Stochastic Effective Core Potentials, toward Efficient Quantum Monte Carlo Simulations of Molecules with Large Atomic Numbers

Jonas Feldt, Roland Assaraf

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

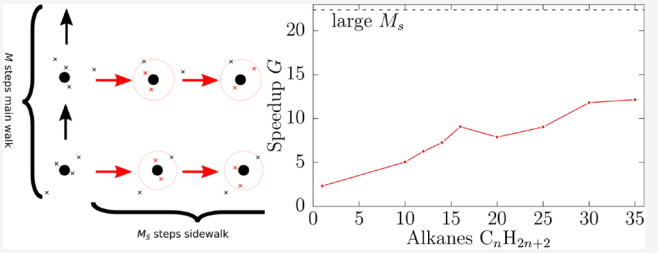
Stochastic Effective Core Potentials, toward Efficient Quantum Monte Carlo Simulations of Molecules with Large Atomic Numbers

Jonas Feldt* and Roland Assaraf*

Cite This: https://dx.doi.org/10.1021/acs.jctc.0c01069 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: We propose a Monte Carlo method which exploits that core regions are physically independent in a molecule to almost remove their contribution to the numerical cost. The method is tantamount to computing an effective core potential on the fly, by efficiently subsampling the core regions with independent sidewalks. The removal of fluctuations in the core region enables also the dynamic in the valence region to be accelerated using a process with two time steps. As a function of the total number of electrons N the numerical overhead $O(N)$ is negligible in comparison to the overall scaling $O(N^3)$ (due to the evaluation of determinants). Tests are presented on atoms, alkane chains, and clusters of silicons. We report a transferability of the parameters of the method from atoms to molecules, enabling a calibration using only single atoms. These tests display a gain in numerical efficiency between one and two orders of magnitude for large N.



1. INTRODUCTION

Quantum Monte Carlo (QMC) methods use a stochastic approach to solve the Schrödinger equation. A great flexibility in the choice of the wave function allows both dynamical as well as static correlation to be efficiently treated. Their scaling with the system size considering the computational costs is very much favorable compared to deterministic quantum chemistry methods. Recently, it has been demonstrated that the scaling of computing a multideterminant expansion and optimizing can also be strongly reduced.[1−3] This is done using the determinant Lemma which allows a Slater determinant to be updated efficiently when only a few columns are modified. QMC has been used extensively for the description of materials including excited states.[4]

A great challenge for QMC methods is the unfavorable scaling with the atomic number Z. Core electrons which are at a distance $O(1/Z)$ from the nucleus have two undesirable consequences. First, they are greatly slowing down the dynamics of the valence degrees of freedom, the ones mostly relevant to chemistry. Second, the core electrons contribute to a large degree to the energy ($O(Z^2)$ for a hydrogenic atom) and, as we will emphasize later, also to the statistical fluctuations. A common way to circumvent these problems is employing effective core potentials (ECP) which remove the core electrons and consequently allow for efficient sampling of the remaining valence electrons. While this is a practical approach it is as well an approximate one which spoils the high accuracy which is expected from QMC. For instance, the widely used Burkatzki−Filippi−Dolg pseudopotentials[5] have been parametrized for Hartree−Fock, completely disregarding the correlation energy, and the error introduced by such empirical ECPs cannot be directly judged. Comparing all-electron and valence-only calculations, it has been shown that the effect is even larger for properties of excited states than for the ground state.[6] For large nuclei (e.g., the third row of the periodic table) pseudopotentials are a practical approach to take into account the important relativistic effect of inner core electrons, but removing more core electrons from the computational cost is still desirable.

So far, most methodological progress regarding the core electrons was focused on improving the sampling to decrease the correlation time or improve the ergodicity. In the variational Monte Carlo method (VMC) a dynamic using spherical coordinates[7] can strongly reduce the slowing down of the dynamics due to the core region. However, this method is not transferable to the diffusion Monte Carlo algorithm for which the dynamics is imposed by the Hamiltonian. For the more accurate fixed-node diffusion Monte Carlo method a spatial discretization using a double grid allows adapting the moves close to nuclei,[8] leading to an acceleration for the valence electrons, with a gain in correlation time up to a factor 10 (for Z = 118).

Received: October 11, 2020

In this paper, the target is to almost remove the contribution of core electrons to the computational cost, both the slowing down in the valence region and the variance. We will show that these two effects are dependent as removing the variance coming from the core region enables the dynamics in the valence region to accelerate drastically. This will be achieved efficiently by exploiting the fact that different core regions are physically separated. First, we build an improved estimator removing the variance coming from the core region which is obtained by subsampling the core region with sidewalks. These sidewalks displace independently a few electrons at a time for a total numerical cost $O(N)$ with a new technique described in Appendix A. This cost is negligible compared to the usual $O(N^3)$ scaling (coming from the evaluation of Slater determinants). The resulting improved estimator (here of the energy) depends only on the valence electron coordinates; it is an effective local energy which includes an effective core potential (ECP).

Because most of the information on the core region is included in the improved estimator, the main walk can be improved using two time steps: a small one for the core electrons and a large one for the valence electrons for which the moves are accelerated (and the correlation factor is lowered). The stochastic process is here the usual drift and diffusion process[9] which has the advantage to be used also in diffusion Monte Carlo calculations.

The combination of these two intertwined gains, in variance and in efficiency of the main walk, removes extensively the effect of the core electrons on the numerical cost for large systems. The gain grows with an increasing number of core electrons, because more information is recovered, with a modest $O(N)$ scaling, and much larger moves can be made in the valence region. We also assess the effect of a minimal Jastrow factor which ensures the electron−electron cusp Kato's conditions. The main result is that the gain in variance is lowered but is compensated for by a speedup of the dynamic in the valence region. One to two orders of magnitude in numerical efficiency can be obtained for alkanes and silicon clusters with or without this Jastrow factor.

2. THEORY: SUBSAMPLING

We want to compute the expectation value of a random variable O on a density π, $\mathbb{E}(O)$. In the variational Monte Carlo framework $\pi = \Psi^2$ is the square of the trial or variational wave function and O can be the local energy for the Schrödinger Hamiltonian H,

$$O = E_L = \frac{H\Psi}{\Psi} \tag{1}$$

We suppose it exists a small region Ω of the probability space which is responsible for a large amount of the variance. This region will correspond to free moves in the core region of an atom with frozen valence electrons. We define in this work a core region as the largest sphere centered on a nucleus which contains n_{core} electrons of a given configuration. The radius of this sphere is the distance of the first valence electron to the nucleus. Note that Ω is a random subspace as it depends on the coordinates v of the valence electrons. We will first consider the conditional expectation value $\mathbb{E}(O|\Omega)$ as an estimator. We recall the meaning of this standard notation in probability theory using the language of integrals. For a given set of valence positions v, the coordinates c of the core electrons are subject to a constraint $c \in C(v)$. This constraint is to be closer to the nucleus than the valence electrons. The conditional expectation value is a number

$$\mathbb{E}(O|\Omega(v)) = \frac{\int_{c\in C(v)} O(c, v)\pi(c, v)\, \mathrm{d}c}{\int_{c\in C(v)} \pi(c, v)\, \mathrm{d}c} \tag{2}$$

which can be interpreted numerically as a partial average of O on a subset of walkers sampling $\pi = \Psi^2$ sharing the same valence configuration v. According to the law of total expectation, the random variable $\mathbb{E}(O|\Omega)$, which depends on v, is an unbiased estimator of $\mathbb{E}(O)$, i.e., $\mathbb{E}(\mathbb{E}(O|\Omega)) = \mathbb{E}(O)$. This standard property in probability theory can be understood also in an integral calculus formulation

$$\begin{aligned}\mathbb{E}(O) &= \int \mathrm{d}v \int_{c\in C(v)} \pi(c, v)O(c, v)\, \mathrm{d}c \\ &= \int \mathrm{d}v \int_{c\in C(v)} \pi(c, v)\mathbb{E}(O|\Omega(v))\mathrm{d}c\end{aligned} \tag{3}$$

The first line is a definition of the expectation value, and the second line which corresponds to the definition of $\mathbb{E}(\mathbb{E}(O|\Omega))$ is easy to check by replacing $\mathbb{E}(O|\Omega(v))$ by eq 2. Since the estimator $\mathbb{E}(O|\Omega)$ depends only on the positions of the valence electrons, this random variable can be seen as the effective valence property which includes an (exact) ECP contribution.

