product.

4 Applications

As briefly discussed in section 1, the adjoint formulation can provide a number of computational benefits, including algorithmic simplicity and considerable computational speedup for certain classes of problems. The latter can be described as problems in which the "detector is small", that is, problems for which the outputs of interest are defined over small regions of physical space, or more generally, phase space. An example of the former is the transient thermoreflectance experiment discussed in the next section, in which the quantity of interest is the temperature at the specimen surface

(which, stricly speaking, has zero volume in three dimensions); an example of the latter is spectrally resolving the contribution of individual phonon modes to the effective thermal conductivity (the detector extends over a small range of frequencies, or, in the most challenging case, features a delta function in frequency). In these problems, in the "forward" Monte Carlo method, the probability for a particle to be found in the detector at a given time is small. The adjoint formulation uses h as a source thus providing an opportunity for alleviating this burden. If the source is larger than the detector, the adjoint formulation ensures that the signal collected by the detector will be enhanced, leading to improved signal (variance reduction). Clearly, the speedup will depend on the size-ratio between the detector and the source; in cases where the detector features a delta function and the source does not, the speedup is theoretically infinite (in practice the forward calculation would smear the delta function into a computational bin in order to collect some samples, thus making the speed-up finite, but introducing error in the process).

Examples of applications of the adjoint formulation are given in the following sections. Note that although the adjoint formulation is indifferent to the numerical implementation (i.e. timestep based, or KMC-type), here we will proceed to demonstrate these methods using the KMC-type algorithm developed in [5] and briefly described in section 2.

4.1 Surface temperature in a transient thermoreflectance experiment

This section illustrates the adjoint formulation using an example of engineering and scientific interest, namely, the pump-probe thermoreflectance experiment [11, 12].

4.1.1 Background

We briefly recall the configuration of the experiment that we consider here; note that several versions of pump-probe thermoreflectance exist, all with their own advantages and shortcomings. More details can be found in [5, 9]. A layer of aluminum (approximately 100nm-thick) lies on a silicon wafer, considered semi-infinite. Figure 1 depicts the system geometry and the coordinate system used in the calculations. At time $t = 0$, the aluminum is heated by a laser pulse. The resulting (deviational) temperature field in the aluminum at $t = 0$ is given by

$$T_i(\mathbf{x}) = \hat{T} \exp\left(-\beta z - \frac{2r^2}{R_0^2}\right) \tag{36}$$

where $\hat{T}$ is taken as 1 K. Here, the penetration depth β^{-1} is taken to be 7nm and the characteristic radius R_0 is taken to be 15 microns. More details on the model parameters, such as the transmission coefficient at the aluminum-silicon interface, can be found in [9] as well as appendix C. Also, we recall that heat transfer by electron transport is neglected in this example.

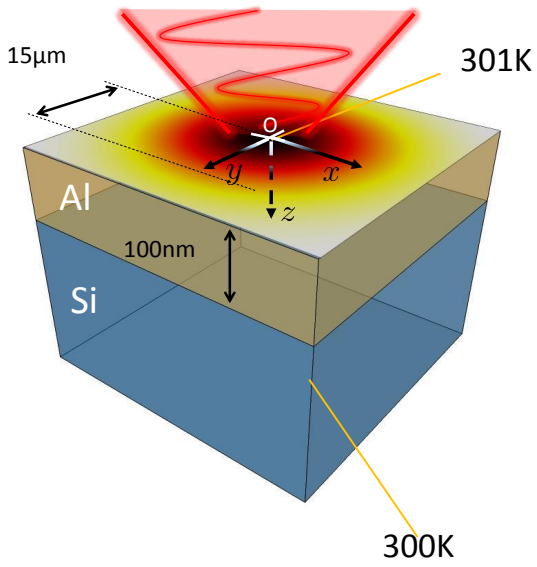

Figure 1: Schematic of a transient thermoreflectance experiment. Point O denotes the center of the heating pulse, also taken to be the origin of the cartesian (x, y, z) set of axes. The system is assumed infinite in the $z > 0$ direction and in the $x - y$ plane.

This problem features only one source term (the initial condition), which can be written as

$$q = T_i(\mathbf{x})\delta(t) \tag{37}$$

The quantity of interest is the surface temperature at time t_j, $j = 1, ..., M$. The function h_j for the corresponding detector is $\mathbb{1}_{\text{disk}}\delta(z)\delta(t-t_j)$; here, we consider the slightly more general case of the temperature in a general and arbitrary volume V

$$h_j = \mathbb{1}_V \frac{1}{VC}\delta(t - t_j) \tag{38}$$

because as shown below, the adjoint formulation lends itself to this generalization naturally. Note here that, for simplicity, we will use the same symbol V to denote the region of interest and its volume.

4.1.2 Adjoint calculation

Let us consider here the case of one sampling time, namely t_M; extension to multiple sampling times is discussed in section 4.1.3. A particle from the corresponding adjoint source (forward detector) h_M is emitted at time t_M and travels backward in time. At $t = 0$, the position $\mathbf{x}_{\text{end}}$ is noted, leading to

$$\mathcal{I}^*_{M,i} = \mathcal{E}^*_{\text{eff}}\hat{T}\exp\left(-\beta z_{\text{end}} - \frac{2r^2_{\text{end}}}{R_0^2}\right) \tag{39}$$

as the contribution of particle i to the estimate of the temperature at time t_M. Here, the weight of each particle, or adjoint effective energy, is given by

$$\mathcal{E}^*_{\text{eff}} = \frac{1}{N}\int \Xi h_j d^2\mathbf{\Omega} d\omega d^3\mathbf{x} dt \tag{40}$$

$$= \frac{1}{N}\frac{1}{VC}\int_{\omega,\mathbf{x}} 4\pi\Xi\mathbb{1}_V d\omega d^3\mathbf{x} = \frac{1}{N} \tag{41}$$

and is independent of the (forward) detector shape. The temperature is thus given by

$$T(t = t_M) = \sum_{i=1}^{N} \mathcal{I}^*_{M,i} \tag{42}$$

4.1.3 Multiple sampling times

To treat other sampling times $t_j \neq t_M$, in principle we have to simulate new particles starting at time $t = t_j$ and measure their position (and contribution) at time $t = 0$. However, by noting that the evolution rules for particles emitted at t_j are the same for all j (only the "internal clock" of the particles differs), we may reuse the information given by the trajectory of the particle emitted at time t_M, by simply recording their contributions at times $t_M - t_j$ for all j.

Ultimately, this process amounts to setting $t = 0$ when the particle is emitted, then to counting the time forward while computing the trajectory. Contributions can then be sampled at times t_j, exactly like in the "forward" Monte Carlo method. Algorithmically, the only difference lies in the exchange of source and detector.

4.1.4 Computational results

Figure 2 shows the temperature variation as a function of time at three locations, as measured by the distance ρ from the origin, on the sample surface (in this case the forward detector is a point on the surface $z = 0$). We note here that the profiles at $\rho = 6\mu$m and $\rho = 12\mu$m were obtained using the same particles as for the $\rho = 0$ calculation by exploiting the translational invariance of the problem. Namely, since translation of the particle source in the x (or y) direction leaves the particle trajectory unaltered, translation of the adjoint detectors should also result in equivalent results. This implies that contributions to the temperature at distance ρ from the origin can be calculated using data from particles from the original calculation using the "shifted" detector

$$\mathcal{I}^*_{M,i} = \frac{\hat{T}}{N} \exp\left(-\beta z_{\text{end}} - 2\frac{(x_{\text{end}} - \rho)^2 + y^2_{\text{end}}}{R_0^2} \right) \tag{43}$$

Figure 3 shows the standard deviation in the temperature measurement in the case of the forward method. In this method, the temperature on the surface is measured in a cylindrical bin of depth (measured from the surface) d. The figure clearly shows that the statistical uncertainty (and thus the statistical accuracy of the results) deteriorates as d is made smaller. On the other hand, d needs to be made as small as possible to minimize *deterministic* errors resulting from averaging over a finite volume (rather than strictly on the surface). In the limiting case where the detector is simply a disk at the surface, calculating the surface temperature using a (forward) Monte Carlo method becomes impossible (unless time discretization is introduced, in which case it is "just" very expensive) since the probability that a particle hits the interface at a specified time is 0. Using the adjoint makes such a calculation possible by switching the source and the detector.

4.1.5 Discussion

Although the above example clearly highlights the computational gains made available by the adjoint method in principle, here we note that the magnitude of the computational gain *in this particular experimental setup* [13] is hard to quantify: the physical detector is a laser probe which theoretically measures the surface temperature by relating it to the surface reflectivity. Since the reflection of photons at a surface involves a penetration to some (small) depth, a more accurate model of this process would take into account that the measured quantity uses a finite depth such as $d \approx 2$ nm (range of the optical skin depth for visible light in aluminum). Consequently, the benefit from

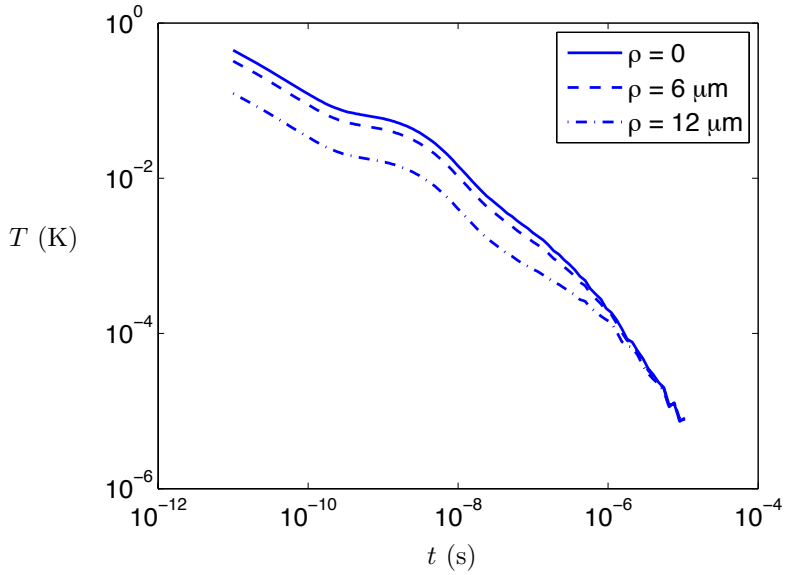

Figure 2: Temperature as a function of time in a transient thermoreflectance experiment. Results shown for the pulse center, 6μm away from the pulse center, and 12μm away from the pulse center.

using the adjoint should be determined by comparing the size of the detector with the size of the source properly adjusted for the above effects.

From a broader perspective, physical detectors usually can only access the surface of a given system, whereas phonons are generated via mechanisms which are inherently volumetric (Joule effect, electron-phonon interaction). For these reasons, accurate and faithful description of physical experiments is, in general, expected to strongly benefit from the adjoint formulation.

Finally, we note that drawing random particles from the distributions derived from the detectors is usually easier than from the forward source terms. For example, a temperature detector usually weighs all samples within a volume equally and thus calls for the creation of a uniform distribution when used as an (adjoint) source. In contrast, the distribution associated with the initial temperature field (36) in the above example is a product of a decaying exponential and a Gaussian. Although this distribution is invertible, this would not necessarily be the case with more general initial conditions.

