# Calculation of space localized properties in correlated quantum Monte Carlo methods with reweighting: The nonlocality of statistical uncertainties

Roland Assaraf and Dominik Domin
*Laboratoire de Chimie Théorique, CNRS-UMR 7616, Université Pierre et Marie Curie Paris VI, Case 137, 4 place Jussieu, 75252 Paris Cedex 05, France*
(Received 3 December 2013; published 10 March 2014)

We study the efficiency of quantum Monte Carlo (QMC) methods in computing space localized ground state properties (properties which do not depend on distant degrees of freedom) as a function of the system size $N$. We prove that for the commonly used correlated sampling with reweighting method, the statistical fluctuations $\sigma^2(N)$ do not obey the locality property. $\sigma^2(N)$ grow at least linearly with $N$ and with a slope that is related to the fluctuations of the reweighting factors. We provide numerical illustrations of these tendencies in the form of QMC calculations on linear chains of hydrogen atoms.

DOI: 10.1103/PhysRevE.89.033304 PACS number(s): 02.70.Ss, 31.15.−p

## I. INTRODUCTION

Many important quantities of chemical or physical interest are localized in space; that is, they do not depend on spatially distant degrees of freedom. For example, the forces exerted on any particular nucleus in a molecule are barely influenced by the presence of neutral molecules that are very far away from the molecule of interest. One would expect that the distant neutral molecules would represent irrelevant degrees of freedom in computation of the force experienced by the nucleus and could be eliminated in the computation of such property. Quantum chemistry exploits locality in many different ways, for example, the Lewis description of covalent bonding [1] and the closely related valence bond theory [2,3]. Deterministic computational methods often exploit locality to tackle large systems. The most obvious strategy would be to study a fragment (for example, in a protein) and rely on the transferability of the results to a larger system. This would also be the idea behind coarse graining and hybrid methods. [4,5]

However, calculations on the full system are more robust since they do not depend on this transferability hypothesis. Exploiting the locality property to lower the cost of a calculation on the entire system is the strategy involved, for example, in local electronic structure methods [6–9]. The latter scale down the computational cost as a function of the system size $N$ (number of particles for a rather homogeneous system at a given scale, or a collection of identical systems). In such methods, molecular orbitals can be localized on single or spatially adjacent atoms according to a variety of criteria [10–12]. For large molecular systems, locality enables local correlation methods to reduce the computational scaling up to a linear dependence with $N$ [13,14].

Quantum Monte Carlo (QMC) methods [15–20], a powerful set of stochastic techniques for solving the Schrödinger equation, are increasingly used for electronic structure calculations on molecular systems. This is primarily because of the moderate computational scaling [$O(N^{3-4})$] and the perceived high accuracy to compute total energies. In the past decade there has been a similar drive to formulate and program linear-scaling QMC algorithms [21–27]. The focus of such works was to generate Monte Carlo sample configurations and evaluate energies with a reduced computational cost. However, another important factor in the numerical efficiency of a Monte Carlo method is the size of the statistical fluctuations in a calculated property. The purpose of this paper is to understand the behavior of QMC statistical fluctuations of spatially localized properties as a function of $N$. We show that conventional methods (correlated sampling methods) do not have the locality property regarding the statistical fluctuations.

Many properties, such as the force on a nucleus, dipole moment, or substitution energy, can be written as a difference of two ground state energies $E_\lambda - E_0$. $E_0$ and $E_\lambda$ are, respectively, the ground state energies of the Hamiltonians $H_0$ and $H_\lambda$. $H_\lambda$ is a small perturbation of $H_0$ and $\lambda$ is a small perturbation parameter,

$$H_\lambda = H + \lambda O. \tag{1}$$

When considering a localized property, $O$ depends mainly on the positions of particles lying in a small region of the space. We then have to compute a small difference in energies $E_\lambda - E_0$ or the energy derivative,

$$\langle O \rangle = \left. \frac{dE_\lambda}{d\lambda} \right|_{\lambda=0} \simeq \frac{E_\lambda - E_0}{\lambda}. \tag{2}$$

Computing these differences (2) from independent energy calculations is particularly inefficient in QMC. Since energy is size extensive, the statistical uncertainty on the energy usually behaves as $\sqrt{N}$. Consequently, the statistical uncertainty on such a calculation of (2) is

$$\sigma_i(\lambda, N, M) \propto \frac{1}{\lambda}\sqrt{\frac{N}{M}}. \tag{3}$$

This formula is valid asymptotically, for small $\lambda$, large system size $N$, and large sample size $M$. It is obvious from (3) that the smaller $\lambda$ is the less efficient independent energy calculations are. The so-called correlated sampling with reweighting methods [28–31], which are popular strategies to compute small differences of energies or properties, seem to be much more suitable methods. As we show later these methods encompass improved estimators which are built using the Hellmann-Feynman theorem [28,32–37]. We prove in this paper that the statistical uncertainty in correlated sampling with reweighting methods behaves as

$$\sigma_c(\lambda, N, M) \propto \sqrt{\frac{N}{M}}. \tag{4}$$

 

These methods are obviously better for small $\lambda$ and can provide derivatives with finite variances (see the section on zero-variance, zero-bias estimators in this paper). However, they have the same large $N$ behavior as independent energy calculations. Both $\sigma_c^2$ and $\sigma_i^2$ grow linearly with the system size. In other words, neither $\sigma_c$ nor $\sigma_i$ satisfy the locality property; they depend not only on the environment of the perturbation but also on irrelevant and distant degrees of freedom. The slope of the function $\sigma_c^2(N)$, however, can be lowered (using, for example, warped coordinates [31,33]).

The outline of this paper is as follows. In the first section we go through the basics of correlated sampling with reweighting method in the variational Monte Carlo (VMC) variant of QMC. We show that Hellmann-Feynman and improved estimators for observables [33] can be considered as a particular case with $\lambda \to 0$. We restrict our study to VMC since it allows for direct comparison with *ab initio* basis set calculations while it does not suffer from issues such as the fixed-node error or small time-step restrictions due to the Suzuki-Trotter expansion. Then we prove the nonlocality property of the statistical uncertainty (4) in the correlated sampling with reweighting method. It is shown that this nonlocal behavior comes from the weights which are introduced in the estimators. When considering Hellmann-Feynman estimators with the so-called zero-variance, zero-bias property, the nonlocality also comes from weights, the latter arising in the so-called Pulay correction. In the subsequent section simulations on hydrogen chains illustrate the size dependence of these methods for forces on a nucleus.

## II. CORRELATED SAMPLING IN VARIATIONAL MONTE CARLO

### A. Variational Monte Carlo

Given a Hamiltonian $H$ and a space of configurations $\mathbf{R}$ representing the coordinates of the particles (in our case the electrons), the VMC method computes the variational energy,

$$E_V[\Psi] = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \tag{5}$$

where $\Psi$ is a trial wave function depending on some variationally optimized parameters. The basic idea of the method is to rewrite the expression (5) in a form that has a probabilistic interpretation, the average of the so-called local energy, over the normalized probability distribution $\Psi^2$,

$$E_V[\Psi] = \frac{\int \Psi^2(\mathbf{R}) \frac{H\Psi}{\Psi}(\mathbf{R}) d\mathbf{R}}{\int \Psi^2(\mathbf{R}) d\mathbf{R}} = \left\langle \frac{H\Psi}{\Psi}(\mathbf{R}) \right\rangle_{\Psi^2} = \langle e \rangle_{\Psi^2}. \tag{6}$$

Note that VMC energies have become significantly more accurate during the last decade due to the development of efficient methods for optimizing the parameters of the wave function in Monte Carlo [38].

