Decorrelation of samples in quantum Monte Carlo calculations and scaling of autocorrelation time in Li and H$_2$O clusters

D. Nissenbaum, B. Barbiellini, and A. Bansil
Physics Department, Northeastern University, Boston, Massachusetts 02115, USA
(Received 24 April 2007; published 23 July 2007)

We have investigated decorrelation of samples in quantum Monte Carlo ground-state energy calculations for large Li and H$_2$O nanoclusters. Binning data as a way of eliminating statistical correlations, as is the common practice, is found to become increasingly impractical as the system size grows. We demonstrate nevertheless that it is possible to perform accurate energy calculations—without decorrelating samples—by exploiting the scaling of the integrated autocorrelation time $\tau$ as a function of the number of electrons in the system.

DOI: 10.1103/PhysRevB.76.033412
PACS number(s): 73.22.--f, 02.70.Ss, 71.10.--w, 71.15.Nc

Quantum Monte Carlo (QMC) methods are becoming increasingly important in electronic structure calculations for a variety of systems of current interest. Although effective QMC methods have been developed in recent years to optimize the parameters of many-body wave functions, QMC calculations are fundamentally limited by the statistical accuracy that can be obtained in a reasonable amount of computation time. Despite the growing maturity of core QMC computations of large clusters become feasible, even in clusters are hampered by increasing correlations in samples due to the Markovian character of the QMC sampling algorithm. We address this problem through an analysis of the integrated autocorrelation time

$$\tau (\epsilon) = \frac{\sigma_0^2}{\sigma_n^2}$$

where $\sigma_0^2$ is the estimate of the variance of $E$ calculated as though the samples were uncorrelated, i.e., $\sigma_0^2 = \sigma_{\text{raw}}^2 / N$, where $\sigma_{\text{raw}}^2$ is the variance of $h(x)$ over $N$ steps. $C_f(i)$, the autocorrelation of $h$ at $i$ steps, is given by $C_f(i) = \langle h(x)h(x+i) \rangle_{\text{eq}} - \langle h(x) \rangle_{\text{eq}}^2$ (normalized to the point $i = 0$), where $j$ is an arbitrary step number in the postequilibration phase and the brackets $\langle \cdot \rangle_{\text{eq}}$ denote the expectation value over an infinite ensemble. The “integrated autocorrelation time” $\tau$ is then defined such that the variance

$$\tau = \frac{\sigma_{\text{raw}}^2}{\sigma_n^2} / (N/\tau),$$

so that the “effective” number of steps taken by the walker is given not by $N$ but by $N/\tau$. Equation (2) generalizes the statistical error whose scaling with system size has been discussed by Williamson et al. by the addition of the quantity $\tau$, the factor which describes the effect of statistical correlations between QMC moves. In the discrete case at hand, we have

$$\tau = \left[ 1 + 2 \sum_{i=1}^{\infty} C_f(i) \right].$$

By comparing Eqs. (1) and (3), we see that

$$\tau = (\sigma / \sigma_0)^2.$$

The ratio in Eq. (4) is sometimes also called the “statistical inefficiency,” see Refs. 21–23. In order to determine the ratio $(\sigma / \sigma_0)^2$, and thus obtain $\tau$, we employed the recently developed dynamic distributable decorrelation algorithm (DDDA) to block data efficiently. Like other data blocking algorithms, the DDDA is a renormalization group method that relies on the appearance of a plateau to determine the
dratically with system size. We emphasize that the total number of points $n_p$ needed to reach the plateau for a given cluster size is much larger than the corresponding value of $\tau$. For example, for the 64 atom cluster the plateau appears in Fig. 1 at the value of 14 or equivalently for $n_p=2^{14}$, while $\tau=(4.3)^2=18.5$. In other words, one would need to skip $2^{14}$ steps in this case if one wanted to decorrelate samples. Note that $\tau$ is an integrated correlation time; therefore, skipping $\tau$ steps will not produce decorrelated samples. Clearly, because $n_p$ grows so large, calculation of the error using reblocking rapidly becomes intractable for large clusters.

