Sign problem of the fermionic shadow wave function

Francesco Calcavecchia, Francesco Pederiva, Malvin H. Kalos, and Thomas D. Kühne

Phys. Rev. E **90**, 053304 — Published 4 November 2014

DOI: 10.1103/PhysRevE.90.053304

On the Sign Problem of the Fermionic Shadow Wave Function

Francesco Calcavecchia*
Institute of Physics, Johannes Gutenberg-University, Staudingerweg 7, D-55128 Mainz, Germany and Graduate School Materials Science in Mainz, Staudingerweg 9, D-55128 Mainz, Germany

Francesco Pederiva
Dipartimento di Fisica, University of Trento, via Sommarive 14, I-38050 Povo, Trento, Italy and INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, Trento, Italy

Malvin H. Kalos
Lawrence Livermore National Laboratory Livermore, California 94550, USA

Thomas D. Kühne†
Institute of Physical Chemistry and Center for Computational Sciences, Johannes Gutenberg University Mainz, Staudinger Weg 7, D-55128 Mainz, Germany and Department of Chemistry, University of Paderborn, Warburger Str. 100, D-33098 Paderborn, Germany

(Dated: October 15, 2014)

We present a whole series of novel methods to alleviate the sign problem of the Fermionic Shadow Wave Function in the context of Variational Monte Carlo. The effectiveness of our new techniques is demonstrated on liquid ^{3}He. We found that although the variance is reduced, the gain in efficiency is restricted by the increased computational cost. Yet, this development not only extends the scope of the Fermionic Shadow Wave Function, but also facilitates highly accurate Quantum Monte Carlo simulations previously thought not feasible.

INTRODUCTION

The difficulty to solve the Schrödinger equation for many interacting particles stems from the fact that it is generally impossible to analytically determine its ground state for more than a few particles. Quantum Monte Carlo techniques [1–4], such as Variational Monte Carlo (VMC) [5], are stochastic methods that enables to approximately solve the many-body Schrödinger equation. The main concepts underlying VMC are the application of the Rayleigh-Ritz variational principle and the use of importance sampled Monte Carlo (MC) to efficiently evaluate the high-dimensional integrals, which are involved in the many different expectation values such as the energy [6–8]. Its great appeal is based upon the low computational complexity, as opposed to quantum-chemical methods [9]. Since many-body correlation effects are taken into account by a prescribed trial wave function, VMC is substantially more accurate than commonly employed mean-field techniques, such as Hartree-Fock and density functional theory [10], and permits to treat even strongly correlated systems. However, since the exact wave function is unknown from the outset, the trial wave function ought to resemble it as closely as possible. Nevertheless, given that the addition of a simple correlation function of the Jastrow form enables to recover most of the correlation effects [11], VMC typically yields sufficiently accurate results.

Here, we consider the Shadow Wave Function (SWF), first introduced by Kalos and coworkers [12, 13], as our trial wave function. The SWF allows to describe all possible condensed phases (gas, liquid and solid) and even phase coexistence within the same trial wave function [14]. Therefore, it is for instance possible to simulate a solid without *a priori* knowing its crystal structure, which instead emerges from the calculation. Moreover, it even permits to describe inhomogeneous systems [15–17]. In addition, the SWF has further advantageous properties, such as for instance that it introduces many-body correlations and obeys a strong similitude with the exact ground state wave function [18].

Since fermions must obey Fermi-Dirac statistics to comply with the Pauli exclusion principle, an antisymmetric version of the SWF is required that changes its sign upon interchanging any two like-spin particles. While these extensions of the SWF to fermionic systems indeed constitute a substantial improvement, when compared to other more conventional trial wave functions, they are plagued by the occurrence of a sign problem, which limits its applicability to rather small systems [19]. Hence, an efficient and accurate method to simulate fermionic systems remains an open and upmost challenging problem. In this paper, we therefore study the origin and nature of the sign problem and present multiple novel methods to alleviate it.

The remaining of the paper is organized as follows. In section we introduce the SWF and its antisymmetric variant, while in the associated sign problem of the latter is described. Sections and describe two kinds of novel approaches to reduce the sign problem, whereas in

section all the methods presented in the previous sections are assessed in terms of their efficiency. The last section contains the conclusions.

THE SHADOW WAVE FUNCTION

Let us begin by defining the SWF that is constructed by introducing auxiliary degrees of freedom $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots \mathbf{s}_N)$ coupled to the particles coordinates, called shadows, and integrating over all of them [12]. The most general form of the SWF reads as

$$\psi_{\text{SWF}}(\mathbf{R}) = \int d\mathbf{S}\, \Gamma(\mathbf{R}, \mathbf{S}), \tag{1}$$

where $\Gamma(\mathbf{R}, \mathbf{S})$ is an arbitrary wave function, while $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots \mathbf{r}_N)$ represents all N particle coordinates.

However, the extension of the SWF to fermionic systems is non-trivial, due to the antisymmetry requirement of the wave function to obey the Pauli exclusion principle. The simplest *ansatz* to achieve this, is known as the Antisymmetric Shadow Wave Function (ASWF) [20]

$$\psi_{\text{ASWF}}(\mathbf{R}) \equiv \Phi(\mathbf{R}) J_p(\mathbf{R}) \int d\mathbf{S}\, \Xi(\mathbf{R}, \mathbf{S}) J_s(\mathbf{S}), \tag{2}$$

where $\Phi(\mathbf{R})$ is a Slater determinant (SD) that satisfies the antisymmetry condition by changing sign upon the exchange of any two like-spin fermions [21]. Two-body correlations between the particles are taken into account by a Jastrow correlation factor $J_p(\mathbf{R}) = e^{-\frac{1}{2}\sum_{i<j} u_{pp}(|\mathbf{r}_i - \mathbf{r}_j|)}$ [11] and likewise interactions between the shadows are introduced via $J_s(\mathbf{S}) = e^{-\sum_{i<j} u_{ss}(|\mathbf{s}_i - \mathbf{s}_j|)}$. The kernel $\Xi(\mathbf{R}, \mathbf{S}) = e^{-\sum_{i=1}^{N_p} u_{ps}(|\mathbf{r}_i - \mathbf{s}_i|)}$ connects the particles with the shadows, even though it can also be interpreted as a Green's function. Therein, u_{pp}, u_{ss} and u_{ss}, respectively, are denoted as two-body pseudopotentials because of their similarity to the potential in the Boltzmann distribution. Here, we have employed

$$u_{pp}(\mathbf{r}_{ij}) = \left(\frac{b}{|\mathbf{r}_i - \mathbf{r}_j|}\right)^5 \tag{3a}$$

$$u_{ss}(\mathbf{s}_{ij}) = c_1 V(c_2 |\mathbf{s}_i - \mathbf{s}_j|) \tag{3b}$$

$$u_{ps}(|\mathbf{r}_i - \mathbf{s}_i|) = C|\mathbf{r}_i - \mathbf{s}_i|^2, \tag{3c}$$

where V is the potential used in the Hamiltonian, while b, c_1, c_2 and C are variational parameters. While Eq. 3a is chosen to satisfy Kato's cusp condition, Eq. 3b is to account for the observation that in a strongly correlated system, the spread of delocalized quantum particles is spatially limited. Lastly, Eq. 3c is motivated by fact that the shadows can be thought of as the centroid of the delocalized quantum particles that are behaving according to the classical Maxwell-Boltzmann distribution. In order to preserve the translational symmetry of the wave function, which is one of the many appealing properties of the SWF, plane wave orbitals are the natural choice to build up $\Phi(\mathbf{R})$. As we have considered an unpolarized system, we have adopted a product of two Slater determinants to describe spin-up and spin-down atoms, respectively, i.e. $\Phi(\mathbf{R}) = \Phi^{\uparrow}(\mathbf{R}^{\uparrow}) \times \Phi^{\downarrow}(\mathbf{R}^{\downarrow})$.

The advantage of the ASWF with respect to conventional trial wave functions is that many-body correlation effects of any order are included from the outset. In fact, even if the shadows are correlated through a two-body function only, the convolution integral permits even higher-order correlation effects to be taken into account. In particular at the presence of phase transitions, where these subtle many-body correlation effects are essential, the ASWF has proven to be superior to ordinary trial wave functions [20]. However, only symmetric correlation effects are taken in account, whereas three-body and backflow correlations are not considered [20, 22–30]. Furthermore, the nodal surface is imposed *a priori* by a single Slater determinant, and as such only improvable within the flexibility of $\Phi(\mathbf{R})$.