This estimator fluctuates less than O because of the variance decomposition theorem (see Appendix B)

$$V(O) = \mathbb{E}(V(O|\Omega)) + V(\mathbb{E}(O|\Omega)) \tag{4}$$

where $V(O|\Omega)$ is the conditional variance on Ω. It is the variance obtained when the valence configuration v is frozen. The expectation value $\mathbb{E}(V(O|\Omega))$ can be interpreted as the contribution of the core electrons to the total variance $V(O)$. Computing $\mathbb{E}(O|\Omega)$ is also equivalent to adding the covariate $\mathbb{E}(O|\Omega) - O$ to O in order to cancel the effect of the core electrons on the statistical fluctuations. For a molecule (i.e., many atoms) the estimator $\mathbb{E}(O|\Omega)$ could be applied with Ω being defined as the union of all the core regions. We prefer instead this estimator

$$\tilde{O} = O + \sum_i (\mathbb{E}(O|\Omega_i) - O) \tag{5}$$

where Ω_i is a set which corresponds to moving only the core electrons of the ith atom and freezing all the other electrons. The motivation is that moving a few electrons at a time will be numerically much cheaper than moving all the electrons of the core regions. However, we expect almost the same variance reduction due to the following physical consideration. Given a valence configuration v, two distant core regions c_i and c_j should be close to being separable for a physical random variable O. Mathematically, denoting by c_i the electronic configuration of the core region of the ith atom, the cores are separable if the core coordinates are independent for a given valence configuration v and we can write $O = \alpha(v) + \sum_i \beta_i(v, c_i)$. This property implies that $\mathbb{E}(O|\Omega_i) - O = \mathbb{E}(\beta_i|\Omega_i) - \beta_i(v, c_i)$. It follows that $\tilde{O} = \alpha(v)$, and of course

$$\tilde{O} = \mathbb{E}(O|\Omega)$$

The conditional expectation values in eq 5 are not known and have to be sampled. In practice, a covariate can be constructed by carrying out M_s additional steps for each of the core regions, the so-called sidewalks. The main walk is carried out in the usual manner. After each sweep of single-electron moves in the main

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

walk, the sidewalks for the cores are started from the current configuration. After completion of the sidewalks, the main walk continues from the original configuration before the sidewalks started (Figure 1). The improved estimator of O is then

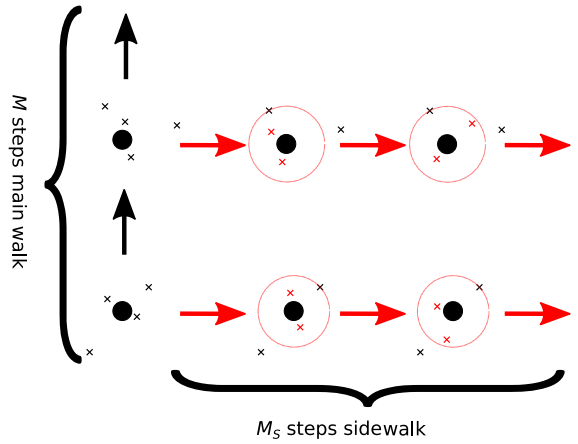


Figure 1. Schematic representation of the core subsampling for an atom ($N = 4$ and $n_{core} = 2$). The core region and the core electrons within the sidewalk are displayed in red.

computed by the subsampling process, i.e., using the following estimator

$$\tilde{O}(\omega, M_s) = O(\omega) + \lambda \sum_i \sum_{k=1}^{M_s} \frac{O(\omega_i^k) - O(\omega)}{M_s} \tag{6}$$

where ω represents a particular configuration in the main walk and ω_i^k is the configuration ω modified by k steps of the ith sidewalk (only the electrons in the ith core region differ between ω_i^k and ω). For the local energy, the control variate $E_L(\omega_i^k) - E_L(\omega)$ can be computed with a cost $O(N)$ for a Jastrow−Slater function (see eq 23 in Appendix A). This cost can be reduced to $O(1)$ by removing terms in the expression of the control variate $E_L(\omega_i^k) - E_L(\omega)$ involving the far environment of the core. Those are zero in the separability limit and should be small in practice. This modification is equivalent to building a generalization of eq 5 using a function parametrized by i

$$\tilde{O} = O + \lambda \sum_i (\mathbb{E}(O^i|\Omega_i) - O^i) \tag{7}$$

In practice, we use the following formula

$$\tilde{O}(\omega, M_s) = O(\omega) + \lambda \sum_i \sum_{k=1}^{M_s} \frac{O^i(\omega_i^k) - O^i(\omega)}{M_s} \tag{8}$$

The term multiplied by λ is of course still a control variate (its expectation value is zero). For the local energy a simplified expression $(E_L^i(\omega) - E_L^i(\omega))$ is presented in eq 24 when Ψ is a Jastrow−Slater function. This form has been found to be much more efficient numerically (the scaling is $O(1)$ and the variance is about the same).

We need to optimize M_s to maximize the efficiency. The efficiency of a Monte Carlo calculation is related to the required time to achieve a given statistical uncertainty σ. With a sample of size M the statistical uncertainty is σ with

$$\sigma^2 = \frac{Vc}{M} \tag{9}$$

where V is the variance and c is a correlation factor ($c \geq 1$) which takes into account that the points in the sample are not independent. The total CPU time is $T = Mt$ where t is the time for one Monte Carlo step (a sweep over all electrons). The method is all the more efficient that the cost parameter

$$\sigma^2 T = Vct \tag{10}$$

is small. Note that this parameter is independent of T for a simulation sufficiently long. Given a random variable $\tilde{O}$ parametrized by M_s all the parameters on the right-hand side of eq 10 are functions of M_s, including t, because the optimal parameters of the main walk may depend on $\tilde{O}$. It is natural to consider that, in the limit $M_s = 0$ (no sidewalk), we have $\tilde{O} = O$. If t_s is the CPU time of the sidewalk with $M_s = 1$, the total computational time for a general value of M_s is $t(M_s) + M_s t_s$. The gain in efficiency is

$$\begin{aligned} G(M_s) &= \frac{V(O)t(0)c(0)}{V(\tilde{O})(t(M_s) + M_s t_s)c(M_s)} \\ &= \frac{1}{r(M_s)\left(1 + M_s \frac{t_s}{t(M_s)}\right)} \frac{c(0)t(0)}{c(M_s)t(M_s)} \end{aligned} \tag{11}$$

where $r(M_s)$ is the reduction of the variance. The parameter $t_s/t(M_s)$ should be negligible for large N, so for a large molecule M_s should approach infinity and asymptotically the gain is

$$G_\infty = \frac{1}{r_\infty} \frac{c_0}{c_\infty} \frac{t_0}{t_\infty} \tag{12}$$

where $r_\infty = \lim_{M_s \to \infty} r(M_s)$ and so on. If the three factors (ratio of variances, ratio of correlation factors, ratio of CPU times of a single step of the main walk) for large M_s are transferable from an atom to a molecule, the gain in numerical efficiency for large molecules can be estimated using single atoms. Next, we will consider isolated atoms before checking this transferability property.

