Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling[⋆]

Emil Løvbak[0000−0003−1520−4922], Bert Mortier[0000−0003−1788−7819],
Giovanni Samaey[0000−0001−8433−4523], and
Stefan Vandewalle[0000−0002−8988−2374]

Department of Computer Science, KU Leuven, Belgium
{emil.loevbak, bert.mortier, giovanni.samaey,
stefan.vandewalle}@kuleuven.be

Abstract. In many applications it is necessary to compute the time-dependent distribution of an ensemble of particles subject to transport and collision phenomena. Kinetic equations are PDEs that model such particles in a position-velocity phase space. In the low collisional regime explicit particle-based Monte Carlo methods simulate these high dimensional equations efficiently, but, as the collision rate increases, these methods suffer from severe time-step constraints. In the high collision regime, asymptotic-preserving particle schemes are able to produce stable results. However, this stability comes at the cost of a bias in the computed results. In earlier work, the multilevel Monte Carlo method was used to reduce this bias by combining simulations with large and small time steps. This multilevel scheme, however, still has large variances when correlating fine and coarse simulations, which leads to suboptimal multilevel performance. In this work, we present an improved correlation approach that decreases the variance when bridging the gap from large time steps to time steps of the order of magnitude of the collision rate. We further demonstrate that this reduced variance results in a sharply reduced simulation cost at the expense of a small bias.

Keywords: Multilevel Monte Carlo · Kinetic Equations · Diffusive Scaling · Particle Methods

1 Introduction

In many application domains, one encounters kinetic equations that model particle behavior in a high dimensional position-velocity phase space. One such motivating application for this work is modelling neutral transport in plasma simulations of Tokamak fusion reactors, as is done in the simulation code EIRENE [19].

Often, one is interested in low-dimensional moments of the particle distribution, which are computed as averages over velocity space. The time-scale at which these quantities of interest change is often much slower than that of the particle dynamics, making these models very stiff: a naive simulation requires small time steps to capture the fast dynamics, but long time horizons to capture the evolution of the macroscopic quantities of interest. The exact nature of the macroscopic behavior depends on the problem scaling, which can be hyperbolic or diffusive [4].

We consider a diffusively-scaled, spatially homogeneous kinetic equation modelling a distribution of particles $f(x, v, t)$ as a function of space $x \in \mathcal{D}_x \subset \mathbb{R}^d$, velocity $v \in \mathcal{D}_v \subset \mathbb{R}^d$ and time $t \in \mathbb{R}^+$. The scaling is captured the parameter ε, which represents the particle mean free path, i.e., the expected time between collisions. Simultaneously re-scaling time by ε, gives a model equation

$$\partial_t f(x, v, t) + \frac{v}{\varepsilon} \nabla_x f(x, v, t) = \frac{1}{\varepsilon^2} \left(\mathcal{M}(v) \rho(x, t) - f(x, v, t) \right), \tag{1}$$

$$\text{with} \quad \rho(x, t) = \int_{\mathcal{D}_v} f(x, v, t) \, \mathrm{d}v \quad \text{the particle density.}$$

The right-hand side of (1) consists of the BGK operator [1], which linearly drives the velocity to a steady-state distribution $\mathcal{M}(v)$ with characteristic velocity $\tilde{v}$.

Individual particles in the distribution $f(x, v, t)$ follow a velocity-jump process, i.e., they travel in a straight line with a fixed velocity for an exponentially distributed time interval $\mathcal{E}(1/\varepsilon)$, at which point a collision is simulated by re-sampling their velocity from the distribution $\mathcal{M}(v)$. Taking the diffusion limit $\varepsilon \to 0$ in (1) drives the collision rate to infinity, while simultaneously increasing the velocity. In this limit, it can be shown that the particle density resulting from (1) converges to the heat equation [11],

$$\partial_t \rho(x, t) = \nabla_x^2 \rho(x, t). \tag{2}$$

Equation (1) can be simulated using a wide selection of methods, which broadly fall into two categories. Deterministic methods solve (1) for $f(x, v, t)$ on a grid over the position-velocity phase space (using, for instance, finite differences or finite volumes). This approach quickly becomes computationally infeasible when the dimension grows, since a grid must be formed over the $2d$-dimensional domain, $\mathcal{D}_x \times \mathcal{D}_v$. Stochastic methods, on the other hand, perform simulations of individual particle trajectories, with each trajectory sampling the probability distribution $f(x, v, t)$. These methods do not suffer from the dimensionality of the phase space, but introduce a statistical error in the computed solution. When using explicit time steps, both approaches become prohibitively expensive for small values of ε due to the time step restriction.

To avoid the issues when ε becomes small, one can use asymptotic-preserving schemes [9]. Such methods preserve the macroscopic limit equation, in our case given by (2), as ε tends to zero, while avoiding time step constraints. Many such methods have been developed in literature for deterministic methods in the

diffusive limit. For more information, we refer to a recent review paper [4] and the references therein. In the particle setting, only a few asymptotic-preserving methods exist, mostly in the hyperbolic scaling [3, 6, 7, 16–18]. In the diffusive scaling, we are only aware of three works [2, 5, 14]. We use the approach taken in [5], which uses operator splitting to produce an unconditionally stable fixed time step particle method. This stability comes at the cost of an extra bias in the model, proportional to the size of the time step.

In [13] the multilevel Monte Carlo method [8] was applied to the scheme presented in [5]. Multilevel Monte Carlo methods first compute an initial estimate, using a large number of samples with a large time step (low cost per sample). This estimate has low variance, but is expected to have a large bias. Next, the bias of the initial estimate is reduced by performing corrections using a hierarchy of simulations with increasingly smaller time steps. Under correct conditions, far fewer samples with small time steps are needed, compared to a direct Monte Carlo simulation with the smallest time step. This approach results in a reduced computational cost for a given accuracy. One of these conditions is the ability to generate samples with different time step sizes which are correlated, i.e., they approximate the same continuous particle trajectory. This paper presents a strategy for improving the correlation developed in [13] and thus further reducing the computational cost of particle based simulations of (1).