### B. The correlated sampling with reweighting method

We want to compute the difference in ground state energies of two systems described by the Hamiltonians $H$ and $H_\lambda$, the latter related to the first by small perturbation as in Eq. (1). We can introduce, respectively, two trial functions, $\Psi$, $\Psi_\lambda$, and two local energies, $e$, $e_\lambda$. The difference in the variational energies is

$$E_\lambda^V - E^V = \langle e_\lambda \rangle_{\Psi_\lambda^2} - \langle e \rangle_{\Psi^2}. \tag{7}$$

The latter expression can be computed either by independent energy calculations or by correlated sampling. The energy difference (7) is of order $\lambda$ and is usually asymptotically independent on the system size $N$ (for large $N$), when the perturbation is space localized. If $E_\lambda^V$ and $E^V$ are computed independently, the relative numerical precision of the difference $E_\lambda^V - E^V$ decreases with the system size. The numerical precision scales like $\frac{\sqrt{N}}{\lambda}$, because the statistical uncertainty on the energy scales as $\sqrt{N}$. Usual correlated methods consist of sampling only one density, for example, $\Psi^2$, and correcting the expectation value by inserting weights in the estimators,

$$E_\lambda^V - E^V = \frac{\left\langle e_\lambda \frac{\Psi_\lambda^2}{\Psi^2}(\mathbf{R}) \right\rangle_{\Psi^2}}{\left\langle \frac{\Psi_\lambda^2}{\Psi^2}(\mathbf{R}) \right\rangle_{\Psi^2}} - \langle e(\mathbf{R}) \rangle_{\Psi^2}, \tag{8}$$

which can be rewritten as

$$E_\lambda^V - E^V = \langle e_\lambda - e \rangle_{\Psi^2} + \frac{\mathrm{cov}\left(e_\lambda, \frac{\Psi_\lambda^2}{\Psi^2}\right)}{\left\langle \frac{\Psi_\lambda^2}{\Psi^2} \right\rangle_{\Psi^2}}. \tag{9}$$

In expression (9) we dropped the $\mathbf{R}$ dependence, and the covariance is defined on the $\Psi^2$ distribution. We underline here that the zero-variance, zero-bias estimators [33] of observables can be included in the same category of correlated sampling methods with reweighting (9). Indeed, the so-called zero-variance, zero-bias estimator of the energy derivative is no more than Eq. (9), with an additional factor $\frac{1}{\lambda}$ and in the limit $\lambda \to 0$. This estimator reads

$$\frac{dE_\lambda^V}{d\lambda} = \langle e'_\lambda \rangle_{\Psi^2} + 2\mathrm{cov}\left(e, \frac{\Psi'_\lambda}{\Psi}\right), \tag{10}$$

where the derivatives are taken with respect to $\lambda$. The first term of Eq. (10) is the so-called "zero-variance" estimator of the expectation value of the observable $O$ on the density $\Psi^2$; its main effect is to lower the statistical fluctuations [32,39]. The second is often called a "Pulay correction." It has the effect of lowering the bias coming from the variational approximation [33].

Expression (9) has the same quadratic zero-variance, zero-bias (ZVZB) property as expression (10): If $\phi$ and $\phi_\lambda$ are, respectively, the exact ground state of $H$ and $H_\lambda$, both the variance and the bias are of order 2 in the differences $(\Psi_\lambda - \phi_\lambda, \Psi - \phi)$.

That means that the statistical uncertainty can be in principle lowered arbitrarily (for a given sample) by appropriately choosing $(\Psi, \Psi_\lambda)$ in Eq. (9). However, as we show in a few sections, despite the usefulness of this variance reduction, the method suffers from analogous deficiencies as independent energy calculations when computing space localized properties for large systems.

### C. Generalization to correlated sampling with warped coordinates

The idea behind correlated sampling using warped coordinates is to compute the integrals in the expression of the

variational energy,

$$E_\lambda^V = \frac{\int \Psi_\lambda^2(\mathbf{R}) \frac{H_\lambda \Psi_\lambda}{\Psi_\lambda}(\mathbf{R}) d\mathbf{R}}{\int \Psi_\lambda^2(\mathbf{R}) d\mathbf{R}}, \tag{11}$$

by a substitution method. For that purpose one introduces warped coordinates,

$$\mathbf{R}_\lambda = T_\lambda(\mathbf{R}) = \mathbf{R} + \lambda \mathbf{v}(\mathbf{R}) + o(\lambda^2), \tag{12}$$

where $T_\lambda(\mathbf{R})$ is a well behaved (bijection and continuously differentiable) transformation, with the Jacobian $J_\lambda(\mathbf{R})$. Introducing the weight

$$w_\lambda(\mathbf{R}) = \frac{\Psi_\lambda^2}{\Psi^2}(T_\lambda[\mathbf{R}]) J_\lambda(\mathbf{R}), \tag{13}$$

one obtains the similar expression as Eq. (9),

$$E_\lambda^V - E^V = \langle e_\lambda(\mathbf{R}_\lambda) - e(\mathbf{R}) \rangle_{\Psi^2} + \frac{\mathrm{cov}(e_\lambda, w_\lambda)}{\langle w_\lambda \rangle_{\Psi^2}}, \tag{14}$$

for which Eq. (9) is a particular case with $\mathbf{R}_\lambda = \mathbf{R}$ and $J_\lambda(\mathbf{R}) = 1$.

The derivative of the last expression leads to one of the most general ZVZB estimator of $O$ proposed in Ref. [33] [Eq. (83)]. This can be written in a similar way as Eq. (10) except that, the $\lambda$ dependence of $\mathbf{R}_\lambda$ has also to be included, i.e.,

$$\left. \frac{dE_\lambda^V}{d\lambda} \right|_{\lambda=0} = \langle (e_\lambda[\mathbf{R}_\lambda])'_{\lambda=0} \rangle_{\psi^2} + \mathrm{cov}\{e, (w_\lambda[\mathbf{R}_\lambda])'_{\lambda=0}\}. \tag{15}$$

The two terms arising in the right-hand side of Eq. (15) have exactly the same nature as in Eq. (10); only the value of the weight is different. Introducing $\tilde{\psi}$, the derivative of $\psi$ with respect to $\lambda$ at $\lambda = 0$,

$$\psi_\lambda = \psi + \lambda \tilde{\psi} + o(\lambda), \tag{16}$$

the two terms of the right-hand side of Eq. (15), will be respectively written as

$$\frac{dE_\lambda^V}{d\lambda} = \langle e'(\tilde{\psi}, \mathbf{v}) \rangle + \mathrm{cov}[e, w'(\tilde{\psi}, \mathbf{v})] \tag{17}$$

$$= ZV(\tilde{\psi}, \mathbf{v}) + P(\tilde{\psi}, \mathbf{v}). \tag{18}$$

Note that with these conventions, $ZV(\tilde{\psi}, \mathbf{v})$ [first term of the right-hand side of Eq. (15)] might differ from the bare estimator expectation value $[ZV(\tilde{\psi} = 0, \mathbf{v} = \mathbf{0})]$, if warped coordinates are used ($\mathbf{v} \neq \mathbf{0}$). The term $P(\tilde{\psi}, \mathbf{v})$ is the covariance between the weight and the local energy. It is a generalized Pulay correction (Pulay correction when $\mathbf{v} = \mathbf{0}$). It includes the dependence on the weight. Having written improved estimators and correlated sampling estimators in the same way [expressions (10),(14),(15)], the subsequent discussion is applicable to both techniques.