3. REDUCTION OF THE VARIANCE FOR SINGLE ATOMS

We consider in this section and the next two sections only a single determinant wave function obtained in a self-consistent field calculation (SCF). It is built of Slater atomic orbitals, and the motivation is to understand the properties of the algorithm with the most basic wave functions in quantum chemistry while retaining the bottleneck of the computational cost we want to lower. We first investigate a series of isolated atoms to understand the properties of the core subsampling as a function of Z and the number of core electrons n_{core}. The subsampling can be done with any method involving the Metropolis scheme but with an additional rejection step when a move is leaving Ω. Such a rejection does not modify the detailed balance property ensuring that $\pi = \Psi^2$ remains the invariant measure of the subsampling process.

The simulations have been carried out for the elements Li ($Z = 3$) to Ar ($Z = 18$) with a varying number of core electrons and $M_s = 100 \cdot n_{\text{core}}$ steps so that they are converged or very close to the converged value $\mathbb{E}(O|\Omega)$. This limit can be better estimated with an hyperbolic fit as detailed in eq 36. The results as a function of the fraction of core electrons x are shown in Figure 2. First, one can see clearly that the variance converges correctly toward zero for an increasing number of core electrons. The zero-variance limit is obtained when all electrons are included in

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

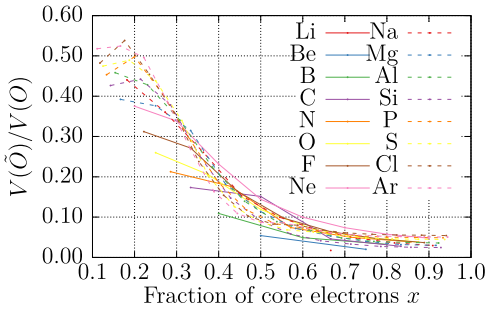


Figure 2. Reduction of the variance $V(\tilde{O})/V(O)$ as a function of the fraction of core electrons x. $M_s = 100 \cdot n_{core}$ for all simulations using an SCF wave function.

the subsampling and for an infinitely long sidewalk. Of course, in this limit the subsampling is exactly equivalent to the main walk itself, and the computational efficiency is not improved. However, it becomes apparent that for all values of Z the two inner electrons contribute to most of the variance, decreasing from 95% for lithium to 50% for argon.

This trend is analyzed in more detail in Figure 3, where results are shown for $n_{core} = 2$ and 10 for $Z = 11-18$, which is equivalent

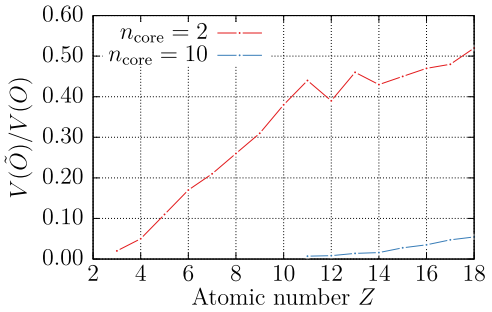


Figure 3. Reduction of the variance $V(\tilde{O})/V(O)$ for various elements with $n_{core} = 2$ ($M_s = 200$) and $n_{core} = 10$ ($M_s = 1000$) using an SCF wave function.

to the chemical core for the second period. One can see that the gain in variance is increased for the larger number of core electrons.

The parameters n_{core} and M_s should be as small as possible to obtain numerically cheap sidewalks, but they also have to be as large as possible to reduce the variance as much as possible, leading to an optimal compromise which has to be determined. In Figure 4 the convergence of $V(\tilde{O})/V(O)$ with M_s is shown for

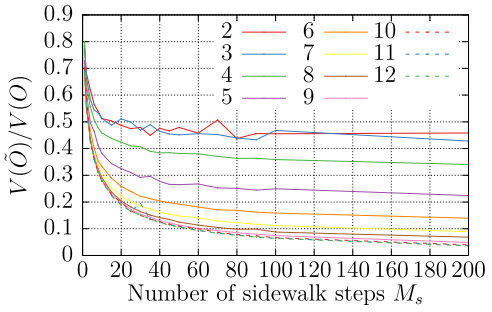


Figure 4. Reduction of the variance $V(\tilde{O})/V(O)$ as a function of the number of subsampling steps M_s for various numbers of core electrons n_{core} for a single aluminum atom (SCF wave function).

a range of numbers of core electrons. Independently of the size of the subsystem, one can distinguish an initial quick decay in the range of 0−20 steps and a $1/M_s$ convergence to the asymptotic limit (see eq 36). In this initial phase most of the reduction in the variance is obtained, and the differently sized subsystems that look very similar for a small number of steps increasingly separate from each other.

4. TRANSFERABILITY TO MOLECULES, GAIN IN VARIANCE, AND COMPUTATIONAL TIME

4.1. Gain in Variance. We first check for systems of many atoms that the gain in the variance for $M_s = \infty$ for a single atom that is transferred to molecular systems. We use linear alkanes C_nH_{2n+2} of increasing length with $n = 1-35$ and an increasingly larger part of a silicon unit cell ($Fd3m$)[10] with 1−8 atoms. The results for the alkanes are shown in Figure 5 in comparison to the

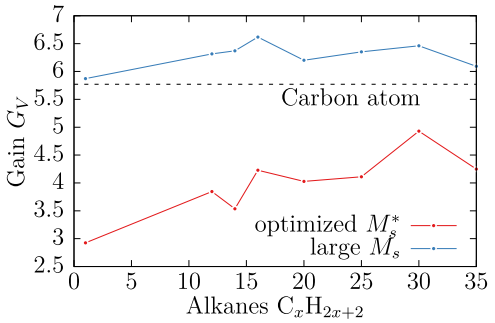


Figure 5. Gain in the variance G_V for alkanes of increasing length for converged sidewalks (large M_s) and for the optimized M_s^*. The dashed line is the gain for a converged sidewalk for a single carbon atom.

gain for a single carbon atom represented by the dashed line. The gain for converged sidewalks, i.e., with large M_s shown in blue, does not change with the length of the alkane chain. Furthermore, one can see that the gain differs by only 5%−10% from the single carbon atom case. This result confirms the transferability of the gain in variance from an atom to a molecule. It is even systematically slightly better (by 5%−10%) for the molecule; this suggests that the separability between the core and the valence regions is enhanced by the chemical bonds. Next, the results for the silicon clusters are shown in Figure 6. Again, the gain in the variance does not change with the system size, and it is about 20% above the gain of 82 for a single silicon atom. Compared to carbon the gain in the variance is about 17 times larger for silicon. Note that the gain on these curves looks

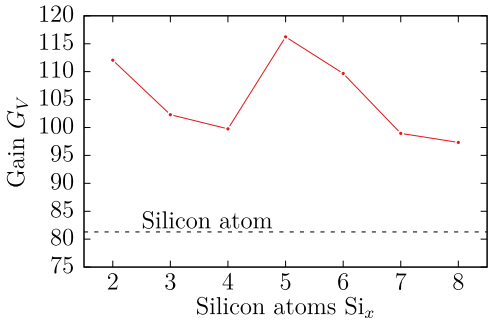


Figure 6. Gain in the variance G_V for silicon clusters of increasing size for converged sidewalks (large M_s). The dashed line is the gain for a converged sidewalk for a single silicon atom.

