Towards Correlated Sampling for the Fixed-Node Diffusion Quantum Monte Carlo Method

Raphael Berner, René Petz, and Arne Lüchow

Institut für Physikalische Chemie, RWTH Aachen University, Landoltweg 2, 52056 Aachen, Germany

Reprint requests to A. L.; E-mail: luechow@rwth-aachen.de

Z. Naturforsch. **69a**, 279 – 286 (2014) / DOI: 10.5560/ZNA.2014-0002
Received December 9, 2013 / published online July 15, 2014

Dedicated to J. Fleischhauer's 75th birthday

Most methods of quantum chemistry calculate total energies rather than directly the energy differences that are of interest to chemists. In the case of statistical methods like quantum Monte Carlo the statistical errors in the absolute values need to be considerably smaller than their difference. The calculation of small energy differences is therefore particularly time consuming. Correlated sampling techniques provide the possibility to compute directly energy differences by simulating the underlying systems with the same stochastic process. The smaller the energy difference the smaller its statistical error. Correlated sampling is well established in variational quantum Monte Carlo, but it is much more difficult to implement in diffusion quantum Monte Carlo because of the fixed node approximation. A correlated sampling formalism and a corresponding algorithm based on a transformed Schrödinger equation having the form of a *Kolmogorov's backward equation* is derived. The numerical verification of the presented algorithm is given for the harmonic oscillator. The extension of the algorithm to electron structure calculations is discussed.

Key words: Quantum Monte Carlo; Kolmogorov Backward Equation; Correlated Sampling.

1. Introduction

Quantum Monte Carlo methods are an efficient and accurate tool to describe the electron structure and to calculate the energy of molecules and molecular aggregates [1]. In the fixed-node diffusion quantum Monte Carlo method (FN-DMC), the Schrödinger equation of a molecule or aggregate is solved with a stochastic process which is simulated in the computer. More precisely, the Schrödinger equation with the nodes of a trial function Ψ as Dirichlet boundary condition is solved exactly although the resulting energy has a statistical error dependent on the simulation length.

In many chemical problems small energy differences between very similar systems are sought. Examples are numerical gradients of the energy or the interaction energy between weakly bound molecules. With correlated sampling techniques, it is possible to obtain energy differences with a substantially smaller error bar than the energies itself. Systematic error cancellation in Monte Carlo methods is achieved with correlated sampling of both systems [2]. While this is well established for variational quantum Monte Carlo (VMC) [3, 4], it becomes very challenging in FN-DMC when the nodes of the two systems do not coincide. The most notable attempt to exploit correlated sampling within the FN-DMC method has been presented by Filippi and Umrigar who employ a space warp transformation to enable correlated sampling [5]. In this work, we demonstrate that correlated sampling is possible without a coordinate transformation if the simulated stochastic process is based on the Kolmogorov backward equation.

Let us denote the wave functions of the two systems A and B, for which the energy difference is to be calculated, as $\Psi^{(A)}$ and $\Psi^{(B)}$, respectively. Correlated sampling can not be implemented easily within the fixed-node approximation if $\Psi^{(A)}$ and $\Psi^{(B)}$ do not have the same nodes. Hence a reweighting scheme as in the VMC case is not feasible [6].

We propose to introduce a third function Ψ_{G} which is positive everywhere and serves as guide function. A suitable ansatz which can be applied to arbitrary molecules has been devised by Ceperley and Alder

in the course of the development of a released node Monte Carlo algorithm [7]. This guide function is constructed such that it is positive everywhere and resembles both $\Psi^{(A)}$ and $\Psi^{(B)}$. Any sample of walkers drawn from the stochastic process constructed with Ψ_G covers the whole configuration space even if the time step approaches zero. The basic idea is to construct a process based on Ψ_G that accounts for the different nodal hypersurfaces of $\Psi^{(A)}$ and $\Psi^{(B)}$. The resulting statistical correlation should decrease the error associated with the energy difference between system A and B.

This paper is organized as follows: First, the formalism of stochastic solutions of partial differential equations (PDEs) is briefly reviewed. Next, the stochastic process enabling correlated sampling and its Monte Carlo simulation is presented, and finally the error reduction for energy differences is shown for simple systems.

2. Stochastic Solution of Partial Differential Equations

Linear homogeneous second-order partial differential equations of the following type are considered:

$$\frac{\partial u(t,x)}{\partial t} = Lu(t,x)\,,\quad x\in\mathbb{R}^n\,,\quad t>0\,, \tag{1}$$

with the initial condition $u_0(x)=u(0,x)$ and

$$L=\frac{1}{2}\Delta+b(x)\cdot\nabla=\frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2}{\partial x_i^2}+\sum_{i=1}^{m}b_i(x)\frac{\partial}{\partial x_i}\,. \tag{2}$$

The differential operator L consists of a constant 'diffusion' term $\frac{1}{2}\Delta$ and a 'drift' term $b(x)\cdot\nabla$ where the vector field b is a function in $\mathbb{R}^n$. L can be considered as the generator of an Itô diffusion defined by the time-homogenous stochastic differential equation

$$\mathrm{d}X_t=\sigma W_t+b(X_t)\,\mathrm{d}t\,, \tag{3}$$

where W_t is a standard n-dimensional Wiener process and $\sigma=1$ [8].