### D. Fully separated model

We consider the limit where the system is composed of two independent subsystems. Mathematically, the total Hamiltonian is the direct sum of two Hamiltonians,

$$H_\lambda = H_\lambda^l + H^u. \tag{19}$$

The Hamiltonian $H_\lambda^l$ acts on a space of particle coordinates $\mathbf{R}^l$, for example, representing the electron positions of an isolated molecule in space. The Hamiltonian $H^u$ would represent an operator acting on a space of coordinates $\mathbf{R}^u$ of distant electrons from those treated by $H_\lambda^l$. The $\lambda$ dependence of the Hamiltonian is purely brought by the first term in Eq. (19), to take into account the spatial locality of the perturbation. In this context, a good variational wave function $\Psi_\lambda$ for the Hamiltonian $H_\lambda$ should obey the same separation property as the exact solution, namely,

$$\Psi_\lambda(\mathbf{R}) = \Psi_\lambda(\mathbf{R}^l, \mathbf{R}^u) = \Psi_\lambda^l(\mathbf{R}^l) \Psi^u(\mathbf{R}^u), \tag{20}$$

where $\Psi_\lambda^l$ and $\Psi^u$ are respectively variational solutions for the Hamiltonians $H_\lambda^l$ and $H^u$. The local energy of the system becomes

$$e_\lambda(\mathbf{R}) = e_\lambda^l(\mathbf{R}^l) + e^u(\mathbf{R}^u). \tag{21}$$

When introducing local warped coordinates, only the coordinates $\mathbf{R}_\lambda^l$ are transformed:

$$\mathbf{R}_\lambda = \left(\mathbf{R}_\lambda^l, \mathbf{R}^u\right) = [T_\lambda(\mathbf{R}^l), \mathbf{R}^u].$$

In the following sections, we omit the dependencies on the electron coordinates; they will be implicit: $\Psi_\lambda = \Psi_\lambda(\mathbf{R}_\lambda)$, $\Psi_\lambda^l = \Psi_\lambda^l(\mathbf{R}_\lambda^l)$, $\Psi^u = \Psi^u(\mathbf{R}^u)$, $e_\lambda^l = e_\lambda^l(\mathbf{R}_\lambda^l)$, $e^u = e^u(\mathbf{R}^u)$.

### E. Statistical fluctuations in a correlated sampling method with reweighting

Given the previous separability hypothesis [Eqs. (19) and (20)], formula (9) can be written as

$$E_\lambda^V - E^V = \left\langle e_\lambda^l - e_0^l \right\rangle + \frac{\mathrm{cov}(e_\lambda, w^l)}{\langle w^l \rangle} \tag{22}$$

$$= \left\langle e_\lambda^l - e_0^l \right\rangle + \frac{\mathrm{cov}\left(e_\lambda^l, w^l\right)}{\langle w^l \rangle} + \frac{\mathrm{cov}(e^u, w^l)}{\langle w^l \rangle}, \tag{23}$$

where the weight is defined as in (13):

$$w^l \equiv \frac{\Psi_\lambda^{l\,2}}{\Psi_0^{l\,2}}\left(\mathbf{R}_\lambda^l\right) J_\lambda(\mathbf{R}^l). \tag{24}$$

In expressions (22) and (23), all expectation values are computed over the sampled distribution ($\Psi^2 = \Psi_0^2$). Let us now evaluate the statistical fluctuations of (23) to analyze how they depend on the system size. The first two terms of the right-hand side of (23) are local, depending only on the positions $\mathbf{R}^l$ of the particles of the isolated fragment. As a consequence, the expectation values and the statistical fluctuations of their estimators do not depend on the system size.

We can now focus on the third term of the right-hand side of Eq. (23), which is nonlocal since it depends on $e^u$. We now evaluate the bias and the statistical fluctuations of this estimator, as a function of the system size.

### F. Estimator of the nonlocal contribution

The nonlocal contribution of Eq. (23) can be written in a simpler notation (we drop the subscripts for readability),

$$P = \frac{\mathrm{cov}(e, w)}{E(w)}, \tag{25}$$

where $w$ is positive and $e$ and $w$ are independent (since $e = e^u$ and $w = w^l$ depend on independent particle coordinates $\mathbf{R^u}$ and $\mathbf{R^l}$). Furthermore, we assume that $e$, $w$ have finite variance. Consider a sample of $M$ independent realizations of the triplet $(e,w,ew)$, namely a set $\{(e_i,w_i,e_iw_i), i \in [1,\ldots,M]\}$. For any set of $M$ independent random variables $X_i$ distributed as $X$ (here $X = e$, $X = w$, or $W = ew$), the mean which is an unbiased expectation value of $E(X)$ is written as

$$\bar{X} = \frac{1}{M}\sum_{i=1}^{M} X_i. \tag{26}$$

$P$ is a function of expectation values,

$$\begin{aligned} P &= f(E(ew),E(e),E(w)), \\ \text{where} \quad f(X,Y,Z) &= \frac{X - YZ}{Z}. \end{aligned} \tag{27}$$

The usual estimator of $P$ is the same function $f$ applied on the estimators of the expectation values,

$$\bar{P} \equiv f(\overline{ew},\bar{e},\bar{w}) = \frac{\overline{ew} - \bar{e}\bar{w}}{\bar{w}}. \tag{28}$$

Now we prove that the bias is zero (for a finite size sample $M$) if $e$ and $w$ are independent. From the expression (28) for $\bar{P}$,

$$E(\bar{P}) = E\left(\frac{\overline{ew}}{\bar{w}} - \bar{e}\right) = E\left(\frac{\sum_{i=1}^{M} e_i w_i}{\sum_{i=1}^{M} w_i}\right) - E(e). \tag{29}$$

Using the independence of $e_i$ and $w_i$, we obtain

$$\begin{aligned} E(\bar{P}) &= \sum_{i=1}^{M} E(e_i) E\left(\frac{w_i}{\sum_{j=1}^{M} w_j}\right) - E(e) \\ &= E(e)\sum_{i=1}^{M} E\left(\frac{w_i}{\sum_{j=1}^{M} w_j}\right) - E(e) \\ &= E(e)E\left(\sum_{i=1}^{M}\frac{w_i}{\sum_{j=1}^{M} w_j}\right) - E(e) = 0. \end{aligned}$$

Since $P = 0$ (because $e$ and $w$ are independent), the bias $E(\bar{P}) - P$ is equal to zero whatever the size $M$ of the sample. For $e$ and $w$ independent, an exact expression for the variance of $\bar{P}$ (cf Appendix) can be shown to be

$$V(\bar{P}) = \frac{V(e)}{M} E\left(\frac{\overline{w^2} - \bar{w}^2}{\bar{w}^2}\right) = \frac{V(e)V(w)}{E(w^2)M} + o\left(\frac{1}{M}\right). \tag{30}$$

The variance of $\bar{P}$ is therefore proportional to $V(e)$, which in turn is usually proportional to $N$. For large system sizes, the statistical fluctuations are dominated by the fluctuations of the nonlocal component [third term of Eq. (23)]. Hence, the variance does not exhibit the locality property.

In practice it is not possible to split exactly the system into independent fragments; thus, we have access to the expression (22) and not to the expression (23). The numerator of the latter is a covariance of nonindependent variables. An evaluation of the bias and the variance can be done in this general case (where $e$ and $w$ are not independent) as a power expansion in $\frac{1}{M}$. We do it in an Appendix for the sake of completeness.