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

like it has rather large fluctuations (∼10%). These fluctuations come mainly from the infinite variance of the estimator of $V(E_L)$, and we cannot rely on the central limit theorem (the rate of convergence is slower than $1/\sqrt{M}$). In Figure 5, the gain in variance is shown for the optimized value M_s^*, which minimizes the computational cost. This gain increases from about 3 for CH_4 to about 5 for 30 carbon atoms.

4.2. Two Time Step Process and Gain in Computational Time. The process we use corresponds to the short-time propagator involved in diffusion Monte Carlo,[9] which is

$$G(\tau) = \Psi e^{-\tau(H-E_L)}\frac{1}{\Psi}$$

and is sampled using a drift and diffusion process on each individual electron i having the three spatial coordinates $\mathbf{r}_i$

$$\mathbf{r_i}(t+\tau) = \mathbf{r_i}(t) + \frac{1}{2}\frac{\nabla_i\Psi}{\Psi}\tau + \sqrt{\tau}\mathbf{W_i} \quad (13)$$

where $\sqrt{\tau}\mathbf{W_i}$ is the three-dimensional Wigner process, the diffusion term which reproduces the kinetic energy term in the Hamiltonian. It is easy to check that Ψ^2 is the stationary density of this process for an infinitesimal τ. A Metropolis accept−reject step is performed to ensure that Ψ^2 remains the stationary density for a finite τ. In the main walk we, however, use two time steps, one small τ_c (which should adapted to the core region) and one large τ_v (for the valence region), with a variable frequency. The two time steps are optimized to maximize the numerical efficiency (see eq 10). The sidewalk is also carried out with the same Metropolis algorithm with its own time step τ_s which can be optimized independently. With the bare or usual estimator (no sidewalk) only one time step is optimal. It is small and determined by the core degrees of freedom. Applying sometimes the process with a larger time step does not improve the correlation factor for the bare estimator. This seems to be counterintuitive because with a large time step the valence electron move much faster, while moves in the core are rejected at a negligible cost ($O(1)$ to $O(N)$ lower than the $O(N^2)$ scaling upon acceptation and updating the wave function with the Sherman−Morrison formula). One may ask why a procedure which displaces the valence electrons much faster does not improve ergodization and efficiency. We will first answer this question before explaining why this is different with the improved estimator.

First, we should be in a situation where the set of valence and core coordinates $\mathbf{R}_v$ and $\mathbf{R}_c$ are not very correlated because of the different scales. If we also have the separation in the local energy expression $E_L = f(\mathbf{R}_c) + g(\mathbf{R}_v)$, with $\mathbf{R}_c$ independent of $\mathbf{R}_v$, large time steps should clearly improve the ergodization. Small time steps would make the process ergodic in the core region, large time steps would make the process ergodic in the valence region, and the correlation factor would be small for both terms representing the local energy. We are clearly not in this situation, but instead E_L can be modeled as a sum of products of the kind $f(\mathbf{R}_c)g(\mathbf{R}_v)$ where $g(\mathbf{R}_v)$ is a low variance mode. E_L moves much faster by displacing the core electrons than the valence electrons. As a result better ergodization in the configuration space does not translate to a better ergodization for the random variable, and large time steps are only increasing the computational time.

For the improved estimator the situation is different since the dependence on $\mathbf{R}_c$ is removed by the subsampling process. Here, large time steps improve the ergodization of sampling the improved estimator; see the optimized time steps in Table 3. In short, the process is hard to improve for the bare estimator but can be easily improved for the improved estimator.

Overall, the numerical gain is larger than the simple gain in the variance; see Figure 7. The speedup, i.e., the gain in efficiency, is

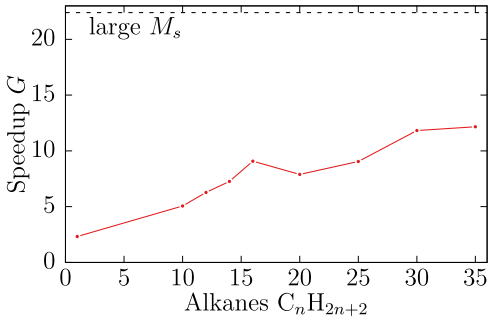


Figure 7. Speedup G for alkanes of increasing length with an SCF wave function. The asymptotic limit (dashed line) is estimated according to eq 12.

the product of the gain in the variance and the gain in the correlation factor, but it includes also a reduction of the computational time of the main walk. In practice, the reduction of the CPU time for a step of the main walk comes from a reduction of the acceptance probability from 0.95 to 0.57 for alkanes and from 0.93 to 0.42 for silicons. These acceptance probabilities do not depend on the size of the molecules within the error bars, which confirms that they are transferable. The corresponding gains in the CPU times t_0/t_∞ are 0.95/0.57 = 1.7 for alkanes and 0.93/0.41 = 2.3 for silicons. Finally, the correlation factors are reduced from 4.4 to 1.8 for the alkanes and 4.5 to 2.1 for the silicons, which results in gains of 2.4 and 2.1, respectively. Again, within the error bars the correlation factor does not depend on the system size and is transferable. Overall, this leads to an asymptotic value of the gain G which is ≃22 for carbon atoms. The numerical gain increases from $G \simeq 2$ for CH_4 to $G_\infty \simeq 12$ for 35 carbon atoms. The gain converges as $O(1/N)$ with the number of nuclei (see eq 44), and the ideal M_s^* increases as a linear function of N in accordance to eq 45. Based on the linear fit in Figure 8 for the alkane chains, $M_s^* = 100$ is for

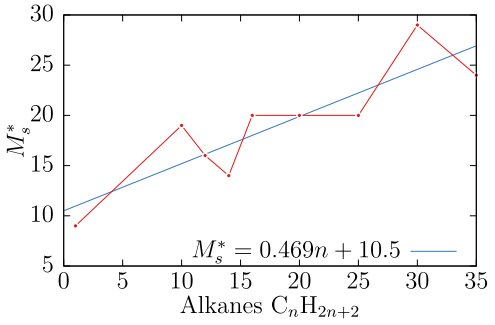


Figure 8. Ideal M_s^* for the alkane chains with an SCF function; a linear fit is done according to eq 45.

example reached for about 192 carbon atoms. One can estimate the asymptotic value of the gain $G_\infty \simeq 420$ for large clusters of silicon. The large reduction of the variance for silicon is due to the large atomic charge Z but also the choice of large cores with 10 electrons. If the asymptotic limit is much better it should also be reached for much larger molecules according to eq 44, not only because r_∞ is smaller but also because t_s is larger (by a factor

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

of 10). The gain of ≃4.5 which we observe for Si_8 is still very far from the asymptotic limit, and it is the same as for an alkane molecule of equal size.

