On efficient simulations of multiscale kinetic transport

Gregg A. Radtke, Jean-Philippe M. Péraud and Nicolas G. Hadjiconstantinou

Supplementary data	"Presentation at Turbulent Mixing and Beyond conference" http://rsta.royalsocietypublishing.org/content/suppl/2012/11/23/rsta.2012.0182.DC1.html
References	**This article cites 55 articles, 1 of which can be accessed free** http://rsta.royalsocietypublishing.org/content/371/1982/201201 82.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections fluid mechanics (123 articles) mechanics (9 articles) statistical physics (47 articles) thermodynamics (10 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Review

CrossMark
click for updates

One contribution of 13 to a Theme Issue 'Turbulent mixing and beyond: non-equilibrium processes from atomistic to astrophysical scales I.'

Author for correspondence:
Nicolas G. Hadjiconstantinou
e-mail: ngh@mit.edu

On efficient simulations of multiscale kinetic transport

Gregg A. Radtke, Jean-Philippe M. Péraud and Nicolas G. Hadjiconstantinou

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

We discuss a new class of approaches for simulating multiscale kinetic problems, with particular emphasis on applications related to small-scale transport. These approaches are based on a decomposition of the kinetic description into an equilibrium part, which is described deterministically (analytically or numerically), and the remainder, which is described using a particle simulation method. We show that it is possible to derive evolution equations for the two parts from the governing kinetic equation, leading to a decomposition that is dynamically and automatically adaptive, and a multiscale method that seamlessly bridges the two descriptions *without introducing any approximation*. Our discussion pays particular attention to stochastic particle simulation methods that are typically used to simulate kinetic phenomena; in this context, these decomposition approaches can be thought of as control-variate variance-reduction formulations, with the nearby equilibrium serving as the control. Such formulations can provide substantial computational benefits in a broad spectrum of applications because a number of transport processes and phenomena of practical interest correspond to perturbations from nearby equilibrium distributions. In many cases, the computational cost reduction is sufficiently large to enable otherwise intractable simulations.

1. Introduction

Computational methods that can efficiently but also accurately capture a wide range of length and time scales remain highly desirable and are the subject of considerable research effort. In the area of kinetic transport, multiscale problems are common in micro- and

nanoscale gas dynamics [1], plasma physics [2], micro- and nanoscale solid-state heat transfer as mediated by phonon transport [3], electron transport [4,5] and neutron transport [6,7]. The primary challenge associated with such problems stems from the computational cost associated with capturing dynamics occurring over a wide range of scales.

One typical example of such problems is thermal response to laser irradiation: the impulsive nature of laser heating means that at early times (compared with the mean time between collisions associated with the carrier) transport is ballistic [8,9]; only at late times the Fourier description becomes appropriate. During the ballistic and transition [1] regimes, a kinetic description is necessary; on the other hand, a kinetic description of this problem for macroscopic times is typically intractable. Multiscale methods aim to provide means for solving such problems without introducing approximations; instead, they seek to combine a number of different descriptions into a formulation that is globally consistent and more computationally efficient than simulation of the problem at the highest level of fidelity (which is often intractable).

Multiscale problems abound in science and engineering. Reviewing methods and applications in this general context is a truly daunting task. For this reason, our discussion in this paper focuses on methods applicable to the primary topic of this review, namely kinetic transport. Moreover, in order to simplify the discussion, technical details will be presented in the context of dilute gases which have been the subject of a considerable part of recent method development efforts. The more general validity of the ideas presented and discussed here is demonstrated in §5b(ii), where their application to nanoscale solid-state heat transfer as mediated by phonons is discussed.

(a) Background

Phenomena involving kinetic transport but in which a limiting description (e.g. a set of continuum equations with Navier–Stokes closure[1]) may be used in part of the computational domain can be treated using hybrid atomistic-continuum methods [10–13]. These methods minimize the computational cost by using the kinetic description only in regions where it is needed, while the remainder of the computational domain is treated by the less expensive continuum description. In addition to a clear spatial separation of kinetic and continuum regions, these methods assume that the region where kinetic effects are important is sufficiently small that the kinetic part of the calculation is feasible. These formulations typically employ domain decomposition techniques [14] to separate the two regions in question and thus coupling of the two descriptions becomes central to the success of the method [15]. In this context, a region where the two descriptions are simultaneously valid—for matching to take place—is required; criteria for transition from one description to the other [12,16] are also necessary. Fortunately, continuum descriptions of kinetic systems are typically well characterized—e.g. in a dilute gas, the molecular distribution function is given by the Chapman–Enskog result in the Navier–Stokes (diffuse transport) limit [17–19]—and thus *rigorous* matching of the two descriptions is possible [12]. This has led to the development of fairly sophisticated hybrid simulation methods [12,13,20] in which the transition from one description to the other is automatic and, in some cases, seamlessly integrated within a mesh refinement process. The situation is not entirely resolved in the case of dense fluids [14,15,21] and solids [22] where the molecular distribution function corresponding to the continuum limit is not known, and thus constructing a microscopic (e.g. molecular) state that exactly corresponds to the continuum solution in the matching region is not *rigorously* possible yet [15].

---

[1]In the literature, it is customary to refer to the Navier–Stokes system of equations as the continuum description, despite the fact that continuum conservation laws, sometimes with other closures, can still be written when the Navier–Stokes closure is not valid. Owing to the long historical precedent, we will also use the term continuum to refer to the Navier–Stokes (more generally, diffusive transport) description valid in the macroscopic limit.

Projective integration [23] also relies on a clear separation of scales. In this method, a microscopic (molecular) description is simulated for a small amount of time, and based on this simulation, the coarse dynamical behaviour can be extracted (e.g. through an eigenvalue analysis) and projected forward in time using timesteps appropriate for a numerical solver of the coarse, macroscopic, description. In direct analogy, one can perform homogeneous simulations of a microscopic description in small, *suitably terminated*, regions of space (e.g. located at the nodes of a macroscopic numerical solver) thus obtaining local closure (constitutive) information which informs a numerical solution of the macroscopic set of conservation equations [23–25]. The above approaches are strictly valid (and worthwhile) when a clear separation between the microscopic and macroscopic time scales and length scales, respectively, exists; under these conditions, treatment of the inner model in a local and/or quasistatic fashion is justified. Examples include formulations that require steady solutions from the microscopic solver (e.g. for steady problems), where the error owing to imprecise initialization of the microscopic solver may be avoided. Another example is the *special case* of kinetic problems in which only initialization of microscopic states corresponding to continuum fields is necessary; in such cases, the existence of Chapman–Enskog-type results [3,17–19] which provide analytical descriptions for the distribution function makes this rigorously possible, as in the case of hybrid Navier–Stokes/kinetic domain-decomposition-based methods discussed above, where coupling happens in a region where both descriptions are valid (see [11,12] for an example). Unfortunately, in the general (and most practically relevant) case where the distribution function is not known analytically, ad hoc initialization and domain termination introduces approximation error [26,27] which is expected to corrupt the closure information subsequently extracted from the microscopic solver.

For a large class of problems of practical interest, the separation of scales is not sufficiently large to justify the approaches described above. In such cases, methods which go beyond spatial decomposition or assumptions of local and/or quasistatic behaviour are needed. The approach reviewed in this paper instead relies on an *algebraic decomposition* of the distribution function. This approach, introduced in §3, is viable because, for kinetic problems, the governing equation is typically sufficiently tractable to allow rigorous descriptions of the evolution of the constituent parts into which the distribution function is decomposed.

The resulting methods are powerful and exhibit a number of desirable features. Let us consider the approach to the Navier–Stokes limit as an example, where it is well known [17] that the distribution can be decomposed into a (local) equilibrium distribution and a (small) correction which describes deviations therefrom. In contrast to standard particle simulation methods, which become increasingly more expensive as this limit is approached (larger length scales imply not only more simulation particles, but also longer evolution time scales), a formulation comprising an equilibrium part and a part simulated by particles becomes more efficient as this limit is approached, because it is able to relegate increasingly more of the description to the equilibrium part, thus *reducing* the number of particles required for the simulation. As we show in §5, such decompositions can also be thought of as variance reduction techniques, which result in simulation methods that exhibit significantly reduced statistical uncertainty [28] compared with traditional particle simulation methods.

We begin our review with an introduction to kinetic transport descriptions and typical methods of simulation.

2. Boltzmann equation

We consider kinetic transport as described by the Boltzmann equation in the absence of external fields, which can be written in the form

$$\frac{\partial f}{\partial t} + \mathbf{c} \cdot \frac{\partial f}{\partial \mathbf{r}} = \left[\frac{\mathrm{d}f}{\mathrm{d}t}\right]_{\mathrm{coll}}, \tag{2.1}$$

where $f(\mathbf{r}, \mathbf{c}, t)$ is the velocity distribution function [17], $\mathbf{r} = (x, y, z)$ is the position vector in physical space, $\mathbf{c} = (c_x, c_y, c_z)$ is the molecular velocity vector, t is time and $[\mathrm{d}f/\mathrm{d}t]_{\mathrm{coll}}$ denotes the collision operator which takes different forms depending on the carrier modelled. Extension of this work to include external fields does not present any additional conceptual challenges.

