

# Multilevel Monte Carlo and control-variate simulation of Coulomb collisions

L. J. Höök, T. Johnson and T. Hellsten

School of Electrical Engineering, Div. Fusion Plasma Physics, Royal Institute of Technology (KTH), Teknikringen 31 SE-100 44 Stockholm, Sweden

E-mail: [joh@kth.se](mailto:joh@kth.se)

**Abstract.** The multilevel Monte Carlo and the control-variate method are reviewed for simulation of Coulomb collisions. The test case considered is fast-ion relaxation from ion-ion and electron-ion collisions against a Maxwellian background. A simplified energy scattering model is derived and related to the Cox-Ingersoll-Ross model for which, an exact time dependent solution can be obtained. The exact model is used as a control-variate for the estimation of the mean energy.

AMS classification scheme numbers: 65C05,65C30,60H35

Submitted to: *Plasma Phys. Control. Fusion*

## 1. Introduction

Simulating particles evolved in time by a stochastic differential equation (SDE) in a Lagrangian frame of reference is a common method for solving diffusion equations in high dimensions. What characterizes stochastic differential equations in contrast to ordinary differential equations is that they have a random term called the Wiener process or Brownian motion. The computational work of particle simulations scales linearly with the number of dimensions and thus do not suffer from the exponential dependence of the dimensionality that the finite difference and the finite element method have. However, due to the finite number of particles the simulation result contains statistical noise. The standard deviation of the statistical uncertainty scales as,

$$\alpha N^{-\beta}, \quad (1)$$

where  $N$  is the number of particles, markers or samples. Numerical techniques for improving the statistical convergence can therefore be separated into two categories: variance reduction techniques that minimize  $\alpha$  and quasi-Monte Carlo techniques that enhance  $\beta$ . Two common methods that minimize  $\alpha$  are the importance-sampling and the control-variate method [1]. The control-variate method reduces  $\alpha$  by solving an auxiliary model, which is correlated with the original model.

For the standard Monte Carlo method the value of  $\beta$  is  $1/2$ . Improving the order of convergence  $\beta$  is slightly more difficult than reducing  $\alpha$ . This is achieved by replacing the pseudo-random numbers with so called quasi-random numbers that are more uniformly distributed over the integration interval. It can be shown [2] that the asymptotic convergence of the statistical error with quasi-random numbers scales as  $\mathcal{O}(\log(N)^s N^{-1})$  where  $s$  is the number of physical dimensions times the number of time steps. In low dimensions the logarithm term is often ignored. However, when  $s \gg 1$  becomes large this term will blow up. Recent work [3] on quasi-Monte Carlo simulation of the pitch and energy scattering processes show that the convergence is close to  $\mathcal{O}(N^{-1})$  for modest number of time steps.

When solving a stochastic differential equation with numerical methods an error from the time discretization is obtained in addition to the statistical error. To minimize the total error a new technique called the multilevel Monte Carlo method based on multigrid ideas was introduced in [4].

The purpose of this paper is to investigate the performance of the control-variate and the multilevel Monte Carlo method for simulation of Coulomb collisions against a local Maxwellian in one-dimension. In section 2, we briefly introduce the link between the stochastic differential equation and the Fokker-Planck equation and discuss the numerical error obtained when simulating particle ensembles. In section 3, we derive a simplified energy scattering model from the *Chandrasekhar* [5] coefficients, which is used as a control-variate. Section 4 introduces the multilevel Monte Carlo method and simulation results are reported. The results are summarized in section 5.

## 2. Approximation of stochastic differential equations

We start of with the definition of the autonomous Itô stochastic differential equation (SDE) for a stochastic variable  $V(t) \in \mathbb{R}$ :

$$dV(t) = A(V(t))dt + \sigma(V(t))dW(t), \quad t_0 \leq t \leq T \quad (2)$$

with  $V(t_0) = v$ . Here  $dW(t) = W(t + dt) - W(t) \in N(0, dt)$  and  $W(t)$  is the Wiener process also known as Brownian motion, which is a normally distributed random process. The Wiener process is the time integral of the Langevin force. Many realizations of the SDE give an ensemble of particles with a distribution function described by the corresponding Fokker-Planck equation,

