Diffusion Monte Carlo: Exponential scaling of computational cost for large systems

Norbert Nemec
Department of Physics, University of Cambridge, CB3 0HE Cambridge, United Kingdom
(Received 2 June 2009; revised manuscript received 9 December 2009; published 26 January 2010)

The computational cost of a Monte Carlo algorithm can only be meaningfully discussed when taking into account the magnitude of the resulting statistical error. Aiming for a fixed error per particle, we study the scaling behavior of the diffusion Monte Carlo method for large quantum systems. We identify the correlation within the population of walkers as the dominant scaling factor for large systems. While this factor is negligible for small and medium sized systems that are typically studied, it ultimately shows exponential scaling. The scaling factor can be estimated straightforwardly for each specific system and we find that is typically only becomes relevant for systems containing more than several hundred atoms.

DOI: 10.1103/PhysRevB.81.035119 PACS number(s): 02.70.Ss, 71.15.Nc, 31.15.−p

I. INTRODUCTION

Today's scientists can choose from a wide range of computational methods for the simulation of quantum mechanical systems. These range from highly efficient semiempirial methods to density functional methods—offering a practical compromise of efficiency and accuracy—all the way to very accurate quantum chemical methods. Besides these deterministic methods, various stochastic quantum Monte Carlo (QMC) methods are gaining ground, offering exact handling of many strongly correlated systems and scaling up to system sizes that are out of reach for the deterministic competitors.

The two major arguments that are typically brought up in the advocacy of QMC are the excellent parallelizability and the good scaling behavior. Depending on the QMC variant that is chosen, the collection of statistical data points can be performed in parallel with little to no communication, making the method well suited for high performance computers of all architectures. The scaling behavior depends greatly on the details of the system and the method, but it is generally found to be significantly better than that of quantum chemical methods and linear scaling algorithms have been reported.[1–8]

A commonly used method for the *ab initio* simulation of electronic structure is diffusion Monte Carlo (DMC),[9–11] typically using the fixed node approximation.[12] For this method, the bulk of the computational effort is spent on the repeated evaluation of a trial wave function for electron positions that change step by step, one electron at a time. The trial wave function is usually expressed as a Slater determinant[13] of single electron orbitals, multiplied by a Jastrow factor[14] to express electron correlations. For single electron orbitals expressed as maximally localized Wannier functions,[15] the local energy can be re-evaluated in constant time after a single electron move, leading to an $O(N)$ algorithm for a complete time step of one configuration.[1–8] As a further refinement to this, trial wave functions for DMC calculations are today commonly expressed in a blip basis,[16] which can be evaluated very efficiently.

In contrast to deterministic methods, however, the computational cost of a Monte Carlo (MC) simulation is meaningless without specifying the statistical error that is achieved. Deterministic methods typically have systematic errors that are either intrinsic or depend on parameters that do not scale with the system size. The statistical error of MC simulations on the other hand scales very simply with the inverse square root of the run time while the scaling with the system size is a nontrivial issue that depends on details of the method and the system of study. Though the unfavorable scaling of the statistical efficiency of DMC has been demonstrated before,[17] it has—to our knowledge—never been studied systematically.

In this paper, we will present a systematic study of the scaling behavior of QMC calculations aiming for a fixed statistical error bar. The main focus will be on the DMC algorithm including branching and population control as described by Umrigar *et al.*,[18] other variants will be briefly discussed as well. The statements that we will derive are expected to hold for DMC calculations in general, but to simplify understanding; we will consider a "typical" system made up of N similar constituents which we simply call "atoms." This could be, for example, a crystal in a simulation cell made up of N primitive cells, a cluster of N atoms or a large organic molecule of N comparable groups.

We will begin by deriving several general quantities and continue by demonstrating these in the case of a simple model of N independent hydrogen atoms. From this model we can numerically extract the missing pieces of the scaling behavior, allowing a quantitative estimate of the scaling limit for arbitrary systems. This limit will then be discussed for a number of different sample systems.

II. SCALING OF COMPUTATIONAL COST

The total computational cost of a DMC calculation (optionally split over a number of parallel CPUs) is

$$t_{\text{total}} = N_{\text{step}} \times N_{\text{pop}} \times t_{\text{step}}, \tag{1}$$

where N_{step} is the number of steps in imaginary time, N_{pop} is the average population size and t_{step} is the CPU time needed for one single all-electron step per configuration. Using a so-called "linear scaling" QMC algorithm,[1] each all-electron move scales as

$$t_{\mathrm{step}} \propto N,$$

assuming that the evaluation of the Slater determinant dominates the computational cost. For the moment, we assume that population fluctuations are negligible and N_{pop} can be treated as an external parameter. The influence of the population control will be discussed later on.