The remainder of this paper is structured as follows. In Section 2 we present a naive particle based Monte Carlo scheme for simulating the kinetic equation (1) and show why this approach fails. In Section 3, we show how combining the asymptotic-preserving scheme from [5] and the multilevel Monte Carlo method can be used to combat these issues. In Section 4, we show where the correlation of the existing method is sub-optimal and introduce our improved correlation. In Section 5, we present experimental results using the new correlation, demonstrating computational speed-up. In Section 6, we summarize our results and list further challenges.

2 Monte Carlo Simulation near the Diffusive Limit

For the sake of exposition, we limit this work to one spatial dimension, i.e., $d = 1$ but, our approach can be applied straightforwardly in more dimensions. Equation (1) now becomes

$$\partial_t f(x, v, t) + \frac{v}{\varepsilon} \partial_x f(x, v, t) = \frac{1}{\varepsilon^2} \left(\mathcal{M}(v) \rho(x, t) - f(x, v, t) \right). \qquad (3)$$

We also restrict this paper to simulations with two discrete velocities, i.e., $\mathcal{M}(v) \equiv \frac{1}{2} \left(\delta_{v, -1} + \delta_{v, 1} \right)$, with δ the Kronecker delta function.

We focus on computing a quantity of interest $Y(t^*)$, which takes the form of an integral over a function $F(x, v)$ of the particle position $X(t)$ and velocity $V(t)$ at time $t = t^*$, with respect to the measure $f(x, v, t) \, \mathrm{d}x \, \mathrm{d}v$, i.e.,

$$Y(t^*) = \mathbb{E} \left[F \left(X(t^*), V(t^*) \right) \right] = \int_{\mathcal{D}_v} \int_{\mathcal{D}_x} F(x, v) f(x, v, t^*) \, \mathrm{d}x \, \mathrm{d}v.$$

Equation (3) can be simulated using a particle scheme with a fixed time step size Δt. Each particle p has a state $\left(X_{p,\Delta t}^n, V_{p,\Delta t}^n\right)$ in the position-velocity phase space at each time step n, with $X_{p,\Delta t}^n \approx X_p(n\Delta t)$ and $V_{p,\Delta t}^n \approx V_p(n\Delta t)$. We represent the distribution $f(x,v,t)$ of particles at time $n\Delta t$ by an ensemble of P particles, with indices $p \in \{1,\ldots,P\}$,

$$\left\{\left(X_{p,\Delta t}^n, V_{p,\Delta t}^n\right)\right\}_{p=1}^P. \tag{4}$$

A classical Monte Carlo estimator $\hat{Y}(t^*)$ for $Y(t^*)$ averages over such an ensemble

$$\hat{Y}(t^*) = \frac{1}{P}\sum_{p=1}^P F\left(X_{p,\Delta t}^N, V_{p,\Delta t}^N\right), \quad t^* = N\Delta t. \tag{5}$$

To generate an ensemble (4) we perform forward time stepping based on operator splitting, which is first order in the time step Δt [15]. For (3), operator splitting results in two actions for each time step:

1. **Transport step.** Each particle's position is updated based on its velocity

$$X_{p,\Delta t}^{n+1} = X_{p,\Delta t}^n + \Delta t V_{p,\Delta t}^n. \tag{6}$$

2. **Collision step.** Between transport steps, each particle's velocity is either left unchanged (no collision) or re-sampled from $\mathcal{M}(v)$ (collision), i.e.,

$$V_{p,\Delta t}^{n+1} = \begin{cases} V_{p,\Delta t}^{n,\prime} \sim \mathcal{M}(v), & \text{with collision probability } p_{c,\Delta t} = \Delta t/\varepsilon^2, \\ V_{p,\Delta t}^n, & \text{otherwise.} \end{cases} \tag{7}$$

Scheme (6)–(7) has a severe time step restriction $\Delta t = \mathcal{O}(\varepsilon^2)$ when approaching the limit $\varepsilon \to 0$. This time step restriction results in unacceptably high simulation costs, despite the well-defined limit (2) [5]. Given that the variance of the estimator $\hat{Y}(t^*)$ scales with $\frac{1}{P}$, a high individual simulation cost will severely constrain the precision with which we can approximate $Y(t^*)$.

3   Multilevel Monte Carlo using AP Schemes

In the previous section, we demonstrated that a time step $\Delta t = \mathcal{O}(\varepsilon^2)$ is needed to resolve the collision dynamics of the kinetic equation (3). As ε becomes small, this causes a naive Monte Carlo estimator to be computationally infeasible. In this section, we first present an asymptotic-preserving Monte Carlo scheme which maintains stability for large time steps by reducing the contribution of collision dynamics for larger time steps. We will then present a multilevel Monte Carlo estimator, using this scheme, to combine simulations with different time step sizes, reducing the overall computational cost of the estimator $\hat{Y}(t^*)$.

3.1   Asymptotic-preserving Monte Carlo scheme

We use the asymptotic-preserving scheme from [5] as an alternative to (6)–(7). This asymptotic-preserving scheme has no time step constraints, so ε and Δt can be chosen independently. The scheme simulates a modified version of (3):

$$\partial_t f + \frac{\varepsilon v}{\varepsilon^2 + \Delta t} \partial_x f = \frac{\Delta t}{\varepsilon^2 + \Delta t} \partial_{xx} f + \frac{1}{\varepsilon^2 + \Delta t} \left(\mathcal{M}_{\Delta t}(v) \rho - f \right). \tag{8}$$

In (8), we have omitted the space, velocity and time dependency of $f(x, v, t)$ and $\rho(x, t)$, for conciseness. Note that the steady-state distribution $\mathcal{M}_{\Delta t}(v)$ now has a time step dependent characteristic velocity $\tilde{v}_{\Delta t} = \dfrac{\varepsilon}{\varepsilon^2 + \Delta t}$ and that the coefficients of (8) explicitly contain the simulation time step Δt.

Equation (8) has the following properties [5]:

1. When $\varepsilon \to 0$, (8) converges to the heat equation (2).
2. When $\Delta t \to 0$, (8) converges to the original kinetic equation (3) $\mathcal{O}(\Delta t)$.
3. When $\Delta t \to \infty$, (8) converges to the heat equation (2) with a rate $\mathcal{O}\left(\frac{1}{\Delta t}\right)$.

