Bond-dependent slave-particle cluster theory based on density matrix expansion

Zheting Jin [1] and Sohrab Ismail-Beigi [1,2,3]

[1] *Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA*

[2] *Department of Physics, Yale University, New Haven, Connecticut 06511, USA*

[3] *Department of Mechanical Engineering and Materials Science, Yale University, New Haven, Connecticut 06511, USA*

(Received 21 September 2022; revised 22 February 2023; accepted 13 March 2023; published 27 March 2023)

Efficient and accurate computational methods for dealing with interacting electron problems on a lattice are of broad interest to the condensed matter community. For interacting Hubbard models, we introduce a cluster slave-particle approach that provides significant computational savings with high accuracy for total energies, site occupancies, and interaction energies. Compared to exact benchmarks using density matrix renormalization group for *d*-*p* Hubbard models, our approach delivers accurate results using two to three orders of magnitude lower computational cost. Our method is based on a slave-particle decomposition with an improved description of particle hoppings, and a density matrix expansion method where the interacting lattice slave-particle problem is turned into a set of overlapping real-space clusters which are solved self-consistently with appropriate physical matching constraints at shared lattice sites between clusters.

DOI: 10.1103/PhysRevB.107.115153

I. INTRODUCTION

One of the outstanding challenges in condensed matter physics is to find computationally efficient and simultaneously accurate methods to describe interacting electron systems for large lattices (e.g., crystalline materials). For the case where localized electronic interactions dominate, the Hubbard model provides a specific theoretical model that can describe key aspects of many important materials such as superconductors, magnets, or metal-insulator systems. Hence, a large amount of research effort has been expended in creating and improving methods for Hubbard systems.

Exact methods for Hubbard systems are limited to low dimensions or finite sizes. The ground state of the one-dimensional (1D) half-filled single-orbital Hubbard model can be solved analytically using the Bethe ansatz [1,2]. In addition, extremely high accurate results can be obtained using numerical methods including the density matrix renormalization group (DMRG) [3,4] and quantum Monte Carlos (QMC) [5]. However, for higher-dimensional models, only small lattices (or small fragments of lattices) can be solved by these numerical methods. Hence, an outstanding challenge is to find methods that work well in low and high dimensions, and (at present) this requires making approximations.

One approximate approach for the Hubbard model is the slave-particle method, also known as the slave or auxiliary or subsidiary boson method (e.g., our previously published boson subsidiary-solver (BOSS) software [6]). It was first proposed [7,8] for analytical calculations in the infinite interaction limit and as an alternative to the Gutzwiller variational approach [9,10]. It was then generalized to finite interactions by a functional integral approach based on the slave-particle representation [11–13]. Since this approach requires one auxiliary slave particle for each possible electronic configuration, whose number grows exponentially with the number of degrees of freedom on each site, the required computations can become expensive for complex materials. Hence, more economical slave-particle methods have been developed. These representations describe the slave particles via electron occupation numbers. Different methods have been developed based on the degrees of freedom treated by the slave particles, such as the slave-rotor method [14,15] which can serve as an impurity solver [16], the slave-spin method [17] which is orbital and spin selective, and a generalized approach that includes the above two (and other variant) methods [18]. An auxiliary symmetry-breaking field approach [19] was then introduced to overcome difficulties in achieving spontaneous symmetry breaking in these approaches.

In slave-particle methods, the interacting electron problem is decomposed into a noninteracting spinon problem on a lattice (easily solved using Bloch's theorem and diagonalization) and an interacting slave-particle problem on a lattice. The most common approach for solving the latter has been to use a single-site approximation. This is very similar in spirit to the local single-site approximation in dynamical mean-field theory (DMFT) [20], a state-of-the-art method for (approximate) solutions to Hubbard models. Single-site slave-particle approaches predict Mott transitions in high-dimensional systems very well [21–26]. However, as a result of stronger fluctuations in low-dimensional systems, the single-site treatments can cause significant errors. For example, the exact solution of the half-filled 1D Hubbard model [2] has no Mott transition for any finite interaction strength U, but a false Mott transition is predicted in the single-site slave-particle theory and DMFT [19,27–29].