The standard error of the total energy can be expressed as

$$\delta E_{\mathrm{total}} = \sqrt{\left(\frac{\tau_{\mathrm{corr}}}{\tau_{\mathrm{step}}}\frac{1}{N_{\mathrm{step}}}\right)\left(\frac{\chi_{\mathrm{pop}}}{N_{\mathrm{pop}}}\right)\sigma_{\mathrm{dmc}}^2}, \tag{2}$$

with the constituents explained in the following.

The raw DMC variance σ_{dmc}^2 is the variance of the local energy of individual configurations over the whole simulation. Being based on the mixed estimator, σ_{dmc}^2 may deviate from the variance σ_{vmc} of the trial wave function obtained in a VMC run. For typical systems, however, we find that $\sigma_{\mathrm{dmc}}^2 \approx \sigma_{\mathrm{vmc}}^2$. Assuming that the trial wave function of the whole system can be optimized to about the same quality as that of a single constituent N, this variance scales as

$$\sigma_{\mathrm{dmc}}^2 \propto N, \tag{3}$$

since the local energy is dominated by the sum of N independent atomic local energies.

The (integrated) correlation time of a series of data points x_i is given by[19]

$$\tau_{\mathrm{corr}} = \tau_{\mathrm{step}}\left(1 + 2\sum_{j=1}^{\infty}\frac{\langle x_i x_{i+j}\rangle_i - \langle x_i\rangle_i^2}{\langle x_i^2\rangle_i - \langle x_i\rangle_i^2}\right), \tag{4}$$

where $\langle\cdot\rangle_i$ denotes the arithmetic mean over the index i. τ_{corr} in units of the time step τ_{step} describes the factor by which the number of steps N_{steps} has to be divided to correct the error bar of the result for serial correlation. To obtain $\tau_{\mathrm{corr}}/\tau_{\mathrm{step}}$, as an alternative to computing Eq. (4) directly, one can also use the reblocking method.[20] Though the system may have various correlation time scales, some of which depend on the system size, we find that the integrated correlation time that is responsible for the reduction in the resulting accuracy is dominated by the shortest correlation times which depend on local properties, such as the kind of nuclei in the system, but not on the size N.

The population correlation factor $\chi_{\mathrm{pop}} \geq 1$ captures the inefficiency of the process due to population correlation and fluctuation. We will treat this factor as an unknown quantity for the moment and discuss it in detail afterwards.

Using Eqs. (1) and (2) along with the discussed scaling laws, we can express the scaling of the total computational cost as

$$t_{\mathrm{total}} \propto \frac{\chi_{\mathrm{pop}}}{\delta E_{\mathrm{atom}}^2}, \tag{5}$$

so we see that—apart from the factor χ_{pop}—DMC is in fact a constant scaling method if a fixed standard error per atom $\delta E_{\mathrm{atom}} = \delta E_{\mathrm{total}}/N$ is required, as it is the case for example in the study of long-ranged correlations in periodic systems. Of course, memory limitations or implementation issues will limit the size of computable systems. Within these limitations, however, the constant scaling behavior is not so surprising, considering that for collecting statistical data, it does not make any difference whether you simulate N weakly interacting systems in parallel or a single system N times as long. Both result in the same statistical error for the same computational cost.

III. POPULATION CORRELATION

For small enough systems, the factor χ_{pop} is close to one, which may be the reason why, to our knowledge, a systematical study has never been attempted before. When scaling up the system size, however, population correlation becomes important and we need a better understanding of its scaling.

The DMC algorithm is based on a drift-diffusion process with branching and killing of configurations due to fluctuations in the local energy. A freshly branched pair of configurations is identical and thereby fully correlated. In the following drift-diffusion process, it takes some time to decorrelate, leading to a fluctuating amount of correlation within the population at any time.

We consider a DMC run over N_{step} time steps i. We will first consider a simplified model with constant population of N_{pop} configurations p, each having a local energy E_p^i. A generalization including population fluctuations will follow in the section below.