The original Boltzmann collision operator was written for a hard-sphere gas [17] in the form

$$\left[\frac{\mathrm{d}f}{\mathrm{d}t}\right]_{\mathrm{coll}} (\mathbf{r}, \mathbf{c}, t) = \int_{\mathcal{R}^3} \int_{\mathcal{S}^2} [f_1'f' - f_1 f] \, ||\mathbf{c} - \mathbf{c}_1|| \sigma \, \mathrm{d}^2\boldsymbol{\Omega} \, \mathrm{d}^3\mathbf{c}_1, \tag{2.2}$$

where $\hat{d}$ is the gas-molecule effective diameter, $\sigma = \hat{d}^2/4$ is the differential collision cross-section for hard spheres and $f' = f(\mathbf{r}, \mathbf{c}', t)$, $f_1 = f(\mathbf{r}, \mathbf{c}_1, t)$, $f_1' = f(\mathbf{r}, \mathbf{c}_1', t)$; here, $\{\mathbf{c}_1, \mathbf{c}\}$ are the pre-collision velocities and $\{\mathbf{c}_1', \mathbf{c}'\}$ are the post-collision velocities, related to the pre-collision velocities through the scattering angle $\boldsymbol{\Omega}$. Integration in velocity space extends over all possible velocities, and the solid angle is integrated over the surface of the unit sphere ($\mathcal{S}^2$).

The global equilibrium distribution

$$f^0(\mathbf{c}) = \frac{n_0}{\pi^{3/2} c_0^3} \exp\left(-\frac{\mathbf{c}^2}{c_0^2}\right) \tag{2.3}$$

for a gas at (constant) reference temperature T_0 and number density n_0 is a special case of a Maxwell–Boltzmann distribution

$$f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}, t) = \frac{n_{\mathrm{MB}}}{\pi^{3/2} c_{\mathrm{MB}}^3} \exp\left(-\frac{||\mathbf{c} - \mathbf{u}_{\mathrm{MB}}||^2}{c_{\mathrm{MB}}^2}\right), \tag{2.4}$$

where in the latter $\mathbf{u}_{\mathrm{MB}} = \mathbf{u}_{\mathrm{MB}}(\mathbf{r}, t)$, $c_{\mathrm{MB}}(\mathbf{r}, t) = \sqrt{2RT_{\mathrm{MB}}(\mathbf{r}, t)}$, $T_{\mathrm{MB}} = T_{\mathrm{MB}}(\mathbf{r}, t)$ and $n_{\mathrm{MB}} = n_{\mathrm{MB}}(\mathbf{r}, t)$. A gas described by (2.3) has a mean free path $\lambda = 1/(\sqrt{2}\pi n_0 \hat{d}^2)$, while the molecular collision time is given by $\tau_{\mathrm{coll}} = \sqrt{\pi}\lambda/(2c_0)$, where $c_0 = \sqrt{2RT_0}$ is the most probable molecular speed, $R = k_{\mathrm{b}}/m$ is the gas constant, k_{b} is Boltzmann's constant and m is the molecular mass. Based on these quantities, we can define a Knudsen number $Kn = \lambda/L$, where L is the characteristic hydrodynamic length scale.

The hard-sphere operator predicts transport coefficients that vary with temperature according to $T^{1/2}$, while real gases exhibit a more general temperature dependence T^ϖ, where the temperature coefficient (ϖ) is typically closer to $\frac{3}{4}$. For this reason, the variable hard sphere (VHS) model [29] has been developed. This model, which finds wide application in engineering studies, achieves the desired temperature dependence (T^ϖ) by introducing a velocity-dependent collision cross-section into the Boltzmann operator (2.2) through a hard-sphere cross-section with a velocity-dependent diameter $\hat{d} = \hat{d}_{\mathrm{r,ref}}(c_{\mathrm{r,ref}}/||\mathbf{c} - \mathbf{c}_1||)^{\varpi - 1/2}$, where $c_{\mathrm{r,ref}} = 4\sqrt{RT_{\mathrm{ref}}/\pi}$ and $\hat{d}_{\mathrm{r,ref}}$ is the reference diameter at temperature T_{ref}.

Owing to the complexity associated with the above collision operators, simpler approximations have been proposed. The most well known, perhaps, is the Bhatnagar–Gross–Krook (BGK) model [17]

$$\left[\frac{\mathrm{d}f}{\mathrm{d}t}\right]_{\mathrm{coll}} (\mathbf{r}, \mathbf{c}, t) = \frac{f^{\mathrm{loc}} - f}{\tau} \tag{2.5}$$

which, by construction, retains (2.3) as the global equilibrium distribution and also preserves mass, momentum and energy conservation (collisional invariants) [17]. This model has also found widespread application in many areas of physics where it is frequently used to model the transport of electrons [4,5], photons [18] and phonons [30], and is more generally referred to as the relaxation-time approximation. In this model, τ is the relaxation time and f^{loc} is the local equilibrium distribution; for a dilute gas, the latter is of the Maxwell–Boltzmann form

$$f^{\mathrm{loc}}(\mathbf{r}, \mathbf{c}, t) = \frac{n_{\mathrm{loc}}}{\pi^{3/2} c_{\mathrm{loc}}^3} \exp\left(-\frac{||\mathbf{c} - \mathbf{u}_{\mathrm{loc}}||^2}{c_{\mathrm{loc}}^2}\right), \tag{2.6}$$

where

$$n_{\rm loc} = n_{\rm loc}(\mathbf{r}, t) = \int_{\mathcal{R}^3} f \, d^3\mathbf{c} \tag{2.7}$$

is the local gas number density,

$$\mathbf{u}_{\rm loc} = \mathbf{u}_{\rm loc}(\mathbf{r}, t) = \frac{1}{n_{\rm loc}} \int_{\mathcal{R}^3} \mathbf{c} f \, d^3\mathbf{c} \tag{2.8}$$

denotes the local flow velocity, and $c_{\rm loc}(\mathbf{r}, t) = \sqrt{2RT_{\rm loc}(\mathbf{r}, t)}$, where

$$T_{\rm loc} = T_{\rm loc}(\mathbf{r}, t) = \frac{1}{3n_{\rm loc}R} \int_{\mathcal{R}^3} ||\mathbf{c} - \mathbf{u}_{\rm loc}||^2 f \, d^3\mathbf{c} \tag{2.9}$$

is the local gas temperature. The molecular mean free path associated with the BGK gas is given by $\lambda = 2c_0\tau/\sqrt{\pi}$.

(a) Particle methods for solving the Boltzmann equation

Owing to the high dimensionality associated with $f(\mathbf{r}, \mathbf{c}, t)$, as well as the need to accurately and *stably* capture the propagation of travelling discontinuities in the distribution function [31], particle methods are generally the method of choice for solving the Boltzmann equation [32]. The most prevalent Boltzmann solution method is a particle method known as direct simulation Monte Carlo (DSMC) [29], which in addition to satisfying the above requirements, combines simplicity with an intuitive formulation which naturally employs importance sampling, and very low memory usage.

DSMC solves [33] the Boltzmann equation by simulating molecular motion as a series of timesteps, each of length Δt $(\ll \lambda/c_0)$ [33,34], and during which collisionless advection and collision substeps are performed [29]. This can be thought of as a splitting scheme in which the collisionless advection substep integrates

$$\frac{\partial f}{\partial t} + \mathbf{c} \cdot \frac{\partial f}{\partial \mathbf{r}} = 0 \tag{2.10}$$

while the collision substep integrates

$$\frac{\partial f}{\partial t} = \left[\frac{df}{dt}\right]_{\rm coll} (\mathbf{r}, \mathbf{c}, t). \tag{2.11}$$

During the advection substep, the positions of all particles are updated according to their velocities. During the collision substep, the distribution function is updated by processing binary collisions between collision partners chosen from the same computational cell of size Δx $(\ll \lambda)$ [33,35] using an acceptance–rejection procedure [36]. This splitting procedure leads to a second-order accurate in time algorithm provided the latter is appropriately symmetrized [34,37]. DSMC is also second-order accurate in space [35,37]. Detailed descriptions of the DSMC algorithm can be found in [36].

3. The Chapman–Enskog expansion

The Navier–Stokes limit can be recovered from the Boltzmann equation through the celebrated Chapman–Enskog procedure, which seeks to expand the distribution in a power series of the Knudsen number Kn. The essential result of this procedure is the expansion [18]

$$f(\mathbf{r}, \mathbf{c}, t) = f^{\rm loc}(\mathbf{r}, \mathbf{c}, t) + Kn \cdot h(\mathbf{r}, \mathbf{c}, t) + O(Kn^2), \tag{3.1}$$

where the first term in the expansion is the *local equilibrium* which gives rise to the Euler set of hydrodynamic equations (no diffusive transport); keeping terms up to and including $O(Kn)$ recovers the Navier–Stokes set of equations [18] (diffusive, linear-gradient transport). More details, as well as a complete mathematical treatment of the hard-sphere collision operator (2.2), can be found in [19]; a less detailed exposition can be found in [18].

4. Algebraic decomposition based on the Chapman–Enskog expansion

The Chapman–Enskog result (3.1) provides a natural framework for reducing the computational cost associated with approaching the $Kn \ll 1$ limit with particle simulations. Specifically, the algebraic decomposition

$$f(\mathbf{r}, \mathbf{c}, t) = f^{loc}(\mathbf{r}, \mathbf{c}, t) + g(\mathbf{r}, \mathbf{c}, t) \tag{4.1}$$

motivates a simulation scheme in which a particle method is used to simulate $g(\mathbf{r}, \mathbf{c}, t)$ and a method more suited to the characteristics of $f^{loc}(\mathbf{r}, \mathbf{c}, t)$ is used to capture the evolution of the latter [38]. In fact, because

$$\left[\frac{\mathrm{d} f^{loc}}{\mathrm{d} t} \right]_{coll} = 0, \tag{4.2}$$

the latter method only needs to update f^{loc} by only considering its advection (2.10). In contrast to traditional particle methods which become stiff in the $Kn \ll 1$ limit, methods based on this decomposition should, in theory, relegate increasingly more of the description to $f^{loc}(\mathbf{r}, \mathbf{c}, t)$ as $Kn \to 0$ (since $g(\mathbf{r}, \mathbf{c}, t) \to 0$) and require a smaller number of particles in this limit.