5. JASTROW FACTOR

In QMC calculations the wave function commonly includes a Jastrow factor. Therefore, we carried out additional simulations for both atoms and molecules to assess the effect on the efficiency of the subsampling approach. We use here a very simple Jastrow J ensuring Kato's electron−electron cusp condition. With that wave function the local energy has no singularity when two electrons are close

$$\Psi = J \det(A) \quad \text{with} \quad \log(J) = \sum_{ij} \frac{\alpha_{ij} r_{ij}}{1 + \beta r_{ij}} \tag{14}$$

where r_{ij} is the distance between electron i and electron j, α_{ij} = 0.5 if the electrons are of different spins, and α_{ij} = 0.25 for same spin electrons. Only the parameter β is optimized to lower the total energy. In Figure 9 one can assess the effect of the Jastrow

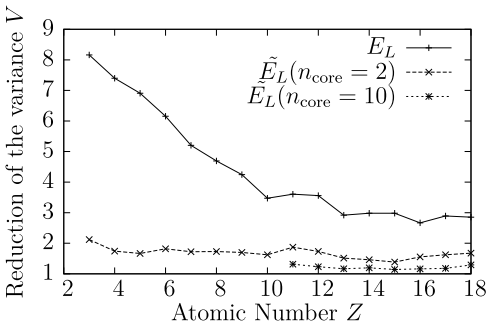


Figure 9. Reduction of the variance V upon multiplication of the wave function with a Jastrow factor.

factor on the variance. The variance of the bare estimator $V(O) = V(E_L)$ is reduced by a factor of about 8 for lithium, decreasing to about 3 for argon. The quality of the Jastrow factor decreases with increasing atomic number Z mainly because we used a very simple Jastrow factor consisting only of an electron−electron term. The Jastrow factor lowers the variance of the improved estimator with large M_s only by a factor of 1.5−2.0 (for $n_{\text{core}} = 2$) without any dependence on Z and 1.1−1.3 (for $n_{\text{core}} = 10$). Consequently, in the asymptotic limit the gain in the variance G_V due to the subsampling is reduced by a factor of 4 (Li) to 1.5 (Ar) for $n_{\text{core}} = 2$ and by a factor of 2−3 for $n_{\text{core}} = 10$. The results show that this simple Jastrow has a decreasing effect with Z while the subsampling method has a stable effect as a function Z for a given number of core electrons. The variance of $\tilde{O}$ is less sensitive to the quality of the wave function than O. This could be particularly interesting for the optimization of wave functions when the variance is very high due to a large number of parameters which can be far from the optimum.

The effect on the overall speedup G is shown on the example of the alkane chains and silicon clusters in Figures 10 and 11. One can see that the gain with and without Jastrow are almost the same and that the gain is even slightly larger for smaller systems. This is because the reduced gain in the variance G_V is compensated for by an increased gain in the correlation factor G_c (see Tables 1 and 2). With the Jastrow factor, the time step of the subsampling τ_s is reduced, and the time step τ_c slightly increases, while τ_v increases substantially up to a factor of 2.3 for silicon (see Table 3). The Jastrow term increases the correlation factor

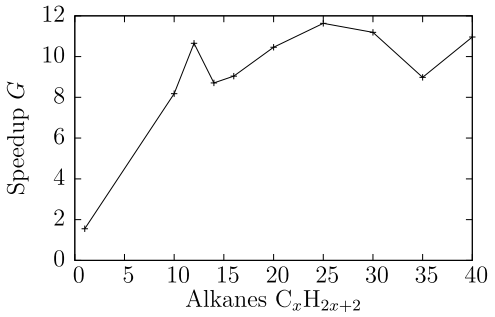


Figure 10. Speedup G for alkanes of increasing length using a Jastrow factor.

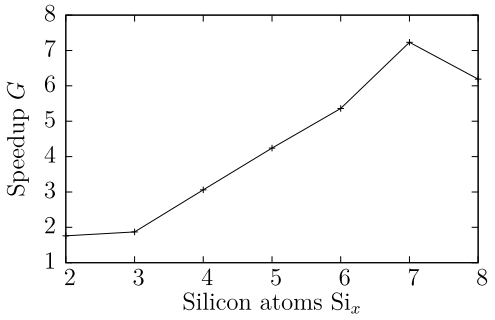


Figure 11. Speedup G for silicon clusters using a Jastrow factor.

of the bare estimator by a factor of 2 for the alkanes and 1.5−2.0 for the silicon clusters, but the correlation factor of the improved estimator is not affected. For the small silicon clusters the computational time is also better due to a reduced M_s. This is because the gain in the variance G_V converges more quickly toward the asymptotic limit according to eq 41 as r_∞ is smaller. The gain in the asymptotic limit G_∞ can be easily estimated to be 13 and 200 for the alkanes and silicon clusters, respectively, about half of the asymptotic gain without Jastrow.

6. COMPUTATIONAL DETAILS

A single iteration (of the main walk or sidewalk) consists of a usual drift (logarithmic derivative of Ψ) and Brownian diffusion (eq 13). We are using two time steps τ_c and τ_v for the main walk and τ_s for the sidewalk. The time steps τ_s and τ_c are small and adapted to the core electrons. The time step τ_v is adapted for the valence electrons and is therefore much larger. The optimal time steps are displayed in Table 3. The frequency ν_c of using the small time step τ_c is also a parameter to be optimized. The complete scheme is displayed in Figure 12. We remember that a main walk with two time steps but without subsampling has not been found to be more efficient than a simple main walk with a single time step. The sidewalk recovers most of the information on the core, which gives us the flexibility to move the core electrons less frequently within the main walk. We applied the following simple protocol to obtain optimized parameters for the three time steps τ_v, τ_s, and τ_c and the ideal number of subsampling steps M_s. All optimizations are carried out with the alternative small time step τ_s. First, the time step of the sidewalk is determined for a sidewalk with large M_s = 100 by minimizing the variance of the improved estimator and with a rough estimate of τ_v and ν_s = 1. Next, the time step τ_v of the main walk (including subsampling) is optimized by minimizing the cost (see eq 10) with ν_s = 1. A single simulation with large M_s allows the results for all possible shorter sidewalks to be extracted and,

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Speedup G, Gain in Variance G_V, Gain in Correlation Factor G_c, and Gain in CPU Time G_t for Alkanes C_nH_{2n+2} without and with the Jastrow Factor

	G		G_V		G_c		G_t	
n	—	Jastrow	—	Jastrow	—	Jastrow	—	Jastrow
1	2.48	1.55	2.93	1.43	2.29	4.69	0.37	0.23
10	5.04	8.18	3.98	2.27	2.47	5.23	0.51	0.69
12	5.95	10.65	3.85	2.54	2.44	5.21	0.63	0.80
14	7.31	8.71	3.54	2.34	2.61	5.00	0.79	0.74
16	8.19	9.04	4.23	2.28	2.59	5.05	0.75	0.79
20	8.52	10.46	4.03	2.27	2.53	5.22	0.84	0.89
25	9.58	11.63	4.11	2.18	2.38	5.41	0.98	0.99
30	13.81	11.19	4.93	2.42	2.88	4.72	0.97	0.95
35	11.10	8.98	4.25	2.01	2.49	4.68	1.05	0.82