4.2 Highly resolved calculations of mode-specific thermal conductivity calculations

Another class of methods where the detector is "small" includes problems for which the quantity of interest needs to be spectrally resolved. Previous work [14] has highlighted the fact that the contribution of low frequency phonon modes is challenging to resolve

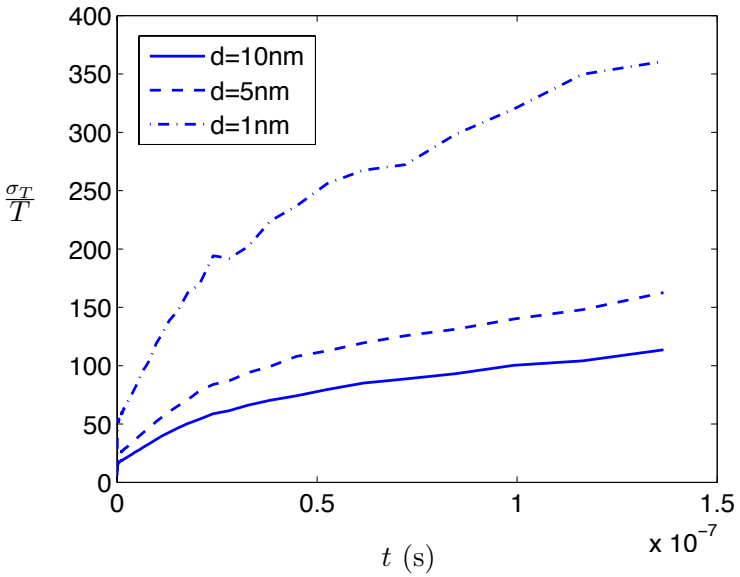

Figure 3: Ratio between the standard deviation in the temperature measurement of σ_T, and the temperature, for 3 cylindrical detectors with height 10nm, 5nm and 1nm.

due to their small densities of states and very large free paths. Unfortunately, due to their low density of states, the forward Monte Carlo technique tends to "under-resolve" estimates of their heat flux contributions, while on the other hand it "over-resolves" the contributions of phonons with the highest densities of states. Increasing the number of samples in order to reach the desired level of resolution at low frequencies will reduce the statistical uncertainty for every frequency, hence wasting computational resources.

The adjoint formulation lends itself naturally to this situation. The quantities of interest here are the heat flux contributions from individual phonon modes (in the isotropic relaxation time approximation, this corresponds to bins in phonon frequency).

To illustrate the method, we will study a nanostructure that has been considered in recent work [8] and calculate the contribution of each phonon frequency to the thermal conductivity. Specifically, we analyze a single period of the porous periodic structure shown in Figure 4. The system is subjected to a temperature gradient, and periodic boundary conditions are applied [8]. As explained in [4,5], applying a spatially variable control with uniform gradient results in strictly periodic boundary conditions for the deviational quantity e^{d} and particles are emitted from the source term

$$Q = \Xi q \quad \text{outside the pore} \tag{44}$$

where $q = -\mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_{\mathrm{eq}}$. To spectrally resolve the effective thermal conductivity, we need to calculate the heat flux for a given frequency "bin" $[\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2]$ and a given polarization. We are interested in the response in the direction of the applied

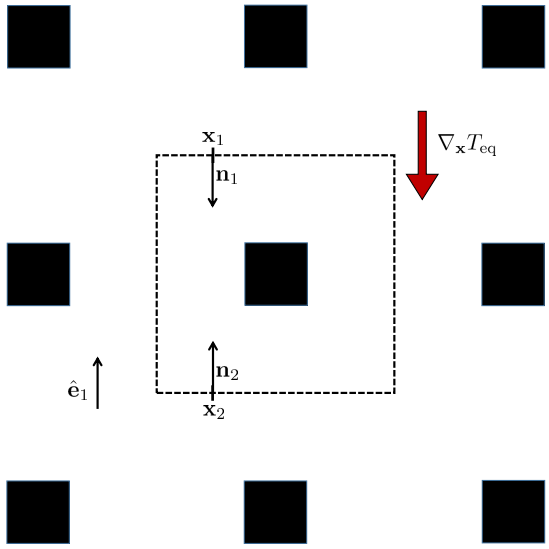

Figure 4: Sketch of the nanoporous structure studied in section 4.2. The dashed square represents the boundary of the computational domain, along which periodic boundary conditions are applied (see Ref. [5, 8]).

temperature gradient. In other words, the characteristic function for this detector is

$$h = \mathbb{1}_{[\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2]} \mathbf{V}_g \cdot \hat{\mathbf{e}}_1 \tag{45}$$

where $\hat{\mathbf{e}}_1$ is the unit vector in the direction of the applied temperature gradient–see Figure 4. The adjoint approach is only valid if

$$\int_{\partial X} \int \mathbf{V}_g \cdot \mathbf{n} \psi \Xi \psi^* d\omega d^2\mathbf{\Omega} = 0 \tag{46}$$

where ∂X refers to the boundary of the square computational domain and the square pore, and where $\mathbf{n}$ is the normal vector pointing inward. The diffuse reflective surface of the pore was treated in 3.1. To show that (46) is true, we may simply notice that the periodic boundary condition imposes $e^{\mathrm{d}}(\mathbf{x}_1, \boldsymbol{\omega}) = e^{\mathrm{d}}(\mathbf{x}_2, \boldsymbol{\omega})$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ are corresponding points of two opposite sides of the periodic boundary condition. As a result, $e^{\mathrm{d}}(\mathbf{x}_1, \boldsymbol{\omega}) \mathbf{V}_g \cdot \mathbf{n}_1 = -e^{\mathrm{d}}(\mathbf{x}_2, \boldsymbol{\omega}) \mathbf{V}_g \cdot \mathbf{n}_2$, leading to the desired result after integration over the boundary domain.

We introduce the adjoint equation and the adjoint source $q^* = h$. Adjoint particles are then emitted from

$$\Xi q^* = \frac{D}{4\pi} \frac{d e^{\mathrm{eq}}_{T_{\mathrm{eq}}}}{dT} \mathbb{1}_{[\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2]} \mathbf{V}_g \cdot \hat{\mathbf{e}}_1 \tag{47}$$

and assigned the weight

$$\frac{1}{N}\int \Xi q^* d\omega d^2\mathbf{\Omega} d^3\mathbf{x} = \frac{1}{N}\frac{D(\omega_0,p)}{4}\left.\frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT}\right|_{\omega_0} V_g(\omega_0,p)\Delta\omega. \tag{48}$$

We note once again that the resulting backward algorithm is nearly identical to the forward one as explained in [4,5] and section 3.3, with the main difference being that the initial frequency/polarization (and the resulting velocity) properties can now be chosen by the practitioner (instead of being randomly drawn from the distribution Ξq).

Figure 5 shows results calculated using the adjoint method for two different pore sizes (25nm and 50nm). These results were obtained using the method for terminating particle trajectories described in [5]; trajectories were terminated after 30 scattering events. These results confirm that low frequency (large free path) phonons may play a critical role in the design of nanostructures for efficient thermoelectric materials. In Figure 6, we show the same result with the forward method using the same overall number of particles. We clearly see how the quality of the results deteriorates in the very low frequency (large free path) regime. In other words, obtaining the insights shown by Figure 5 with the forward method is significantly more costly. For example, we found that the statistical uncertainty associated with the contributions of particles with free paths of 10 μm, 100 μm and 1 mm were 10 times, 25 times and 70 times smaller, respectively, when calculated using the adjoint method rather than the forward method. These uncertainties correspond to speedup factors of approximately 100, 600 and 5000. In reality, the observed speedup will be somewhat smaller due to the following considerations: first, because trajectories of particles with long free paths–which are larger than the system periodicity and thus require more operations to calculate–are more frequently sampled in the backward calculation (for a given number of particles), this method is approximately 5 times more expensive than the forward method. Second, the quality of the solution using the backward method is worse than that of the forward method for high frequencies. The latter can be rectified at small computational cost by customizing the number of particles used for each particular frequency range. This is possible in the backward case because the detector is no longer frequency specific, ensuring that all particles emitted in a particular frequency range will contribute. The forward method does not allow such a flexibility. The results in figure 5 were calculated using the same number of particles for each frequency bin.

5 Spatially variable control temperature in the adjoint framework

So far we have discussed situations where the control, e_{control}, is a constant. Previous work, both in the rarefied gas domain [15, 16] and the phonon domain [4, 5], has used spatially variable controls as a means of accelerating the computation (variance reduction) [16] or introducing externally imposed driving forces (e.g. a temperature gradient

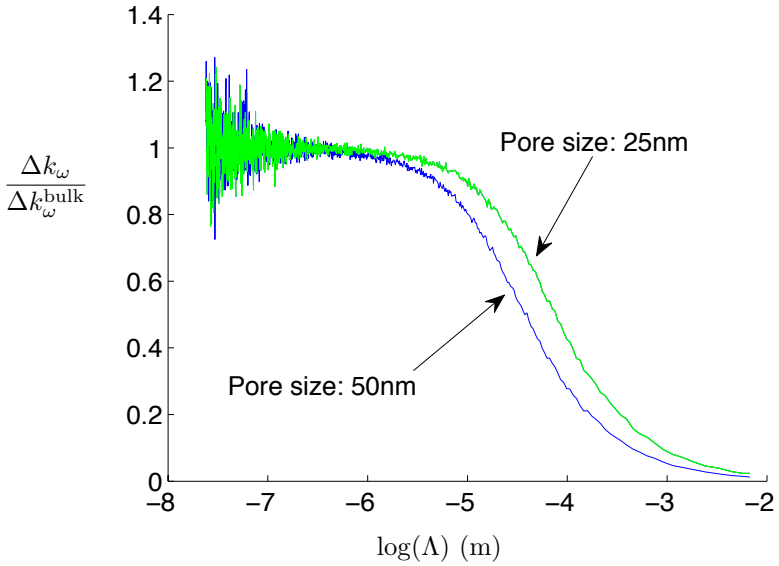

Figure 5: Frequency-resolved differential contribution to the thermal conductivity (measured heat flux per unit temperature gradient) from the longitudinal acoustic (LA) modes in the problem defined in figure 4. Result is normalized by the corresponding frequency-resolved differential contribution to the bulk thermal conductivity, and calculated using the adjoint method. Results shown for square pores of side 25nm and 50nm; the spacing between the pores is 2 microns in both cases. These calculations used 28000 particles per frequency cell, for a total of 1399 frequency cells (for a total of approximately 40 million particles).

