Computer Physics Communications 191 (2015) 119–124

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Mean square displacements with error estimates from non-equidistant time-step kinetic Monte Carlo simulations

Mikael Leetmaa *, Natalia V. Skorodumova

Multiscale Materials Modelling, Materials Science and Engineering, School of Industrial Engineering and Management, KTH - Royal Institute of Technology, Brinellvägen 23, 100 44 Stockholm, Sweden
Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 26 August 2014
Received in revised form
6 February 2015
Accepted 10 February 2015
Available online 17 February 2015

Keywords:
KMC
Kinetic
Monte Carlo
Diffusion
Simulation
Errors
Mean square displacement
MSD

ABSTRACT

We present a method to calculate mean square displacements (MSD) with error estimates from kinetic Monte Carlo (KMC) simulations of diffusion processes with non-equidistant time-steps. An analytical solution for estimating the errors is presented for the special case of one moving particle at fixed rate constant. The method is generalized to an efficient computational algorithm that can handle any number of moving particles or different rates in the simulated system. We show with examples that the proposed method gives the correct statistical error when the MSD curve describes pure Brownian motion and can otherwise be used as an upper bound for the true error.

1. Introduction

Diffusion processes play a hugely important role in all aspects of human activity, from the very biological foundations of our existence to a myriad of technological applications. A detailed understanding of diffusion processes is therefore crucial in research fields as diverse as energy production and catalysis, defect diffusion and segregation in steels and alloys, and in biological and pharmaceutical sciences.

Detailed insight into diffusion processes can be obtained from atomic scale simulations in the form of calculated velocity auto-correlation functions, or from mean square displacement (MSD) data for the diffusing particles. For molecular-dynamics (MD) simulations there exist well established techniques to calculate both velocity auto-correlation functions and MSD, and to estimate the errors of these and derived quantities [1,2]. Related work is done in the context of single particle tracking experiments, where trajectories of measured particle positions provide data for MSD analysis and subsequent extraction of diffusion constants [3–5]. Common for these simulations and experimental techniques is that the time-steps in the generated particle trajectories are equidistant.

Obtaining the MSD curve as a function of time is straightforward from trajectories with equidistant time-steps. The position of each particle is registered at each time-frame and the squared displacement for each particle is calculated and registered for all possible time differences. The resulting squared displacements are averaged for each time difference and all diffusing particles to yield the MSD data [6]. Estimating the errors can be done using the block-average technique of Flyvbjerg and Petersen [7] applied to each bin in the MSD data, or, as in the context of single particle tracking experiments, by using the analytical expressions for the error estimates derived by Qian et al. [3].

Kinetic Monte Carlo (KMC) simulations is a powerful tool to investigate slow processes in atomic systems, such as rare diffusion events, that cannot easily be studied with conventional MD techniques. The time-steps in most versions of KMC are, however, non-equidistant [8], making the standard techniques for calculating diffusion far more elaborate to apply. Several diffusion studies with KMC that calculate MSD curves exist in the literature (see e.g. [9–12]). However it is challenging to find work where MSD curves are presented with error estimates together with a detailed description of the procedure used for obtaining them. Unless proper

* Corresponding author at: Multiscale Materials Modelling, Materials Science and Engineering, School of Industrial Engineering and Management, KTH - Royal Institute of Technology, Brinellvägen 23, 100 44 Stockholm, Sweden.
E-mail address: leetmaa@kth.se (M. Leetmaa).

http://dx.doi.org/10.1016/j.cpc.2015.02.006
0010-4655/

error estimates are available, subsequent extraction of diffusion constants from the MSD data will of course suffer the same lack of error estimates. The procedure of placing the trajectory from a KMC simulation on an equidistant time-grid, calculating the MSD curve from the equidistant trajectory, and estimate the errors for each MSD bin using a block-averages [7] or a jackknife [13] approach is appropriate, but computationally somewhat demanding and in practice rarely, if ever, used.

We will here describe a much simpler procedure for calculating the MSD curve with error estimates from non-equidistant time-step KMC simulations. The method we propose calculates the MSD as an equidistant histogram directly from the non-equidistant time step KMC simulation trajectory. We have then re-worked the expressions from Qian et al. [3] for the error estimates of MSD data for single particle tracking experiments, to be applicable to the MSD histogram from KMC. Using the block average algorithm of Flyvbjerg and Petersen [7] applied to each bin in the MSD data as a reference, we show that our method for estimating the errors is highly reliable, to a fraction of the computational cost of the block average approach. Our method gives the correct statistical error when the MSD curve describes pure Brownian motion and can otherwise be used as an upper bound for the true error. Our method has furthermore been implemented as an on-the-fly analysis option in the publicly available *KMCLib* program [14]. With our simple method readily available it should be straight forward for anyone presenting MSD curves from KMC simulations to include proper error estimates.