An effective population size $N_{\mathrm{pop}}^{\mathrm{eff}}$ can be defined as the number of configurations that would result in the same variance of the average as the correlated population. For a long DMC run, the raw DMC variance can be estimated from the averages over all configurations at all time steps as

$$\sigma_{\mathrm{dmc}}^2 = \langle\langle (E_p^i)^2\rangle_p\rangle_i - \langle\langle E_p^i\rangle_p\rangle_i^2, \tag{6}$$

while the variance of the population average is defined as

$$\sigma_{\mathrm{pop}}^2 = \langle\langle E_p^i\rangle_p^2\rangle_i - \langle\langle E_p^i\rangle_p\rangle_i^2, \tag{7}$$

with the averages abbreviated as $\langle\cdot\rangle_p = \sum_{p=1}^{N_{\mathrm{pop}}}\cdot/N_{\mathrm{pop}}$ and $\langle\cdot\rangle_i = \sum_{i=1}^{N_{\mathrm{step}}}\cdot/N_{\mathrm{step}}$.

In the case of an uncorrelated population, we would find $\sigma_{\mathrm{pop}}^2 = \sigma_{\mathrm{dmc}}^2/N_{\mathrm{pop}}$, so we can measure the amount of correlation by defining an effective population size as the ratio

$$N_{\mathrm{pop}}^{\mathrm{eff}} = \sigma_{\mathrm{dmc}}^2/\sigma_{\mathrm{pop}}^2, \tag{8}$$

where $N_{\mathrm{pop}}^{\mathrm{eff}} \leq N_{\mathrm{pop}}$ with equality only in the case of a completely uncorrelated population. By collecting the necessary data during a DMC run, $N_{\mathrm{pop}}^{\mathrm{eff}}$ can be computed at negligible cost. Figure 1 displays the scaling of the effective population size with increasing system size for a sample system.

Note that the average population N_{pop} is typically slightly lower than the target population $N_{\mathrm{pop}}^{\mathrm{target}}$, because the population control implemented in CASINO uses the linear average $\langle E_p^i\rangle_p$ of the energy instead of the exponential average $\ln\langle\exp(E_p^i)\rangle_p$ which determines the actual growth of the population. Apart from reducing the population size, this has no effect on the statistics or the result.

IV. POPULATION FLUCTUATIONS

Keeping the population size completely fixed as we had assumed in the previous model gives rise to a population

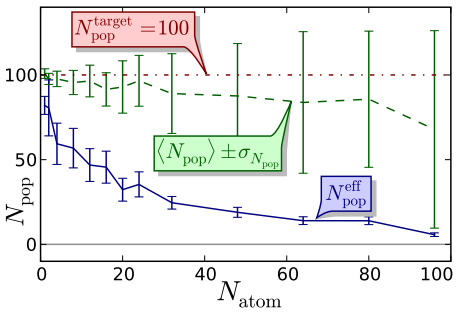

FIG. 1. (Color online) Scaling of the effective population size $N_{\rm pop}^{\rm eff}$ [see Eq. (8)] in a sample system [$\alpha=1.5$, see Eq. (13)] with increasing number of atoms $N_{\rm atom}$. The target population size is fixed, the true population size $N_{\rm pop}$ fluctuates around a slightly lower average (see text). The "error bars" of $N_{\rm pop}$ visualize the increasing population fluctuations $\sigma_{N_{\rm pop}}$. The effective population drops exponentially, due to increasing correlations within the population.

control bias. To reduce this bias, the population size $N_{\rm pop}^i$ is allowed to fluctuate and only weakly controlled. The average over all time steps then needs to be weighted. In the simple population control mechanism considered here,[18] the weights are simply defined by the population size $w_i=N_{\rm pop}^i$ for each time step i. The resulting total energy average of a DMC run is then

$$E_{\rm tot}=\frac{1}{\sum_i w_i}\sum_{i=1}^{N_{\rm step}} w_i\langle E_p^i\rangle_p.$$

In more sophisticated schemes, w_i and $N_{\rm pop}^i$ may be decoupled. To estimate the variance of this weighted average, we can split off the correlation time into a factor and use the estimator of the variance of a weighted average. Viewing the local energies E_p^i as random variables and the weights w_i as constants given by a long DMC run, we can express this as

$$\mathrm{var}[E_{\rm tot}]=\frac{\tau_{\rm corr}}{\tau_{\rm step}}\frac{1}{\left(\sum_i w_i\right)^2}\sum_{i=1}^{N_{\rm step}} w_i^2\,\mathrm{var}[\langle E_p^i\rangle_p]. \tag{9}$$