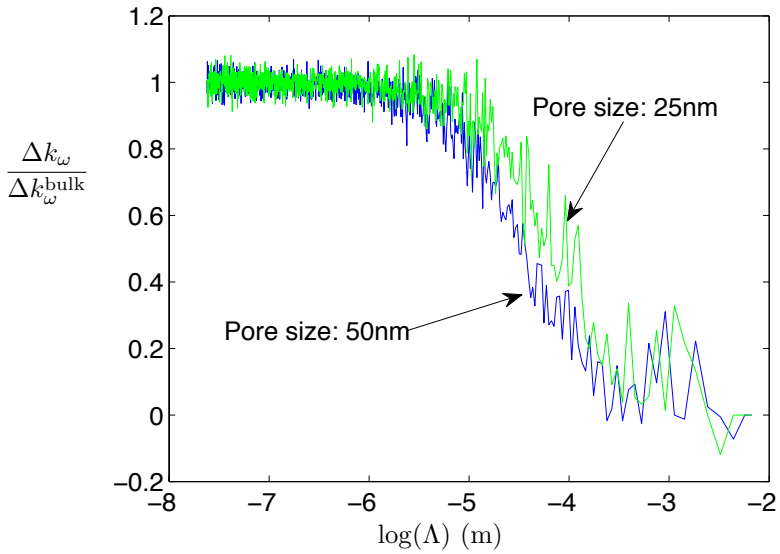

Figure 6: Frequency-resolved differential contribution to the thermal conductivity (measured heat flux per unit temperature gradient) from the longitudinal acoustic (LA) modes in the problem defined in figure 4. Result is normalized by the corresponding frequency-resolved differential contribution to the bulk thermal conductivity, and calculated using the forward method. Results shown for square pores of side 25nm and 50nm; the spacing between the pores is 2 microns in both cases. These calculations used a total of 40 million particles.

for the calculation of the effective thermal conductivity of a material [5]). As an example, consider a variable control of the form $e_{\text{control}}(\mathbf{x}) = e^{\text{eq}}_{T_{\text{eq}}} + T_0(\mathbf{x}) de^{\text{eq}}_{T_{\text{eq}}}/dT$ which results in the following equation governing e^{d}:

$$\frac{\partial e^{\text{d}}}{\partial t} + \mathbf{V}_g \cdot \nabla_{\mathbf{x}} e^{\text{d}} = \frac{\mathcal{L}(e^{\text{d}}) - e^{\text{d}}}{\tau} - \mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0 \frac{de^{\text{eq}}_{T_{\text{eq}}}}{dT} \tag{49}$$

Under these dynamics, particles are emitted from the distribution $\Xi \mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0$ in the bulk and from the distribution $e_b(\mathbf{x}_b) - e_{\text{control}}(\mathbf{x}_b)$ at the system boundaries. In the linearized setting, we may write $e_b(\mathbf{x}_b) = e^{\text{eq}}_{T_{\text{eq}}} + T_b de^{\text{eq}}_{T_{\text{eq}}}/dT$. Thus, $e_b(\mathbf{x}_b) - e_{\text{control}}(\mathbf{x}_b) = (T_b - T_0(\mathbf{x}_b)) de^{\text{eq}}_{T_{\text{eq}}}/dT$. To simplify the discussion, we will assume that, by choice, $T_0(\mathbf{x})$ obeys Dirichlet boundary conditions for prescribed temperature boundaries and Von Neumann boundary conditions for reflective boundaries. This allows us to eliminate the boundary effects from the present discussion, although extending the conclusions of this paragraph to more general choices of T_0 is straightforward.

Drawing particles from $\Xi \mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0$ can be a significant programming burden if the spatial dependence of T_0 is complicated. Let us explore the implications of applying an adjoint approach to this situation; we consider here the steady state case. Particles are emitted from the detector function, while the quantity of interest is given by

$$\mathcal{I}^* = \int \left[-\mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0(\mathbf{x}) \right] \Xi \psi^* d\omega d^2\mathbf{\Omega} d^3\mathbf{x} dt. \tag{50}$$

The contribution of particle j, with weight $\dot{\mathcal{E}}_{\text{eff}}$, to the final estimate can be written as

$$\mathcal{I}^*_j = \dot{\mathcal{E}}^*_{\text{eff}} \int_{t=0}^{t_{\text{end}}} -\mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0(\mathbf{x}(t)) dt \tag{51}$$

where the time t is only used formally to parametrize the line integral along the particle trajectory (see Refs. [5, 10]). In other words, the value of the line integral does not depend of the direction of the time parametrization, which is consistent with the fact that the time is absent from the steady state adjoint equation. We may considerably simplify this expression by introducing the particle coordinates at the scattering points $\mathbf{x}_i$. Since the trajectory is a series of N_{seg} linear segments delimited by the points $\mathbf{x}_i$, expression (51) becomes

$$\mathcal{I}^*_j = \dot{\mathcal{E}}^*_{\text{eff}} \sum_{i=0}^{i=N_{\text{seg}}-1} \left(T_0(\mathbf{x}_{i+1}) - T_0(\mathbf{x}_i) \right). \tag{52}$$

The fact that particles travel in the opposite direction of $\mathbf{V}_g$ is important for deriving the above expression, since the line integral over a segment may be written as

$$\int_{t_i}^{t_{i+1}} -\mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0 \left(\mathbf{x}_i - (t - t_i)\mathbf{V}_g \right) dt = \left[T_0 \left(\mathbf{x}_i - (t - t_i)\mathbf{V}_g \right) \right]_{t_i}^{t_{i+1}} \tag{53}$$

$$= T_0(\mathbf{x}_{i+1}) - T_0(\mathbf{x}_i) \tag{54}$$

Equation (52) straightforwardly simplifies into

$$\mathcal{I}_j^* = \dot{\mathcal{E}}_{\text{eff}}^* \left(T_0(\mathbf{x}_{N_{\text{seg}}}) - T_0(\mathbf{x}_0) \right) \tag{55}$$

This result appears powerful in the sense that the source q can be handled by evaluating the value of T_0 at the emission and termination points (with the latter usually given by boundary conditions) instead of generating random samples from the distribution $\Xi \mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0$; this represents a considerable simplification in most cases.

However, this result needs to be put into context by comparing with the case of the adjoint algorithm with fixed control. In the case of fixed control (and no other sources– i.e. the same problem studied above) the source term only includes the prescribed temperature boundaries. In other words, the same development as above leads to

$$\mathcal{I}_j^* = \dot{\mathcal{E}}_{\text{eff}}^* T_0(\mathbf{x}_{N_{\text{seg}}}) \tag{56}$$

which only differs from (55) by the term $T_0(\mathbf{x}_0)$. The latter term is usually fixed (when the estimate is calculated at one point only).

This means that, although the adjoint formulation led to considerable simplification (removing the need to sample the source term in equation (49)), the statistical uncertainty of the adjoint formulation with spatially variable control is not smaller– in fact, it may be higher–than the statistical uncertainty of the adjoint calculation with a fixed control. In other words, the adjoint formulation with the source term $e_{\text{source}} = e_{T_{\text{eq}}}^{\text{eq}} + T_0(\mathbf{x}) de_{T_{\text{eq}}}^{\text{eq}}/dT$ is not expected to provide improved variance reduction compared to the adjoint formulation with a fixed control, in contrast to forward, time-step based algorithms where additional variance reduction is observed when a (suitably chosen) spatially variable control is used [16]. On the other hand, the control $e_{\text{source}} = e_{T_{\text{eq}}}^{\text{eq}} + T_0(\mathbf{x}) de_{T_{\text{eq}}}^{\text{eq}}/dT$ remains useful for imposing a temperature gradient for effective thermal conductivity calculations [4].

Fortunately, significantly reduced variance is indeed possible with a spatially variable control within the adjoint formulation. In fact, as we show below, the improved variance can be achieved while retaining the simplification resulting from avoiding the generation of samples from complex distributions. Such formulations are discussed in the following section.

6 Implementing asymptotically-derived controls through the adjoint approach

Previous work using spatially variable controls for improved variance reduction [15, 16] utilized the local equilibrium–based on real-time (cell-based) estimates of its parameters– as a control. One drawback of this approach is that the resulting discontinuities in the control (at cell boundaries) require particle generation at cell boundaries, which becomes cumbersome in higher dimensions [16]. Here, we introduce a new approach which uses asymptotic solutions of the Boltzmann equation as controls and show how the adjoint

formulation can make such approaches more efficient as well as simpler to code. This section considers steady state problems only, although extension to transient problems will be considered in future work.

6.1 Asymptotic control for steady multiscale problems

Let $T_0(\mathbf{x})$ be the solution of Laplace's equation with *ad hoc* boundary conditions. It is shown in [17–19] that

$$e_1^{\mathrm{d}}(\mathbf{x}) = \frac{de_{T_{\mathrm{eq}}}^{\mathrm{eq}}}{dT}\left[T_0(\mathbf{x}) + \langle \mathrm{Kn}\rangle \left(T_K(\mathbf{x}) + T_{G1}(\mathbf{x})\right) - \tau \mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0(\mathbf{x})\right] \tag{57}$$

is a first-order asymptotic solution of the steady Boltzmann equation in the expansion parameter $\langle \mathrm{Kn}\rangle$ (assumed small) and subject to arbitrary kinetic (Boltzmann) boundary conditions. In this expression, T_{G1} denotes a solution of the heat equation with boundary conditions that are determined, self-consistently, by the asymptotic analysis once the kinetic boundary condition is specified. The same analysis determines T_K, a kinetic boundary layer in the vicinity of the boundaries which blends the equilibrium distribution at the wall with the bulk distribution (which is clearly non equilibrium). The interested reader is referred to [17–19] for more details.

Here, we adopt a heuristic approach which amounts to including the most readily available first order term of (57) in the control, namely choose

$$e_{\mathrm{control}} = e_0 - \mathbf{V}_g \tau \cdot \nabla_{\mathbf{x}} T_0 \frac{de_{T_{\mathrm{eq}}}^{\mathrm{eq}}}{dT} \tag{58}$$

By noting that $\mathcal{L}(\mathbf{V}_g \tau \cdot \nabla_{\mathbf{x}} T_0) = 0$, the BTE for $e^{\mathrm{d}} = e - e_{\mathrm{control}}$ can now be written in the form

$$\mathbf{V}_g \cdot \nabla_{\mathbf{x}} e^{\mathrm{d}} = \frac{\mathcal{L}(e^{\mathrm{d}}) - e^{\mathrm{d}}}{\tau} + \tau \mathbf{V}_g \cdot \nabla\left(\mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0(\mathbf{x})\right) \frac{de_{T_{\mathrm{eq}}}^{\mathrm{eq}}}{dT} \tag{59}$$

The source term that appears is composed of all the second order derivatives of T_0. It can be explicitly written as the double sum

$$V_g^2 \tau \sum_i \sum_j \Omega_i \Omega_j \frac{\partial^2 T_0}{\partial x_i \partial x_j} \tag{60}$$

Drawing particles from such a distribution as is required in forward frameworks is very challenging. In addition to this volumetric source, other source terms appear at the boundaries, from the mismatch (anisotropy) between the control and the boundary condition. For instance, for a prescribed temperature boundary, the modified boundary condition reads:

$$e^{\mathrm{d}}(\omega, p, \mathbf{\Omega} \cdot \mathbf{n} > 0) = \mathbf{V}_g \cdot \nabla_{\mathbf{x}} T_0 \frac{de_{T_{\mathrm{eq}}}^{\mathrm{eq}}}{dT} \tag{61}$$

On the other hand, in the case of the adjoint formulation, for the source given in (60), and using the same procedure used for (50) to (55), the contribution of trajectory (particle) j can be shown to be

$$\mathcal{I}_j^* = \dot{\mathcal{E}}_{\text{eff}}^* \sum_{i=0}^{N_{\text{seg}}-1} \tau_i \mathbf{V}_{g,i} \cdot (\nabla_{\mathbf{x}} T_0(\mathbf{x}_i) - \nabla_{\mathbf{x}} T_0(\mathbf{x}_{i+1})) \tag{62}$$

where τ_i and $\mathbf{V}_{g,i}$ respectively refer to the characteristic relaxation time and the velocity vector of the particle on segment i. A source term of type (61) is treated by adding $\mathbf{V}_g \tau \cdot \nabla_{\mathbf{x}} T_0(\mathbf{x}_{N_{\text{seg}}})$ to the contribution. This cancels the last term of expression (62) for $i = N_{\text{seg}} - 1$ (all trajectories terminate at the boundaries).

Finally, we need to recall that the final result will be obtained by adding the stochastic estimate to the deterministic value represented by the control. The deterministic value for the temperature is T_0. In the case of the heat flux, the deterministic heat flux associated with the control is $k_{\text{bulk}} \nabla_{\mathbf{x}} T_0$.