Nevertheless, a more intriguing way to devise an antisymmetric version of the SWF is to introduce a SD as a function of $\mathbf{S}$, i.e. $\Phi(\mathbf{S})$. The resulting Fermionic Shadow Wave Function (FSWF) [18, 19, 31] reads as

$$\psi_{\text{FSWF}}(\mathbf{R}) = J_p(\mathbf{R}) \int d\mathbf{S}\, \Xi(\mathbf{R}, \mathbf{S}) \Phi(\mathbf{S}) J_s(\mathbf{S}). \tag{4}$$

In fact, given an arbitrary like-spin odd-particle permutation operator $\mathcal{P}$, and exploiting that $\Xi(\mathcal{P}\mathbf{R}, \mathbf{S}) = \Xi(\mathbf{R}, \mathcal{P}\mathbf{S})$,

$$\begin{aligned}
\psi_{\text{FSWF}}(\mathcal{P}\mathbf{R}) &= J_p(\mathcal{P}\mathbf{R}) \int d\mathbf{S}\, \Xi(\mathcal{P}\mathbf{R}, \mathbf{S}) \Phi(\mathbf{S}) J_s(\mathbf{S}) \\
&= J_p(\mathbf{R}) \int d\mathbf{S}\, \Xi(\mathbf{R}, \mathcal{P}\mathbf{S}) \Phi(\mathbf{S}) J_s(\mathbf{S}) \\
&= J_p(\mathbf{R}) \int d(\mathcal{P}\mathbf{S})\, \Xi(\mathbf{R}, \mathcal{P}\mathbf{S})\, (-\Phi(\mathcal{P}\mathbf{S})) \\
&\quad \times J_s(\mathcal{P}\mathbf{S}) \\
&= -\psi_{\text{FSWF}}(\mathbf{R}).
\end{aligned} \tag{5}$$

The FSWF has several advantages over the ASWF: (i) It closer resembles the projection onto the exact fermionic ground state, (ii) in the limits of high and low density, the exact asymptotic nodal structure is correctly reproduced and (iii) backflow correlation effects are naturally included [18].

As can be seen in Table I, the FSWF provides a much improved variational ground state energy of liquid ^{3}He, even though with an admitted large statistical uncertainty. The corresponding computational details are given in [34]. But, as we are going to explain in detail in the next section, the FSWF entails a serious sign problem that makes it computationally expensive to obtain reliable results for large systems. This system size limitation not only restricts the applicability, but also the reliability of the FSWF in the thermodynamic limit, due to the

Trial wave function	Energy per particle	N
J-SD[27]	-1.08 ± 0.03 K	38
JB-SD[27]	-1.55 ± 0.04 K	38
JT-SD[27]	-1.61 ± 0.03 K	38
JBT-SD[27]	-1.91 ± 0.04 K	38
DMC[32]	-2.299 ± 0.005 K	54
ASWF	-1.222 ± 0.006 K	66
FSWF	-1.8 ± 0.2 K	66
Exp.[33]	-2.47 K	∞

Table I. The ground state energy per particle of liquid ^{3}He as obtained by VMC using different trial wave functions. The Jastrow-Slater Determinant (J-SD) trial wave function is defined as $\psi_{\text{J-SD}} \equiv \Phi(\mathbf{R})J_p(\mathbf{R})$ and may be augmented by backflow (B) and three-body (T) correlation. The values from Diffusion Monte Carlo (DMC) and experiment are shown for comparison.

presence of significant finite-size effects [35]. Therefore, it would be highly desirable to solve - or at least alleviate - the sign problem, and to facilitate accurate simulations employing the FSWF, though with many more particles than presently feasible.

THE SIGN PROBLEM OF THE FSWF

We will illustrate the sign problem of the FSWF when computing the ground state energy E, which in VMC is estimated by

$$E \simeq \frac{1}{M}\sum_{i=1}^{M} \mathrm{H}^{\mathrm{loc}}(\mathbf{R}_i), \tag{6}$$

where $\mathrm{H}^{\mathrm{loc}}(\mathbf{R}_i) = \frac{\mathrm{H}\psi(\mathbf{R}_i)}{\psi(\mathbf{R}_i)}$ is the local energy and M the number of sampling points. To this end, the positions $\mathbf{R}_i$ of the particles are sampled from the probability density function (pdf) $|\psi^2(\mathbf{R})|$, where ψ is the preassigned normalized trial wave function. The required positiveness of $|\psi^2(\mathbf{R})|$ is satisfied by definition.

In conjunction with the previously introduced SWF, the energy reads as

$$E = \frac{\int d\mathbf{R}\, d\mathbf{S}_1 d\mathbf{S}_2\, \Gamma(\mathbf{R},\mathbf{S}_1)\mathrm{H}\Gamma(\mathbf{R},\mathbf{S}_2)}{\int d\mathbf{R}\, d\mathbf{S}_1 d\mathbf{S}_2\, \Gamma(\mathbf{R},\mathbf{S}_1)\Gamma(\mathbf{R},\mathbf{S}_2)}, \tag{7}$$

where $\mathbf{R}_i$, $\mathbf{S}_{1i}$ and $\mathbf{S}_{2i}$ should in principle be sampled from $\Gamma(\mathbf{R},\mathbf{S}_1)\times\Gamma(\mathbf{R},\mathbf{S}_2)$. However, due to the fact that the FSWF is evaluated using two different shadows, $\mathbf{S}_1$ and $\mathbf{S}_2$, the necessary positiveness requirement of the sampled function is no longer fulfilled. As a consequence, it is not possible to sample $\mathbf{R}_i$, $\mathbf{S}_{1i}$, and $\mathbf{S}_{2i}$ directly from $\Gamma(\mathbf{R},\mathbf{S}_1)\times\Gamma(\mathbf{R},\mathbf{S}_2)$.

In spite of that, it is feasible to sample from $|\Gamma(\mathbf{R},\mathbf{S}_1)\times\Gamma(\mathbf{R},\mathbf{S}_2)|$ by introducing the weights $w(\mathbf{R},\mathbf{S}_1,\mathbf{S}_2) = \mathrm{sign}(\Gamma(\mathbf{R},\mathbf{S}_1)\times\Gamma(\mathbf{R},\mathbf{S}_2))$ and estimating the energy as

$$E \simeq \frac{\sum_{i=1}^{M} \frac{w_i}{2}\left(\mathrm{H}_{1i}^{\mathrm{loc}} + \mathrm{H}_{2i}^{\mathrm{loc}}\right)}{\sum_{i=1}^{M} w_i}, \tag{8}$$

where

$$\mathrm{H}_1^{\mathrm{loc}} \equiv \frac{\mathrm{H}\Gamma(\mathbf{R},\mathbf{S}_1)}{\Gamma(\mathbf{R},\mathbf{S}_1)} \quad \text{and} \quad \mathrm{H}_2^{\mathrm{loc}} \equiv \frac{\mathrm{H}\Gamma(\mathbf{R},\mathbf{S}_2)}{\Gamma(\mathbf{R},\mathbf{S}_2)}.$$

However, due to the sign $w(\mathbf{R}_i,\mathbf{S}_{1i},\mathbf{S}_{2i})$, the sum in Eq. 8 is typically very slowly converging. This issue is particularly severe for disordered systems, such as liquid ^{3}He [36].

In order to evaluate the mean value of E and its unbiased error bar σ we have employed the so-called blocking technique [7, 37]. To that extent, the data set is divided into n_{block} disjoint blocks (typical values for n_{block} are between 4 and 50), each one with its corresponding average value E_j^{block}. Hence, the average energy $\langle E\rangle_{\text{block}}$ and the corresponding variance σ^2_{block} can be computed as

$$\langle E\rangle_{\text{block}} = \frac{1}{n_{\text{block}}}\sum_{j=1}^{n_{\text{block}}} E_j^{\text{block}} \text{and} \tag{9a}$$

$$\sigma^2_{\text{block}} = \frac{1}{n_{\text{block}}-1}\sum_{j=1}^{n_{\text{block}}}\left(E_j^{\text{block}} - \langle E\rangle_{\text{block}}\right)^2. \tag{9b}$$

The straightforward evaluation of the standard deviation σ provides only a biased estimate that may severely underestimate the true error bar due to the presence of serial correlation between successive data points. Nevertheless, given that the length of each block $\frac{M}{n_{\text{block}}}$ is large enough, serial correlation between the block averages E_j^{block} becomes arbitrarily small with the result that σ can after all be correctly estimated. In fact, when plotting σ_{block} as a function of n_{block} and assuming that M is sufficiently large, a plateau that corresponds to the correct estimation of the unbiased error bar is emerging. We remark that, mathematically speaking, $\langle E\rangle_{\text{block}}$ may vary for different values of n_{block}, but as long as M is large enough, each value E_j^{block} is very close to E, so that eventually $\langle E\rangle_{\text{block}}$ is independent from the choice of n_{block}.