$$\frac{\partial f}{\partial t}(v, t) = -\frac{\partial}{\partial v}A(v)f + \frac{1}{2}\frac{\partial^2}{\partial v^2}\sigma^2(v)f, \quad (3)$$

where  $A$  and  $\sigma^2/2$  are the drift and diffusion coefficient respectively. Analytical solution of SDEs are rare and depend strongly on the structure of the coefficients. Therefore it is common to integrate SDEs with numerical schemes. The simplest scheme available is the Euler-Maruyama scheme, which read:

$$\tilde{V}^{i+1} = \tilde{V}^i + A(\tilde{V}^i)\Delta t + \sigma(\tilde{V}^i)\Delta t^{1/2}Z^i, \quad (4)$$

where  $Z^i$  is a normally distributed random number with zero mean and unit variance. Here  $i$  is the time increment index on the grid  $\{t_0, \dots, t_0 + i\Delta t, \dots, T\}$ . Two common measures of the error are the strong and the weak errors [6]. The strong error measures convergence of the paths of individual samples w.r.t. the time step, while the weak error is defined as

$$|\epsilon_{\Delta t}| = |\mathbb{E}[g(V(T))] - \mathbb{E}[g(\tilde{V}(T))]| \leq K_q \Delta t^q, \quad (5)$$

and measures the convergence in distribution (moments). Given a numerical solution of the SDE (2), assume that the goal is to estimate the expected value,  $E[g(\tilde{V}(T))]$ . The function  $g(\cdot)$  measures an unspecified property of the distribution function. This function can be any type of measurement on the distribution function, e.g. a mean value  $g(x) = x$  or a indicator function, used for creating histograms. The concept of the Monte Carlo method is to approximate the expected value with the sample mean from  $N$  independent simulations,

$$\mathbb{E}[g(\tilde{V}(T))] = \int g(\tilde{v})f(\tilde{v})d\tilde{v} \approx \frac{1}{N} \sum_{l=1}^N g(\tilde{V}(\omega_l)(T)), \quad (6)$$

where  $\tilde{V}(\omega_l)$  are realizations (particles) at the final time  $T$ . The statistical error,

$$\epsilon_{stat} = \mathbb{E}[g(\tilde{V}(T))] - \frac{1}{N} \sum_{l=1}^N g(\tilde{V}(\omega_l)(T)), \quad (7)$$

of this estimator is a stochastic variable with zero mean and variance,

$$\text{Var}[\epsilon_{stat}] = \text{Var}[g(\tilde{V}(T))]N^{-1}. \quad (8)$$

This gives the well known  $N^{-1/2}$  behavior of the confidence interval. Methods that reduce the statistical error are called variance reduction methods. The total error is the sum of the statistical error and the bias from the numerical time integration,

$$\epsilon_{tot} = \mathbb{E}[g(V(T))] - \frac{1}{N} \sum_{l=1}^N g(\tilde{V}(\omega_l)(T)) = \epsilon_{\Delta t} + \epsilon_{stat}, \quad (9)$$

and has a mean square error,

$$\mathbb{E}[\epsilon_{tot}^2] = \text{Var}[\epsilon_{stat}] + \epsilon_{\Delta t}^2 \approx \text{Var}[g(\tilde{V}(T))]N^{-1} + K_q \Delta t^{2q}. \quad (10)$$

Note that this is a weak error since it measures the difference of the mean and not the mean of the difference. In the next section we will present methods that minimize the statistical error.

### 3. The control-variate method

One of the more well known variance reduction technique available is the control-variate method. To illustrate this method we here consider the problem of estimating the  $\mathbb{E}[g(V)]$  for a function  $g(\cdot)$  and a stochastic process  $V(t)$  evolving according to (2). Next construct an new stochastic differential equation with coefficients  $A_0 \approx A$  and  $\sigma_0 \approx \sigma$  for a process  $V_0$ , called the *control variate*, such that  $m = \mathbb{E}[g(V_0)]$  is known. Thus  $V$  and  $V_0$  are described by the SDEs,