The total time of a computation can be written as $T=t_s N$, where $t_s$ is the time per step and $N$ is the total number of steps. From Eq. (2), we see that in order to maintain a given accuracy $\sigma$ for QMC calculations of different clusters, the number of steps $N$ must change to balance changes in $\sigma_{\{raw}{raw}}^2$ and $\tau$. For quadratically scaling $\tau$, the number of steps $N$, and therefore the total computation time, will have an additional scaling factor of up to $M^2$, where $M$ denotes the number of atoms in the cluster, in order to compensate for the scaling of $\tau$. This additional scaling factor for $\tau$ is present whether the quantity being calculated is the energy for the entire system or the energy per atom. The scaling factors due to $\sigma_{\{raw}{raw}}^2$ and $t_s$ have been considered in previous work\textsuperscript{14} and are not considered in this study.

A QMC calculation is complete only when accompanied by a measure $\sigma$ of its statistical uncertainty, which is given by the onset of the decorrelation plateau in a plot such as that of Fig. 1. The failure to obtain this onset indicates that it is impossible to decorrelate samples during the run. This “stickiness” of correlated samples in QMC runs impedes our ability to determine the statistical error $\sigma$ of the computation. Often, in order to decorrelate samples, the problem of stickiness is overcome by obtaining a large number of sample points and discarding intermediate correlated points. However, as we have already pointed out above, it becomes increasingly difficult in large clusters to obtain the plateau and decorrelate samples. It may in fact be impossible to decorrelate even a single sample during the run. In sharp contrast, our procedure requires that all steps are sampled, with no steps discarded, in order to first calculate $\sigma_0$ and then obtain $\sigma^2$ by rescaling with an extrapolated value of $\tau$. In this way, our scheme provides a route for circumventing the problem of decorrelating samples in large clusters.

As an example of an energy calculation obtained without observing the onset of a decorrelation plateau, we were able to obtain the energy (per atom) for the Li\textsubscript{64} cluster to within $\sigma=0.1$ eV ($\pm 10\%$) using only $10^5$ QMC steps.\textsuperscript{31} We have verified our estimate of the accuracy of $\sigma$ for this energy calculation by performing multiple, independently equilibrated runs and observing the spread of the energy results.

We turn now to consider H\textsubscript{2}O clusters, where we have studied clusters of 2, 4, 8, 9, 14, and 20 molecules.\textsuperscript{32} These clusters have similar electron counts to those of the Li clusters discussed above. Figure 2 shows decorrelation plots as a function of block size. All clusters display the onset of the decorrelation plateau. A comparison of the insets in Figs. 1 and 2 shows that the value of $\sqrt{\tau}=\sigma/\sigma_0$ varies significantly less rapidly with system size in H\textsubscript{2}O clusters. It is striking that all curves in Fig. 2 collapse onto an essentially universal curve up to five block transformations. Beyond this block...
size, however, it appears that other secondary longer-distance correlations play a noticeable role so that the curves begin to rise again and settle into new, weakly size-dependent plateaux. If it were not for this secondary feature, all H$_2$O clusters would achieve their plateaux in the same small number of blocking steps, and the stickiness induced by correlated sampling would not be important. Because of these secondary features, H$_2$O clusters exhibit quadratic scaling of $\tau$ with $M$, but the influence is significantly less severe than for Li. In fact, if $a+bN_{el}$ describes a linear fit for the data in the inset for H$_2$O, we see that for the cluster sizes we studied, the constant term in the quadratic expression $\tau=(\sqrt{\tau})^2=(a+bN_{el})^2$ dominates. This may explain why the phenomenon was not observed in Ref. 14. For Li, the linear term dominates and adds to the scaling of the total computation time for the cluster sizes we studied.

