

Improved diffusion Monte Carlo for bosonic systems using time-step extrapolation “on the fly”

P. Håkansson and M. Mella^{a)}

School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, United Kingdom

(Received 11 December 2006; accepted 19 January 2007; published online 13 March 2007)

A diffusion Monte Carlo algorithm employing “on the fly” extrapolation with respect to the time step is implemented and demonstrated simulating realistic systems. Significant advantages are obtained when using on the fly extrapolation, leading to reduced systematic and statistical errors. The sound theoretical basis of extrapolation on the fly is discussed and compared to justifications for the *a posteriori* extrapolation. © 2007 American Institute of Physics. [DOI: 10.1063/1.2647227]

I. INTRODUCTION

Diffusion Monte Carlo (DMC) is a valuable computational tool in the study of (0 K) bosonic systems allowing “exact” ground state properties to be computed.¹ DMC normally consists of two main parts, importance sampling (IS), with points diffusing in configuration space guided by a trial wave function (ψ_T), and branching, which changes number of configurations (walkers) according to a computed weight. Whereas IS greatly reduces the statistical error, the branching ensures that ground state properties are sampled in the limits $t \rightarrow \infty$ and $\tau \rightarrow 0$ (where τ is the time step in the simulation and t is the length of the imaginary time). With the trial wave function introduced, the Schrödinger equation takes the form²

$$-\frac{\partial f(x,t)}{\partial t} = -\underbrace{\frac{1}{2}\nabla^T D \nabla f(x,t) + D \nabla \cdot [F(x)f(x,t)]}_{=Lf(x,t)} + E_L(x)f(x,t), \quad (1)$$

where $f(x,t) = \psi_0 \psi_T$, $E_L(x) = \psi_T(x)^{-1} H \psi_T(x)$ is the local energy, H is the N -dimensional Hamiltonian, $F(x) = \psi_T(x)^{-1} \nabla \psi_T(x)$ is the quantum force (QMF), and D is a diagonal diffusion matrix. L , acting on $f(x,t)$, is the Fokker-Planck (Schmoluchowski) operator whose stationary solution is $\psi_T(x)^2$ and is responsible for the importance sampling during a DMC simulation. In obtaining a numerical scheme to sample Eq. (1), a common approach is to start from the formal solution of Eq. (1) and let the exponential operator, $\exp(-\tau(L+E_L))$, determine the structure of the stochastic differential equations (SDEs) to be solved.^{3,4} An alternative starting point is to consider the following set of stochastic equations:⁵

$$X_{n+1}^{(k)} = X_n^{(k)} + \int_{t_n}^{t_{n+1}} F^{(k)}(X_s) ds + \int_{t_n}^{t_{n+1}} dW_s^{(k)},$$

$$I_{n+1} = I_n + \int_{t_n}^{t_{n+1}} E_L(X_s) ds, \quad (2)$$

where X is a vector of dimension N representing the position of all particles in the system (a walker), W_t is a Brownian motion in \mathbb{R}^N , and I_n is a random variate that depends on the realization of W_t . Solving Eq. (2) does provide a solution to Eq. (1) as can be proven using the Feynman-Kac and Itô formulas.⁵ Note that Eq. (2) represents an $(N+1)$ -dimensional Itô equation that can be approximated using a weak numerical scheme.⁵ This way of rephrasing the DMC algorithm is exactly equivalent to the one based on the Baker-Campbell-Hausdorff formula.⁶ If branching is done at each time step τ ($\tau = t_{n+1} - t_n$), then an initial condition $I_n = 0$ is imposed and the weight $w = \exp(-I_{n+1})$ is used to kill/duplicate walkers in the population. In this respect, a recent work of Håkansson *et al.*⁷ compared second order numerical schemes in the time step τ , either consistent with an Itô-Taylor expansion⁵ of Eq. (2) or based on exponential formulas, that could be used to solve Eq. (2). There, the explicit second order (E2) and the predictor corrector (PC2) schemes, both based on the Itô-Taylor expansion, were found to have smaller time step bias and longer range of convergence compared to other second order schemes proposed in the literature.⁷ This finding allowed us to propose these schemes as suitable algorithms to solve the Schmoluchowski equation in the DMC method, thus simulating more efficiently bosonic systems at 0 K. In this respect, it is perhaps worth pointing out that the application of similar approaches to fermionic systems is hampered by few theoretical difficulties, mainly associated with nonanalyticities in $\ln(\psi_T)$ such as the divergence of the quantum force on a nodal surface and the nonanalytical behavior due to the presence of a cusp. As for the latter, no problem appears to be present when one limits himself to the calculation of the expectation value of the local energy using a ψ_T that exactly satisfies the cusp conditions. Even so, the convergence of the sampled density in a region around a nonanalytical point does not follow the polynomial order with respect to the time step predicted using theoretical considerations.⁴