$$dV(t) = A(V(t))dt + \sigma(V(t))dW(t) \quad (11)$$

$$dV_0(t) = A_0(V_0(t))dt + \sigma_0(V_0(t))dW(t). \quad (12)$$

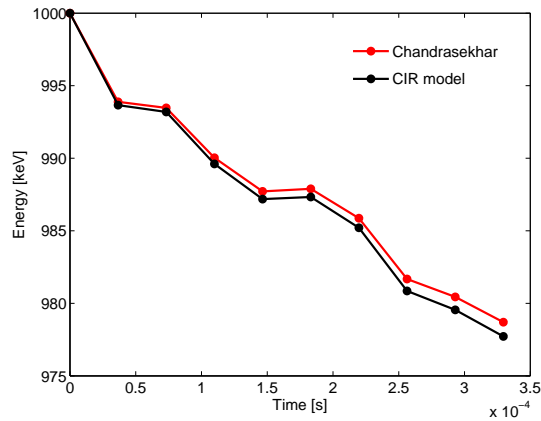
Finally we construct a new process,

$$Z = g(V) - c(g(V_0) - m), \quad (13)$$

for a scalar  $c \in [0, 1]$  determined below. Comparing  $Z$  with  $g(V)$  we note that they have the same mean,  $\mathbb{E}[Z] = \mathbb{E}[g(V)]$ , thus our problem could be solved by simulating  $Z$ . The variance of  $Z$  and  $g(V)$  are however different,

$$\text{Var}[Z] = \text{Var}[g(V)] + c^2 \text{Var}[g(V_0)] - 2c \text{Cov}[g(V), g(V_0)]. \quad (14)$$

Minimizing this variance with respect to  $c$  yields  $c = \text{Cov}[g(V), g(V_0)]/\text{Var}[g(V_0)]$ . Note that when  $g(V)$  and  $g(V_0)$  are strongly correlated, then  $\text{Var}[Z] \ll \text{Var}[g(V)]$ . According to (7) these variances describe the numerical error obtained in Monte Carlo evaluations of  $\mathbb{E}[Z]$  and  $\mathbb{E}[g(V)]$ . Thus, for strong correlation between  $g(V)$  and  $g(V_0)$  the evaluation of  $\mathbb{E}[Z]$  has significantly smaller statistical error than an evaluation of  $\mathbb{E}[g(V)]$ . Replacing  $V_0$  by  $V$  give  $c = 1$  and the variance of  $Z$  is zero. In the following section we will derive a simplified model, which can work as a control variable for simulation of the energy scattering process.



**Figure 1.** Particle paths of the exact model (15) in red and the approximate model (17) in black.

### 3.1. A control-variate model for the energy scattering operator

Modeling energy scattering against a Maxwellian background is achieved with the following SDE,

$$dE = (\alpha(2mE)^{1/2} + m\beta/2) dt + (2mE\beta)^{1/2} dW, \quad (15)$$

where  $\alpha, \beta$  are the Coulomb coefficients derived by *Chandrasekhar* [5]. Exact expressions for the coefficients are found in [7]. Neglecting the ion-ion collisions we can find approximate expressions for  $\alpha$  and  $\beta$  valid in the region  $\max(V_\alpha, V_\beta) \ll v \ll v_{th}^e$ :

$$\alpha \approx -\frac{v}{t_s}, \quad \beta \approx \frac{2kT_e}{mt_s}, \quad (16)$$

where  $t_s$  is the slowing down time and  $V_{\alpha, \beta}$  are given in [7]. Inserting these expressions in (15) gives,

$$dE_{cv} = \frac{2}{t_s} \left( \frac{kT_e}{2} - E_{cv} \right) dt + 2 \left( \frac{kT_e}{t_s} E_{cv} \right)^{1/2} dW. \quad (17)$$

This SDE is the Cox-Ingersoll-Ross model (CIR) [8] and describes a squared Ornstein-Uhlenbeck process [9] with the mean value reverting to  $kT_e/2$  for large times. The CIR model has a time dependent analytical solution of the distribution function [10] and can be simulated exactly [1, p.124]. The expected value of (17) satisfies,

$$m(t) = \frac{kT_e}{2} + \left( m_0 - \frac{kT_e}{2} \right) \exp \left( -\frac{2t}{t_s} \right), \quad (18)$$

where  $m(t) = \mathbb{E}[E_{cv}]$ . We next analyze the performance of (15) and (17) for the estimation of the mean energy using  $Z = E - c(E_{cv} - m(t))$ . The simulated plasma