The block average-energy $\langle E\rangle_{\text{block}}$ and the corresponding standard deviation σ_{block} as a function of n_{block} from a FSWF simulation of ^{3}He are shown in Fig. 1. However, the expected plateau of σ_{block} and the estimated energy $\langle E\rangle_{\text{block}}$ can only be observed when the lengths of the individual blocks is rather large. As a consequence, the statistical uncertainty of the mean value $\langle E\rangle_{\text{block}} = 1.8(2)$ K is relatively large, which is a clear manifestation of the sign problem of the FSWF. Nevertheless, it has to be said that the present example represents a worst-case scenario

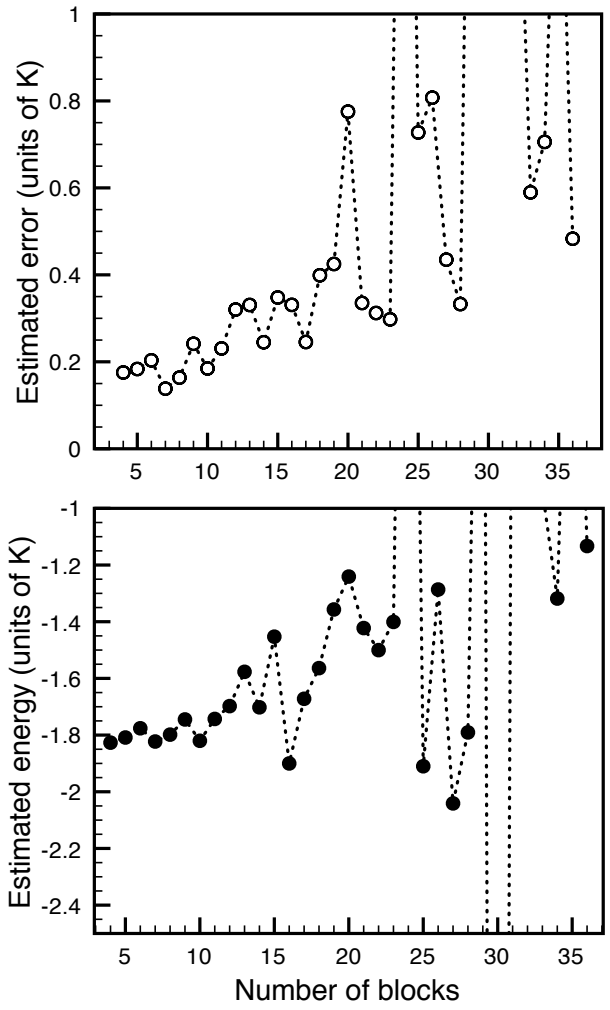

Figure 1. The block average-energy $\langle E\rangle_{\rm block}$ and the corresponding error bar $\sigma_{\rm block}$ from a FSWF simulation of ^{3}He with $N = 66$ atoms and $M = 128 \cdot 10^8$ sampling points as a function of $n_{\rm block}$.

for the FWSF and that the sign problem is in this case particularly severe. In fact, if the parameter C of Eq. 3c is large, $\mathbf{S}_1$ and $\mathbf{S}_2$ are confined around $\mathbf{R}$, which causes that $\Phi(\mathbf{S}_1)$ as well as $\Phi(\mathbf{S}_2)$ have actually the same sign.

In any case, it is important to emphasize that the sign problem of the FSWF differs from the infamous fermion sign problem of projection methods such as Green's Function [38] or Diffusion Monte Carlo [39], which is due to the intrinsic difficulty to sample from a positive pdf generated by a squared antisymmetric function. Nevertheless, contrary to the latter, there is no categorical reason that prohibits to solve the former and to efficiently evaluate a largely fluctuating integral. Nevertheless, the convergence is drastically reduced, so that in many cases, such as the one we have just illustrated, it is practically impossible to obtain reliable results for any reasonable large number of particles. Apart from that, we would like to point out that an antisymmetric component in the integral always entails convergence problems using MC techniques, so that the present sign problem can be viewed as a particular case of a more general class of integrals.

ANTITHETIC VARIATES

In order to accelerate the convergence, one can examine the behavior of the integrand, and sum over those contributions that immediately leads to a better approximation of the average [40]. The following example is intended to clarify this concept.

Suppose that we are interested in numerically evaluating the integral

$$I = \int_0^\infty dx \underbrace{ae^{-ax}}_{f(x)} \underbrace{\sin(\pi x)}_{g(x)} = \frac{\pi a}{\pi^2 + a^2}. \tag{10}$$

A MC procedure is to sample x_k from the pdf $f(x_k)$ and form the average of $g(x_k)$. The variance of this estimator reads as

$$\begin{aligned}\sigma^2 &= \int f(x)g^2(x)dx - I^2 \\ &= \frac{\pi^2(2\pi^4 + a^4)}{(4\pi^2 + a^2)(\pi^2 + a^2)^2},\end{aligned} \tag{11}$$

whereas the quotient

$$\frac{I^2}{\sigma^2} = \frac{a^2(4\pi^2 + a^2)}{(2\pi^4 + a^4)} \tag{12}$$

is a measure of the "signal-to-noise" ratio and approaches

$$2\frac{a^2}{\pi^2} + O\left(\frac{a^4}{\pi^4}\right) \text{ as } a \to 0. \tag{13}$$

That is to say that the procedures becomes very inefficient if a small. This is a simple example of a "sign problem".

Nevertheless, the problem can be completely eliminated by various forms of correlated estimates, e.g. by correlating a negative lobe with the previous positive lobe. The most effective correlation (and easiest to analyze) is to sample x only on $[0, 1]$, but include all $x + n$, where $n = 0 \ldots \infty$, with the factor $(-1)^n e^{-na}$. Since

$$\sum_{n=0}^{\infty} (-1)^n e^{-na} = \frac{e^a}{1 + e^a}, \tag{14}$$

we can recast the integral as

$$\begin{aligned}I &= \int_0^1 dx \underbrace{\left\{\frac{ae^{-ax}}{1 - e^{-a}}\right\}}_{f_A(x)} \underbrace{\left\{\frac{e^a - 1}{e^a + 1}\right\} \sin(\pi x)}_{g_A(x)} \\ &= \frac{\pi a}{\pi^2 + a^2},\end{aligned} \tag{15}$$

where again $f_A(x)$ and $g_A(x)$ are the modified functions to sample and average over, respectively. The associated variance reads as

$$\sigma_A^2 = \int_0^1 \left\{\frac{ae^{-ax}}{1-e^{-a}}\right\}\left[\left\{\frac{e^a-1}{e^a+1}\right\}\sin(\pi x)\right]^2 dx - I^2 = \frac{2\pi^2\tanh^2\left(\frac{a}{2}\right)}{a^2+4\pi^2} - \frac{\pi^2 a^2}{\left(a^2+\pi^2\right)^2} \tag{16}$$

and the signal-to-noise ratio

$$\lim_{a\to 0}\frac{I^2}{\sigma_A^2} = \frac{8}{\pi^2-8} + \frac{2(2\pi^2-21)}{3(\pi^2-8)^2}a^2 + O(a^4) \tag{17}$$

is now finite, at variance to Eq. 13.

In the case of the FSWF, the underlying idea is that the integral

$$\int d\mathbf{S}\,\Xi(\mathbf{R},\mathbf{S})\Phi(\mathbf{S})J_s(\mathbf{S}) \tag{18}$$

has both positive and negative contributions, and that summing pairs of positive and negative values speeds up the convergence. To that extent two promising geometrical transformations that take advantage of the antithetic contributions are proposed in the following: permutations and reflections.

Permutations

In all of the presented methods belonging to this category, pair permutations of like-spin shadows are employed to induce the desired antithetic contributions.