In order to improve the performance of quantum Monte Carlo simulations avoiding the necessity of using small time

^{a)}Electronic mail: mellam@cardiff.ac.uk

steps to obtain expectation values with a small discretization error, at least two approaches are available. On one hand, one could employ a time-independent algorithm such as Green's function MC proposed by Kalos *et al.*⁸ and recently revisited by Schmidt *et al.*⁹ Alternatively, and it is the path we will follow in this work, one may attempt to further improve the order of the discretization error (time-step bias), considering numerical schemes of order higher than two derived via the exponential operator^{10,11} or via an Itô-Taylor expansion.⁵ In latter two cases, however, this will require the evaluation of derivatives of quantum force/local energy (or potential energy if no importance sampling is done), with the net effect of increasing the complexity of the code, and perhaps the cost of the simulations as well.

In this respect, an attractive alternative is to consider extrapolation methods such as the Richardson-Romberg extrapolation, where two or more simulation results (done with time steps of different lengths) are linearly combined so that the lowest order error cancels.¹² There are, however, two issues in applying such extrapolation, and these are addressed in this paper. First, there is a computational overload connected with obtaining a small statistical error if simulation results are extrapolated *a posteriori*, i.e., combining independent simulation results. This originates from the fact that, for a given computational effort, the shortest time-step simulation will be the least precise and also have the largest prefactor in the extrapolation [cf. Eq. (3)]. So, the statistical error for the computed result may be significantly scaled up by the extrapolation. Besides, only an empirical foundation appears to be currently available for extrapolating *a posteriori* DMC results (including branching). Instead, the extrapolation of expectation values computed from an integrated SDE is mathematically motivated by a global error expansion,¹³ which can be applied to our specific instance, DMC simulations. To show that this is exactly the case, we specialize our discussion to the extrapolation of second order schemes and start by considering an initial distribution of walkers at time t_n , fulfilling the accuracy requirement $|\text{error}_{t_n}| \leq C\tau^2$, $C > 0$, where error_{t_n} refers to the error for any expectation value computed over the walker distribution. A central component in demonstrating the order τ^γ ($\gamma > 2$) of extrapolated result is proving the existence of an error expansion $|\text{error}_{t_{n+1}} - \{e_2(t_{n+1})\tau^2 + e_3(t_{n+1})\tau^3 + \dots + e_{\gamma-1}(t_{n+1})\tau^{\gamma-1}\}| \leq C(t_{n+1})\tau^\gamma$, where e_i and C are real numbers independent of time step τ . This is what is done in Refs. 12 and 13 considering the continuous Itô process given in our case by the $(N+1)$ -dimensional equation [Eq. (2)]. Notice that, in this way, the error expansion need *not* consider the discontinuity that occurs when branching is employed at time t_{n+1} because the calculation of expectation values could be made with the proper discretized form of $\exp(-I_{n+1})$. Thus, if the integration scheme is formally second order (e.g., E2 and PC2), the distribution at time t_{n+1} would satisfy the error expansion in Refs. 12 and 13, hence allowing for a Richardson-Romberg extrapolation with respect to τ of both X_{n+1} and I_{n+1} . It is now necessary to stress that the application of the branching procedure after having extrapolated X_{n+1} and I_{n+1} does not introduce further complications since it induces no change of order in $|\text{error}_{t_{n+1}}|$.