In the same interpretation of E_p^i as random variables, we can replace the averages over time steps in Eqs. (6) and (7) by statistical expectation values $\langle\cdot\rangle$ and write for each single time step

$$\mathrm{var}[\langle E_p^i\rangle_p]=\sigma_{\rm dmc}^2-\langle\langle (E_p^i)^2\rangle_p-\langle E_p^i\rangle_p^2\rangle.$$

Substituting this into Eq. (9) results in a sum over expectation values, so the $\mathrm{var}[E_{\rm tot}]$ itself can be written as the expectation value of a single expression which can be expressed as a product

$$\mathrm{var}[E_{\rm tot}]=\left\langle \frac{\tau_{\rm corr}}{\tau_{\rm step}}\times\frac{1}{N_{\rm step}^{\rm eff}}\times\frac{1}{N_{\rm pop}^{\rm eff}}\times\sigma_{\rm dmc}^2\right\rangle,$$

with generalized expressions for the effective step number and population size,

$$\frac{1}{N_{\rm step}^{\rm eff}}=\frac{\sum_i (w_i^2)}{\left(\sum_i w_i\right)^2},$$

$$\frac{1}{N_{\rm pop}^{\rm eff}}=1-\frac{1}{\sigma_{\rm dmc}^2}\frac{\sum_i w_i^2(\langle (E_p^i)^2\rangle_{\rm pop}-\langle E_p^i\rangle_{\rm pop}^2)}{\sum_i w_i^2},$$

where the quadratic appearance of the weights makes the effective population size sensitive to population fluctuations as well.

To estimate the statistical efficiency of the DMC algorithm, it is most useful to combine both quantities into the definition

$$\chi_{\rm pop}=\frac{N_{\rm step}\times N_{\rm pop}}{N_{\rm step}^{\rm eff}\times N_{\rm pop}^{\rm eff}}. \tag{10}$$

For a DMC run that is sufficiently long that the set of weights w_i is a good representation of the statistical distribution, the quantity $\chi_{\rm pop}$ is an unbiased estimator. It corresponds exactly to the quantity in Eq. (2) and remains directly proportional to the total CPU cost according to Eq. (5).

V. ASYMPTOTICS OF $\chi_{\rm pop}$

The asymptotic behavior of $\chi_{\rm pop}$ for weak population correlation can be derived by a few simple arguments. Assume, for a moment, the local energy $E_{\rm loc}$ of configurations to be normally distributed as

$$p(E_{\rm loc})=\frac{1}{\sigma_{\rm dmc}\sqrt{2\pi}}\exp\left(-\frac{E_{\rm loc}^2}{2\sigma_{\rm dmc}^2}\right).$$

A single configuration with $E_{\rm loc}<0$ will branch at a rate of $-E_{\rm loc}$. Integrated over the distribution of $E_{\rm loc}$, this leads to a branching rate per configuration of

$$\tau_{\rm branch}^{-1}=\int_{-\infty}^{0} dE_{\rm loc}p(E_{\rm loc})(-E_{\rm loc})=\frac{\sigma_{\rm dmc}}{\sqrt{2\pi}}.$$

It is, of course, known that the true distribution of the local energy is far from normal.[21] However, we can significantly relax the previous assumption, considering that we essentially obtained the ratio between standard deviation and mean absolute deviation which holds approximately for a wide range of distributions.[22]

Starting from an initially uncorrelated population of size $N_{\rm pop}$, the population after branching is $N_{\rm pop}+1$ with two identical configurations. The population mean is equivalent to that over $N_{\rm pop}-1$ correlations of single weight and one of double weight which has the variance

$$(\sigma_{\rm pop}^2)'=\sigma_{\rm dmc}^2\frac{(N_{\rm pop}-1)\times 1^2+1\times 2^2}{[(N_{\rm pop}-1)\times 1+1\times 2]^2}.$$

For $N_{\rm pop}\gg 1$, the effective population size after the branching is therefore $(N_{\rm pop}^{\rm eff})'=N_{\rm pop}-1$.

To keep the population stable, the branching and killing rates have to be equal. For a weakly correlated population, it

is trivial to see that a killing event reduces the (effective) population by 1 as well.

After branching, the two copies evolving independently take an effective time of $\tau_{\mathrm{corr}}/2$ to decorrelate. (Since τ_{corr} measures the amount of statistical data that is lost due to serial correlation, which is exactly the quantity that we want to measure for population correlation as well.)