Unfortunately, decomposition (4.1) is no longer preserved by the advection operator. This can be illustrated by considering the form of the distribution function

$$f(\mathbf{r}, \mathbf{c}, t + \Delta t) = f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) + g(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) \tag{4.3}$$

following the action of (2.10) over a timestep Δt. In general, $f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t)$ is not a Maxwell–Boltzmann distribution and thus $f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) \neq f^{loc}(\mathbf{r}, \mathbf{c}, t + \Delta t)$. This implies that a scheme based on (4.1) will, in general, require some form of reconstruction process that restores

$$f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) + g(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) \to f^{loc}(\mathbf{r}, \mathbf{c}, t + \Delta t) + g(\mathbf{r}, \mathbf{c}, t + \Delta t) \tag{4.4}$$

after every time the advection operator is applied (e.g. every timestep). Perhaps the simplest approach would be to generate particles by sampling $f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) - f^{loc}(\mathbf{r}, \mathbf{c}, t + \Delta t)$ and add them to the particles which represent the distribution $g(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t)$. This requires the generalization [39] of the particle description by including negative particles, since $f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) - f^{loc}(\mathbf{r}, \mathbf{c}, t + \Delta t)$ will, in general, be of indeterminate sign. Such approaches are discussed in §§5 and 6; in the next section, we discuss a number of approaches [38,40,41] where the decomposition (4.1) is instead constrained to be a convex combination of two positive functions.

(a) Applications

In one class of approaches aimed at problems in the $Kn \ll 1$ limit [38,40,41], decomposition (4.1) is constrained to be of the form

$$f(\mathbf{r}, \mathbf{c}, t) = w(\mathbf{r}, \mathbf{c}, t) f^{loc}(\mathbf{r}, \mathbf{c}, t) + g(\mathbf{r}, \mathbf{c}, t), \quad g(\mathbf{r}, \mathbf{c}, t) \geq 0 \tag{4.5}$$

with $w(\mathbf{r}, \mathbf{c}, t) \in [0, 1]$ determining the 'fraction' of the equilibrium distribution f^{loc} to be 'separated out'. The existence of w is a result of the requirement that $g(\mathbf{r}, \mathbf{c}, t) \geq 0$, which allows the latter to be straightforwardly interpreted as a probability density of positive particles, but means that, in general, $g(\mathbf{r}, \mathbf{c}, t) \neq Kn \cdot h(\mathbf{r}, \mathbf{c}, t) + O(Kn^2)$, i.e. the existence of a local equilibrium distribution is not fully taken advantage of. On the other hand, in the $Kn \to 0$ limit, the correct limiting distribution is recovered ($w \to 1$ and $f \to f^{loc}$).

A number of implementations of this procedure have been reported. In [38], spatially homogeneous relaxation problems were considered. In [40,41], one-dimensional shock problems were studied. As discussed above, in spatially dependent problems, the advection operator does not preserve the form of (4.5), but instead produces

$$f(\mathbf{r}, \mathbf{c}, t + \Delta t) = w(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) + g(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) \tag{4.6}$$

which requires some form of reconstruction. Reconstruction schemes are first discussed in [38], where it is proposed to convert $w(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t) f^{loc}(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t)$ into particles and add those to the population representing $g(\mathbf{r} - \mathbf{c}\Delta t, \mathbf{c}, t)$—in other words, create an intermediate representation with $w = 0$. From this intermediate representation, $w(\mathbf{r}, \mathbf{c}, t + \Delta t)$ and $f^{loc}(\mathbf{r}, \mathbf{c}, t + \Delta t)$ can be determined using a number of approaches, the most direct of which would be to sample the particle population (measure its moments) to determine the parameters of $f^{loc}(\mathbf{r}, \mathbf{c}, t + \Delta t)$.

In [40], both components of f, namely $w(\mathbf{r}, \mathbf{c}, t) f^{loc}(\mathbf{r}, \mathbf{c}, t)$ and $g(\mathbf{r}, \mathbf{c}, t)$, are simulated by particles. The algorithm for advecting $w(\mathbf{r}, \mathbf{c}, t) f^{loc}(\mathbf{r}, \mathbf{c}, t)$ is found to be 'somewhat clumsy' [40]; it is proposed that a Navier–Stokes-based approach for the advection step of $w(\mathbf{r}, \mathbf{c}, t) f^{loc}(\mathbf{r}, \mathbf{c}, t)$ is preferable. Various approaches using a numerical (e.g. finite volume) solver to perform the advection step of $w(\mathbf{r}, \mathbf{c}, t) f^{loc}(\mathbf{r}, \mathbf{c}, t)$ are presented in [41,42], where also a number of reconstruction schemes are discussed and evaluated. Results demonstrate accurate simulation of shockwave propagation in the range $10^{-5} \leq Kn \leq 10^{-1}$.

In these studies, algebraic decomposition is coupled to a time-relaxed [43] time-integration scheme which removes the stiffness associated with time integration of (2.11) in the $Kn \to 0$ limit. The time-relaxed scheme uses the formal representation of the solution to (2.11) in the form of an infinite Wild expansion [44] to emulate an implicit integration method, which is stable for all timestep sizes, and approaches the correct limiting distribution ($f^{loc}(\mathbf{r}, \mathbf{c}, t)$) for $\Delta t \to \infty$; however, complexity considerations limited implementations to a first-order scheme, obtained by *approximating* the high-order terms in the expansion by a Maxwellian distribution [38], which amounts to thermalizing particles that have had more than a single collision. More recently, higher-order schemes for the VHS model have been presented [45], but not implemented within an algebraic decomposition framework.

5. Algebraic decomposition and variance reduction

The property

$$\left[\frac{d f^{MB}}{dt} \right]_{coll} = 0, \tag{5.1}$$

where f^{MB} is *any* Maxwellian distribution, means that this distribution plays a central role in kinetic problems, even away from the $Kn \ll 1$ limit. It is particularly useful in cases where the deviation from a well-defined equilibrium state is small. In such cases, it can be used [46,47] for obtaining solutions to the Boltzmann equation; the solution proceeds by linearizing the governing equation and boundary conditions about the reference state $f^0(\mathbf{c})$, resulting in a linear problem in terms of the deviation from equilibrium, $f^d(\mathbf{r}, \mathbf{c}, t)$, defined by

$$f(\mathbf{r}, \mathbf{c}, t) = f^0(\mathbf{c}) + f^d(\mathbf{r}, \mathbf{c}, t). \tag{5.2}$$

Small deviation from equilibrium typically implies [46] that the variation of local conditions throughout the domain remains sufficiently small that f^0 may be based on any convenient (but physically relevant) reference properties (e.g. uniform initial condition if applicable, system average properties, etc).

In the interest of brevity, we will denote small deviation from equilibrium by $\epsilon \ll 1$, where ϵ is a problem-dependent measure of the departure from equilibrium which usually can be taken to be the Mach number Ma ($= \sqrt{2/\gamma} U / c_0$) or a dimensionless temperature difference $\Delta T / T_0$, where U is the characteristic flow speed, ΔT is a characteristic temperature difference, and γ is the ratio of specific heats ($= \frac{5}{3}$ for a monatomic gas).

We emphasize here that a small f^d does *not* require $Kn \ll 1$. We can see this by noticing that $\epsilon \ll 1$ does not involve a length scale restriction. For example, in the case of gas flow, the von-Karman relation $Ma \approx Re\, Kn$ implies that the Mach number can be small for $Kn \gg 1$, provided the

Reynolds number Re is sufficiently small, which can be realized when the characteristic length scale L is sufficiently small. Here, $Re = \rho U L / \mu$, where μ is the gas viscosity and $\rho\ (= mn)$ is the gas density.

(a) Variance reduction

Decomposition (5.2) can be used [28] for reducing the statistical uncertainty of direct Monte Carlo simulations of the Boltzmann equation by describing part of the solution, namely $f^0(\mathbf{c})$, deterministically. The need for this arises because modern applications associated with nanoscale science and engineering typically involve small deviations from equilibrium [1,47], which translates to small hydrodynamic signals (e.g. flow velocity, heat flux, etc.). In this limit, DSMC, *as well as all molecular methods* [48], become very expensive because resolution of the hydrodynamic signals requires very large numbers of samples. Specifically, in the limit of small deviation from equilibrium, statistical fluctuations are dominated by the equilibrium spectrum (which can be quantified using equilibrium statistical mechanics [48]) that does not depend on ϵ and makes resolution of smaller signals increasingly difficult [48].

Decomposition (5.2) lends itself to a control variate approach for reducing the statistical uncertainty of Monte Carlo integration. In control variate Monte Carlo integration, we write

$$\int_{\mathcal{R}^3} W(\mathbf{c}) f(\mathbf{c})\, \mathrm{d}^3 \mathbf{c} = \int_{\mathcal{R}^3} W(\mathbf{c}) [f(\mathbf{c}) - \xi(\mathbf{c})]\, \mathrm{d}^3 \mathbf{c} + \int_{\mathcal{R}^3} W(\mathbf{c}) \xi(\mathbf{c})\, \mathrm{d}^3 \mathbf{c}, \qquad (5.3)$$

where the function $\xi(\mathbf{c})$ fulfils the following requirements:

— captures most of the variation of $f(\mathbf{c})$ (i.e. $\xi(\mathbf{c}) \approx f(\mathbf{c})$) and
— $\int_{\mathcal{R}^3} W(\mathbf{c}) \xi(\mathbf{c}) d^3 \mathbf{c}$ can be deterministically (analytically or otherwise) evaluated.