We now provide a possible explanation for the divergence of the autocorrelation time $\tau$ in ground-state computations of Li clusters. We recall that bulk Li metal contains discontinuities in the momentum density of the electron gas at the Fermi momentum. Such a sharp cutoff in the momentum distribution can be expected to generally cause oscillations of electron configurations involved in the QMC algorithm and thereby slow down the statistical convergence in clusters like those of Li atoms, which must eventually approach the bulk metallic state with increasing size. These considerations suggest that smearing or broadening the underlying “Fermi surface” (formally, of course, there is no Fermi surface in a finite system) in metallic nanoclusters using a trial wave function such as an AGP (Ref. 33) or a Pfaffian (Ref. 34) might result in shorter autocorrelation times. Notably, Baroni and Moroni$^{35}$ have discussed a theoretical connection between the autocorrelation time and the spectrum of an auxiliary quantum Hamiltonian, which might help explain the divergence of the autocorrelation time in metallic systems.

The integrated autocorrelation time $\tau$ is related to the second largest eigenvalue $\lambda$ of the transition matrix by $\tau=1+2\lambda/(1-\lambda)$. Very recent work regarding the scaling of $\lambda$ in the Metropolis algorithm demonstrates that for logarithm-concave distributions $|\Phi_0(x)|^2$, the scaling of $\tau$ with $M$ is at most $M^2$. For Langevin dynamics, the divergence of $\tau$ is usually less severe than in the Metropolis algorithm. In principle, the appearance of nodes in the wave function can violate the logarithm concavity of $|\Phi_0(x)|^2$ and might modify the scaling properties of $\tau$. Although our wave functions contain nodes, the fact that we obtain a clear $M^2$ scaling of $\tau$ suggests that the $M^2$ scaling remains maximal for typical QMC distributions $|\Phi_0(x)|^2$ more generally.

In conclusion, we have investigated the scaling of the integrated autocorrelation time $\tau$ in QMC computations of Li and H$_2$O clusters. The quantity $\tau$, which directly yields a measure of sampling correlations in the calculation, is found to diverge quadratically with system size $M$ in both cases, more severely for Li than for H$_2$O, and becomes increasingly difficult to calculate for large clusters. We have shown that brute force decorrelation of samples is simply not feasible for these large clusters. Our study highlights the importance of correlated sampling in QMC computations of large clusters and provides a route for obtaining accurate error estimates by exploiting the scaling properties of $\tau$ with system size, without requiring explicit decorrelation of samples. The present results should be relevant for the growing scientific community interested in the convergence properties of Markov chains and QMC calculations of properties of nanoclusters. Further work towards understanding the effects of excitation gaps on the convergence of the QMC algorithm might be valuable.

We acknowledge useful discussions with Lubos Mitas, Chip Kent, Dario Bressanini, and Erich Novak, and we thank R. Rousseau and D. Marx for sending us atomic coordinates for Li clusters. This work was made possible by the support of the U.S. DOE Contracts No. DE-FG02-07ER46352 and No. DE-AC03-76SF00098. We benefited from the allocation of supercomputer time at NERSC and Northeastern University’s Advanced Scientific Computation Center (ASCC).

---

We ran preliminary calculations of small Li clusters using single-electron moves. These calculations revealed similar modified autocorrelation times (Ref. 9) as global moves. Also, the QMC step size scales linearly with the inverse of the number of electrons for the Li clusters we studied, so that the total distance the electrons move remains the same, whether global or single-electron moves are used.

We performed a detailed study of Li$_{64}$ and (H$_2$O)$_{20}$ equilibration by running a massively parallel calculation (1024 processors), assuming the points were equilibrated (even when they were not), and noticing that the variational theorem demands that the energy be higher for unequilibrated systems. Therefore, only when the error bars for our massively parallel run overlapped the (known) HF energy did we consider our samples to be equilibrated.

The geometry of H$_2$O clusters can be found at http://www-wales.ch.cam.ac.uk/~wales/CCD/TIP4P-water.html. The use of the HF wave function is adequate for our study, aimed at understanding convergence properties of QMC calculations rather than obtaining an accurate value of the correlation energy. Also, by duplicating a known HF result, we can be certain that the system is ergodic.