Gaussian Determinant

The first approach is to directly sum over all permuted terms, which eventually translates into a determinant consisting of Gaussians. To illustrate this, we sum over all pair permutations $\mathcal{P}_{ij}^{(2)}$, which leads to

$$\psi_{\text{FSWF}}(\mathbf{R}) = J_p(\mathbf{R})\int d\mathbf{S}\left[\Xi(\mathbf{R},\mathbf{S}) - \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Xi(\mathbf{R},\mathcal{P}_{ij}^{(2)}\mathbf{S})\right]\Phi(\mathbf{S})J_s(\mathbf{S}), \tag{19}$$

as well as over all 3-term permutations $\mathcal{P}_{ijk}^{(3)}$, i.e.

$$\psi_{\text{FSWF}}(\mathbf{R}) = J_p(\mathbf{R})\int d\mathbf{S}\left[\Xi(\mathbf{R},\mathbf{S}) + \sum_{i=1}^{N-2}\sum_{j=i+1}^{N-1}\sum_{k=j+1}^{N}\Xi(\mathbf{R},\mathcal{P}_{ijk}^{(3)}\mathbf{S})\right]\Phi(\mathbf{S})J_s(\mathbf{S}). \tag{20}$$

This is to say that in general an even number of permutations results in a change of sign, while an odd number does not. It is now clear that the sum over all the possible permutations can be expressed in a compact way as a matrix determinant $\det\left(e^{-C(\mathbf{r}_\alpha-\mathbf{s}_\beta)^2}\right)$, where α and β denotes the matrix rows and columns, respectively. As a consequence,

$$\psi_{\text{FSWF}}(\mathbf{R}) = J_p(\mathbf{R})\int d\mathbf{S}\,\Phi(\mathbf{R},\mathbf{S})\Phi(\mathbf{S})J_s(\mathbf{S}) \tag{21}$$

that we refer to as Gaussian determinant (GD) $\Phi(\mathbf{R},\mathbf{S})$. This representation is particularly convenient, because, similar to the Slater determinant, it permits the summation over all $N!$ terms with a computational cost of $\mathcal{O}(N^3)$, where N is the number of atoms [21].

In other words, in this first scheme, the Gaussian product $\Xi(\mathbf{R},\mathbf{S})$ is replaced by the GD $\Phi(\mathbf{R},\mathbf{S})$. The corresponding results are illustrated in Fig. 2. Due to the fact that the parameter C of Eq. 3c is related to the mutual confinement of the particles and shadows, it is large in the crystalline phase, while for a liquid it is relatively small. As expected, the GD method reduces the variance, in particular for small values of C.

However, a potential limitation of the present scheme may arise due to sampling a sum of permuted terms, which means that only one of them is sampled efficiently, regardless of all the others.

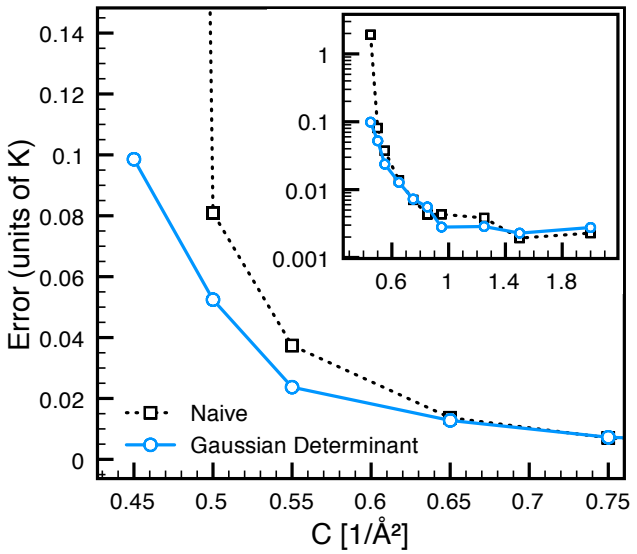

Figure 2. (Color online) Estimated error with and without the adoption of the GD approach, for different values of C. Results were obtained for $N = 38$ and $M = 64 \cdot 10^7$.

Explicit pair permutation term: Duet and Quartet

Therefore, an alternative approach is to add the contributions, which are due to a single pair permutation

$\mathcal{P}_{ab}$, to the original integrand, i.e.

$$\begin{aligned}\psi_{\text{FSWF}}(\mathbf{R}) &= \int d\mathbf{S}\ (\Gamma(\mathbf{R},\mathbf{S}) + \Gamma(\mathbf{R},\mathcal{P}_{ab}\mathbf{S})) \\ &= J_p(\mathbf{R}) \int d\mathbf{S}\, J_s(\mathbf{S})\Phi(\mathbf{S}) \\ &\times (\Xi(\mathbf{R},\mathbf{S}) - \Xi(\mathbf{R},\mathcal{P}_{ab}\mathbf{S}))\,. \end{aligned} \tag{22}$$

Compared with the GD method, the latter has the advantage of allowing to sample from a product of permuted terms, instead of from a sum:

$$\begin{aligned}\rho(\mathbf{R},\mathbf{S}_1,\mathbf{S}_2) &= \sqrt{\Gamma(\mathbf{R},\mathbf{S}_1)\Gamma(\mathbf{R},\mathcal{P}_{ab}\mathbf{S}_1)} \\ &\times \sqrt{\Gamma(\mathbf{R},\mathbf{S}_2)\Gamma(\mathbf{R},\mathcal{P}_{ab}\mathbf{S}_2)} \end{aligned} \tag{23}$$

At variance to the GD, the original as well as the permuted configuration are given equal importance in the sampling function, so that eventually their contributions will be of the same order, which results in a more effective mutual annihilation. The permuted term can either be added to $\Gamma(\mathbf{R},\mathbf{S}_1)$ alone or to both $\Gamma(\mathbf{R},\mathbf{S}_1)$ and $\Gamma(\mathbf{R},\mathbf{S}_2)$, respectively. Due to the fact that this results in two or four terms in the expression for the energy, we will refer to these schemes as the Duet and Quartet techniques, respectively.

To assess the effectiveness of this concept, we have performed a calculation for $N = 14$ and $M = 16 \cdot 10^6$ using the Duet technique. However, at variance to the naive approach, the fluctuations were so high that it was difficult to estimate the error. Therefore, the sampling function of Eq. 23 seems to be not optimal for integrating either $\Gamma(\mathbf{R},\mathbf{S})$ or $\Gamma(\mathbf{R},\mathcal{P}_{ab}\mathbf{S})$.

This is to say that a single permutation does not provide an effective antithetic contribution, as otherwise the product of Γ with its permuted term would exhibit its maximum in the same region where the function itself has its maximum, so that the sampling problem causing the ineffectiveness of the present schemewould not have emerged in the first place.

Permutation move

As we have just seen, even if a permutation implies a sign change, it is not necessarily resulting in an optimal antithetic contribution, which is due to the presence of the kernel that breaks the symmetry.

We can therefore infer that after performing a permutation, a specific translation needs to be added in order to obtain an effective antithetic contribution leading to mutual cancelation. Even though this translation is evidently unknown, it is yet possible to allow a walker to diffuse after a permutation, so that it can spontaneously move to the correct antithetic point. To implement this idea we need to consider permutations as proposed moves for the walkers in the context of the $\mathrm{M(RT)}^2$ algorithm [41]. To that extend, it is of upmost importance to take all possible pair permutations into account, and to select the most favorable one in order to maximize the acceptance rate.

Specifically, given a certain $\mathbf{R}$ and $\mathbf{S}$, we evaluate $\Xi(\mathbf{R},\mathcal{P}_{ij}\mathbf{S})$ for all the possible i and j and propose a permutation (a,b) according to the transition probability

$$T(\mathbf{S} \to \mathcal{P}_{ab}\mathbf{S}) = \frac{\Xi(\mathbf{R},\mathcal{P}_{ab})}{\sum_{(i,j)} \Xi(\mathbf{R},\mathcal{P}_{ij})}. \tag{24}$$

Thereafter, the acceptance probability has to be modified and reads as

$$A(\mathbf{S} \to \mathcal{P}_{ab}\mathbf{S}) = \frac{|\Gamma(\mathbf{R},\mathcal{P}_{ab}\mathbf{S})|T(\mathcal{P}_{ab}\mathbf{S} \to \mathbf{S})}{|\Gamma(\mathbf{R},\mathbf{S})|T(\mathbf{S} \to \mathcal{P}_{ab}\mathbf{S})}. \tag{25}$$

The permutation moves are proposed always after the evaluation of the estimator, in order to allow the walkers to diffuse before the next evaluation of $\mathrm{H}^{\mathrm{loc}}(\mathbf{R})$.