Using a Monte Carlo method to evaluate *only the term* $\int_{\mathcal{R}^3} W(\mathbf{c})[f(\mathbf{c}) - \xi(\mathbf{c})] \mathrm{d}^3 \mathbf{c}$ results in significantly reduced statistical uncertainty, because most of the statistical uncertainty is removed through the deterministic evaluation of $\int_{\mathcal{R}^3} W(\mathbf{c}) \xi(\mathbf{c})\, \mathrm{d}^3 \mathbf{c}$ [49].

As stated above, decomposition (5.2) lends itself naturally to control variate integration, since it clearly identifies a control $\xi = f^0$ that is guaranteed to satisfy the above requirements ((1) $f^0 \approx f$ and (2) equilibrium moments are easy to calculate deterministically, if not intuitively obvious); moreover, it has a number of desirable attributes which alleviate the limitations arising from using a time-dependent equilibrium distribution (4.1), discussed in §4. Specifically, since f^0 does not vary in time, it does not need to be evolved by any 'auxiliary' numerical method, nor does it need to be reconstructed every timestep.

A simulation method using the above construct reduces to a method for simulating the dynamics of the deviation from equilibrium $f^d = f - f^0$; such a method has been developed [32,50–52] and is referred to as low-variance deviational simulation Monte Carlo (LVDSMC), which highlights its connection to DSMC to which it is related. LVDSMC proceeds in time using the standard splitting scheme (2.10) and (2.11) on the 'deviational population' of *signed* particles (i.e. positive or negative) sampling f^d. Specifically, it follows directly from (2.10) and (5.2) that

$$\frac{\partial f^d}{\partial t} + \mathbf{c} \cdot \frac{\partial f^d}{\partial \mathbf{r}} = 0, \qquad (5.4)$$

and thus the motion of deviational particles is identical to that of physical particles. The effect of collisions is implemented by interpreting the collision integral in terms of source and sink terms

$$\left[\frac{\mathrm{d}f}{\mathrm{d}t} \right]_{\text{coll}} = \left[\frac{\mathrm{d}f^d}{\mathrm{d}t} \right]_{\text{coll}} = \underbrace{\frac{f^{\text{loc}} - f^0}{\tau}}_{\text{generation}} \quad \underbrace{-\frac{f^d}{\tau}}_{\text{deletion}}, \qquad (5.5)$$

which modify the distribution by generating signed particles (with sign $\mathrm{sgn}(f^{\text{loc}} - f^0)$) from the distribution $|f^{\text{loc}} - f^0|$ and deleting particles from the deviational distribution.
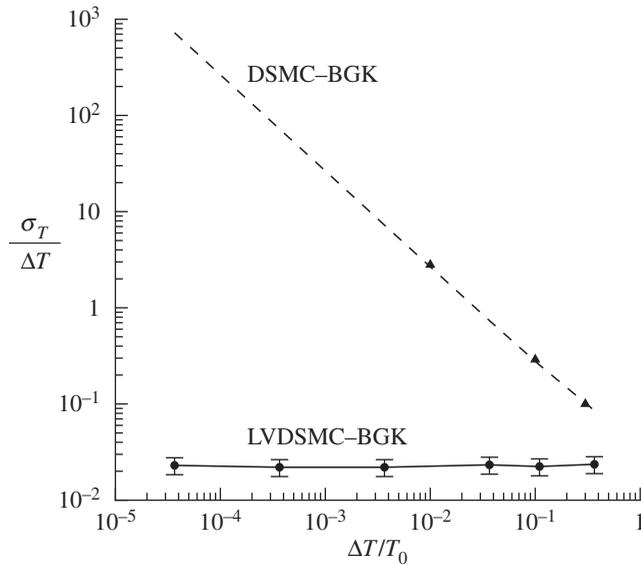
Figure 1. The relative statistical uncertainty in temperature, $\sigma_T/\Delta T$, as a function of $\Delta T/T_0$, with $T_0 = 273$ K, for steady-state heat exchange between two fully accommodating parallel plates at different temperatures a distance $L = \sqrt{\pi}\lambda/2$ apart. BGK simulation results (symbols with error bars on solid line) are presented and compared with the theoretical prediction [48] for DSMC (dashed line), which serves as a canonical case of a 'non-deviational' method. Triangles show actual DSMC (BGK) results verifying that equilibrium theory is reliable up to at least $\Delta T \approx 0.3 T_0$.

In the interest of simplicity, the above discussion considered the BGK model with a constant relaxation time; more details for this implementation can be found in [51]. An example with a variable relaxation time is discussed in §5b(ii). The hard-sphere case was first treated in [32], while recently, collision algorithms with no timestep error for the more general VHS model were developed [53] and implemented [52]. A discussion of stability problems associated with integrating $[df^d/dt]_{coll}$ by extending DSMC collisional processes, rather than the generation/deletion interpretation discussed above, can be found in [32,39]. Applications to microscale transport problems are discussed in §5b.

An interesting feature of these algorithms is that by removing the fluctuations associated with equilibrium, the resulting methods exhibit statistical uncertainty which is proportional to the non-equilibrium signal. In other words, the relative statistical uncertainty (standard deviation in the signal over the signal characteristic magnitude) is independent of the signal magnitude; this means that these methods can resolve, at a constant relative statistical uncertainty, arbitrarily small deviations from equilibrium at a cost that does not increase as the signal decreases. This is demonstrated in figure 1, which shows the relative statistical uncertainty, $\sigma_T/\Delta T$, in the case of heat transfer between two parallel plates at temperatures T and $T + \Delta T$, respectively, as a function of ΔT [51]; in the figure, σ_T is the standard deviation in the temperature estimate. The figure also shows the relative statistical uncertainty for DSMC, which grows as ΔT decreases, since in that case σ_T is dominated by the equilibrium fluctuations [48,51] (and is thus a constant). We note that the theoretical speedup scales as the square of the ratio of the two relative statistical uncertainties.

(b) Applications

Below, we review a few examples that illustrate the computational savings of variance-reduced computational approaches based on algebraic decomposition of the distribution function.
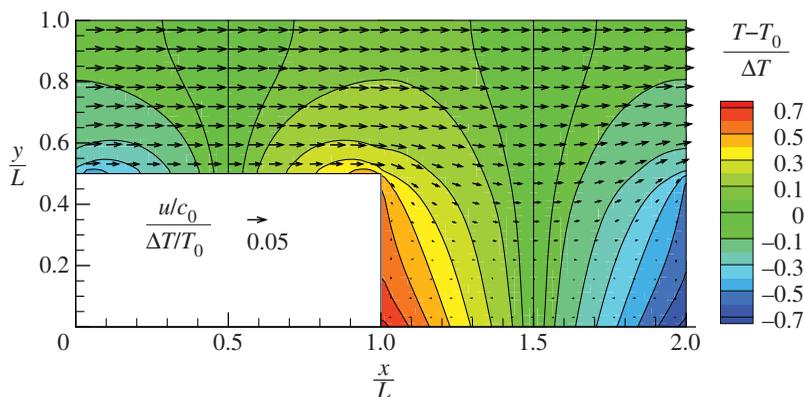
Figure 2. LVDSMC simulation of a simple Knudsen compressor with $Kn = 1$ and arbitrarily small temperature gradients. The arrows are velocity vectors and the contours are isothermal lines. (Online version in colour.)

(i) Simulation of Knudsen compressors

Figure 2 shows an LVDSMC simulation of thermal transpiration flow of a hard-sphere gas [52] through a single stage of a Knudsen compressor at a Knudsen number $Kn = \lambda/L = 1$; the lower boundary is diffusely reflecting with a temperature (T_w) given by

$$\frac{T_w - T_0}{\Delta T} = \begin{cases} 2\left(\frac{x}{L}\right) - 1 & \text{for } x < L, \\ 3 - 2\left(\frac{x}{L}\right) & \text{for } x > L, \end{cases}$$

while the upper boundary is specularly reflecting, imposing a symmetry boundary condition. The system is periodic in the axial (flow) direction ($T(x, y) = T(x + 2L, y)$, etc.).

Knudsen compressors are devices with no moving parts, which exploit the kinetic phenomenon of thermal transpiration—the tendency of a rarefied ($Kn > 0$) gas in contact with a wall along which a temperature gradient exists to move from cold regions to hot regions [46]—to pump or compress a gas and more recently to separate gases [54]. These devices generally consist of many stages, each with a capillary section ($x < L$ in the figure) with a positive streamwise temperature gradient, followed by a connector section ($x > L$ in the figure) with a negative temperature gradient. Simulation of these devices is challenging because of the high cost associated with obtaining noise-free results for representative (moderate) temperature differences [55]. LVDSMC overcomes this limitation by simulating only the deviation from equilibrium.

(ii) Phonon transport for microscale heat transfer

Applications of the methods discussed above can also be found in the field of microscale solid-state heat transfer. Heat carriers (phonons) feature a mix of ballistic and diffusive behaviour at scales comparable to the phonon mean free path (O(100 nm) for Si at 300 K), which manifests itself through the breakdown of Fourier's law. In spite of the conceptual and physical differences between molecular gases and so-called phonon gases (Bose–Einstein distribution, existence of a dispersion relation, frequency dependent scattering rate, etc.), direct Monte Carlo simulations are widely used to solve phonon transport problems [56–60]. Practical applications of such problems are drawn from thermoelectrics, manufacturing, design and cooling of microelectronic devices, and more generally from any application involving microscale solid-state thermal management. Recent applications typically involve small temperature differences, making variance reduction very useful.