The first property states that (8) has the same asymptotic limit in ε as (3). The second and third properties provide us with an intuitive understanding of (8). We can interpret the modified equation as a combination of the heat equation (2) and the original kinetic equation (3), where the two contributions are weighted in function of Δt. For large time steps the heat equation dominates over the kinetic equation. At the particle level, the heat equation corresponds with Brownian motion, which has no time constraints, hence the stability of (8).

Particle trajectories are now simulated as follows:

1. **Transport-diffusion step.** The position of the particle is updated based on its velocity and a Brownian increment

$$X_{p,\Delta t}^{n+1} = X_{p,\Delta t}^n + V_{p,\Delta t}^n \Delta t + \sqrt{2\Delta t}\sqrt{D_{\Delta t}}\xi_p^n, \tag{9}$$

   in which we generate $\xi_p^n \sim \mathcal{N}(0,1)$ and introduce a Δt-dependent velocity $V_{p,\Delta t}^n \sim \mathcal{M}_{\Delta t}(v)$ and a diffusion coefficient $D_{\Delta t} = \frac{\Delta t}{\varepsilon^2 + \Delta t}$.

2. **Collision step.** During collisions, each particle's velocity is updated as:

$$V_{p,\Delta t}^{n+1} = \begin{cases} V_{p,\Delta t}^{n,\prime} \sim \mathcal{M}_{\Delta t}(v), & \text{with probability } p_{c,\Delta t} = \dfrac{\Delta t}{\varepsilon^2 + \Delta t}, \\ V_{p,\Delta t}^n, & \text{otherwise.} \end{cases} \tag{10}$$

   The considered time step dependent distribution $\mathcal{M}_{\Delta t}(v)$ can be decomposed as a characteristic velocity multiplied with a time step independent velocity distribution $\bar{\mathcal{M}}(\bar{v})$ with unit variance and expected value zero

$$\mathcal{M}_{\Delta t}(v) = \tilde{v}_{\Delta t}\bar{\mathcal{M}}(\bar{v}). \tag{11}$$

The scheme (9)–(10) stably generates trajectories for large Δt. However, these trajectories are biased with $\mathcal{O}(\Delta t)$ compared to those generated by (6)–(7).

3.2   Multilevel Monte Carlo

The key idea behind the multilevel Monte Carlo method (MLMC) [8] is to combine simulations with different time step sizes to simulate many trajectories (low variance), while also simulating accurate trajectories (low simulation bias).

First, a coarse Monte Carlo estimator $\hat{Y}_0(t^*)$ with a time step Δt_0 is used

$$\hat{Y}_0(t^*) = \frac{1}{P_0} \sum_{p=1}^{P_0} F\left(X_{p,\Delta t_0}^{N_0}, V_{p,\Delta t_0}^{N_0}\right), \quad t^* = N_0 \Delta t_0. \tag{12}$$

Estimator (12) is based on a large number of trajectories P_0, but has a low computational cost as few time steps N_0 are required to reach time t^*.

Next, $\hat{Y}_0(t^*)$ is refined upon by a sequence of L difference estimators at levels $\ell = 1, \ldots, L$. Each difference estimator uses an ensemble of P_ℓ particle pairs

$$\hat{Y}_\ell(t^*) = \frac{1}{P_\ell} \sum_{p=1}^{P_\ell} \left(F\left(X_{p,\Delta t_\ell}^{N_\ell}, V_{p,\Delta t_\ell}^{N_\ell}\right) - F\left(X_{p,\Delta t_{\ell-1}}^{N_{\ell-1}}, V_{p,\Delta t_{\ell-1}}^{N_{\ell-1}}\right) \right), \quad t^* = N_\ell \Delta t_\ell. \tag{13}$$

Each correlated particle pair consists a particle with a fine time step Δt_ℓ and a particle with a coarse time step $\Delta t_{\ell-1} = M\Delta t_\ell$, with M a positive integer. The particles in each pair undergo correlated simulations, which intuitively can be understood as making both particles follow the same qualitative trajectory for two different simulation accuracies. One can interpret the difference estimator (13) as using the fine simulation to estimate the bias in the coarse simulation.

Given a sequence of levels $\ell \in \{0, \ldots, L\}$, with decreasing step sizes, and the corresponding estimators given by (12)–(13), the multilevel Monte Carlo estimator for the quantity of interest $Y(t^*)$ is computed by the telescopic sum

$$\hat{Y}(t^*) = \sum_{\ell=0}^{L} \hat{Y}_\ell(t^*). \tag{14}$$

It is clear that the expected value of the estimator (14) is the same as that of (5), with the finest time step $\Delta t = \Delta t_L$. Given a sufficiently fast reduction in the number of simulated (pairs of) trajectories P_ℓ as ℓ increases, it can be shown that the multilevel Monte Carlo estimator requires a lower computational cost than a classical Monte Carlo estimator to achieve the same mean square error.

3.3   Term-by-term correlation approach

The differences in (13) will only have low variance if the simulated paths up to $X_{p,\Delta t_\ell}^{N_\ell}$ and $X_{\Delta t_{p,\ell-1}}^{N_{\ell-1}}$, with time steps related by $\Delta t_{\ell-1} = M\Delta t_\ell$ are correlated. To discuss these correlated pairs of trajectories, we define a sub-step index $m \in \{1, \ldots, M\}$, i.e., $X_{p,\Delta t_\ell}^{n,m} \approx X_p(n\Delta t_{\ell-1} + m\Delta t_\ell) \equiv X_p((nM+m)\Delta t_\ell)$. In order to

span a time interval of size $\Delta t_{\ell-1}$, the coarse simulation requires a single time step of size $\Delta t_{\ell-1}$ while the fine simulation requires M time steps of size Δt_ℓ:

$$\begin{cases} X^{n+1}_{p,\Delta t_{\ell-1}} = X^n_{p,\Delta t_{\ell-1}} + \Delta t_{\ell-1} V^n_{p,\Delta t_{\ell-1}} + \sqrt{2\Delta t_{\ell-1}}\sqrt{D_{\Delta t_{\ell-1}}}\xi^n_{p,\ell-1} \\[2mm] X^{n+1,0}_{p,\Delta t_\ell} = X^{n,0}_{p,\Delta t_\ell} + \sum_{m=1}^{M}\left(\Delta t_\ell V^{n,m}_{p,\Delta t_\ell} + \sqrt{2\Delta t_\ell}\sqrt{D_{\Delta t_\ell}}\xi^{n,m}_{p,\ell}\right) \end{cases},$$

with $\xi^n_{p,\ell-1}, \xi^{n,m}_{p,\ell} \sim \mathcal{N}(0,1)$, $V^n_{p,\Delta t_{\ell-1}} \sim \mathcal{M}_{\Delta t_{\ell-1}}(v)$ and $V^{n,m}_{p,\Delta t_\ell} \sim \mathcal{M}_{\Delta t_\ell}(v)$.