The stochastic solution of (1) is given by

$$u(t,x)=\mathcal{E}^x[u(0,X_t)]\,, \tag{4}$$

where $\mathcal{E}^x[\cdot]$ denotes the expectation for the process X_t with initial position $X_0=x$. Note the time reversal; the initial function $u(0,x)$ is evaluated at time t. Equation (1) is called *Kolmogorov's backward equation* [9].

The adjoint operator L^* of L yields the following PDE:

$$\begin{aligned}\frac{\partial g(t,y)}{\partial t}&=L^*g(t,y)=\frac{1}{2}\Delta g-\nabla(b\cdot g)\\&=\frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 g}{\partial y_i^2}-\sum_{i=1}^{m}\frac{\partial}{\partial y_i}\big(b_i(y)\cdot g\big)\,.\end{aligned} \tag{5}$$

It is referred to as *Kolmogorov's forward equation* or Fokker–Planck equation [10]. This equation is the basis of most variational and diffusion quantum Monte Carlo methods.

Let ϕ_i be the eigenfunctions of the time-independent Schrödinger equation

$$H\phi_i=E_i\phi_i \tag{6}$$

with the eigenvalues E_i. The Hamiltonian contains only the kinetic energy operator $-\frac{1}{2}\Delta$, the potential energy V, and and a reference energy E_ref that allows a shift of the eigenvalue spectrum (which is bounded from below) such that $E_0\approx 0$:

$$H=-\frac{1}{2}\Delta+V-E_\mathrm{ref}\,. \tag{7}$$

With a positive guide function Ψ_G, a transformed Hamiltonian is constructed,

$$\tilde{H}=\Psi_\mathrm{G}^{-1}H\Psi_\mathrm{G}\,, \tag{8}$$

having the same eigenvalue spectrum and the eigenfunctions $\tilde{\phi}_i=\Psi_\mathrm{G}^{-1}\phi_i$. The operator

$$\tilde{H}^*=\Psi_\mathrm{G}H\Psi_\mathrm{G}^{-1} \tag{9}$$

is the adjoint operator with the eigenfunctions $\phi_i\Psi_\mathrm{G}$.

The corresponding time-dependent PDEs read as follows:

$$\frac{\partial u}{\partial t}=-\tilde{H}u=\frac{1}{2}\Delta u+b\nabla u-\Big(E_L-E_\mathrm{ref}\Big)u \tag{10}$$

and

$$\frac{\partial g}{\partial t}=-\tilde{H}^*g=\frac{1}{2}\Delta g+\nabla(b\cdot g)-\Big(E_L-E_\mathrm{ref}\Big)g \tag{11}$$

with the drift term $b=\Psi_\mathrm{G}^{-1}\nabla\Psi_\mathrm{G}$ and the local energy $E_L=\frac{1}{2}\Psi_\mathrm{G}^{-1}\Delta\Psi_\mathrm{G}+V$.

A glance at (1) and (5) reveals that (10) can be considered as a Kolmogorov backward equation with an additional source/sink term and (11) similarly as

a Kolmogorov forward equation. Note that the latter equation is the basis of fixed-node diffusion quantum Monte Carlo with importance sampling (see [6, and references therein]). The solution of (10) can be expressed easily in terms of the eigenfunctions of H. With $u(0,x) = \sum_{i=0} a_i \phi_i(x)/\Psi_G(x)$, the solution is

$$u(t,x) = \sum_{i=0} a_i \, \mathrm{e}^{-(E_i - E_{\mathrm{ref}})t} \, \frac{\phi_i(x)}{\Psi_G(x)} \,, \tag{12}$$

and in the long-time limit only the lowest eigenfunction survives, thus $u(t,x)$ becomes proportional to $\phi_0(x)/\Psi_G(x)$:

$$\begin{aligned} u(t,x) &= a_0 \, \mathrm{e}^{-(E_0 - E_{\mathrm{ref}})t} \, \frac{\phi_0(x)}{\Psi_G(x)} \,, \\ &\text{for } t \gg 1/|E_1 - E_0| \,, \;\; a_0 \neq 0 \,. \end{aligned} \tag{13}$$

This equation is the basis for the determination of the ground state energy E_0 (see below).

Using the stochastic process governed by (3), the solution to (10) can be expressed as [8]

$$\begin{aligned} u(t,x) = \mathcal{E}^x \Bigg[&\exp\left(- \int_0^t \Big(E_L\big(X_s\big) - E_{\mathrm{ref}} \Big) \, \mathrm{d}s \right) \\ &\cdot u(0, X_t) \Bigg] \,. \end{aligned} \tag{14}$$

This equation corresponds to the well-known Feynman–Kac method [11, 12]. It has recently been used by Korzeniowski et al. to solve the imaginary time-dependent Schrödinger equation in a Monte Carlo framework [13]. Note that this procedure differs from standard DMC as the solution is obtained pointwise rather than as distribution for the whole function. In DMC, the backward approach is employed in forward-walking techniques [14, 15].