consists of protons and electrons, both with densities  $3 \times 10^{19} \text{ m}^{-3}$  at temperature 4 [keV]. The initial energy is 1 [MeV]. The fast protons above and below  $mV_\alpha^2/2 = 60$  [keV] collide primarily with the electrons and with the thermal protons respectively. The simulation was run to 0.5 and to 1.5 slowing down times ( $t_s$ ). The number of time steps were 2048 and 3072, respectively. Equation (15) was discretized with the Euler-Maruyama scheme and (17) was discretized with,

$$\tilde{E}_{cv}^{i+1} = \frac{kT_e}{2} + e^{-\frac{2}{t_s}\Delta t} \left( \tilde{E}_{cv}^i - \frac{kT_e}{2} + 2 \left( \frac{kT_e}{t_s} \tilde{E}_{cv}^i \Delta t \right)^{1/2} Z^i \right), \quad (19)$$

where  $Z^i \sim N(0, 1)$  [11, 4]. Figure 1 illustrates the thermalization of two particles, one from the approximate model in black (17) and the other simulated with the exact model in red (15). The figure shows that the paths are correlated and are slowly separating as they go down in energy. This is also seen in figures 2(a), 2(b) where the approximate model is plotted against the exact model. At  $0.5t_s$  the correlation between  $E$  and  $E_{cv}$  is strong, which means that the realizations of  $E - E_{cv}$  is small and hence the distribution of  $Z$  becomes very narrow as illustrated in figures 2(c), 2(d). Continuing the simulation to  $1.5t_s$  the correlation is now weak at energies around and below 60 [keV], which is where the ion-ion collisions start to dominate in the exact model. The weak correlation give a distribution of  $Z$  with a larger variance as seen in figure 2(d).

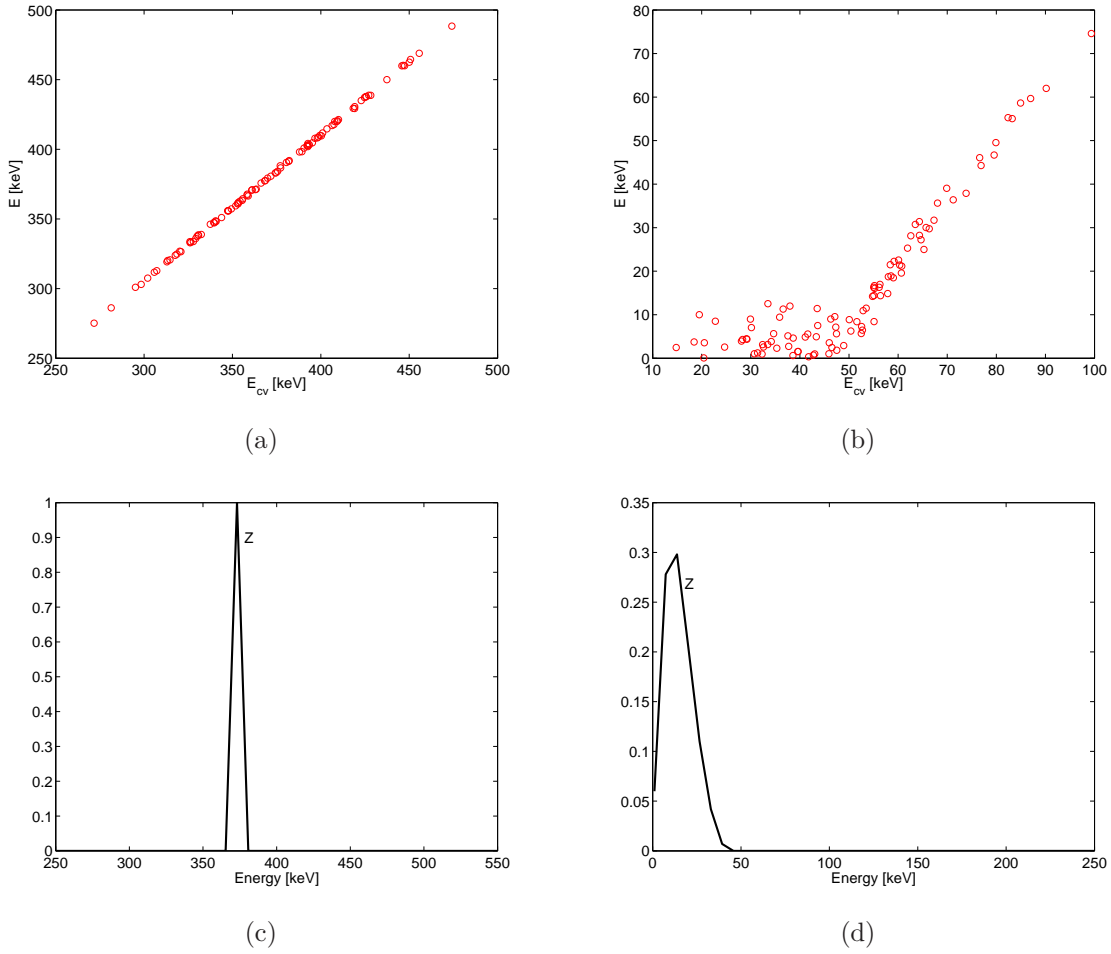
Calculations of the batch mean energy with 90% confidence interval are given in table 1. For a fixed statistical error, the variance reduction in (8) give a smaller  $N$  and a smaller value of the computational complexity, which is proportional to  $\mathcal{O}(N/\Delta t)$ . Therefore we can measure the speedup, of a simulation, as the ratio of the batch variance of the control-variate estimator and the standard Monte Carlo estimator. The speedup is about 4000 at  $0.5t_s$ , which means that we can obtain the same confidence interval with 4000 times fewer particles than needed by the standard Monte Carlo. A speedup of 6 is achieved for  $1.5t_s$ .