For $\tau_{\mathrm{branch}} \gg \tau_{\mathrm{corr}}$, the effective population is reduced by 1 at a rate of $2N_{\mathrm{pop}}\tau_{\mathrm{branch}}^{-1}$ and restored to N_{pop} within $\tau_{\mathrm{corr}}/2$. On average this gives

$$N_{\mathrm{pop}}^{\mathrm{eff}} = N_{\mathrm{pop}} - 2N_{\mathrm{pop}}\tau_{\mathrm{branch}}^{-1}\tau_{\mathrm{corr}}/2,$$

or, since population fluctuations can be neglected,

$$(\chi_{\mathrm{pop}})_{\tau_{\mathrm{corr}}\sigma_{\mathrm{dmc}}\to 0} \to 1 + \frac{\tau_{\mathrm{corr}}\sigma_{\mathrm{dmc}}}{\sqrt{2\pi}}. \tag{11}$$

VI. MODEL DMC PROCESS

To study the dependence of χ_{pop} on the system parameters beyond the perturbative regime, we have implemented the full DMC algorithm on top of a minimal model of a correlated diffusion process. Each configuration is reduced to a single random variable with a simple exponential autocorrelation so that σ_{dmc} and τ_{corr} are free parameters. The value of the variable is used directly as the local energy for the branching process.

The population control mechanism described by Umrigar *et al.* introduces an additional parameter τ_{ceref} with the dimension of time (implemented in CASINO as parameter cerefdmc, used in updating the reference energy E_{ref}; $\tau_{\mathrm{ceref}}/\tau_{\mathrm{step}}$ corresponds to g in the original publication Ref. 18). To avoid frequent population instabilities for extreme parameter settings, we have restricted the population size to a window around the target population and decoupled the total weight from the total population size outside of this window, allowing recovery from explosions or starvation without introducing any additional bias. All the results presented below are in the regime where this mechanism has no significant impact on the efficiency.

With each run, one obtains the factor χ_{pop} as a function of the parameters σ_{dmc}, τ_{corr}, τ_{ceref}, and τ_{step}. The result must be dimensionless, reducing the number of relevant parameters by one. Furthermore, one is interested in the limit $\tau_{\mathrm{step}} \to 0$. We find that for $\tau_{\mathrm{step}} \lesssim 0.1 \times \min(\tau_{\mathrm{corr}}, \tau_{\mathrm{ceref}}, 1/\sigma_{\mathrm{dmc}})$, the inefficiency factor χ_{pop} becomes practically independent of τ_{step} in all cases. We can combine the remaining free parameters into $\sigma_{\mathrm{dmc}}\tau_{\mathrm{corr}}$ and $\tau_{\mathrm{ceref}}/\tau_{\mathrm{corr}}$, leading to the results displayed in Fig. 2.

Most significantly, we find that Eq. (11) is not only confirmed in the perturbative limit, but its exponential continuation gives a strict lower limit for the inefficiency factor

$$\chi_{\mathrm{pop}} \geq \exp(\sigma_{\mathrm{dmc}}\tau_{\mathrm{corr}}/\sqrt{2\pi}) \tag{12}$$

where the deviation from this exponential depends on the ratio $\tau_{\mathrm{ceref}}/\tau_{\mathrm{corr}}$.

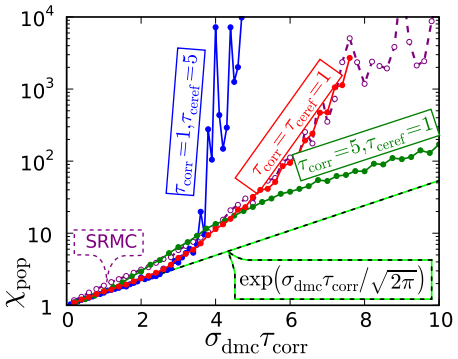

FIG. 2. (Color online) Inefficiency factor χ_{pop} computed for a model DMC process (see text) in dependence of the two relevant parameters $\sigma_{\mathrm{dmc}}\tau_{\mathrm{corr}}$ and $\tau_{\mathrm{ceref}}/\tau_{\mathrm{corr}}$. The exponential law extrapolated from the perturbative limit is found to give a strict lower limit for χ_{pop}. Solid circles refer to standard DMC with branching. Hollow circles refer to minimal stochastic reconfiguration MC (SRMC, Refs. 23 and 24).