In the fixed-node DMC method, additional boundary conditions have to be fulfilled. The forward equation allows correlated sampling with different (but similar) boundaries. As is common practice in quantum Monte Carlo, the boundary condition is given implicitly by the nodes of a trial function, here $\Psi_T^{(A)}$ and $\Psi_T^{(B)}$, respectively, where the solution has to vanish (fixed-node approximation). Let $D \subset \mathbb{R}^n$ be the corresponding domain and ∂D its boundary. The time when the diffusion process first hits the boundary ∂D is denoted *first exit time* τ_D

$$\tau_D = \inf \Big\{ t > 0 \,; \;\; X_t \in \partial D \Big\} \,. \tag{15}$$

To implement the fixed-node approximation, we confine our further examination to the special case that the solution $u(t,x)$ vanishes beyond the boundary by introducing the indicator function χ:

$$\chi = \begin{cases} 1 \,, & t \leq \tau_D \\ 0 \,, & t > \tau_D \end{cases} \,. \tag{16}$$

Equation (14) is extended to take into account the fixed-node Dirichlet boundary condition as follows [16]:

$$\begin{aligned} u(t,x) = \mathcal{E}^x \Bigg[&\exp\left(- \int_0^t \Big(E_L\big(X_s\big) - E_{\mathrm{ref}} \Big) \, \mathrm{d}s \right) \\ &\cdot \chi_{(t<\tau_D)} u(0, X_t) \Bigg] \,. \end{aligned} \tag{17}$$

In a Monte Carlo simulation, the time is discretized and the expectation is evaluated by weighting the function value of the initial function $u(0,x)$ at the position of the stochastic trajectory at the final time t. The weight is given by the discretized integral evaluated along the trajectory. All trajectories that have hit the boundary before time t count as zero in the expectation. The eigenvalue E_0 corresponding to the state of the symmetry imposed by the boundary condition can be calculated as

$$\begin{aligned} E_0 = -\lim_{t\to\infty} \frac{1}{t} \ln \mathcal{E}^x \Bigg[&\exp\left(- \int_0^t \Big(E_L(X_s) - E_{\mathrm{ref}} \Big) \, \mathrm{d}s \right) \\ &\cdot \chi_{(t<\tau_D)} u(0, X_t) \Bigg] \,. \end{aligned} \tag{18}$$

Numerically, it is more efficient not to integrate over all times, but only after stationarity is achieved. The growth estimator based on (13) is defined with a time lag δt as

$$E_0 = E_{\mathrm{ref}} - \lim_{t\to\infty} \frac{1}{\delta t} \ln \left(\frac{u(t+\delta t, x)}{u(t,x)} \right) \tag{19}$$

independent of x where u is obtained from (17) [13].

3. Correlated Sampling Formalism

Correlated sampling is based on the idea of introducing a correlation between the stochastic processes describing two different systems. The trial functions

for the latter are referred to as $\Psi^{(A)}$ for system A and $\Psi^{(B)}$ for system B. The appropriate DMC runs yield the energies $E^{(A)}$ and $E^{(B)}$, both associated with a certain statistical error, while the property of interest is the energy difference $E^{(B)} - E^{(A)}$ with the variance

$$\begin{aligned}\mathrm{var}\left(E^{(B)} - E^{(A)}\right) = \mathrm{var}\left(E^{(A)}\right) + \mathrm{var}\left(E^{(B)}\right) \\ - 2\mathrm{cov}\left(E^{(A)}, E^{(B)}\right).\end{aligned} \tag{20}$$

A large covariance thus reduces the variance of the difference. The simplest way to assure a large correlation between $E^{(A)}$ and $E^{(B)}$ is to calculate both quantities by means of the same stochastic process. This becomes complicated when different boundary conditions have to be fulfilled.

In this paper, we consider two systems with Hamiltonians differing in the potential but not in the kinetic operator:

$$\hat{H}^{(A)} = -\frac{1}{2}\Delta + V^{(A)} \text{ and } \hat{H}^{(B)} = -\frac{1}{2}\Delta + V^{(B)}. \tag{21}$$

This covers for instance the important application to weak interacting systems where the non-interacting system can be obtained by deleting the interaction part of the potential in the Hamiltonian. The corresponding eigenfunctions are denoted $\{\phi_k^{(A)}\}$ and $\{\phi_k^{(B)}\}$. Transformation of the Hamiltonians with the guide function Ψ_G yields the Hamiltonians $\tilde{H}^{(A)}$ and $\tilde{H}^{(B)}$:

$$\tilde{H}^{(A)} = \Psi_\mathrm{G}^{-1} H^{(A)} \Psi_\mathrm{G} \text{ and } \tilde{H}^{(B)} = \Psi_\mathrm{G}^{-1} H^{(B)} \Psi_\mathrm{G}. \tag{22}$$