There are two sources of stochastic behavior in the asymptotic-preserving scheme (9)–(10). On the one hand, a new Brownian increment ξ^n_p is generated at each time step. On the other hand, each time step has a finite collision probability. Collisions produce a new particle velocity $V^{n+1}_{p,\Delta t}$ for the next time step.

To correlate these simulations we first perform M independent time steps of the fine simulation. We then combine the M sets of random numbers in the fine path into one set of random numbers which can be used in a single time step for the coarse simulation. If the fine simulation random numbers are combined in such a way that the resulting coarse simulation random numbers are distributed as if they had been generated independently, then the coarse simulation statistics will be preserved, while also being dependent on the fine simulation.

In [13], an approach was developed for independently correlating Brownian increments and collision phenomena, which we briefly describe here. For a full derivation and an analysis of the correlation we refer to [12, 13].

Correlating Brownian increments In each fine sub-step $m \in \{1,\dots,M\}$ a Brownian increment $\xi^{n,m}_{p,\ell} \sim \mathcal{N}(0,1)$ is generated. The sum of these M increments is distributed as $\mathcal{N}(0,M)$. This means that a Brownian increment for the corresponding coarse simulation step can be generated as

$$\xi^n_{p,\ell-1} = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\xi^{n,m}_{p,\ell} \sim \mathcal{N}(0,1). \tag{15}$$

Correlating collisions Implementing (10) requires simulating a collision in a fine sub-step $m \in \{1,\dots,M\}$, with the collision probability $p_{c,\Delta t_\ell}$. In practice this is achieved by drawing a uniformly distributed random number $u^{n,m}_{p,\ell} \sim \mathcal{U}([0,1])$ and performing a collision if this number is larger than the probability $p_{nc,\Delta t_\ell} = 1 - p_{c,\Delta t_\ell}$ that no collision has occurred in the the time step

$$u^{n,m}_{p,\ell} \geq p_{nc,\Delta t_\ell} = \frac{\varepsilon^2}{\varepsilon^2 + \Delta t_\ell}. \tag{16}$$

At least one collision has taken place in the fine simulation if the largest of the generated $u^{n,m}_{p,\ell}$, $m \in \{1,\dots,M\}$, satisfies (16), i.e.,

$$u^{n,\max}_{p,\ell} = \max_m u^{n,m}_{p,\ell} \geq p_{nc,\Delta t_\ell}. \tag{17}$$

When (17) is satisfied, the probability of a collision taking place in the correlated coarse simulation should be large. This is the case if we compare

$$u_{p,\ell-1}^n = \left(u_{p,\ell}^{n,\mathrm{max}}\right)^M \sim \mathcal{U}([0,1]), \tag{18}$$

with $p_{nc,\Delta t_{\ell-1}}$ in the coarse simulation.

If a collision is performed in the simulation at level $\ell-1$ using (18), then the coarse simulation velocity should be correlated with that of the fine simulation, at the end of the time interval. We consider the decomposition of $\mathcal{M}(v)$ given in (11). In those of the M fine sub-steps that contain a collision, we draw a time step independent velocity $\bar{v}_{p,\ell}^{n,m}$ to generate $V_{p,\ell}^{n,m}$. Given that the $\bar{v}_{p,\ell}^{n,m}$ are i.i.d., we can select one freely to use as $\bar{v}_{p,\ell-1}^n$. To maximize the correlation of the velocities at the end of the time interval, we take the last generated $\bar{v}_{p,\ell}^{n,m}$, i.e.,

$$\bar{v}_{p,\ell-1}^n = \bar{v}_{p,\ell}^{n,i}, \quad i = \operatorname*{arg\,max}_{1 \le m \le M} \left(m \middle| u_{p,\ell}^{n,m} \ge p_{nc,\Delta t_\ell}\right). \tag{19}$$

4 Improved Coupling of Particle Trajectories

4.1 Limitations of term-by-term correlation

In [12], it was demonstrated that a multilevel Monte Carlo scheme, based on the correlation in Section 3.3 achieves an asymptotic speed-up $\mathcal{O}\left(E^{-2}\log^2(E)\right)$ in the root mean square error bound E for the following sequence of levels:

1. At level 0, generate an initial estimate of $\hat{Y}(t^*)$ by simulating with $\Delta t_0 = t^*$.
2. At level 1, perform correlated simulations to t^* using $\Delta t_0 = t^*$ and $\Delta t_1 = \varepsilon^2$.
3. Continue to generate a geometric sequence of levels $\Delta t_l = \varepsilon^2 M^{1-l}$ for $l > 1$ until an acceptably low bias has been achieved.

This result only holds when $\Delta t_\ell \ll \varepsilon^2$, however, meaning that the scheme only significantly reduces computational cost when calculating very accurate results.

To demonstrate this phenomenon, we present a multilevel simulation with a two-speed velocity distribution using the scheme from Section 3.3 to estimate the squared particle displacement $F(x,v) = x^2$. The results for $\varepsilon = 0.1$, a RMSE $E = 0.1$, $f(x,v,0) \equiv \delta_{x,0}\frac{1}{2}\left(\delta_{v,-1} + \delta_{v,1}\right)$, and $t^* = 0.5$ are shown in Table 1. In this table, we list the fine time step size Δt_ℓ, number of samples P_ℓ, expected value $\mathbb{E}\left[F_\ell - F_{\ell-1}\right]$ and variance $\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$ of the differences of simulations, estimated variance of the estimator $\mathbb{V}[\hat{Y}_\ell]$, cost per sample C_ℓ and level cost $P_\ell C_\ell$. We define $F_{-1} \equiv 0$ and define C_ℓ relative to a simulation with $\Delta t = \varepsilon^2$.

