

# Global space-time update

V.G. Rousseau<sup>1</sup> and D. Galanakis<sup>2</sup>

<sup>1</sup>*Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA*

<sup>2</sup>*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371*

The Stochastic Green Function (SGF) algorithm is able to simulate any Hamiltonian that does not suffer from the so-called “sign problem”. We propose a new global space-time update scheme for the SGF algorithm which, in addition to being simpler than the previous formulation, reduces auto-correlation times. Using as a concrete example the extended Bose-Hubbard model and the complex Hamiltonian with six-site ring-exchange interactions which was recently studied in arXiv:1206.2566v1, we present a comprehensive review of the SGF algorithm and the new updating scheme. Measurements of non-trivial physical quantities are presented in detail. While the SGF algorithm works in the canonical ensemble by nature, we give a simple extension that allows to perform simulations in the grand-canonical ensemble too. We also discuss an optimized implementation which allows for access to large system sizes.

PACS numbers: 02.70.Uu, 05.30.Jp

## I. INTRODUCTION

Monte Carlo methods were first introduced more than 60 years ago [1] to solve various classical problems. The development of quantum Monte Carlo (QMC) [2–4] allowed the extension of those methods to quantum systems. For fermions, the power of QMC methods remains limited by the so-called *sign problem*, which prevents the treatment of systems with large sizes. While the Determinant Quantum Monte Carlo algorithm [4] allows to treat fermions exactly in some particular cases, it is often necessary to make use of approximate methods such as the fixed-node approximation [5], and the dynamical cluster approximation [6, 7]. On the other hand there are many interesting bosonic systems which do not suffer from the sign problem, and which can be described efficiently using a worldline representation [8]. Consequently, over the last two decades, there have been tremendous advances in boson QMC methods, such as the development of the Stochastic Series Expansion (SSE) algorithm [9], the loop algorithm [10] which makes use of global updates in the grand-canonical ensemble, or the Reptation Quantum Monte Carlo algorithm [11]. In recent years the canonical worm (CW) algorithm [12], which works in the canonical ensemble with global updates, was proposed. The method was generalized later to a wider class of Hamiltonians, leading to the Stochastic Green Function (SGF) algorithm [13], and improved with a new type of *directed* updates [14].

Recently, the generality of the SGF algorithm made it possible to perform exact studies of complex systems, such as multispecies systems with interspecies conversions [15–19], systems with fully connected graphs [20], and systems described by Hamiltonians with six-site coupling terms [21].

In this paper we review the theory of the SGF algorithm and propose a new *global space-time update* with reduced auto-correlation times. The paper is organized as follows: In section II we describe the properties and the framework of the SGF algorithm. In section III we

present the new global space-time update scheme and derive the expression of all associated probabilities that satisfy detailed balance. We detail the differences between this new update and the updates that are used in the SSE and CW algorithms. Section IV gives full details on measurements. In particular we describe how non-trivial quantities can be measured, such as the specific heat, the imaginary dynamical structure factor, the entropy, or  $n$ -point Green functions. We give in section V a simple extension that allows the algorithm to simulate exactly the grand-canonical ensemble. We propose in Section VI an efficient implementation of the algorithm. We illustrate in section VII the exactness of the algorithm by making comparisons between the SGF algorithm and exact diagonalizations, and show that the new global space-time update leads to smaller auto-correlation times. Finally we conclude in section VIII.

## II. GENERALITIES

### A. Properties

The SGF algorithm is characterized by the following properties:

1. It can be applied to any sign-problem-free Hamiltonian.
2. It is completely independent of the structure of the Hamiltonian, no particular decomposition is required.
3. As a corollary, it is possible to write a single computer code that can simulate any sign-problem-free Hamiltonian.
4. The acceptance rate of every update is 100%. The benefit of that is efficiency, as no cpu time is wasted with useless rejected updates and implementation simplicity because the changes made in the configuration during an update do not need to be stored.

5. It can simulate both the canonical and the grand-canonical ensembles.
6. It makes use of a global space-time update (see section III), which reduces the auto-correlation time (see section VII).
7. It allows for the measurement of high-order correlation functions, such as  $n$ -point Green functions.
8. It works in continuous imaginary time and is exact (no errors beyond statistical errors).
9. It ensures the ergodicity. In particular, the winding is sampled.

The above properties 1 – 7 illustrate the differences between the SGF algorithm and other existing algorithms. In particular, while other algorithms can treat only a specific class of Hamiltonians (usually Hamiltonians that can be written as a sum of two-site coupling terms), the SGF algorithm can be directly applied to any sign-problem-free Hamiltonian “as is”.

In order to give a concrete illustration of the above abilities of the SGF algorithm, we consider in the following two Hamiltonians. The first Hamiltonian is the one-dimensional extended Bose-Hubbard model, which is familiar:

$$\hat{\mathcal{H}} = -t \sum_{\langle i,j \rangle} (a_i^\dagger a_j + H.c.) + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) + V \sum_{\langle i,j \rangle} \hat{n}_i \hat{n}_j \quad (1)$$

The operators  $a_i^\dagger$  and  $a_i$  are the creation and annihilation operators of a boson on site  $i$ , which satisfy the usual boson commutation rules, and  $\hat{n}_i = a_i^\dagger a_i$  is the number operator on site  $i$ . The sum  $\sum_{\langle i,j \rangle}$  is over all pairs of first-neighboring sites  $i$  and  $j$ . The second Hamiltonian describes soft-core bosons on a two-dimensional kagome lattice interacting via an onsite repulsion potential and a sextic ring-exchange term. The Hamiltonian takes the form:

$$\hat{\mathcal{H}} = -t \sum_{\langle i,j \rangle} (a_i^\dagger a_j + H.c.) - K \sum_{\bigcirc} (a_{\bigcirc 1}^\dagger a_{\bigcirc 3}^\dagger a_{\bigcirc 5}^\dagger a_{\bigcirc 6} a_{\bigcirc 4} a_{\bigcirc 2} + H.c.) + U \sum_i \hat{n}_i (\hat{n}_i - 1) \quad (2)$$

The sum  $\sum_{\bigcirc}$  is over all hexagons of the lattice. The indices  $\bigcirc 1 - \bigcirc 6$  in the sextic term denote the sites of a given hexagon  $\bigcirc$ . This Hamiltonian has been studied recently in reference [21] in the hard-core case. Because of the sextic term which couples six sites at a time, simulating this Hamiltonian with usual methods is cumbersome. We demonstrate below that it is straightforward to simulate it with the SGF algorithm.

## B. The (extended) partition function and the Green operator

Like many other QMC algorithms, the SGF algorithm operates directly on physical states. An occupation number basis is chosen,  $\mathcal{B} = \{|\psi\rangle\}$ , where  $|\psi\rangle$  is a configuration in the occupation number representation. We consider a Hamiltonian written in the form

$$\hat{\mathcal{H}} = \hat{\mathcal{V}} - \hat{\mathcal{T}}, \quad (3)$$

where  $\hat{\mathcal{V}}$  is diagonal in the basis  $\mathcal{B}$ , and  $\hat{\mathcal{T}}$  is off-diagonal and assumed to have positive matrix elements. The SGF algorithm does not require any further assumptions on the Hamiltonian. Defining the inverse temperature  $\beta$ , the purpose of the algorithm is to sample the partition function

$$\mathcal{Z}(\beta) = \text{Tr } e^{-\beta \hat{\mathcal{H}}}. \quad (4)$$

To this end, we define the *Green operator*  $\hat{\mathcal{G}}$  by its matrix elements for all pairs of states  $|L\rangle$  and  $|R\rangle$ ,

$$\langle L | \hat{\mathcal{G}} | R \rangle = g(n), \quad (5)$$

where  $n$  is the number of creations and annihilations needed to transform the state  $|R\rangle$  into the state  $|L\rangle$ , and  $g$  is an arbitrary function (see section III), with the constraint  $g(0) = 1$ , that is all diagonal matrix elements of  $\hat{\mathcal{G}}$  are equal to 1. In the following,  $n$  will be referred to as the *offset* of the Green operator. By breaking up the exponential in (4) at imaginary time  $\tau$  and introducing the Green operator between the two parts, we can define an *extended* partition function:

$$\mathcal{Z}(\beta, \tau) = \text{Tr } e^{-(\beta-\tau)\hat{\mathcal{H}}} \hat{\mathcal{G}} e^{-\tau\hat{\mathcal{H}}} \quad (6)$$

Defining  $\hat{\mathcal{T}}(\tau) = e^{\tau\hat{\mathcal{V}}} \hat{\mathcal{T}} e^{-\tau\hat{\mathcal{V}}}$  and  $\hat{\mathcal{G}}(\tau) = e^{\tau\hat{\mathcal{V}}} \hat{\mathcal{G}} e^{-\tau\hat{\mathcal{V}}}$ , and using the equality

$$e^{-\beta(\hat{\mathcal{V}}-\hat{\mathcal{T}})} = e^{-\beta\hat{\mathcal{V}}} \overleftarrow{\mathbf{T}}_{\tau'} e^{\int_0^\beta \hat{\mathcal{T}}(\tau') d\tau'}, \quad (7)$$

where  $\overleftarrow{\mathbf{T}}_{\tau'}$  denotes the time-ordering operator over the variable  $\tau'$  with time increasing from the right to the left, the extended partition function takes the form:

$$\mathcal{Z}(\beta, \tau) = \text{Tr } e^{-\beta\hat{\mathcal{V}}} \overleftarrow{\mathbf{T}}_{\tau'} \left[ e^{\int_\tau^\beta \hat{\mathcal{T}}(\tau') d\tau'} \hat{\mathcal{G}}(\tau) e^{\int_0^\tau \hat{\mathcal{T}}(\tau') d\tau'} \right] \quad (8)$$

By expanding the exponentials in (8) and ordering the operators in imaginary time, we get:

$$\begin{aligned} \mathcal{Z}(\beta, \tau) = \text{Tr } e^{-\beta\hat{\mathcal{V}}} \sum_{n \geq 0} \int_{0 < \tau_1 < \dots < \tau_n < \beta} \hat{\mathcal{T}}(\tau_n) \dots \hat{\mathcal{T}}(\tau_{L_2}) \hat{\mathcal{T}}(\tau_{L_1}) \hat{\mathcal{T}}(\tau_L) \hat{\mathcal{G}}(\tau) \\ \times \hat{\mathcal{T}}(\tau_R) \hat{\mathcal{T}}(\tau_{R_1}) \hat{\mathcal{T}}(\tau_{R_2}) \dots \hat{\mathcal{T}}(\tau_1) d\tau_1 \dots d\tau_n \end{aligned} \quad (9)$$

Note that we have used the labels  $L, L_1, L_2, \dots$  (resp.  $R, R_1, R_2, \dots$ ) for the first, second, third, ... time indices

that appear on the left (resp. right) of  $\hat{\mathcal{G}}(\tau)$ . By introducing  $n$  complete sets of states  $\sum_k |k\rangle\langle k|$  with  $k = 1 \dots n$  between each  $\hat{\mathcal{T}}$  and  $\hat{\mathcal{G}}$  operators, and an extra set with  $k = 0$  for the trace, the extended partition function takes the final form

$$\begin{aligned} \mathcal{Z}(\beta, \tau) = & \sum_{n \geq 0} \sum_{\{|k\rangle\}} e^{-\beta V_0} \int_{0 < \tau_1 < \dots < \tau_n < \beta} \langle 0 | \hat{\mathcal{T}}(\tau_n) | n \rangle \dots \\ & \times \langle L_1 | \hat{\mathcal{T}}(\tau_L) | L \rangle \langle L | \hat{\mathcal{G}}(\tau) | R \rangle \langle R | \hat{\mathcal{T}}(\tau_R) | R_1 \rangle \\ & \times \dots \langle 1 | \hat{\mathcal{T}}(\tau_1) | 0 \rangle d\tau_1 \dots d\tau_n, \end{aligned} \quad (10)$$

where we have used the notation  $V_k$  (here with  $k = 0$ ) to denote the matrix element  $\langle k | \hat{\mathcal{V}} | k \rangle$ . As a result, a configuration of the extended partition function (10) is determined by a set of  $n$  time indices and  $n+1$  states  $|k\rangle$ . Since the states  $|k\rangle$  are connected to each other by single  $\hat{\mathcal{T}}$  operators, it is actually simpler to generate the set of  $n+1$  states starting from the two states  $|L\rangle$  and  $|R\rangle$  and a set of  $n$  particular terms  $\hat{\mathcal{T}}_k$  of the operator  $\hat{\mathcal{T}}$ . In the following, a configuration for which  $|L\rangle = |R\rangle$  will be referred to as a *diagonal configuration*, and a particular term  $\hat{\mathcal{T}}_k$  will be called *term with space index  $k$* .

In our example (2), the  $\hat{\mathcal{V}}$  and  $\hat{\mathcal{T}}$  operators are identified as

$$\hat{\mathcal{V}} = U \sum_i \hat{n}_i (\hat{n}_i - 1), \quad (11)$$

$$\begin{aligned} \hat{\mathcal{T}} = & t \sum_{\langle i, j \rangle} (a_i^\dagger a_j + H.c.) \\ & + K \sum_{\square} (a_{\square 1}^\dagger a_{\square 3}^\dagger a_{\square 5}^\dagger a_{\square 6} a_{\square 4} a_{\square 2} + H.c.), \end{aligned} \quad (12)$$

and a term  $\hat{\mathcal{T}}_k$  can be either a kinetic or a ring-exchange term. Fig. 1 shows a possible configuration for the extended partition function associated with this example.

The extended partition function  $\mathcal{Z}(\beta, \tau)$  is a sum of diagonal configurations that belongs to the actual partition function  $\mathcal{Z}(\beta)$ , and non-diagonal configurations. More precisely,

$$\begin{aligned} \mathcal{Z}(\beta, \tau) = & \mathcal{Z}(\beta) \\ & + \sum_{L \neq R} \text{Tr} e^{-(\beta - \tau) \hat{\mathcal{H}}} |L\rangle \langle L | \hat{\mathcal{G}} | R\rangle \langle R | e^{-\tau \hat{\mathcal{H}}}. \end{aligned} \quad (13)$$

The purpose of the algorithm is to evolve from a diagonal configuration to another one, via non-diagonal configurations. The role of the Green operator  $\hat{\mathcal{G}}$  is to allow the transition from a configuration to another one by propagating across the operator string and inserting or removing  $\hat{\mathcal{T}}$  operators while the offset fluctuates. By satisfying detailed balance (See section III), the configurations of the extended partition function (10) can be generated with an *extended Boltzmann weight*. Then all quantities of interest can be estimated using those configurations (See section IV).

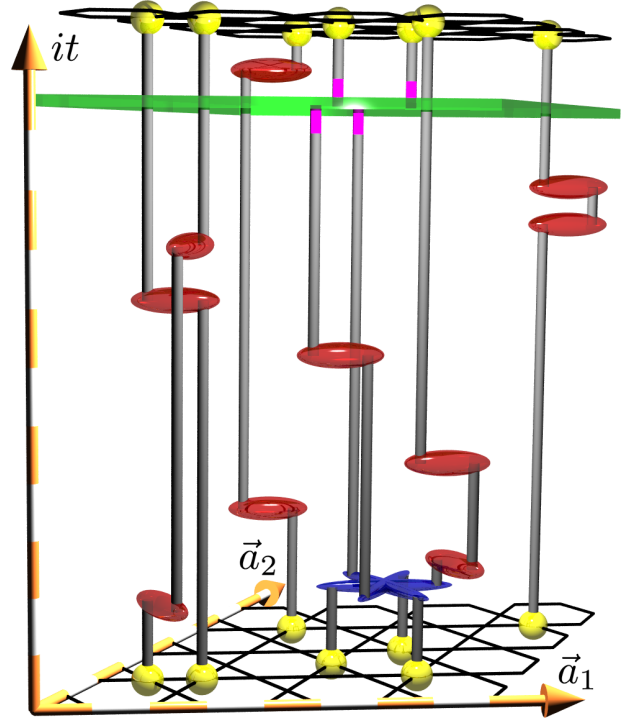


FIG. 1. (Color online) A possible configuration of the extended partition function  $\mathcal{Z}(\beta, \tau)$  for the Hamiltonian (2). The kagome lattice is represented in black with its primary directions  $\vec{a}_1$  and  $\vec{a}_2$  at imaginary times 0 and  $\beta$ . The particles (yellow) represent the current state of the trace (same state at times 0 and  $\beta$ ). Some kinetic (red) and ring-exchange (blue) terms are distributed in both space and time. The Green operator (green) reconnects the broken worldlines (gray). In the present case, the offset (purple) is  $n = 4$ .