The corresponding forward PDEs are

$$\begin{aligned}\frac{\partial u^{(A)}}{\partial t} = -\tilde{H}^{(A)} u^{(A)} = \frac{1}{2}\Delta u^{(A)} + b \cdot \nabla u^{(A)} \\ - \left(E_L^{(A)} - E_\mathrm{ref}^{(A)}\right) u^{(A)},\end{aligned} \tag{23}$$

$$\begin{aligned}\frac{\partial u^{(B)}}{\partial t} = -\tilde{H}^{(B)} u^{(B)} = \frac{1}{2}\Delta u^{(B)} + b \cdot \nabla u^{(B)} \\ - \left(E_L^{(B)} - E_\mathrm{ref}^{(B)}\right) u^{(B)}\end{aligned}. \tag{24}$$

The local energies are calculated as

$$\begin{aligned}E_L^{(A)} &= -\frac{1}{2}\frac{\Delta\Psi_\mathrm{G}}{\Psi_\mathrm{G}} + V^{(A)} \text{ and} \\ E_L^{(B)} &= -\frac{1}{2}\frac{\Delta\Psi_\mathrm{G}}{\Psi_\mathrm{G}} + V^{(B)}.\end{aligned} \tag{25}$$

At this point it is important to recognize that not only the kinetic energy but also the drift $b = \Psi_\mathrm{G}^{-1}\nabla\Psi_\mathrm{G}$ is the same in (23) and (24). Therefore, the generator L and thus the corresponding stochastic processes are identical. This is the basis of the correlated sampling. The solutions $u^{(A)}$ and $u^{(B)}$ can be obtained with the Feynman–Kac formula (see (17))

$$\begin{aligned}u^{(A)}(t,x) = \mathcal{E}^x\left[\exp\left(-\int_0^t \left(E_L^{(A)}(X_s) - E_\mathrm{ref}^{(A)}\right)\mathrm{d}s\right)\right. \\ \left.\cdot\, \chi_{(t<\tau_D)}^{(A)} u(0,X_t)\right],\end{aligned} \tag{26}$$

$$\begin{aligned}u^{(B)}(t,x) = \mathcal{E}^x\left[\exp\left(-\int_0^t \left(E_L^{(B)}(X_s) - E_\mathrm{ref}^{(B)}\right)\mathrm{d}s\right)\right. \\ \left.\cdot\, \chi_{(t<\tau_D)}^{(B)} u(0,X_t)\right],\end{aligned} \tag{27}$$

where it has been assumed that both have the same initial function $u(0,X_t)$. Because both expressions are based on the same stochastic process, the expectations are calculated with the same trajectories X_t assuring high correlation. The expectations differ in the weighting factors and the contributing trajectories due to different boundaries accounted for by the indicator functions $\chi_{(t<\tau_D)}^{(A)}$ and $\chi_{(t<\tau_D)}^{(B)}$. The energies are calculated with the growth estimator (19).

4. Monte Carlo Simulation

Monte Carlo simulations are based on time discretization. We introduce a constant time step Δt with

$$\Delta t = t_{i+1} - t_i. \tag{28}$$

The stochastic process can be realized most easily by means of the Euler scheme:

$$Y_{n+1} = Y_n + \sqrt{\Delta t}\cdot\Delta W + b(Y_n)\cdot\Delta t, \tag{29}$$

where ΔW is a random variable with standard normal distribution. The sequence of realizations of Y_n describes a trajectory. Due to the finite time step there is a bias which is of order $\mathcal{O}(\Delta t)$.

During a Monte Carlo calculation, $u^{(A)}$ and $u^{(B)}$ are obtained in the following manner. A sample of N walkers is propagated according to (29). After each propagation step, the desired quantities are calculated as sample mean

$$u^{(A)}(n\Delta t,x) = \frac{1}{N}\sum_{i=1}^{N} w_i^{(A)}(n\Delta t)\cdot\chi_i^{(A)}\cdot f\left(Y_n^{(i)}\right) \tag{30}$$

and

$$u^{(B)}(n\Delta t,x) = \frac{1}{N}\sum_{i=1}^{N} w_i^{(B)}(n\Delta t)\cdot\chi_i^{(B)}\cdot f\left(Y_n^{(i)}\right), \tag{31}$$

where the initial function $f(x) = u(0,x)$ which is a constant in this work. The term $w_i^{(A)}(n\Delta t)$ represents an approximation to the integral in (26):

$$w_i^{(A)}(n\Delta t) \approx \exp\left(-\int_0^{n\Delta t}\left(E_L^{(A)}(X_s^{(i)}) - E_{\text{ref}}^{(A)}\right)\mathrm{d}s\right) \tag{32}$$

by application of the trapezoidal rule

$$w_i^{(A)}(n\Delta t) = \prod_{k=1}^{n} w_{i,k}^{(A)}(\Delta t) \tag{33}$$

with

$$w_{i,k}^{(A)}(\Delta t) = \exp\left(-\left[\frac{1}{2}\left(E_L^{(A)}\left(Y_k^{(i)}\right) + E_L^{(1)}\left(Y_{k-1}^{(i)}\right)\right) - E_{\text{ref}}^{(A)}\right]\Delta t\right). \tag{34}$$

The error for the integral is of order $\mathcal{O}(\Delta t^2)$. The formulae for system B are analogous.