Looking at the values of $\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$ in Table 1, it is clear that the variance at level 1 is almost as large as that at level 0. Comparing the values of $\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$ with those of $\mathbb{E}\left[F_\ell - F_{\ell-1}\right]$, we see, contrastingly, that the expected value of the estimator at level 1 is an order of magnitude smaller than that at level 0. This indicates that the lack of decreasing variance from level 0 to level 1 is due to insufficient correlation between the trajectories used in the

Table 1. Computing $\hat{Y}(t^*)$ for $\varepsilon = 0.1$ and $E = 0.1$.

Level	Δt_ℓ	P_ℓ	$\mathbb{V}\left[\hat{F}_\ell\right]$	$\mathbb{E}\left[\hat{F}_\ell - F_{\ell-1}\right]$	$\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$	$\mathbb{V}[\hat{Y}_\ell]$	C_ℓ	$P_\ell C_\ell$
0	5.00×10^{-1}	17 716	2.02	1.01×10^{0}	2.02×10^{0}	1.14×10^{-4}	0.02	354
1	1.00×10^{-2}	2 157	1.43	-1.17×10^{-1}	1.48×10^{0}	6.85×10^{-4}	1.02	2 200
2	5.00×10^{-3}	602	1.49	-1.36×10^{-2}	3.71×10^{-1}	6.16×10^{-4}	3	1 806
3	2.50×10^{-3}	489	1.65	3.55×10^{-2}	4.10×10^{-1}	8.39×10^{-4}	6	2 934
4	1.25×10^{-3}	345	1.44	-3.94×10^{-2}	4.43×10^{-1}	1.29×10^{-3}	12	4 140
5	6.25×10^{-4}	185	1.61	3.88×10^{-2}	1.84×10^{-1}	9.94×10^{-4}	24	4 440
6	3.13×10^{-4}	50	2.87	-5.41×10^{-2}	1.31×10^{-1}	2.62×10^{-3}	48	2 400
$\sum$				8.60×10^{-1}		7.16×10^{-3}		18 274

estimator at level 1. As a result a large number of samples P_1 are needed with cost C_1, making the total cost $P_1 C_1$ of level 1 much higher than that of level 0.

The reason for this poor correlation can be found in the intuitive interpretation given to equation (8) in Section 3.1. For large time steps, the diffusion term becomes dominant over the transport and collision effects, whereas the inverse is true for smaller time steps. At level 1, the coarse diffusive simulation is being generated from a simulation with a $\Delta t = \varepsilon^2$, which contains a significant transport component. The correlation approach in Section 3.3 correlates the Brownian increments independently from the collisions and resulting velocity changes. Hence the transport-collision effects from the fine simulation, which form a significant part of the particle behavior, are ignored in the coarse simulation.

4.2 Improved correlation approach

In this section, we present an improved correlation approach for particles simulating the scheme (9)–(10). This scheme will let the velocities generated in the fine simulation $\bar{v}_{p,1}^{n,m}$ influence the Brownian increments $\xi_{p,0}^{n,m}$ in the correlated coarse simulation, avoiding the correlation issues described in Section 4.1.

We start from [10], which shows how one can simulate Brownian motion in the weak sense, using an approximate weak Euler scheme. The lower order moments of the approximate Brownian increments in this scheme must be bounded by

$$\left|\mathbb{E}\left[\xi_{p,\ell-1}^n\right]\right| + \left|\mathbb{E}\left[\left(\xi_{p,\ell-1}^n\right)^3\right]\right| + \left|\mathbb{E}\left[\left(\xi_{p,\ell-1}^n\right)^2\right] - 1\right| \leq K \Delta t_{\ell-1}, \qquad (20)$$

for some constant K. If this condition is satisfied, the same weak convergence is achieved as the classical Euler-Maruyama scheme. The new, improved, correlation approach generates $\xi_{p,0}^n$ as a weighted sum of a contribution $\xi_{p,\ell-1,W}^n$ from the fine simulation diffusion and a contribution $\xi_{p,\ell-1,T}^n$ from the fine simulation re-scaled velocities $\bar{v}_{p,\ell}^{n,m}$

$$\xi_{p,\ell-1}^n = \sqrt{\theta_\ell}\, \xi_{p,\ell-1,W}^n + \sqrt{1-\theta_\ell}\, \xi_{p,\ell-1,T}^n, \qquad (21)$$

with $\theta_\ell \in [0,1]$. If both $\xi_{p,\ell-1,W}^n$ and $\xi_{p,\ell-1,T}^n$ have mean zero and variance one, the same holds for $\xi_{p,\ell-1}^n$. This means that $K = 0$ in (20). The correlation of the

transport increments (18)–(19) is left unaltered, to maintain the correlation of the simulation velocities. We now discuss each value in (21):

Diffusive contribution $\xi^n_{p,\ell-1,W}$ The coupling of Brownian increments is done in a way identical to (15) from the correlation approach in Section 3.3:

$$\xi^n_{p,\ell-1,W} = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \xi^{n,m}_{p,\ell}. \tag{22}$$

Transport contribution $\xi^n_{p,\ell-1,T}$ We can generate a value $\xi^n_{p,\ell-1,T}$ with expected value zero and unit variance from the $\bar{v}^{n,m}_{p,\ell}$ as

$$\xi^n_{p,\ell-1,T} = \left(\mathbb{V}\left[\sum_{m=1}^{M} \bar{v}^{n,m}_{p,\ell} \right] \right)^{-\frac{1}{2}} \sum_{m=1}^{M} \bar{v}^{n,m}_{p,\ell}.$$