2. Algorithm

We propose the following method for calculating the MSD curve as an equidistant histogram from a non-equidistant time-step KMC simulation. For each step in the simulation the time and particle position for each particle of interest is recorded. The difference in simulation time between the current step and the N previous steps in the simulation are calculated, along with the corresponding squared differences in particle positions. The N previous steps will be referred to as the history window. The calculated time differences are binned on a fine grid in t, and the corresponding squared differences in particle positions are added to the corresponding bins in a histogram $D(t)$. The number of squared displacements added to each bin are recorded in a bin-count histogram $H(t)$, and the MSD curve $\rho(t)$ is calculated for each bin i as

$$\rho_i = D_i/H_i \tag{1}$$

with the corresponding time values

$$t_i = i\tau + \tau/2 \tag{2}$$

where τ is the bin size. Note that it does not matter if this procedure is carried out over a stored trajectory or if it performed on-the-fly during the simulation. The rest of this paper will be concerned with determining the error estimates and reliability of the MSD data $\rho(t)$. We will determine a cutoff t_c, that depends on the size of the history window N and the total rate in the simulated system. Above this cutoff the collected MSD data cannot be used. We will furthermore, based on the work for equidistant time-step single particle tracking by Qian et al. [3], derive an expression for the standard deviation of $\rho(t)$ from non-equidistant time-step KMC simulations. We will finally demonstrate with examples that our proposed algorithm for determining the cutoff and standard deviation can be used also for simulations of an arbitrary number of moving particles and elementary process rate constants. In the case of non-Brownian motion the obtained error estimates can be used as an upper limit of the true statistical error, while the error is correctly estimated for pure Brownian motion.

2.1. Analytical expression for the bin-count histogram

It is well known from probability theory that the distribution of the sum of n random independent variables with the same distribution $f(t)$ is given by the *n-fold* convolution $f^{*n}(t)$ [15]. With the exponentially decreasing distribution

$$f(t) = a\exp(-bt) \tag{3}$$

and the normalization condition

$$\int_0^\infty f^{*n}(t)dt = \int_0^\infty f(t)dt \tag{4}$$

and by defining $f^{*0}(t) = f(t)$ we have

$$f^{*n}(t) = \frac{b^n t^n}{n!} a\exp(-bt) \tag{5}$$

for KMC a simulation of one moving particle with the total rate-constant b kept fixed during the simulation the distribution of consecutive time-steps are given by (5), with $n = 0$ for the distribution of adjacent time-steps, $n = 1$ for the distribution of pairs of adjacent time-steps, etc. The constant a is related to the length of the simulation. With M steps in the simulation $a = b(M-1)$.

In the above described algorithm, the analytical equivalent $\mathcal{H}(t)$ of the MSD bin-count histogram $H(t)$ is, using a history window of size $(N+1)$, given by the sum

$$\mathcal{H}_N(t) = \sum_{n=0}^{N} f^{*n}(t) = a\exp(-bt)\sum_{n=0}^{N}\frac{b^n t^n}{n!} \tag{6}$$

where we recognize the well known series

$$\lim_{N\to\infty}\sum_{n=0}^{N}\frac{b^n t^n}{n!} = \exp(bt) \tag{7}$$

and write for the asymptotic expression when the size of the history window goes to infinity

$$\lim_{N\to\infty}\mathcal{H}_N(t) = a \quad \text{for all } t. \tag{8}$$

2.2. Truncation and simulation start-up effects

The limit expression (8) holds for a history window and simulation time approaching infinity. We will now examine what to expect when the sum in (6) is truncated. This is important since the simulations are always finite and the history window used is typically much smaller than the length of the simulation.

Fig. 1(a) shows $f^{*n}(t)$ and $\mathcal{H}_N(t)$ for $n = 0, 1, 2, \ldots, 15$ and $N = 15$, with $a = 20$ and $b = 30$. We know from probability theory that the distribution of the form of (5) can be approximated with a Gaussian for large values of n, centered at larger t for larger n. The bin-count histogram $\mathcal{H}_N(t)$ will thus stay constant according to (8) up to the point where $f^{*(N+1)}(t)$ should have started to contribute significantly and we see a rapid drop to zero in $\mathcal{H}_N(t)$. A cutoff t_c defined as the point where the last contributing $f^{*n}(x)$ curve reaches 1% of its maximum value is indicated with a blue vertical line in Fig. 1 at $t_c = 0.21$. A MSD curve sampled according to $\mathcal{H}(t)$ will above the t_c cutoff suffer from incorrect sampling since there will be important contributions missing from higher successive convolutions, while below t_c the collected MSD data can safely be used.

Fig. 1(b) shows the corresponding curves from a KMC simulations of one particle on a 1D lattice. The simulation was run for 10^6 elementary steps and the total rate was at each time $b = 30$. $f^{*n}(t)$ and $H_N(t)$ have here been scaled down a factor (1/3000) to normalize against the analytical expressions in Fig. 1(a). The