For higher accuracy, one must go beyond the single-site approximation and consider a local cluster of interacting lattice sites. Several cluster extension methods [30,31] were proposed for slave-particle problems based on a cluster mean-field approximation. If one sets certain quasiparticle renormalization factors to unity, the cluster slave-particle

theory can be simplified [27,32] to the density matrix embedding theory (DMET) [33,34]. However, setting the renormalization factors to unity leads to the appearance of the bare intersite hopping in the embedding spinon Hamiltonian, and this cannot reproduce interaction-induced band narrowing, e.g., as predicted by *GW* or DMFT calculations [35–38]. In addition, all existing cluster slave-particle, cluster DMFT, and cluster DMET methods describe the cluster as a multisite impurity connected to an averaged external bath. Inevitably, some chemical bonds with large associated hoppings are approximated as intercluster hoppings between different fragments of the systems, and the cutting and modification of these bonds to form the clusters can cause large errors [39–41]. Consequently, the finite cluster size effects lead to a tradeoff between cluster size and errors in these cluster methods [27,39,40,42,43]. The finite-size errors are even more difficult to control in higher-dimensional systems since the cluster surface grows with cluster radius r as r^{d-1} in d dimensions.

In this work, we introduce a slave-particle method that addresses many prior shortcomings. We present a bond-dependent slave-particle theory along with a cluster decomposition based on a density matrix expansion. It has the following key features:

(i) The degrees of freedom involved in each slave bond are orbital+spin+bond, and the free parameters in the slave operators are designed so that all unphysical, particle nonconserving intersite hoppings are forbidden (this was impossible in prior slave-particle approaches).

(ii) Instead of coupling the interacting problem (site or cluster) to a mean-field bath, a density matrix expansion approach is used to reduce the interacting slave-particle lattice problem to a set of separate cluster problems solved under appropriate constraints. The clusters overlap with each other to span the whole lattice, so that they connect via the interacting density matrices on the shared sites instead of via a mean-field bath.

(iii) The resulting numerical method is highly efficient and parallelizable: All the benchmark tests on d-p Hubbard models described below take on the order of one CPU minute of serial computation to complete on a modern commodity laptop computer. In addition, for a general d-dimensional lattice and for a fixed cluster size, the computational cost only grows quadratically with the number of clusters in the whole system. The separate clusters can be solved in parallel, so that generalizations to large systems, higher dimensions, and multiple orbitals per lattice site will have reasonable computational costs.

In the following, we describe the theory and then present numerical results based on its implementation. As we will see, our theory reproduces remarkably accurate results with low computational cost compared to our benchmark results either from exact diagonalization or DMRG.

II. SLAVE-BOND REPRESENTATION

In this section, we introduce our slave-bond representation and explain how it generalizes standard site-based slave-particle methods in a manner that allows one to avoid all unphysical particle-number-violating hopping processes in the slave-particle problem. We compare our method to more familiar existing methods at each step to allow for a clear comparison in the reader's mind.

We consider Hubbard Hamiltonians of the form

$$\hat{H} = -\sum_{\alpha\beta} t_{\alpha\beta}\hat{c}^\dagger_\alpha\hat{c}_\beta + \sum_\alpha \epsilon_\alpha \hat{n}_\alpha + \hat{H}^{\text{int}}, \tag{1}$$

where greek-letter indices α, β combine the site indices i, j, orbital indices m, m', and spin indices σ, σ' together, ranging over all sites, orbitals, and spins in the system, i.e., $\alpha \equiv im\sigma$. The $\hat{c}_\alpha$ is a fermion annihilation operator removing an electron from localized state α, and $\hat{n}_\alpha = \hat{c}^\dagger_\alpha\hat{c}_\alpha$ is the fermion number counting operator for state α. The $t_{\alpha\beta}$ and ϵ_α denote hopping and onsite energies, respectively. The electron-electron interaction term $\hat{H}^{\text{int}}$ is the sum of local operators at each site $\hat{H}^{\text{int}}_i$, which in the simplest case are given by the classic Hubbard "U" form

$$\hat{H}^{\text{int}} = \sum_i \hat{H}^{\text{int}}_i = \sum_{i,m} U_{im}\hat{n}_{im\uparrow}\hat{n}_{im\downarrow}, \tag{2}$$

where each site and spatial orbital can, in principle, have its unique interaction strength U_{im}. Additional local interactions that depend on the electron counts $\hat{n}_\alpha$ are completely straightforward to include requiring no change of formalism [18,19].