6.2 Validation and accuracy

In this section we validate the method described in section 6.1, which we will refer to as asymptotically controlled adjoint (ACA), using the following two-dimensional problem. We consider an infinitely long slab of material of thickness $2L$. We denote the coordinate in the infinite direction by x_1 and the coordinate in the other direction by x_2. At $x_2 = L$ the material is held at a prescribed (deviational) temperature $T_{\text{b}} = T_{\text{eq}} \epsilon \cos(2\pi x_1/(3L))$; at $x_2 = -L$ the deviational temperature is given by $T_b = -T_{\text{eq}} \epsilon \cos(2\pi x_1/(3L))$. Here, ϵ denotes a small quantity; in other words we are interested in the linear regime around a reference temperature $T_{\text{eq}} = 300$K. By linearity, the discussions that follow do not depend on the dimensionless coefficient ϵ; all calculations were performed with $\epsilon = 1/300$.

The system was chosen because the solution of Laplace's equation can be obtained analytically:

$$T_0(x_1, x_2) = \epsilon T_{\text{eq}} \cos\left(\frac{2\pi x_1}{3L}\right) \frac{\sinh\left(\frac{2\pi x_2}{3L}\right)}{\sinh\left(\frac{2\pi}{3L}\right)} \tag{63}$$

This will allow us to focus our validation on the stochastic error only. This solution is plotted in Figure 7.

This phonon transport problem can be easily solved using either the forward Monte Carlo method or the adjoint method. In Fig. 8, we show the temperature calculated on 51 equispaced points of the line parametrized by $x_1 = 0$ and $0 \leq x_2 \leq 1$, using both the adjoint method with uniform control, and the ACA method presented above, for $\langle \text{Kn} \rangle = 0.5$ and $\langle \text{Kn} \rangle = 0.1$. In the interest of simplicity, both calculations used the single mean free path model (constant relaxation time and Debye model); in other words, $\Lambda_{\omega,p} = V_g \tau = \Lambda = \textit{constant}$. The agreement between the two methods is excellent.

Figure 9 shows the statistical uncertainty associated with the calculation of the x_2 component of the heat flux at point $(0, 0)$. It clearly reveals that, in the ACA method,

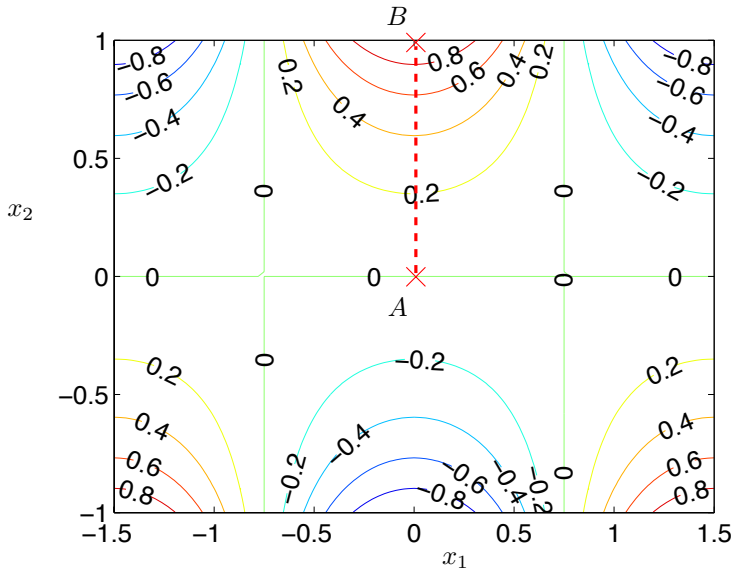

Figure 7: Contour plot of the solution (63) of Laplace's equation in a thin film with sinusoidal Dirichlet boundary conditions.

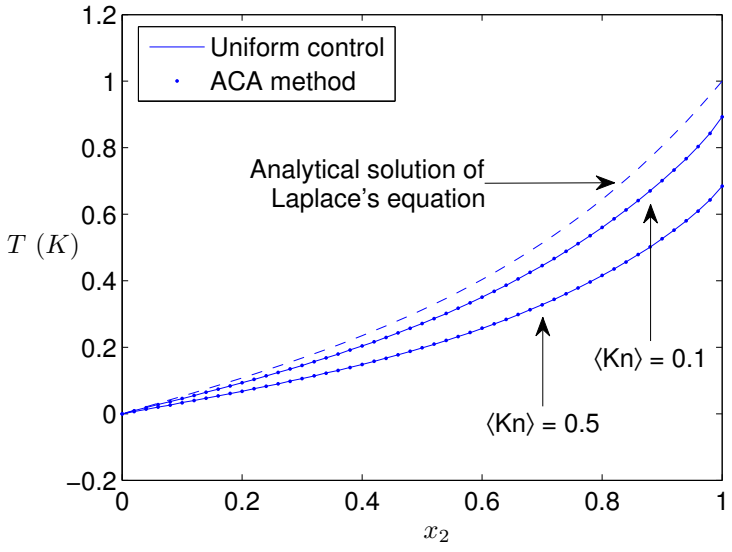

Figure 8: Temperature profile along the segment AB in Figure 7. Comparison between the solution to the BTE calculated using the adjoint method of section 3.1, and the ACA method of section 6.1. The analytical solution of Laplace's equation is also shown.

the standard deviation scales linearly with the Knudsen number, while in the adjoint method with uniform control it is approximately constant. In appendix B we provide a mathematical explanation for the scaling observed in the ACA case. In particular, we highlight the major difference that arises, in terms of statistical properties, when a temperature field other than the 0th order solution, T_0, is used as control and show that using a solution of Laplace's equation is key to this result.

This result is of great importance for multiscale simulations because it means that, for a fixed uncertainty, low Knudsen number systems (large lengthscales) can be simulated using the ACA technique at a *fixed computational cost as* $\langle \mathrm{Kn} \rangle$ *decreases*. This follows from the fact that the cost of computing a single particle trajectory increases proportionally to $\langle \mathrm{Kn} \rangle^{-2}$ as $\langle \mathrm{Kn} \rangle \to 0$, since characteristic transport timescales follow a diffusive scaling in this regime. On the other hand, the statistical uncertainty scales as $\sigma/\sqrt{N}$, where σ is the standard deviation of particle contributions to the estimate and N is the number of particles used; therefore for a fixed statistical uncertainty the scaling $\sigma \propto \langle \mathrm{Kn} \rangle$ requires $N \propto \langle \mathrm{Kn} \rangle^2$. Since the overall cost per simulation scales with the product of the number of particles times the cost of a single trajectory, we obtain a constant cost. In contrast, the cost of methods which have a constant statistical uncertainty as a function of $\langle \mathrm{Kn} \rangle$ (such as traditional MC methods, as well as forward deviational methods) increases as $\langle \mathrm{Kn} \rangle^{-2}$ in the $\langle \mathrm{Kn} \rangle \to 0$ limit, a manifestation of the kinetic description becoming stiff in this limit. In other words, *the ACA formulation overcomes this stiffness and results in a computational method that can simulate large systems as efficiently as small systems, a highly desirable feature of any multiscale method* [20]. In fact, for classes of problems for which the cost of computing a single trajectory scales as $\langle \mathrm{Kn} \rangle^{-1}$, such as the case of solving for the heat flux at a location close to the boundary[1] discussed in the next section, the cost of the ACA formulation is expected to scale as $\langle \mathrm{Kn} \rangle$ and therefore *decrease as lengthscales increase*.

We note that the above features pertain to problems for which the adjoint method is primarily suited for, namely problems in which the solution of interest is the transport field in a small region of space (see section 7 for further discussion). It should also be noted that the above scaling estimates for the cost refer to the case where the acceptable uncertainty level is prescribed in an absolute sense. In some cases, for instance when calculating the heat flux (which is formally a quantity that also scales with $\langle \mathrm{Kn} \rangle$ [17–19]), it may be more appropriate to consider the uncertainty in a relative sense – namely, the ratio between the uncertainty and the calculated heat flux. A multiscale method that features a constant cost as a function of $\langle \mathrm{Kn} \rangle$ at constant relative uncertainty would have to use a control which includes the higher order terms presented in [17] and is a direct extension of the present work.

Figure 10 shows the uncertainty associated with the calculation of the x_2 component of the heat flux at points $(0, 0)$ and $(0, L)$ when using a material model with frequency-dependent free path (for a description of the material model see Appendix C). This

[1] This scaling may be arrived at by applying the optional stopping theorem to the martingale representing the transverse coordinate of the particle position (see for example chapter 12 in [21])

figure reveals that, when the variable free path model is used, the computational advantage associated with the ACA method is beneficial only for low Knudsen numbers ($\langle \mathrm{Kn} \rangle \leq 0.02$ in 10a and even lower for 10b). The reason for the breakdown of the efficiency for large Knudsen numbers lies in the fact that the contribution (62) to the estimate is a sum of terms that scale with $\langle \mathrm{Kn} \rangle$. While this feature contributes to variance reduction at low Knudsen numbers, it becomes a hindrance in the ballistic limit. We also note that even though the small values of the Knudsen number might create the impression that a diffusive approximation might be sufficient (i.e. the problem can be solved using Fourier's Law), due to the large variation in mean free paths this is not the case (significant discrepancies exist between the Boltzmann and Fourier solutions, and thus Boltzmann solutions are still necessary); this is further quantified in the following section, which also lays out an approach for recovering and in fact enhancing some of the computational benefits lost in the presence of widely variable free paths.

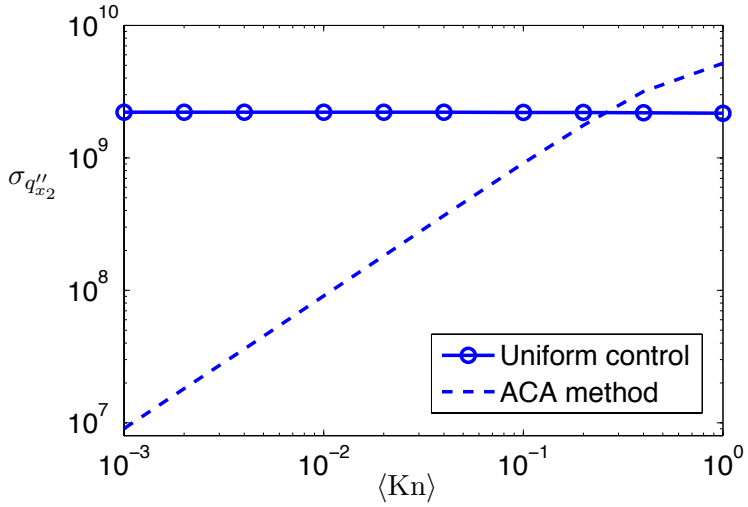

Figure 9: Standard deviation $\sigma_{q''_{x_2}}$ of the particle contributions to the estimate of the heat flux at point A of Figure 7 in the x_2 direction, in the single mean free path model. In the ACA method, the standard deviation is proportional to $\langle \mathrm{Kn} \rangle$. The latter outperforms the adjoint method with fixed control for $\langle \mathrm{Kn} \rangle \leq 0.2$.

6.3 Using a "hybrid" control for models with widely variable free paths

In section 4.2, we showed that the adjoint method is well suited to the case where mean free paths cover a wide range and when we seek to calculate the contribution of each individual mode to a given quantity of interest (typically, the heat flux or the thermal conductivity). On the other hand, as was shown in Figure 10, in the presence of a large variation in free paths, the benefits associated with the ACA method

become significant only for very low knudsen numbers. In this section we show that this limitation can be overcome with a slight modification of the ACA formulation; in fact, the modified formulation, referred to here as hybrid, improves the performance of the ACA formulation for all $\langle \mathrm{Kn} \rangle$ at almost no additional cost.