At an equilibrium state with temperature T, phonons follow Bose–Einstein statistics leading to an equilibrium distribution given by

$$\hat{f}_T^{eq}(\omega) = \frac{1}{\exp(\hbar\omega/k_b T) - 1}. \tag{5.6}$$

Non-equilibrium transport of phonons is well approximated by a Boltzmann transport equation (BTE) that is very similar to the BGK model presented in §2; however, in the phonon case, the distribution function $\hat{f} = \hat{f}(\mathbf{r}, \mathbf{k}, t, p)$ represents the occupation number of the energy levels and is a function of space, the phonon wavevector ($\mathbf{k}$), time and polarization p; the main polarizations for three-dimensional crystals are longitudinal acoustic, transverse acoustic, longitudinal optical and transverse optical [3]. Although phonon wave vectors are quantized, the energy levels can be treated as a quasi-continuum by introducing the density of states which accounts for degeneracies of energy levels and therefore allows sums over $\mathbf{k}$ to be transformed into integrals. Owing to the presence of a dispersion relation that is typically of the form $\omega = \omega(k, p)$, where ω is the phonon angular frequency and $k = |\mathbf{k}|$, it is usually more convenient to use $\hat{f} = \hat{f}(\mathbf{r}, \omega, \boldsymbol{\Omega}, t, p)$ where $\boldsymbol{\Omega}$ is the solid angle associated with the representation of $\mathbf{k}$ in spherical coordinates. The density of states $D(\omega, p)$ then allows one to write $\hat{f}(\mathbf{r}, \mathbf{k}, t, p) d^3\mathbf{k}/(2\pi)^3 = D(\omega, p)\hat{f}(\mathbf{r}, \omega, \boldsymbol{\Omega}, t, p)\, d\omega\, d\boldsymbol{\Omega}/(4\pi)$. The resulting Boltzmann equation reads

$$\frac{\partial \hat{f}}{\partial t} + \mathbf{v}_g \cdot \frac{\partial \hat{f}}{\partial \mathbf{r}} = \frac{\hat{f}^{loc} - \hat{f}}{\tau}, \tag{5.7}$$

where $\mathbf{v}_g = \mathbf{v}_g(\omega, p)$ is the phonon group velocity, $\tau = \tau(\omega, p, T_{loc})$ is the relaxation time and T_{loc} is the (local) temperature defined through the implicit relation

$$\int_{\mathcal{S}^2} \int_{\omega=0}^{\omega_{max}} \sum_p \hbar\omega \frac{D(\omega, p)}{4\pi} \hat{f}\, d\omega\, d\boldsymbol{\Omega} = \int_{\omega=0}^{\omega_{max}} \sum_p \hbar\omega D(\omega, p) \hat{f}_{T_{loc}}^{eq}\, d\omega. \tag{5.8}$$

In contrast to the BGK model (see §2), here the scattering rate depends on frequency and on temperature. As a result, the local equilibrium distribution $\hat{f}^{loc}$ in the collision term is evaluated [17,60] at the pseudo-local temperature $\tilde{T}$ (i.e. $\hat{f}^{loc} = \hat{f}_{\tilde{T}}^{eq}$), which ensures that energy is conserved during scattering, namely

$$\int_{\mathcal{S}^2} \int_{\omega=0}^{\omega_{max}} \sum_p \frac{\hbar\omega D(\omega, p)\hat{f}}{4\pi \tau(\omega, p, T_{loc})}\, d\omega\, d\boldsymbol{\Omega} = \int_{\omega=0}^{\omega_{max}} \sum_p \frac{\hbar\omega D(\omega, p)\hat{f}_{\tilde{T}}^{eq}}{\tau(\omega, p, T_{loc})}\, d\omega. \tag{5.9}$$

Given that the only conserved quantity during scattering is energy, it is preferable [61,62] to multiply the BTE (5.7) by $\hbar\omega$ (energy of a phonon) and to consider particles that simulate phonon bundles with fixed energy. Using this approach, energy conservation is automatically ensured if we conserve the number of computational particles, whereas in previous formulations [58,59], a phonon deletion/creation process was required.

Variance reduction can again be achieved by considering the difference between the actual distribution and an appropriate nearby equilibrium described by $\hat{f}^0 = \hat{f}_{T_0}^{eq}$. The equation obeyed by the difference $\hat{f}^d = \hat{f} - \hat{f}^0$ can be readily written down as

$$\frac{\partial \hbar\omega D\hat{f}^d}{\partial t} + \mathbf{v}_g \cdot \frac{\partial \hbar\omega D\hat{f}^d}{\partial \mathbf{r}} = \frac{\hbar\omega D(\hat{f}^{loc} - \hat{f}^0)}{\tau} - \frac{\hbar\omega D\hat{f}^d}{\tau} \tag{5.10}$$

and deviational particles can be used to simulate it. In this case, particles representing the distribution $\hbar\omega D\hat{f}^d$ are deleted at a frequency-dependent rate $\tau(\omega, p, T_{loc})^{-1}$, while particles of sign $\text{sgn}(\hat{f}^{loc} - \hat{f}^0)$ are generated from the distribution $|\hbar\omega D(\hat{f}^{loc} - \hat{f}^0)|/\tau$.

As discussed in §5a, this formulation removes the statistical uncertainty associated with equilibrium fluctuations which can become prohibitive when the deviation from equilibrium is small. An additional advantage of this formulation manifests itself in problems where a large
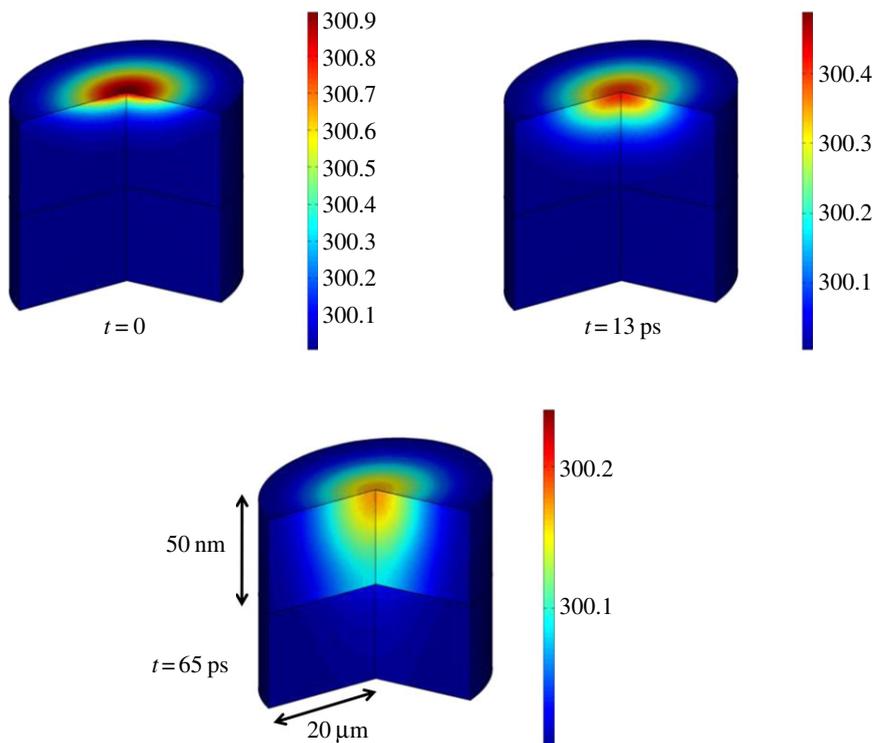
Figure 3. System composed of a 50 nm slab of aluminium on a semi-infinite silicon substrate. At $t = 0$, a laser pulse induces a small and localized rise in the temperature at the top surface. The subsequent evolution of the temperature field is computed by assuming the top surface to be adiabatic and that the Al–Si interface obeys the diffuse mismatch model [63]. The figure shows the resulting temperature field at three different times. (Online version in colour.)

fraction of the computational domain remains in equilibrium, such as the transient problem discussed briefly below and shown in figure 3. In this problem, a thin slab of crystalline material, initially at uniform temperature and resting on a silicon substrate, is subjected to a localized heating by a laser source. As a result of the localized heating, large parts of the computational domain are in thermal equilibrium, especially for early times. Algebraic decomposition enables the use of particles only in regions where $\hat{f}^d \neq 0$, thus significantly reducing the computational cost compared with the standard approach.

Figure 3 shows three snapshots of the temperature field solution of the laser heating problem discussed above obtained using the variance-reduced method. More details on the problem formulation and simulation method can be found in [62]. Despite the very small signal (temperature differences are typically much smaller than 1 K and of the order of 10^{-3} K at late times), the time-dependent temperature field is resolved to very small statistical uncertainty. The computational savings compared with a standard Monte Carlo simulation method have been estimated [62] to be of the order of 10^9. In other words, simulation of this problem (at the same statistical uncertainty) is infeasible using traditional methods with our present computational resources.