As a consequence of the above discussion, the aim of this work is to illustrate how the problems with an *a posteriori* extrapolation can be circumvented by implementing what we call extrapolation "on the fly" with respect to the time step. Here, the discussion is limited only to the case of two solutions (both with second order time-step error) starting from the same configuration Y_n and propagated for the interval τ , while driven by the same Brownian motion (W_t). To extrapolate to third order on the fly, a single time step τ is computed by first generating random vectors (δW and $\delta \tilde{W}$) that are normally distributed with variance $\tau/2$ and mean zero and computing $\Delta W = \delta W + \delta \tilde{W}$. The walker initial position is then propagated in two ways, one using ΔW and the other in a two-step process using δW and $\delta \tilde{W}$. The branching weight and expectation values are then computed using the stochastic variates

$$w_3(t_{n+1}) = 1/3\{4w_2^{\tau/2}(t_{n+1}) - w_2^\tau(t_{n+1})\}, \quad (3)$$

$$\langle A \rangle = \sum_i w_{3,i} A_{3,i} / \sum_i w_{3,i}, \quad (4)$$

where subscript denotes the order of discretization error, superscript the time step used, and the sums in Eq. (4) extend over all simulated configurations. Before the next time step, final (t_{n+1}) walker positions are extrapolated to third order [Eq. (3)] and finally the branching is performed. In Eq. (3), both processes are highly correlated and this may significantly reduce the statistical error when compared with an *a posteriori* extrapolation, as illustrated in the following. Furthermore, by extrapolating before the branching step we make sure that the mathematical foundations for the extrapolated results are valid (see Ref. 12 for error expansion of expectation values computed with higher order SDE integrator).

In this work, we apply the extrapolation on the fly to solve high dimensional problems using DMC. This scheme has never been applied to DMC simulations before and the efficiency gain that may be obtained in comparison with *a posteriori* extrapolation was never evaluated; to illustrate the performance of the algorithm, examples from cluster calculations with [Eq. (1)] or without importance sampling are used. In the latter case, the direct sampling of the Schrödinger equation,

$$-\frac{\partial \psi(x,t)}{\partial t} = -\frac{1}{2} \nabla^T D \nabla \psi(x,t) + V(x)\psi(x,t), \quad (5)$$

is performed without a trial wave function guiding the distribution. This latter case is of interest in situations when accurate ψ_T are difficult to find.¹¹ All results presented are calculated without the Metropolis acceptance/rejection step. This has the advantage that the dependence on τ is known for all calculated expectation values and one can deduce when discretization error (dependent on τ) is less than the desired standard error. A discussion on the relative performance of schemes including the Metropolis acceptance/rejection step was given in Ref. 7.

As a last comment in this Introduction, we note that it has been shown previously that *a posteriori* extrapolation

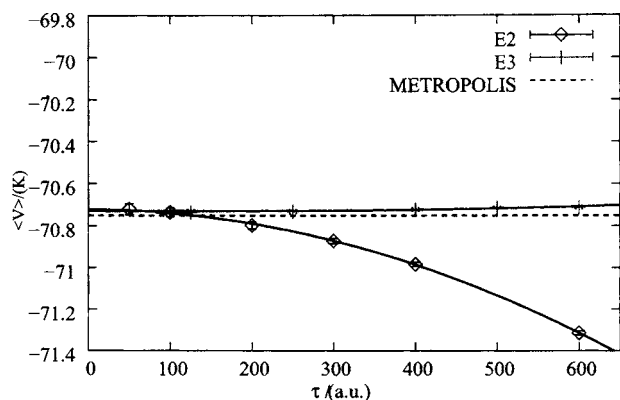


FIG. 1. Expectation value of the interaction potential $\langle V \rangle$ (K) as a function of the time step (a.u.) for a VMC simulation of MgHe₁₂. The Metropolis result with standard error (± 0.02) (dashed line) represents the limiting value for $\tau \rightarrow 0$ that should be obtained by all simulation algorithms. The continuous lines are quadratic (E2) or cubic (E3) fits to the numerical results.