In order to motivate the hybrid version, we first explain why the ACA formulation fairs poorly as $\langle \mathrm{Kn} \rangle$ increases. When this method is applied to a phonon model with highly variable free paths, such as the one described in Appendix C, the terms that compose the sum (62) span several orders of magnitude. The presence of phonon modes with large free path (exceeding 100μm) tends to increase the variance of the estimate because the prefactor $V_g\tau$ becomes comparatively large. We can overcome this limitation by adapting the expression of the control. Since the problem is caused by the prefactor $V_g\tau$ which appears in (62), we propose a control which uses (58) for the small mean free path modes only, namely

$$e_{\mathrm{control}} = e_0 - \mathbb{1}_{\mathrm{Kn}_{\omega,p}<c} \mathbf{V}_g \tau \cdot \nabla_{\mathbf{x}} T_0 \frac{de^{\mathrm{eq}}_{T_{\mathrm{eq}}}}{dT}. \tag{64}$$

In words, according to this definition, the "hybrid" control uses e_0 for modes with large free paths ($\Lambda_{\omega,p} \geq cL$), while it introduces (58) for small free paths ($\Lambda_{\omega,p} < cL$). There is some degree of freedom in the choice of the constant c, although some trial-and-error revealed that, for the model that we used and the problem tested, $c \approx 0.4$ is close to optimal. Repeating the derivation procedure of the previous section, we can show that, algorithmically, the adjoint routine stays nearly the same, apart from the following two changes:

- The contribution of a particle j to the estimate is now:

$$\mathcal{I}^*_j = \dot{\mathcal{E}}^*_{\mathrm{eff}} \sum_{i=0}^{i=N_{\mathrm{seg}}-1} \mathcal{F}_i \tag{65}$$

  where

$$\mathcal{F}_i = \begin{cases} T_0(\mathbf{x}_{i+1}) - T_0(\mathbf{x}_i) & \text{if } \Lambda_{\omega,p} \geq cL \\ \tau_i \mathbf{V}_{g,i} \cdot (\nabla_{\mathbf{x}} T_0(\mathbf{x}_i) - \nabla_{\mathbf{x}} T_0(\mathbf{x}_{i+1})) & \text{if } \Lambda_{\omega,p} < cL \end{cases} \tag{66}$$

  Similarly to the ACA approach, the second case of equation (66) must account for the mismatch between the control and the boundary conditions by adding $\dot{\mathcal{E}}^*_{\mathrm{eff}} \tau_{N_{\mathrm{seg}-1}} \mathbf{V}_{g,N_{\mathrm{seg}}-1} \nabla_{\mathbf{x}} T_0(\mathbf{x}_{N_{\mathrm{seg}}})$ if the particle encounters the boundary with a mode obeying the criterion $\Lambda_{\omega,p} < cL$.

- The deterministic quantity associated with the final estimate needs to be calculated using the hybrid control.

We emphasize that implementing these changes only requires minor modifications since the core of the algorithm, i.e. the calculation of particle trajectory, remains the same. Only the values assigned to the estimates change. In fact, in the comparison

of the three approaches in Figures 9 and 10, all results were obtained using the same random numbers (all three methods were evaluated using the same particle trajectories).

Our results show that the hybrid method outperforms the two other approaches for all average Knudsen numbers. The amount of computational savings is however problem dependent. The results presented in Fig. 10a correspond to a favorable case where the heat flux is calculated at a surface point (point B in Figure 7). In this case, the length of a trajectory – and therefore computational time per particle – is proportional to $\langle \mathrm{Kn} \rangle^{-1}$ enabling us to accurately resolve the asymptotic behavior of the *standard deviation of the solution* all the way to $\langle \mathrm{Kn} \rangle = 0.001$ (in general, the standard deviation of a given quantity, as a higher moment of the distribution, is more expensive to resolve than the actual quantity).

At the crossing point of the constant-control and the ACA method, $\langle \mathrm{Kn} \rangle \approx 0.02$, the hybrid approach already reduces the standard deviation by a factor of 10, which corresponds to a speedup of around 100. Such a Knudsen number might appear small in the sense that kinetic effects might be expected to be negligible at such scales (10μm). This point of view would be incorrect since the free paths of low-frequency modes, known to significantly contribute to the heat flux [14], do not behave diffusively. Our calculations corroborate this claim; we find that, at this scale, the normal heat flux near a boundary still differs from the heat flux calculated using Fourier's Law by 30%. At $\langle \mathrm{Kn} \rangle \approx 0.002$, the speedup is close to a factor 2000 (standard deviation improvement of almost 45) and, although the system is quite close to the diffusive limit, we find a difference of almost 10% with respect to the Fourier solution. In addition, kinetic effects near boundaries and interfaces [17–19] can not be captured by the Fourier description at any $\langle \mathrm{Kn} \rangle$.

The results presented in Fig. 10b show a less favorable case in which the heat flux is calculated in the middle of the domain (point A in Figure 7). In this case, the length of particle trajectories is proportional to $\langle \mathrm{Kn} \rangle^{-2}$, making accurate resolution of the standard deviation in the solution more expensive. As a result, we are unable to study the asymptotic behavior of the standard deviation of the ACA method for $\langle \mathrm{Kn} \rangle \lesssim 0.01$. We also observe that the asymptotic behavior of the hybrid method for $\langle \mathrm{Kn} \rangle \to 0$ has not reached the expected $\sigma \propto \mathrm{Kn}$. We attribute this to the presence of a wide range of free paths which causes some phonons to behave ballistically even at these small Knudsen numbers (at this Knudsen number, the discrepancy between the Fourier solution and the simulation result is 13%) delaying the onset of the asymptotic behavior. A study showing the progressive delay of the onset of the asymptotic behavior as the range of mean free path increases can be found in [19]. Moreover, Fig. 10a shows, that the scaling $\sigma \propto \langle \mathrm{Kn} \rangle$ is still valid for the hybrid case at a surface point (for a discussion of the effect of the dependence of trajectory length on $\langle \mathrm{Kn} \rangle$–$\langle \mathrm{Kn} \rangle^{-1}$ vs $\langle \mathrm{Kn} \rangle^{-2}$–as well as the validity of the mathematical justification to the variable free path case, see Appendix B). We finally note that despite the fact that the hybrid approach has not reached the asymptotic regime at the smallest Knudsen number considered here, $\langle \mathrm{Kn} \rangle \approx 0.001$, the speedup provided by the hybrid method compared to uniform control is appreciable,

namely a factor of 25.

The hybrid approach, which takes advantage of the fact that the modes with small free paths behave diffusively, shares connections with the work in [22], in which these modes are assumed diffusive and treated by a Fourier-based description. The key difference is that in our method no Fourier model (approximation) is used; instead the proximity of these modes to the diffusive regime is used for switching between two modes of variance reduction (for numerically solving the same equation); diffusive behavior or particular modes of interaction between the long- and short-free-path modes is at no point assumed.

In this section, we only used the most readily available asymptotic solution of the Boltzmann equation, namely, the order 0 temperature field and its gradient. We expect that using higher order approximations, as derived in [17–19], would contribute even further to reducing the cost. Including such higher-order terms would be very complicated in a forward particle Monte Carlo. It would be close to straightforward in the adjoint framework since, as already demonstrated in this section, only the values assigned to the particle contributions would be modified, while the adjoint particle trajectories would remain unchanged.

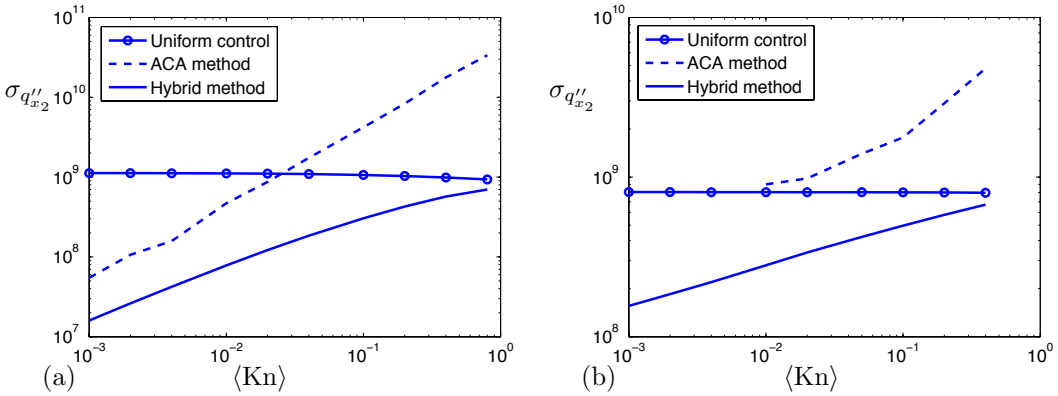

Figure 10: (a) Standard deviation $\sigma_{q''_{x_2}}$ of the particle contributions to the estimate of the heat flux at point B of Figure 7 in the x_2 direction, in the variable mean free path model. (b) Standard deviation $\sigma_{q''_{x_2}}$ of the particle contributions to the estimate of the heat flux at point A of Figure 7 in the x_2 direction, in the variable mean free path model.

7 Discussion

We developed an adjoint formulation for the linearized Boltzman transport equation for phonons in the relaxation-time approximation. We showed that, similarly to what is found in the fields of radiation, neutron transport, or computer graphics, the adjoint

approach is particularly suited to situations where the detector is small and the source is large. In the case of phonons, this is not only often true in a spatial sense, but also in a spectral sense. The free paths of phonons in semiconductors are known to cover a very broad range and, for this reason, the ability to discriminate individual phonon-mode contributions, as shown in Figure 5, is a very powerful feature of the adjoint framework. Although the precise speedup will depend on the relative size of the detector and source, in the examples considered here, speedups ranging from one to three orders of magnitude were observed. We note that, in consultation with the authors, Chengyun Hua and Austin Minnich [23] *applied* the proposed adjoint formulation to the investigation of boundary scattering in nanocrystalline materials. The method allowed them to show that low frequency phonons, in spite of the nanocrystalline structure, still carry a significant proportion of the heat and that, as a consequence, design of efficient thermoelectric materials should account for such effects.

An additional strength of the adjoint approach is its simplicity: the forward linearized approach relies on a cell-based approach, where quantities need to be sampled in computational cells of specific geometries. Sampling the contribution of a particle trajectory requires to study the overlap between the cell geometry and the trajectory geometry, which may be complicated. Unless the original source term is complicated itself, the adjoint alleviates this problem. We also note that the adjoint formulation proposed here is sufficiently general to be applicable to both timestep-based MC and KMC-type algorithms.

We showed that by using a control inspired by asymptotic solution of the Boltzmann equation, steady state problems of arbitrarily low Knudsen number can be treated at constant cost. This last feature results from an absolute statistical uncertainty that is proportional to $\langle \mathrm{Kn} \rangle$ (see Appendix B). The associated quadratic savings balance the quadratically increased cost caused by the calculation of longer trajectories in the low Knudsen number limit. As a result, simulations of structures or devices with lengthscales ranging from nanometers to hundreds of microns (see Figure 9) are not only possible, but also efficient. Extension to unsteady problems directly follows.

One weakness of the adjoint method is that each detector has to be replaced by an adjoint source. As a result, the more detectors, the more complex and thus less desirable the adjoint method becomes. Although exceptions sometimes occur (for instance, we saw in section 4.1.3 that multiple time detectors may be treated the same way as in the forward problem), in practice, the adjoint method is best suited to problems requiring high resolution (low statistical uncertainty) in *small* regions of phase space. One example is the recent use of adjoint formulation to validate the jump coefficients of the asymptotic theory developed and presented in [17–19]. These validations required a high level of accuracy for low Knudsen numbers, which was made possible by the method outlined here.

To our knowledge, this is the first time that the self-adjoint property of the phonon relaxation-time scattering operator is discussed and the adjoint formulation is used. In the field of neutron and gas transport, studies of the adjoint BTE have yielded results

whose application extended well beyond Monte Carlo simulations [24]. We hope that the present study will stimulate research efforts in this direction and lead to new insights for better understanding of phonon transport in general.