## III. NUMERICAL RESULTS

### A. Model

Linear chains of hydrogen atoms are a convenient test bed for studying the large $N$ behavior of QMC methods in the context of highly correlated many electron problems. We provide illustrations on two different geometries. The first geometry is the one that minimizes the Hartree-Fock energy. The corresponding chain is then made of molecules of $H_2$ (hydrogen intermolecular distances around 1.4 a.u.) separated by distances of typically 6.5 a.u. This system is nonmetallic, displaying short range correlations with the electronic correlations decreasing exponentially fast as a function of the interparticle distance. The second geometry is made of equally separated hydrogen nuclei, with a distance equal to the equilibrium geometry of the $H_2$ molecule (1.4 a.u.). This chain is known to have metallic properties, i.e., long range correlations between electrons.

The most commonly employed wave function form in VMC is a product of a sum of antisymmetric Slater determinants multiplied by a symmetric correlation function (which contain explicit electron-electron, electron-nuclei, and higher order terms) [40–42]. Such wave functions provide a compact representation of the important static and dynamic electronic correlation effects compared to configurational interaction (CI) expansions, but generally result in integrals that cannot be integrated analytically, and must be performed numerically by Monte Carlo integration. In the present case since we are only interested in the system size dependence [43] of the statistical uncertainty, we limit ourselves to simple single determinant wave functions, obtained from restricted Hartree-Fock (RHF) calculations with the GAMESS package. This enables direct comparison with energies and gradient components obtained using *ab initio* deterministic methods. The basis set used is a Slater basis set expanded as a large expansion of Gaussian basis functions to facilitate treatment by the GAMESS *ab initio* package. All the VMC calculations are performed, using an overdamped Langevin process on 480 walkers with a time simulation of 6400 a.u. The statistics is done on 400 blocks of 480 trajectories that each span 16 a.u. of simulation time.

### B. Energies

Figure 1 shows the RHF energies as a function of the number of atoms in a chain for both geometries (metallic and insulating). The RHF energies display a linear behavior as a function of the number of atoms in the chain. This curve illustrates that the size extensivity holds almost perfectly even for the metallic chains and from a small number $N$ of atoms ($\sim$4–6). Note that, as expected, the energies obtained in VMC do not differ within the error bars from the deterministic *ab initio* calculations.

### C. Force estimators

The space localized property we focus on in this paper is the force component on the first nucleus of the chain in the direction of the chain axis ($z$ axis). The expression (1) of the perturbation ($\lambda O$) holds with $O$ being the $z$ component of the Coulombic force on the first nucleus, $\lambda$ being the magnitude of the displacement. This property is also the derivative of the

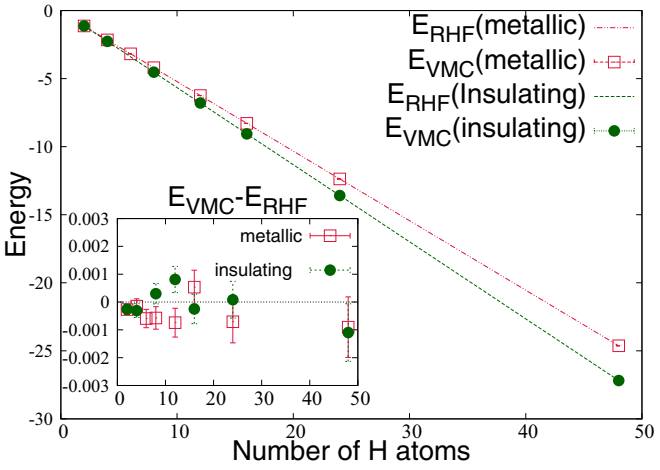


FIG. 1. (Color online) Total energies for linear $H_n$ molecules, insulating and metallic. The two curves represent spline interpolations of the RHF energies that were calculated with the *ab initio* deterministic package GAMESS. The points with the error bars are the VMC energies. The inset shows the deviation of the VMC energies from the RHF energies.

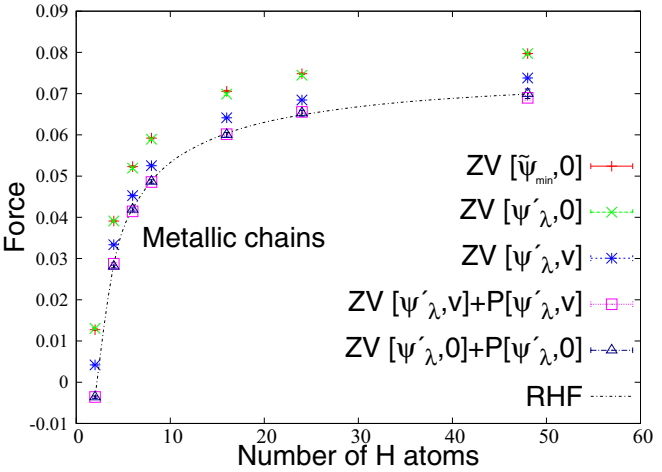


FIG. 2. (Color online) Energy derivative, different estimators, metallic chains.

energy with respect to the displacement of the first nucleus in the chain. We use different estimators coming from the general expression (18). We take two kinds of auxiliary function states $\tilde{\psi}$, $\tilde{\psi} = \tilde{\psi}_{\rm min}$ for a "minimal" choice [32,33] for which the estimator $e'(\tilde{\psi}_{\rm min},0)$ has a finite variance. Its expectation value $ZV(\tilde{\psi}_{\rm min},0)$ is the same that of the bare estimator (the usual Coulombic force).

When one chooses $\tilde{\psi} = \psi'_\lambda$, that is the derivative of the self-consistent solution $\psi$ with respect to the first nucleus $z$ component, then $e'(\psi'_\lambda,0)$ is also the derivative of the local energy with respect to the nucleus $z$ component. Also note that its expectation value, $ZV(\psi'_\lambda,0)$, does not differ from that of the bare estimator. Regarding the warped coordinates ($\mathbf{v} \neq \mathbf{0}$), we use the expression used in Ref. [33] with parameters optimized to lower the fluctuations of the weight $w'(\tilde{\psi},\mathbf{v})$. The generalized Pulay correction [second term of the right-hand side of (17)] is either $P(\psi'_\lambda,0)$ (the usual Pulay correction) or $P(\psi'_\lambda,\mathbf{v})$ (the generalized Pulay correction including the Jacobian of the warped coordinates).

### D. Force expectation values

Forces computed with different estimators are reported in Figs. 2 and 3. Similar results are observed for both metallic and nonmetallic systems. (i) For estimators that have the locality property for the statistical fluctuations: the zero variance with no warped coordinates [$ZV(\tilde{\psi}_{\rm min},0)$, $ZV(\psi'_\lambda,0)$] estimates are found to be the same within the error bars. These estimators should coincide with the target, i.e., the RHF forces; however, they are very different in practice. This is because we do not deal with the exact RHF solution, only an approximation coming from the incomplete basis used in the self-consistent field (SCF) calculation. Such behavior illustrates the sensitivity of the bare force to the quality of the variational wave function.

The zero-variance calculations with warped coordinates $ZV(\psi'_\lambda,\mathbf{v})$ are, as one would expect, different from the bare result. Interestingly, they are found to be closer to the target by a factor 2. We interpret this result as follows. In the limit where the fluctuations of the weight $w'(\tilde{\psi},\mathbf{v})$ are zero, the Pulay correction should be zero. In general, a substantial gain in the fluctuations of the weight should improve the accuracy of $ZV(\psi'_\lambda,\mathbf{v})$. In other words, warped coordinates lower the sensitivity of the expectation value of the modified zero-variance estimator $ZV(\psi'_\lambda,\mathbf{v})$, to the quality of the variational wave function. With the addition of the Pulay correction, we recover the derivative of the variational energy with respect to the nucleus position (the SCF force).