As subsequent $\bar{v}^{n,m}_{p,\ell}$ are correlated, computing this variance is slightly more involved than in the case of $\xi^n_{p,\ell-1,W}$, but still straightforward

$$\mathbb{V}\left[\sum_{m=1}^{M} \bar{v}^{n,m}_{p,\ell} \right] = \sum_{m=1}^{M} \mathbb{V}\left[\bar{v}^{n,m}_{p,\ell} \right] + 2 \sum_{m=1}^{M-1} \sum_{m'=m+1}^{M} \mathrm{Cov}\left(\bar{v}^{n,m}_{p,\ell}, \bar{v}^{n,m'}_{p,\ell} \right)$$

$$= M + 2 \sum_{\Delta m=1}^{M} (M - \Delta m) \left(p_{nc,\Delta t_\ell} \right)^{\Delta m}$$

$$= M + 2 \frac{p_{nc,\Delta t_\ell} \left((p_{nc,\Delta t_\ell})^M - M(p_{nc,\Delta t_\ell} - 1) - 1 \right)}{(p_{nc,\Delta t_\ell} - 1)^2}$$

$$= M + 2 \frac{p_{nc,\Delta t_\ell} \left((p_{nc,\Delta t_\ell})^M + M p_{c,\Delta t_\ell} - 1 \right)}{(p_{c,\Delta t_\ell})^2}, \tag{23}$$

where $p_{c,\Delta t_\ell}$ is the collision probability as defined in (10) and $p_{nc,\Delta t_\ell} = 1 - p_{c,\Delta t_\ell}$. We want to improve the correlation of level 1 where a large time step Δt_0 is correlated with $\Delta t_1 = \varepsilon^2$. In that case $p_{nc,\Delta t_1} = p_{c,\Delta t_1} = \frac{1}{2}$ so (23) simplifies to

$$\mathbb{V}\left[\sum_{m=1}^{M} \bar{v}^{n,m}_{p,1} \right] = 3M + 2^{2-M} - 4 \approx 3M - 4, \tag{24}$$

where the approximation quickly becomes more accurate for larger values of M.

Contribution weight θ_ℓ What remains is to choose the weight between the two contributions from the simulation with Δt_ℓ to the diffusion in the simulation with $\Delta t_{\ell-1}$. We aim to pick θ_ℓ so that (22) has an equally large contribution when simulating over a time interval of size $\Delta t_{\ell-1}$ at both level ℓ and $\ell - 1$.

At level ℓ the variance contribution of (22) relative to the variance of the total simulation spanning $\Delta t_{\ell-1}$ with M steps is given by

$$\frac{\mathbb{V}\left[\sum_{m=1}^{M}\sqrt{2\Delta t_\ell}\sqrt{D_\ell}\xi_{p,\ell}^{n,m}\right]}{\mathbb{V}\left[\sum_{m=1}^{M}\sqrt{2\Delta t_\ell}\sqrt{D_\ell}\xi_{p,\ell}^{n,m}+\Delta t_\ell\tilde{v}_\ell\bar{v}_{p,\ell}^{n,m}\right]}. \tag{25}$$

The contribution of (22) when spanning the same time interval of size $\Delta t_{\ell-1}$ with a single step is given by

$$\frac{\mathbb{V}\left[\sqrt{2\Delta t_{\ell-1}}\sqrt{D_{\ell-1}}\sqrt{\theta_\ell}\xi_{p,\ell-1,W}^{n}\right]}{\mathbb{V}\left[\sqrt{2\Delta t_{\ell-1}}\sqrt{D_{\ell-1}}\xi_{p,\ell-1}^{n}+\Delta t_{\ell-1}\tilde{v}_{\ell-1}\bar{v}_{p,\ell-1}^{n}\right]}. \tag{26}$$

Equating (25) and (26) and solving for θ_ℓ gives

$$\theta_\ell=\frac{\mathbb{V}\left[\sum_{m=1}^{M}\sqrt{2\Delta t_\ell}\sqrt{D_\ell}\xi_{p,\ell}^{n,m}\right]\mathbb{V}\left[\sqrt{2\Delta t_{\ell-1}}\sqrt{D_{\ell-1}}\xi_{p,\ell-1}^{n}+\Delta t_{\ell-1}\tilde{v}_{\ell-1}\bar{v}_{p,\ell-1}^{n}\right]}{\mathbb{V}\left[\sqrt{2\Delta t_{\ell-1}}\sqrt{D_{\ell-1}}\xi_{p,\ell-1,W}^{n}\right]\mathbb{V}\left[\sum_{m=1}^{M}\sqrt{2\Delta t_\ell}\sqrt{D_\ell}\xi_{p,\ell}^{n,m}+\Delta t_\ell\tilde{v}_\ell\bar{v}_{p,\ell}^{n,m}\right]}.$$

Given (23) and the unit variance of $\xi_{p,\ell-1,W}^{n},\xi_{p,\ell-1}^{n},\xi_{p,\ell}^{n,m}$ and $\bar{v}_{p,\ell-1}^{n}$, we get

$$\theta_\ell=\frac{D_\ell\left(2\Delta t_{\ell-1}D_{\ell-1}+\Delta t_{\ell-1}^2\tilde{v}_{\ell-1}^2\right)}{D_{\ell-1}\left(2M\Delta t_\ell D_\ell+\Delta t_\ell^2\tilde{v}_\ell^2\left(M+2\dfrac{p_{nc,\Delta t_\ell}(p_{nc,\Delta t_\ell}^M+Mp_{c,\Delta t_\ell}-1)}{p_{c,\Delta t_\ell}^2}\right)\right)}. \tag{27}$$

At level 1, a very large time step Δt_0 is correlated with $\Delta t_1=\varepsilon^2$. Here, we can use approximations $D_0\approx 1,\tilde{v}_0\approx\frac{\varepsilon}{\Delta t_0},p_{nc,\Delta t_0}\approx\frac{\varepsilon^2}{\Delta t_0}$ and (24) to simplify (27) to

$$\theta_1\approx\frac{4M+1}{7M-4}.$$

It is important to note that (21) does not produce true Brownian increments, in contrast to the independent simulation at level 0. On the one hand, the distribution of (21) is not Gaussian. On the other hand, subsequent coarse time steps will be correlated. Both issues become more pronounced for higher levels and for that reason we only apply this approach in level 1. In the next section, numerical experiments asses whether this mismatch produces biased results.

5   Experimental Results

5.1   Application to a low-precision case

We re-do the simulation in Table 1, using the new correlation (21) at level 1. Comparing Tables 1–2 we see that the variance $\mathbb{V}\left[F_1-F_0\right]$ reduces by a factor of 30. This reduced variance causes the overall computational cost of the simulation to decrease by a factor of 3.9. Given the low tolerance $E=0.1$, it is not possible to observe whether the improved correlation produces biased results. For this, more accurate simulations are needed.

Table 2. Repeating the simulation in Table 1 with the new correlation at level 1.