can reduce discretization error in path integral quantum Monte Carlo simulations¹⁴ and DMC.⁷ Thus, one of the aims of this report is, instead, to investigate the effectiveness of the on the fly extrapolation in the framework of DMC simulations, as well as to evaluate its efficiency when compared with *a posteriori* extrapolation. In this respect, we also notice that one dimensional examples of extrapolating on the fly SDE simulations were already provided in Refs. 5 and 15, where the on the fly method implemented here (i.e., using different solutions driven by the same W_i) can be found as a coded example.¹⁵ However, the efficiency gain that could be obtained in realistic simulations with respect to *a posteriori* extrapolation has never been observed or commented upon.

II. METHODS, RESULTS, AND DISCUSSION

In this work, second order schemes and extrapolation to third order are considered, but the methodology is general and it should be obvious how extrapolation to an order higher than three may be carried out following the proposed scheme using a higher order formula.¹²

A. Importance sampled methods

The main ingredient for the extrapolation is an efficient second order numerical scheme and we suggest to start from the numerical schemes, E2 or PC2, which have proven to have second order convergence for a wide range of time steps.⁷ In this work, we used the E2 scheme and a trial wave function for MgHe₁₂ with a form taken from Ref. 16 but with parameters slightly modified to generate a less smooth local energy. This increases the time-step error helping to clarify the difference between numerical schemes; it also mimics the expected error increase obtained in larger systems due to less accurate ψ_T .

To show the effect of on the fly extrapolation on the integration of the Schmoluchowski equation [i.e., Eq. (1) with no branching] for the above system, the potential energy expectation value computed using variational Monte Carlo (VMC) simulations is shown in Fig. 1 and compared with the exact Metropolis result for the system MgHe₁₂. In this case, all weights $w_3(t_{n+1})$ are identically equal to 1, and the ex-

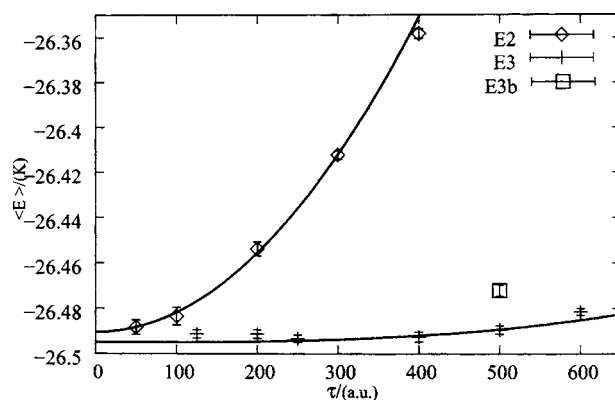


FIG. 2. Expectation value of the local energy $\langle E_L \rangle$ (K) as a function of the time step (a.u.) for DMC simulations of MgHe₁₂. The continuous lines are quadratic (E2) and cubic (E3) fits to the numerical results. The single data point E3b has six QMFs per time step (see text).

trapolation [cf. Eq. (3)] is applied to the potential energy and the walker coordinates (as discussed below). Clearly, the explicit third order (E3) extrapolation on the fly using the E2 scheme does provide third order convergence and gives a significant improvement, as compared with the E2 result.