The energy difference ΔE is calculated by adaptation of the growth estimator [13]. Calculating u at two times $t_1 = t$ and $t_2 = t + n_g\Delta t$ which are n_g steps apart, one obtains for sufficiently long t

$$\Delta E = \Delta E_{\text{Ref}} - \frac{1}{n_g\Delta t}\cdot\ln\left(\frac{u^{(A)}(t+n_g\Delta t,x)\cdot u^{(B)}(t,x)}{u^{(B)}(t+n_g\Delta t,x)\cdot u^{(A)}(t,x)}\right) \tag{35}$$

with $\Delta E_{\text{Ref}} = E_{\text{Ref}}^{(B)} - E_{\text{Ref}}^{(A)}$. After an equilibration period, ΔE is calculated at each step. The mean over all the computed energy differences yields an estimate for ΔE.

The algorithm presented so far has two disadvantages compared to standard DMC calculations, both are associated with setting $\chi_i^{(A,B)} = 0$ for walkers that have crossed the nodes of $\Psi^{(A)}$ and/or $\Psi^{(B)}$. The first problem is due to a decreasing sample size. A look at (30) and (31) illustrates that after a long simulation time most of the walkers have $\chi_i^{(A,B)} = 0$ and therefore just the few which have *survived* still contribute to the estimate. The equilibration time has thus to be chosen carefully and rather than one very long calculation, many short calculations have to be carried out.

The second disadvantage is the time step error. Due to the *killing* of walkers at the boundary, a bias is introduced into the distribution function $u(t,x)$ which scales with $\mathcal{O}(\sqrt{\Delta t})$. Note that this differs from standard DMC where the hits of the boundary vanish as the time step vanishes which is not the case with a positive Ψ_{G}. On the other hand, a positive Ψ_{G} is necessary to allow the stochastic process to sample the full domain of both systems. Since the energy is calculated from $u(t,x)$ its time step error is anticipated to have the same scaling as the function itself. Due to the lack of a suitable proof within the mathematical literature, the latter assumption is going to be verified numerically by plotting the data $E(\Delta t)$ against $\sqrt{\Delta t}$:

$$E(\Delta t) = E(0) + a\cdot\sqrt{\Delta t}, \tag{36}$$

with $E(0)$ being the energy for the time step approaching zero and a a positive or negative constant. This equation holds for $E^{(A)}$, $E^{(B)}$, and ΔE, respectively.

In this context, one has to assess the statistical error of $E(0)$. It should be emphasized that there is no possibility to maintain a constant sample size. Within the algorithm, the boundary condition is fullfilled by deleting those walkers which have crossed it. Since the regions describing system A and B may be different in size, the ratio of the individual populations $(N_A(t)/N_B(t))$ can not be prevented from going to infinity or zero, respectively.

5. Numerical Examples of Correlated Sampling

The following section aims to give a numerical verification of the formulae derived above. In this spirit, the harmonic oscillator is very well suited as model system since the corresponding Schrödinger equation has the form of (1). Because the interest is in systems containing nodes, all calculations being presented in the further course of this paper correspond to the second excited state with quantum number $\upsilon = 2$. Additionally, we set $\hbar = \mu = 1$ and $\omega = \sqrt{k_{\text{f}}}$ where μ refers to the reduced mass and k_{f} to the force constant. We consider two harmonic oscillators A and B with slightly differing frequencies $\omega^{(A)}$ and $\omega^{(B)}$ by means of the same stochastic process. The energy of the second excited state is $E = \frac{5}{2}\omega$ and therefore

$$\Delta E = E^{(B)} - E^{(A)} = \frac{5}{2} \cdot \left(\omega^{(B)} - \omega^{(A)} \right). \quad (37)$$

The location of the two exact nodes is $x = \pm(2\omega)^{-1}$, and the simulation is to be carried out between the two nodes. For the guide function the following ansatz has been chosen:

$$\Psi_G(x) = \left(4\omega_G x^2 - 2\right) \mathrm{e}^{-\frac{\omega}{2}x^2}. \quad (38)$$

If $\omega_G < \min(\omega^{(A)}, \omega^{(B)})$ then Ψ_G is positive at the nodes of both systems A and B. The exponential term in the last equation governs the shape of the function. Hence it is most suggestive to use the mean:

$$\omega = \frac{1}{2}\left(\omega^{(A)} + \omega^{(B)} \right). \quad (39)$$

In the following, we employ $\omega^{(A)} = \frac{1}{18}$ (with nodes at $x = \pm 3$), and $\omega_G = \frac{1}{22}$. $\omega^{(B)}$ will take variable values in the range from $0.05518 = 18.12^{-1}$ to $0.05 = 20^{-1}$.