Following this procedure for $N = 14$ and $M = 16 \cdot 10^6$, we obtained $E = -1.997(11)$ K, which has to be compared to $E = -1.986(18)$ K using the naive algorithm. These values differ from those of Table I for $N = 66$ due to the presence of single-particle finite-size effects. The acceptance rate for the permutation moves was roughly 2.5%. This implies that it is actually possible to employ permutation moves, since the acceptance rate is significant, and that they indeed systematically reduce the variance. But, although our novel permutation moves lower the variance, this is largely due to the smaller correlation between successive steps, and thus cannot be a definitive solution to the sign problem, for which negative correlation ought to be introduced.

Reflections

In general, the integral over $\mathbf{S}$ is centered around $\mathbf{R}$ by the Gaussian term. This means that if the parameter C is small, there is a significant delocalization, and $\mathbf{S}$ is more likely to cross the nodal surface defined by $\Phi(\mathbf{S})$. However, if we additionally consider the contributions that are arising from a *reflected shadow* $\mathbf{S}' = 2\mathbf{R} - \mathbf{S}$, we will possibly obtain an opposite contribution. The underlying notion of this approach is illustrated in Fig. 3.

To take advantage of this idea, we sample from the usual pdf

$$\rho(\mathbf{R},\mathbf{S}_1,\mathbf{S}_2) = |\Gamma(\mathbf{R},\mathbf{S}_1)\Gamma(\mathbf{R},\mathbf{S}_2)|, \tag{26}$$

but sum up the contributions that are originating from the reflected shadows in the energy estimator

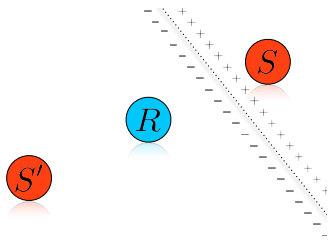

Figure 3. (Color online) Illustration of a reflected shadow $\mathbf{S}'$. Since $\mathbf{R}$ is close to the nodal surface defined by the SD, $\mathbf{S}$ and $\mathbf{S}'$ are in different nodal pockets.

$$E = \frac{\frac{1}{2}\sum_{i=1}^{M}\left\{\left[\frac{\big(\Gamma(\mathbf{R},\mathbf{S}_1)+\Gamma(\mathbf{R},\mathbf{S}'_1)\big)\mathrm{H}\big(\Gamma(\mathbf{R},\mathbf{S}_2)+\Gamma(\mathbf{R},\mathbf{S}'_2)\big)}{|\Gamma(\mathbf{R},\mathbf{S}_1)\Gamma(\mathbf{R},\mathbf{S}_2)|}+\frac{\big(\Gamma(\mathbf{R},\mathbf{S}_2)+\Gamma(\mathbf{R},\mathbf{S}'_2)\big)\mathrm{H}\big(\Gamma(\mathbf{R},\mathbf{S}_1)+\Gamma(\mathbf{R},\mathbf{S}'_1)\big)}{|\Gamma(\mathbf{R},\mathbf{S}_1)\Gamma(\mathbf{R},\mathbf{S}_2)|}\right]\right\}_i}{\sum_{i=1}^{M}\left\{\frac{\big(\Gamma(\mathbf{R},\mathbf{S}_1)+\Gamma(\mathbf{R},\mathbf{S}'_1)\big)\big(\Gamma(\mathbf{R},\mathbf{S}_2)+\Gamma(\mathbf{R},\mathbf{S}'_2)\big)}{|\Gamma(\mathbf{R},\mathbf{S}_1)\Gamma(\mathbf{R},\mathbf{S}_2)|}\right\}_i}. \tag{27}$$

But, in a realistic calculation $\Gamma(\mathbf{R},\mathbf{S}')$ turns out to be considerably smaller than $\Gamma(\mathbf{R},\mathbf{S})$, which implies that its contribution is essentially negligible. Specifically, using 14 particles, the estimated energy and error, with and without the adoption of the GD method, improved by less a factor of 10^{-4}. This marginal enhancement immediately suggests that our initial suggestion to generate antithetic contributions by employing reflected shadows needs to be reconsidered. Finally, we remark that the presented reflection method suffers from an infinite variance problem, which can be effectively eliminated by removing the zero values from the sampling function.

Constrained Domains

Eventually, it is possible to use symmetry arguments to constrain the domain of the integrals over $\mathbf{R}$, $\mathbf{S}_1$ and $\mathbf{S}_2$, which results in a significant reduction of the integration space. First of all, due to the antisymmetry requirement of the FSWF, $\mathbf{R}$ is constrained to the positive (or negative) domain of the corresponding $\Phi(\mathbf{R})$. No particular form is required for $\Phi(\mathbf{R})$: the present method is correct independent of its choice. The second symmetry argument is only valid in conjunction with the GD method. In fact, if we sum over all the permutations of $\mathbf{S}$, it is possible to integrate $\mathbf{S}_1$ and $\mathbf{S}_2$ only in the positive (or negative) domains, i.e. where $\Phi(\mathbf{S}_1)$ and $\Phi(\mathbf{S}_2)$ are positive (or negative). We stress that these restrictions do not imply that Γ must always be positive (or negative), as the Gaussian determinant permits a change of sign.

A simulation using the above described constrained domains method, with $N = 38$ and $M = 60 \cdot 10^6$, yielded $E = -2.5(10)$ K, whereby 5% and 0.5% of the moves for $\mathbf{R}$ and $\mathbf{S}$, respectively, were rejected due to the constraints. Comparing this with $E = -1.9(1)$ K using the bare GD technique without any restrictions, it is clear that no reduction of the sign fluctuations has been achieved. However, the fact that error increases by a factor of 10 is surprising, but might be explained to be most likely a consequence that when integrating on a restrained domain the efficiency decreases near its borders.

THE GROUPING TECHNIQUE AND THE MARGINAL DISTRIBUTION

A completely different approach can be devised by analyzing the expression for the energy with a special attention on the integrals over $\mathbf{S}_1$ and $\mathbf{S}_2$, i.e.

$$E = \frac{\int d\mathbf{R}\,\left(\int d\mathbf{S}_1\,\Gamma(\mathbf{R},\mathbf{S}_1)\right)\left(\int d\mathbf{S}_2\,\mathrm{H}\Gamma(\mathbf{R},\mathbf{S}_2)\right)}{\int d\mathbf{R}\,\underbrace{\left(\int d\mathbf{S}_1\,\Gamma(\mathbf{R},\mathbf{S}_1)\right)}_{\Omega_1(\mathbf{R})}\underbrace{\left(\int d\mathbf{S}_2\,\Gamma(\mathbf{R},\mathbf{S}_2)\right)}_{\Omega_2(\mathbf{R})}}. \tag{28}$$

We point out that $\Omega_1(\mathbf{R}) = \Omega_2(\mathbf{R}) = \psi_{\mathrm{FSWF}}(\mathbf{R})$. From this it follows that knowing $\psi_{\mathrm{FSWF}}(\mathbf{R})$, i.e. knowing the analytical solution of the integral over $\mathbf{S}$, the sign problem ceases to exist, since $\Omega_1(\mathbf{R})\Omega_2(\mathbf{R}) = \psi^2_{\mathrm{FSWF}}(\mathbf{R}) \geq 0$. The fact that $\psi_{\mathrm{FSWF}}(\mathbf{R})$ is apparently unknown has the following two major consequences. First, $\Omega_1(\mathbf{R})$ and $\Omega_2(\mathbf{R})$ needs to be approximated by sampling $\Gamma(\mathbf{R},\mathbf{S}_1)$ and $\Gamma(\mathbf{R},\mathbf{S}_2)$, respectively. Due to the fact that the estimates may have have different signs, the local energy is weighted by an extremely noisy function. Second, $\mathbf{R}$ is not efficiently sampled from its marginal distribution $\psi^2_{\mathrm{FSWF}}(\mathbf{R})$. The following example is intended to illustrate this subtle point. Suppose that $\mathbf{R}_0$ is on the nodal surface, i.e. $\psi_{\mathrm{FSWF}}(\mathbf{R}_0) = 0$. Even though this configuration should never be sampled, since its probability is identical to zero by its very definition, it will, nevertheless, be

M_s	Efficiency
1	0.63
10	0.36
100	0.36
1000	0.23

Table II. Efficiency in units of $1/(\sec \times K^2)$ of the improved grouping technique in conjunction with the GD approach for $N = 14$ as a function of the block size M_s. The efficiency is defined as 1/(simulation time × variance).

sampled with a finite probability of $\Gamma(\mathbf{R}_0, \mathbf{S}_1)\Gamma(\mathbf{R}_0, \mathbf{S}_2)$, which is a manifestation that that integrand and the integral may greatly differ from each other. Moreover, if $\psi_{\mathrm{FSWF}} \simeq 0$, $\Gamma(\mathbf{R}, \mathbf{S}_1)$ and $\Gamma(\mathbf{R}, \mathbf{S}_2)$ are contributing only with noise.