**Table 1.** The batch averaged correlation for  $T = 0.5t_s$  is 99.98% and the batch averaged speedup is 4159. The correlation at  $T = 1.5t_s$  is 91.15% and the speedup is 6.

$T = 0.5t_s$	Batch mean ( 90% confidence )	Batch variance
Control-variate	$378.17 \pm 0.016$	0.00076
Standard MC	$378.64 \pm 0.545$	0.88
$T = 1.5t_s$		
Control-variate	$18.064 \pm 0.178$	0.095
Standard MC	$18.091 \pm 0.367$	0.403

#### 4. The Multilevel Monte Carlo method

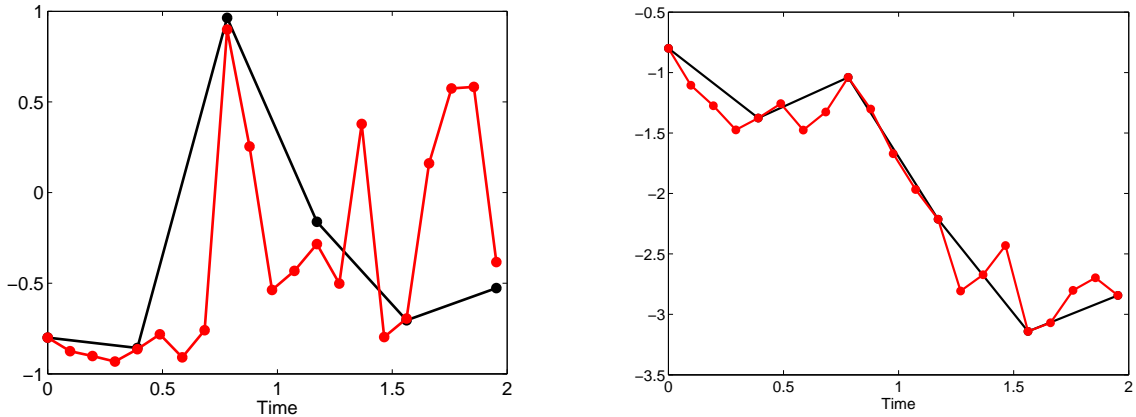
The multilevel Monte Carlo method is closely related to the control variate method, but there are a number of differences. Perhaps the most fundamental difference is that



**Figure 2.** Scatter plot of  $(E, E_{cv})$  (top panel) for  $T = 0.5t_s$  (left column) and  $T = 1.5t_s$  (right column). Bottom panel show the distributions of  $Z$ .