6. Variance reduction using a local equilibrium

As shown in §5, a deviational algorithm based on a global equilibrium distribution provides significant acceleration in the form of variance reduction for $\epsilon \ll 1$, but also computational savings in cases where part of the simulation domain is in equilibrium. Moreover, this algorithm is

reasonably simple and does not require methods for evolving or reconstructing f^0. However, although as $\epsilon \to 0$ the variance reduction improves (see also figure 1), the variance reduction does not improve as $Kn \to 0$ (for a constant ϵ) because in this limit $f^d = f - f^0 \to f^{\mathrm{loc}} - f^0 \sim \epsilon$. In other words, in the $Kn \to 0$ limit, the degree of variance reduction can further be improved using the decomposition

$$f(\mathbf{r}, \mathbf{c}, t) = f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}) + f^d(\mathbf{r}, \mathbf{c}, t), \tag{6.1}$$

where f^{MB} varies in space. We note, however, that $f^{\mathrm{MB}} \neq f^{\mathrm{MB}}(t)$ and thus *no fluid-dynamic solver is required for its evolution.*

The seemingly impossible task of approximating $f^{\mathrm{loc}}(t)$ by $f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}) \neq f^{\mathrm{MB}}(t)$ can be achieved [32] by updating f^{MB} once every timestep (e.g. before the start of a new timestep). In other words, after every timestep, the update seeks to find a new $f^{\mathrm{MB}}_{t+\Delta t} \approx f^{\mathrm{loc}}(t + \Delta t)$ without trying to construct the complete distribution from its samples, in contrast to the approach of §4a. This 'change of basis' can be achieved [32] by introducing a change $\Delta f^{\mathrm{MB}} = f^{\mathrm{MB}}_{t+\Delta t} - f^{\mathrm{MB}}_t \approx f^{\mathrm{loc}}(t + \Delta t) - f^{\mathrm{loc}}(t)$ such that at time $t + \Delta t$ decomposition (6.1) is given by

$$f(\mathbf{r}, \mathbf{c}, t + \Delta t) = f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}) + \Delta f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}) + f^d(\mathbf{r}, \mathbf{c}, t + \Delta t) - \Delta f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}) \tag{6.2}$$

$$= f^{\mathrm{MB}}_{t+\Delta t}(\mathbf{r}, \mathbf{c}) + f^d(\mathbf{r}, \mathbf{c}, t + \Delta t) - \Delta f^{\mathrm{MB}}(\mathbf{r}, \mathbf{c}), \tag{6.3}$$

where

$$\Delta f^{\mathrm{MB}} \approx \left[\frac{\Delta n_{\mathrm{MB}}}{n_{\mathrm{MB}}} + 2\boldsymbol{\zeta} \cdot \frac{\Delta \mathbf{u}_{\mathrm{MB}}}{c_{\mathrm{MB}}} + (2\zeta^2 - 3)\frac{\Delta c_{\mathrm{MB}}}{c_{\mathrm{MB}}} \right] f^{\mathrm{MB}} \tag{6.4}$$

and $\boldsymbol{\zeta} = (\mathbf{c} - \mathbf{u}_{\mathrm{MB}})/c_{\mathrm{MB}}$. Once the parameters $\Delta n_{\mathrm{MB}}, \Delta \mathbf{u}_{\mathrm{MB}}, \Delta c_{\mathrm{MB}}$ are determined, the change of basis can proceed ($f^{\mathrm{MB}}_{t+\Delta t} = f^{\mathrm{MB}}_t + \Delta f^{\mathrm{MB}}$); at the same time, deviational particles that sample $-\Delta f^{\mathrm{MB}}$ need to be generated and added to the f^d population.

It was shown in [32,50] that it is best to combine this step with the particle generation step of the collision algorithm, which for the BGK model can be summarized as

$$\left[\frac{\mathrm{d}f}{\mathrm{d}t} \right]_{\mathrm{coll}} \Delta t = \left[\frac{\mathrm{d}f^d}{\mathrm{d}t} \right]_{\mathrm{coll}} \Delta t = \underbrace{\frac{(f^{\mathrm{loc}} - f^{\mathrm{MB}})\Delta t}{\tau} - \Delta f^{\mathrm{MB}}}_{\text{generation}} \underbrace{+\Delta f^{\mathrm{MB}}}_{\text{change in equilibrium}} \underbrace{-\frac{f^d \Delta t}{\tau}}_{\text{deletion}}. \tag{6.5}$$

This formulation is advantageous because in addition to not requiring reconstruction of $f^{\mathrm{MB}}_{t+\Delta t}$ from its samples every timestep, it exploits analytical cancellation (between positive and negative particles) by sampling only the net function $(f^{\mathrm{loc}} - f^{\mathrm{MB}})\Delta t/\tau - \Delta f^{\mathrm{MB}}$ (as opposed to generating particles from the distributions $(f^{\mathrm{loc}} - f^{\mathrm{MB}})\Delta t/\tau$ and $-\Delta f^{\mathrm{MB}}$ separately). The reduction in complexity is immediately obvious, while the computational savings are considerable, provided Δf^{MB} is chosen appropriately (see below for a discussion): in the special case of the linearized BGK model, a choice of Δf^{MB} exists for which the cancellation is perfect (($f^{\mathrm{loc}} - f^{\mathrm{MB}})\Delta t/\tau - \Delta f^{\mathrm{MB}} = 0$) and therefore no particle generation is required; this could perhaps be expected from the physical interpretation of the BGK model (the action of the collision operator is to move the distribution towards local equilibrium) [50].

The above algorithm depends on an appropriate choice for Δf^{MB} which has the objective of tracking $f^{\mathrm{loc}}(t)$ using f^{MB}_t. In [32,50], this was reliably achieved by requiring that Δf^{MB} be chosen such that it absorbs the mass, momentum and energy change *due to the action of the particle-generation part of the collision integral*; in other words, Δf^{MB} is determined from

$$\int_{\mathcal{R}^3} \Delta f^{\mathrm{MB}} \phi \, \mathrm{d}^3 \mathbf{c} = \Delta t \int_{\mathcal{R}^3} \phi \left(\left[\frac{\mathrm{d}f^d}{\mathrm{d}t} \right]_{\mathrm{coll}} - \frac{f^d}{\tau} \right) \mathrm{d}^3 \mathbf{c}, \tag{6.6}$$

where $\phi = \{1, \mathbf{c}, ||\mathbf{c}||^2\}$. The rationale for this requirement is as follows [32]: in problems where the final distribution is of the Maxwell–Boltzmann type (say $\check{f}^{\mathrm{MB}}$), if f^d is constrained to have no net mass, momentum and energy, it has to be zero since it is also constrained to be equal to the difference of two Maxwell–Boltzmann distributions ($f^d = f - f^{\mathrm{MB}} = \check{f}^{\mathrm{MB}} - f^{\mathrm{MB}}$). In other words,
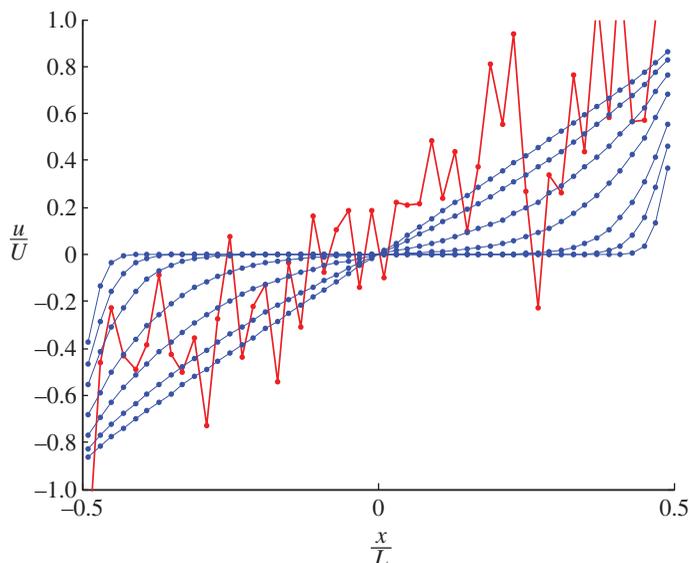
Figure 4. Transient (start-up) Couette flow of a BGK gas at $Kn = 0.2/\sqrt{\pi}$. The smooth lines show LVDSMC results at various times. The noisy result denotes the steady-state solution obtained via DSMC simulation using the same computational resources. (Online version in colour.)

it is reasonable to expect that for problems where a local equilibrium solution exists, through this formulation, the algorithm will be able to arrive at this solution (i.e. $f^{MB} = \check{f}^{MB}$, $f^d = 0$). Given the above, one may expect that in general problems, this formulation drives f^{MB} towards an appropriate Maxwell–Boltzmann distribution so as to make f^d small.

As stated above, in the case of the linearized BGK model this formulation reduces [50] to

$$\left[\frac{df^d}{dt}\right]_{coll} \Delta t = \underbrace{\Delta f^{MB}}_{\text{change in equilibrium}} \underbrace{-\frac{f^d \Delta t}{\tau}}_{\text{deletion}} \tag{6.7}$$

with

$$\Delta n_{MB} = (n_{loc} - n_{MB})\frac{\Delta t}{\tau}, \tag{6.8}$$

$$\Delta \mathbf{u}_{MB} = (\mathbf{u}_{loc} - \mathbf{u}_{MB})\frac{\Delta t}{\tau} \tag{6.9}$$

and

$$\Delta T_{MB} = (T_{loc} - T_{MB})\frac{\Delta t}{\tau}, \tag{6.10}$$

which is both simple and in line with the physical interpretation of the BGK model, namely that the effect of the collision operator is to drive f^{MB} towards f^{loc} while eliminating deviational particles.

The resulting simulation process is very efficient and can approach the $Kn \ll 1$ limit without experiencing the stiffness problems associated with traditional molecular methods. This can be illustrated in a number of ways. Figure 4 shows the time evolution of a start-up Couette flow at $Kn = 0.2/\sqrt{\pi}$; this essentially noise-free solution required a CPU time of 70 s (on one core of a 3 GHz Core2 Quad). The steady-state flow field obtained via DSMC calculation at $Ma = 0.02$ (based on the wall velocity U) using the same CPU time is also shown for comparison.