The first problem has been previously addressed in a study of the vacancy formation energy in solid ^{3}He using the grouping technique that is based on the aforementioned blocking scheme [17]. We have modified this approach in our improved grouping technique by sampling many successive shadows in order to obtain a rough estimate of $\Omega_1(\mathbf{R})$ and $\Omega_2(\mathbf{R})$. The present algorithm then reads as:

1. Start from a configuration $\mathbf{R}_{(0)}$, $\mathbf{S}_{1(0,1)}$, as well as $\mathbf{S}_{2(0,1)}$ and set $i = 1$
2. Sample $\mathbf{R}_{(i)}$ from the pdf
$$\rho(\mathbf{R}) = |\Gamma(\mathbf{R}, \mathbf{S}_{1(i-1,1)})\Gamma(\mathbf{R}, \mathbf{S}_{2(i-1,1)})|$$
3. Sample M_s points $\left(\mathbf{S}_{1(i,1)}, \ldots, \mathbf{S}_{1(i,M_s)}\right)$ from
$$\rho(\mathbf{S}_1) = |\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_1)\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{2(0,1)})|$$
and analog $\left(\mathbf{S}_{2(i,1)}, \ldots, \mathbf{S}_{2(i,M_s)}\right)$ from the pdf
$$\rho(\mathbf{S}_2) = |\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{1(i,M_s)})\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_2)|$$
4. Evaluate $\Omega_1(\mathbf{R}_{(i)}) = \sum_{j=1}^{M_s} \frac{\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{1(i,j)})}{|\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{1(i,j)})|}$ and $\Omega_2(\mathbf{R}_i) = \sum_{j=1}^{M_s} \frac{\Gamma(\mathbf{R}_i, \mathbf{S}_{2(i,j)})}{|\Gamma(\mathbf{R}_i, \mathbf{S}_{2(i,j)})|}$
5. Evaluate $\mathrm{H}_1^{\mathrm{loc}}(\mathbf{R}_i) = \sum_{j=1}^{M_s} \frac{\mathrm{H}\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{1(i,j)})}{|\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{1(i,j)})|}$ and $\mathrm{H}_2^{\mathrm{loc}}(\mathbf{R}_i) = \sum_{j=1}^{M_s} \frac{\mathrm{H}\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{2(i,j)})}{|\Gamma(\mathbf{R}_{(i)}, \mathbf{S}_{2(i,j)})|}$
6. Set $i = i + 1$
7. Repeat the steps 2 to 6 M times
8. Compute $E = \frac{\frac{1}{2}\sum_{i=1}^{M}\left(\Omega_1(\mathbf{R}_i)\mathrm{H}_2^{\mathrm{loc}}(\mathbf{R}_i) + \Omega_2(\mathbf{R}_i)\mathrm{H}_1^{\mathrm{loc}}(\mathbf{R}_i)\right)}{\sum_{i=1}^{M}\Omega_1(\mathbf{R}_i)\Omega_2(\mathbf{R}_i)}$

We have repeated the calculation using the blocking technique to ensure that successively sampled $\mathbf{R}$ values were actually decorrelated and thus finding the optimal number of diffusive steps $\mathrm{M}_{\mathrm{diff}}$. However, we found that even though the improved grouping technique indeed stabilizes the sign of $\Omega_1(\mathbf{R})\Omega_2(\mathbf{R})$, the computational time required to do so is not entirely compensated by the reduced variance. This can be seen in Table II, where the efficiency that is defined as 1/(simulation time×variance) is reported for different values of M_s. Nevertheless, although this scheme alone does not improve the efficiency for the liquid phase, it potentially does for the solid state, where the sign problem is less severe.

In order to make further progress, we will focus on the second problem, whose solution may also solve the first one *en passant*. To this end we are going to propose two different methods: In the first approach the marginal distribution is approximated analytically, whereas in the second scheme it is estimated numerically instead.

J-SD approximation

In our first approach, the J-SD trial wave function is employed as an approximation for $\psi_{\mathrm{FSWF}}(\mathbf{R})$ to sample $\mathbf{R}$, which is why we call this technique J-SD approximation. However, in this way $\mathbf{R}$ would be sampled independent from its shadows, which would require to relax $\mathbf{S}_1$ and $\mathbf{S}_2$ whenever $\mathbf{R}$ changes. To avoid this, we have decided to use the same sampling function for $\mathbf{R}$ and $\mathbf{S}$. At variance to the just described algorithm based on the blocking technique, $\Gamma(\mathbf{R}, \mathbf{S})$ is replaced by $\Gamma'(\mathbf{R}, \mathbf{S})$ together with appropriate weights in the energy estimator. Specifically, the following forms are proposed here, which all incorporates $\Phi(\mathbf{R})$ into the sampling function for $\mathbf{R}$:

- $\Gamma'_1(\mathbf{R}, \mathbf{S}) = \Gamma(\mathbf{R}, \mathbf{S})\left(\Phi(\mathbf{R})^2 + \Lambda^2\right)^{1/4}$
- $\Gamma'_2(\mathbf{R}, \mathbf{S}) = \Gamma(\mathbf{R}, \mathbf{S})\left(\Phi(\mathbf{R})^2 + \Lambda^2\right)$
- $\Gamma'_3(\mathbf{R}, \mathbf{S}) = \Gamma(\mathbf{R}, \mathbf{S})\left(\Phi(\mathbf{R})^2 + \Lambda^2\right)^{1/2}$
- $\Gamma'_4(\mathbf{R}, \mathbf{S}) = \Gamma(\mathbf{R}, \mathbf{S})\left(\Phi(\mathbf{R})^4 + \Lambda^4\right)^{1/4}$

To prevent the infinite variance problem, we have introduced an auxiliary factor Λ, whose optimal value is expected to be of the same order as $\Phi(\mathbf{R})$.

From Fig. 4 we can conclude that by incorporating $\Phi(\mathbf{R})$ into the sampling function it is possible to reduce the variance by up to a factor of 3/2. Among the various sampling functions we proposed, Γ'_4 appears to be the most effective, which demonstrates that, in the absence of the extra term, $\mathbf{R}$ is not efficiently sampled from its marginal distribution.

S-averaged marginal distribution

An alternative possibility that we have investigated here is to employ a numeric estimate of Ω_1 and Ω_2