VII. HYDROGEN SAMPLE SYSTEM

To demonstrate our result in a real calculation, we have performed various DMC runs using the CASINO program.[25] We chose a system of N hydrogen atoms placed several thousand atomic units apart to make them effectively independent. As a trial wave function, we used the exact ground state with a detuning parameter α and an additional term to satisfy the Kato cusp condition,[26] centered on each hydrogen atom

$$\Psi_\alpha(r) = \alpha e^{-\alpha r} + (1-\alpha)e^{-(\alpha+1)r}. \tag{13}$$

We performed a large variety of runs on this model system with system sizes $N \in \{1,2,4,8,\ldots,64\}$, and detuning parameters $\alpha \in \{1.1,\ldots,3.0\}$ and target population $N_{\mathrm{pop}}=200$. The DMC time step was set to $\tau_{\mathrm{step}}=0.02$ in all cases.

The variance σ_{dmc}^2 was found to be equal to σ_{vmc}^2 within the statistical error in all cases. In each case, the population correlation factor χ_{pop} was determined from the variances using Eq. (10).

Obtaining a precise value for the correlation time τ_{corr} takes an extremely large amount of data in either of the two methods described above. For a reasonably precise value, we performed a very long DMC run ($N_{\mathrm{step}} > 10^7$) on a single atom for each value of α. Several tests on larger systems confirmed that the same value holds for larger numbers of atoms N. Since τ_{corr} is independent of τ_{step} only if $\tau_{\mathrm{corr}} \gg \tau_{\mathrm{step}}$, we performed these runs for the same step size as the main calculations.

As in Fig. 2, we plot the population correlation factor χ_{pop} from all our calculations as a function of the product $\sigma_{\mathrm{dmc}}\tau_{\mathrm{corr}}$ and again find the exponential lower bound described by Eq. (12), as displayed in Fig. 3.

VIII. ANALYSIS OF VARIOUS SAMPLE SYSTEMS

The exponential law in Eq. (12) has severe implications for the scaling of the DMC method. Following Eqs. (3) and (5), the total CPU cost becomes

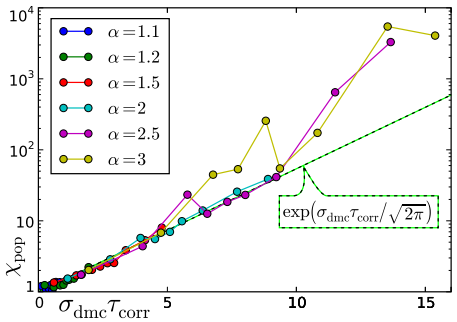

FIG. 3. (Color online) Data from many different calculations on a model system of N independent hydrogen atoms with a detunable trial wave function [Eq. (13)]. $\sigma_{\rm dmc}$ and $\chi_{\rm pop}$ were directly obtained from each run. $\tau_{\rm corr}$ was determined from a single, very long run for each type of atomic wave function. Population control is kept at the default $t_{\rm ceref}=1$.

$$t_{\rm total} \propto \frac{\exp(X\sqrt{N})}{\delta E_{\rm atom}^2},$$

or worse. So, even if population correlation may not be an issue yet for most applications, it will eventually lead to an exponential scaling of the cost. The factor X can be reduced by optimizing the wave function, but the gain that is possible with reasonable effort is very limited.

Table I lists a selection of sample systems showing the size at which the exponential scaling becomes observable. The integrated correlation time $\tau_{\rm corr}$ [via Eq. (4)] and the raw variance $\sigma_{\rm dmc}^2$ were computed for very small systems and Eq. (12) was then used to estimate the size at which a comparable system would show significant population correlation. All values should be understood as rough estimates. The trial wave functions were either taken from a library of examples or optimized with moderate effort. Further optimizations could certainly reduce $\sigma_{\rm dmc}^2$ and thereby shift the onset of significant population correlation. Typically, however, significant effort is necessary even for minor improvements using optimizations beyond the standard Jastrow terms.

TABLE I. Estimated values for various sample systems. The last column gives the system size based on Eq. (12) at which the population correlation becomes significant with $\chi_{\rm pop}=2$. Beyond this size, the DMC method must be expected to become exponentially inefficient. The first three categories are based on either all-electron (ae) or pseudopotential (pp) wave functions with optimized Jastrow factors. All numbers should be understood as rough estimates based on moderately optimized trial wave functions. Reducing $\sigma_{\rm dmc}^2$ by further optimization will shift the onset of the inefficiency by the same factor.