Table 2. Speedup G, Gain in Variance G_V, Gain in Correlation Factor G_c, and Gain in CPU Time G_t for Silicon Clusters with n Atoms without and with the Jastrow Factor

	G		G_V		G_c		G_t	
n	—	Jastrow	—	Jastrow	—	Jastrow	—	Jastrow
2	1.82	1.76	31.19	8.63	2.43	3.22	0.02	0.06
3	0.75	1.87	16.93	9.16	1.31	3.04	0.03	0.07
4	2.47	3.06	23.63	12.25	2.83	2.98	0.04	0.08
5	2.22	4.24	16.09	12.04	2.13	3.50	0.06	0.10
6	3.93	5.36	25.33	14.79	2.74	3.52	0.06	0.10
7	3.99	7.23	21.20	15.09	2.41	4.24	0.08	0.11
8	4.44	6.19	18.46	14.84	2.39	3.20	0.10	0.13

Table 3. Optimal Time Steps for the Simple Main Walk and the Two-Time Step Subsampling Scheme

	alkanes		silicons	
time steps	—	Jastrow	—	Jastrow
$\tau = \tau_c$	0.0198	0.0299	0.0065	0.0073
τ_v	0.47	0.83	0.47	1.08
τ_s	0.0110	0.0081	0.0065	0.0048

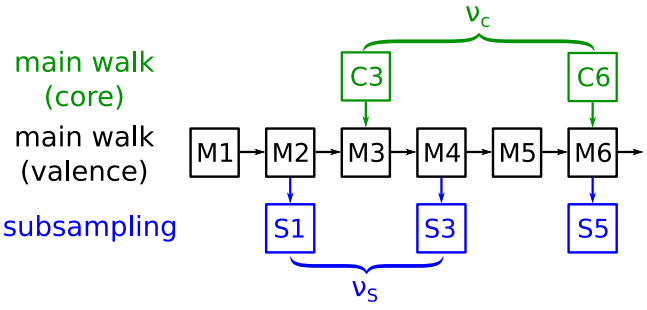


Figure 12. Complete scheme of the subsampling with two time steps with the main walk valence steps (**M**), the subsampling (**S**) with frequency ν_s, and the main walk core steps (**C**) with frequency ν_c.

therefore, M_s to be optimized. The correlation factor of the improved estimator is generally between 1 and 2 for these optimized parameters. Therefore, the sidewalk length M_s has been determined for the subsampling frequency ν_s of 1 and 2 (for the alkanes 10−16), but in all instances the former turns out to be more efficient. The parameters for molecular systems can be transferred from single atoms or small model systems; e.g., the parameters for a carbon atom in arbitrary alkanes can be determined from the CH_4 molecule. For heavier elements like silicon, the time step of the subsampling is simply identical to the time step of an ideal main walk without subsampling. The wave function Ψ comes from SCF calculations performed with Quantum Package.[11] The atomic basis is made of Slater atomic orbitals from ref 12, TZP for alkanes, and SZ for silicon clusters. The coefficients of molecular orbitals on the slater funtions have been computed using a large sum of Gaussians to be treated by Quantum Package.

7. CONCLUSION

In this work we are exploiting that a large number of the statistical fluctuations come from the core region and that the core regions are separable, to efficiently remove that amount. This is done using independent sidewalks for each core region, i.e., moving the core electrons while freezing the environment. This reduction of the variance triggers another gain in efficiency, and larger moves adapted to the valence electrons lower the correlation factor for the improved estimator with a reduced computational time. Unsurprisingly, with a simple Jastrow factor describing mostly the wave function in the core region, the information to be recovered by the subsampling is lower, and we observe a reduced gain in variance. However, this is compensated for by a better reduction of the correlation factors for a smaller numerical cost. For very large alkane chains and clusters of silicons, the overall gains can be estimated to be about 22 and 420, respectively, without the Jastrow factor. Including the Jastrow factor, the asymptotic overall gain is about half. The convergence to the asymptotic limit appears to be faster with the Jastrow factor; as our tests do not display a reduction of efficiency, we observe an improvement even on small silicon clusters. Overall, the presented method is a proof of concept for removing the effect of core electrons on the numerical cost in QMC calculations by sampling conditional expectation values. There are many more aspects left to explore for further improvements, like using alternative definitions of the core region, more elaborate wave functions, and different dynamics. One obvious next step to obtain physical meaningful results is to

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

adapt the scheme to other types of wave functions and to properties other than the energy, for example, its derivatives to optimize Ψ. Very intriguing is the perspective to extend this work to diffusion Monte Carlo.

APPENDIX A: SUBSAMPLING AND UPDATING SLATER DETERMINANTS

The wave function is built on p functions of $\chi^i(\mathbf{r})$ where $\mathbf{r}$ represents the three spatial coordinates of an electron and the spin $(\pm\frac{1}{2})$. Because they are usually centered on each atom, these functions are called atomic spin−orbitals. We suppose them to be localized; that is, they reduce to zero if the distance from a given atom is larger than a threshold. Given the configuration ω that is the N positions $\mathbf{r}_i$ of the electrons, we define X as the $N \times p$ rectangular matrix of spin−orbitals.

$$X_{ij} = \chi^j(\mathbf{r}_i) \tag{15}$$

A Slater determinant is

$$\Phi(X) = \det(XC) \tag{16}$$

where C is a $p \times N$ matrix of the so-called molecular orbital coefficients. The local energy like the drift can be written as a logarithmic derivative of Φ.[2] That property holds also if the Jastrow factor is included. Here we choose to separate the kinetic energy from the local potential energy

$$E_L = \partial_\lambda \ln \Phi\left(X - \frac{\lambda}{2}\Delta X\right) + \sum_{ij}\frac{1}{r_{ij}} + \sum_{iA}\frac{Z_A}{r_{iA}} \tag{17}$$

The first term is the kinetic energy, the second term is the electron−electron potential, and the third is the electron-nuclei potential (Z_A is the nuclear charge of the atom A). r_{ij} is of course the distance between electron i and electron j while r_{iA} is the distance between the electron i and the nucleus A. If C and X depend on a parameter λ

$$\partial_\lambda \ln \Phi = \mathrm{tr}(D\partial_\lambda X) + \mathrm{tr}((XC)^{-1}X\partial_\lambda C) \tag{18}$$

where

$$D \equiv C(XC)^{-1} \tag{19}$$

represents the logarithmic gradient of Φ with respect to X. A given configuration ω defines Ω which is subsampled by moving a few electrons of ω evolving in this way to $\omega' \in \Omega$. Correspondingly, if X' differs from X by a few lines, the determinant and its derivatives can be updated with efficient formulas. First we define the operator P, which when applied on the left selects those lines, and PX' are the lines which may differ from the lines of PX. We also define the operator Q^T which, when applied on the right of PX or PX', removes zero columns of PX and PX'.

$$\bar{X} \equiv PX'Q^T$$

Using the determinant lemma

$$\begin{aligned}\Phi(X') &= \det(XC)\det(PX'C(XC)^{-1}P^T)\\ &= \det(XC)\det(PX'Q^TQC(XC)^{-1}P^T)\\ &= \det(XC)\det(\bar{X}\bar{C})\end{aligned} \tag{20}$$

where $\bar{X}$ and $\bar{C}$ are submatrices of X' and D.

$$\bar{X} \equiv PX'Q^T \tag{21}$$

$$\bar{C} \equiv QC(XC)^{-1}P^T = QDP^T \tag{22}$$

The second term of the right-hand side of eq 20 is a Slater determinant for the subsystem with reduced numbers of electrons and atomic orbitals; the matrix $\bar{C}$ represents effective molecular orbitals. Equation 20 performs an update of the Slater determinant for the full system using a reduced Slater determinant.