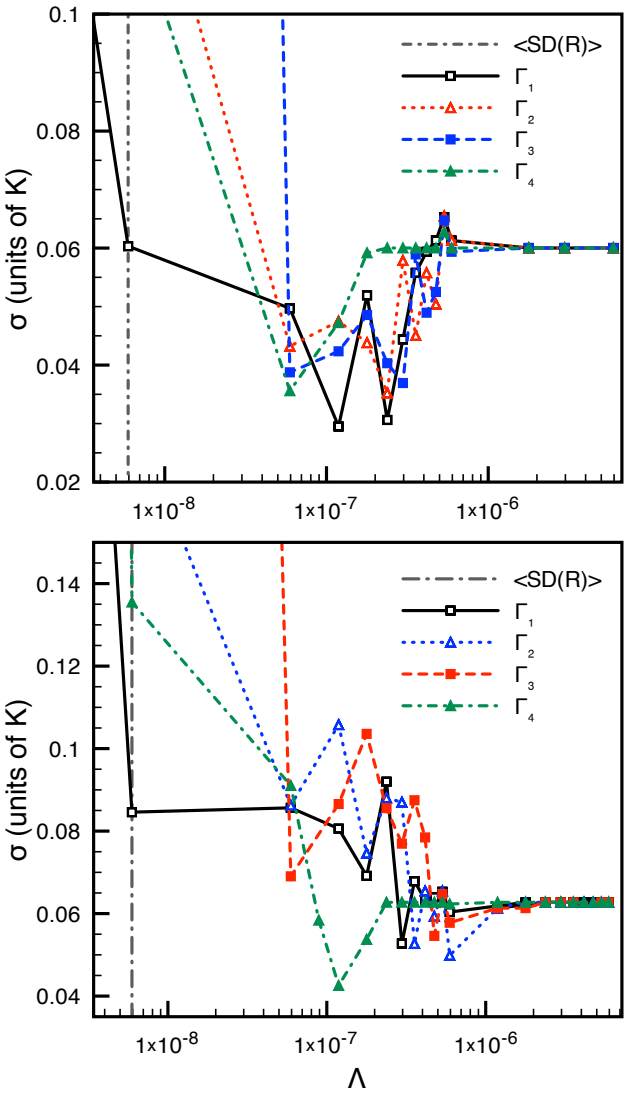

Figure 4. (Color online) Comparison of the error for $N = 38$ with respect to $\Lambda(\mathbf{R})$ as obtained using the GD approach together with the modified grouping technique in conjunction with the J-SD approximation, i.e. $\Gamma'(\mathbf{R}, \mathbf{S})$ instead of $\Gamma(\mathbf{R}, \mathbf{S})$. In the upper panel, the blocking technique with $M_s = 100$ was used and averaged over $M = 96 \cdot 10^4$ points. Instead, in the lower panel $M = 48 \cdot 10^6$, though without utilizing the blocking technique. The expectation value $\langle \Phi(\mathbf{R}) \rangle$ was evaluated by sampling from the pdf $\Gamma(\mathbf{R}, \mathbf{S}_1)\Gamma(\mathbf{R}, \mathbf{S}_2)$.

as the sampling function. To that extent we assume $2N_S$ shadows $\mathbf{S}_{1,1}, \mathbf{S}_{1,2}, \ldots, \mathbf{S}_{1,N_S}, \mathbf{S}_{2,1}, \mathbf{S}_{2,2}, \ldots, \mathbf{S}_{2,N_S}$, and replace

$$\int d\mathbf{S}_1\, \Gamma(\mathbf{R}, \mathbf{S}_1) \quad \text{with} \quad \sum_{i=1}^{N_S} \int d\mathbf{S}_{1,i}\, \Gamma(\mathbf{R}, \mathbf{S}_{1,i}), \quad (29a)$$

N_S	Error
1	0.024
2	0.025
4	0.034
10	0.029

Table III. The error for $N = 14$ and $M = 8{\cdot}10^6$ as obtained by the $\mathbf{S}$-averaged marginal distribution approach in conjunction with the GD method with respect to the number of shadows $2N_S$.

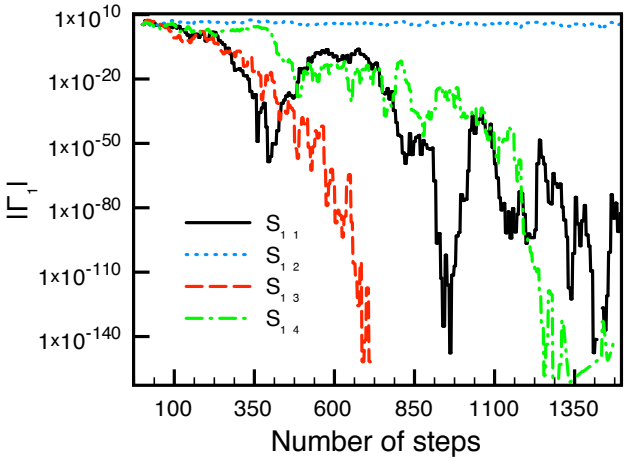

Figure 5. (Color online) Trend of the various Γ corresponding to $\mathbf{S}_{1,1}, \ldots, \mathbf{S}_{1,4}$, during the progression of a simulation. The line associated with $\mathbf{S}_{1,3}$ is interrupted when, because of the double precision float limitations, the value of Γ is so small that it is numerically identical to zero and is no longer visualizable on a logarithmic scale.

and

$$\int d\mathbf{S}_2\, \Gamma(\mathbf{R}, \mathbf{S}_2) \quad \text{with} \quad \sum_{i=1}^{N_S} \int d\mathbf{S}_{2,i}\, \Gamma(\mathbf{R}, \mathbf{S}_{2,i}). \quad (29b)$$

From this it follows that our sampling function will take the form

$$\rho(\mathbf{R}, \mathbf{S}_{1,1}, \ldots, \mathbf{S}_{1,N_S}, \mathbf{S}_{2,1}, \ldots, \mathbf{S}_{2,N_S}) = \left| \left(\sum_{i=1}^{N_S} \Gamma(\mathbf{R}, \mathbf{S}_{1,i}) \right) \left(\sum_{i=1}^{N_S} \Gamma(\mathbf{R}, \mathbf{S}_{2,i}) \right) \right|. \quad (30)$$

In this way, $\mathbf{R}$ is sampled from a more accurate approximation of $\Omega_1(\mathbf{R})\Omega_2(\mathbf{R})$ than in the J-SD approximation.

The results, which are shown in Table III, imply that the introduction of additional shadows does not have any statistical significant influence on the variance. From Fig 5 it is clear why assuming multiple shadows is not effective: During the sampling only one of the $\mathbf{S}_{1,i}$ and one of the $\mathbf{S}_{2,i}$ become significant, whereas all the others

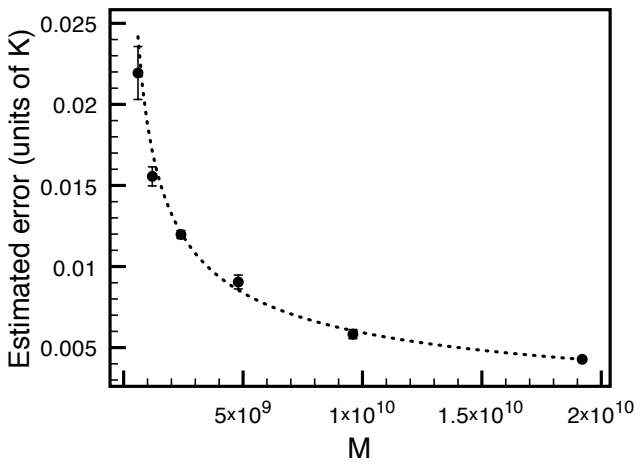

Figure 6. Fitting procedure to estimate the efficiency more accurately. Here we present the data obtained with the naive method, simulating 38 ^{3}He atoms. Each estimated error was calculated averaging over the eight independent simulations. The fitted function $f(x) = A/\sqrt{M}$ is represented with a dotted line. The chi-squared test was successfully fulfilled.

tend to zero. As a result, the algorithm returns to its original form, which is hence an inherent consequence of sampling from a sum of $\Gamma(\mathbf{R}, \mathbf{S}_{1,i})$ and $\Gamma(\mathbf{R}, \mathbf{S}_{2,i})$, respectively.

DISCUSSION

The reduced variance of the presented methods, specifically the GD and the J-SD approaches, has to be related to the required computational effort. Thus, for the purpose of assessing the various techniques presented here, we have summarized their corresponding efficiencies in Table IV. All the presented results were obtained by means of the M(RT)^2 algorithm [41], where single-particle or single-shadow moves were proposed at random without any additional drift term. In this process, the step lengths were adjusted to yield an acceptance ratio of $\sim 50\%$. Due to the fact that the Duet, Quartet, and *R-S* domain constraint methods are apparently inefficient, no error bars for the efficiency were calculated. On the contrary, the J-SD approximation somewhat reduces the variance, though this is largely eroded by the additional computational cost. As a consequence, the J-SD method is only marginally more efficient. The GD technique, however, does indeed exhibit a sizable variance reduction. In spite of its increased computational cost to evaluate the Gaussian determinants, it is yet very competitive with the original approach, though generally not significantly more efficient either. Nevertheless, for $C < 0.5$ Å^{-2} the GD method is clearly superior.

Eventually, the combination of the GD and J-SD approximation methods turned out to be the best among the various technique we have devised here. For this reason we have reviewed the accuracy of our results, in order to exclude the possibility that our outcomes were affected by an ergodicity problem. To that extend we have performed several additional calculations for different values of M. Then we have fitted the obtained errors to the function $f(x) = A/\sqrt{M}$, which is the expected asymptotic behavior, as shown in Fig. 6. The fact that the chi-squared test was passed successfully indicates that M was large enough to ensure ergodicity. In addition, the eventual parameter A can be used as a more accurate estimate for the efficiency instead of the variance. The final results are reported in Table V.