### III. GLOBAL SPACE-TIME UPDATE AND DETAILED BALANCE

In previous formulations of the SGF algorithm [13, 14], the updating procedure was affecting only space indices and a maximum of two time indices of the operators of the extended partition function (10). In this section, we propose a new *directed update* procedure, the “global space-time update”, that is able to globally update both space and time indices of the operators of the extended partition function. This represents a new improvement, not only over past versions of the SGF algorithm, but also over other algorithms. For example, in the SSE algorithm a single loop can update only a restricted region of space (Fig. 2), and an additional procedure is necessary to update the time indices of the operators. In the CW algorithm, while a single update is able to affect the entire space, it is unable to update more than two time indices (Fig. 3).

The new global space-time update we propose overcomes those weaknesses by generating a completely new portion of the operator string. In addition the update is directed, that is to say, the Green operator can be prop-

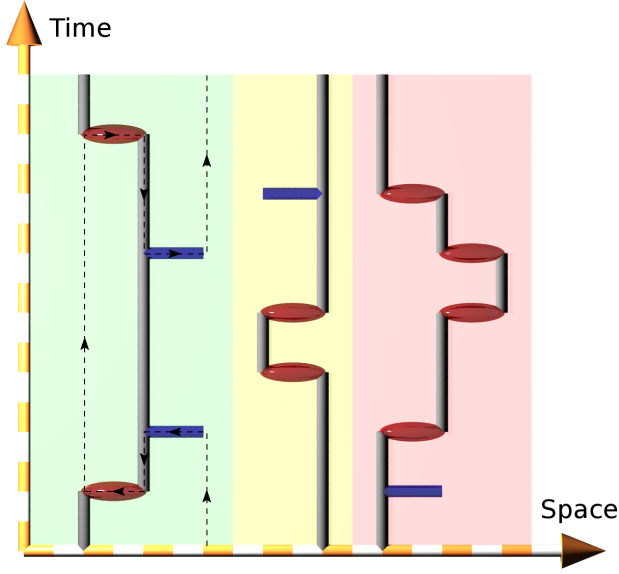


FIG. 2. (Color online) A typical configuration in the SSE representation with a loop update. Diagonal operators (blue) do not affect the worldlines (gray), while non-diagonal operators (red) allow to break the worldlines. In order to update the configuration, a loop (black dotted line) is constructed, the occupation numbers of the sites that are visited are increased or decreased, and the vertices are updated with diagonal or non-diagonal operators. The loop can jump vertically between two vertices, or horizontally across the same vertex. Because horizontal jumps are not permitted between two different vertices, a loop that opens in a given region (cyan, yellow, or pink) must close in the same region. As a result, only a restricted region of space can be updated with a single loop. An additional procedure is necessary in order to update the time indices of the vertices, and redistribute diagonal operators over the space to give the next loop a chance to update a new region of space.

agated in the same direction over a time window  $[\tau_i; \tau_f]$  whose width can be controlled via an optimization parameter. Also, the length of the operator string can fluctuate by any number with a single update. Therefore the auto-correlation time between different configurations is reduced, resulting in a better sampling (See section VII).

### A. Global space-time update scheme

An easy way to generate configurations of the extended partition function consists in starting from an initial configuration where  $\hat{G}$  is the only operator present in the operator string (10), with an arbitrary imaginary time index  $\tau$  and an arbitrary initial state,  $|L\rangle = |R\rangle$ . Then, while propagating to the left (increasing time) or to the right (decreasing time), the Green operator can either drop a  $\hat{T}$  operator behind (creation) or pick up a  $\hat{T}$  operator ahead (destruction). Each creation introduces a new projector  $|\psi\rangle\langle\psi|$ , and selects a single term of  $\hat{T}$ . Each destruction removes a projector  $|\psi\rangle\langle\psi|$  and a sin-

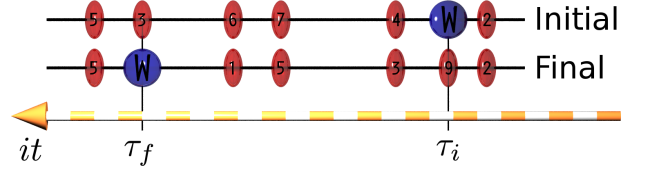


FIG. 3. (Color online) Typical initial and final configurations in the representation of the CW algorithm. Non-diagonal operators (red) appear with given space (black labels) and time indices. In order to update the configuration, a worm operator (blue) is propagated in time and changes the space indices of the operators that are visited. The worm operator can create an operator with a new time index only at the beginning of the propagation, and destroy an operator only at the end. As a result, a single update can affect only a maximum of two time indices, and the length of the operator string can fluctuate only by  $+1, 0, -1$ .

gle term of  $\hat{T}$ . More precisely, assuming a propagation of  $\hat{G}$  to the left, a creation corresponds to the following transition

$$\langle L|\hat{G}|R\rangle \rightarrow \langle L|\hat{G}|\psi\rangle\langle\psi|\hat{T}|R\rangle, \quad (14)$$

and a destruction corresponds to

$$\langle L_1|\hat{T}|L\rangle\langle L|\hat{G}|R\rangle \rightarrow \langle L_1|\hat{G}|R\rangle. \quad (15)$$

The idea of the global space-time update is to give to the Green operator a chance to perform several creations and destructions while propagating in the same direction, and allow both space and time indices of the terms of the  $\hat{T}$  operators to be updated. Thus, while propagating in the same direction, the Green operator can update the system over the whole space in a time window  $[\tau_i; \tau_f]$  whose average width can be controlled. The scheme for this global space-time update is shown in Fig. 4. A direction of propagation to the left or to the right is chosen for the Green operator. Then an action, creation or destruction of a  $\hat{T}$  operator, is chosen. If creation is chosen, then a new imaginary time index  $\tau'$  is chosen for  $\hat{G}$  in the open time window  $]\tau_R; \tau_L[$ , a state for a new projector is chosen, and a  $\hat{T}(\tau')$  operator is created behind  $\hat{G}$ . If destruction is chosen, then a  $\hat{T}$  operator is destroyed ahead of  $\hat{G}$ . After the action has been performed, a decision to loop in the same direction or to stop is made. The algorithm continues to loop until it chooses to stop. Then a new time index is chosen in the new open time window  $]\tau_R; \tau_L[$  for  $\hat{G}$ , and the update is over.

Note that when the Green operator crosses the periodic imaginary time boundaries, the state of the trace is updated. This update ensures the ergodicity of the algorithm, since any permitted term of the  $\hat{T}$  operator can be inserted in the operator string at any imaginary time index. Reciprocally, any term encountered can be removed. Thus any configuration is accessible from a given initial configuration in a finite number of iterations.

The advantage of this global space-time update is that it is able to fully update a controllable portion of the

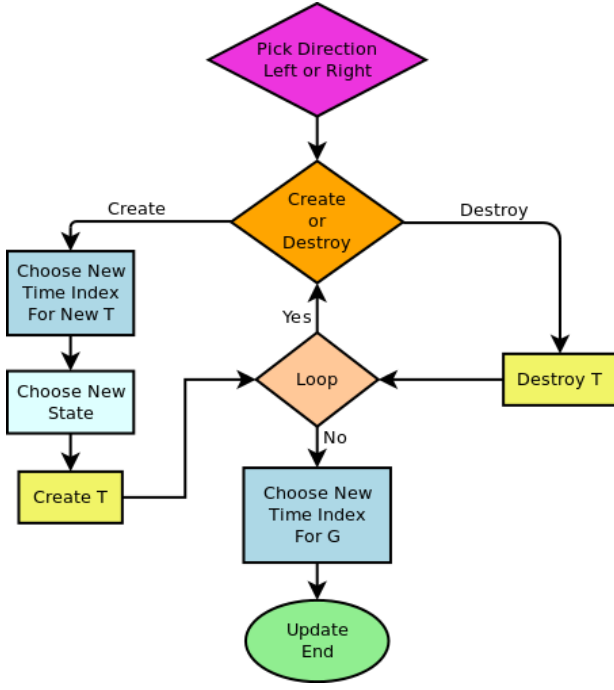


FIG. 4. (Color online) The global space-time update scheme. A  $\hat{T}$  operator can be created with any space and time indices, and any  $\hat{T}$  operator that is encountered can be destroyed. Sequences consisting of several creations and destructions can be performed while propagating in the same direction (see text for details).

operator string. Fig. 5 shows two configurations of the operator string associated with the Hamiltonian (2) that are accessible from each other with a single update.

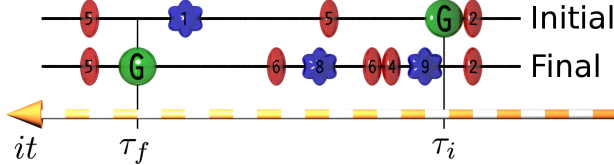


FIG. 5. (Color online) An example of global space-time update for the SGF algorithm in the case of the Hamiltonian (2). Kinetic (red) and ring-exchange (blue) operators can be created or destroyed. The labels on the operators correspond to a particular term (space index). The Green operator (green) is propagated to the left from  $\tau_i$  to  $\tau_f$  and can update both space and time indices of the kinetic and ring-exchange operators that are encountered. In addition, the length of the operator string can fluctuate by any number and the average width of the time window  $[\tau_i; \tau_f]$  is controllable (directed update).

### B. Detailed balance

In order to generate operator strings with the (extended) Boltzmann weight, detailed balance must be sat-

isfied. For this purpose we consider a transition between an initial and a final configuration.

Defining the Boltzmann weight  $P_i$  (and  $P_f$ ) of the initial (and final) configuration, and the probability  $W_{i \rightarrow f}$  (and  $W_{f \rightarrow i}$ ) to make a transition from the initial to the final (and from the final to the initial) configuration, the detailed balance takes the form:

$$P_i W_{i \rightarrow f} = P_f W_{f \rightarrow i} \quad (16)$$

The probabilities  $W_{i \rightarrow f}$  and  $W_{f \rightarrow i}$  to make a transition can be factorized as

$$W_{i \rightarrow f} = S_{i \rightarrow f} A_{i \rightarrow f}, \quad (17)$$

$$W_{f \rightarrow i} = S_{f \rightarrow i} A_{f \rightarrow i}, \quad (18)$$

where  $S_{i \rightarrow f}$  (resp.  $S_{f \rightarrow i}$ ) is the probability to propose the transition from the initial to the final (resp. from the final to the initial) configuration, and  $A_{i \rightarrow f}$  (resp.  $A_{f \rightarrow i}$ ) is the probability to accept the proposed transition. In the following, we make use of the Metropolis solution

$$A_{i \rightarrow f} = \min(1, q_{i \rightarrow f}), \quad (19)$$

$$A_{f \rightarrow i} = \min(1, q_{f \rightarrow i}), \quad (20)$$

where the acceptance factors  $q_{i \rightarrow f}$  and  $q_{f \rightarrow i}$  are given by

$$q_{i \rightarrow f} = \frac{P_f S_{f \rightarrow i}}{P_i S_{i \rightarrow f}}, \quad (21)$$

$$q_{f \rightarrow i} = \frac{P_i S_{i \rightarrow f}}{P_f S_{f \rightarrow i}}. \quad (22)$$

### C. Determination of probabilities

The global space-time update involves several probability functions that need to be determined. The choice of these functions is arbitrary, the only requirements are that ergodicity and detailed balance must be satisfied. Thus we have some freedom that allows us to make a choice that will be convenient and efficient. We consider the following probabilities:

- $P^{LR}(\sigma)$ , with  $\sigma = \leftarrow, \rightarrow$ , the probability to choose a propagation of the Green operator in the  $\sigma$  direction, conditioned by the states and time indices with labels  $L$  and  $R$ .
- $P_{\sigma}^{LR}(\dagger)$ , the probability to choose a creation, conditioned by  $L, R, \sigma$ .
- $P^{LR}(\tau)$ , the probability to choose a new time index for a  $\hat{T}$  operator or for the Green operator, conditioned by  $L, R$ .
- $P_{\sigma}^{LR}(\psi)$ , the probability to choose the state  $\psi$  for a new projector  $|\psi\rangle\langle\psi|$ , conditioned by  $L, R, \sigma$ .
- $P_{\sigma}^{LR}(\odot)$ , the probability to loop, conditioned by  $L, R, \sigma$ .

In order to satisfy detailed balance, we must be able to evaluate the acceptance factor  $q_{\sigma}^{\alpha_1\alpha_2\cdots\alpha_n}$  for any sequence consisting of  $n$  actions of type  $\alpha_i$  in the direction  $\sigma$ , where  $\alpha_i$  is either a creation ( $c$ ) or a destruction ( $d$ ). Our purpose is to make a suitable choice for the above probabilities, such that all acceptance factors associated with all possible sequences in any direction reduce to a single acceptance factor  $q$ .

For this purpose, let us consider a sequence that consists of a propagation to the left with a single creation. The corresponding acceptance factor is  $q_{\leftarrow}^c$ . In the following, we use primed (non-primed) labels for the final (initial) configuration. The Boltzmann weight  $P_i$  of the initial configuration is proportional to

$$P_i \propto \langle L|\hat{G}(\tau)|R\rangle \propto \langle L|\hat{G}|R\rangle e^{\tau(V_L-V_R)}. \quad (23)$$

The hidden matrix elements of the operator string are unchanged by the update. The Boltzmann weight of the final configuration is proportional to

$$P_f \propto \langle L'|\hat{G}(\tau')|R'\rangle \langle R'|\hat{T}(\tau_{R'})|R_1'\rangle \propto \langle L'|\hat{G}|R'\rangle \langle R'|\hat{T}|R_1'\rangle \times e^{\tau'(V_{L'}-V_{R'})} e^{\tau_{R'}(V_{R'}-V_{R_1'})}. \quad (24)$$

The probability  $S_{i \rightarrow f}$  to propose the transition from the initial configuration to the final configuration is the probability to choose a propagation to the left, times the probability to do a creation, times the probability to choose the time index  $\tau_{R'}$  for the new  $\hat{T}$ , times the probability to choose the new state  $|R'\rangle$ , times the probability to stop the update, times the probability to choose the time index  $\tau'$  for  $\hat{G}$ :

$$S_{i \rightarrow f} = P^{LR}(\leftarrow) P_{\leftarrow}^{LR}(\dagger) P^{LR}(\tau_{R'}) P_{\leftarrow}^{LR}(R') \times (1 - P_{\leftarrow}^{L'R'}(\odot)) P^{L'R'}(\tau') \quad (25)$$

The probability to propose the reverse update from the final configuration to the initial configuration is simply the probability to choose a propagation to the right, times the probability to do a destruction, times the probability to stop the update, times the probability to choose the time index  $\tau$  for  $\hat{G}$ :

$$S_{f \rightarrow i} = P^{L'R'}(\rightarrow) (1 - P_{\rightarrow}^{L'R'}(\dagger)) \times (1 - P_{\rightarrow}^{LR}(\odot)) P^{LR}(\tau) \quad (26)$$

Putting everything together, and realizing that  $V_{L'} = V_L$ , and  $V_{R_1'} = V_R$ , the acceptance factor can be written as:

$$q_{\leftarrow}^c = \frac{\langle L|\hat{G}|R'\rangle \langle R'|\hat{T}|R\rangle}{\langle L|\hat{G}|R\rangle P_{\leftarrow}^{LR}(R')} \times \frac{P^{LR}(\tau)}{e^{\tau(V_L-V_R)}} \frac{e^{\tau'(V_{L'}-V_{R'})}}{P^{L'R'}(\tau')} \frac{e^{\tau_{R'}(V_{R'}-V_{R_1'})}}{P^{LR}(\tau_{R'})} \times \frac{P^{L'R'}(\rightarrow) (1 - P_{\rightarrow}^{L'R'}(\dagger)) (1 - P_{\rightarrow}^{LR}(\odot))}{P^{LR}(\leftarrow) P_{\leftarrow}^{LR}(\dagger) (1 - P_{\leftarrow}^{L'R'}(\odot))} \quad (27)$$

We notice that the initial and the final times of the Green operator,  $\tau$  and  $\tau'$ , appear only in the second factor of (27). This suggests that it is possible to make the acceptance factor independent of those times by using an exponential distribution for the time probability,

$$P^{LR}(\tau) = \frac{\Delta V e^{\tau \Delta V}}{e^{\tau_L \Delta V} - e^{\tau_R \Delta V}}, \quad (28)$$

where we have defined  $\Delta V = V_L - V_R$  and  $\Delta \tau = \tau_L - \tau_R$ .

In order to favor the states that are important, we can choose the new states with a probability that is proportional to the weight of the new matrix elements. Thus we use the following distributions:

$$P_{\leftarrow}^{LR}(\psi) = \frac{\langle L|\hat{G}|\psi\rangle \langle \psi|\hat{T}|R\rangle}{\langle L|\hat{G}\hat{T}|R\rangle} \quad (29)$$

$$P_{\rightarrow}^{LR}(\psi) = \frac{\langle L|\hat{T}|\psi\rangle \langle \psi|\hat{G}|R\rangle}{\langle L|\hat{T}\hat{G}|R\rangle} \quad (30)$$

Injecting (28) and (29) in (27), the acceptance factor takes the form

$$q_{\leftarrow}^c = \frac{\langle L|\hat{G}\hat{T}|R\rangle (1 - P_{\rightarrow}^{LR}(\odot))}{\langle L|\hat{G}|R\rangle P^{LR}(\leftarrow) P_{\leftarrow}^{LR}(\dagger)} \times \frac{P^{L'R'}(\rightarrow) (1 - P_{\rightarrow}^{L'R'}(\dagger)) (e^{\Delta \tau' \Delta V'} - 1)}{\Delta V' (1 - P_{\leftarrow}^{L'R'}(\odot))}, \quad (31)$$

and is written as a quantity that depends on the initial configuration, times a quantity that depends on the final configuration. Note that the reverse update corresponds to a propagation to the right with a destruction, hence  $q_{\leftarrow}^d = 1/q_{\leftarrow}^c$ . As a result, the acceptance factor of a left propagation with a destruction,  $q_{\leftarrow}^d$ , is obtained by inverting (31), switching the direction, and exchanging the primed labels with the non-primed ones:

$$q_{\leftarrow}^d = \frac{\Delta V (1 - P_{\rightarrow}^{LR}(\odot))}{P^{LR}(\leftarrow) (1 - P_{\leftarrow}^{LR}(\dagger)) (1 - e^{-\Delta \tau \Delta V})} \times \frac{\langle L'|\hat{G}|R'\rangle P^{L'R'}(\rightarrow) P_{\rightarrow}^{L'R'}(\dagger)}{\langle L'|\hat{T}\hat{G}|R'\rangle (1 - P_{\leftarrow}^{L'R'}(\odot))} \quad (32)$$

For a uniform sampling, we can impose the acceptance factor of a left propagation and creation to be equal to the acceptance factor of a left propagation and destruction,  $q_{\leftarrow}^c = q_{\leftarrow}^d$ . This allows us to determine the probability of creation

$$P_{\leftarrow}^{LR}(\dagger) = \frac{\langle L|\hat{G}\hat{T}|R\rangle}{\langle L|\hat{G}|R\rangle} \frac{1}{f_{\leftarrow}^{LR}} \quad (33)$$

$$P_{\rightarrow}^{LR}(\dagger) = \frac{\langle L|\hat{T}\hat{G}|R\rangle}{\langle L|\hat{G}|R\rangle} \frac{1}{f_{\rightarrow}^{LR}}, \quad (34)$$

where we have defined

$$f_{\leftarrow}^{LR} = \frac{\langle L|\hat{G}\hat{T}|R\rangle}{\langle L|\hat{G}|R\rangle} + \frac{\Delta V}{1 - e^{-\Delta \tau \Delta V}} \quad (35)$$

$$f_{\rightarrow}^{LR} = \frac{\langle L|\hat{T}\hat{G}|R\rangle}{\langle L|\hat{G}|R\rangle} - \frac{\Delta V}{1 - e^{\Delta \tau \Delta V}}. \quad (36)$$

The acceptance factors for a creation or a destruction become:

$$q_{\leftarrow}^c = q_{\leftarrow}^d = \frac{(1 - P_{\rightarrow}^{LR}(\odot))f_{\leftarrow}^{LR}}{P_{\leftarrow}^{LR}(\leftarrow)} \frac{P_{\rightarrow}^{L'R'}(\rightarrow)}{(1 - P_{\leftarrow}^{L'R'}(\odot))f_{\rightarrow}^{L'R'}} \quad (37)$$

Let us consider now a propagation to the left and a sequence of two creations, with a corresponding acceptance factor  $q_{\leftarrow}^{cc}$ . The probability of the initial configuration is given by (23). The probability of the final configuration is:

$$\begin{aligned} P_f &\propto \langle L' | \hat{G}(\tau') | R' \rangle \langle R' | \hat{T}(\tau_{R'}) | R'_1 \rangle \langle R'_1 | \hat{T}(\tau_{R'_1}) | R'_2 \rangle \\ &\propto \langle L' | \hat{G} | R' \rangle \langle R' | \hat{T} | R'_1 \rangle \langle R'_1 | \hat{T} | R'_2 \rangle \\ &\times e^{\tau'(V_{L'} - V_{R'})} e^{\tau_{R'}(V_{R'} - V_{R'_1})} e^{\tau_{R'_1}(V_{R'_1} - V_{R'_2})} \end{aligned} \quad (38)$$

The probability to propose a transition from the initial configuration to the final configuration is the product:

$$\begin{aligned} S_{i \rightarrow f} &= P_{\leftarrow}^{LR}(\leftarrow) P_{\leftarrow}^{LR}(\dagger) P_{\rightarrow}^{LR}(\tau_{R'_1}) P_{\leftarrow}^{LR}(R'_1) \\ &\times P_{\leftarrow}^{LR'_1}(\odot) P_{\leftarrow}^{LR'_1}(\dagger) P_{\rightarrow}^{LR'_1}(\tau_{R'}) P_{\leftarrow}^{LR'_1}(R') \\ &\times (1 - P_{\leftarrow}^{L'R'}(\odot)) P_{\leftarrow}^{L'R'}(\tau') \end{aligned} \quad (39)$$

In the same way, the probability to propose the reverse transition from the final configuration to the initial configuration takes the form:

$$\begin{aligned} S_{f \rightarrow i} &= P_{\rightarrow}^{L'R'}(\rightarrow) (1 - P_{\rightarrow}^{L'R'}(\dagger)) P_{\rightarrow}^{L'R'}(\odot) (1 - P_{\rightarrow}^{L'R'}(\dagger)) \\ &\times (1 - P_{\rightarrow}^{LR}(\odot)) P_{\rightarrow}^{LR}(\tau) \end{aligned} \quad (40)$$

Using our previous definitions and the fact that  $V_{L'} = V_L$  and  $V_{R'_2} = V_R$ , the acceptance factor takes the form:

$$\begin{aligned} q_{\leftarrow}^{cc} &= \frac{(1 - P_{\rightarrow}^{LR}(\odot))f_{\leftarrow}^{LR}}{P_{\leftarrow}^{LR}(\leftarrow)} \\ &\times \frac{P_{\rightarrow}^{LR'_1}(\odot)f_{\leftarrow}^{LR'_1}}{P_{\leftarrow}^{LR'_1}(\odot)f_{\rightarrow}^{LR'_1}} \\ &\times \frac{P_{\rightarrow}^{L'R'}(\rightarrow)}{(1 - P_{\leftarrow}^{L'R'}(\odot))f_{\rightarrow}^{L'R'}} \end{aligned} \quad (41)$$

and is written as a product of quantities that depend on the initial, intermediate, and final configurations, respectively. One can see from the above expression that a suitable choice for  $P_{\sigma}^{LR}(\odot)$  allows us to make the acceptance factor independent of the intermediate configuration. A possible solution is

$$P_{\sigma}^{LR}(\odot) = \alpha \min \left( 1, \frac{f_{\sigma}^{LR}}{f_{\bar{\sigma}}^{LR}} \right), \quad (42)$$

where  $\bar{\sigma}$  is the opposite direction of  $\sigma$ , and  $\alpha$  is an optimization parameter to be chosen in  $[0; 1[$ . With this choice, the acceptance factor of any update becomes totally independent of the sequence of creations and destructions, and reads

$$q_{\sigma} = \frac{P_{\rightarrow}^{L'R'}(\bar{\sigma})Q^{LR}(\sigma)}{P_{\leftarrow}^{LR}(\sigma)Q^{L'R'}(\bar{\sigma})}, \quad (43)$$

with

$$Q^{LR}(\sigma) = f_{\sigma}^{LR} \left( 1 - \alpha \min(1, f_{\bar{\sigma}}^{LR}/f_{\sigma}^{LR}) \right). \quad (44)$$

Finally we can impose the acceptance factor of a propagation to the left to be equal to the acceptance factor of a propagation to the right,  $q_{\leftarrow} = q_{\rightarrow}$ . This is realized if

$$P^{LR}(\sigma) = \frac{Q^{LR}(\sigma)}{Q^{LR}(\sigma) + Q^{LR}(\bar{\sigma})}, \quad (45)$$

and, defining  $Q^{LR} = Q^{LR}(\leftarrow) + Q^{LR}(\rightarrow)$ , we are left with a single acceptance factor for any update:

$$q = \frac{Q^{LR}}{Q^{L'R'}} \quad (46)$$

Because the acceptance factor (46) is written as a ratio of a quantity that depends only on the initial configuration and a quantity that depends only on the final configuration, an ultimate simplification can be done. Using (46) and defining  $Q_i = Q^{LR}$  and  $Q_f = Q^{L'R'}$ , we can rewrite (21) as:

$$\frac{P_f Q_f S_{f \rightarrow i}}{P_i Q_i S_{i \rightarrow f}} = 1 \quad (47)$$

The above equation can be interpreted as follows: Accepting all transitions with a probability of 1 is equivalent to sampling the partition function with the pseudo-Boltzmann weight  $PQ$  instead of the true Boltzmann weight  $P$ . The statistical average of any operator  $\hat{\mathcal{A}}$  can be obtained with a simple renormalization:

$$\langle \hat{\mathcal{A}} \rangle_P = \frac{\langle \hat{\mathcal{A}}/Q \rangle_{PQ}}{\langle 1/Q \rangle_{PQ}} \quad (48)$$

By construction, the function  $Q$  never diverges nor vanishes. Hence the renormalization is well defined in any case. Note that we have explicitly excluded the value 1 from the allowed values for  $\alpha$ . This prevents the probability of looping,  $P_{\sigma}^{LR}(\odot)$ , from being systematically equal to the unity in diagonal configurations, otherwise no measurements would be possible. The advantage of accepting all transitions with a probability of 1 is that no CPU time is wasted with useless rejected updates, and there is no need to record the changes made in the operator string during the update. Also, accepting a “bad” transition from time to time may help the system to escape from a local minimum of the energy.

It is worth to emphasize the importance of interpreting correctly the meaning of  $\Delta\tau = \tau_L - \tau_R$ , especially at high temperature where the system is dominated by configurations with zero or a single  $\hat{T}$  operator.  $\Delta\tau$  corresponds to the time length over which the Green operator can be shifted without encountering a  $\hat{T}$  operator. On the one hand, when there is a single  $\hat{T}$  operator in the string, then  $\tau_L = \tau_R$  and the Green operator is able to move everywhere over the imaginary time axis, except over the



single point  $\tau_L = \tau_R$ . Thus the difference  $\tau_L - \tau_R$  must include the periodicity of the imaginary time axis, leading to the value  $\Delta\tau = \beta$ . On the other hand, when the operator string is empty, there is no limit of the time length over which the Green operator can be shifted. So the value  $\Delta\tau = +\infty$  must be used, which ensures that the probability of creation of a  $\hat{T}$  operator is  $P_\sigma^{LR}(\dagger) = 1$ . In this limit, the probability distribution for the time index of the new  $\hat{T}$  operator or the Green operator becomes defined in the interval  $] -\infty; +\infty[$  with the value  $P^{LR}(\tau) \rightarrow 0$ . Taking into account the periodicity of the imaginary time axis allows us to restrict the distribution over the finite range  $[0; \beta[$  with  $P^{LR}(\tau) = 1/\beta$ . However it is crucial to use  $\Delta\tau = +\infty$  in all other equations where  $\Delta\tau$  is involved, in order to have them correct.

By adjusting the value of  $\alpha$  for the probability to loop, one can tune the “directionality” of the update, that is to say the average length of the sequence of creations and destructions of the update. Having long sequences of creations and destructions reduces the probability for an update to undo the changes made in the configuration with the previous update. This also increases the width of the time window  $[\tau_i, \tau_f]$  of the operator string that is updated, and reduces the auto-correlation time (see section VII).

The choice of the function  $g(n)$  determines how the extended space of configurations is sampled. In section IV, it is shown that it is necessary to generate diagonal configurations in order to make measurements. Thus  $g(n)$  must be a function that decreases with  $n$  sufficiently fast in order to have a chance to generate diagonal configurations. However, non-diagonal configurations are needed in order to update the system. The choice of  $g(n)$  must be done in such a way that all non-diagonal terms have a comparable probability to be introduced in the operator string. In practice, we find that the choice  $g(n) = 1/L^{n/2}$  where  $L$  is the number of lattice sites is a good choice for Hamiltonians for which the highest order of non-diagonal terms is 2, like the Hamiltonian (1). For our non-trivial example (2), a better choice is:

$$g(n) = \begin{cases} 1 & n = 0 \\ 1/L & n = 1, 2, 3, 4 \\ 1/L^{n-3} & n \geq 5 \end{cases} \quad (49)$$

#### D. Summary of probabilities

The total weight of creation and destruction for a particular direction of motion,  $\sigma = \rightarrow, \leftarrow$ , and for particular  $|L\rangle$  and  $|R\rangle$  states, can be parameterized as:

$$W_\sigma = W_\sigma^+ + W_\sigma^-, \quad (50)$$

where  $W_\sigma^+$  is the weight of creation,

$$W_\sigma^+ = \frac{\langle L | \hat{G} \hat{T} | R \rangle}{\langle L | \hat{G} | R \rangle} \quad (51)$$

$$W_\sigma^+ = \frac{\langle L | \hat{T} \hat{G} | R \rangle}{\langle L | \hat{G} | R \rangle} \quad (52)$$

and  $W_\sigma^-$  is the weight of destruction,

$$W_\sigma^- = \frac{s_\sigma \Delta V}{1 - e^{s_\sigma \Delta\tau \Delta V}}, \quad (53)$$

where we introduced the symbol  $s_\sigma$  with  $s_{\leftarrow} = -1$  and  $s_{\rightarrow} = +1$ .

In this notation the absolute creation probability is

$$P_\sigma(+) = \frac{W_\sigma^+}{W_\sigma} \quad (54)$$

Equivalently the destruction probability is

$$P_\sigma(-) = \frac{W_\sigma^-}{W_\sigma} \quad (55)$$

The probability to loop is

$$P_\sigma(\odot) = \alpha \min \left( 1, \frac{W_\sigma}{W_{\bar{\sigma}}} \right). \quad (56)$$

The *relative* probability to pick a particular direction reads

$$Q(\sigma) = W_\sigma (1 - P_{\bar{\sigma}}(\odot)). \quad (57)$$

As shown in Fig. 4, the algorithm starts by picking a particular direction with relative probability  $Q(\sigma)$ . Then it decides if it will destroy or create an operator with relative probabilities  $W_\sigma^+$  and  $W_\sigma^-$  respectively. If creation is chosen the new time index is chosen within the open interval  $] \tau_R, \tau_L[$  with relative probability distribution  $e^{\tau \Delta V}$  and the space index with relative probability given by Eq. (29) and (30). At the end of the creation or destruction the algorithm decides if it wants to continue, with absolute probability  $P_\sigma(\odot)$ , or stop and start over. All the update probabilities are summarized in table I and shown graphically in Fig. 6.

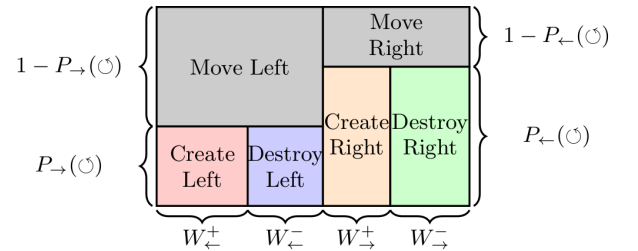


FIG. 6. (color online) Graphical representation of the update probabilities. First the algorithm picks a direction of motion with a relative probability represented by the gray tiles. Their total area is the renormalization weight. Once the direction is fixed it will pick a destruction or creation of an operator with relative probabilities represented by the colored tiles of the corresponding side. The height of those tiles represents the probability to keep moving in the same direction after a creation or destruction.



|                        |  |
|------------------------|--|
| Destruction weight     | $W_{\sigma}^{-} = \frac{s_{\sigma} \Delta V}{1 - e^{s_{\sigma} \Delta \tau \Delta V}}$   |
| Creation weight        | $W_{\leftarrow}^{+} = \frac{\langle L   \hat{G} \hat{\tau}   R \rangle}{\langle L   \hat{G}   R \rangle}$ $W_{\rightarrow}^{+} = \frac{\langle L   \hat{\tau} \hat{G}   R \rangle}{\langle L   \hat{G}   R \rangle}$   |
| Loop probability       | $P_{\sigma}(\odot) = \alpha \min \left( 1, \frac{W_{\sigma}^{-} + W_{\sigma}^{+}}{W_{\sigma}^{-} + W_{\sigma}^{+}} \right)$  |
| Direction weight       | $Q(\sigma) = (W_{\sigma}^{-} + W_{\sigma}^{+}) (1 - P_{\sigma}(\odot))$  |
| Renormalization weight | $Q = Q(\leftarrow) + Q(\rightarrow)$   |
| Time index weight      | $P(\tau) = \frac{\Delta V e^{\tau \Delta V}}{e^{\tau_L \Delta V} - e^{\tau_R \Delta V}}$   |
| Space index weight     | $P_{\leftarrow}^{LR}(\psi) = \frac{\langle L   \hat{G}   \psi \rangle \langle \psi   \hat{\tau}   R \rangle}{\langle L   \hat{G} \hat{\tau}   R \rangle}$ $P_{\rightarrow}^{LR}(\psi) = \frac{\langle L   \hat{\tau}   \psi \rangle \langle \psi   \hat{G}   R \rangle}{\langle L   \hat{\tau} \hat{G}   R \rangle}$ |

TABLE I. Summary of the update probabilities.

#### IV. MEASUREMENTS

##### A. Notation

We will systematically use the notations  $\tilde{\mathcal{A}}(\tau)$  and  $\hat{\mathcal{A}}(\tau)$  for the Heisenberg representation and the interaction representation of an operator  $\hat{\mathcal{A}}$ , respectively:

$$\tilde{\mathcal{A}}(\tau) = e^{\tau \hat{\mathcal{H}}} \hat{\mathcal{A}} e^{-\tau \hat{\mathcal{H}}} \quad (58)$$

$$\hat{\mathcal{A}}(\tau) = e^{\tau \hat{\mathcal{V}}} \hat{\mathcal{A}} e^{-\tau \hat{\mathcal{V}}} \quad (59)$$

It is important to avoid confusion between different types of statistical averages. We use the two following notations and definitions:

$$\langle \hat{\mathcal{A}} \rangle^h = \frac{1}{\mathcal{Z}} \text{Tr} \hat{\mathcal{A}} e^{-\beta \hat{\mathcal{H}}} \quad (60)$$

$$\langle \hat{\mathcal{A}} \rangle^i = \frac{1}{\mathcal{Z}} \text{Tr} e^{-\beta \hat{\mathcal{V}}} \overleftarrow{\mathbf{T}}_{\tau} \left[ \hat{\mathcal{A}} e^{\int_0^{\beta} \hat{\mathcal{T}}(\tau) d\tau} \right] \quad (61)$$

These definitions ensure that

$$\langle \tilde{\mathcal{A}}(\tau) \rangle^h = \langle \hat{\mathcal{A}}(\tau) \rangle^i, \quad (62)$$

and in particular, for an operator  $\hat{\mathcal{A}}$  independent of imaginary time, the two statistical averages are equivalent and the superscript can be omitted,

$$\langle \hat{\mathcal{A}} \rangle = \langle \hat{\mathcal{A}} \rangle^h = \langle \hat{\mathcal{A}} \rangle^i. \quad (63)$$

We show in appendix that, for any product of operators  $\hat{\mathcal{A}}_1, \dots, \hat{\mathcal{A}}_R$  and any set of time indices  $\tau_1, \dots, \tau_R$ , we have:

$$\langle \overleftarrow{\mathbf{T}}_{\tau} \tilde{\mathcal{A}}_R(\tau_R) \dots \tilde{\mathcal{A}}_1(\tau_1) \rangle^h = \langle \hat{\mathcal{A}}_R(\tau_R) \dots \hat{\mathcal{A}}_1(\tau_1) \rangle^i \quad (64)$$

##### B. Measuring quantities with the extended Boltzmann weight

By definition of the extended partition function (6), the states  $|L\rangle$  and  $|R\rangle$  of the Green operator are associated with the *extended* Boltzmann weight:

$$P(L, R) = \frac{\langle L | \hat{\mathcal{G}} | R \rangle \langle R | e^{-\beta \hat{\mathcal{H}}} | L \rangle}{\mathcal{Z}(\beta, \tau)} \quad (65)$$

This extended Boltzmann weight can be used to measure any operator  $\hat{\mathcal{A}}$ . Indeed, consider the expectation value of  $\hat{\mathcal{A}}$ :

$$\begin{aligned} \langle \hat{\mathcal{A}} \rangle &= \frac{\text{Tr} \hat{\mathcal{A}} e^{-\beta \hat{\mathcal{H}}}}{\mathcal{Z}(\beta)} \\ &= \frac{\sum_{\psi} \langle \psi | \hat{\mathcal{A}} e^{-\beta \hat{\mathcal{H}}} | \psi \rangle}{\mathcal{Z}(\beta)} \end{aligned} \quad (66)$$

By introducing a complete set of states, and using the fact that all diagonal matrix elements of  $\hat{\mathcal{G}}$  are equal to 1, Eq. (66) can be rewritten as

$$\begin{aligned} \langle \hat{\mathcal{A}} \rangle &= \frac{\sum_{L,R} \langle L | \hat{\mathcal{A}} | R \rangle \langle R | e^{-\beta \hat{\mathcal{H}}} | L \rangle}{\mathcal{Z}(\beta)} \\ &= \frac{\sum_{L,R} \frac{\langle L | \hat{\mathcal{A}} | R \rangle}{\langle L | \hat{\mathcal{G}} | R \rangle} P(L, R)}{\sum_{L,R} \delta_{L,R} P(L, R)}, \end{aligned} \quad (67)$$

where  $\delta_{L,R}$  is the Kröneckner delta. By performing an importance sampling (denoted  $S_{LR}$ ) over  $S$  samples of states  $|L\rangle$  and  $|R\rangle$  with the distribution  $P(L, R)$ , the expectation value reduces to

$$\begin{aligned} \langle \hat{\mathcal{A}} \rangle &= \lim_{S \rightarrow \infty} \frac{\sum_{S_{LR}} \frac{\langle L | \hat{\mathcal{A}} | R \rangle}{\langle L | \hat{\mathcal{G}} | R \rangle}}{\sum_{S_{LR}} \delta_{L,R}} \\ &= \lim_{S \rightarrow \infty} \frac{1}{N_d} \sum_{S_{LR}} \frac{\langle L | \hat{\mathcal{A}} | R \rangle}{\langle L | \hat{\mathcal{G}} | R \rangle}, \end{aligned} \quad (68)$$

where  $N_d$  is the number of diagonal configurations in the set of samples.

If we make the choice of accepting all updates with a probability of 100%, it is necessary to perform the renormalization (48). The estimator for  $\hat{\mathcal{A}}$  becomes:

$$\langle \hat{\mathcal{A}} \rangle = \lim_{S \rightarrow \infty} \frac{\sum_{S_{LR}} \frac{\langle L | \hat{\mathcal{A}} | Q | R \rangle}{\langle L | \hat{\mathcal{G}} | R \rangle}}{\sum_{S_{\psi}} \langle \psi | 1/Q | \psi \rangle} \quad (69)$$

It is important to note that measurements can be done only at the end of a global space-time update, when the loop is over, that is to say when detailed balance is satisfied.

### C. Quantities represented by diagonal operators

For diagonal operators, only diagonal configurations  $|L\rangle = |R\rangle$  have a non-vanishing contribution, and Eq.(69) reduces to:

$$\langle \hat{\mathcal{A}} \rangle = \lim_{S_\psi \rightarrow \infty} \frac{\sum_{S_\psi} \langle \psi | \hat{\mathcal{A}} | Q | \psi \rangle}{\sum_{S_\psi} \langle \psi | 1 | Q | \psi \rangle} \quad (70)$$

This is the case for the diagonal energy  $\hat{\mathcal{V}}$  (IV F 1), the superfluid density (IV F 2 and IV F 3), imaginary time-dependent density-density correlation functions (IV F 4), and the imaginary dynamical structure factor (IV F 5).

### D. Quantities represented by non-diagonal operators

Non-diagonal operators can be measured “on the fly” with Eq.(69), while exploring the extended space of configurations. The numerator has contributions from non-diagonal configurations  $|L\rangle \neq |R\rangle$ , while the denominator is evaluated in diagonal configurations  $|L\rangle = |R\rangle$ . This is the case for Green functions and the momentum distribution function (IV F 6). The non-diagonal energy  $\hat{\mathcal{T}}$  can also be measured by decomposing it as a sum of Green functions. However it is possible to measure it using diagonal configurations only (IV F 1), which is simpler and more efficient.

### E. Integrals over imaginary time

#### 1. Diagonal operators

Many quantities of interest are defined as integrals of the form

$$\hat{\mathcal{I}} = \left\langle \int_0^\beta \hat{\mathcal{O}}_d(\tau) f(\tau) d\tau \right\rangle^i, \quad (71)$$

where  $\hat{\mathcal{O}}_d$  is a diagonal operator and  $f(\tau)$  an arbitrary function. This is the case, for instance, for the improved estimator of the diagonal energy (see IV F 1) and the Fourier transform of the local density (see IV F 5). Therefore, it is necessary to understand how integrals like (71) can be evaluated exactly within the SGF framework.

For this purpose, consider a configuration  $\mathcal{C}_n$  of the operator string with length  $n$ , imaginary time indices  $\tau_1, \tau_2, \dots, \tau_n$ , and the convention that  $\tau_0 = 0$  and  $\tau_{n+1} = \beta$ . For any time index  $\tau \in [0; \beta[$ , there exists a unique  $k$  such that  $\tau_k \leq \tau < \tau_{k+1}$ . This implies the identity

$$\sum_{k=0}^n \Theta(\tau_k \leq \tau < \tau_{k+1}) = 1 \quad \forall \tau \in [0; \beta[, \quad (72)$$

where  $\Theta(\arg) = 1$  if  $\arg$  is true, 0 otherwise. We denote the value of the integral (71) in this particular configuration by:

$$\left\langle \int_0^\beta \hat{\mathcal{O}}_d(\tau) f(\tau) d\tau \right\rangle_{\mathcal{C}_n}^i \quad (73)$$

This integral receives only contributions from diagonal configurations, where all worldlines are straight between two consecutive time indices,  $\tau_k$  and  $\tau_{k+1}$ . As a result,  $\hat{\mathcal{O}}_d$  has a constant expectation value between  $\tau_k$  and  $\tau_{k+1}$  and, using (72), the integral of  $\hat{\mathcal{O}}_d$  in the configuration  $\mathcal{C}_n$  takes the form

$$\begin{aligned} \left\langle \int_0^\beta \hat{\mathcal{O}}_d(\tau) f(\tau) d\tau \right\rangle_{\mathcal{C}_n}^i &= \sum_{k=0}^n \langle \psi_k^{k+1} | \hat{\mathcal{O}}_d | \psi_k^{k+1} \rangle \\ &\times \int_{\tau_k}^{\tau_{k+1}} f(\tau) d\tau, \end{aligned} \quad (74)$$

where  $\psi_k^{k+1}$  labels the state between the time indices  $\tau_k$  and  $\tau_{k+1}$ .

#### 2. Non-diagonal operators

The integral over the imaginary time axis of a non diagonal operator  $\hat{\mathcal{O}}_{nd}$  can be easily evaluated if  $\hat{\mathcal{O}}_{nd}$  is a particular Hamiltonian term,  $\hat{\mathcal{O}}_{nd} = \hat{\mathcal{T}}_k$ . In this case, for a particular configuration  $\mathcal{C}_n$ , we have

$$\left\langle \int_0^\beta \hat{\mathcal{T}}_k(\tau) d\tau \right\rangle_{\mathcal{C}_n}^i = \frac{n_k}{\beta}, \quad (75)$$

where  $n_k$  is the number of occurrences of  $\hat{\mathcal{T}}_k$  in the operator string. This is a particular case of the theorem presented in next paragraph.

#### 3. Theorem of integration

We propose here a convenient theorem that allows us to easily measure time integrals of multi-point correlation functions. We consider a set of  $M$  arbitrary diagonal operators  $\hat{\mathcal{D}}_k$ , and a set of  $N$  distinct non-diagonal Hamiltonian terms  $\hat{\mathcal{T}}_k$  each being associated with an integer  $p_k$ . For a given operator  $\hat{\mathcal{O}}$ , we define the integrals:

$$\tilde{\mathcal{I}}(\hat{\mathcal{O}}) = \frac{1}{\beta} \int_0^\beta \tilde{\mathcal{O}}(\tau) d\tau \quad (76)$$

$$\hat{\mathcal{I}}(\hat{\mathcal{O}}) = \frac{1}{\beta} \int_0^\beta \hat{\mathcal{O}}(\tau) d\tau \quad (77)$$

Then we have

$$\begin{aligned} &\left\langle \hat{\mathbf{T}}_\tau \left[ \tilde{\mathcal{I}}(\hat{\mathcal{D}}_M) \dots \tilde{\mathcal{I}}(\hat{\mathcal{D}}_1) (\tilde{\mathcal{I}}(\hat{\mathcal{T}}_N))^{p_N} \dots (\tilde{\mathcal{I}}(\hat{\mathcal{T}}_1))^{p_1} \right] \right\rangle^h \\ &= \frac{\left\langle \hat{\mathcal{I}}(\hat{\mathcal{D}}_M) \dots \hat{\mathcal{I}}(\hat{\mathcal{D}}_1) \frac{n_N!}{(n_N - p_N)!} \dots \frac{n_1!}{(n_1 - p_1)!} \right\rangle^i}{\beta^{p_N + \dots + p_1}}, \end{aligned} \quad (78)$$

where  $n_k$  is the number of  $\hat{T}_k$  terms in the configuration. Each integral  $\hat{\mathcal{I}}(\hat{\mathcal{D}}_k)$  in the RHS of (78) is given by (74).

This theorem states that it is equivalent to measure a time-ordered product of operators in the Heisenberg representation and to measure the same product of operators in the interaction representation. The simplification provided by the theorem is that the measurement of the time integral of a non-diagonal Hamiltonian term reduces to counting the number of occurrences of that term in the operator string. More precisely, the factor  $\frac{n_k!}{(n_k - p_k)!}$  corresponds the number of distinct possibilities to select  $p_k$  operators of type  $\hat{T}_k$  in the operator string. The exponent of  $\beta$  in the denominator is the total number of integrals of non-diagonal Hamiltonian terms.

The proof of this theorem is given in appendix, and a direct application of it allows us to measure the energy of the system (see IV F 1) and the specific-heat (see IV F 7).

## F. Particular measurements

### 1. The energy

The diagonal energy  $\hat{\mathcal{V}}$  can be measured directly in diagonal configurations by using (70). However, a better estimator can be constructed by taking advantage of the invariance by imaginary time translation, and integrating over the full imaginary time axis. Using the theorem (78) with  $M = 1$ ,  $N = 0$ , and  $\hat{\mathcal{D}} = \hat{\mathcal{V}}$ , we have:

$$\begin{aligned} \langle \hat{\mathcal{V}} \rangle &= \frac{1}{\beta} \langle \overleftarrow{\mathbf{T}}_\tau \int_0^\beta \hat{\mathcal{V}}(\tau) d\tau \rangle^h \\ &= \frac{1}{\beta} \langle \int_0^\beta \hat{\mathcal{V}}(\tau) d\tau \rangle^i \end{aligned} \quad (79)$$

The integral in (79) can be evaluated using (74) with  $\hat{\mathcal{O}}_d = \hat{\mathcal{V}}$  and  $f(\tau) = 1$ .

The non-diagonal energy  $\hat{\mathcal{T}}$  can be measured easily by averaging the length of the operator string. More generally, the energy associated with any non-diagonal Hamiltonian term can be measured by counting the number of occurrences of that term in the operator string with diagonal configurations. Indeed, consider the decomposition:

$$\hat{\mathcal{T}} = \sum_k \hat{\mathcal{T}}_k \quad (80)$$

Using (78) with  $M = 0$  and  $N = 1$ , the expectation value of a particular term  $\hat{\mathcal{T}}_k$  is given by:

$$\begin{aligned} \langle \hat{\mathcal{T}}_k \rangle &= \frac{1}{\beta} \langle \overleftarrow{\mathbf{T}}_\tau \int_0^\beta \hat{\mathcal{T}}_k(\tau) d\tau \rangle^h \\ &= \frac{1}{\beta} \langle n_k \rangle^i \end{aligned} \quad (81)$$

In particular, the non-diagonal energy  $\hat{\mathcal{T}}$  is given by averaging the total length  $n$  of the operator string:

$$\langle \hat{\mathcal{T}} \rangle = \frac{1}{\beta} \langle n \rangle^i \quad (82)$$

### 2. The superfluid density at finite temperature

The superfluid density can be easily measured via the winding number. For a  $d$ -dimensional system, the winding number is a vectorial operator,  $\hat{W}$ , whose components measure the number of times that the worldlines cross the boundaries of the system in the primary directions of the lattice. Considering a configuration  $\mathcal{C}_n$  of the operator string with length  $n$  and imaginary time indices  $\tau_1, \tau_2, \dots, \tau_n$ , we define the associated *pseudo-current* in the  $\zeta$  direction at time  $\tau$  as

$$\hat{j}_\zeta(\tau) = \sum_{k=1}^n \Delta_k^\zeta \delta(\tau - \tau_k), \quad (83)$$

where  $\Delta_k^\zeta$  measures the discontinuity of the worldlines in the  $\zeta$  direction introduced by the  $\hat{\mathcal{T}}$  operator acting at time  $\tau_k$ . In our example (2), a kinetic term acting in the  $\zeta$  direction gives  $\Delta_k^\zeta = \pm 1$ , while a ring-exchange term systematically gives  $\Delta_k^\zeta = 0$ . By definition, the winding in the  $\zeta$  direction is obtained by integrating the pseudo-current

$$\begin{aligned} \hat{W}_\zeta &= \frac{1}{L_\zeta} \int_0^\beta \hat{j}_\zeta(\tau) d\tau \\ &= \frac{1}{L_\zeta} \sum_{k=1}^n \Delta_k^\zeta, \end{aligned} \quad (84)$$

where  $L_\zeta$  is the number of sites in the  $\zeta$  direction. Thus the winding can be easily measured in the configuration  $\mathcal{C}_n$  by using Eq.(84). For identical particles in a cubic lattice with  $L^d$  sites, it has been shown [22] that the dimensionless superfluid density is given by

$$\rho_s = \frac{L^{2-d} \langle \hat{W}^2 \rangle}{2dt\beta}, \quad (85)$$

where  $t = \frac{\hbar^2}{2ma^2}$ ,  $m$  is the mass of one particle, and  $a$  is the lattice constant.

### 3. Improved estimator for the zero-temperature superfluid density

As we have seen above, the superfluid density  $\rho_s$  can be measured via the winding number  $\hat{W}$ . However  $\rho_s$  can show a strong dependence in temperature, which makes it difficult to estimate its zero-temperature value. Measuring the zero-temperature superfluid density requires, in principle, to perform simulations with increasing values of the inverse temperature  $\beta$ , which is computationally expensive, and then perform an extrapolation to  $\beta \rightarrow \infty$ . We propose here an improved estimator that has a faster convergence to the zero-temperature superfluid density, thus making simulations easier. This improved estimator has been proposed by Batrouni and Scalettar for the discrete time Worldline algorithm [8]. The generalization to

continuous time is straightforward using our continuous-time definition of the pseudo current (83). The improved estimator is actually for the winding number, and we determine the superfluid density using (85).

To this end, we define the Fourier transform of the pseudo-current

$$\begin{aligned}\dot{j}_\zeta(\omega_p) &= \int_0^\beta \dot{j}_\zeta(\tau) e^{-i\omega_p \tau} d\tau \\ &= \sum_{k=1}^n \Delta_k^\zeta e^{-i\omega_p \tau_k},\end{aligned}\quad (86)$$

with  $\omega_p = \frac{2\pi p}{\beta}$ . One notices that  $\dot{W}_\zeta = \dot{j}_\zeta(0)/L_\zeta$ . The idea of the improved estimator  $\dot{W}_\zeta^i$  is based on the fact that the discrete frequencies  $\omega_p$  become continuous in the limit  $\beta \rightarrow \infty$ , and the “numerical observation” that the values of  $|\dot{j}_\zeta(\omega_p)|^2$  for  $p \geq 1$  scale linearly with  $\beta$  at finite but low temperature. As a result, the value of  $|\dot{j}_\zeta(0)|^2$  at zero-temperature can be estimated at finite temperature by performing a linear extrapolation.

Therefore we define the improved estimator for the zero-temperature winding number as:

$$W_\zeta^{i2} = \frac{1}{L_\zeta^2} \left( 2|\dot{j}_\zeta(\omega_1)|^2 - |\dot{j}_\zeta(\omega_2)|^2 \right) \quad (87)$$

Since both  $\omega_1$  and  $\omega_2$  vanish in the limit  $\beta \rightarrow \infty$ , we have  $\lim_{\beta \rightarrow \infty} \dot{W}_\zeta^i = \dot{W}_\zeta$ . It turns out that  $\langle \dot{W}_\zeta^{i2} \rangle^i$  shows a quasi-linear dependence in  $\beta$  at low temperature. Thus, when injected into (85), the dependence in  $\beta$  is canceled and the measured superfluid density becomes independent of the temperature.

Figure 7 shows the finite-temperature superfluid density as a function of  $\beta$  and the improved estimator for the zero-temperature superfluid density for the extended Bose-Hubbard model (1). As  $\beta$  increases, the value given by the improved estimator converges to the zero-temperature limit much faster than the finite-temperature superfluid density.

#### 4. Imaginary time-dependent density-density correlation function

Quantities like  $\langle \hat{\mathbf{T}}_\tau \tilde{n}_{\vec{r}'}(\tau') \tilde{n}_{\vec{r}}(\tau) \rangle^h = \langle \hat{n}_{\vec{r}'}(\tau') \hat{n}_{\vec{r}}(\tau) \rangle^i$  can be measured directly in diagonal configurations, since a given configuration  $\mathcal{C}_n$  fully determines the values of  $\hat{n}_{\vec{r}'}(\tau')$  and  $\hat{n}_{\vec{r}}(\tau)$ . However, in order to reduce the statistical fluctuations, we might want to take advantage of the invariance by imaginary time translation and perform an average over the imaginary time axis. Also, if the Hamiltonian is invariant by space translation, the correlation function will depend only on  $\vec{r}' - \vec{r}$ , and fluctuations can be reduced further by summing over the space.

As a result we consider the correlation function

$$C_{\vec{r}}(\tau) = \frac{1}{L\beta} \sum_{\vec{r}'} \int_0^\beta \langle \hat{n}_{\vec{r}+\vec{r}'}(\tau + \tau') \hat{n}_{\vec{r}'}(\tau') \rangle^i d\tau', \quad (88)$$

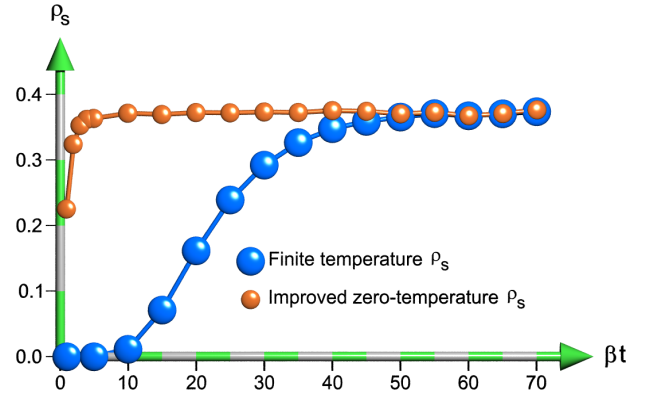


FIG. 7. (Color online) The finite-temperature superfluid density and the improved estimator of the zero-temperature superfluid density as functions of the inverse temperature  $\beta$ , for the extended Bose-Hubbard model (1). The lattice size is  $L = 100$ , the onsite repulsion  $U = 14t$ , the nearest neighbor interaction  $V = 0$ , and the number of particles  $N = 50$ . As  $\beta$  increases, the value given by the improved estimator converges to the zero-temperature limit much faster than the finite-temperature superfluid density.

which depends only on the translation  $\vec{r}$  and time  $\tau$  between the two correlated points. By introducing two identities (72), we get

$$\begin{aligned}C_{\vec{r}}(\tau) &= \frac{1}{L\beta} \sum_{\vec{r}'} \sum_{k=0}^n \sum_{q=0}^n \langle \hat{n}_{\vec{r}+\vec{r}'}(\tau_k) \hat{n}_{\vec{r}'}(\tau_q) \rangle_{\mathcal{C}_n}^i \\ &\times [(M_1 - m_1)\Theta(m_1 < M_1) + (M_2 - m_2)\Theta(m_2 < M_2)],\end{aligned}\quad (89)$$

with

$$\begin{cases} m_1 = \max(\tau_k, \tau_q - \tau) \\ m_2 = \max(\tau_k, \tau_q + \beta - \tau) \\ M_1 = \min(\tau_{k+1}, \tau_{q+1} - \tau) \\ M_2 = \min(\tau_{k+1}, \tau_{q+1} + \beta - \tau) \end{cases} \quad (90)$$

Again, Eq. (89) can be directly evaluated for any given configuration  $\mathcal{C}_n$ . However it should be pointed that evaluating Eq. (89) for a single vector  $\vec{r}$  and time  $\tau$  is a process that has a  $L\beta^2$  complexity. If one is interested in getting full information on density-density correlations, it must be evaluated for all possible vectors  $\vec{r}$  and a number of time samples that is proportional to  $\beta$ . Thus the total complexity of the measurement process becomes  $L^2\beta^3$ . We show in subsection IV F 5 that it is actually possible to obtain full information on density-density correlations with a  $L^2\beta + L\beta^2$  complexity by first evaluating correlations in the Fourier space, then performing an invert Fourier transform.

#### 5. Imaginary dynamical structure factor

We define the imaginary dynamical structure factor as the space-time Fourier transform of the imaginary time-

dependent density-density correlation function (88),

$$S(\vec{k}, \omega_p) = \frac{1}{L\beta} \sum_{\vec{r}} \int_0^\beta C_{\vec{r}}(\tau) e^{i(\vec{k} \cdot \vec{r} - \omega_p \tau)} d\tau \quad (91)$$

with  $\omega_p = \frac{2\pi p}{\beta}$ . As discussed in subsection IV F 4, evaluating directly (88) for all vectors  $\vec{r}$  and time  $\tau$  has a  $L^2\beta^3$  complexity, thus evaluating (91) for all vectors  $\vec{k}$  and frequencies  $\omega_p$  may appear to be very difficult. Actually, this can be done with a  $L^2\beta + L\beta^2$  complexity by working directly in the  $\omega$ -space. For this purpose, we consider the Fourier transform of the imaginary time-dependent local density operator  $\hat{n}_{\vec{r}}(\tau)$ , which can be directly evaluated in a configuration  $\mathcal{C}_n$  by using (74) with  $\hat{\mathcal{O}}_d = \hat{n}_{\vec{r}}$  and  $f(\tau) = e^{-i\omega_p \tau}$ ,

$$\begin{aligned} \langle \hat{n}_{\vec{r}}(\omega_p) \rangle_{\mathcal{C}_n}^i &= \frac{1}{\beta} \langle \int_0^\beta \hat{n}_{\vec{r}}(\tau) e^{-i\omega_p \tau} d\tau \rangle_{\mathcal{C}_n}^i \\ &= \frac{1}{\beta} \sum_{k=0}^n \langle \psi_k^{k+1} | \hat{n}_{\vec{r}} | \psi_k^{k+1} \rangle \\ &\quad \times \frac{e^{-i\omega_p \tau_k} - e^{-i\omega_p \tau_{k+1}}}{i\omega_p}. \end{aligned} \quad (92)$$

Evaluating (92) for all vectors  $\vec{r}$  and frequencies  $\omega_p$  is a process that has a  $L\beta^2$  complexity. Consider now the time Fourier transform of (88):

$$\begin{aligned} \hat{C}_{\vec{r}}(\omega_p) &= \frac{1}{\beta} \int_0^\beta C_{\vec{r}}(\tau) e^{-i\omega_p \tau} d\tau \\ &= \frac{1}{L} \sum_{\vec{r}'} \langle \hat{n}_{\vec{r}}^*(\omega_p) \hat{n}_{\vec{r}+\vec{r}'}(\omega_p) \rangle^i \end{aligned} \quad (93)$$

Knowing  $\hat{n}_{\vec{r}}(\omega_p)$ , we can obtain  $\hat{C}_{\vec{r}}(\omega_p)$  for all  $\vec{r}$  and  $\omega_p$  with an additional  $L^2\beta$  complexity. Finally, the imaginary dynamical structure factor can be obtained by performing a space Fourier transform,

$$S(\vec{k}, \omega_p) = \frac{1}{L} \sum_{\vec{r}} \hat{C}_{\vec{r}}(\omega_p) e^{i\vec{k} \cdot \vec{r}}, \quad (94)$$

which again has a  $L^2\beta$  complexity. So the total complexity for measuring the dynamical structure factor is  $L\beta^2 + L^2\beta$ . In addition, the imaginary time-dependent density-density correlation function (88) can be obtained from (93) for all vectors  $\vec{r}$  and an arbitrary set of time indices by performing an invert time Fourier transform,

$$C_{\vec{r}}(\tau) = \sum_p \hat{C}_{\vec{r}}(\omega_p) e^{i\omega_p \tau}, \quad (95)$$

again with a total complexity  $L\beta^2 + L^2\beta$ .

## 6. Green functions and the momentum distribution function

Green functions can be measured directly by using Eq.(69). For example, the 4-point Green function  $G_{ijkl} = a_i^\dagger a_j^\dagger a_k a_l$  receives contributions from non-diagonal configurations for which the matrix element

$\langle L | G_{ijkl} / Q | R \rangle$  is non-zero. In the contributing configurations, the denominator  $\langle L | \hat{\mathcal{G}} | R \rangle$  is equal to  $g(4)$ , so the choice of the function  $g(n)$  must be suitable for having a good frequency of measurement and a value of  $g(4)$  that is not too small, in order to avoid strong fluctuations. However  $g(4)$  should not be too big, in order to have a good chance of generating diagonal configurations that are needed for the normalization of the measurements.

The momentum distribution function, defined as  $\hat{n}(\vec{k}) = \langle \hat{a}_{\vec{k}}^\dagger \hat{a}_{\vec{k}} \rangle$ , measures the average number of particles with momentum  $\hbar\vec{k}$ . For a  $d$ -dimensional lattice with  $L$  sites, the creation and annihilation operators  $\hat{a}_{\vec{k}}^\dagger$  and  $\hat{a}_{\vec{k}}$  of a particle with momentum  $\vec{k}$  are defined as the Fourier transforms

$$\hat{a}_{\vec{k}}^\dagger = \frac{1}{\sqrt{L}} \sum_p \hat{a}_p^\dagger e^{i\vec{k} \cdot \vec{r}_p}, \quad (96)$$

$$\hat{a}_{\vec{k}} = \frac{1}{\sqrt{L}} \sum_q \hat{a}_q e^{-i\vec{k} \cdot \vec{r}_q}, \quad (97)$$

where  $\hat{a}_p^\dagger$  and  $\hat{a}_q$  are the creation and annihilation operators of a particle on sites  $p$  and  $q$ , respectively. The momentum distribution function takes the form

$$\hat{n}(\vec{k}) = \frac{1}{L} \sum_{p,q} \langle \hat{a}_p^\dagger \hat{a}_q \rangle e^{i\vec{k} \cdot (\vec{r}_p - \vec{r}_q)}, \quad (98)$$

and can be obtained from 2-point Green functions, or the so-called *one-particle density matrix*,  $\rho_{pq} = \langle \hat{a}_p^\dagger \hat{a}_q \rangle$ .

## 7. The specific heat

The specific heat  $C_v$  is defined as the rate of change of the energy  $E = \langle \hat{\mathcal{H}} \rangle$  with temperature  $T$ ,  $C_v = \left. \frac{\partial E}{\partial T} \right|_V$ , keeping the volume (number of sites) constant. A simple way to approximate  $C_v$  consists in performing a numerical symmetric derivative of the energy,  $C_v \approx (E(T+\delta T) - E(T-\delta T)) / 2\delta T$ . This method works well, but requires to perform two simulations at temperatures  $T-\delta T$  and  $T+\delta T$  in addition to the simulation at the temperature of interest  $T$ , and several attempts are usually needed in order to determine the “best” value for  $\delta T$ .

It is actually possible to evaluate  $C_v$  exactly with a single simulation, via the fluctuation-response theorem,

$$C_v = \frac{\partial \langle \hat{\mathcal{H}} \rangle}{\partial T} = \beta^2 (\langle \hat{\mathcal{H}}^2 \rangle - \langle \hat{\mathcal{H}} \rangle^2). \quad (99)$$

In principle all terms in Eq.(99) can be measured as described in this section. However this requires to build the list of all terms present in  $\hat{\mathcal{H}}^2$ , whose number scales as  $L^2$  for Hamiltonians with finite-range interactions ( $L^4$  otherwise). In addition to being computationally expensive, the sum of these terms may suffer from strong fluctuations. We show below that it is actually possible to evaluate  $\langle \hat{\mathcal{H}}^2 \rangle$  globally by performing diagonal measurements only.

Since  $\hat{\mathcal{H}}$  commutes with its exponential  $e^{-\beta\hat{\mathcal{H}}}$ , we can write

$$\begin{aligned}\langle \hat{\mathcal{H}}^2 \rangle &= \langle \hat{\mathbf{T}}_\tau (\tilde{\mathcal{I}}(\hat{\mathcal{H}}))^2 \rangle^h \\ &= \langle \hat{\mathbf{T}}_\tau (\tilde{\mathcal{I}}(\hat{\mathcal{V}}))^2 \rangle^h + \sum_{k_1, k_2} \langle \hat{\mathbf{T}}_\tau [\tilde{\mathcal{I}}(\hat{\mathcal{T}}_{k_1}) \tilde{\mathcal{I}}(\hat{\mathcal{T}}_{k_2})] \rangle^h \\ &\quad - 2 \sum_k \langle \hat{\mathbf{T}}_\tau [\tilde{\mathcal{I}}(\hat{\mathcal{V}}) \tilde{\mathcal{I}}(\hat{\mathcal{T}}_k)] \rangle^h\end{aligned}\quad (100)$$

By applying the theorem (78) for each term in (100), it follows that the specific heat can be measured as

$$\begin{aligned}C_v &= \langle n(n-1) \rangle^i + \left\langle \left( \int_0^\beta \hat{\mathcal{V}}(\tau) d\tau \right)^2 \right\rangle^i \\ &\quad - 2 \left\langle n \int_0^\beta \hat{\mathcal{V}}(\tau) d\tau \right\rangle^i \\ &\quad - \left\langle \left( \int_0^\beta \hat{\mathcal{V}}(\tau) d\tau - n \right)^2 \right\rangle^i,\end{aligned}\quad (101)$$

where  $n$  is the total length of the operator string and the integrals are computed using (74) with  $\hat{\mathcal{O}}_d = \hat{\mathcal{V}}$  and  $f(\tau) = 1$ .

#### 8. The Entropy and the thermal susceptibility

The statistical Von Neumann entropy is given by  $\mathcal{S} = k(\ln \mathcal{Z} + \beta \langle \hat{\mathcal{H}} \rangle)$ , where  $k$  is the Boltzmann constant. Because the actual value of the partition function  $\mathcal{Z}$  is unknown in QMC simulations, a direct measurement of  $\mathcal{S}$  is not possible. In order to overcome this problem,  $\mathcal{S}$  is usually evaluated by performing a numerical integration of the specific heat  $C_v$  over the temperature  $T$ ,

$$\mathcal{S}(T) = \int_0^T \frac{C_v}{T'} dT', \quad (102)$$

which requires a set of simulations at different temperatures.

We propose here an alternative which consists in performing a set of simulations in the grand-canonical ensemble (see section V) at different values of the chemical potential. For simplicity, we consider here a Hamiltonian  $\hat{\mathcal{H}}$  that describes identical particles, for example Eq.(1) or Eq.(2). The grand-canonical partition function takes the form

$$\mathcal{Z} = \text{Tr} e^{-\beta(\hat{\mathcal{H}} - \mu \hat{\mathcal{N}})}, \quad (103)$$

where  $\hat{\mathcal{N}}$  is the operator that measures the total number of particles, and  $\mu$  is the chemical potential. Our thermodynamic control parameters are the temperature  $T$ , the volume  $V$  (number of sites  $L$ ), and the chemical potential  $\mu$ . We define the *thermal susceptibility*  $\chi_{th}$  as the rate of change of the average number of particles  $N = \langle \hat{\mathcal{N}} \rangle$  with temperature  $T$ :

$$\chi_{th}(T, V, \mu) = \left. \frac{\partial N}{\partial T} \right|_{V, \mu} \quad (104)$$

By substituting  $N = \frac{1}{\mathcal{Z}} \text{Tr} \hat{\mathcal{N}} e^{-\beta(\hat{\mathcal{H}} - \mu \hat{\mathcal{N}})}$  in Eq. (104), and assuming that  $[\hat{\mathcal{H}}, \hat{\mathcal{N}}] = 0$ , we get an expression for the thermal susceptibility that can be directly measured,

$$\chi_{th} = \beta^2 [\langle \hat{\mathcal{N}}(\hat{\mathcal{H}} - \mu \hat{\mathcal{N}}) \rangle - \langle \hat{\mathcal{N}} \rangle \langle \hat{\mathcal{H}} - \mu \hat{\mathcal{N}} \rangle], \quad (105)$$

in a way similar to the energy as explained in IV F 1. Considering the energy  $E = \langle \hat{\mathcal{H}} \rangle$  and the associated differential  $dE = Td\mathcal{S} - PdV + \mu dN$ , where the pressure is defined as  $P = -\frac{\partial E}{\partial V}|_{\mathcal{S}, N}$ , and performing a Legendre transformation over the variables  $\mathcal{S}$  and  $N$ , we can define the grand-canonical potential  $\Omega$  that depends only on our natural variables,  $\Omega(T, V, \mu) = E - T\mathcal{S} - \mu N = -PV$ . Its differential takes the form

$$d\Omega = -\mathcal{S}dT - PdV - Nd\mu. \quad (106)$$

We can then extract a useful Maxwell relation,

$$\left. \frac{\partial \mathcal{S}}{\partial \mu} \right|_{V, T} = \left. \frac{\partial N}{\partial T} \right|_{V, \mu}, \quad (107)$$

so the entropy can be easily obtained by integrating the thermal susceptibility over the chemical potential and keeping the temperature and the volume constant,

$$\mathcal{S}(T, V, \mu) = \int_{\mu_0}^{\mu} \chi_{th}(T, V, \mu') d\mu', \quad (108)$$

where  $\mu_0$  is the critical value of the chemical potential below which the average number of particles  $N$  and the thermal susceptibility  $\chi_{th}$  are vanishing. Recently, this method has been applied successfully to Hamiltonians describing diagonal and off-diagonal confinement [23].

### V. SIMULATION OF THE GRAND-CANONICAL ENSEMBLE

Consider a general Hamiltonian  $\hat{\mathcal{H}}$  that describes  $S$  species of particles, and let  $\hat{N}_s$  be the number of particles of a given species  $s$  in an initial state. If it is possible to define  $P$  charges  $\hat{\mathcal{N}}_p$  that are conserved by the Hamiltonian,

$$\begin{cases} \hat{\mathcal{N}}_1 = \gamma_{1,1}\hat{N}_1 + \gamma_{1,2}\hat{N}_2 + \cdots + \gamma_{1,S}\hat{N}_S \\ \hat{\mathcal{N}}_2 = \gamma_{2,1}\hat{N}_1 + \gamma_{2,2}\hat{N}_2 + \cdots + \gamma_{2,S}\hat{N}_S \\ \vdots \\ \hat{\mathcal{N}}_P = \gamma_{P,1}\hat{N}_1 + \gamma_{P,2}\hat{N}_2 + \cdots + \gamma_{P,S}\hat{N}_S, \end{cases} \quad (109)$$

where  $\gamma_{i,j}$  are integers, then the Hamiltonian commutes with  $\hat{\mathcal{N}}_1, \hat{\mathcal{N}}_2, \dots, \hat{\mathcal{N}}_P$ , and the *canonical ensemble* is defined as the set of all states that contain exactly the same number of charges as the initial state.

By nature the SGF algorithm samples the canonical ensemble, since all states are generated from an initial state by successive applications of  $\hat{\mathcal{T}}$  operators. Simulating the canonical ensemble can be convenient, especially

for systems with several species [15, 16]. However it is sometimes useful to work in the *grand-canonical ensemble* [21], that is to say in the ensemble of states that contain any number of particles, especially for magnetic systems. Thus, the best solution is to have an algorithm that can simulate both ensembles. We describe below a simple extension that allows the SGF algorithm to simulate exactly the grand-canonical ensemble.

The idea is to add a non-conservative part  $\hat{\mathcal{H}}_{nc}$  to the Hamiltonian,

$$\hat{\mathcal{H}}_{nc} = \lambda \sum_{j,s} (a_{js}^\dagger + a_{js}), \quad (110)$$

where  $j$  runs over all lattice sites and  $s$  runs over all species, and  $\lambda$  is an optimization parameter. This non-conservative part allows the number of particles to fluctuate, so simulating the Hamiltonian  $\hat{\mathcal{H}}' = \hat{\mathcal{H}} + \hat{\mathcal{H}}_{nc}$  will sample the grand-canonical ensemble. In order to make measurements that correspond to the actual Hamiltonian  $\hat{\mathcal{H}}$ , we can perform a *restricted* simulation of  $\hat{\mathcal{H}}'$  by applying the following conditions:

1. We allow at most one term of  $\hat{\mathcal{H}}_{nc}$  at a time in the operator string.
2. We make measurements only if the operator string contains no  $\hat{\mathcal{H}}_{nc}$  terms.

The first condition is not required, but it allows to increase the probability to make a measurement. The second condition ensures that the actual Hamiltonian  $\hat{\mathcal{H}}$  is exactly simulated in the grand-canonical ensemble. Indeed, the extended partition function  $\mathcal{Z}'(\beta, \tau)$  associated with  $\hat{\mathcal{H}}'$  is a sum over configurations that contain any number of  $\hat{\mathcal{H}}_{nc}$  terms. Ignoring configurations with  $\hat{\mathcal{H}}_{nc}$  terms is equivalent to performing a renormalization of  $\mathcal{Z}'(\beta, \tau)$ , such that the resulting extended partition function  $\mathcal{Z}(\beta, \tau)$  is only the sum of configurations that do not contain  $\hat{\mathcal{H}}_{nc}$  terms. As a result, configurations with no  $\hat{\mathcal{H}}_{nc}$  terms are generated with the correct Boltzmann weight, and the grand-canonical ensemble associated with the Hamiltonian  $\hat{\mathcal{H}}$  is simulated exactly.

The value of  $\lambda$  can be adjusted in order to tune the proportion of unphysical configurations with  $\hat{\mathcal{H}}_{nc}$  terms.  $\lambda$  should be large enough to allow a fast decorrelation of the number of particles between different configurations. But using a value too big reduces the probability of having a physical configuration and making a measurement. In practice we find that a good choice is  $\lambda = 1/L$ , where  $L$  is the number of lattice sites.

Finally, the average number of particles  $\langle \hat{N}_s \rangle$  of species  $s$  can be adjusted via a chemical potential  $\mu_s$  by adding the term  $-\mu_s \hat{N}_s$  to the Hamiltonian. An example of successful use of the SGF algorithm in both canonical and grand-canonical ensembles is given in [21].

## VI. IMPLEMENTATION

Any implementation of the SGF algorithm will have to efficiently evaluate all the update probabilities that appear in table I. In particular it will have to efficiently trace the values of the matrix elements  $\langle L | \hat{\mathcal{G}} \hat{\mathcal{T}} | R \rangle$ ,  $\langle L | \hat{\mathcal{T}} \hat{\mathcal{G}} | R \rangle$ , the left and right matrix elements  $\langle \psi | \hat{\mathcal{T}} | R \rangle$  and  $\langle L | \hat{\mathcal{T}} | \psi \rangle$  and the potential energy difference  $\Delta V$ . For a Hamiltonian with arbitrary non-diagonal terms, for example with long range hopping, the only general way of keeping track of these quantities is to calculate on the fly those matrix elements. Therefore the cpu time of each update will scale linearly with the number of non-diagonal terms in the Hamiltonian. This scaling severely limits the system sizes that are accessible with implementations that rely on recalculating all the probabilities at each update.

Fortunately, a major simplification is possible for the most physically relevant Hamiltonians, for which the non-diagonal terms couple a constant number of sites, as in the models of Eq. (1) and Eq. (2). In both these Hamiltonians the number of non-diagonal terms increases linearly in size and each non-diagonal term involves only a few neighboring site indices. In this section we will discuss an update algorithm where the cpu time per update scales logarithmically with the number of non-diagonal terms.

Without loss of generality we will consider the case with an insertion of a non-diagonal term to the right of  $\hat{\mathcal{G}}$ , while the latter is moving to the left. The relative probability of the new state  $|R'\rangle$  after the insertion is

$$W^{LR}(R') = \langle L | \hat{\mathcal{G}} | R' \rangle \langle R' | \hat{\mathcal{T}} | R \rangle, \quad (111)$$

where the new matrix element of the green operator,  $\langle L | \hat{\mathcal{G}} | R' \rangle$ , is evaluated using Eq. (5). This expression depends only on the updated offset of the Green operator, that is to say the updated number of broken lines,

$$n_{LR'} = \left| \mathbf{n}^L - \mathbf{n}^{R'} \right| = \sum_i \left| n_i^L - n_i^{R'} \right|, \quad (112)$$

where the sum is over all sites  $i$ , and  $n_i^L$  and  $n_i^{R'}$  are the corresponding occupancies of states  $|L\rangle$  and  $|R'\rangle$  respectively. In most physical models each term,  $\hat{\mathcal{T}}_k$ , in  $\hat{\mathcal{T}}$  is a product of just a few creation/annihilation operators, such as  $a_i^\dagger a_j$ . Therefore the occupancies of  $|R'\rangle$  and  $|R\rangle$  will differ only over few indices. We can separate out those indices in Eq. (112) and write the updated number of broken lines as

$$n_{LR'} = n_{LR} + \delta n_{LR} \left[ \hat{\mathcal{T}}_k \right], \quad (113)$$

where  $\delta n_{LR} \left[ \hat{\mathcal{T}}_k \right]$  is the additional number of lines broken by  $\hat{\mathcal{T}}_k$  and is given by

$$\delta n_{LR} \left[ \hat{\mathcal{T}}_k \right] = \sum_{i \text{ in } \hat{\mathcal{T}}_k} \left( \left| n_i^L - n_i^R + \delta n_i(\hat{\mathcal{T}}_k) \right| - \left| n_i^L - n_i^R \right| \right), \quad (114)$$



where  $n_i^L$  and  $n_i^R$  are the occupancies of the  $i^{th}$  index of  $|L\rangle$  and  $|R\rangle$  respectively,  $\delta n_i(\hat{T}_k)$  is the change in occupancy caused by  $\hat{T}_k$  on index  $i$ , and the summation extends only over the indices of the operator  $\hat{T}_k$ . We will refer to  $\delta n_{LR}[\hat{T}_k]$  as the “offset” of  $\hat{T}_k$ . Eq. (113) and (114) suggest that the total number of broken lines can be updated by summing only over the usually few indices of  $\hat{T}_k$  rather than all the indices of the configuration.

Furthermore we can rewrite Eq. (111) as

$$W^{LR}(R') = g(n_{LR} + \delta n_{LR}[\hat{T}_k]) T_k^R, \quad (115)$$

where  $T_k^R \equiv \langle R' | \hat{T}_k | R \rangle$  is the matrix element of the term. The offset is an integer that can take only few values, for example it can only take the values -2, 0, and +2 for the model (1) and the values -6, -4, -2, 0, +2, +4, and +6 for the model (2). It can therefore be used to categorize the non-diagonal terms. Let  $S(\delta n)$  be the group of non-diagonal terms with the same offset,  $\delta n$ . For all those terms in  $S(\delta n)$  the matrix element of the Green operator,  $g(n_{LR} + \delta n)$ , is the same.

The key idea is that the choice of an operator can be done in two steps, first we choose an offset and then a term from the corresponding group. The relative probability for choosing an offset is the sum of the relative probabilities of all the terms with the same offset and can be written as

$$W^{LR}(\delta n) = g(n_{LR} + \delta n) \sum_{\hat{T}_k \text{ in } \hat{T}} T_k^R \delta_{\delta n, \delta n_{LR}(\hat{T}_k)}, \quad (116)$$

where the summation extends over all the non-diagonal terms and the Kröneckers delta selects the terms with a particular offset  $\delta n$ . After the offset a non-diagonal term is chosen from the group  $S(\delta n)$  with relative probability  $T_k^R$ . The normalization of all the probabilities,  $\langle L | \hat{G} \hat{T} | R \rangle$ , reads

$$\langle L | \hat{G} \hat{T} | R \rangle = \sum_{\delta n} W^{LR}(\delta n). \quad (117)$$

So far we discussed about the insertion of a term  $\hat{T}_k$  to the right. Removing a term from the right is equivalent to inserting  $\hat{T}_k^\dagger$  as far as the configuration is concerned. Similarly an operator can be added to the left which is the same as adding  $\hat{T}_k^\dagger$  to the left or removed from the left which is equivalent to removing  $\hat{T}_k$ .

When a term is added or removed, it will affect the offsets and the matrix elements,  $T_k^R$  of the terms that share an index with it. An optimization occurs if instead of updating the matrix elements and offsets of every term we update only the few affected terms. In most physical models the number of non-diagonal terms that share any particular index does not change with the system size. For example if all non-diagonal terms are of the form  $a_i^\dagger a_j$ , such as in the Bose-Hubbard model (1), then in

$d$  dimensions, there are exactly  $4d$  terms with any particular index. Therefore if such a term is inserted with indices  $i$  and  $j$  there are  $8d-2$  terms that will be affected.

This observation leads to the development of a fast-update algorithm for SGF. When a term is inserted or removed, a few non-diagonal terms will need to be removed from the group of their original offset and added to the group of their final offset. The pairs of  $(\hat{T}_k, T_k^R)$  for each group can be stored in a binary search tree, one for each offset  $\delta n$  and direction of insertion relative to  $\hat{G}$ . The binary search tree supports searching, inserting and removing of terms that scales logarithmically in the number of terms. There are many potential implementations of such a binary search tree. One possibility is that the leaves correspond to the terms and their weight or zero if they are not. The nodes have weights which are equal to the sum of the weights of their children. To choose a term one starts from the root and selects one of the children randomly with their weight as the relative probability and proceeds similarly until a leaf is reached. When a term is inserted/removed, only the weights of the nodes that are immediate ancestors need to be updated. Therefore the logarithmic scaling is guaranteed. A simple picture of a binary search tree with 4 terms is shown in Fig. 8.

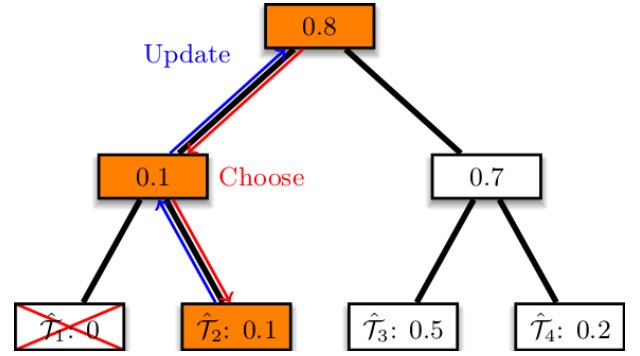


FIG. 8. A binary search tree for a model with 4 non-diagonal terms. There is one such tree for each group of terms with the same offset  $\delta n$ . The leaves represent the terms and their weights,  $T_k^R$ . The nodes are the sum of the weights of all their children. In this example, the term  $\hat{T}_1$  is not part of the group because it has zero weight. The colored boxes represent the nodes that will be accessed if term  $\hat{T}_2$  is chosen, or when it is inserted, removed from the tree or its weight changes. To choose an operator, one starts from the root and proceeds to the leaves, by making a binary choice at every level with relative probabilities equal to the stored weights of each node. When the weight of a term is updated during an insertion, removal, or a change of occupancy, the change is communicated to all its parents all the way up to the root.

## VII. TEST OF THE ALGORITHM

In this section we illustrate the exactness of the SGF algorithm with global space-time update by making comparisons between QMC results and exact diagonalizations. We also show that tuning the directionality of the new global space-time update allows to minimize the auto-correlation time. Finally we show that the optimized auto-correlation time is smaller than the one obtained with the previous formulation of the directed update [14].

### A. Exactness of the SGF algorithm

Figure 9 shows a comparison between SGF results and an exact diagonalization for the non-trivial model (2) in the hard-core limit, for  $2 \times 2$  hexagons (12 sites),  $\beta t = 4$ , at half-filling. The total energy and the superfluid density are measured as explained in subsections IV F 1 and IV F 2, respectively. The agreement for the energy and the superfluid density illustrates the exactness of the SGF algorithm in both small and large  $K/t$  limits.

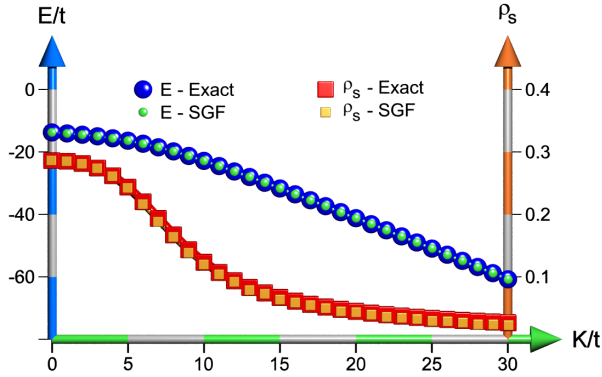


FIG. 9. (Color online) Comparison between the SGF algorithm with global space-time update and an exact diagonalization for the model with six-site ring-exchange terms in the hard-core limit. The agreement for the energy and the superfluid density is good in both small and large  $K/t$  limits.

The specific heat, measured as explained in subsection IV F 7, is shown on Fig. 10 as a function of temperature  $T$  for  $K/t = 5$  at half-filling. Again, the agreement is excellent both at low and at high temperature.

### B. Auto-correlation time

Let us consider a random variable  $X$  and a set of  $M$  successive samples  $X(1), X(2), \dots, X(M)$ . We define the auto-correlation function of  $X$  at separation  $\tau \in \mathbb{N}$  (shift

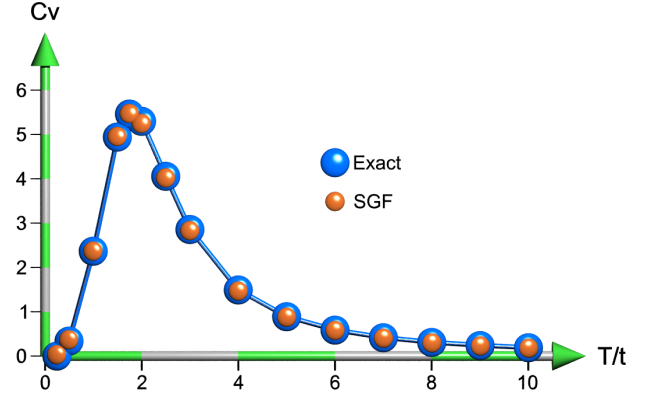


FIG. 10. (Color online) The specific heat  $C_v$  as a function of temperature  $T$ . Comparison between the SGF algorithm with global space-time update and an exact diagonalization for the model with six-site ring-exchange terms in the hard-core limit. The two curves show a perfect agreement.

in the sample indices) by

$$C_X(\tau) = \frac{1}{M-\tau} \sum_{t=1}^{M-\tau} (X(t+\tau) - \langle X \rangle)(X(t) - \langle X \rangle), \quad (118)$$

where  $\langle X \rangle$  is the average value of  $X$ . If the samples  $X(t)$  are independent from each other, then  $C_X(\tau)$  is vanishing for all  $\tau > 0$ , namely  $C_X(\tau) \propto \delta_{\tau,0}$ . If the samples are correlated, then the auto-correlation function is a decreasing function of  $\tau$ . The decay law (exponential, power, ...) depends on the system's Hamiltonian and dimensionality, and whether one is close to a transition point or not. Thus we need a definition of the auto-correlation time that is independent of these details. As a result, we define the auto-correlation time  $\tau_c$  by the CPU time needed for the auto-correlation function to drop by one order of magnitude with respect to its value at zero time separation:

$$C_X(\tau_c) = \frac{C_X(0)}{10} \quad (119)$$

The auto-correlation time depends on the physical quantity  $X$  that is measured. In the following we consider the model (1) with  $L = 100$  sites and the inverse temperature  $\beta t = 20$ . We set the hopping parameter to  $t = 1$ , and vary the onsite repulsion  $U$  and the first-nearest neighbor interaction  $V$  in order to drive the system in a superfluid phase, a Mott insulating phase, and a charge density wave phase. We calculate the auto-correlation functions for the potential energy  $E_p$ , the kinetic energy  $E_k$ , the winding number  $W$ , and a non-trivial six-point correlation function  $C_{cor}^6$  defined by

$$C_{cor}^6 = \left\langle \overleftarrow{T}_\tau [I_1 I_2 I_3] \right\rangle^h \quad (120)$$

$$I_1 = \frac{1}{\beta} \int_0^\beta a_2^\dagger(\tau) a_3(\tau) d\tau \quad (121)$$

$$\mathcal{I}_2 = \frac{1}{\beta} \int_0^\beta a_3^\dagger(\tau) a_4(\tau) d\tau \quad (122)$$

$$\mathcal{I}_3 = \frac{1}{\beta} \int_0^\beta a_7^\dagger(\tau) a_6(\tau) d\tau, \quad (123)$$

which is measured by making use of (78).

As an illustration, Fig. 11 shows the normalized auto-correlation function of the six-point correlation function (120) as a function of CPU time for three different sets of parameters corresponding to a Mott insulator (top panel), a superfluid (middle panel), and a charge density wave (bottom panel). The inset of the Mott insulator case with logarithmic-linear axes shows that the decay is nearly exponential, while the inset of the charge density wave case with logarithmic-logarithmic axes shows a power law decay. This justifies our definition of the auto-correlation time (119) which is independent of the decay law.

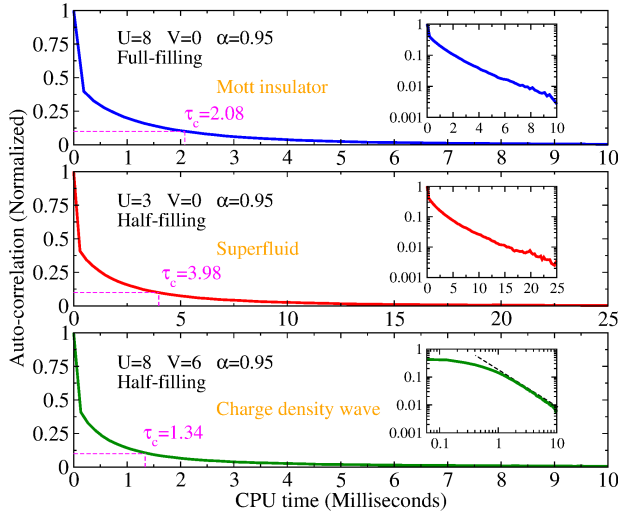


FIG. 11. (Color online) The normalized auto-correlation function of the six-point correlation function (120) as a function of CPU time for three different sets of parameters. For the Mott insulator case (top panel), the inset with logarithmic-linear axes reveals a nearly exponential decay. In the case of the charge density wave (bottom panel), the inset with logarithmic-logarithmic axes shows that the decay follows a power law.

### 1. Effect of the directionality

The parameter  $\alpha$  allows us to control the directionality of the global space-time update, that it to say the average width of the time window that is visited by the Green operator during a directed update. In other words, increasing  $\alpha$  makes the steps of the random walk of the Green operator in imaginary time larger, so the time needed to visit the full operator string is smaller. But this also increases the CPU time needed to perform a directed update. The competition between these two effects results

in the existence of an optimal value of  $\alpha$  that gives the smallest auto-correlation time for a given observable.

Figure 12 shows the auto-correlation time of  $E_p$ ,  $E_k$ ,  $W$  and  $C_{cor}^6$  as functions of the directionality  $\alpha$  for the superfluid phase (top panel), the Mott insulating phase (middle panel), and the charge density wave phase (bottom panel). Note that for the Mott insulator and the charge density wave it is not possible to define an auto-correlation time for the winding number  $W$ , because it is systematically vanishing for all samples. It is worth to emphasize how the auto-correlation time is highly reduced for  $C_{cor}^6$  from  $\alpha = 0$  to  $\alpha = 0.99$  by a factor of  $\approx 20$  in the Mott insulating phase.

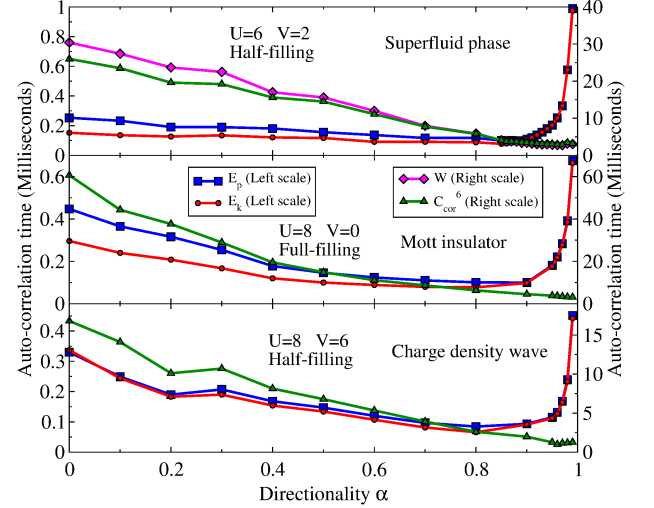


FIG. 12. (Color online) The auto-correlation time of  $E_p$ ,  $E_k$ ,  $W$  and  $C_{cor}^6$  as functions of the directionality  $\alpha$  in the superfluid phase (top panel), the Mott insulating phase (middle panel), and the charge density wave phase (bottom panel).

As a result, there exist an optimal value for  $\alpha$  that gives the smallest auto-correlation time. But this optimal value depends on the quantity that is measured and the phase in which the system is. In general we find that a good compromise is to choose  $\alpha \approx 0.90$ , which gives nearly optimized auto-correlation times for all observables in all phases of the system.

### 2. Global space-time update versus previous formulation of directed update

We compare here the auto-correlation times obtained with the global space-time update and those obtained with the previous formulation of directed update [14]. The auto-correlation times given here are obtained by using the optimal values of  $\alpha$ , which are determined graphically by using plots similar to Fig. 12 for each set of parameters.

We vary the values of the onsite potential  $U$ , the interaction between nearest neighbors  $V$ , and the density of particles  $\rho$ . These parameters cover the superfluid phase,

the Mott insulating phase, and the charge density wave phase. Results for the potential energy  $E_p$ , the kinetic energy  $E_k$ , the winding number  $W$ , and the correlation function  $C_{cor}^6$  are shown on Fig. 13 as scatter plots of the auto-correlation times. One can notice that all points fall in the lower right part of the graphs, which indicates that the auto-correlations times obtained with the global space-time update are smaller than those obtained with the previous directed update in all cases.

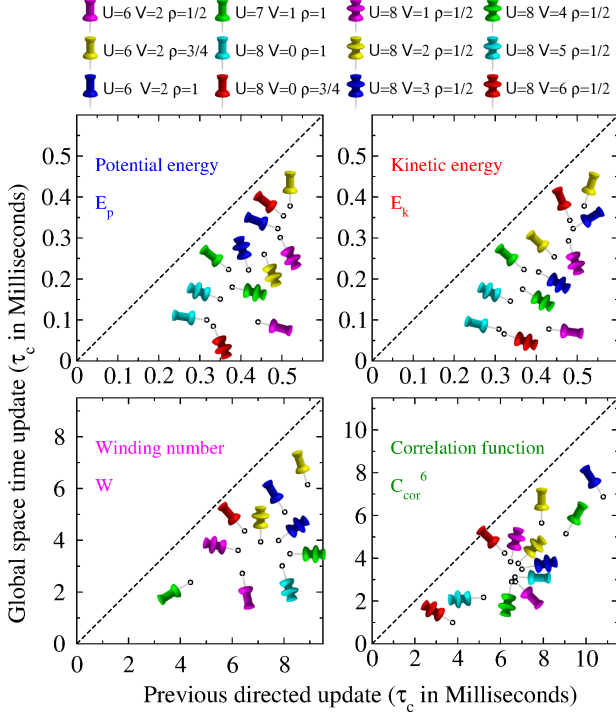


FIG. 13. (Color online) Comparison between the auto-correlation times obtained with the global space-time update and those obtained with the previous formulation of the directed update [14]. All points fall in the lower right part of the graphs, which indicates that the auto-correlations times obtained with the global space-time update are smaller than those obtained with the previous directed update in all cases.

### VIII. CONCLUSION

We have presented the Stochastic Green Function (SGF) algorithm and showed that it is able to simulate any Hamiltonian that does not suffer from the so-called “sign problem”. We have proposed a new *global space-time update* scheme which, in addition to being *directed*, has the advantage of reducing the auto-correlation time of the samples of measured quantities. The SGF algorithm is the first quantum Monte Carlo (QMC) method that does not make any assumption on the form of the Hamiltonian. As a result, it can be directly applied “as is” to any Hamiltonian. We have presented an optimized implementation where each update scales logarithmically

with the system size, and which allows access to larger systems. We have illustrated the capabilities of the SGF algorithm by applying it to a Hamiltonian that includes six-site coupling terms, which is challenging for other QMC methods. In addition, we have shown that the SGF algorithm can work in both canonical and grand-canonical ensembles. Finally, we have shown that various quantities of interest can be measured by the algorithm, such as  $n$ -point Green functions, imaginary time-dependent correlation functions, the imaginary dynamical structure factor, and the specific heat.

### Appendix A: Proof of (64) and theorem (78)

For convenience, we introduce the operator

$$\hat{\mathcal{U}}(b, a) = \overleftarrow{\mathbf{T}}_{\tau} e^{\int_a^b \hat{\mathcal{T}}(\tau) d\tau}, \quad (\text{A1})$$

which satisfies the property

$$\hat{\mathcal{U}}(c, b) \hat{\mathcal{U}}(b, a) = \hat{\mathcal{U}}(c, a) \quad (\text{A2})$$

for  $a \leq b \leq c$ . Using (7) and (A1), it follows that any set of arbitrary operators  $\hat{\mathcal{A}}_1 \cdots \hat{\mathcal{A}}_R$  satisfies

$$\begin{aligned} e^{-\beta \hat{\mathcal{H}}} \tilde{\mathcal{A}}_R(\tau_R) \cdots \tilde{\mathcal{A}}_1(\tau_1) &= e^{-\beta \hat{\mathcal{V}}} \hat{\mathcal{U}}(\beta, \tau_R) \hat{\mathcal{A}}_R(\tau_R) \\ &\quad \times \hat{\mathcal{U}}(\tau_R, \tau_{R-1}) \cdots \hat{\mathcal{U}}(\tau_2, \tau_1) \\ &\quad \times \hat{\mathcal{A}}(\tau_1) \hat{\mathcal{U}}(\tau_1, 0), \end{aligned} \quad (\text{A3})$$

for any set of time-ordered indices  $0 < \tau_1 < \cdots < \tau_R < \beta$ . By noticing that all  $\tilde{\mathcal{A}}$ ,  $\hat{\mathcal{A}}$ , and  $\hat{\mathcal{U}}$  operators in (A3) are written in the chronological order, we can enclose them under a single time-ordering operator, which allows us to combine all  $\hat{\mathcal{U}}$  operators into a single one. This leads to

$$\begin{aligned} e^{-\beta \hat{\mathcal{H}}} \overleftarrow{\mathbf{T}}_{\tau} [\tilde{\mathcal{A}}_R(\tau_R) \cdots \tilde{\mathcal{A}}_1(\tau_1)] \\ = e^{-\beta \hat{\mathcal{V}}} \overleftarrow{\mathbf{T}}_{\tau} [\hat{\mathcal{A}}_R(\tau_R) \cdots \hat{\mathcal{A}}(\tau_1) e^{\int_0^{\beta} \hat{\mathcal{T}}(\tau) d\tau}], \end{aligned} \quad (\text{A4})$$

which is now valid for any set of unsorted time indices  $\tau_1, \dots, \tau_R \in [0, \beta]$ . Taking the trace and normalizing with  $\mathcal{Z}$  leads to (64). If in addition we integrate over all time indices, we get

$$\begin{aligned} \left\langle \overleftarrow{\mathbf{T}}_{\tau} [\tilde{\mathcal{I}}(\hat{\mathcal{A}}_R) \cdots \tilde{\mathcal{I}}(\hat{\mathcal{A}}_1)] \right\rangle^h \\ = \frac{1}{\mathcal{Z}} \text{Tr} e^{-\beta \hat{\mathcal{V}}} \overleftarrow{\mathbf{T}}_{\tau} [\hat{\mathcal{I}}(\hat{\mathcal{A}}_R) \cdots \hat{\mathcal{I}}(\hat{\mathcal{A}}_1) e^{\int_0^{\beta} \hat{\mathcal{T}}(\tau) d\tau}] \end{aligned} \quad (\text{A5})$$

Consider now the case where the operators  $\hat{\mathcal{A}}_1 \cdots \hat{\mathcal{A}}_R$  are all equal to the same non-diagonal Hamiltonian term  $\hat{\mathcal{T}}_k$ , then (A5) reduces to:

$$\begin{aligned} \left\langle \overleftarrow{\mathbf{T}}_{\tau} [(\tilde{\mathcal{I}}(\hat{\mathcal{T}}_k))^R] \right\rangle^h \\ = \frac{1}{\mathcal{Z}} \text{Tr} e^{-\beta \hat{\mathcal{V}}} \overleftarrow{\mathbf{T}}_{\tau} [(\hat{\mathcal{I}}(\hat{\mathcal{T}}_k))^R e^{\int_0^{\beta} \hat{\mathcal{T}}(\tau) d\tau}] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\mathcal{Z}} \text{Tr} e^{-\beta \hat{\mathbf{V}}} \overleftarrow{\mathbf{T}}_{\tau} \left[ e^{\int_0^{\beta} \sum_{q \neq k} \hat{\mathcal{T}}_q(\tau) d\tau} \sum_{n_k \geq 0} \frac{\left( \int_0^{\beta} \hat{\mathcal{T}}_k(\tau) d\tau \right)^{n_k+R}}{n_k! \beta^R} \right] \\
&= \sum_{n_k \geq 0} \frac{n_k!}{\beta^R (n_k - R)!} \\
&\quad \times \underbrace{\frac{1}{\mathcal{Z}} \text{Tr} e^{-\beta \hat{\mathbf{V}}} \overleftarrow{\mathbf{T}}_{\tau} \left[ e^{\int_0^{\beta} \sum_{q \neq k} \hat{\mathcal{T}}_q(\tau) d\tau} \frac{\left( \int_0^{\beta} \hat{\mathcal{T}}_k(\tau) d\tau \right)^n}{n_k!} \right]}_{\text{Probability of a string with } n_k \text{ operators of type } \hat{\mathcal{T}}_k} \\
&= \frac{1}{\beta^R} \left\langle \frac{n_k!}{(n_k - R)!} \right\rangle^i \tag{A6}
\end{aligned}$$

It is easy to check that the above derivation is still valid when mixing up arbitrary operators  $\hat{\mathcal{A}}_k$  with any num-

ber of non-diagonal Hamiltonian terms  $\hat{\mathcal{T}}_k$ , hence (78) is proven.

## ACKNOWLEDGMENTS

VR is supported by NSF OISE-0952300 and DG by Lee Kuan Yew Fellowship of Singapore. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number DMR100007. We are grateful to Kalani Hettiarachchilage for useful discussions, and Mark Jarrell and Juana Moreno for encouraging and supporting the development of the fast-update implementation.

- 
- [1] N. Metropolis and S. Ulam, J. Am. Stat. Assoc. 44, 335 (1949).
  - [2] D.C. Handscomb, Proc. Cambridge Philos. Soc. 58, 594 (1962).
  - [3] M.H. Kalos, Phys. Rev. 128, 1791 (1962).
  - [4] R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D 24, 2278 (1981).
  - [5] J. B. Anderson, J. Chem. Phys. 65:41214127 (1976).
  - [6] M.H. Hettler, A.N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H.R. Krishnamurthy, Phys. Rev. B 58, R7475-R7479 (1998).
  - [7] M.H. Hettler, M. Mukherjee, M. Jarrell, and H.R. Krishnamurthy, Phys. Rev. B 61, 12739-12756 (2000).
  - [8] G.G. Batrouni and R.T. Scalettar, Phys. Rev. B 46, 9051 (1992).
  - [9] A.W. Sandvik and J. Kurkijarvi, Phys. Rev. B 43, 5950 (1991).
  - [10] H.G. Evertz, G. Lana and M. Marcu, Phys. Rev. Lett. 70 (1993) 875.
  - [11] G. Carleo, F. Becca, S. Moroni, and S. Baroni, Phys. Rev. E 82, 046710 (2010).
  - [12] K. Van Houcke, S.M.A. Rombouts, and L. Pollet, Phys. Rev. E 73, 056703 (2006).
  - [13] V.G. Rousseau, Phys. Rev. E 77, 056705 (2008).
  - [14] V.G. Rousseau, Phys. Rev. E 78, 056707 (2008).
  - [15] V.G. Rousseau and P.J.H. Denteneer, Phys. Rev. Lett. 102, 015301 (2009).
  - [16] G.G. Batrouni, V.G. Rousseau, and R.T. Scalettar, Phys. Rev. Lett. 102, 140402 (2009).
  - [17] L. de Forges de Parny, M. Traynard, F. Hébert, V.G. Rousseau, R.T. Scalettar, and G.G. Batrouni, Phys. Rev. A 82, 063602 (2010).
  - [18] L. de Forges de Parny, F. Hébert, V.G. Rousseau, R.T. Scalettar, and G.G. Batrouni, Phys. Rev. B 84, 064529 (2011).
  - [19] L. de Forges de Parny, F. Hébert, V.G. Rousseau, and G.G. Batrouni, Eur. Phys. J. B. (2012) 85:169.
  - [20] G. Carleo, M. Tarzia2, and F. Zamponi, Phys. Rev. Lett. 103, 215302 (2009).
  - [21] V.G. Rousseau, K. Tam, M. Jarrell, and J. Moreno, arXiv:1206.2566v1.
  - [22] E.L. Pollock, and D.M. Ceperley, Phys. Rev. B 36, 8343 (1987).
  - [23] V.G. Rousseau, K. Hettiarachchilage, M. Jarrell, J. Moreno, and D.E. Sheehy, Phys. Rev. A 82, 063631 (2010).