Atoms (ae)	$\tau_{\rm corr}$	$\sigma_{\rm dmc}^2$/atom	$\chi_{\rm pop}=2$
He	0.5	0.0044	2700 atoms
C	0.4	0.16	140 atoms
Ar	0.04	8.0	250 atoms
Molecules (ae)	$\tau_{\rm corr}$	$\sigma_{\rm dmc}^2$/molec	$\chi_{\rm pop}=2$
H_2O	0.1	0.58	550 molecules
CH_4	0.3	0.24	120 molecules
C_2H_4	0.4	0.51	38 molecules
SO_2	0.06	7.5	105 molecules
Crystals	$\tau_{\rm corr}$	$\sigma_{\rm dmc}^2$/atom	$\chi_{\rm pop}=2$
Diamond (pp)	0.15	0.23	630 atoms
Diamond (ae)	0.1	2.3	133 atoms
Graphite (pp)	0.3	0.20	135 atoms
Silicon (pp)	0.4	0.052	328 atoms
Electron gas	$\tau_{\rm corr}$	$\sigma_{\rm dmc}^2$/elec.	$\chi_{\rm pop}=2$
$3d$ crystal ($r_s=1$)	0.2	0.26	193 electrons
$3d$ fluid ($r_s=5$)	5	4.2×10^{-4}	330 electrons
$3d$ fluid ($r_s=10$)	16	5.1×10^{-5}	242 electrons
$2d$ crystal ($r_s=1$)	0.4	0.038	570 electrons
$2d$ fluid ($r_s=1$)	0.3	0.033	1154 electrons

IX. ALTERNATIVE VARIANTS OF QMC

To this point the discussion was centered on the conventional DMC algorithm including drift and branching. In the following, we will briefly discuss a number of alternative QMC algorithms in view of the population correlation scaling.

First, it is clear that population correlation can only be caused by some form or branching. The variational MC (VMC) algorithm, which samples an explicitly known wave function, clearly does not have this feature. A variant of DMC with branching switched off (sometimes referred to as "pure" DMC) also features a completely uncorrelated population. If all configurations are fixed to the same statistical weight, this process produces the same distribution of configurations and thereby the same total energy as VMC and can therefore be seen as a variant of the former.

If, on the other hand, each configuration in a pure DMC run carries a statistical weight evolving with the fluctuations in the local energy, the effective population size $N_{\rm pop}^{\rm eff}$ is reduced in the same way as it would be when branching were allowed, with the only difference that decorrelation does not happen and the method becomes exponentially unstable with simulation time.[24]

A number of variants of the DMC algorithms keep the population size fixed and include branching in form of stochastic reconfiguration, duplicating some configurations and deleting others.[23,24,27] While the population is fixed, the total weight is allowed to fluctuate independently and the population control is replaced by a weight control mechanism.

Our definition of the population correlation factor $\chi_{\rm pop}$ in Eq. (10) is already kept general enough to capture the effects of weight fluctuations within the population along the correlations within the population and to capture the fluctuations of the total weight along with the population fluctuations. Tests on several variants of the branching strategy confirmed that these have no influence on the exponential lower bound of $\chi_{\rm pop}$ but only affect how far the actually measured $\chi_{\rm pop}$

exceeds the predicted exponential scaling (see Fig. 2).

One remaining option to limit χ_{pop} is the use of strong population control on a small population, accepting a significant population bias. The extreme case of this strategy would be a single walker with the weight renormalized after every step, leading exactly to the VMC distribution. We can, therefore, tune between exponentially scaling statistical inefficiency and a population bias that ultimately leads to recovering the VMC algorithm, which—as we know—does not suffer from exponential scaling.

Variants of QMC such as path integral MC (PIMC) (Ref. 28) or reptation Monte Carlo (RMC) (Ref. 29) are somewhat related to DMC in the sense that they are based on a drift-diffusion process in imaginary time. Unlike DMC, however, these methods are based on a true Metropolis algorithm without the need for branching. Population correlation does not occur and the statistical weight fluctuations are not a problem. Instead, an analysis of the statistical efficiency of these methods would need to take into account the correlation time and its scaling with system size.

X. CONCLUSIONS

To conclude, we have derived an expression for the scaling behavior of DMC calculations when aiming at a fixed statistical precision per particle. Using a linear scaling algorithm for an individual time step, constant scaling of the total computational cost for the energy per particle is possible in principle, except for a factor χ_{pop}, which quantifies the correlation within the population of walkers. The exact value of χ_{pop} was derived in the perturbative limit, depending only on the correlation time and the raw variance of the DMC process. Based on numerical evidence, we demonstrated that an exponential extrapolation of the perturbative law gives a strict lower bound to the inefficiency factor χ_{pop}. From this, it follows that the DMC algorithm generally scales at least exponentially in the square root of the system size. The numbers for actual sample systems indicate that this exponential scaling should not even be observable in most DMC based studies done so far, leaving plenty of room to do interesting research with the DMC method.