Future work will consider extension of the adjoint methodology to more realistic material models, ranging from models which include anisotropic dispersion relations [25] to the Boltzmann equation with (linearized) ab initio scattering [26]. The possibility of developing adjoint formulations for treating coupled electron-phonon transport will also be investigated.

A Proof of the fundamental relation (21) for the prescribed-temperature boundary conditions

A boundary with prescribed (deviational) temperature T_b is modeled as a black body. Any particle incident on the boundary is absorbed. At the same time, the boundary emits particles from the equilibrium (Bose-Einstein) distribution with temperature parameter T_b. The classical model consists of simply defining the boundary condition by specifying the incoming distribution at the wall for incoming particles:

$$\psi_\mathrm{b}(\omega, p, \mathbf{x}_\mathrm{b}, \mathbf{\Omega}\cdot\mathbf{n} > 0) = T_\mathrm{b}. \tag{67}$$

Here, following the general methodology developed in section 3.1, this boundary condition is expressed in terms of a combination of source terms. Emission of particles by the boundary can be represented by the source term:

$$q_\mathrm{b} = \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}T_\mathrm{b} \tag{68}$$

where H is the Heaviside function defined by

$$H(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases} \tag{69}$$

In addition to this source term that is independent of ψ and which replaces the thermalized region beyond the boundary, we need to use a source term that absorbs particles incident on the boundary. Such source term can be written [27] in the form:

$$\delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(-\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi \tag{70}$$

Since ψ appears explicitly in the above expression, we write the linearized BTE (12) in the form:

$$\frac{\partial\psi}{\partial t} + \mathbf{V}_g\cdot\nabla\psi = \frac{\mathcal{L}(\psi)-\psi}{\tau} + q + \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(-\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi \tag{71}$$

where q includes q_b and any other sources that do not depend on ψ. By analogy, the adjoint BTE is given by:

$$-\frac{\partial\psi^*}{\partial t} - \mathbf{V}_g\cdot\nabla\psi^* = \frac{\mathcal{L}(\psi^*)-\psi^*}{\tau} + h - \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi^* \tag{72}$$

We now repeat the integration by parts procedure of section 3.2 by writing

$$\mathcal{I}^* = \int q\Xi\psi^* d^3\mathbf{x}d^2\mathbf{\Omega}d\omega dt \tag{73}$$

$$= \int \left[\frac{\partial\psi}{\partial t} + \mathbf{V}_g\cdot\nabla\psi - \frac{\mathcal{L}(\psi)-\psi}{\tau} - \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(-\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi\right]\Xi\psi^* d^3\mathbf{x}d^2\mathbf{\Omega}d\omega dt \tag{74}$$

$$\mathcal{I}^* = \int_{\partial X}\int \mathbf{V}_g\cdot\mathbf{n}\psi\Xi\psi^* d^2\mathbf{x}d^2\mathbf{\Omega}d\omega dt + \int \psi\Xi\left[-\frac{\partial\psi^*}{\partial t} - \mathbf{V}_g\cdot\nabla\psi^* - \frac{\mathcal{L}(\psi^*)-\psi^*}{\tau}\right]d^3\mathbf{x}d^2\mathbf{\Omega}d\omega dt$$
$$- \int \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(-\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi\Xi\psi^* d^3\mathbf{x}d^2\mathbf{\Omega}d\omega dt \tag{75}$$

By noting that

$$\int_{\partial X}\int \mathbf{V}_g\cdot\mathbf{n}\psi\Xi\psi^* d^2\mathbf{x}d^2\mathbf{\Omega}d\omega dt = \int \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(-\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi\Xi\psi^* d^3\mathbf{x}d^2\mathbf{\Omega}d\omega dt$$
$$+ \int \delta(\mathbf{x}-\mathbf{x}_\mathrm{b})H(\mathbf{V}_g\cdot\mathbf{n})\mathbf{V}_g\cdot\mathbf{n}\psi\Xi\psi^* d^3\mathbf{x}d^2\mathbf{\Omega}d\omega dt$$

we obtain $\mathcal{I}^* = \mathcal{I}$.

B On the convergence rate of the ACA method: mathematical justification and discussion

In section 6.1, we find that using the spatially variable control

$$e_\mathrm{control} = \frac{de^\mathrm{eq}_{T_\mathrm{eq}}}{dT}\left(T_0 - \tau\mathbf{V}_g\cdot\nabla_\mathbf{x}T_0\right) \tag{76}$$

yields estimates whose standard deviations scale with the Knudsen number $\langle\mathrm{Kn}\rangle$, *provided that the temperature field T_0 is a solution to Laplace's equation*. In this section, we provide a mathematical explanation for this assertion. In the interest of simplicity, we consider here the case of a constant free path $\Lambda_{\omega,p} = \Lambda$. The case of variable free path can be treated by simple extension of this approach and is expected to yield similar results. This is further discussed below.

In the linearized algorithm, each particle, i, is associated with a contribution y_i which is a realization of a random variable Y such that, ultimately, the quantity estimated is the average $\mathcal{I}^*_N = \sum y_i/N$ which converges to $E(Y)$ in the limit $N\to\infty$. The standard deviation of Y and $\mathcal{I}^*_N$, respectively σ_Y and $\sigma_{\mathcal{I}^*_N}$, are related by $\sigma_{\mathcal{I}^*_N} = \sigma_Y/\sqrt{N}$. Showing that the standard deviation of $\sigma_{\mathcal{I}^*_N}$ scales with $\langle\mathrm{Kn}\rangle$ amounts to showing that σ_Y scales in the same manner (with $\langle\mathrm{Kn}\rangle$). The random variable Y is a sum of random variables, $Y = Z_1 + Z_2 + ... + Z_{N_\mathrm{seg}}$, where each variable Z_j corresponds to the contribution of a

single segment of trajectory, as shown in section 6.1. Let us first recall that Z_j is given by

$$Z_j = N\dot{\mathcal{E}}^*_{\text{eff}}\tau_{j-1}\mathbf{V}_{g,j-1}\cdot(\nabla_{\mathbf{x}}T_0(\mathbf{x}_{j-1}) - \nabla_{\mathbf{x}}T_0(\mathbf{x}_j)) \tag{77}$$

except for $j = N_{\text{seg}}$. Note that $\mathcal{E}^*_{\text{tot}} \equiv N\dot{\mathcal{E}}^*_{\text{eff}}$ is independent of $\langle\text{Kn}\rangle$. When $\langle\text{Kn}\rangle$ is small, Z_j may be written as

$$Z_j = -\mathcal{E}^*_{\text{tot}}\Lambda l_j \left.\frac{\partial^2 T_0}{\partial x_m \partial x_n}\right|_{\mathbf{x}_j} \Omega_m\Omega_n - \mathcal{E}^*_{\text{tot}}\frac{\Lambda l_j^2}{2}\left.\frac{\partial^3 T_0}{\partial x_m\partial x_n\partial x_q}\right|_{\mathbf{x}_j}\Omega_m\Omega_n\Omega_q + \text{h.o.t.} \tag{78}$$

where $l_j(\approx\Lambda)$ is the length of that segment of the trajectory and h.o.t. denotes higher order terms. This can be rearranged in the form

$$Z_j = -\mathcal{E}^*_{\text{tot}}\langle\text{Kn}\rangle^2\frac{l_j}{\Lambda}\left.\frac{\partial^2 T_0}{\partial x'_m\partial x'_n}\right|_{\mathbf{x}_j}\Omega_m\Omega_n - \mathcal{E}^*_{\text{tot}}\langle\text{Kn}\rangle^3\frac{l_j^2}{2\Lambda^2}\left.\frac{\partial^3 T_0}{\partial x'_m\partial x'_n\partial x'_q}\right|_{\mathbf{x}_j}\Omega_m\Omega_n\Omega_q + \text{h.o.t.} \tag{79}$$

$$\equiv\tilde{Z}_j + \mathcal{O}\left(\frac{l_j^2}{\Lambda^2}\langle\text{Kn}\rangle^3\right) \tag{80}$$

where $\mathbf{x}'_j$ is the dimensionless coordinate defined by $\mathbf{x}'_j = \mathbf{x}_j/L$. First we note that due to the isotropy of the post-scattering distribution, $E(\Omega_i) = 0$. Moreover, here we are examining the case $\nabla^2_{\mathbf{x}'}T_0 = 0$. From these two observations it directly follows that $E(\tilde{Z}_j) = 0$ and therefore

$$\tilde{Y}_n = \sum_{i=1}^{n}\tilde{Z}_i \tag{81}$$

defines a martingale [21] with (optional) stopping time $n = N_{\text{seg}} - 1$. In what follows, we denote $\tilde{Y}_{N_{\text{seg}}-1}$ by $\tilde{Y}$.

In summary $Y = \tilde{Y} + Z_{N_{\text{seg}}} + \zeta$, where ζ represents the contribution of $N_{\text{seg}} - 1$ order 3 terms and

$$Z_{N_{\text{seg}}} = \dot{\mathcal{E}}^*_{\text{tot}}\langle\text{Kn}\rangle\left(\Omega_j\left.\frac{\partial T_0}{\partial x'_j}\right|_{\mathbf{x}_{N_{\text{seg}}-1}}\right). \tag{82}$$

The variance of Y is therefore:

$$\text{Var}(Y) = \text{Var}(\tilde{Y}) + \text{Var}(\zeta) + \text{Var}(Z_{N_{\text{seg}}}) + 2\text{Cov}(\tilde{Y}, Z_{N_{\text{seg}}}) + 2\text{Cov}(\zeta, Z_{N_{\text{seg}}}) + 2\text{Cov}(\tilde{Y},\zeta) \tag{83}$$

Below, we examine each of the terms of the above expression and show that they all scale with $\langle\text{Kn}\rangle^2$.

- Variance of $\tilde{Y}$: By applying the optional stopping theorem to the martingale $S_n = \tilde{Y}_n^2 - \sum_{i=1}^{n}\text{Var}(\tilde{Z}_i)$ (see [21]), which implies that $E(S_n) = 0$, we obtain

$$\text{Var}(\tilde{Y}) = E\left(\sum_{i=1}^{N_{\text{seg}}-1}\text{Var}(\tilde{Z}_i)\right) \tag{84}$$

We note that, provided the second derivatives of T_0 are bounded, we can find a positive constant M_1 such that the variances of $\tilde{Z}_i$ are all smaller than $M_1\langle \mathrm{Kn}\rangle^4$. It follows that:

$$\mathrm{Var}(\tilde{Y}) \leq E\left(\sum_{i=1}^{N_{\mathrm{seg}}-1} M_1\langle \mathrm{Kn}\rangle^4\right), \tag{85}$$

and therefore

$$\mathrm{Var}(\tilde{Y}) \leq M_1\langle \mathrm{Kn}\rangle^4 E(N_{\mathrm{seg}}) \tag{86}$$

Finally, since the average number of jumps is asymptotically proportional to $\langle \mathrm{Kn}\rangle^{-2}$:

$$\mathrm{Var}(\tilde{Y}) = \mathcal{O}\left(\langle \mathrm{Kn}\rangle^2\right) \tag{87}$$

- Variance of ζ:

  The variance of ζ is defined as $E(\zeta^2) - E(\zeta)^2$, where:

$$E(\zeta^2) = E\left(\left[\sum_{i=1}^{N_{\mathrm{seg}}-1} \mathcal{O}(\frac{l_i^2}{\Lambda^2}\langle \mathrm{Kn}\rangle^3)\right]^2\right) \tag{88}$$

$$E(\zeta^2) = E\left(\sum_{i=1}^{N_{\mathrm{seg}}-1}\sum_{j=1}^{N_{\mathrm{seg}}-1} \mathcal{O}\left(\frac{l_i^2 l_j^2}{\Lambda^4}\langle \mathrm{Kn}\rangle^6\right)\right). \tag{89}$$