The standard approaches for slave particles [14–19,30,31,44] replace the physical electron operators on each site by a combination of a noninteracting fermion (called a spinon) and an interacting auxiliary or slave particle in a completely local manner:

$$\hat{c}_\alpha \to \hat{f}_\alpha\hat{O}_\alpha,$$

where $\hat{f}_\alpha$ is the noninteracting fermion annihilation operator, and $\hat{O}_\alpha$ is the lowering ladder operator for the slave particles. Accordingly, the original physical electron Hilbert space is mapped onto a larger Hilbert space $\mathcal{H} \to \mathcal{H}_f \otimes \mathcal{H}_s$ where $\mathcal{H}_f$ and $\mathcal{H}_s$ are the Hilbert spaces of the spinons and slave particles, respectively. The spinon Hilbert space $\mathcal{H}_f$, being a fermionic one, contains the same degrees of freedom as the original electron problem, but the physical modes for the slave particles vary based on the type of slave model chosen. For example, if all degrees of freedom α are explicitly described in slave-particle Hilbert space, it is known as the "slave-spin" or "spin+orbital" method [17,18].

The eventual goal of any slave-particle method is to have the spinons carry the fermionic spin of the original electron, while the slave particles carry the charge of the electron, and by decoupling them one has two easier problems to solve (see the next section). However, at this stage, one is still considering an exact reformulation, so the spinons and slave particles always move together during a hopping process in a correlated manner. Mathematically, it means that in the enlarged Hilbert space $\mathcal{H}_f \otimes \mathcal{H}_s$, there is a subset of physical states where the number of spinons and slave particles are equal for each state α which form a faithful one-to-one representation of the original states of the physical electrons $|n_\alpha\rangle$:

$$|n_\alpha\rangle \to |n^f_\alpha = n_\alpha; N_\alpha = n_\alpha\rangle, \tag{3}$$

where $n_\alpha, n^f_\alpha, N_\alpha$ are the occupation numbers of the physical electron, spinon, and slaves, respectively, and can take the

values of 0 or 1. We will call these states the "physical" or "number-matching" states in what follows.

With this short review concluded, we now define our slave-bond formalism. Our formalism lives in the same Hilbert space as before $\mathcal{H}_f \otimes \mathcal{H}_s$, and the difference is in how we choose to define the slave-particle operators. Our approach is based on using the full microscopic set of local quantum numbers $\alpha \equiv im\sigma$ for the slave-particle description. We take a pair of localized states α, β to define the bond $\alpha\beta$ with associated hopping operators $\hat{c}^\dagger_\alpha \hat{c}_\beta$ which moves an electron from β to α. We define our slave-bond representation by using the spinon fermionic operators $\hat{f}_\alpha$ as before but defining slaves-particle operators that have a bond index $\langle\alpha\beta\rangle$, so the operator replacement is now given by

$$\hat{c}^\dagger_\alpha \hat{c}_\beta \to \hat{f}^\dagger_\alpha \hat{f}_\beta \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle}. \tag{4}$$

To make the action of $\hat{f}^\dagger_\alpha \hat{f}_\beta \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle}$ on the physical states $|n^f_\alpha = n_\alpha; N_\alpha = n_\alpha\rangle$ identical to that of $\hat{c}^\dagger_\alpha \hat{c}_\beta$ on the original electron states $|n_\alpha\rangle$, Appendix A shows that the lowering operator $\hat{O}_{\alpha\langle\alpha\beta\rangle}$ must take the form

$$\hat{O}_{\alpha\langle\alpha\beta\rangle} = \begin{pmatrix} 0 & 1 \\ c_{\alpha\langle\alpha\beta\rangle} & 0 \end{pmatrix}, \tag{5}$$

with the basis ordered as $\{|N_\alpha = 0\rangle, |N_\alpha = 1\rangle\}$. The number $c_{\alpha\langle\alpha\beta\rangle}$ is often called the "gauge" and its value is arbitrary at present since we are dealing with the exact problem with no approximations within the physical subspace. The constraints determining its value will be described in Sec. III. Appendix A also shows that the form of Eq. (5) also guarantees that anticommutation relations are obeyed for the collective bond spinon+slave operator:

$$\{\hat{f}_\beta \hat{O}_{\beta\langle\alpha\beta\rangle}, \hat{f}^\dagger_\alpha \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\} = \delta_{\alpha\beta}. \tag{6}$$