Figure 2, instead, shows the convergence properties of the local energy expectation value computed as a function of time step obtained using importance sampled DMC simulations on the same MgHe₁₂ cluster. In this respect, the E2 scheme requires two QMF calculations/time step (the major computational bottleneck); thus the third order extrapolation of the weights requires six QMF calculations/time step if we follow Ref. 5. However, this strategy would give walker positions accurate only to second order; even so, the improvement obtained by extrapolating the weight may still be substantial. In this work, the final walker configuration is extrapolated to third order and stored as the initial position for the next time step (E3 scheme); this adds another QMF calculation stored as initial QMF in the next time step. With seven QMFs, third order convergence is obtained (see E3 in Fig. 2); this is also the case with six QMFs, even though a larger prefactor is obtained, as it is illustrated by the data point E3b in Fig. 2. Given the substantial improvement, in our opinion, the seventh QMF calculation is well worth the extra effort. Note that a time step roughly ten times longer can be used in E3 as compared with E2, translating to an efficiency gain of ≈ 2.8 due to a faster decorrelation of the data.

To discuss the relative efficiency of the on the fly extrapolation with respect to the *a posteriori* one, we start by mentioning that the standard errors $\sigma(\tau)$ in E2 and E3 simulations with equal number of configurations are (empirically) found to be almost identical. Note also that approximately the double number of blocks will be needed to make $\sigma_{\tau/2}^2$ equal to σ_τ^2 . From this, the extra effort needed in a third order *a posteriori* extrapolation can be derived and compared to an E3 simulation. Consider now the statistical error propagation in an *a posteriori* extrapolation

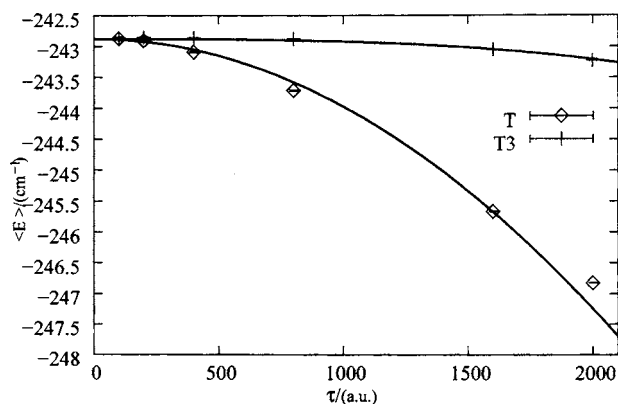


FIG. 3. Expectation value of the total energy $\langle E \rangle$ (cm^{-1}) of Ne_7 as a function of the time step (a.u.) for the T and T3 DMC algorithms. The continuous lines are quadratic (T) or cubic (T3) fits to the numerical results.

$$\sigma_{\text{extr}} = \sqrt{\frac{16}{9}\sigma_{\tau/2}^2 + \frac{1}{9}\sigma_{\tau}^2}. \quad (6)$$

If the computer resources were distributed wisely for it, it is deduced that we should use four times more blocks of E2 $\tau/2$ simulations to make σ_{extr} equal to σ_{E3} . If instead the number of blocks in the two E2 simulations is similar, there should be 11 or 3 times more blocks in both E2 simulations, as opposed to an E3 simulation to make $\sigma_{\text{extr}} = \sigma_{\text{E3}}$. Finally, accounting for the number of force evaluations in the two approaches, an efficiency gain of 1.4–2.1 is deduced for the E3 simulations. Furthermore, it is seen in Fig. 2 that E2 simulations were possible only if $\tau \leq 400$ a.u., thus limiting *a posteriori* extrapolation to a smaller range of time steps than can be used on the fly.

B. Simulations without importance sampling

In sampling Eq. (5), the weight is computed using a SDE equivalent of the Trotter (T) formula,³

$$Y_{n+1} = Y_n + \Delta W, \quad (7)$$

$$w_{n+1} = \exp(-\tau/2\{V(Y_{n+1}) + V(Y_n)\}), \quad (8)$$

this scheme having formally a second order discretization error.⁵ We note that, in simulating Eq. (7) using two solutions with time step τ and $\tau/2$ on the fly, the final position of the walkers will be exactly the same; as a consequence, only three potential values/time step are needed. In Fig. 3 the energy expectation value for a Ne_7 cluster is shown. In these simulations, population bias was circumvented with the use of a sufficiently large population of 30 000 walkers.¹¹ It is seen that the Trotter (T) scheme follows second order convergence and that the third order “extrapolation on the fly” (T3) follows third order convergence behavior. A time step of 800 a.u. in T3 the simulation gives an equally well converged result as a T time step of 100 a.u. Whereas in the T3