Level	Δt_ℓ	P_ℓ	$\mathbb{V}\left[\hat{F}_\ell\right]$	$\mathbb{E}\left[\hat{F}_\ell - F_{\ell-1}\right]$	$\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$	$\mathbb{V}[\hat{Y}_\ell]$	C_ℓ	$P_\ell C_\ell$
0	5.00×10^{-1}	8 123	1.88	9.76×10^{-1}	1.88×10^{0}	2.31×10^{-4}	0.02	162
1	1.00×10^{-2}	189	1.74	-1.44×10^{-1}	4.89×10^{-2}	2.59×10^{-4}	1.02	193
2	5.00×10^{-3}	315	1.76	-1.74×10^{-2}	4.30×10^{-1}	1.36×10^{-3}	3	945
3	2.50×10^{-3}	152	1.28	-5.32×10^{-2}	2.71×10^{-1}	1.78×10^{-3}	6	912
4	1.25×10^{-3}	102	3.16	6.78×10^{-2}	1.58×10^{-1}	1.55×10^{-3}	12	1 224
5	6.25×10^{-4}	50	1.08	3.61×10^{-2}	1.76×10^{-1}	3.52×10^{-3}	24	1 200
$\sum$				8.65×10^{-1}		8.70×10^{-3}		4 636

5.2   Application to a high-precision case

In Table 3, we repeat the simulations in Tables 1 and 2 for $E = 0.01$. Despite the variance decrease at level 1, both simulation costs are of the same order of magnitude. This is due to the high cost of levels $\ell > 1$, reducing the impact of the improvement at level 1. At first sight, this seems a major setback, however the penultimate column of Table 3 shows that $\mathbb{V}\left[F_\ell - F_{\ell-1}\right] > \mathbb{V}\left[F_1 - F_0\right]$ for many of these levels. This indicates that we can skip levels by correlating Δt_0 with $\Delta t_1 \ll \varepsilon^2$ at level 1. In Table 4, we do this with $\Delta t_1 = \frac{\varepsilon^2}{128}$. For a fair comparison, we also apply the same sequence of levels to the term-by-term correlation.

Table 3. Computing $\hat{Y}(t^*)$ for $\varepsilon = 0.1$ and $E = 0.01$.

Level	Δt_ℓ	Term-by-term correlation				Improved correlation			
		P_ℓ	$\mathbb{E}\left[F_\ell - F_{\ell-1}\right]$	$\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$	$P_\ell C_\ell$	P_ℓ	$\mathbb{E}\left[F_\ell - F_{\ell-1}\right]$	$\mathbb{V}\left[F_\ell - F_{\ell-1}\right]$	$P_\ell C_\ell$
0	5.0×10^{-1}	3 420 716	9.91×10^{-1}	1.96×10^{0}	68 414	3 436 293	9.89×10^{-1}	1.96×10^{0}	68 726
1	1.0×10^{-2}	406 158	-1.23×10^{-1}	1.42×10^{0}	414 281	78 377	-1.36×10^{-1}	5.25×10^{-2}	79 945
2	5.0×10^{-3}	130 658	8.34×10^{-3}	4.36×10^{-1}	391 974	132 725	1.06×10^{-2}	4.38×10^{-1}	398 175
3	2.5×10^{-3}	88 941	2.57×10^{-2}	4.02×10^{-1}	533 646	90 310	3.02×10^{-2}	4.03×10^{-1}	541 860
4	1.3×10^{-3}	55 692	3.05×10^{-2}	3.12×10^{-1}	668 304	55 124	2.65×10^{-2}	3.04×10^{-1}	661 488
5	6.3×10^{-4}	31 140	2.29×10^{-2}	1.96×10^{-1}	747 360	30 906	2.49×10^{-2}	1.91×10^{-1}	741 744
6	3.1×10^{-4}	16 680	8.53×10^{-3}	1.15×10^{-1}	800 640	16 510	9.44×10^{-3}	1.10×10^{-1}	792 480
7	1.6×10^{-4}	8 006	9.38×10^{-3}	5.21×10^{-2}	768 576	8 559	4.06×10^{-3}	5.98×10^{-2}	821 664
8	7.8×10^{-5}	3 842	1.88×10^{-3}	2.35×10^{-2}	737 664	4 514	2.64×10^{-3}	3.12×10^{-2}	866 688
9	3.9×10^{-5}	2 175	3.07×10^{-4}	1.23×10^{-2}	835 200	2 754	5.54×10^{-3}	1.86×10^{-2}	1 057 536
10	2.0×10^{-5}	1 000	3.54×10^{-3}	1.74×10^{-2}	768 000	1 000	1.88×10^{-3}	1.08×10^{-2}	768 000
$\sum$			9.79×10^{-1}		6 734 059		9.69×10^{-1}		6 798 305

Comparing Tables 3 and 4, shows that leaving out levels reduces the computational cost when using the improved correlation at level 1, while the term-by-term correlation becomes more expensive. The simulation with term-by-term correlation in Table 4 requires a factor 2 less computation than that with improved correlation in Table 3.

As a reference for the bias, we note that a simulation with $E = 0.001$ in [12], computed an estimate of 9.80×10^{-1} for $\hat{Y}(t^*)$. Tables 3 and 4 indicate that the results using the term-by-term correlation are consistently within the set

Improved MLMC for kinetic equations in the diffusive scaling   13

tolerance $E = 0.01$. Those using the new correlation are consistently too small. Twenty repetitions of the simulation with term-by-term correlation in Table 3 and that with improved correlation in Table 4 estimatess a bias of 2.31×10^{-2} with standard deviation 1.14×10^{-2}. This shows that the improved correlation comes with the cost of an additional bias.

6   Conclusion

We presented a new, improved particle based multilevel Monte Carlo scheme for simulating the Boltzmann-BGK equation in diffusive regimes. This scheme improves upon earlier work by introducing a correlation between the kinetic transport in the fine simulation and the diffusion in the coarse simulation. The approach reduces the difference estimator variance at level 1 by an order of magnitude. In turn, the reduced variance significantly decreases the computation required for low accuracy simulations. For more accurate simulations, the reduced variance allows for several computationally expensive levels to be skipped, resulting in a greatly improved computational efficiency. A caveat is the inter time-step correlation and non-normal distribution of the coarse random numbers, resulting in biased results. Resolving these inconsistencies is the topic of future work, which will aid the extension of the improved correlation to all levels.