All of the computations presented below have been conducted with the following parameters: the initial sample size is 100 000 and any calculation comprises a total simulation time of $t = 30$ a.u. The first 10 a.u. are discarded to allow equilibration. For each time step 19 independent computations have been conducted to obtain reliable statistics.

Since the positive guide function leads to a different time-step behaviour compared to standard DMC, we demonstrate the correctness of the approach by calculating the energies of both systems before discussing the correlated sampling results. In Figure 1 the energy is calculated with the growth estimator for a series of

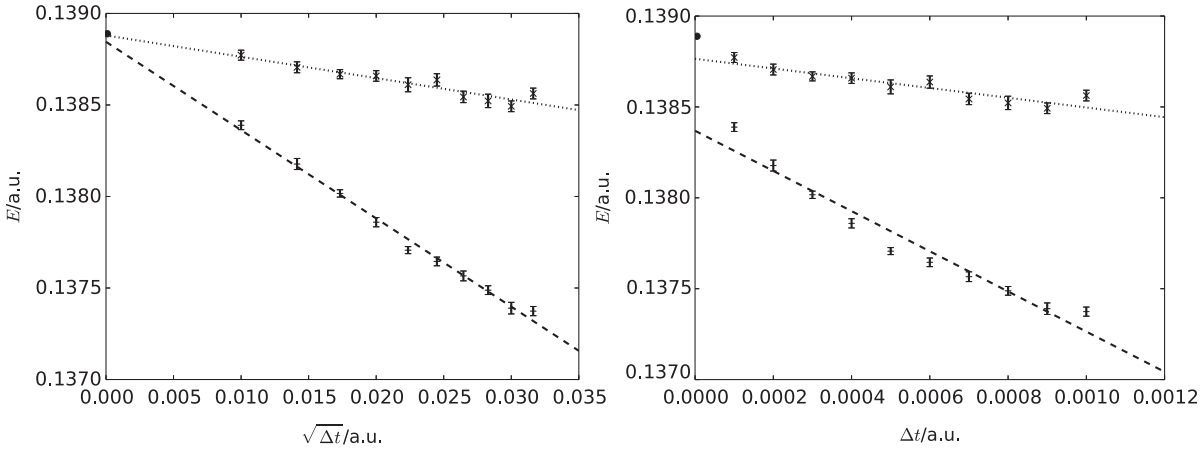

Fig. 1 (colour online). Time step extrapolation for system A: extrapolation E vs. $\sqrt{\Delta t}$ (left) and E vs. Δt (right). Results are shown for the simple estimator (dashed line) and for the improved estimator including the first exit time (dotted line).

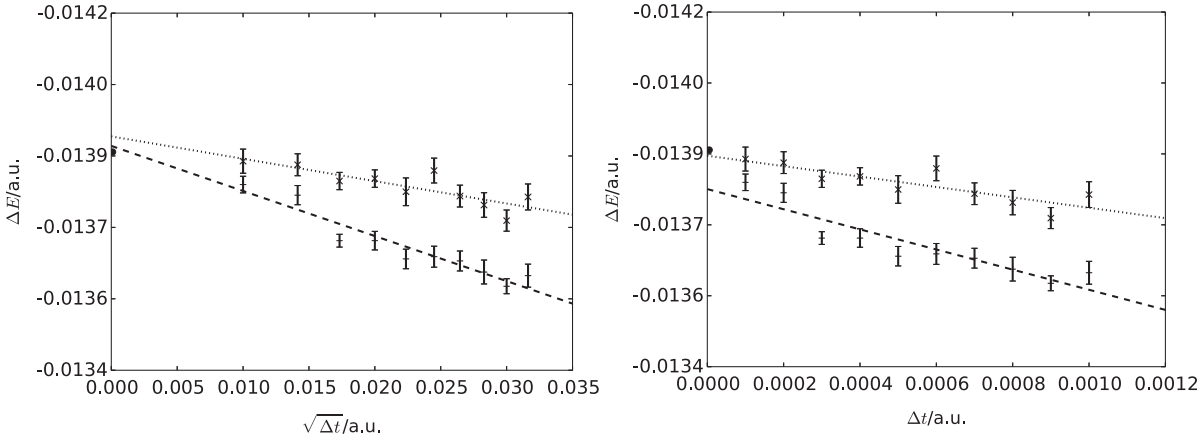

Fig. 2 (colour online). Time step extrapolation for ΔE: extrapolation E vs. $\sqrt{\Delta t}$ (left) and E vs. Δt (right). Results are shown for the simple estimator (dashed line) and for the improved estimator including the first exit time (dotted line).