Alternative schemes for branching and population control that have been suggested[23,24,27] may certainly influence the efficiency of the algorithm. The exponential lower bound of the statistical inefficiency, however, may at best be shifted over toward an exponentially scaling population control bias.[30]

It must be stressed that this exponential scaling factor is specific to the DMC method and does not occur in other methods like VMC. It is not linked to the more fundamental fermion sign problem[31] and it is not limited to certain observables.[32]

In fact, the exponential scaling may be a symptom of the very nature of the DMC process. In general, Markov-chain MC methods such as VMC exhibit excellent scaling behavior. DMC however, is not based on a Markov process but rather on the simulation of a time-dependent stochastic diffusion process. As such, it must be expected to suffer from the exponential accumulation of errors inherent in the simulation of time evolution in nonintegrable systems. The population control that is necessary to stabilize the process might then necessarily lead to exponential scaling either in the bias or the efficiency of the process. The only way to overcome this problem might then be to resort to alternative QMC methods like VMC, PIMC, or RMC that are based the stochastic computation of a multidimensional integral in the original spirit of Markov-chain MC methods.

ACKNOWLEDGMENTS

We acknowledge fruitful discussions with Richard Needs, Neil Drummond, and Matthew Foulkes. This work was funded by the DAAD and the EPSRC. The computations were performed using the facilities of the University of Cambridge High Performance Computing Service.

[1] A. J. Williamson, R. Q. Hood, and J. C. Grossman, Phys. Rev. Lett. **87**, 246406 (2001).
[2] S. Manten and A. Lüchow, *Recent Advances in Quantum Monte Carlo Methods*, edited by J. William, A. Lester, S. M. Rothstein, and S. Tanaka (World Publishing, Singapore, 2002), Vol. II.
[3] S. Manten and A. Lüchow, J. Chem. Phys. **119**, 1307 (2003).
[4] D. Alfè and M. J. Gillan, J. Phys.: Condens. Matter **16**, L305 (2004).
[5] F. A. Reboredo and A. J. Williamson, Phys. Rev. B **71**, 121105(R) (2005).
[6] A. Aspuru-Guzik, R. Salomón-Ferrer, B. Austin, J. William, and A. Lester, J. Comput. Chem. **26**, 708 (2005).
[7] J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **128**, 134104 (2008).
[8] P. K. V. V. Nukala and P. R. C. Kent, J. Chem. Phys. **130**, 204105 (2009).
[9] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).
[10] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[11] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[12] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., J. Chem. Phys. **77**, 5593 (1982).
[13] J. C. Slater, Phys. Rev. **34**, 1293 (1929).
[14] R. Jastrow, Phys. Rev. **98**, 1479 (1955).
[15] G. H. Wannier, Phys. Rev. **52**, 191 (1937).
[16] D. Alfè and M. J. Gillan, Phys. Rev. B **70**, 161101(R) (2004).
[17] S. Moroni (unpublished, lecture at QMC School, ICTP Trieste, Feb. 2008).
[18] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).
[19] U. Wolff, Comput. Phys. Commun. **156**, 143 (2004).
[20] H. Flyvbjerg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).
[21] J. R. Trail, Phys. Rev. E **77**, 016703 (2008).
[22] S. Gorard, Br. J. Educ. Stud. **53**, 417 (2005).

[23] M. Calandra Buonaura and S. Sorella, Phys. Rev. B **57**, 11446 (1998).
[24] R. Assaraf, M. Caffarel, and A. Khelif, Phys. Rev. E **61**, 4566 (2000).
[25] R. J. Needs, M. D. Towler, N. D. Drummond, and P. López Ríos, J. Phys.: Condens. Matter **22**, 023201 (2009).
[26] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).
[27] A. Jones, A. Thompson, J. Crain, M. H. Müser, and G. J. Martyna, Phys. Rev. B **79**, 144119 (2009).
[28] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
[29] S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4745 (1999).
[30] N. D. Drummond, R. J. Needs, A. Sorouri, and W. M. C. Foulkes, Phys. Rev. B **78**, 125106 (2008).
[31] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).
[32] G. L. Warren and R. J. Hinde, Phys. Rev. E **73**, 056706 (2006).