Table 4. Computing $\hat{Y}(t^*)$ for $\varepsilon = 0.1$ and $E = 0.01$ with skipped levels.

Level	Δt_ℓ	Term-by-term correlation				Improved correlation			
		P_ℓ	$\mathbb{E}[F_\ell - F_{\ell-1}]$	$\mathbb{V}[F_\ell - F_{\ell-1}]$	$P_\ell C_\ell$	P_ℓ	$\mathbb{E}[F_\ell - F_{\ell-1}]$	$\mathbb{V}[F_\ell - F_{\ell-1}]$	$P_\ell C_\ell$
0	5.0×10^{-1}	4 937 922	9.91×10^{-1}	1.96×10^0	98 758	1 848 883	9.89×10^{-1}	1.95×10^0	36 978
1	7.8×10^{-5}	85 415	-4.93×10^{-3}	3.76×10^0	10 934 828	3 171	-3.12×10^{-2}	3.76×10^{-2}	405 951
2	3.9×10^{-5}	3 683	1.55×10^{-4}	1.57×10^{-2}	1 414 272	1 114	-5.25×10^{-3}	1.42×10^{-2}	427 776
3	2.0×10^{-5}	1 000	3.01×10^{-3}	3.16×10^{-3}	768 000	1 146	2.68×10^{-4}	2.58×10^{-2}	880 128
4	9.8×10^{-6}	-	-	-	-	1 000	-1.25×10^{-3}	6.75×10^{-3}	1 536 000
$\sum$			9.89×10^{-1}		13 215 858		9.51×10^{-1}		3 286 833

References

1. Bhatnagar, P.L., Gross, E.P., Krook, M.: A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems. Physical Review **94**(3), 511–525 (may 1954). https://doi.org/10.1103/PhysRev.94.511
2. Crestetto, A., Crouseilles, N., Lemou, M.: A particle micro-macro decomposition based numerical scheme for collisional kinetic equations in the diffusion scaling. Communications in Mathematical Sciences **16**(4), 887–911 (2018). https://doi.org/10.4310/CMS.2018.v16.n4.a1
3. Degond, P., Dimarco, G., Pareschi, L.: The moment-guided Monte Carlo method. International Journal for Numerical Methods in Fluids **67**(2), 189–213 (sep 2011). https://doi.org/10.1002/fld.2345

4. Dimarco, G., Pareschi, L.: Numerical methods for kinetic equations. Acta Numerica **23**, 369–520 (may 2014). https://doi.org/10.1017/S0962492914000063
5. Dimarco, G., Pareschi, L., Samaey, G.: Asymptotic-Preserving Monte Carlo Methods for Transport Equations in the Diffusive Limit. SIAM Journal on Scientific Computing **40**(1), A504–A528 (jan 2018). https://doi.org/10.1137/17M1140741
6. Dimarco, G., Pareschi, L.: Hybrid Multiscale Methods II. Kinetic Equations. Multiscale Modeling & Simulation **6**(4), 1169–1197 (jan 2008). https://doi.org/10.1137/070680916
7. Dimarco, G., Pareschi, L.: Fluid Solver Independent Hybrid Methods for Multiscale Kinetic Equations. SIAM Journal on Scientific Computing **32**(2), 603–634 (jan 2010). https://doi.org/10.1137/080730585
8. Giles, M.B.: Multilevel Monte Carlo Path Simulation. Operations Research **56**(3), 607–617 (2008). https://doi.org/10.1287/opre.1070.0496
9. Jin, S., Pareschi, L., Toscani, G.: Uniformly Accurate Diffusive Relaxation Schemes for Multiscale Transport Equations. SIAM Journal on Numerical Analysis **38**(3), 913–936 (jan 2000). https://doi.org/10.1137/S0036142998347978
10. Kloeden, P.E., Platen, E.: Weak Taylor Approximations. In: Numerical Solution of Stochastic Differential Equations, pp. 457–484. Springer Berlin Heidelberg, Berlin, Heidelberg (1992). https://doi.org/10.1007/978-3-662-12616-5_14
11. Lapeyre, B., Pardoux, É., Sentis, R., Craig, A.W., Craig, F.: Introduction to Monte Carlo methods for transport and diffusion equations, vol. 6. Oxford University Press (2003)
12. Løvbak, E., Samaey, G., Vandewalle, S.: A multilevel Monte Carlo method for asymptotic-preserving particle schemes. arXiv:1907.04610 (jul 2019)
13. Løvbak, E., Samaey, G., Vandewalle, S.: A Multilevel Monte Carlo Asymptotic-Preserving Particle Method for Kinetic Equations in the Diffusion Limit. In: L'Ecuyer, P., Tuffin, B. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2018. In press. Springer (2020)
14. Mortier, B., Baelmans, M., Samaey, G.: Kinetic-diffusion asymptotic-preserving Monte Carlo algorithms for plasma edge neutral simulation. Contributions to Plasma Physics p. e201900134 (dec 2019). https://doi.org/10.1002/ctpp.201900134
15. Pareschi, L.: Hybrid Multiscale Methods for Hyperbolic and Kinetic Problems. ESAIM: Proceedings **15**, 87–120 (nov 2005). https://doi.org/10.1051/proc:2005024
16. Pareschi, L., Caflisch, R.E.: An Implicit Monte Carlo Method for Rarefied Gas Dynamics. Journal of Computational Physics **154**(1), 90–116 (sep 1999). https://doi.org/10.1006/jcph.1999.6301
17. Pareschi, L., Russo, G.: An introduction to Monte Carlo method for the Boltzmann equation. ESAIM: Proceedings **10**, 35–75 (aug 2001). https://doi.org/10.1051/proc:2001004
18. Pareschi, L., Trazzi, S.: Numerical solution of the Boltzmann equation by time relaxed Monte Carlo (TRMC) methods. International Journal for Numerical Methods in Fluids **48**(9), 947–983 (jul 2005). https://doi.org/10.1002/fld.969
19. Reiter, D., Baelmans, M., Börner, P.: The EIRENE and B2-EIRENE Codes. Fusion Science and Technology **47**(2), 172–186 (feb 2005). https://doi.org/10.13182/FST47-172