Used technique	**Energy** [K]	**Efficiency** [sec^{-1}K^{-2}]
Naive	-1.949 ± 0.016	30 ± 3
Gaussian determinant	-1.943 ± 0.014	22 ± 3
Duet	-2.7 ± 4.3	≈ 0
Quartet	-2.2 ± 1.6	≈ 0
Permutation moves	-2.012 ± 0.016	25 ± 3
Reflections†	-1.945 ± 0.016	15 ± 2
R-S domain constraint†	-1.974 ± 0.028	≈ 5
R domain constraint†	-1.975 ± 0.017	24 ± 3
J-SD approximation	-1.953 ± 0.015	31 ± 3
J-SD approximation†	-1.949 ± 0.017	35 ± 7

Table IV. Average energy, associated error and efficiency of all the presented methods. In each case, we have performed eight independent simulations with $N = 38$ and $M = 12 \cdot 10^8$. For the J-SD approximation method we set $M_s = 1$. We did not estimate the efficiency error for certain methods, because they clearly proved unable to provide an improvement. The symbol † denotes that the technique was used in conjunction with the GD method. The efficiency of the Duet and Quartet approaches was extremely low (minor than 10^{-3}) and therefore approximated to be zero. The *R-S* domain constraint scheme refers to the algorithm described above, whereas for the *R* domain constraint technique, we limited the constraint only to *R*. For the sake of readability, we multiplied the efficiency by an arbitrary value of 100.

CONCLUSIONS

To summarize, beside revisiting the FSWF and demonstrating the origin and implications of the corresponding sign problem, we have proposed two families of novel

Technique	N=16 - Efficiency	N=38 - Efficiency	N=54 - Efficiency
Naive	1375 ± 50	3.03 ± 0.06	0.0210 ± 0.0005
Gaussian Determinant	1220 ± 30	2.54 ± 0.14	0.0216 ± 0.0032
J-SD approximation*	-	3.23 ± 0.22	-

Table V. Accurate efficiency estimates of the GD and J-SD techniques, for different number of atoms N. We decided to pass over evaluating the efficiency of the "J-SD approximation*" method for $N > 38$, since there are no reason to expect significantly different results.

methods to solve it: Antithetic variates and an improved marginal distribution to sample from. Several specific implementations of these ideas were presented. Even though the GD and J-SD methods are indeed rather effective in reducing the variance, the gain in efficiency is limited due to increased computational cost associated with them.

We thus conclude that although the presented techniques alleviate the sign problem and allow for accurate accurate calculations of fermionic systems up to 66 particles, at least when using state of the art supercomputers, a general solution of the sign problem is still outstanding.

Financial support from the IDEE project of the Carl Zeiss Foundation is kindly acknowledged. M.K. and F.C. would like to acknowledge the hospitality of the Lawrence Livermore National Laboratory. T.D.K and F.C. would like to thank the Graduate School of Excellence MAINZ and the Max-Planck Graduate Center for financial support and the Gauss Center form Supercomputing (GCS) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS share of the supercomputer JUQUEEN at the Jülich Supercomputing Center (JCS). This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

* francesco.calcavecchia@gmail.com
† tdkuehne@mail.upb.de

[1] M. A. Morales, R. Clay, C. Pierleoni, and D. M. Ceperley, Entropy **16**, 287 (2014).
[2] B. M. Austin, D. Y. Zubarev, and W. A. Lester, Chem. Rev. **112**, 263 (2012).
[3] J. Kolorenc and L. Mitas, Rep. Prog. Phys. **74**, 026502 (2011).
[4] R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Rios, J. Phys.: Condens. Matter **22**, 023201 (2010).
[5] W. L. McMillan, Phys. Rev. **138**, A442 (1965).
[6] K. Binder, Rep. Prog. Phys. **60**, 487 (1997).
[7] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley-VCH, Weinheim, 2008).
[8] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2013).
[9] J. A. Pople, Rev. Mod. Phys. **71**, 1267 (1999).
[10] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2013).
[11] R. Jastrow, Phys. Rev. **98**, 1479 (1955).
[12] S. Vitiello, K. Runge, and M. H. Kalos, Phys. Rev. Lett. **60**, 1970 (1988).
[13] L. Reatto and G. L. Masserini, Phys. Rev. B **38**, 4516 (1988).
[14] F. Pederiva, A. Ferrante, S. Fantoni, and L. Reatto, Phys. Rev. Lett. **72**, 2589 (1994).
[15] F. Pederiva, G. V. Chester, S. Fantoni, and L. Reatto, Phys. Rev. B **56**, 5909 (1997).
[16] F. Operetto and F. Pederiva, Phys. Rev. B **69**, 024203 (2004).
[17] L. Dandrea, F. Pederiva, S. Gandolfi, and M. H. Kalos, Phys. Rev. Lett. **102**, 255302 (2009).
[18] M. H. Kalos and L. Reatto, in *Progress in Computational Physics of Matter*, edited by L. Reatto and F. Manghi (World Scientific, Singapore, 1995).
[19] F. Calcavecchia, F. Pederiva, and T. D. Kühne, Journal of Unsolved Questions **1**, 13 (2011).
[20] F. Pederiva, S. A. Vitiello, K. Gernoth, S. Fantoni, and L. Reatto, Phys. Rev. B **53**, 15129 (1996).
[21] J. C. Slater, Phys. Rev. **34**, 1293 (1929).
[22] C. Lhuillier and D. Levesque, Phys. Rev. B **23**, 2203 (1981).
[23] R. P. Feynman, Phys. Rev. **94**, 262 (1954).
[24] R. P. Feynman and M. Cohen, Phys. Rev. **102**, 1189 (1956).
[25] K. E. Schmidt and V. R. Pandharipande, Phys. Rev. B **19**, 2504 (1979).
[26] M. A. Lee, K. E. Schmidt, M. H. Kalos, and G. V. Chester, Phys. Rev. Lett. **46**, 728 (1981).
[27] K. E. Schmidt, M. A. Lee, M. H. Kalos, and G. V. Chester, Phys. Rev. Lett. **47**, 807 (1981).
[28] S. Moroni, S. Fantoni, and G. Senatore, Europhys. Lett. **30**, 93 (1995).
[29] Y. Kwon, D. M. Ceperley, and R. M. Martin, Phys. Rev. B **58**, 6800 (1998).
[30] M. Holzmann, B. Bernu, and D. M. Ceperley, Phys. Rev. B **74**, 104510 (2006).
[31] F. Pederiva and G. V. Chester, J. Low Temp. Phys. **113**, 741 (1998).
[32] S. Moroni, S. Fantoni, and G. Senatore, Phys. Rev. B **52**, 13547 (1995).
[33] R. De Bruyn Ouboter and C. N. Yang, Physica B+C **144**, 127 (1987).
[34] Note1, We considered an unpolarized 3D system of ^{3}He at a density of 0.016588 Å^{-3} (liquid phase) using the

Aziz potential HFDHE2 [42, 43], and employing periodic boundary conditions in order to mimic an essentially infinite system. We remark that whenever a SD of simple plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$ is used, the occurrence of a drift (i.e. $\Sigma_\beta \mathbf{k}_\beta \neq 0$), as well as anisotropy has to be explicitly taken into account. The simplest way to remedy this is to consider only magic numbers for N, i.e. numbers that fill the momenta shell. For a 3D polarized system, they are 1, 7, 19, 27, 33, etc. and for an unpolarized system 2, 14, 38, 54, 66, etc. With this choice, $\Phi(\mathbf{R})$ can be assumed to be real. We have set the variational parameters for the SWF to be $b = 2.76$ Å$, c_1 = 0.11$ K$^{-1}, c_2 = 0.88$ and $C = 0.55$ Å^{-2}, as suggested in [20], whereas for the J-SD trial wave function we have employed $b = 2.9$ Å. .

[35] S. Azadi, W. M. C. Foulkes, and T. D. Kühne, New J. Phys. **15**, 113005 (2013).

[36] J.-P. Bouchaud and C. Lhuillier, Europhys. Lett. **3**, 1273 (1987).

[37] H. Flyvbjerg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).

[38] M. H. Kalos, D. Levesque, and L. Verlet, Phys. Rev. A **9**, 2178 (1974).

[39] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[40] J. M. Hammersley and K. W. Morton, Proc. Cam. Phil. Soc. **52**, 449 (1956).

[41] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[42] R. A. Aziz, V. P. S. Nain, J. S. Carley, W. L. Taylor, and G. T. McConville, J. Chem. Phys. **70**, 4330 (1979).

[43] S. A. Sofianos, S. A. Rakityansky, and S. E. Massen, Phys. Rev. A **60**, 337 (1999).