### E. Statistical fluctuations and size dependencies

The uncertainties obtained with zero-variance estimators (estimators without a Pulay correction) are reported in Fig. 4. Estimators containing the auxiliary function $\tilde{\psi} = \psi'_\lambda$ display large fluctuations in the statistical uncertainty, due to the formal infinite variance of these estimators (the nodes of $\psi'_\lambda$ differ from the $\psi$'s). However, the overall behavior does not

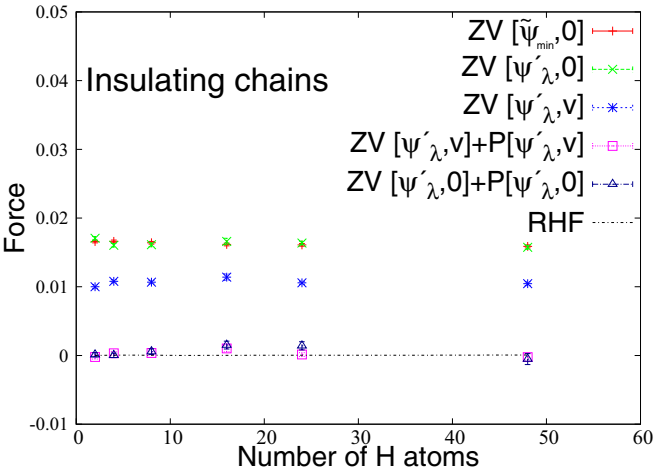


FIG. 3. (Color online) Energy derivative, different estimators, insulating chains.

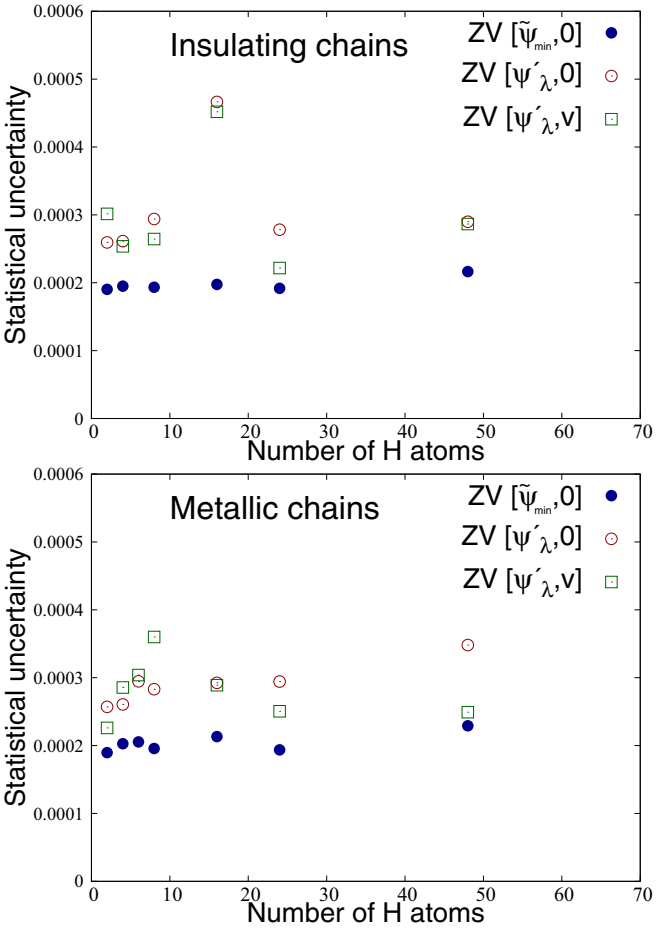


FIG. 4. (Color online) Statistical uncertainties for estimators without a Pulay correction. (Top) Insulating $H_n$ chains; (bottom) metallic $H_n$ chains.

appear to grow with system size. With the minimal auxiliary function $\psi_{\min}$, the independence of the statistical uncertainty with respect to the system size is more obvious (that comes from the finite variance of the estimator).

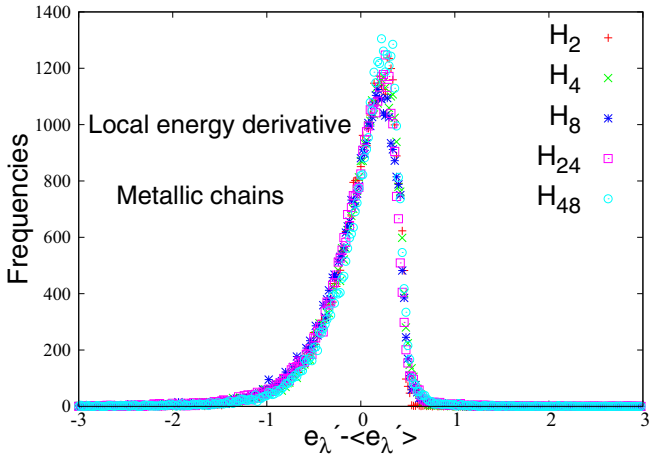


FIG. 5. (Color online) Histograms of the local energy derivative (ZV estimator) for the $H_n$ metallic chains.

FIG. 6. (Color online) Histograms of the local energy for the $H_n$ metallic chains.

To confirm the trend in a more quantitative way, we provide the histograms of the derivatives of the local energy with and without warped coordinates. The histogram of the local energy derivative (which has an infinite variance) for the metallic chain is presented in Fig. 5. The histograms converge rapidly with the number $N$ of atoms; they appear to be numerically

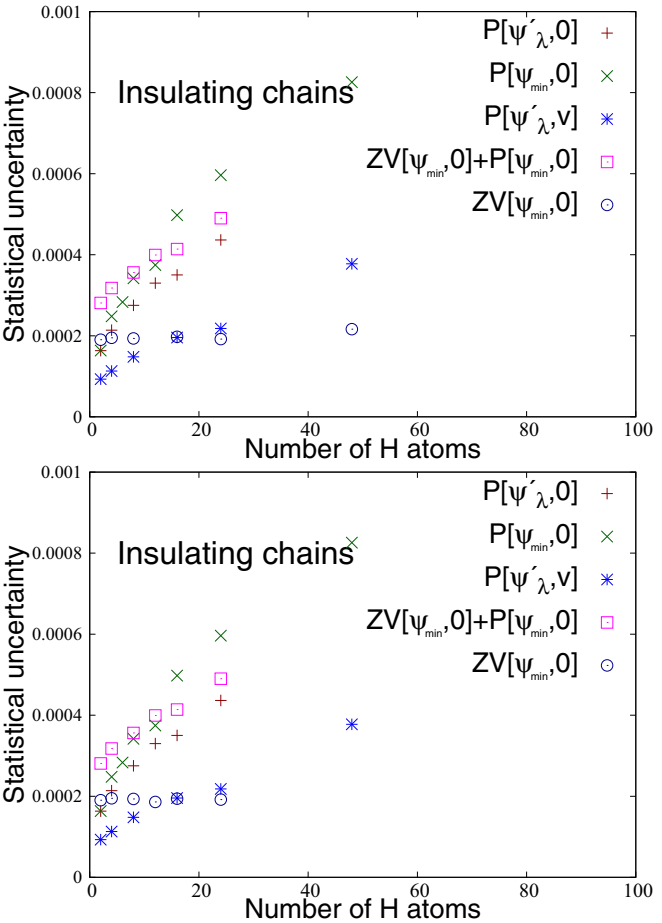


FIG. 7. (Color online) Statistical uncertainties on the Pulay correction. (Top) Insulating chains; (bottom) metallic chains.