Table 1. Impact of the node distance Δ_{node} on the standard deviation of the energy difference ΔE. Energy given in E_h, standard deviation in $10^{-5}E_h$. σ and $\sigma^{(\text{cs})}$ denote the standard errors obtained from error propagation and correlated sampling, respectively.

$\left(\omega^{(B)}\right)^{-1}$	ΔE	Δ_{node}	σ	$\sigma^{(\text{cs})}$	$\sigma/\sigma^{(\text{cs})}$
20	−0.013889	0.16228	3.4	2.1	1.6
19	−0.007310	0.08221	3.8	2.0	1.9
18.5	−0.003754	0.04138	4.3	1.3	3.3
18.25	−0.001903	0.02076	4.5	1.1	4.1
18.12	−0.000920	0.00998	4.5	0.8	5.6

time steps. The exact result is indicated with a black dot on the energy axis. It is obvious, that the energy is linear in $\sqrt{\Delta t}$ but not in Δt. The linear extrapolation to zero time step yields 0.138842 ± 0.000028 a.u., compared to the exact result of $E = \frac{5}{2}\frac{1}{18} = 0.138889$ a.u. An improvement of the time-step behaviour will be discussed below.

As a first test for the correlated sampling, we choose $\omega^{(B)} = \frac{1}{20}$ with $E = 0.125$ a.u. and $\Delta E = 0.013889$ a.u. or a tenth of $E^{(A)}$. This is much more than the relative energy difference in typical electron structure problems. With time-step extrapolation, we obtain $E^{(B)} = 0.12494 \pm 0.000019$ a.u. The energy difference obtained from independent calculations is thus $\Delta E = 0.01390 \pm 0.000034$ a.u. The correlated calculations are shown in Figure 2. Again, the linear extrapolation is valid only when plotting ΔE vs. $\sqrt{\Delta t}$. The result is $\Delta E = 0.013898 \pm 0.000021$ a.u. The ratio of the statistical error is 1.6 corresponding to a central processing unit (CPU) time ratio of $1.6^2 = 2.6$.

This ratio is expected to depend on how close the two systems are. Let Δ_{node} define the distance between the nodes of A and B (the same for right and left node). The data in Table 1 show that Δ_{node} has a strong impact on the standard error for correlated sampling $\sigma^{(\text{cs})}$ when the errors for systems A and B are kept nearly constant. As expected, the standard error of the correlated sampling energy differences decrease quickly as system B gets closer to system A. Note that the CPU ratios are the squared error ratios.

6. Improving the Time-Step Behaviour

The main drawback of the formalism presented so far is that the random trajectory does not contain any information about a possible vicinity of the boundary [17]. This gives rise to a poor sampling near the nodes. A major reason is the so called *recrossing error* that has first been described by Anderson [18]. It is based on the assumption that within the time step Δt a certain trajectory could have already left and then reentered the region defined by the nodes. A suitable treatment of the recrossing error corresponds directly to the calculation of the first exit time, the time the trajectory hits the node for the first time. To calculate the first exit time, one has to account for two different possibilities [19, 20]:

(i) Both the starting and the end point of the trajectory are within the region enclosed by the boundary (first exit time is designated $\mathcal{T}_1$).
(ii) The starting point resides within but the end point of the trajectory outside the region enclosed by the boundary (first exit time is designated $\mathcal{T}_2$).

A detailed mathematical derivation exists only for b in (3) being constant everywhere.

Let

$$z = y + b(y)\cdot\Delta t + \Delta W\cdot\sqrt{\Delta t}\,, \tag{40}$$

i. e. y is a realization of Y_n and z the value of Y_{n+1} if no crossing occurred. The resulting recrossing probability is known for a constant right boundary at c (i. e. $y < c$, $z < c$, and $z > y$)

$$\mathcal{T}_1 = -\frac{2(c-y)(c-z)}{\log\xi}\,. \tag{41}$$

For $z < y$, c is the corresponding left boundary. ξ denotes a uniform random variate between 0 and 1. If $\mathcal{T}_1 < \Delta t$, the diffusion is killed, if $\mathcal{T}_1 > \Delta t$, the new value z is accepted ($Y_{n+1} := z$) [19, 20].

The first exit time for the second case is expressed as follows [20]:

$$\mathcal{T}_2 = \frac{\Delta t\cdot s}{1+s} \quad \text{with} \quad s \sim \mathcal{IG}[\mu,\lambda]\,, \tag{42}$$

where $\mathcal{IG}$ represents an inverse gaussian distribution with its parameters being calculated as

$$\mu = \frac{(c-y)^2}{\Delta t} \tag{43}$$

and

$$\lambda = \frac{c-y}{z-c}\,. \tag{44}$$

The first exit time is included into the calculation of the path integral as

$$w_i^{(A)}(k,\mathcal{T}_1) = \exp\left[-\left(E_L^{(A)}\left(Y_{k-1}^{(i)}\right) - E_{\mathrm{ref}}^{(A)}\right)\mathcal{T}_1\right], \quad (45)$$

$$w_i^{(A)}(k,\mathcal{T}_2) = \exp\left[-\left(E_L^{(A)}\left(Y_{k-1}^{(i)}\right) - E_{\mathrm{ref}}^{(A)}\right)\mathcal{T}_2\right], \quad (46)$$

where the $\mathcal{T}_1$ weight is employed when the path is killed. $w_i^{(B)}(k,\mathcal{T}_1)$ and $w_i^{(B)}(k,\mathcal{T}_2)$ are calculated in the same manner.