simulation there are three potential energy calculations in each time step, only one is done in the T simulation; this translates into 2.7 times less computational effort for the T3 scheme to reach the same standard error. We have also verified that similar gains are present when simulating molecular complexes such as $\text{H}^+(\text{H}_2\text{O})_n$.¹⁷

III. CONCLUSIONS

In this work, the efficiency gain obtained using a third order extrapolation “on the fly” (E3 and T3) is quantified with numerical examples. A substantial gain provided by extrapolated schemes, as compared to both underlying (second order) schemes and to *a posteriori* extrapolation, is confirmed. The extrapolation methodology used does provide a general and efficient class of algorithms and will be valuable not only for the specific DMC methods tested here. The related reptation algorithm¹⁸ is a method where the extrapolation on the fly is directly applicable. Furthermore, whenever a direct computation of the path integral Monte Carlo (PIMC) density matrix is of interest [see Eq. (4.1) in Ref. 19], the stochastic process [Eq. (7)] can be replaced with a Brownian bridge and the density matrix obtained using extrapolation on the fly. Similarly, the PIMC approach (see Appendix A in Ref. 20), where a trial density matrix is used, can benefit from the use of extrapolation on the fly.

ACKNOWLEDGMENT

One of the authors (M.M.) acknowledges EPSRC, Grant Nos. EP/C528301/1 and GR/R77803/2.

- ¹J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).
- ²P. J. Reynolds, D. M. Ceperley, B. J. Alder, and J. W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
- ³M. Suzuki, J. Math. Phys. **26**, 601 (1985).
- ⁴S. A. Chin, Phys. Rev. A **42**, 6991 (1990).
- ⁵P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, New York, 1992).
- ⁶H. Kobayashi, N. Hatano, and M. Suzuki, Physica A **250**, 535 (1998).
- ⁷P. Håkansson, M. Mella, D. Bressanini, G. Morosi, and M. Patrone, J. Chem. Phys. **125**, 184106 (2006).
- ⁸M. H. Kalos, D. Levesque, and L. Verlet, Phys. Rev. A **9**, 2178 (1974).
- ⁹K. H. Schmidt, P. Niyaz, A. Vaught, and M. A. Lee, Phys. Rev. E **71**, 016707 (2005).
- ¹⁰H. A. Forbert and S. A. Chin, Phys. Rev. B **63**, 144518 (2001).
- ¹¹S. Chiesa, M. Mella, G. Morosi, and D. Bressanini, J. Chem. Phys. **119**, 5601 (2003).
- ¹²P. E. Kloeden, E. Platen, and N. Hofmann, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **32**, 1919 (1995).
- ¹³D. Talay and L. Tubaro, Stoch. Anal. Appl. **8**, 483 (1990).
- ¹⁴L. Brualla, K. Sakkos, J. Boronat, and J. Casulleras, J. Chem. Phys. **121**, 636 (2004).
- ¹⁵P. E. Kloeden, E. Platen, and H. Shurz, *Numerical Solution of SDE Through Computer Experiments* (Springer, New York, 1997).
- ¹⁶M. Mella, G. Calderoni, and F. Cargnoni, J. Chem. Phys. **123**, 054328 (2005).
- ¹⁷M. Mella and D. C. Clary, J. Chem. Phys. **119**, 10048 (2003).
- ¹⁸S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4575 (1999).
- ¹⁹D. M. Ceperley, Rev. Mod. Phys. **67**, 2 (1995).
- ²⁰E. L. Pollock and D. M. Ceperley, Phys. Rev. B **30**, 5 (1984).