Figure 5 shows the total number of samples required to resolve the flow velocity (to a fixed level of statistical uncertainty) in the cell adjacent to the wall for a steady Couette flow. Here,
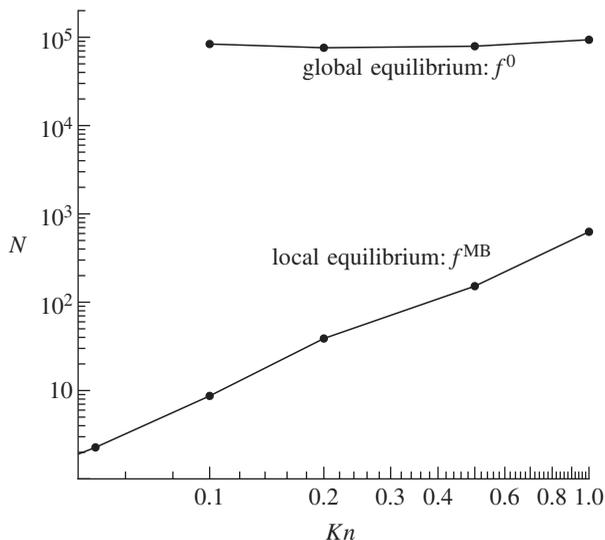
Figure 5. Total number of samples required (in arbitrary units) to achieve a fixed level of statistical error in the cell near a wall for steady-sate Couette flow simulations. Deviational approaches based on global and local equilibrium distributions are compared for a range of Knudsen numbers.

the number of samples is defined as the number of particles in the simulation multiplied by the number of independent statistical ensembles. The figure verifies that deviational simulations based on the local equilibrium require increasingly fewer samples (an indication of computational effort) than simulations based on a global equilibrium as the $Kn \to 0$ limit is approached.

Note on the advection step: when $f^{MB} = f^{MB}(\mathbf{r}, \mathbf{c})$, (2.10) requires the deviational distribution function to obey

$$\frac{\partial f^{d}}{\partial t} + \mathbf{c} \cdot \frac{\partial f^{d}}{\partial \mathbf{r}} = -\mathbf{c} \cdot \frac{\partial f^{MB}}{\partial \mathbf{r}} \qquad (6.11)$$

during the advection substep. In the implementation discussed above, f^{MB} is local to each cell and thus varies in a piecewise constant fashion from one cell to the next. This leads to discontinuities of this function at the boundaries between cells which leads to fluxes of deviational particles across these boundaries. The solution to (6.11) can be written as a superposition of a free molecular advection for the deviational particles (solution of the homogeneous equation) and a correction term accounting for the contribution of the right-hand side. The latter, which can be written analytically [32], can be implemented as a deviational-particle flux term at cell boundaries; more details can be found in [32,50].

One disadvantage of this implementation is that particle generation at cell interfaces becomes cumbersome in high-dimensional problems. An alternative formulation that may alleviate this limitation would consist of a continuously variable f^{MB}, whereby particle generation at cell boundaries will be replaced by volumetric particle generation, which will reduce the complexity associated with the dependence of the number of cell interfaces on dimensionality.

7. Discussion

Multiscale kinetic phenomena lend themselves naturally to algebraic decomposition of the distribution function into a part treated by a particle method and a part that can be treated using an appropriate limiting description, either numerical (§4) or analytical (§§5 and 6). These approaches are very promising, and in some cases have already been used to solve problems that

had been otherwise intractable *without introducing approximations*. We note that ideas similar in spirit have been used [64,65] in the context of modelling plasma dynamics, namely the simulation of gyrokinetic equations using particle-in-cell approaches.

We also note a recently proposed approach for reducing the statistical uncertainty of DSMC simulations by coupling them with a deterministic solver of an extended Euler system of equations through a moment-matching procedure. The rationale behind this formulation, referred to as moment-guided Monte Carlo [66], is that constraining the lower moments of the DSMC solution to match a significantly less noisy solution field injects information into the computation that should reduce the statistical uncertainty of the overall solution. In the first implementation [66], moment matching between the two simulations is achieved by a rather arbitrary additive and multiplicative readjustment of particle velocities in DSMC until the desired mean and variance is achieved. Applications to problems where the distribution function is not known will in general require rigorous methods for matching the moments of the two simulations without affecting the particle dynamics in the DSMC procedure.

Our discussion has mostly focused on the spatial aspects of the multiscale problem, primarily because the majority of recent work was devoted to these challenges. Despite this, developing methods that can span a wide range of temporal scales is equally important and challenging. In the context of kinetic problems, algebraic decomposition (4.1) can be exploited for this purpose: Cheremisin [67] developed a *deterministic* method for solving the Boltzmann equation which uses such a decomposition to remove the stiffness associated with integrating the Boltzmann equation in the limit $Kn \to 0$. For particle methods, time-relaxed [43] integration schemes, briefly described in §4, provide a promising new direction. These schemes remove the stiffness associated with time integration of (2.11) in the $Kn \to 0$ limit by emulating a number of desirable features of implicit integration methods, namely stability for all timestep sizes, and the ability to approach the correct limiting distribution in the limit of large timesteps. Although originally developed for the BGK model and limited to first-order accuracy, higher order schemes for the VHS model have recently been presented [45].

The ultimate challenge, both in the context of spatial and temporal scales, is the extension of algebraic decomposition methods to more complex systems (e.g. dense fluids) where governing equations for the distribution function reside in higher-dimensional spaces. Such systems are more challenging because the governing equation is typically more complex, making it harder to derive explicit expressions for the dynamics of the equilibrium part and the deviation therefrom; moreover, although nearby equilibrium states are easy to identify based on physical considerations, mathematical descriptions of those states will be significantly more complex.

References

1. Hadjiconstantinou NG. 2006 The limits of Navier–Stokes theory and kinetic extensions for describing small-scale gaseous hydrodynamics. *Phys. Fluids* **18**, 111301. (doi:10.1063/1.2393436)
2. Caflisch R, Wang CM, Dimarco G, Cohen B, Dimits A. 2008 A hybrid method for accelerated simulation of Coulomb collisions in a plasma. *Multiscale Model. Simul.* **7**, 865–887. (doi:10.1137/070704939)
3. Chen G. 2005 *Nanoscale energy transfer and conversion*. New York, NY: Oxford University Press.
4. Harrison WA. 1980 *Solid state theory*. New York, NY: Dover.
5. Lundstrom M. 2000 *Fundamentals of carrier transport*, 2nd edn. Cambridge, UK: Cambridge University Press.
6. Dumonteil E, Courau T. 2010 Dominance ratio assessment and Monte Carlo criticality simulations: dealing with high dominance ratio systems. *Nucl. Technol.* **172**, 120–131.

7. Dumonteil E, Diop CM. 2011 Biases and statistical errors in Monte Carlo burnup calculations: an unbiased stochastic scheme to solve Boltzmann/Bateman coupled equations. *Nucl. Sci. Eng.* **167**, 165–170.

8. Manela A, Hadjiconstantinou NG. 2007 On the motion induced in a gas confined in a small-scale gap due to instantaneous boundary heating. *J. Fluid Mech.* **593**, 453–462. (doi:10.1017/S0022112007008658)

9. Manela A, Hadjiconstantinou NG. 2008 Gas motion induced by unsteady boundary heating in a small-scale slab. *Phys. Fluids* **20**, 117104. (doi:10.1063/1.3010759)

10. Bourgat J, Le Tallec P, Tidriri M. 1996 Coupling Boltzmann and Navier–Stokes equations by friction. *J. Comput. Phys.* **127**, 227–245. (doi:10.1006/jcph.1996.0172)

11. Garcia AL, Bell JB, Crutchfield WY, Alder BJ. 1999 Adaptive mesh and algorithm refinement using direct simulation Monte Carlo. *J. Comput. Phys.* **154**, 134–155. (doi:10.1006/jcph.1999.6305)

12. Wijesinghe HS, Hornung R, Garcia AL, Hadjiconstantinou NG. 2004 Three-dimensional hybrid continuum-atomistic simulations for multiscale hydrodynamics. *J. Fluids Eng.* **126**, 768–777. (doi:10.1115/1.1792275)

13. Schwartzentruber TE, Boyd ID. 2007 A modular particle–continuum numerical method for hypersonic non-equilibrium gas flows. *J. Comput. Phys.* **225**, 1159–1174. (doi:10.1016/j.jcp.2007.01.022)

14. Hadjiconstantinou NG. 1999 Hybrid atomistic-continuum formulations and the moving contact-line problem. *J. Comput. Phys.* **154**, 245–265. (doi:10.1006/jcph.1999.6302)

15. Hadjiconstantinou NG. 2005 Discussion of recent developments in hybrid atomistic-continuum methods for multiscale hydrodynamics. *Bull. Pol. Acad. Sci. Tech. Sci.* **53**, 335–342.

16. Wang W, Boyd ID. 2003 Predicting continuum breakdown in hypersonic viscous flows. *Phys. Fluids* **15**, 91–100. (doi:10.1063/1.1524183)

17. Cercignani C. 1988 *The Boltzmann equation and its applications*. New York, NY: Springer. (doi:10.1007/978-1-4612-1039-9)

18. Vincenti WG, Kruger CH. 1965 *Introduction to physical gas dynamics*. New York, NY: Wiley.

19. Chapman S, Cowling TG. 1991 *The mathematical theory of non-uniform gases*. Cambridge, UK: Cambridge University Press.