it considers not only the statistical error, but the total error described in (10). The auxiliary model in the multilevel method is not a simplified model of the stochastic differential equation, as in the control-variate method, but an approximate Monte Carlo solution of the original equation evaluated on a sparse time grid. Thus, the auxiliary model will give a large time discretization error. However, since it has been evaluated on a sparse time grid (a few long time steps) it is relatively cheap to evaluate this solution with many particles and thus with a small statistical error. The next step in the multilevel method is a *corrector step* and involves estimating the time discretization error on the sparse time grid. This corresponds to the term  $\mathbb{E}[g(V_1) - g(V_0)]$  in the control-variate method and is evaluated by letting  $V_0$  be a solution on a sparse time grid and  $V_1$  be a solution on a refined grid. Note that  $V_0$  and  $V_1$  have to be calculated from the same realization of the Wiener process  $W(t)$ . This is illustrated in figure 3; the two processes  $V_0$  and  $V_1$  in 3(a) are simulated using the two time discretizations from the same realization of the Wiener process, shown in 3(b). Since  $V_0$  and  $V_1$  are strongly correlated the variance  $\text{Var}[g(V_1) - g(V_0)]$  is small and relatively few particles are needed

to evaluate the correction  $\mathbb{E}[g(V_1) - g(V_0)]$ . The *corrector step* described above can be



(a) Sample trajectories for  $V_0$  (black) and  $V_1$  (red) (b) Underlying Brownian motion sampled on coarse (black) and fine (red) time grids

**Figure 3.** Illustration of the strong convergence property for coarse (black) and fine (red) time discretization levels.

applied several times to generate a sequence of solutions with more and more refined time grids [4], forming a telescoping sum,

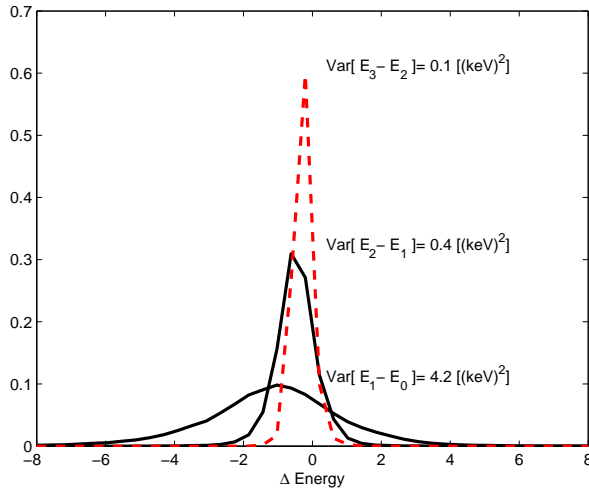
$$\begin{aligned} \mathbb{E}[g(V_L)] &= \mathbb{E}[g(V_0)] + \mathbb{E}[g(V_1) - g(V_0)] + \mathbb{E}[g(V_2) - g(V_1)] + \dots \\ &= \mathbb{E}[g(V_0)] + \sum_l^L \mathbb{E}[g(V_l) - g(V_{l-1})]. \end{aligned} \quad (20)$$

Here each level  $l$  represents a time grid (the higher the value of  $l$  the finer the grid) and each term in the sum is a corrector step. In order for the different corrector steps not to correlate, each corrector step has to be evaluated with a different realization of the Wiener process. However, within one corrector step  $V_l$  and  $V_{l-1}$  has to be evaluated with two time discretizations from the *same* realization of the Wiener process. Note that for each corrector step the time steps gets smaller and thus the processes  $V_l$  and  $V_{l-1}$  will be more and more correlated, as illustrated in figure 4. Thus, the variance of  $\text{Var}[g(V_l) - g(V_{l-1})]$  will decrease and hence the fewer and fewer particles are needed to evaluate the corrector. An important part of designing a multilevel algorithm is therefore to optimize how the sequence of time grids and the number of particles to use in each corrector step are selected. For further reading on the subject see [4] and [12].