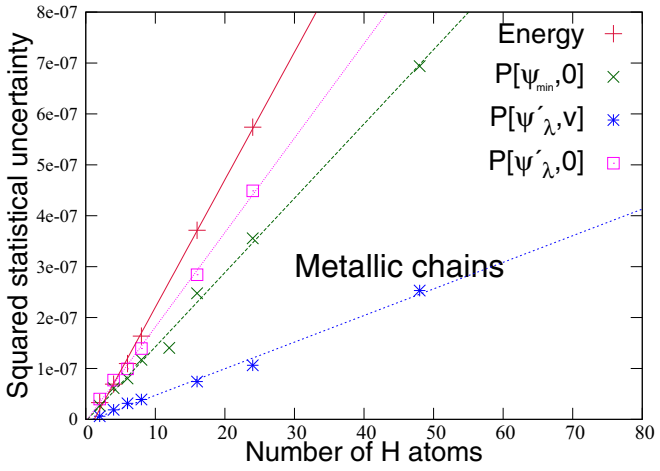


FIG. 8. (Color online) Square of statistical uncertainties as a function of the number of atoms (metallic chains). The lines are linear fits on the energy and minimal Pulay correction uncertainties squared.

the same for the $H_2$ and the $H_{48}$ molecules. This contrasts with the behavior of the local energy (see Fig. 6) which, as expected, has larger tails as $N$ increases. We conclude that the locality property of the statistical fluctuations of the zero-variance estimator is not limited to systems with short range correlations.

The uncertainties obtained in the Pulay corrections are reported in Fig. 7. Three different Pulay corrections are presented: the ones $[P(\psi'_\lambda,\mathbf{v})$ and $P(\psi'_\lambda,0)]$ that have already been introduced, but also a minimal Pulay correction $P(\tilde{\psi}_{\min},0)$ which has a finite variance. The Pulay corrections for both the metallic and insulating hydrogen chains appear to grow as a function of the system size. They grow as $\sqrt{N}$, just like the variational energy. This is illustrated in Fig. 8, where the square of the statistical uncertainties appears to be a linear function of $N$. Hence, this behavior which has been proven here for independent (noninteracting) systems, also holds for (strongly interacting) metallic hydrogen chains and thus is expected to hold for more general systems.

As a consequence, the statistical uncertainty on the ZVZB estimators are then dominated by the Pulay correction for large $N$, which grow as $\sqrt{N}$. In practice one can see on the Fig. 7 that the statistical fluctuations on the Pulay correction becomes dominant from 2 to 4 atoms without warped coordinates and from about 10 atoms (metallic chain) or 15 atoms (insulating chain) with warped coordinates. We note that the main role of warped coordinates is to lower the slope of the variance as a function of $N$.

## IV. CONCLUSION

Many physical properties are local in space, meaning that their expectation values converge as a function of the system size $N$. We have analyzed in this paper the statistical fluctuations as a function of system size displayed by usual QMC estimators of space local properties.

Correlated sampling techniques with reweighting (to compute small differences of energies) and ZVZB estimators (to compute energy derivatives) were handled in the same way. Both types of estimators were written as a sum of two terms. The first term is a difference (or derivative) of two local energies. In the context of improved estimators for observables (energy derivatives), this term is the so-called zero-variance estimator (which has the same expectation value as the bare estimator but a lower variance). When warped coordinates are introduced this difference involves two local energies computed at two different locations in the space of configurations, and the expectation value is different from the bare estimator. The second involves the covariance of the local energy and a weight and possibly includes a Jacobian (if warped coordinates are used). This second term is interpreted as a generalized Pulay correction.

In this paper, we have proved that for a sum of independent Hamiltonians (noninteracting fragments), the variance of the first term has the locality property, meaning that the fluctuations do not depend on irrelevant degrees of freedom. We have observed numerically for linear hydrogen chains that this property holds also for both interacting and noninteracting systems.

The variance of the second term (Pulay) does not have the locality property. This is because it is proportional to the variance of the local energy which usually grows linearly as a function of the system size $N$. The consequence is that the usual reported $N^2$-$N^3$ scaling of QMC methods (for the energy) has to multiplied by $N$ when local properties are to be computed. The variance of the Pulay term can be huge, if there is, for example, a very heavy atom far from the perturbation site. The heavy atom would not contribute to the expectation value but would dramatically increase the variance (since this heavy atom would have a large contribution to the variance of the local energy).

Note that this behavior might also be responsible for a deterioration of the accuracy for large $N$ when optimizing trial functions, since estimators of the derivatives of the energy involve a covariance of the weights and the local energies.

In summary, calculations of properties, energy derivatives or energy differences using available methods lead to statistical fluctuations depending on irrelevant degrees of freedom, and which grow with the system size. The slope of this growth, however, can be limited using warped coordinates as we have discussed and illustrated numerically.

To avoid these artificial statistical fluctuations, we see two possibilities. (i) The first is to limit the slope of the growth via warped coordinates and/or a very careful optimization of the wave function to lower the variance of the local energy. Here we did not improve the trial function for the hydrogen chains beyond the SCF solution. When we do improve it with a simple Jastrow enforcing some of the cusp conditions, a gain in the variance of the local energy by a factor 10 is typically achieved, further decreasing the slope of the variance on the Pulay correction. That means that the Pulay correction becomes only dominant for 10 times the size of the chain (here 50–100 atoms, which might be large enough for practical purposes). For a system with larger atoms, in an all-electron calculation, the contribution to the variance of a single large atom is much larger (even after optimization), and so is the slope; the predominance of fluctuations in the Pulay correction would arise for much smaller system sizes.

(ii) The second strategy would be to use the zero-variance estimator, i.e., dropping the Pulay correction. The inconvenience is that the bias is of order one (see Ref. [33]), rather than order two. Using optimized warped coordinates, however, is liable to lower the prefactor of the bias. Of course, a new method having the locality property for the statistical uncertainty while keeping a second order systematic error on physical properties would be potentially much more efficient.

## ACKNOWLEDGMENTS

We warmly thank Professor William A. Lester, Jr., for valuable discussions and making this collaborative project possible. We thank the France Berkeley Fund and the Mairie de Paris for financial support.

## APPENDIX A: ESTIMATOR OF THE PULAY CORRECTION—BIAS AND STANDARD ERROR

### 1. Estimator of the Pulay correction

We introduce the property

$$P = \frac{\mathrm{cov}(e,w)}{E(w)}, \tag{A1}$$

where $e$ and $w$ are random variables defined on some probability measure. We suppose that $w$ is positive and that $e$, $w$, and $ew$ have a finite variance.

Introducing a sample of $M$ independent realizations of the triplet $(e,w,ew)$, namely $\{(e_i,w_i,e_iw_i), i \in [1,\ldots,M]\}$. We note that for each random variable $X \in \{e,w,ew\}$ the mean

$$\bar{X} = \frac{1}{M}\sum_{i=1}^{M} X_i \tag{A2}$$

is an unbiased estimator of $E(X)$.

$P$ is a function of expectation values

$$P = f(E(ew),E(e),E(w)), \quad \text{where} \quad f(X,Y,Z) = \frac{X - YZ}{Z}. \tag{A3}$$

A usual estimator of $P$ is the same function applied on the estimators of the expectation values,

$$\bar{P} \equiv f(\overline{ew},\bar{e},\bar{w}) = \frac{\overline{ew} - \bar{e}\bar{w}}{\bar{w}}. \tag{A4}$$

This is the estimator we consider here.

### 2. Evaluation of the bias

The estimator (A4) has the following bias:

$$E(\bar{P}) - P = E[f(\overline{ew},\bar{e},\bar{w}) - f(E(ew),E(e),E(w))]. \tag{A5}$$

We evaluate (A5) and prove that it is generally nonzero. For that purpose we write the development of the function $f(X,Y,Z)$ around a point $(X_0,Y_0,Z_0)$,

$$f(X,Y,Z) = -Y + \frac{X}{Z_0}\sum_{n=0}^{\infty}(-1)^n\frac{(Z-Z_0)^n}{Z_0^n}. \tag{A6}$$

Using (A6) and making the identifications $(X_0,Y_0,Z_0) = [E(ew),E(e),E(w)]$, $(X,Y,Z) = (\overline{ew},\bar{e},\bar{w})$, we have

$$\begin{aligned}\bar{P} - P &= f(X,Y,Z) - f(X_0,Y_0,Z_0)\\ &= -[\bar{e} - E(e)] + \frac{\overline{ew} - E(ew)}{E(w)}\\ &\quad + \frac{\overline{ew}}{E(w)}\sum_{n=1}^{\infty}(-1)^n\frac{[\bar{w} - E(w)]^n}{E(w)^n}.\end{aligned} \tag{A7}$$

Note that this expression is well defined provided that $M$ is sufficiently large [44].

Writing $\bar{w}\bar{e} = E(\overline{we}) + [\overline{we} - E(\overline{we})]$ and reordering the sum order by order we have

$$\begin{aligned}\bar{P} - P &= -[\bar{e} - E(e)] + \frac{\overline{ew} - E(ew)}{E(w)} - \frac{E(ew)}{E(w)^2}[\bar{w} - E(w)]\\ &\quad + \sum_{n=2}^{\infty}(-1)^n E(w)^{-n}\left(\frac{E(ew)}{E(w)}\{[\bar{w} - E(w)]^n\}\right.\\ &\quad \left. - \{[\overline{ew} - E(ew)][\bar{w} - E(w)]^{n-1}\}\right).\end{aligned} \tag{A8}$$

Hence, the bias [Eq. (A5)] reads

$$\begin{aligned}E(\bar{P}) - P &= \sum_{n=2}^{\infty}(-1)^n E(w)^{-n}\left(\frac{E(ew)}{E(w)}E\{[\bar{w} - E(w)]^n\}\right.\\ &\quad \left. - E\{[\overline{ew} - E(ew)][\bar{w} - E(w)]^{n-1}\}\right).\end{aligned} \tag{A9}$$

To proceed further we need to evaluate $E\{[\bar{w} - E(w)]^n\}$ and $E\{[\overline{ew} - E(ew)][\bar{w} - E(w)]^{n-1}\}$. Both terms can be written in the form $(\bar{X}\bar{Y}^{n-1})$, where $\bar{X}$ and $\bar{Y}$ are centered random variables:

$$\begin{aligned}E(\bar{X}\bar{Y}^{n-1}) &= E\left[\frac{\sum_{i=1}^{M} X_i}{M}\left(\frac{\sum_{i=1}^{M} Y_i}{M}\right)^{n-1}\right]\\ &= \frac{1}{M^{n-1}}E\left[X_1\left(\sum_{i=1}^{M} Y_i\right)^{n-1}\right].\end{aligned} \tag{A10}$$

In the last expression, we used the hypothesis that the $X_i$ are equidistributed.

For $n = 1$ this reduces trivially to zero. We restrict the discussion to the case $n > 1$. Developing the $(n-1)$th power we obtain

$$\begin{aligned}E(\bar{X}\bar{Y}^{n-1}) &= \frac{1}{M^{n-1}}\sum_{\sum_{i=1}^{M} k_i = n-1}\frac{(n-1)!}{k_1!k_2!\cdots k_n!}\\ &\quad \times E\left(X_1 Y_1^{k_1} Y_2^{k_2}\cdots Y_M^{k_n}\right).\end{aligned} \tag{A11}$$

Using the independence of random variables for two different indices $i$ and the hypothesis that $Y_i$ are equidistributed, this expression reads

$$\begin{aligned}E(\bar{X}\bar{Y}^{n-1}) &= \frac{1}{M^{n-1}}\sum_{\sum_{i=1}^{M} k_i = n-1, k_1\neq 0, i>0 \Rightarrow k_i \neq 1}\frac{(n-1)!}{k_1!k_2!\cdots k_n!}\\ &\quad \times E(XY^{k_1})E(Y^{k_2})\cdots E(Y^{k_n}).\end{aligned} \tag{A12}$$

The indices $k_i$ in this summation are positive integers (including 0). The restrictions $k_1 \neq 0$ and $k_i \neq 1$ can be done since the corresponding terms are zero (all variables are centered). For the first values of $n$ one would obtain from this formula

$$
\begin{aligned}
E(\bar{X}\bar{Y}) &= \frac{1}{M}E(XY),\\
E(\bar{X}\bar{Y}^2) &= \frac{1}{M^2}E(XY^2),\\
E(\bar{X}\bar{Y}^3) &= \frac{1}{M^3}[E(XY^3)+3(M-1)E(XY)E(Y^2)],\\
E(\bar{X}\bar{Y}^4) &= \frac{1}{M^4}\{E(XY^4)+(M-1)[6E(XY^2)E(Y^2)\\
&\quad +4E(XY)E(Y^3)]\},\\
E(\bar{X}\bar{Y}^5) &= \frac{1}{M^5}\{E(XY^5)+(M-1)[10E(XY^3)E(Y^2)\\
&\quad +10E(XY^2)E(Y^3)+5E(XY)E(Y^4)]\\
&\quad +15(M-1)(M-2)E(XY)E(Y^2)^2\}.
\end{aligned}
$$

More generally, the number of terms in the brackets depends only on $n$. Each term is a product of some number $k$ of expectation values times a term growing like $M^{k-1}$ for large $M$. For a fixed even $n$ the largest value of $k$, is $n/2$ corresponding to the largest nonzero product, namely $E(XY)E(Y^2)^{\frac{n-2}{2}}$. For a fixed odd $n \geqslant 5$ the largest value of $k$ is $\frac{n-1}{2}$, corresponding to the product $E(XY)E(Y^2)^{\frac{n-5}{2}}E(Y^3)$.

Under the hypothesis that $E(XY) \neq 0$, then for large $M$

$$
\begin{aligned}
&E(\bar{X}\bar{Y}^{n-1}) \propto M^{-\frac{n}{2}}E(XY)E(Y^2)^{\frac{n-2}{2}}+o\big(M^{-\frac{n}{2}}\big), \quad \text{for } n \text{ even},\\
&E(\bar{X}\bar{Y}^{n-1}) \propto M^{-\frac{n+1}{2}}E(XY)E(Y^2)^{\frac{n-5}{2}}E(Y^3)+o\big(M^{-\frac{n+1}{2}}\big),\\
&\qquad n \geqslant 5, \ n \text{ odd}. \qquad \text{(A13)}
\end{aligned}
$$

Applying this result, the dominant term in expression (A9) as a function of $M$ is obtained for the index $n = 2$:

$$
\begin{aligned}
E(\bar{P})-P &= \frac{1}{M}E(w)^{-2}\left(\frac{E(ew)}{E(w)}E\{[w-E(w)]^2\}\right.\\
&\quad \left.-E\{[ew-E(ew)][w-E(w)]\}\right)+o\left(\frac{1}{M}\right)\\
&= \frac{1}{M}E(w)^{-3}[E(ew)E(w^2)-E(ew^2)E(w)]\\
&\quad +o\left(\frac{1}{M}\right)\\
&= \frac{1}{M}[\mathrm{cov}(ew,w^2)-\mathrm{cov}(ew^2,w)]+o\left(\frac{1}{M}\right). \qquad \text{(A14)}
\end{aligned}
$$

This term is, in general, nonzero. For example, if one takes $e = \frac{1}{w}$, one has $E(\bar{P})-P = \frac{1}{N}E(w)^{-3}[E(w^2)-E(w)^2] = \frac{1}{N}E(w)^{-3}V(w)$, which is general strictly positive.

We have proved that the dominant term of the bias as a function of $N$ is, in general, nonzero. Hence, the bias is, in general, nonzero. Furthermore, it is decreasing like the inverse of the size of the sample.

### 3. Standard error of the estimator

The standard error on the estimator of $P(\bar{P})$ is usually given by the square root of its variance provided that $M$ is large enough. The variance reads as

$$
\begin{aligned}
V(\bar{P}) &= E[\bar{P}-E(\bar{P})^2] = E[\bar{P}-P+P-E(\bar{P})^2]\\
&= [P-E(\bar{P})]^2+E[(\bar{P}-P)^2]. \qquad \text{(A15)}
\end{aligned}
$$

The last term $E(\bar{P}-P)^2$ is the square of the bias. Therefore, it is at least of order $\frac{1}{M^2}$. The previous term is the square of the average of (A8). It can be written as an expansion of powers of $\frac{1}{M}$. For the same justifications as before, the leading term is the one of order 2. This term is the square of the term of order 1 in (A15)

$$
E[(\bar{P}-P)^2] = \frac{1}{M}V\left[\frac{ew}{E(w)}-\frac{E(ew)}{E(w)^2}w-e\right]+o\left(\frac{1}{M}\right). \qquad \text{(A16)}
$$

The last term of (A15) is the square of the bias; thus, it is behaving as $\frac{1}{M^2}$. We finally have

$$
V(\bar{P}) = \frac{1}{M}V\left[\frac{ew}{E(w)}-\frac{E(ew)}{E(w)^2}w-e\right]+o\left(\frac{1}{M}\right). \qquad \text{(A17)}
$$


[1] G. N. Lewis, J. Am. Chem. Soc. **38**, 762 (1916).
[2] J. Gerratt, D. L. Cooper, P. B. Karadakov, and M. Raimondi, Chem. Soc. Rev. **26**, 87 (1997).
[3] S. Shaik and P. C. Hiberty, in *Reviews in Computational Chemistry* (Wiley & Sons, New York, 2004), pp. 1–100.
[4] P. W. Anderson, Science **177**, 393 (1972).
[5] W. G. Noid, J. Chem. Phys. **139**, 090901 (2013).
[6] P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[7] R. K. Nesbet, in *Advances in Chemical Physics* (Wiley & Sons, New York, 2007), pp. 321–363.
[8] O. Sinanoglu, in *Advances in Chemical Physics* (Wiley & Sons, New York, 2007), pp. 315–412.
[9] C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).
[10] C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).
[11] W. von Niessen, J. Chem. Phys. **56**, 4290 (1972).
[12] B. Jansik, S. Host, K. Kristensen, and P. Jorgensen, J. Chem. Phys. **134**, 194104 (2011).
[13] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[14] D. R. Bowler, T. Miyazaki, and M. J. Gillan, J. Phys.: Condens. Mater. **14**, 2781 (2002).
[15] B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, in *World Scientific Lecture and Course Notes in Chemistry* (World Scientific, Singapore, 1994), Vol. 1.

[16] D. M. Ceperley and L. Mitas, in *New Methods in Computational Quantum Mechanics*, edited by I. Prigogine and S. A. Rice (Wiley & Sons, New York, 1996), Vol. 93, pp. 1–38.
[17] A. Aspuru-Guzik and W. A. Lester, , in *Handbook of Numerical Analysis*, edited by C. Le Bris (Elsevier, Amsterdam, 2003), Vol. 10, pp. 485–527.
[18] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[19] J. B. Anderson, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (Wiley & Sons, New York, 1999), Vol. 13, pp. 133–182.
[20] W. A. Lester, Jr., L. Mitas, and B. Hammond, Chem. Phys. Lett. **478**, 1 (2009).
[21] A. Aspuru-Guzik, R. Salomon-Ferrer, B. Austin, and W. A. Lester, Jr., J. Comput. Chem. **26**, 708 (2005).
[22] B. Austin, A. Aspuru-Guzik, R. Salmon-Ferrer, and W. A. Lester, Jr., in *Recent Advances in Quantum Monte Carlo*, edited by S. Rothstein and J. Anderson (Oxford University Press, Oxford, UK, 2007).
[23] J. Kussmann, H. Riede, and C. Ochsenfeld, Phys. Rev. B **75**, 165107 (2007).
[24] A. J. Williamson, R. Q. Hood, and J. C. Grossman, Phys. Rev. Lett. **87**, 246406 (2001).
[25] F. A. Reboredo and A. J. Williamson, Phys. Rev. B **71**, 121105 (2005).
[26] S. Manten and A. Luchow, J. Chem. Phys. **119**, 1307 (2003).
[27] D. Alfe and M. J. Gillan, J. Phys.: Condens. Mater. **16**, L305 (2004).
[28] J. Vrbik and S. M. Rothstein, J. Chem. Phys. **96**, 2071 (1992).
[29] P. Belohorec, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **98**, 6401 (1993).
[30] P. Langfelder, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **107**, 8525 (1997).
[31] C. Filippi and C. J. Umrigar, Phys. Rev. B **61**, R16291 (2000).
[32] R. Assaraf and M. Caffarel, J. Chem. Phys. **113**, 4028 (2000).
[33] R. Assaraf and M. Caffarel, J. Chem. Phys. **119**, 10536 (2003).
[34] M. Casalegno, M. Mella, and A. M. Rappe, J. Chem. Phys. **118**, 7193 (2003).
[35] A. Badinski, J. R. Trail, and R. J. Needs, J. Chem. Phys **129**, 224101 (2008).
[36] S. Chiesa, D. M. Ceperley, and S. Zhang, Phys. Rev. Lett. **94**, 036404 (2005).
[37] A. Badinski and R. J. Needs, Phys. Rev. B **78**, 035134 (2008).
[38] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, Phys. Rev. Lett. **98**, 110201 (2007).
[39] R. Assaraf and M. Caffarel, Phys. Rev. Lett. **83**, 4682 (1999).
[40] N. D. Drummond, M. D. Towler, and R. J. Needs, Phys. Rev. B **70**, 235119 (2004).
[41] K. Schmidt and J. Moskowitz, Stat. Phys. **43**, 1027 (1986).
[42] Z. Sun, S.-Y. Huang, R. N. Barnett, and W. A. Lester, Jr., J. Chem. Phys. **93**, 3326 (1990).
[43] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).
[44] This is because the power series $\sum_{n=1}^{\infty} \frac{(-1)^n}{E(w)^n} x^n$ has a finite convergence radius $[E(w)]$, and by the virtue of the law of large numbers $x = w - E(w)$ converges almost surely to zero as function of $M$.