It should be emphasized that the impact of $\mathcal{T}_1$ is much larger than the one of $\mathcal{T}_2$. Since the latter is calculated for trajectories which are about to be deleted, $\mathcal{T}_2$ provides an one-time correction of the path integral (see (46)). On the other hand, the magnitude of $\mathcal{T}_1$ governs if a certain trajectory is to be deleted or not.

In Figures 1 and 2, the energies for the improved weighting scheme are plotted in comparison with the simple scheme. One can see that the recrossing correction reduces the time step error by a factor of about five. Judging from the plot of the energy vs. Δt it appears that the $\sqrt{\Delta t}$ dependence has almost but not completely vanished. For the energy difference ($\omega^{(B)} = \frac{1}{20}$) the linear extrapolation with Δt yields the correct result possibly due to a cancellation of the $\sqrt{\Delta t}$ dependence. From the linear extrapolation we obtain with correlated sampling 0.0139326 ± 0.000028 a.u. at zero time step while we get 0.013938 ± 0.000034 a.u. based on the difference of the extrapolated energies.

7. Conclusion and Outlook

A new approach for correlated sampling in DMC within the fixed node approximation based on the *Kolmogorov backward equation* has been presented. The resulting algorithm has been implemented and tested for the harmonic oscillator. According to the fixed node approximation, walkers having crossed a node are deleted from the sample. The resulting time step error is proportional to $\mathcal{O}(\sqrt{\Delta t})$ due to the positive guide function. A possibility to reduce that error is given by accounting for the first exit time, namely $\mathcal{T}_1$ and $\mathcal{T}_2$. Especially $\mathcal{T}_1$ has a significant influence since it directly corrects the walker distribution in the vicinity of the nodes. While the time step error is greatly reduced, the scaling with $\mathcal{O}(\sqrt{\Delta t})$ does not vanish for calculations of the system individually. However, for the energy difference, being the quantity of interest, the time step error is not only greatly reduced but also shown to be proportional to Δt. The model calculations of this paper demonstrate the feasibility of the correlated approach for energy difference using the forward equation. The extension of the method to electronic ground states of molecules is currently investigated. The proposed recrossing correction will be necessary for successful correlated sampling calculations.

[1] A. Lüchow, WIREs Comput. Mol. Sci. **1**, 388 (2011).
[2] M. H. Kalos and P. A. Whitlock, Monte Carlo Methods, Wiley Interscience, New York 1986.
[3] R. E. Lowther and R. L. Coldwell, Phys. Rev. A **22**, 14 (1980).
[4] C. J. Umrigar, Int. J. Quantum Chem. Symp. **23**, 217 (1989).
[5] C. Filippi and C. J. Umrigar, Phys. Rev. B **61**, R16291 (2000).
[6] B. L. Hammond, W. A. Lester, and P. J. Reynolds, Monte Carlo Methods, in: Ab Initio Quantum Chemistry (Ed. S. H. Lin), World Scientific Publishing Co. Pte. Ltd., Singapore 1994.
[7] D. M. Ceperley and B. J. Alder, J. Chem. Phys. **81**, 5833 (1984).
[8] P. E. Kloeden and E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, Berlin 1999.
[9] B. Øksendal, Stochastic Differential Equations, Springer, Berlin 2007.
[10] H. Risken, The Fokker–Planck Equation, Springer, Berlin 1989.
[11] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[12] M. Kac, Trans. Amer. Math. Soc. **65**, 1 (1949).
[13] A. Korzeniowski, J. L. Fry, D. E. Orr, and N. G. Fazleev, Phys. Rev. Lett. **69**, 893 (1992).
[14] P. J. Reynold, R. N. Barnett, B. L. Hammond, and W. A. Lester, Jr., J. Stat. Phys. **43**, 1017 (1986).
[15] I. Bosá and S. M. Rothstein, J. Chem. Phys. **121**, 4486 (2004).
[16] M. Freidlin, Functional Integration And Partial Differential Equations, Princeton University Press, Princeton 1985.
[17] R. Mannella, Phys. Lett. A **254**, 257 (1999).
[18] J. B. Anderson, J. Chem. Phys. **65**, 4121 (1976).
[19] F. M. Buchmann, J. Comp. Phys. **202**, 446 (2005).
[20] F. M. Buchmann and W. P. Peterson, Siam. J. Sci. Comput. **28**, 1153 (2006).