It is important to clarify that the individual site-based slave-particle lowering operators $\hat{O}_{\alpha\langle\alpha\beta\rangle}$ resemble but are *not* bosonic field operators. They are defined in such a way to obey Eq. (6). Hence, our theory is a slave-particle theory and not a slave-boson theory.

In contrast to prior slave-particle theories, the slave-bond operators in our theory are nonlocal. Figure 1 shows examples of bonds (double-arrow lines) on a checkerboard lattice, where the bonds can correspond to hopping processes in the Hamiltonian, e.g., nearest-neighbor hopping (orange) and next-nearest-neighbor hopping (green). However, one can also consider longer-ranged slave bonds (purple). Therefore, in principle, there are a huge number of bonds in a crystalline system. But, in practice, only a small subset contributes to observables like the Hamiltonian. For example, only bonds with nonzero hoppings $t_{\alpha\beta} \neq 0$ contribute to the total energy, and these are typically only the nearest and next-nearest neighbors. From a pragmatic viewpoint, in our work below we only need to define slave-bond operators on the bonds that appear with $t_{\alpha\beta} \neq 0$ in Eq. (1).

To compare to previous slave-particle methods, for a given bond $\langle\alpha\beta\rangle$, our slave-bond approach can be viewed as a recipe similar to the prior site-based slave-particle approaches where one does the replacement

$$\hat{c}_\alpha \to \hat{f}_\alpha \hat{O}_{\alpha\langle\alpha\beta\rangle}. \tag{7}$$

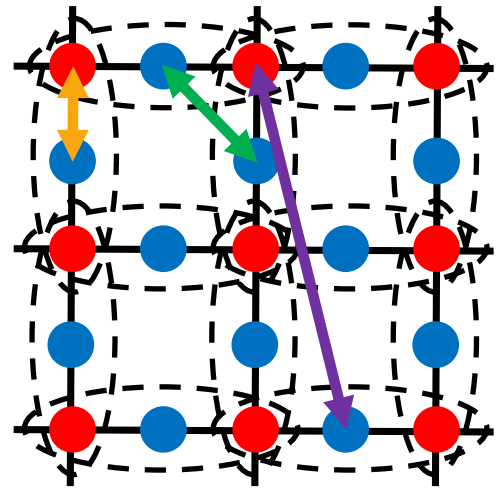

FIG. 1. An illustration of a checkerboard lattice structure representing a metal oxide 2D layer. Red circles represent correlated d sites (transition metals), and blue circles are noninteracting p sites (oxygens). The double-arrow lines illustrate examples of bonds used to define the slave-particle operators. The dashed black ellipses indicate the d-p-d clusters in the layer which overlap with each other on the correlated d sites.

Even though the mapping of Eq. (7) is mathematically correct (as detailed in Appendix A), it shows an index mismatch from both sides. Physically, it originates from the fact that the slave bonds are nonlocal; operationally, what it means is that mapping of Eq. (7) only makes sense in the context of the hopping part of the Hamiltonian which is the sum over bonds: For each bond $\alpha\beta$, one can do the mapping in Eq. (7) without confusion.

III. SLAVE-PARTICLE DECOMPOSITION

With the slave-bond representation, the Hamiltonian of Eq. (1) in the enlarged Hilbert space turns into

$$\hat{H} = -\sum_{\alpha\beta} t_{\alpha\beta} \hat{f}^\dagger_\alpha \hat{f}_\beta \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} + \sum_\alpha \epsilon_\alpha \hat{n}_\alpha + \sum_i \hat{H}^{\rm int}_i, \tag{8}$$

where, again, i is a site index and greek-letter indices combine site, orbital, and spin together. The local interaction term in the original electron Hamiltonian can be described by the slave particles alone through $\hat{H}^{\rm int}_i = \sum_m U_{im} \hat{N}_{im\uparrow} \hat{N}_{im\downarrow}$. The difficulty is that in addition to the physical or number-matching states, the enlarged Hilbert space $\mathcal{H}_f \otimes \mathcal{H}_s$ also contains numerous unphysical states, e.g., a state such as $|n^f_\alpha = 0; N_\alpha = 1\rangle$ where the number of spinons and slaves do not match for localized state α. In an exact treatment of the interacting problem, these states are excluded. However, to make practical progress, one must make approximations.

The first approximation common to all slave-particle approaches, including ours, is to decouple the spinon and slave problems. The simplest way forward is to approximate the density matrix for the joint spinon+slave system $\rho_{\rm tot}$ by a product of a spinon density matrix ρ_f and a slave density matrix ρ_s, i.e., $\rho_{\rm tot} = \rho_f \otimes \rho_s$. This decouples the two problems at the cost of losing the concerted or correlated description

of the spinon and slave particles during the hopping process along each bond. The best one can do is to ensure agreement on average. Hence, one enforces a matching of the expectation values of the number operators for each state α,

$$\langle \hat{n}_\alpha \rangle_f = \langle \hat{N}_\alpha \rangle_s. \tag{9}$$

Here, expectations are defined in the standard way: For any operator $\hat{A}$ acting in the spinon space, we have $\langle \hat{A} \rangle_f = \mathrm{Tr}(\hat{A}\hat{\rho}_f)$; similarly, for any operator $\hat{B}$ acting in the slave space, we have $\langle \hat{B} \rangle_f = \mathrm{Tr}(\hat{B}\hat{\rho}_s)$. Here $\hat{n}_\alpha$ and $\hat{N}_\alpha$ are number operators in spinon and slave Hilbert space, and we will drop the f superscript on the spinon number operator going forward.

Given this approximate density matrix, the average of a hopping process along a bond in Eq. (4) factorizes as

$$\langle \hat{f}^\dagger_\alpha \hat{f}_\beta \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle = \langle \hat{f}^\dagger_\alpha \hat{f}_\beta \rangle_f \langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle_s. \tag{10}$$

Most generally, the expectation of the product operator $\hat{A}\hat{B}$ of a spinon operator $\hat{A}$ and slave operator $\hat{B}$ factorizes as

$$\begin{aligned} \langle \hat{A}\hat{B} \rangle &= \mathrm{Tr}(\hat{A}\hat{B}\hat{\rho}_f \otimes \hat{\rho}_s) \\ &= \mathrm{Tr}(\hat{A}\hat{\rho}_f)\mathrm{Tr}(\hat{B}\hat{\rho}_s) = \langle \hat{A} \rangle_f \langle \hat{B} \rangle_s. \end{aligned} \tag{11}$$

The decoupling in Eq. (10) results in two simpler problems to be solved instead of the original electron problem. The easiest way to achieve this is to consider the variational problem of minimizing the total energy $E = \langle H \rangle$ under the occupation number constraints of Eq. (9) as well as more obvious constraints of the normalization of the density matrices $\mathrm{Tr}(\hat{\rho}_f) = \mathrm{Tr}(\hat{\rho}_s) = 1$. Using the Lagrange multiplier approach, we consider the unconstrained minimization of the function F:

$$\begin{aligned} F = \langle \hat{H} \rangle &- \sum_\alpha h_\alpha (\nu_\alpha - \langle \hat{N}_\alpha \rangle_s) - \sum_\alpha h'_\alpha (\nu_\alpha - \langle \hat{n}_\alpha \rangle_f) \\ &- \epsilon_f \big[\mathrm{Tr}(\hat{\rho}_f) - 1 \big] - \epsilon_s [\mathrm{Tr}(\hat{\rho}_s) - 1]. \end{aligned} \tag{12}$$

Here h_α, h'_α are Lagrange multipliers for the mean occupation numbers of the slaves and spinons, while ϵ_f, ϵ_s are the Lagrange multipliers for the normalization constraints. For convenience, we have enforced occupation number matching via separate matching to target spinon occupancies ν_α, whose values are determined variationally [19].