Introducing

$$\bar{\alpha} \equiv (\bar{X}\bar{C})^{-1}$$

$$\bar{D} \equiv \bar{C}(\bar{X}\bar{C})^{-1}$$

the logarithmic derivative of eq 20 is

$$\begin{aligned}\partial_\lambda \ln \Phi(X') &= \mathrm{tr}(D\partial_\lambda X) + \mathrm{tr}(\bar{D}\partial_\lambda \bar{X}) + \mathrm{tr}(\bar{\alpha}\bar{X}\partial_\lambda \bar{C})\\ &= \mathrm{tr}(D\partial_\lambda X) + \mathrm{tr}(\bar{D}\partial_\lambda \bar{X}) - \mathrm{tr}(\bar{\alpha}\bar{X}QD\partial_\lambda XDP^T)\end{aligned}$$

Note that this expression does not depend on the lines of $\partial_\lambda X$ which are replaced by the lines of $\partial_\lambda X'$; in other words, $\partial_\lambda X$ can be replaced by $(1 - P^TP)X$ in eq 23. This property can be checked algebraically and will be used later. The control variate for the local energy is

$$\begin{aligned}E_L(X') - E_L(X) = \mathrm{tr}(\bar{D}\partial_\lambda \bar{X}) - \mathrm{tr}(\bar{\alpha}\bar{X}QD\partial_\lambda XDP^T)\\ + \sum_{i\in\Omega,j}\left(\frac{1}{r'_{ij}} - \frac{1}{r_{ij}}\right) + \sum_{i\in\Omega,A}\left(\frac{Z_A}{r'_{iA}} - \frac{Z_A}{r_{iA}}\right)\end{aligned} \tag{23}$$

where r'_{ij} and r'_{iA} represent distances from the electron i in the new configuration $\omega' \in \Omega$. The matrix $QD\,\partial_\lambda X\,DP^T$ is computed for a $O(N^3)$ cost and stored once for the sidewalk. Computing the control variate (eq 23) has an $O(N)$ scaling because of the two last Coulombic terms. Indeed the index i runs only on the electrons of the subsystem Ω (core electrons), but there are $N-1$ electrons j and $O(N)$ atoms A.

We can take instead an Ω-dependent approximation of E_L by considering only the interaction of the electrons in Ω with particles within a fixed distance from the center of Ω. This reduces the Coulombic sum to an $O(1)$ numerical cost; however, we expect a little effect on the statistical fluctuations. This is because we only neglect interactions with distant particles, distant dipoles, quadrupoles, or higher moments.

We propose also to remove the kinetic energy of particles which are not in the core region which defines Ω. Physically, if this core region is independent of the rest of the system, we can replace the one-body terms outside of Ω without modifying the difference (eq 23). Mathematically canceling the kinetic energy outside the core region is equivalent to replacing $(1 - P^TP)X$ by 0 in eq 23. This leads to the (zero-expectation-value) control variate

$$\begin{aligned}E_L^A(X') - E_L^A(X) = tr(\bar{D}\partial_\lambda \bar{X}) + \sum_{i\in\Omega,j}\left(\frac{1}{r'_{ij}} - \frac{1}{r_{ij}}\right)+\\ \sum_{i\in\Omega,A}\left(\frac{Z_A}{r'_{iA}} - \frac{Z_A}{r_{iA}}\right)\end{aligned}$$

where the sums over j and A are restricted to the electrons and the atoms in a sphere around the center of Ω. This expression is simpler as it has the same form as the expression of the local energy for the full system and is computationally less

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

demanding. These formulas apply with a Jastrow factor since the latter only modifies the definition of the derivative $\partial_\lambda X$.[2]

APPENDIX B: CONVERGENCE AS A FUNCTION OF M_s, THE SIZE OF A SIDEWALK

Given a set of random variables Ω, $\mathbb{E}(O|\Omega)$ is an unbiased estimator of $\mathbb{E}(O)$ since $\mathbb{E}(O) = \mathbb{E}(\mathbb{E}(O|\Omega))$. Let us prove that it is a variance-reduced estimator. The conditional variance is

$$V(O|\Omega) = \mathbb{E}(O^2|\Omega) - \mathbb{E}(O|\Omega)^2 \tag{24}$$

Now taking the expectation value of the two sides of this equation and isolating $\mathbb{E}(O)^2$ on the left-hand side we find

$$\mathbb{E}(O^2) = \mathbb{E}(V(O|\Omega)) + \mathbb{E}(\mathbb{E}(O|\Omega)^2) \tag{25}$$

which becomes, after removing $\mathbb{E}(O)^2$ on the two sides of this identity,

$$V(O) = \mathbb{E}(V(O|\Omega)) + V(\mathbb{E}(O|\Omega)) \tag{26}$$

The variance of the conditional estimator $\mathbb{E}(O|\Omega)$ is then lower than $V(O)$.

Here, O is the local energy for an atom and Ω is the set of coordinates of the valence electrons. In practice we perform a main walk and sidewalks to sample Ω, i.e., moving the core electrons while freezing the valence region

$$\tilde{O} = \frac{1}{M_s}\sum_k O(\omega_k) \tag{27}$$

For a given Ω the variance of $\tilde{O}$ is

$$V(\tilde{O}|\Omega) = \frac{V(O|\Omega)c_s(\Omega)}{M_s} \tag{28}$$

where c_s is a correlation factor which takes into account that the points on a sidewalk are not independent. We assume here that c_s depends only on Ω and not on M_s. This property holds in a regime where M_s is sufficiently large. Remember that

$$V(\tilde{O}|\Omega) = \mathbb{E}(\tilde{O}^2|\Omega) - \mathbb{E}(O|\Omega)^2 \tag{29}$$

We combine the two last equations and apply the expectation value

$$\begin{aligned}\frac{1}{M_s}\mathbb{E}(V(O|\Omega)\bar{c}_s) &= \mathbb{E}(\tilde{O}^2) - \mathbb{E}(\mathbb{E}(O|\Omega)^2)\\ &= \mathbb{E}(\tilde{O}^2) - V(\mathbb{E}(O|\Omega)) - \mathbb{E}(O)^2\\ &= V(\tilde{O}) - V(\mathbb{E}(O|\Omega))\end{aligned} \tag{30}$$

$$V(\tilde{O}) = V(\mathbb{E}(O|\Omega)) + \frac{1}{M_s}\mathbb{E}(\mathbb{E}(V(O|\Omega)c_s) \tag{31}$$