20. Degond P, Dimarco G, Mieussens L. 2010 A multiscale kinetic-fluid solver with dynamic localization of kinetic effects. *J. Comput. Phys.* **229**, 4907–4933. (doi:10.1016/j.jcp.2010.03.009)

21. Werder T, Walther JH, Koumoutsakos P. 2005 Hybrid atomistic-continuum method for the simulation of dense fluid flows. *J. Comput. Phys.* **205**, 373–390. (doi:10.1016/j.jcp.2004.11.019)

22. Miller RE, Tadmor EB. 2009 A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. *Modelling Simul. Mater. Sci. Eng.* **17**, 053001. (doi:10.1088/0965-0393/17/5/053001)

23. Kevrekidis IG, Gear CW, Hyman JM, Kevrekidis PG, Runborg O, Theodoropoulos C. 2003 Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.* **1**, 715–762.

24. Laso M, Ottinger HC. 1993 Calculation of viscoelastic flow using molecular-models: the CONNFFESSIT approach. *J. Non-Newtonian Fluid Mech.* **47**, 1–20. (doi:10.1016/0377-0257(93)80042-A)

25. Weinan E, Engquist B. 2003 The heterogeneous multiscale method. *Commun. Math. Sci.* **1**, 87–133.

26. Al-Mohssen HA, Hadjiconstantinou NG, Kevrekidis IG. 2007 Acceleration methods for coarse-grained numerical solution of the Boltzmann equation. *J. Fluids Eng.* **129**, 908–912. (doi:10.1115/1.2742725)

27. Kessler DA, Oran ES, Kaplan CR. 2010 Towards the development of a multiscale, multiphysics method for the simulation of rarefied gas flows. *J. Fluid Mech.* **661**, 262–293. (doi:10.1017/S0022112010002934)

28. Baker LL, Hadjiconstantinou NG. 2005 Variance reduction for Monte Carlo solutions of the Boltzmann equation. *Phys. Fluids* **17**, 051703. (doi:10.1063/1.1899210)

29. Bird GA. 1994 *Molecular gas dynamics and the direct simulation of gas flows*. Oxford, UK: Clarendon Press.

30. Joshi AA, Majumdar A. 1993 Transient ballistic and diffusive phonon heat transport in thin films. *J. Appl. Phys.* **74**, 31–39. (doi:10.1063/1.354111)

31. Aoki K, Takata S, Aikawa H, Golse, F. 2001 A rarefied gas flow caused by a discontinuous wall temperature. *Phys. Fluids* **13**, 2645–2661. (doi:10.1063/1.1416500)

32. Homolle TMM, Hadjiconstantinou NG. 2007 A low-variance deviational simulation Monte Carlo for the Boltzmann equation. *J. Comput. Phys.* **226**, 2341–2358. (doi:10.1016/j.jcp.2007.07.006)

33. Wagner W. 1992 A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation. *J. Statist. Phys.* **66**, 1011–1044. (doi:10.1007/BF01055714)

34. Hadjiconstantinou NG. 2000 Analysis of discretization in the direct simulation Monte Carlo. *Phys. Fluids* **12**, 2634–2638. (doi:10.1063/1.1289393)

35. Alexander FJ, Garcia AL, Alder BJ. 1998 Cell size dependence of transport coefficients in stochastic particle algorithms. *Phys. Fluids* **10**, 1540–1542. (doi:10.1063/1.869674)

36. Alexander FJ, Garcia AL. 1997 The direct simulation Monte Carlo method. *J. Comput. Phys.* **11**, 588–593. (doi:10.1063/1.168619)

37. Rader DJ, Gallis MA, Torczynski JR, Wagner W. 2006 Direct simulation Monte Carlo convergence behavior of the hard-sphere-gas thermal conductivity for Fourier heat flow. *Phys. Fluids* **18**, 077102. (doi:10.1063/1.2213640)

38. Caflisch RE, Pareschi L. 2004 Towards a hybrid Monte Carlo method for rarefied gas dynamics. In *Transport in transition regimes*, vol. 15, pp. 57–74. New York, NY: Springer.

39. Baker LL, Hadjiconstantinou NG. 2008 Variance-reduced particle methods for solving the Boltzmann equation. *J. Comput. Theoret. Nanosci.* **5**, 165–174.

40. Caflisch R, Chen H, Luo E, Pareschi L. 2006 A hybrid method that interpolates between DSMC and CFD. AAIA paper 2006–987.

41. Dimarco G, Pareschi L. 2008 Hybrid multiscale methods. II. Kinetic equations. *Multiscale Model. Simul.* **6**, 1169–1197. (doi:10.1137/070680916)

42. Dimarco G, Pareschi L. 2010 A fluid solver independent hybrid method for multiscale kinetic equations. *SIAM J. Sci. Comput.* **32**, 603–634. (doi:10.1137/080730585)

43. Pareschi L, Russo G. 2000 Asymptotic preserving Monte Carlo methods for the Boltzmann equation. *Transp. Theoret. Statist. Phys.* **29**, 415–430. (doi:10.1080/00411450008205882)

44. Wild E. 1951 On Boltzmann's equation in the kinetic theory of gases. *Proc. Cambridge Phil. Soc.* **47**, 602–609. (doi:10.1017/S0305004100026992)

45. Trazzi S, Pareschi L, Wennberg B. 2009 Adaptive and recursive time relaxed Monte Carlo methods for rarefied gas dynamics. *SIAM J. Sci. Comput.* **31**, 1379–1398. (doi:10.1137/07069119X)

46. Sone Y. 2007 *Molecular gas dynamics: theory, techniques, and applications*. Boston, MA: Birkhauser.

47. Cercignani C. 2006 *Slow rarefied flows*. Basel, Switzerland: Birkhauser.

48. Hadjiconstantinou NG, Garcia AL, Bazant MZ, He G. 2003 Statistical error in particle simulations of hydrodynamic phenomena. *J. Comput. Phys.* **187**, 274–297. (doi:10.1016/S0021-9991(03)00099-8)

49. Hammersley JM, Handscomb DC. 1964 *Monte Carlo methods*. New York, NY: Wiley.

50. Radtke GA, Hadjiconstantinou NG. 2009 Variance-reduced particle simulation of the Boltzmann transport equation in the relaxation-time approximation. *Phys. Rev. E* **79**, 056711. (doi:10.1103/PhysRevE.79.056711)

51. Hadjiconstantinou NG, Radtke GA, Baker LL. 2010 On variance reduced simulations of the Boltzmann transport equation for small-scale heat transfer applications. *J. Heat Transfer* **132**, 112401. (doi:10.1115/1.4002028)

52. Radtke GA, Hadjiconstantinou NG, Wagner W. 2011 Low-noise Monte Carlo simulation of the variable hard sphere gas. *Phys. Fluids* **23**, 030606. (doi:10.1063/1.3558887)

53. Wagner W. 2008 Deviational particle Monte Carlo for the Boltzmann equation. *Monte Carlo Methods Appl.* **14**, 191–268. (doi:10.1515/MCMA.2008.010)

54. Takata S, Sugimoto H, Kosuge S. 2007 Gas separation by means of the Knudsen compressor. *Eur. J. Mech. B* **26**, 155–181. (doi:10.1016/j.euromechflu.2006.05.002)

55. Alexeenko AA, Gimelshein SF, Muntz EP, Ketsdever A. 2005 Modeling of thermal transpiration flows for Knudsen compressor optimization. AIAA paper 2005-0963.

56. Peterson RB. 1994 Direct simulation of phonon-mediated heat-transfer in a Debye crystal. *J. Heat Transfer* **116**, 815–822. (doi:10.1115/1.2911452)

57. Mazumder S, Majumdar A. 2001 Monte Carlo study of phonon transport in solid thin films including dispersion and polarization. *J. Heat Transfer* **123**, 749–759. (doi:10.1115/1.1377018)

58. Lacroix D, Joulain K, Lemonnier D. 2005 Monte Carlo transient phonon transport in silicon and germanium at nanoscales. *Phys. Rev. B* **72**, 064305. (doi:10.1103/PhysRevB.72.064305)

59. Jeng MS, Yang R, Song D, Chen G. 2008 Modeling the thermal conductivity and phonon transport in nanoparticle composites using Monte Carlo simulation. *J. Heat Transfer* **130**, 042410. (doi:10.1115/1.2818765)

60. Hao Q, Chen G, Jeng MS. 2009 Frequency-dependent Monte Carlo simulations of phonon transport in two-dimensional porous silicon with aligned pores. *J. Appl. Phys.* **106**, 114321. (doi:10.1063/1.3266169)

61. Péraud J-PM. 2011 Low variance methods for Monte Carlo simulation of phonon transport. MS thesis, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA.

62. Péraud J-PM, Hadjiconstantinou NG. 2011 Efficient simulation of multidimensional phonon transport using energy-based variance-reduced Monte Carlo formulations. *Phys. Rev. B* **84**, 205331. (doi:10.1103/PhysRevB.84.205331)

63. Swartz ET, Pohl RO. 1987 Thermal resistance at interfaces. *Appl. Phys. Lett.* **51**, 2200–2202. (doi:10.1063/1.98939)

64. Denton RE, Kotschenreuther M. 1995 δf Algorithm. *J. Comput. Phys.* **119**, 283–294. (doi:10.1006/jcph.1995.1136)

65. Dimits AM, Lee WW. 1993 Partially linearized algorithms in gyrokinetic particle simulation. *J. Comput. Phys.* **107**, 309–323. (doi:10.1006/jcph.1993.1146)

66. Degond P, Dimarco G, Pareschi L. 2011 The moment-guided Monte Carlo method. *Int. J. Numer. Meth. Fluids* **67**, 189–213. (doi:10.1002/fld.2345)

67. Cheremisin FG. 2000 Solving the Boltzmann equation in the case of passing to the hydrodynamic flow regime. *Dokl. Phys.* **45**, 401–404. (doi:10.1134/1.1310733)