  Wald's equation [21] applies to the latter expression and yields

$$E(\zeta^2) = E\left(\sum_{i=1}^{N_{\mathrm{seg}}-1}\sum_{j=1}^{N_{\mathrm{seg}}-1} E\left[\mathcal{O}\left(\frac{l_i^2 l_j^2}{\Lambda^4}\langle \mathrm{Kn}\rangle^6\right)\right]\right) \tag{90}$$

  We can find a positive constant M_2 such that:

$$E(\zeta^2) \leq M_2 E\left((N_{\mathrm{seg}}-1)^2\right)\langle \mathrm{Kn}\rangle^6 \tag{91}$$

  In other words:

$$E(\zeta^2) = \mathcal{O}\left(\langle \mathrm{Kn}\rangle^2\right) \tag{92}$$

  Also:

$$E(\zeta) = E\left(\sum_{i=1}^{N_{\mathrm{seg}}-1} \mathcal{O}\left(\frac{l_j^2}{\Lambda^2}\langle \mathrm{Kn}\rangle^3\right)\right) \tag{93}$$

$$E(\zeta) \leq M_3 E\left(N_{\mathrm{seg}}-1\right)\langle \mathrm{Kn}\rangle^3 \tag{94}$$

$$E(\zeta) = \mathcal{O}\left(\langle \mathrm{Kn}\rangle\right) \tag{95}$$

  We finally find that $\mathrm{Var}(\zeta) = \mathcal{O}(\langle \mathrm{Kn}\rangle^2)$.

- Variance of $Z_{N_{\text{seg}}}$:

  From the definition of $Z_{N_{\text{seg}}}$, we immediately find that $\text{Var}\left(Z_{N_{\text{seg}}}\right) = \mathcal{O}(\langle \text{Kn} \rangle^2)$.

- Covariance of $\tilde{Y}$ and $Z_{N_{\text{seg}}}$:

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = E\left(\tilde{Y} Z_{N_{\text{seg}}}\right) - E(\tilde{Y})E(Z_{N_{\text{seg}}}) \tag{96}$$

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = E\left(\tilde{Y} Z_{N_{\text{seg}}}\right) \tag{97}$$

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = \int_{\tilde{y}} E\left(\tilde{Y} Z_{N_{\text{seg}}} | \tilde{Y} = \tilde{y}\right) P(\tilde{Y} = \tilde{y}) d\tilde{y} \tag{98}$$

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = \int_{\tilde{y}} \tilde{y} E\left(Z_{N_{\text{seg}}} | \tilde{Y} = \tilde{y}\right) P(\tilde{Y} = \tilde{y}) d\tilde{y} \tag{99}$$

  The martingale central limit theorem for $\tilde{Y}$ states that $P(\tilde{Y} = \tilde{y})$ tends asymptotically to a Gaussian with standard deviation $\sigma = \sqrt{\text{Var}(\tilde{Y})}$. Also, due to isotropy associated with the scattering process $E\left(Z_{N_{\text{seg}}} | \tilde{Y} = \tilde{y}\right) \sim E\left(Z_{N_{\text{seg}}}\right) = \mathcal{O}(\langle \text{Kn} \rangle)$. Hence:

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = \mathcal{O}\left(\langle \text{Kn} \rangle \int_{\tilde{y}} |\tilde{y}| \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\tilde{y}^2}{2\sigma^2}\right) d\tilde{y}\right) \tag{100}$$

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = \mathcal{O}\left(\langle \text{Kn} \rangle \sigma\right) \tag{101}$$

$$\text{Cov}\left(\tilde{Y}, Z_{N_{\text{seg}}}\right) = \mathcal{O}\left(\langle \text{Kn} \rangle^2\right) \tag{102}$$

- Covariance of $\tilde{Y}$ and ζ:

  We note that the value of ζ is obtained using the same random numbers as $\tilde{Y}$. We may still obtain an upper bound for the covariance using:

$$\text{Cov}\left(\tilde{Y}, \zeta\right) = E\left(\tilde{Y} \zeta\right) - E(\tilde{Y})E(\zeta) \tag{103}$$

$$\text{Cov}\left(\tilde{Y}, \zeta\right) = E\left(\tilde{Y} \zeta\right) \tag{104}$$

$$\text{Cov}\left(\tilde{Y}, \zeta\right) = \int_{\tilde{y}} \tilde{y} E\left(\zeta | \tilde{Y} = \tilde{y}\right) P(\tilde{Y} = \tilde{y}) d\tilde{y} \tag{105}$$

$$\text{Cov}\left(\tilde{Y}, \zeta\right) = \mathcal{O}\left(\langle \text{Kn} \rangle \int_{\tilde{y}} |\tilde{y}| \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\tilde{y}^2}{2\sigma^2}\right) d\tilde{y}\right) \tag{106}$$

$$\text{Cov}\left(\tilde{Y}, \zeta\right) = \mathcal{O}\left(\langle \text{Kn} \rangle^2\right) \tag{107}$$

  where, since $E(\zeta) \sim \sqrt{Var(\zeta)} \sim \mathcal{O}\left(\langle \text{Kn} \rangle\right)$, we estimate $E\left(\zeta | \tilde{Y} = \tilde{y}\right)$ as of $\mathcal{O}\left(\langle \text{Kn} \rangle\right)$.

- Covariance of ζ and $Z_{N_{\text{seg}}}$:

$$\text{Cov}\left(Z_{N_{\text{seg}}}, \zeta\right) = E\left(Z_{N_{\text{seg}}}\zeta\right) - E(Z_{N_{\text{seg}}})E(\zeta) \tag{108}$$

$$\text{Cov}\left(Z_{N_{\text{seg}}}, \zeta\right) = E\left(\mathcal{O}\left[\sum_{i=1}^{N_{\text{seg}}-1} \frac{l_j^2 l_{N_{\text{seg}}}}{\Lambda^3}\langle\text{Kn}\rangle^4\right]\right) + \mathcal{O}(\langle\text{Kn}\rangle^2) \tag{109}$$

$$\text{Cov}\left(Z_{N_{\text{seg}}}, \zeta\right) = \mathcal{O}\left(\langle\text{Kn}\rangle^2\right) \tag{110}$$

  which is again obtained using Wald's equation.

In summary, this shows that

$$\text{Var}(Y) = O(\langle\text{Kn}\rangle^2), \tag{111}$$

which implies that the standard deviation associated with the estimate $\mathcal{I}_N^*$, $\sigma_{\mathcal{I}_N^*}$, scales linearly with $\langle\text{Kn}\rangle$.

We note here the following:

1 Even in cases where $E[N_{\text{seg}}] = O(\langle\text{Kn}\rangle^{-1})$ (and thus $\text{Var}(\tilde{Y}) = O(\langle\text{Kn}\rangle^3)$), such as in the proximity of a boundary, the leading order term in $\text{Var}(Y)$ is still of $O(\langle\text{Kn}\rangle^2)$, yielding the same result.

2 In the above development, $\nabla^2_{\mathbf{x}'} T_0 = 0$ comes as a *necessary* condition for the first-order scaling, ($\sigma_{\mathcal{I}_N^*} \propto \text{Kn}$), to be true. To see this, let us imagine that, in a given region of space, this condition is not satisfied. Then, in such a region, $\tilde{Y}$ is no longer a martingale because the expected value of $\tilde{Z}_i$ is no longer 0. As a consequence, the result (84) can no longer be used. More precisely, it follows from (79) that $\text{Cov}(\tilde{Z}_i, \tilde{Z}_j) = \mathcal{O}(\langle\text{Kn}\rangle^4(\nabla^2_{\mathbf{x}'} T_0)^2)$; as a result if $\nabla^2_{\mathbf{x}'} T_0 \neq 0$, correlations between $\tilde{Z}_i$ and $\tilde{Z}_j$ will cause $\text{Var}(\sum \tilde{Z}_i) = \mathcal{O}(N_{\text{seg}}^2\langle\text{Kn}\rangle^4) \neq \mathcal{O}(\langle\text{Kn}\rangle^2)$ and under most conditions $\text{Var}(Y) = \mathcal{O}(\langle\text{Kn}\rangle^0)$ (see below for an example).

To illustrate the second point, let us consider the example discussed in section 6.2, namely, using the adjoint method to find the x_2-component of the heat flux at point $(0, 0)$. Figures 11 and 12 show $\sum_{i=1}^n Z_i$ as a function of the index n, where Z here corresponds to the heat flux contribution. Figure 11 shows the result obtained using

$$T_0(x_1, x_2) = \epsilon T_{\text{eq}} \cos\left(\frac{2\pi x_1}{3L}\right) \frac{\sinh\left(\frac{2\pi x_2}{3L}\right)}{\sinh\left(\frac{2\pi}{3L}\right)} \tag{112}$$

as a control which is a solution of Laplace's equation with Dirichlet boundary conditions; Figure 12 shows the result obtained using

$$T_0(x_1, x_2) = \epsilon T_{\text{eq}} \cos\left(\frac{2\pi x_1}{3L}\right)\left(1 - \frac{x_2}{L}\right). \tag{113}$$

as a control, which is not a solution of Laplace's equation. Figure 11 shows that when T_0 satisfies the Laplace equation, very little correlation between each segment of a trajectory exists and, as a result, the standard deviation of the contributions is proportional to $\langle \mathrm{Kn} \rangle$, in agreement with (111). In contrast, Figure 12 shows that when T_0 does not satisfy the Laplace equation, correlation between different segments of the trajectory makes the contribution of each particle independent of the Knudsen number (proportional to $\langle \mathrm{Kn} \rangle^0$ as discussed in item 2 above), leading to a standard deviation that also scales as $\langle \mathrm{Kn} \rangle^0$.

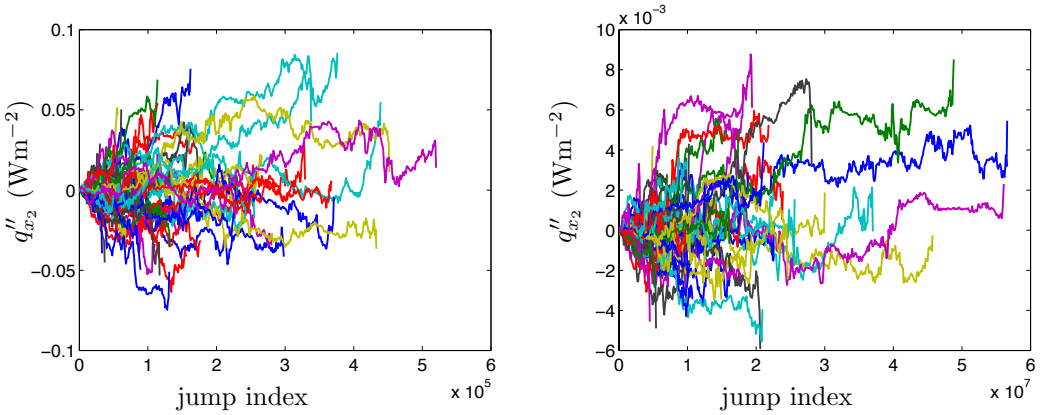

Figure 11: Evolution of the contributions of particles to the final estimate, when the adjoint method is used with the control (76) along with a temperature field which is the solution of Laplace's equation for $\langle \mathrm{Kn} \rangle = 0.01$ (Left) and $\langle \mathrm{Kn} \rangle = 0.001$ (Right). This calculation was performed using the single-mean-free-path model.