In the calculation we do not use $\mathbb{E}(O|\Omega)$ as an improved estimator; we use instead $\tilde{O}$ which converges to $\mathbb{E}(O|\Omega)$ for large M_s (ergodicity theorem). Equation 31 tells us that the variance of $\tilde{O}$ converges hyperbolically to the variance of $\mathbb{E}(O|\Omega)$. The variance of $\tilde{O}$ is a fraction $r(M_s) \leq 1$ of $V(O)$

$$r(M_s) = \frac{V(\tilde{O})}{V(O)} \tag{32}$$

which becomes the full gain only in the limit $M_s \to \infty$. Introducing the mean correlation time $\bar{c}_s$

$$\bar{c}_s \equiv \frac{\mathbb{E}(V(O|\Omega)c_s)}{\mathbb{E}(V(O|\Omega))} \tag{33}$$

eq 31 becomes

$$\begin{aligned}r(M_s) &= r_\infty + \frac{1}{M_s}\frac{\mathbb{E}(V(O|\Omega))}{V(O)}\bar{c}_s\\ &= r_\infty + \frac{1}{M_s}(1 - r_\infty)\bar{c}_s\end{aligned} \tag{34}$$

where we used eq 26 for the last expression. A hyperbolic fit of the function $r(M_s)$ can provide the two parameters r_∞ and $\bar{c}_s$. One can also use two values M_s and αM_s

$$r_\infty = \frac{r(M_s) - \alpha r(\alpha M_s)}{1 - \alpha} \tag{35}$$

$$\bar{c}_s = M_s\frac{r(M_s) - r_\infty}{1 - r_\infty} \tag{36}$$

For example, if $\alpha = \frac{1}{2}$

$$r_\infty = 2r(M_s) - r(M_s/2) \tag{37}$$

$$\bar{c}_s = M_s\frac{r(M_s) - r_\infty}{1 - r_\infty} \tag{38}$$

The explicit dependence on M_s should not make us forget that $\bar{c}_s$ is converging to a constant when M_s is sufficiently large.

APPENDIX C: OPTIMAL GAIN IN EFFICIENCY/COMPUTATIONAL TIME

Combining eqs 11 and 12, we can write when M_s is sufficiently large and the three gain factors do not depend anymore on M_s

$$G(M_s) = \frac{G_\infty}{r(M_s)\left(1 + M_s\frac{t_s}{t_\infty}\right)} \tag{39}$$

Maximizing the efficiency in eq 11 is equivalent to minimizing the function

$$\frac{1}{G(M_s)} = \left(1 + M_s\frac{t_s}{t_\infty}\right)\left(1 + \frac{1}{M_s}\frac{1 - r_\infty}{r_\infty}\bar{c}_s\right)\frac{1}{G_\infty} \tag{40}$$

The optimal value of M_s is

$$M_s^* = \sqrt{\frac{t_\infty}{t_s}\frac{\bar{c}_s(1 - r_\infty)}{r_\infty}} \tag{41}$$

for an optimal gain $G(M_s^*)$ such that

$$\frac{1}{G(M_s^*)} = \frac{1}{G_\infty}\left(1 + \sqrt{\frac{t_\infty}{t_s}\frac{\bar{c}_s(1 - r_\infty)}{r_\infty}}\right)^2 \tag{42}$$

For a large molecule, $\frac{t_s}{t_\infty}$ is small, and a first order Taylor expansion gives for $r_\infty \neq 0$

$$\frac{1}{G(M_s^*)} = \frac{1}{G_\infty}\left[1 + 2\sqrt{\frac{t_s}{t_\infty}\frac{\bar{c}_s(1 - r_\infty)}{r_\infty}} + o\left(\sqrt{\frac{t_s}{r_\infty t_\infty}}\right)\right] \tag{43}$$

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

When computing the variational energy with a Jastrow−Slater determinant and using the simplified control variate in eq 24, we have $t_s/t_\infty = O(N^{-2})$ and

$$\frac{1}{G(M_s^*)} = \frac{1}{G_\infty}\left(1 + o\left(\frac{1}{N\sqrt{r_\infty}}\right)\right) \tag{44}$$

Regarding the corresponding optimal value of M_s the Taylor expansion gives

$$M_s^* = O\left(N\frac{1}{\sqrt{r_\infty}}\right) \tag{45}$$

which means that the number of side steps (of the subsampling) for each core is optimally proportional to the number of cores.

AUTHOR INFORMATION

Corresponding Authors

Jonas Feldt − *Laboratoire de Chimie Théorique - UMR7616, Sorbonne Université & CNRS, 75005 Paris, France;* orcid.org/0000-0002-8361-6569; Email: jfeldt.theochem@gmail.com

Roland Assaraf − *Laboratoire de Chimie Théorique - UMR7616, Sorbonne Université & CNRS, 75005 Paris, France*; Email: assaraf@lct.jussieu.fr

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c01069

Funding

J.F. acknowledges the Deutsche Forschungsgemeinschaft (DFG) for financial support (Grant FE 1898/1-1).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Clark, B. K.; Morales, M. A.; McMinis, J.; Kim, J.; Scuseria, G. E. Computing the energy of a water molecule using multideterminants: A simple, efficient algorithm. *J. Chem. Phys.* **2011**, *135*, 244105.

(2) Filippi, C.; Assaraf, R.; Moroni, S. Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. *J. Chem. Phys.* **2016**, *144*, 194105.

(3) Assaraf, R.; Moroni, S.; Filippi, C. Optimizing the Energy with Quantum Monte Carlo: A Lower Numerical Scaling for Jastrow−Slater Expansions. *J. Chem. Theory Comput.* **2017**, *13*, 5273−5281.

(4) Feldt, J.; Filippi, C. In *Quantum chemistry and dynamics of excited states methods and applications*; Gonzalez, L., Lindh, R., Eds.; Wiley: 2020; pp 247−276.

(5) Burkatzki, M.; Filippi, C.; Dolg, M. Energy-consistent pseudopotentials for quantum Monte Carlo calculations. *J. Chem. Phys.* **2007**, *126*, 234105.

(6) Scemama, A.; Caffarel, M.; Benali, A.; Jacquemin, D.; Loos, P.-F. Influence of pseudopotentials on excitation energies from selected configuration interaction and diffusion Monte Carlo. *Results Chem.* **2019**, *1*, 100002.

(7) Umrigar, C. J. Accelerated Metropolis method. *Phys. Rev. Lett.* **1993**, *71*, 408−411.

(8) Nakano, K.; Maezono, R.; Sorella, S. Speeding up ab initio diffusion Monte Carlo simulations by a smart lattice regularization. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2020**, *101*, 155106.

(9) Toulouse, J.; Assaraf, R.; Umrigar, C. J. In *Electron Correlation in Molecules - ab initio Beyond Gaussian Quantum Chemistry*; Hoggan, P. E., Ozdogan, T., Eds.; Advances in Quantum Chemistry; Academic Press: 2016; Vol. 73, pp 285−314.

(10) Wyckoff, R. W. G. *Crystal Structures*, 2nd ed.; Interscience Pub./John Wiley & Sons: 1963.

(11) Garniron, Y.; Applencourt, T.; Gasperich, K.; Benali, A.; Ferté, A.; Paquier, J.; Pradines, B.; Assaraf, R.; Reinhardt, P.; Toulouse, J.; Barbaresco, P.; Renon, N.; David, G.; Malrieu, J.-P.; Véril, M.; Caffarel, M.; Loos, P.-F.; Giner, E.; Scemama, A. Quantum Package 2.0: An Open-Source Determinant-Driven Suite of Programs. *J. Chem. Theory Comput.* **2019**, *15*, 3591−3609.

(12) Van Lenthe, E.; Baerends, E. J. Optimized Slater-type basis sets for the elements 1−118. *J. Comput. Chem.* **2003**, *24*, 1142−1156.

https://dx.doi.org/10.1021/acs.jctc.0c01069
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX