

Supporting Information for:
“Boosting the accuracy and speed of quantum Monte Carlo: size-consistency and time-step”

Andrea Zen^{1,2,3}, Sandro Sorella⁴, Michael J. Gillan^{1,2,3}, Angelos Michaelides^{1,2,3}, and Dario Alfè^{1,2,3,5*}
¹ *London Centre for Nanotechnology, Gordon St., London WC1H 0AH, UK* ² *Thomas Young Centre, University College London, London WC1H 0AH, UK* ³ *Dept. of Physics and Astronomy, University College London, London WC1E 6BT, UK* ⁴ *International School for Advanced Studies (SISSA), Via Beirut 2-4, 34014 Trieste, Italy and INFN Democritos National Simulation Center, Trieste, Italy* ⁵ *Dept. of Earth Sciences, University College London, London WC1E 6BT, UK*
(Dated:)

PACS numbers:

In the first section of this Supporting Information we provide a short review of the DMC method, followed by a description of the DMC algorithm, the problem of the divergences in proximity of the nodal surface, the instabilities in DMC simulations and the size-consistency issue met when DMC is stabilized by slightly modifying the algorithm. All this is used to contextualize the methodological improvements of this work. The last three sections provide further details on the three examples shown in the paper, including the atomic coordinates. Finally we report results on the relative computational cost of the old and new methods.

I. REVIEW OF DMC

DMC energy evaluations are mostly concerned with the *mixed estimator*, defined as:

$$E_{mix} = \frac{\langle \phi | \hat{H} | \psi_G \rangle}{\langle \phi | \psi_G \rangle} \quad (1)$$

where ψ_G is the *guiding function* (a parametrized wave function optimized within VMC schemes in order to be as close as possible to the ground state) and ϕ is the *exact ground state* wave function of the Hamiltonian \hat{H} . As long as ψ_G has a non-zero overlap with ϕ , E_{mix} is equivalent to the *pure estimator* $\frac{\langle \phi | \hat{H} | \phi \rangle}{\langle \phi | \phi \rangle}$.

The exact wave function $\phi(\mathbf{R})$ can be obtained from the solution $\Phi(\mathbf{R}, t)$ of the imaginary time Schrödinger equation

$$-\frac{\partial \Phi(\mathbf{R}, t)}{\partial t} = -\frac{1}{2} \nabla^2 \Phi(\mathbf{R}, t) - (E_T - V_P(\mathbf{R})) \Phi(\mathbf{R}, t) \quad (2)$$

where t is the time, $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ specifies the coordinates of the N electrons, V_P is the potential energy and E_T is an energy offset. Given the boundary condition $\Phi(\mathbf{R}, 0) = \psi_G(\mathbf{R})$, for time $t \rightarrow \infty$ the imaginary time solution converges to the ground state:

$$\lim_{t \rightarrow \infty} \Phi(\mathbf{R}, t) = \phi(\mathbf{R}).$$

It is often convenient to write the time evolution of Φ in terms of the Green function $G(\mathbf{R} \leftarrow \mathbf{R}'; t)$:

$$\Phi(\mathbf{R}, t_0 + t) = \int G(\mathbf{R} \leftarrow \mathbf{R}'; t) \Phi(\mathbf{R}', t_0) d\mathbf{R}'. \quad (3)$$

The Green function $G(\mathbf{R} \leftarrow \mathbf{R}'; t)$, which satisfies an equation analogous to that of Φ , prescribes how to propagate further in time the distribution Φ . Formally, we can write:

$$G(\mathbf{R} \leftarrow \mathbf{R}'; t) = \langle \mathbf{R} | e^{-t(\hat{H} - E_T)} | \mathbf{R}' \rangle. \quad (4)$$

*Electronic address: d.alfè@ucl.ac.uk

Unfortunately, $G(\mathbf{R} \leftarrow \mathbf{R}'; t)$ is not exactly known for realistic systems. However, by considering that the time interval t can be divided in n smaller intervals of time $\tau = t/n$, and iteratively using Eq. 3 to write $\Phi(\mathbf{R}, t_i)$ in terms of $\Phi(\mathbf{R}, t_{i-1})$, with $i = 1, \dots, n$ and $t_i = t_0 + i\tau$, we obtain the following expression for the Green function:

$$G(\mathbf{R} \leftarrow \mathbf{R}'; t) = \int G(\mathbf{R} \leftarrow \mathbf{R}_1; \tau) \dots G(\mathbf{R}_{n-1} \leftarrow \mathbf{R}', \tau) d\mathbf{R}_1 \dots d\mathbf{R}_{n-1}. \quad (5)$$

For a small enough time step τ , the Green function can be approximated using the Trotter-Suzuki formula, which results in:

$$G(\mathbf{R} \leftarrow \mathbf{R}'; \tau) \approx G_b(\mathbf{R} \leftarrow \mathbf{R}'; \tau) G_d(\mathbf{R} \leftarrow \mathbf{R}'; \tau) \quad (6)$$

where

$$G_d(\mathbf{R} \leftarrow \mathbf{R}'; \tau) = (2\pi\tau)^{-\frac{3}{2}N} \exp \left[-\frac{(\mathbf{R} - \mathbf{R}')^2}{2\tau} \right]$$

is a diffusion term, and

$$G_b(\mathbf{R} \leftarrow \mathbf{R}'; \tau) = \exp \left[\tau \frac{2E_T - V_P(\mathbf{R}) - V_P(\mathbf{R}')}{2} \right]$$

is a branching term. The DMC algorithm is a stochastic realization of Eq. 3, in which a series of *walkers* initially distributed as some $\Phi(\mathbf{R}, 0)$ is propagated ahead in time with the short time approximation to the Green function in Eq. 6. In the long time limit the walkers become distributed as $\phi(\mathbf{R})$.

The method works perfectly well for bosons, as the ground state of the Hamiltonian is node-less. However, the fermionic ground state is generally difficult to calculate, because it is an excited state of the Hamiltonian. The difficulty comes from the fact that in the time evolution of Eq. 2 the weight of the ground state becomes exponentially dominant compared to excited states, and so the fermionic signal is quickly lost into noise. The common solution is to embrace the *fixed node approximation*: $\Phi(\mathbf{R}, t)$ is constrained to have the same nodal surface of some guiding function $\psi_G(\mathbf{R})$. The constraint makes DMC only approximate, and the variational principle then implies that the fixed-node DMC energy is an upper bound of the true fermionic ground state energy. If the nodal surface of the guiding function is exact then also the fixed-node DMC energy is exact.

The fixed-node constraint is conveniently implemented by introducing the mixed distribution $f(\mathbf{R}, t) = \psi_G(\mathbf{R})\Phi(\mathbf{R}, t)$, which satisfies the equation:

$$-\frac{\partial f(\mathbf{R}, t)}{\partial t} = -\frac{1}{2}\nabla^2 f(\mathbf{R}, t) + \nabla \cdot [\mathbf{V}(\mathbf{R})f(\mathbf{R}, t)] - S(\mathbf{R})f(\mathbf{R}, t) \quad (7)$$

(Eq. 1 in the letter), where $\mathbf{V}(\mathbf{R}) \equiv \nabla \log |\psi_G(\mathbf{R})|$ is the *drift velocity*, or local gradient, and $S(\mathbf{R}) \equiv E_T - E_L(\mathbf{R})$ is the *branching* term, with $E_L(\mathbf{R}) = \psi_G(\mathbf{R})^{-1} \hat{H} \psi_G(\mathbf{R})$ the *local energy*. Note that in Eq. 7 there is an additional drift term that was not present in the original imaginary time Schrödinger equation for Φ . The mixed distribution f has the border condition $f(\mathbf{R}, 0) = \psi_G(\mathbf{R})^2$ and, in the limit of large time t :

$$\lim_{t \rightarrow \infty} f(\mathbf{R}, t) = \psi_G(\mathbf{R})\phi(\mathbf{R}).$$

Thus, the mixed estimator can be written as:

$$E_{mix} = \lim_{t \rightarrow \infty} \frac{\int E_L(\mathbf{R})f(\mathbf{R}, t)d\mathbf{R}}{\int f(\mathbf{R}, t)d\mathbf{R}}. \quad (8)$$

It is convenient to write the time evolution of f in terms of the Green function $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; t)$, which prescribes how to propagate further in time the distribution f :

$$f(\mathbf{R}, t_0 + t) = \int \tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; t) f(\mathbf{R}', t_0) d\mathbf{R}', \quad (9)$$

where $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; t)$ satisfies an equation analogous to that of f , and formally can be written as:

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; t) = \frac{\psi_G(\mathbf{R})}{\psi_G(\mathbf{R}')} \langle \mathbf{R} | e^{-t(\hat{H} - E_T)} | \mathbf{R}' \rangle. \quad (10)$$

Again, $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; t)$ is not exactly known for realistic systems, but we can use the same trick of splitting t in n time steps of length $\tau = t/n$. We obtain the following expression for the Green function:

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; t) = \int \tilde{G}(\mathbf{R} \leftarrow \mathbf{R}_{n-1}; \tau) \dots \tilde{G}(\mathbf{R}_1 \leftarrow \mathbf{R}', \tau) d\mathbf{R}_1 \dots d\mathbf{R}_{n-1}. \quad (11)$$

For a small enough time step τ , $\tilde{G}(\mathbf{R}_i, \mathbf{R}_{i+1}; \tau)$ is approximated by the Green functions for purely drift, diffusion and branching processes. This leads to:

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}'; \tau) \approx \tilde{G}_b(\mathbf{R} \leftarrow \mathbf{R}'; \tau) \tilde{G}_d(\mathbf{R} \leftarrow \mathbf{R}'; \tau) \quad (12)$$

where

$$\tilde{G}_d(\mathbf{R} \leftarrow \mathbf{R}'; \tau) = (2\pi\tau)^{-\frac{3}{2}N} \exp \left[-\frac{(\mathbf{R} - \mathbf{R}' - \tau\mathbf{V}(\mathbf{R}'))^2}{2\tau} \right]$$

is the drift-diffusion term, and

$$\tilde{G}_b(\mathbf{R} \leftarrow \mathbf{R}'; \tau) = \exp \left[\tau \frac{S(\mathbf{R}) + S(\mathbf{R}')}{2} \right]$$

is the branching term.

Eq. 7 also introduces *importance sampling*. Beside concentrating the sampling in the important part of the phase space, an additional advantage of importance sampling over simple sampling is that the branching term depends on the local energy $E_L(\mathbf{R})$, and not on the potential energy $V_P(\mathbf{R})$. Since $E_L(\mathbf{R})$ is much smoother than $V_P(\mathbf{R})$, and it is constant in the limit of $\psi_G \sim \phi$, the stability of the DMC simulation is greatly enhanced. The error on this approximate expression for $\tilde{G}(\mathbf{R}_i, \mathbf{R}_{i+1}; \tau)$ can be evaluated using the Zassenhaus formula [1], and the leading correction is of order $\mathcal{O}(\tau^2)$. This translates into an error of order $\mathcal{O}(\tau)$ on $\tilde{G}(\mathbf{R}, \mathbf{R}'; t)$ (see Eq. 11). In the limit of $\tau \rightarrow 0$ the error on the Green function is zero, but the computational cost is $\propto 1/\tau$ because $\tilde{G}_b(\mathbf{R} \leftarrow \mathbf{R}'; t)$ is split in $n = t/\tau$ terms.

A. DMC algorithm

We discuss here how the DMC algorithm actually works. At each time t the distribution $f(\mathbf{R}, t)$ can be represented by a discrete set $\{\mathbf{R}^\alpha(t), w^\alpha(t)\}_{\alpha=1, \dots, n_w(t)}$ of walkers (*i.e.* sampling points \mathbf{R}^α with a weight w^α), such that $f(\mathbf{R}, t) \sim \sum_{\alpha} w^\alpha \delta(\mathbf{R} - \mathbf{R}^\alpha) / \sum_{\alpha} w^\alpha$. By using the Metropolis algorithm we can easily generate an ensemble of configurations $\{\mathbf{R}^\alpha\}_{\alpha=1, \dots, n_w}$ (*i.e.*, a set of walkers with unit weight) that correspond to the initial distribution $f(\mathbf{R}, 0) = \psi_G(\mathbf{R}_n)^2$. In DMC we need to project forward in time the walkers in order to calculate the mixed distribution for $f(\mathbf{R}, t \rightarrow \infty)$.

If in Eq. 8 we express the mixed distribution $f(\mathbf{R}, t)$ as in Eq. 9 (with initial distribution $f(\mathbf{R}, 0) = \psi_G(\mathbf{R})^2$), and we expand the Green function as in Eq. 11 (with $t = n\tau$), we obtain that the mixed estimator is rewritten in the following way:

$$E_{mix} = \lim_{n \rightarrow \infty} \frac{\int E_L(\mathbf{R}_n) \tilde{G}(\mathbf{R}_n \leftarrow \mathbf{R}_{n-1}; \tau) \dots \tilde{G}(\mathbf{R}_1 \leftarrow \mathbf{R}_0, \tau) \psi_G(\mathbf{R}_n)^2 d\mathbf{R}_0 \dots d\mathbf{R}_n}{\int \tilde{G}(\mathbf{R}_n \leftarrow \mathbf{R}_{n-1}; \tau) \dots \tilde{G}(\mathbf{R}_1 \leftarrow \mathbf{R}_0, \tau) \psi_G(\mathbf{R}_n)^2 d\mathbf{R}_0 \dots d\mathbf{R}_n}, \quad (13)$$

and using the approximation in Eq. 12 for the Green function with small τ we have:

$$E_{mix} \simeq \lim_{n \rightarrow \infty} \frac{(2\pi\tau)^{-\frac{3}{2}nN} \int E_L(\mathbf{R}_n) \prod_{i=0}^{n-1} \left\{ \exp \left[-\frac{(\mathbf{R}_{i+1} - \mathbf{R}_i - \tau\mathbf{V}(\mathbf{R}_i))^2}{2\tau} \right] \exp \left[\tau \frac{S(\mathbf{R}_{i+1}) + S(\mathbf{R}_i)}{2} \right] \right\} \psi_G(\mathbf{R}_n)^2 d\mathbf{R}_1 \dots d\mathbf{R}_n}{(2\pi\tau)^{-\frac{3}{2}nN} \int \prod_{i=0}^{n-1} \left\{ \exp \left[-\frac{(\mathbf{R}_{i+1} - \mathbf{R}_i - \tau\mathbf{V}(\mathbf{R}_i))^2}{2\tau} \right] \exp \left[\tau \frac{S(\mathbf{R}_{i+1}) + S(\mathbf{R}_i)}{2} \right] \right\} \psi_G(\mathbf{R}_n)^2 d\mathbf{R}_1 \dots d\mathbf{R}_n}. \quad (14)$$

Thus, according to the RHS of Eq. 14, each walker evolves in time according to a branching-drift-diffusion process: given the configuration \mathbf{R}_i^α and weight w_i^α at time $t = i * \tau$, the walker drift-diffuses as follows:

$$\mathbf{R}_i^\alpha \rightarrow \mathbf{R}_{i+1}^\alpha = \mathbf{R}_i^\alpha + \tau\mathbf{V}(\mathbf{R}_i^\alpha) + \sqrt{\tau}\eta, \quad (15)$$

where η is a $3N$ -dimensional random vector generated from a normal distribution with zero mean and unit variance, and the walker weight evolves as:

$$w_i^\alpha \rightarrow w_{i+1}^\alpha = w_i^\alpha * \exp \left[\tau \frac{S(\mathbf{R}_{i+1}^\alpha) + S(\mathbf{R}_i^\alpha)}{2} \right]. \quad (16)$$

The evolution of the weight is efficiently realized by using a branching (birth/death) algorithm, where walkers with small weight are killed and walkers with high weight are replicated [2]. Moreover, a Metropolis acceptance/rejection move is usually introduced after the drift-diffusion step [3, 4], in order to satisfy the detailed balance and reduce the time-step error, and with that an efficient time-step τ_{eff} , which rescales the nominal time-step τ taking into account the acceptance probability, is used in Eq. 16 in place of τ .

Finally, given the chosen time-step τ and a sufficiently large number n of DMC steps, the mixed energy is calculated as:

$$E_{mix}^{\tau} = \frac{\langle E_L(\mathbf{R}_n^{\alpha})w_n^{\alpha} \rangle_{\alpha}}{\langle w_n^{\alpha} \rangle_{\alpha}}, \quad (17)$$

where $\langle \cdot \rangle_{\alpha}$ is the average over all the walkers. Clearly, this evaluation is affected by a stochastic error inversely proportional to the square root of the number n_w of walkers. In order to increase the precision of the evaluations it is not necessary to use a huge number of walkers; it is much more efficient, because of the equilibration time, to propagate further in time the walkers and to use the following expression to evaluate the mixed energy:

$$E_{mix}^{\tau} = \frac{1}{M} \sum_{m=1}^M \frac{\langle E_L(\mathbf{R}_{n+m}^{\alpha})w_{n+m}^{\alpha} \rangle_{\alpha}}{\langle w_{n+m}^{\alpha} \rangle_{\alpha}}. \quad (18)$$

Notice that in Eq. 18 the walkers provide almost independent evaluations, but the local energies are instead serially correlated, with a correlation time proportional to τ . Thus, in evaluating the stochastic error for the mixed energy it is important to get rid of the serial correlation, for instance by using the “blocking method” [5]. Sometimes the estimator actually used can be slightly different from Eq. 18 – for instance some corrections are sometimes introduced in order to correct for the finite population bias (*i.e.*, having a finite number of walkers can introduce a bias) – but the size-consistency issue here addressed is unaffected by these corrections.

B. Divergences in proximity of the nodal surface

Close to the nodal surface Σ_G of the guiding function ψ_G the approximation in Eq. 12 is problematic, because a configuration \mathbf{R} at a distance δ from Σ_G has both the local gradient $\mathbf{V}(\mathbf{R})$ and the local energy $E_L(\mathbf{R})$ (and consequently the branching term $S(\mathbf{R})$) diverging in modulus as $1/\delta$, leading to instabilities and big finite time step errors. This problem has been tackled both by DePasquale et al. [6] and Umrigar et al. [4], who proposed modifications for $\mathbf{V}(\mathbf{R})$ and for $S(\mathbf{R})$ for \mathbf{R} close to Σ_G to eliminate these divergences. These modifications are strictly related to the size-inconsistency issue addressed in this work.

C. DMC instabilities

DMC instabilities are uncontrolled walker population fluctuations (*i.e.*, weights w_i^{α} experiencing huge changes in a single step $i \rightarrow i + 1$, see Eq. 16), which jeopardize the DMC energy evaluations and makes the simulation unfeasible. They are mainly due to walkers reaching regions of diverging local energy (because of the pseudo-potential or proximity to the nodal surface), and in particular for $E_L(\mathbf{R}) \rightarrow -\infty$ the branching term $S(\mathbf{R})$ leads to proliferation of walkers from just one problematic configuration. Instabilities are strictly related with time step τ : with small τ instabilities are usually under control, but as larger and larger values of τ are considered instabilities are more often observed. The reason is that the diffusion step is random and proportional to $\sqrt{\tau}$, see Eq. 15, and if the time-step is too large there is some chance to fall into the problematic regions, because the drift step is unable to keep electrons away for the divergences. A small enough τ allows the drift step to recover from a “bad” diffusion step. As a matter of fact, DMC simulations with no modifications to the drift and branching terms are stable only for tiny values of τ , making schemes as those proposed by DePasquale et al. [6] or Umrigar et al. [4] necessary in actual calculations, the latter being much more stable than the former. The new limiting scheme proposed in this work (which is the same of Umrigar et al. [4] for the drift, Eq. 3 of the letter, and the one in Eq. 5 of the letter for the branching) appears as effective as the limiting scheme of Umrigar et al. [4] (see Eqs. 3 and 4 of the letter), if not better, in keeping the DMC simulation stable.

A pragmatic way to recover from a diverging population count (population explosion) is to back-track the simulation to a region far from the instability, run the random number generator idle for a number of cycles, and resume the DMC simulation. Often this procedure sends the simulation to a different region of phase space, avoiding the instability. However, if the instabilities are too frequent, the simulation becomes impractical or even impossible. To highlight

the improvement in the stability of the calculations using the new limiting procedure, consider for example the CH₄ - H₂O dimer in the bound configuration. Using the UNR limiting procedure and $\tau = 0.05$ a.u. we encountered 32 population explosions in $\sim 26,000$ steps (population size: 20,480 walkers). No simulations were possible with any larger value of time step. By contrast, using the new limiting procedure we observed no instabilities in $\sim 176,000$ steps at $\tau = 0.05$ a.u., and also no instabilities in $\sim 250,000$ steps at $\tau = 0.1$ a.u..

D. Size-consistency in DMC

As discussed in the letter, a method is size-consistent if the energy E_{AB} of any system AB constituted by the two non-interacting subsystems A and B , is equal to the sum $E_A + E_B$ of the energies of individual subsystems. As in the letter, we assume here to deal with systems that are size-consistent when described with a single Slater determinant (so, also with a Jastrow correlated single Slater determinant). In this section we show that the fixed-node DMC with importance sampling (*i.e.*, Eq. 17) is size-consistent for any τ , but if the modifications to the branching proposed by Umrigar et al. [4] are used DMC is size-consistent only in the limit of $\tau \rightarrow 0$.

Clearly, any configuration $\mathbf{R}^{[AB]}$ of the systems AB is given by the configurations $\mathbf{R}^{[A]}$ and $\mathbf{R}^{[B]}$ of the subsystems A and B , because any electron in AB belongs either to the subsystem A or to B . Mathematically, this means that the vectorial space where the configurations $\mathbf{R}^{[AB]}$ live is the direct sum of the two vectorial spaces where $\mathbf{R}^{[A]}$ and $\mathbf{R}^{[B]}$ live, and we can write (with a little abuse of notation):

$$\mathbf{R}^{[AB]} = \mathbf{R}^{[A]} \oplus \mathbf{R}^{[B]}. \quad (19)$$

As discussed in the letter, the guiding wave function factorizes, *i.e.*:

$$\psi_G^{[AB]}(\mathbf{R}^{[AB]}) = \psi_G^{[A]}(\mathbf{R}^{[A]}) \otimes \psi_G^{[B]}(\mathbf{R}^{[B]}) \quad (20)$$

whenever A and B are far away. From the properties of the hamiltonian operator it follows that the local energy is additive:

$$E_L^{[AB]}(\mathbf{R}^{[AB]}) = E_L^{[A]}(\mathbf{R}^{[A]}) + E_L^{[B]}(\mathbf{R}^{[B]}), \quad (21)$$

which proves that VMC is size-consistent. Moreover, considering that the drift velocity is the local gradient, it is easy to show that:

$$\mathbf{V}^{[AB]}(\mathbf{R}^{[AB]}) = \mathbf{V}^{[A]}(\mathbf{R}^{[A]}) \oplus \mathbf{V}^{[B]}(\mathbf{R}^{[B]}), \quad (22)$$

where the symbol \oplus is used in the same way as in Eq. 19.

In order to address the properties of the DMC mixed energy E_{mix}^τ evaluated for a finite value τ of the time-step, we can consider Eq. 17. According to Eq. 16, the weight is $w_n^\alpha = \exp[\tau \sum_i^n S(\mathbf{R}_i^\alpha)]$ (here, for simplicity, we have slightly simplified the expression, neglecting that the first and last step have a weight that is 1/2) and including that the branching term $S(\mathbf{R}_i^\alpha) = E_T - E_L(\mathbf{R}_i^\alpha)$, it is straightforward to see that:

$$E_{mix}^\tau = \frac{\left\langle E_L(\mathbf{R}_n^\alpha) e^{-\tau \sum_i^n E_L(\mathbf{R}_i^\alpha)} \right\rangle_\alpha}{\left\langle e^{-\tau \sum_i^n E_L(\mathbf{R}_i^\alpha)} \right\rangle_\alpha}. \quad (23)$$

By using Eq. 23, the additivity of the local energy (Eq. 21) and of the drift velocity (Eq. 22), and some algebra, it is easy to prove that:

$$\text{DMC with no modifications:} \quad E_{mix}^{\tau [AB]} = E_{mix}^{\tau [A]} + E_{mix}^{\tau [B]}, \quad (24)$$

for any value of the time-step τ , and of course also for $\tau \rightarrow 0$. The main point of the proof is that the additivity of the local energy imply the factorization of the weight, *i.e.*:

$$e^{-\tau \sum_i^n E_L^{[AB]}(\mathbf{R}_i^{[AB]})} = e^{-\tau \sum_i^n E_L^{[A]}(\mathbf{R}_i^{[A]})} * e^{-\tau \sum_i^n E_L^{[B]}(\mathbf{R}_i^{[B]})} \quad \text{for any walker } \alpha. \quad (25)$$

In principle, it could be explicitly tested that DMC with no modifications satisfy the size-consistency for any finite time-step, but in practice it can be done only for very small values of τ because of the instabilities discussed in Section I C.

The UNR modification to the drift, as reported in Eq. 3 of the letter, does not affect the additivity of the drift (because the correction is performed independently for each electron), and we have that:

$$\text{DMC with UNR modifications:} \quad \bar{\mathbf{V}}^{[AB]}(\mathbf{R}^{[AB]}) = \bar{\mathbf{V}}^{[A]}(\mathbf{R}^{[A]}) \oplus \bar{\mathbf{V}}^{[B]}(\mathbf{R}^{[B]}). \quad (26)$$

which clearly does not affect the size-consistency of the method. The source of the size-inconsistency is instead the UNR modification to the branching term, see Eq. 4 in the letter, because we have that:

$$\text{DMC with UNR modifications:} \quad \bar{S}^{[AB]}(\mathbf{R}^{[AB]}) \neq \bar{S}^{[A]}(\mathbf{R}^{[A]}) + \bar{S}^{[B]}(\mathbf{R}^{[B]}). \quad (27)$$

because of the term \bar{V}/V appearing in the expression of \bar{S} . This imply that the weight of a DMC realization does not factorize any more, that is:

$$\text{DMC with UNR modifications:} \quad e^{\tau \sum_i^n \bar{S}^{[AB]}(\mathbf{R}_i^{\alpha[AB]})} \neq e^{\tau \sum_i^n \bar{S}^{[A]}(\mathbf{R}_i^{\alpha[A]})} * e^{\tau \sum_i^n \bar{S}^{[B]}(\mathbf{R}_i^{\alpha[B]})}. \quad (28)$$

However, in the limit of $\tau \rightarrow 0$ we have that $\bar{V} \rightarrow V$ and $\bar{S} \rightarrow S$, thus UNR approaches asymptotically the case of no modifications, where size-consistency is proven.

The scheme proposed in this work (named here ZSGMA, from authors' names), see Eqs. 5 and 6 in the letter, is exactly size-consistent for $E_{\text{cut}} \rightarrow \infty$ (namely, for $\alpha \rightarrow \infty$ or $\tau \rightarrow 0$), because the branching \bar{S} becomes equivalent to S , which factorizes exactly, so we recover the unmodified DMC algorithm. The method is only approximated for finite E_{cut} ; the modified branching term is not exactly additive, *i.e.* $\bar{S}^{[AB]}(\mathbf{R}^{[AB]}) \neq \sum_i^n \bar{S}^{[A]}(\mathbf{R}^{[A]}) + \sum_i^n \bar{S}^{[B]}(\mathbf{R}^{[B]})$, but what we approximatively satisfy is that:

$$\text{DMC with ZSGMA modifications:} \quad \sum_i^n \bar{S}^{[AB]}(\mathbf{R}^{[AB]}) \sim \sum_i^n \bar{S}^{[A]}(\mathbf{R}^{[A]}) + \sum_i^n \bar{S}^{[B]}(\mathbf{R}^{[B]}), \quad (29)$$

at least when E_{cut} is large enough. This happens because, assuming that E_T is properly set, we have that \bar{S} can be seen as a random variable of zero mean and a variance proportional to \sqrt{N} . In order to satisfy Eq. 29, at least approximatively, we require that the number of times we perform a cut on \bar{S} is independent on the size of the system and with a random sign. This implies a value of $E_{\text{cut}} \propto \sqrt{\text{VAR}(\bar{S})}$.

II. WATER-METHANE DIMER

The atomic coordinates for the bonded and the separated water-methanol complexes follows.

Bonded water-methanol complex

8 atoms

C	8.074716	8.896609	8.452975
H	8.154933	9.264317	9.473808
H	8.072329	9.738234	7.763478
H	8.921481	8.249858	8.232870
H	7.150122	8.334028	8.341744
O	5.000000	7.147077	7.650971
H	4.068066	6.942975	7.563761
H	5.380237	6.896963	6.807984

Separated water-methanol complex

8 atoms

C	15.583540	9.429598	9.707369
H	16.359724	10.130817	10.006659
H	14.662537	9.659431	10.239040
H	15.413947	9.511962	8.635829
H	15.897952	8.416183	9.947948
O	5.000000	6.655767	6.366148
H	5.807082	6.512788	5.870226
H	4.461206	5.887241	6.174226

In Fig. 1 we display the energy of the dimer, $E_{\text{CH}_4-\text{H}_2\text{O}(\text{shifted})}$ as well and the energies of the monomers, E_{CH_4} and $E_{\text{H}_2\text{O}}$, computed in independent calculations performed with simulation cells containing either the $\text{CH}_4\text{-H}_2\text{O}(\text{shifted})$ dimer or the isolated CH_4 and H_2O monomers, respectively.

Single particle wavefunctions were obtained using a plane-wave cutoff of 300 Ry, and re-expanded in terms of B-splines with the natural grid spacing $a = \pi/G_{\text{max}}$, where G_{max} is the magnitude of the largest plane wave in the expansion. The Jastrow factor used in the trial wavefunction of the system included a two-body electron-electron (e-e) term; three different two-body electron-nucleus (e-n) terms for C, O and H, respectively; and three different three-body electron-electron-nucleus (e-e-n) terms, for C, O and H. Of course, for the isolated CH_4 and H_2O systems we only included the e-n and the e-e-n terms for C, H and O, H, respectively, but a part from this difference the Jastrow factors were exactly the same in all systems. The cutoff radii of the e-e, e-n, and e-e-n terms were all lower than 3.5 Å, and the large distance between the two molecules guarantees that the overlap between their respective orbitals is effectively zero. Therefore the trial wavefunction of the dimer $\psi_{\text{CH}_4-\text{H}_2\text{O}(\text{shifted})}$, is effectively the appropriately antisymmetrised product of the trial wavefunctions ψ_{CH_4} and $\psi_{\text{H}_2\text{O}}$ of the CH_4 and the H_2O sub-systems, respectively: $\psi_{\text{CH}_4-\text{H}_2\text{O}(\text{shifted})} = \psi_{\text{CH}_4} \otimes \psi_{\text{H}_2\text{O}}$. The variances of the local energy with the variational Monte Carlo (VMC) distributions were $\sim 0.72, 0.26$ and 0.45 Ha^2 for the $\text{CH}_4\text{-H}_2\text{O}$, CH_4 and H_2O systems, respectively.

As seen in the paper, the finite time-step error in the binding energy, whenever the E_b evaluation is used, is mostly due to the size consistency error. The speedup obtained by using present work prescriptions for the branching factor in comparison with UNR branching factor is of two orders of magnitude, as it is shown in Fig. 4(left). In this system there is the possibility to use E_{b_s} and to alleviate the size-consistency issue of the UNR prescription for the branching factor. However, when big clusters or molecular crystals are considered, E_{b_s} could be an unfeasible choice.

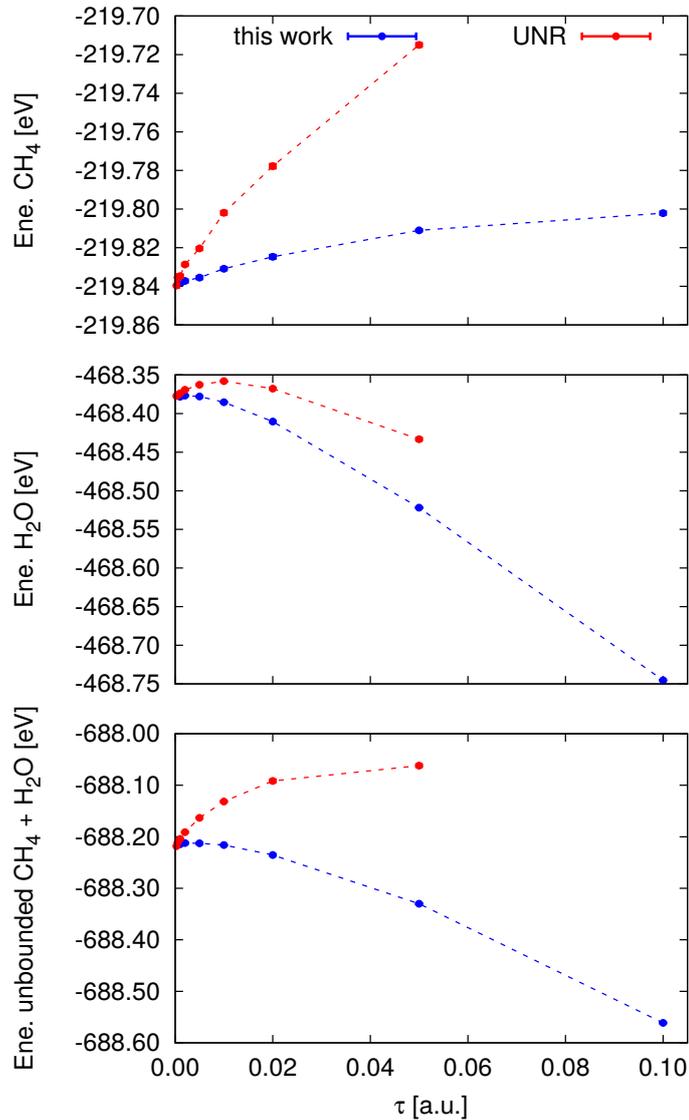


FIG. 1: Energies of the CH_4 (top panel), H_2O (middle), and unbounded $\text{CH}_4\text{-H}_2\text{O}$ (bottom) systems as function of time step τ , calculated using the UNR and present work prescriptions for the limits on the branching factor. Error bars are smaller than the size of the symbols.

III. THE $\text{C}_{60}\text{-C}_{60}\text{H}_{28}$ COMPLEX

The atomic coordinates of the buckyball catcher in the bond and separated configurations follows.

```
Buckyball catcher, bound configuration
148 atoms
C 11.3381 9.61627 5.65661
C 10.5019 10.2959 6.55609
C 10.5019 11.7041 6.55609
C 11.3381 12.3837 5.65661
C 12.1514 11.6972 4.75968
C 12.1514 10.3028 4.75968
C 9.70687 12.5177 7.51144
C 10.38 13.4171 8.33717
C 9.72521 14.3467 9.15701
C 8.27479 14.3467 9.15701
```

C 7.62005 13.4171 8.33717
C 8.29313 12.5177 7.51144
C 10.4758 15.2618 10.0256
C 9.71491 16.2459 10.6215
C 8.28509 16.2459 10.6215
C 7.5242 15.2618 10.0256
C 7.85139 16.9685 11.7592
C 9 17.4111 12.4634
C 10.1486 16.9685 11.7592
C 9 17.6539 13.8283
C 7.68932 17.6257 14.4384
C 6.56242 17.1973 13.7483
C 6.62708 16.7486 12.3758
C 5.74743 15.8711 11.6461
C 6.17772 15.16 10.5294
C 10.3107 17.6257 14.4384
C 11.4376 17.1973 13.7483
C 11.3729 16.7486 12.3758
C 12.2526 15.8711 11.6461
C 11.8223 15.16 10.5294
C 7.49814 11.7041 6.55609
C 7.49814 10.2959 6.55609
C 6.66189 9.61627 5.65661
C 5.8486 10.3028 4.75968
C 5.8486 11.6972 4.75968
C 6.66189 12.3837 5.65661
C 9.70687 9.48234 7.51144
C 10.38 8.58287 8.33717
C 9.72521 7.6533 9.15701
C 8.27479 7.6533 9.15701
C 7.62005 8.58287 8.33717
C 8.29313 9.48234 7.51144
C 7.5242 6.73817 10.0256
C 8.28509 5.75408 10.6215
C 9.71491 5.75408 10.6215
C 10.4758 6.73817 10.0256
C 10.1486 5.03152 11.7592
C 9 4.58894 12.4634
C 7.85139 5.03152 11.7592
C 6.17772 6.84002 10.5294
C 5.74743 6.12895 11.6461
C 6.62708 5.25145 12.3758
C 6.56242 4.80269 13.7483
C 7.68932 4.3743 14.4384
C 9 4.34607 13.8283
C 11.8223 6.84002 10.5294
C 12.2526 6.12895 11.6461
C 11.3729 5.25145 12.3758
C 11.4376 4.80269 13.7483
C 10.3107 4.3743 14.4384
H 4.75135 15.6673 12.0312
H 5.49478 14.432 10.1
H 12.5052 14.432 10.1
H 13.2486 15.6673 12.0312
H 12.3718 17.1048 14.2968
H 10.4112 17.8505 15.4974
H 7.58878 17.8505 15.4974
H 5.62821 17.1048 14.2968

H 5.62821 4.89524 14.2968
H 7.58878 4.14953 15.4974
H 10.4112 4.14953 15.4974
H 12.3718 4.89524 14.2968
H 13.2486 6.33265 12.0312
H 12.5052 7.56798 10.1
H 5.49478 7.56798 10.1
H 4.75135 6.33265 12.0312
H 11.4664 13.4142 8.31724
H 6.53362 13.4142 8.31724
H 11.4664 8.58581 8.31724
H 6.53362 8.58581 8.31724
H 6.66653 13.4703 5.6653
H 5.22065 12.248 4.06543
H 5.22065 9.75197 4.06543
H 6.66653 8.52969 5.6653
H 11.3335 13.4703 5.6653
H 12.7794 12.248 4.06543
H 12.7794 9.75197 4.06543
H 11.3335 8.52969 5.6653
C 5.51767 11.6983 14.9166
C 5.9663 12.4253 13.7406
C 6.69431 13.5983 14.1912
C 6.69371 13.6004 15.6439
C 7.82578 14.0247 13.4951
C 6.39823 11.7269 12.6094
C 5.96742 12.4247 16.0929
C 5.51767 10.3017 14.9166
C 6.39931 11.7265 17.2228
C 7.82388 14.0324 16.3428
C 9 14.4733 14.2211
C 8.27369 13.301 12.3195
C 9 14.4786 15.6164
C 8.2731 13.3054 17.5179
C 7.57488 12.1754 17.9491
C 6.39931 10.2735 17.2228
C 5.96742 9.57525 16.0929
C 5.9663 9.57471 13.7406
C 6.39823 10.2731 12.6094
C 7.57539 12.1739 11.884
C 10.1742 14.0247 13.4951
C 9.72631 13.301 12.3195
C 7.57539 9.82611 11.884
C 8.30111 11 11.4301
C 6.69371 8.3996 15.6439
C 6.69431 8.4017 14.1912
C 7.57488 9.82458 17.9491
C 8.30173 11 18.3978
C 9.7269 13.3054 17.5179
C 10.1761 14.0324 16.3428
C 11.3063 13.6004 15.6439
C 11.3057 13.5983 14.1912
C 10.4246 12.1739 11.884
C 9.69889 11 11.4301
C 7.82578 7.97528 13.4951
C 8.27369 8.69902 12.3195
C 8.2731 8.6946 17.5179
C 7.82388 7.96759 16.3428

C 9.69827 11 18.3978
C 10.4251 12.1754 17.9491
C 12.0337 12.4253 13.7406
C 11.6018 11.7269 12.6094
C 10.4246 9.82611 11.884
C 9.72631 8.69902 12.3195
C 9 7.52671 14.2211
C 9 7.5214 15.6164
C 9.7269 8.6946 17.5179
C 10.4251 9.82458 17.9491
C 11.6007 11.7265 17.2228
C 12.0326 12.4247 16.0929
C 11.6007 10.2735 17.2228
C 12.4823 11.6983 14.9166
C 11.6018 10.2731 12.6094
C 10.1742 7.97528 13.4951
C 10.1761 7.96759 16.3428
C 11.3063 8.3996 15.6439
C 12.0326 9.57525 16.0929
C 12.4823 10.3017 14.9166
C 12.0337 9.57471 13.7406
C 11.3057 8.4017 14.1912

Buckyball catcher, separated by 10 Angstrom
148 atoms

C 11.3381 9.61627 5.65661
C 10.5019 10.2959 6.55609
C 10.5019 11.7041 6.55609
C 11.3381 12.3837 5.65661
C 12.1514 11.6972 4.75968
C 12.1514 10.3028 4.75968
C 9.70687 12.5177 7.51144
C 10.38 13.4171 8.33717
C 9.72521 14.3467 9.15701
C 8.27479 14.3467 9.15701
C 7.62005 13.4171 8.33717
C 8.29313 12.5177 7.51144
C 10.4758 15.2618 10.0256
C 9.71491 16.2459 10.6215
C 8.28509 16.2459 10.6215
C 7.5242 15.2618 10.0256
C 7.85139 16.9685 11.7592
C 9 17.4111 12.4634
C 10.1486 16.9685 11.7592
C 9 17.6539 13.8283
C 7.68932 17.6257 14.4384
C 6.56242 17.1973 13.7483
C 6.62708 16.7486 12.3758
C 5.74743 15.8711 11.6461
C 6.17772 15.16 10.5294
C 10.3107 17.6257 14.4384
C 11.4376 17.1973 13.7483
C 11.3729 16.7486 12.3758
C 12.2526 15.8711 11.6461
C 11.8223 15.16 10.5294
C 7.49814 11.7041 6.55609
C 7.49814 10.2959 6.55609
C 6.66189 9.61627 5.65661
C 5.8486 10.3028 4.75968

C 5.8486 11.6972 4.75968
C 6.66189 12.3837 5.65661
C 9.70687 9.48234 7.51144
C 10.38 8.58287 8.33717
C 9.72521 7.6533 9.15701
C 8.27479 7.6533 9.15701
C 7.62005 8.58287 8.33717
C 8.29313 9.48234 7.51144
C 7.5242 6.73817 10.0256
C 8.28509 5.75408 10.6215
C 9.71491 5.75408 10.6215
C 10.4758 6.73817 10.0256
C 10.1486 5.03152 11.7592
C 9 4.58894 12.4634
C 7.85139 5.03152 11.7592
C 6.17772 6.84002 10.5294
C 5.74743 6.12895 11.6461
C 6.62708 5.25145 12.3758
C 6.56242 4.80269 13.7483
C 7.68932 4.3743 14.4384
C 9 4.34607 13.8283
C 11.8223 6.84002 10.5294
C 12.2526 6.12895 11.6461
C 11.3729 5.25145 12.3758
C 11.4376 4.80269 13.7483
C 10.3107 4.3743 14.4384
H 4.75135 15.6673 12.0312
H 5.49478 14.432 10.1
H 12.5052 14.432 10.1
H 13.2486 15.6673 12.0312
H 12.3718 17.1048 14.2968
H 10.4112 17.8505 15.4974
H 7.58878 17.8505 15.4974
H 5.62821 17.1048 14.2968
H 5.62821 4.89524 14.2968
H 7.58878 4.14953 15.4974
H 10.4112 4.14953 15.4974
H 12.3718 4.89524 14.2968
H 13.2486 6.33265 12.0312
H 12.5052 7.56798 10.1
H 5.49478 7.56798 10.1
H 4.75135 6.33265 12.0312
H 11.4664 13.4142 8.31724
H 6.53362 13.4142 8.31724
H 11.4664 8.58581 8.31724
H 6.53362 8.58581 8.31724
H 6.66653 13.4703 5.6653
H 5.22065 12.248 4.06543
H 5.22065 9.75197 4.06543
H 6.66653 8.52969 5.6653
H 11.3335 13.4703 5.6653
H 12.7794 12.248 4.06543
H 12.7794 9.75197 4.06543
H 11.3335 8.52969 5.6653
C 5.51767 11.6983 24.9166
C 5.9663 12.4253 23.7405
C 6.69431 13.5983 24.1912
C 6.69371 13.6004 25.6439

C 7.82578 14.0247 23.4951
 C 6.39823 11.7269 22.6094
 C 5.96742 12.4247 26.0929
 C 5.51767 10.3017 24.9166
 C 6.39931 11.7265 27.2228
 C 7.82388 14.0324 26.3428
 C 9 14.4733 24.2211
 C 8.27369 13.301 22.3195
 C 9 14.4786 25.6164
 C 8.2731 13.3054 27.5179
 C 7.57488 12.1754 27.9491
 C 6.39931 10.2735 27.2228
 C 5.96742 9.57525 26.0929
 C 5.9663 9.57471 23.7405
 C 6.39823 10.2731 22.6094
 C 7.57539 12.1739 21.884
 C 10.1742 14.0247 23.4951
 C 9.72631 13.301 22.3195
 C 7.57539 9.82611 21.884
 C 8.30111 11 21.4302
 C 6.69371 8.3996 25.6439
 C 6.69431 8.4017 24.1912
 C 7.57488 9.82458 27.9491
 C 8.30173 11 28.3978
 C 9.7269 13.3054 27.5179
 C 10.1761 14.0324 26.3428
 C 11.3063 13.6004 25.6439
 C 11.3057 13.5983 24.1912
 C 10.4246 12.1739 21.884
 C 9.69889 11 21.4302
 C 7.82578 7.97528 23.4951
 C 8.27369 8.69902 22.3195
 C 8.2731 8.6946 27.5179
 C 7.82388 7.96759 26.3428
 C 9.69827 11 28.3978
 C 10.4251 12.1754 27.9491
 C 12.0337 12.4253 23.7405
 C 11.6018 11.7269 22.6094
 C 10.4246 9.82611 21.884
 C 9.72631 8.69902 22.3195
 C 9 7.52671 24.2211
 C 9 7.5214 25.6164
 C 9.7269 8.6946 27.5179
 C 10.4251 9.82458 27.9491
 C 11.6007 11.7265 27.2228
 C 12.0326 12.4247 26.0929
 C 11.6007 10.2735 27.2228
 C 12.4823 11.6983 24.9166
 C 11.6018 10.2731 22.6094
 C 10.1742 7.97528 23.4951
 C 10.1761 7.96759 26.3428
 C 11.3063 8.3996 25.6439
 C 12.0326 9.57525 26.0929
 C 12.4823 10.3017 24.9166
 C 12.0337 9.57471 23.7405
 C 11.3057 8.4017 24.1912

As for the water-methane dimer, single particle wavefunctions were obtained using a plane-wave cutoff of 300 Ry, and re-expanded in terms of B-splines with the natural grid spacing $a = \pi/G_{\max}$. The Jastrow factor (e-e), (e-n) and

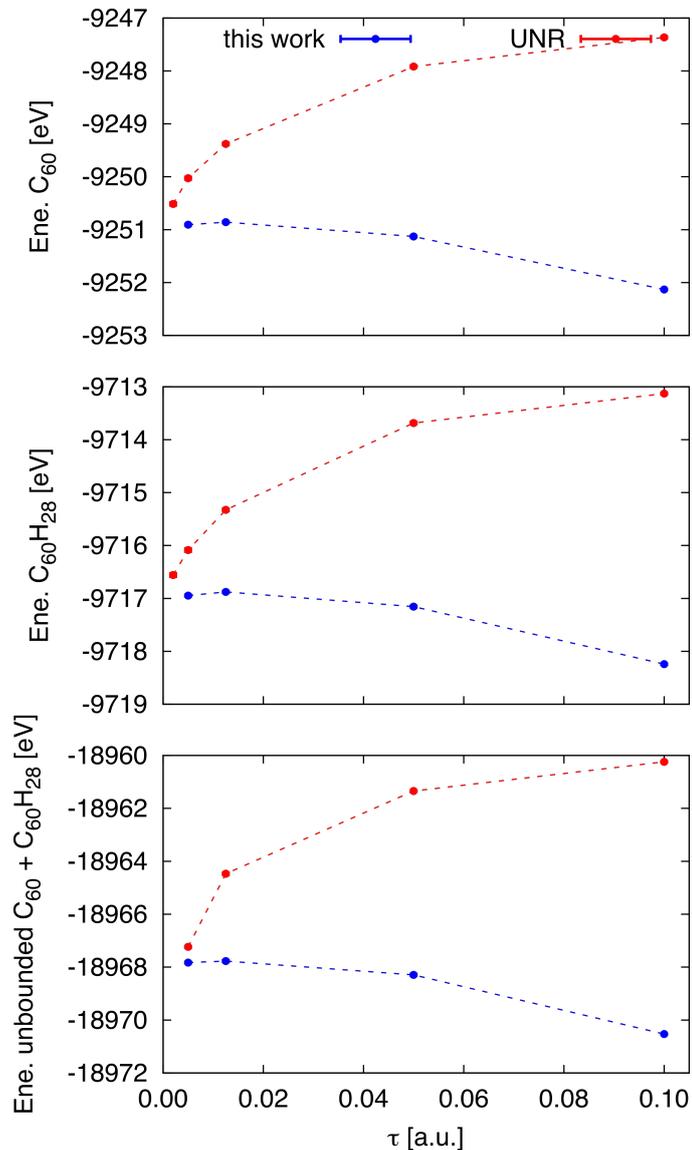


FIG. 2: Energies of C_{60} (top panel), $C_{60}H_{28}$ (middle), and unbonded $C_{60}-C_{60}H_{28}$ (bottom) systems as function of time step τ , calculated using the UNR and present work prescriptions for the limits on the branching factor. Error bars are smaller than the size of the symbols.

(e-e-n) terms, and was constructed with the same procedure as in the water-methane system, i.e. by ensuring that it is the same in all systems. The variances of the VMC local energies were $\sim 11, 5.4$ and 5.8 Ha^2 for the $C_{60}-C_{60}H_{28}$, C_{60} and $C_{60}H_{28}$ systems, respectively.

In Fig. 2 we display the energy of the supramolecular system, $E_{C_{60}-C_{60}H_{28}}$ as well as the energies of the monomers, $E_{C_{60}}$ and $E_{C_{60}H_{28}}$, computed in independent calculations performed with simulation cells containing either the isolated C_{60} and $C_{60}H_{28}$ molecules, respectively.

The improved accuracy of present work prescriptions for the branching factor in comparison with the UNR branching factor can be appreciated in Fig. 4(center).

IV. TWO DIMENSIONAL SQUARE ICE

We considered a monolayer of flat square ice of water, that is a system with 2-dimensional periodicity that is attaining considerable attention [7, 8]. The atomic coordinates of the 2D ice and of the reference water molecule, used for the calculations of the cohesive energy, follows.

```

Two-dimensional ice
primitive unit cell
vector 1:  5.6474896423   0.0000000000   0.0000000000
vector 2:  0.0000000000   5.6474577703   0.0000000000
12 atoms
H   3.1763611983   4.3182010792  10.3550914468
H   4.1530093801   5.2948451664   9.6449029889
H   2.4711284440   1.3292566911  10.3550914468
H   1.4944802622   0.3526156375   9.6449029889
H   0.3526163772   4.1529855762  10.3550914468
H   1.3292645589   3.1763414891   9.6449029889
H   5.2948732651   1.4944721941  10.3550914468
H   4.3182250834   2.4711162812   9.6449029889
O   4.0902046808   4.3810037401   9.9999995643
O   1.5572849615   1.2664540302   9.9999995643
O   1.2664598597   4.0901829153   9.9999995643
O   4.3810297826   1.5572748550   9.9999995643

Reference water molecule
3 atoms
H   7.563300       7.376408       8.620742
H   6.623736       7.956615       7.550648
O   7.445309       7.447677       7.658192

```

The unit cell include four water molecules, and here we considered a 4×4 supercell, for a total of 64 waters in the system. The cohesive energy is obtained by subtracting the energy of the relevant number of isolated water molecules. Single particle wavefunctions were obtained using a plane-wave cutoff of 600 Ry, and re-expanded in terms of B-splines with the natural grid spacing $a = \pi/G_{\max}$. The larger plane-wave cutoff used for these calculations resulted in a lower variance of the VMC local energies, which was $\sim 0.28 \text{ Ha}^2$ for the isolated molecule, and $\sim 19.8 \text{ Ha}^2$ for the square ice (corresponding to $\sim 0.31 \text{ Ha}^2$ per water molecule). At the VMC level of theory the evaluated cohesive energy is $-0.108(4) \text{ eV}$, that is severely underestimated (by a factor 4) with respect to the DMC evaluations.

In Fig. 3 we display the energy of the isolated water molecule, as well as the energy per water in the square lattice 2-dimensional system. A comparison with Fig. 1 shows that the higher quality of the trial wavefunctions for this system results in a lower time step error.

The speedup obtained with present work prescriptions for the branching factor in comparison with the UNR branching factor can be appreciated in Fig. 4(left).

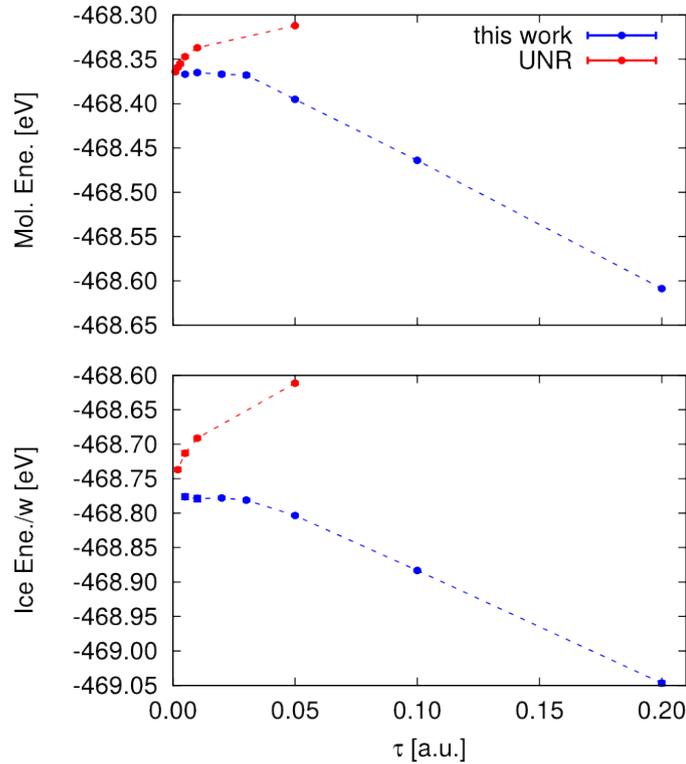


FIG. 3: Energies of an isolated water molecule (top), and of a water molecule in a periodic two dimensional squalice (bottom) systems as function of time step τ , calculated using the UNR and present work prescriptions for the limits on the branching factor. Error bars are smaller than the size of the symbols.

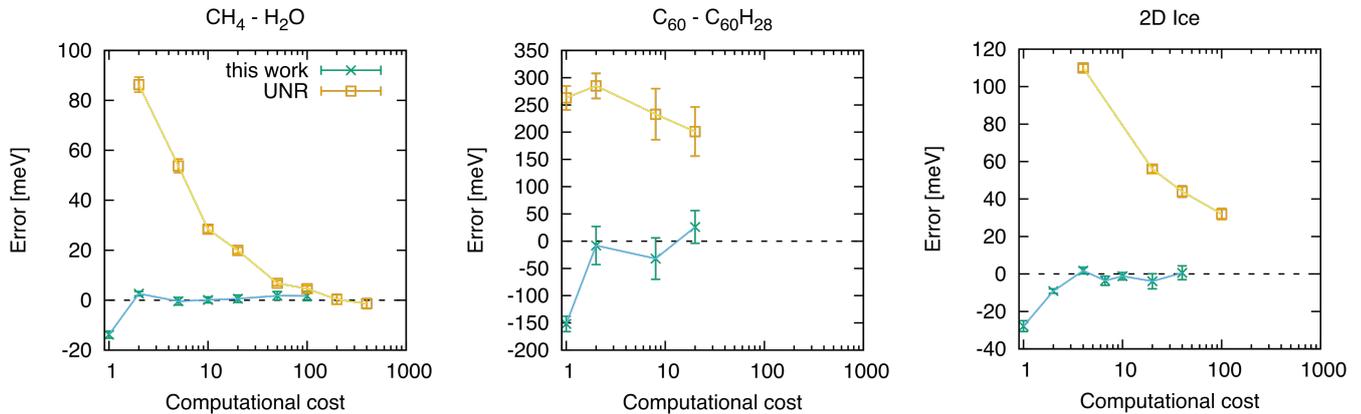


FIG. 4: Size-consistency error as a function of the computational cost in DMC calculations, for the three systems considered in the work, obtained by using UNR and present work prescriptions for the branching factor. Computational cost is in arbitrary units.

-
- [1] M. Suzuki, Commun. Math. Phys. **57**, 193 (1977).
 - [2] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
 - [3] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
 - [4] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).
 - [5] H. Flyvbjerg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).
 - [6] M. F. DePasquale, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **89**, 3629 (1988).
 - [7] J. Chen, G. Schusteritsch, C. J. Pickard, C. G. Salzmann, and A. Michaelides, Phys. Rev. Lett. **116**, 025501 (2016).

- [8] G. Algara-Siller, O. Lehtinen, F. C. Wang, R. R. Nair, U. Kaiser, H. A. Wu, A. K. Geim, and I. V. Grigorieva, *Nature* **519**, 443 (2015).