C Material models

Dispersion relations (and the resulting group velocities) and scattering rates for the simulations in this paper were calculated using the material model described in detail in [9]; the description is reproduced here for convenience. We note that since the primary focus of the present paper is the development of a numerical approach, this material model choice represents a balance between simplicity and fidelity. In other words, although this material model is sufficiently realistic to capture a number of important features that have a large influence on the computational method (e.g. wide range of free paths), it includes a number of simplifying assumptions (e.g. isotropic dispersion relation [25]) that may need to be re-examined when used to model transport in nanostructures. Extension to more realistic material models of varying complexity, including the ab-initio scattering operator, will be considered in future work.

In the present model, dispersion relations are adapted from the experimentally measured dispersion relation in the [100] direction ([28] for Al, [29] for Si). Note that extension to more realistic dispersion relations such as the one presented in [30] is

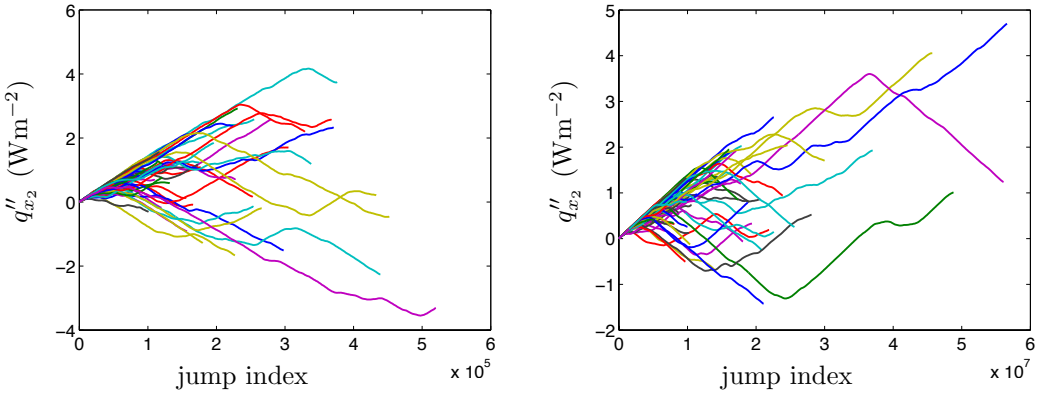

Figure 12: Evolution of the contributions of particles to the final estimate, when the adjoint method is used with the control (76) along with a temperature field which is *not* a solution of Laplace's equation, for $\langle \mathrm{Kn} \rangle = 0.01$ (Left) and $\langle \mathrm{Kn} \rangle = 0.001$ (Right). This calculation was performed using the single-mean-free-path model.

straightforward, as long as the post-scattering traveling directions are assumed isotropic, as assumed in [30]. From the dispersion relation, the density of states may be derived using

$$D(\omega, \mathrm{LA}) = \frac{1}{2} \frac{\omega^2}{\pi^2 c(\omega, \mathrm{LA})^2 V_g(\omega, \mathrm{LA})} \tag{114}$$

$$D(\omega, \mathrm{TA}) = \frac{\omega^2}{\pi^2 c(\omega, \mathrm{TA})^2 V_g(\omega, \mathrm{TA})} \tag{115}$$

where $c(\omega, p)$ refers to the phase velocity (given by ω/k, k being the wavenumber). Note the absence of the factor $1/2$ for the TA modes due to the presence of two such modes which, in this model, share the same properties.

For aluminum in the TTR calculation, a constant relaxation time is used; it is chosen to fit the desired lattice thermal conductivity (as in [29]) and is given by:

$$\tau_{\mathrm{Al}} = 10^{-11} \mathrm{s} \tag{116}$$

For silicon, the expressions are taken from [31], with constants from [29]. For acoustic modes, these are

phonon-phonon scattering, LA	$\tau_L^{-1} = A_L \omega^2 T^{1.49} \exp\left(-\frac{\theta}{T}\right)$
phonon-phonon scattering, TA	$\tau_T^{-1} = A_T \omega^2 T^{1.65} \exp\left(-\frac{\theta}{T}\right)$
impurity scattering	$\tau_I^{-1} = A_I \omega^4$
boundary scattering	$\tau_B^{-1} = w_b$

where the constants take the following values

Parameter	A_L	A_T	θ	A_I	w_b
Value (in SI units)	$2.09 \times 10^{-19}/(2\pi)^2$	$1.23 \times 10^{-19}/(2\pi)^2$	80	3×10^{-45}	1.2×10^6

For a given polarization we obtain the total relaxation time by applying the Matthiessen rule

$$\tau^{-1} = \sum_i \tau_i^{-1} \tag{117}$$

For simplicity, we use Einstein's model to treat optical phonons and consider them as immobile. Their behavior is therefore purely capacitive. Einstein's model states that the contribution of optical phonons to the vibrational energy per unit volume in a crystal is given by [3]

$$U = \frac{N_p N' \hbar \omega_E}{V[\exp(\hbar\omega_E/k_b T) - 1]} \tag{118}$$

where $N_p = 3$ is the number of polarizations, $N' = 1$ is the number of optical states per lattice point, ω_E is the Einstein radial frequency ($\omega_E = 9.1 \times 10^{13} s^{-1}$ [29]), V is the volume of a lattice point (with a lattice constant $a = 5.43\mathring{A}$, $V = a^3/4 = 4 \times 10^{-29} m^3$).

For the relaxation time of optical phonons at 300K, we use the value [32]

$$\tau_O = 3 \times 10^{-12}\text{s} \tag{119}$$

D Acknowledgements

The authors wish to thank Colin Landon, Austin Minnich and Konstantin Turitsyn for useful discussions and comments. Research on the hybrid control and the convergence rate of the ACA method was supported as part of the Solid-State Solar-Thermal Energy Conversion Center (S3TEC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award# DE-SC0001299 / DE-FG02-09ER46577.

References

[1] 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. *Journal of Applied Physics*, 93:793–818, 2003.

[2] David G. Cahill, Paul V. Braun, Gang Chen, David R. Clarke, Shanhui Fan, Kenneth E. Goodson, Pawel Keblinski, William P. King, Gerald D. Mahan, Arun Majumdar, Humphrey J. Maris, Simon R. Phillpot, Eric Pop, and Li Shi. Nanoscale thermal transport. ii. 20032012. *Applied Physics Reviews*, 1(1):–, 2014.

[3] G. Chen. *Nanoscale Energy Transport and Conversion*. Oxford University Press, New York, NY, 2005.

[4] J.-P. M. Péraud, C. D. Landon, and N. G. Hadjiconstantinou. Monte Carlo methods for solving the Boltzmann transport equation. In *Annual Review of Heat Transfer*, volume 17, pages 205–265. Begell House, 2014.

[5] J.-P. M. Péraud and N. G. Hadjiconstantinou. An alternative approach to efficient simulation of micro/nanoscale phonon transport. *Applied Physics Letters*, 101:153114, 2012.

[6] J. Spanier and E. M. Gelbard. *Monte Carlo Principles and Neutron Transport Problems*. Addison-Wesley, Reading, MA, 1969.

[7] K. M. Case and P. F. Zweifel. *Linear Transport Theory*. Addison-Wesley, Reading, MA, 1967.

[8] Q. Hao, G. Chen, and M. S. Jeng. Frequency-dependent Monte Carlo simulation of phonon transport in two-dimensional porous silicon with aligned pores. *Journal of Applied Physics*, 106:114321, 2009.

[9] J.-P. M. Péraud and N. G. Hadjiconstantinou. Efficient simulation of multidimensional phonon transport using energy-based variance-reduced Monte Carlo formulations. *Physical Review B*, 84:205331, 2011.

[10] J.-P. M. Péraud and N. G. Hadjiconstantinou. Deviational phonons and thermal transport at the nanoscale. *The 2012 International Mechanical Engineering Congress and Exposition*, (paper number IMECE2012-87547), 2012.

[11] A. J. Minnich. Determining phonon mean free paths from observation of quasiballistic thermal transport. *Physical Review Letters*, 109:205901, 2012.

[12] A. Schmidt. Pump probe thermoreflectance. In *Annual Review of Heat Transfer*, volume 16, pages 159–181. Begell House, 2013.

[13] A. J. Schmidt, X. Chen, and G. Chen. Pulse accumulation, radial heat conduction, and anisotropic thermal conductivity in pump-probe transient thermoreflectance. *Review of Scientific Instruments*, 79:114902, 2008.

[14] F. Yang and C. Dames. Mean free path spectra as a tool to understand thermal conductivity in bulk and nanostructures. *Physical Review B*, 87:035437, 2013.

[15] T. M. M. Homolle and N. G. Hadjiconstantinou. A low-variance deviational simulation Monte Carlo for the Boltzmann equation. *Journal of Computational Physics*, 226:2341–2358, 2007.

[16] G. A. Radtke and N. G. Hadjiconstantinou. Variance-reduced particle simulation of the Boltzmann transport equation in the relaxation-time approximation. *Physical Review E*, 79:056711, 2009.

[17] J.-P. M. Péraud and N. G. Hadjiconstantinou. in preparation.

[18] J.-P. M. Péraud and N. G. Hadjiconstantinou. On the equations and boundary conditions governing phonon-mediated heat transfer in the small mean free path limit. An asymptotic solution of the Boltzmann equation. *The 2014 International Mechanical Engineering Congress and Exposition*, (paper number IMECE2014-36475), in press.

[19] J.-P. M. Péraud. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2015.

[20] G. A. Radtke, J.-P. M. Péraud, and N. G. Hadjiconstantinou. On efficient simulations of multiscale kinetic transport. *Philosophical Transactions of the Royal Society A*, 371:20120182, 2013.

[21] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes, 3rd edition*. Oxford University Press, 2001.

[22] J. M. Loy, J. Y. Murthy, and D. Singh. A fast hybrid Fourier-Boltzmann transport equation solver for nongray phonon transport. *Journal of Heat Transfer*, 135:011008, 2013.

[23] C. Hua and A. J. Minnich. Importance of frequency-dependent grain boundary scattering in nanocrystalline silicon and silicon-germanium thermoelectrics. *Semiconductor Science Technology*, in press.

[24] S. Takata. Symmetry of the unsteady linearized Boltzmann equation in a fixed bounded domain. *Journal of Statistical Physics*, 140(5):985–1005, 2010.

[25] Z. Chen and Z. Wei and Y. Chen and C. Dames. Anisotropic model for the thermal boundary conductance. *Physical Review B*, 87, 2013.

[26] C.D. Landon and N. G. Hadjiconstantinou. Deviational simulation of phonon transport in graphene ribbons with ab initio scattering. *Journal of Applied Physics*, 16:163502, 2014.

[27] R. Sanchez and N. J. McCormick. Review of neutron transport approximations. *Nuclear Science and Engineering*, 80(4):481–535, 1982.

[28] R. Stedman and G. Nilsson. Dispersion relations for phonons in Aluminum at 80 and 300K. *Physical Review*, 145(2):492–500, MAY 1966.

[29] A. J. Minnich. *Exploring electron and phonon transport at the nanoscale for thermoelectric energy conversion*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2011.

[30] K. Kukita and Y. Kamakura. Monte Carlo simulation of phonon transport in silicon using a realistic dispersion relation. *Journal of Applied Physics*, 114:154312, 2013.

[31] A. S. Henry and G. Chen. Spectral phonon transport properties of silicon based on molecular dynamics simulations and lattice dynamics. *J. Computational and Theoretical Nanoscience*, 5(1–12), 2008.

[32] P. G. Klemens. *Physical Review*, 148(2):845–848, Aug 1966.