### *Thermalization of fast ions*

We have tested the multilevel Monte Carlo method, the randomized quasi-Monte Carlo method described in [3] and the standard Monte Carlo method on the one-dimensional energy scattering operator (15). The tests was conducted in Matlab<sup>TM</sup> version R2012a on a twelve core Intel computer @ 2.67 GHz with 62 gigabyte of RAM. In the multilevel





**Figure 4.** Distributions of the coarse fine estimator  $E_l - E_{l-1}$ .

Monte Carlo method we used  $2^6$  time steps on the coarsest level ( not a single time step  $\Delta t = T$  as in [4] ). The simulation parameters are the same as for the control-variate case, but with initial energy at 100 [keV] and  $T = t_s$  [s]. The time grid was refined by a factor of 4 on each level and the measured sample averages was run 10 times. Two moments were considered, the mean energy and the energy squared. The root mean square error is measured against a Monte Carlo simulation with  $N = 10^7$  and  $\Delta t = 2^{-12}t_s$ . The results are given in table 2. The multilevel Monte Carlo method is about 18 times faster than the standard Monte Carlo and is about 21 times faster when combined with the control-variate estimator. Note that the simulation time of the randomized quasi-Monte Carlo method is misleading, since we have excluded the wall-clock time to generate the quasi-random numbers, which was quite long compared to the other methods. The reason for this is because the quasi-random generator has not been optimized and cannot compete with the pseudo-random generator in Matlab.

## 5. Conclusions

We have derived an approximate analytical energy relaxation model and used it as a control-variate for studying fast ion relaxation. The derived model is the Cox-Ingersoll-Ross model from finance, which has a known analytical solution of the time evolved distribution function. The performance of the derived model is very good when electron collisions dominate, but breaks down when ion-ion collisions are important. The multilevel Monte Carlo method was tested on the same model and compared with the randomized quasi-Monte Carlo method and the standard Monte Carlo method. The multilevel Monte Carlo method is up to 21 times faster, measured in wall-clock simulation time, than the standard Monte Carlo method for similar mean error. The method requires fewer particles to be simulated than the randomized quasi-Monte Carlo

**Table 2.** Simulation results from the multilevel Monte Carlo method (MLMC), the randomized quasi-Monte Carlo method (RQMC) and the standard Monte Carlo (MC). Reference values obtained from a standard Monte Carlo simulation with  $10^7$  particles and  $\Delta t = 2^{-12}t_s$  are 6.0254 [keV] for  $\mathbb{E}[E]$  and 60.4624 [(keV)<sup>2</sup>] for  $\mathbb{E}[E^2]$ . Note that the time for generating the quasi-random numbers has been excluded for RQMC.

Type	$\mathbb{E}[E]$ [keV]	r.m.s.	wall-clock time [s]	$N$
MC	6.054	0.0285	280	$10^6$
MLMC	6.032	0.0202	15	$\approx 7 \times 10^5$
MLMC+CV	6.038	0.0218	13	$\approx 7 \times 10^5$
RQMC	6.051	0.0278	19	$10^5$
Type	$\mathbb{E}[E^2]$ [(keV) <sup>2</sup> ]			
MC	60.872	0.419	280	$10^6$
MLMC	60.585	0.384	17	$\approx 10^5$
RQMC	60.873	0.432	15	$10^5$

method.

## Acknowledgments

This work has been partially conducted within the framework of the EFDA Integrated Tokamak Modelling task force and the EFDA Goal Oriented Training Programme GOTiT.

## 6. References

- [1] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Applications of Mathematics. Springer, 2004.
- [2] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
- [3] L J Höök, T Johnson, and T Hellsten. Randomized quasi-monte carlo simulation of fast-ion thermalization. *Computational Science & Discovery*, 5(1):014010, 2012.
- [4] Michael B. Giles. Multilevel monte carlo path simulation. *Operations Research*, 56(3):607–617, May/June 2008.
- [5] L. Spitzer. *Physics of fully ionized gases*. Interscience tracts on physics and astronomy. Interscience Publishers, 1962.
- [6] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Applications of mathematics. Springer-Verlag, 1992.
- [7] T.H. Stix. *Waves in plasmas*. American Institute of Physics, 1992.
- [8] John C. Cox, Jr. Ingersoll, Jonathan E., and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):pp. 385–407, 1985.
- [9] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841, Sep 1930.
- [10] William Feller. Two singular diffusion problems. *The Annals of Mathematics*, 54(1):pp. 173–182, 1951.
- [11] Christian Kahl and Peter Jäckel. Fast strong approximation monte carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(6):513–536, 2006.
- [12] Waterhouse B.J. Giles M.B. Multilevel quasi-monte carlo path simulation. *Radon Ser. Comput. Appl. Math.*, 8:1–18, 2009. cited By (since 1996) 1.