The differential of F versus the two independent variables $\hat{\rho}_f$ and $\hat{\rho}_s$ takes the form

$$dF = \mathrm{Tr}([\hat{H}_f - \epsilon_f \hat{I}] d\hat{\rho}_f) + \mathrm{Tr}([\hat{H}_s - \epsilon_s \hat{I}] d\hat{\rho}_s),$$

where the effective Hamiltonians for spinons ($\hat{H}_f$) and slaves ($\hat{H}_s$) are

$$\begin{aligned} \hat{H}_f &= -\sum_{\alpha\beta} t_{\alpha\beta} \langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle_s \hat{f}^\dagger_\alpha \hat{f}_\beta + \sum_\alpha (\epsilon_\alpha + h'_\alpha) \hat{n}_\alpha, \\ \hat{H}_s &= -\sum_{\alpha\beta} t_{\alpha\beta} \langle \hat{f}^\dagger_\alpha \hat{f}_\beta \rangle_f \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} + \sum_\alpha h_\alpha \hat{N}_\alpha + \sum_i \hat{H}^{\mathrm{int}}_i. \end{aligned} \tag{13}$$

Therefore, the original interacting electron problem is turned into a noninteracting spinon problem with symmetry-breaking field (onsite energies) h'_α and an interacting slave problem with onsite energies h_α. Both Hamiltonian problems contain hopping terms that are renormalized by averages of the other problem, and the renormalization factors are to be determined self-consistently at the minimum energy configuration. The expectation values of observables are then described by the minimizing spinon and slave density matrices.

Since we have created an approximation to the problem, the choice of gauge numbers $c_{\alpha\langle\alpha\beta\rangle}$ will now matter. In previous slave-particle methods [14,17,18], the gauge is chosen to ensure that in the noninteracting limit, the spinon system alone will faithfully describe the original electron problem. In these prior approaches, the purely local replacement $\hat{c}_\alpha \to \hat{f}_\alpha \hat{O}_\alpha$ goes hand in hand with a single-site slave-particle approximation where $\hat{\rho}_s$ is approximated as a product over single-site density matrices $\hat{\rho}^i_s$ as $\hat{\rho}_s \approx \bigotimes_i \hat{\rho}^i_s$. Then the slave expectation over each bond factorizes as well $\langle \hat{O}^\dagger_\alpha \hat{O}_\beta \rangle_s \approx \langle \hat{O}^\dagger_\alpha \rangle_s \langle \hat{O}_\beta \rangle_s$. The gauge numbers c_α are then chosen to ensure $\langle \hat{O}_\alpha \rangle_s = 1$ when $\hat{H}^{\mathrm{int}}_i = 0$, and thus in the noninteracting limit, the spinon Hamiltonian $\hat{H}_f$ of Eq. (13) will become identical to the original electron Hamiltonian since the slave hopping renormalization factor $\langle \hat{O}^\dagger_\alpha \hat{O}_\beta \rangle_s$ is replaced by unity. As we explain below, this methodology permits unphysical particle-number-violating hopping process which creates significant errors in the total energy.

One of the virtues of our bond-based approach is that it eliminates such unphysical processes. Consider the original electron Hamiltonian: The hopping on a given bond $\langle\alpha\beta\rangle$ is described by the term $-t_{\alpha\beta}\hat{c}^\dagger_\alpha \hat{c}_\beta + \mathrm{H.c.}$ This term is Hermitian and particle conserving. In the Fock space, its only nonzero matrix elements are between the two states $|n_\alpha = 0; n_\beta = 1\rangle$ and $|n_\alpha = 1; n_\beta = 0\rangle$. Within the exact slave-particle description in the physical subspace, the same is true of the corresponding hopping operator $-t_{\alpha\beta}\hat{f}^\dagger_\alpha \hat{f}_\beta \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} + \mathrm{H.c.}$ because the fermionic spinon part $\hat{f}^\dagger_\alpha \hat{f}_\beta$ alone can ensure that all other matrix elements are zero. However, once we approximately separate the spinon and slave problems, the slave-particle Hamiltonian $\hat{H}_s$ of Eq. (13) does not necessarily conserve particle number. The hopping along bond $\langle\alpha\beta\rangle$ in $\hat{H}_s$ is proportional to

$$\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} + \mathrm{H.c.} = \begin{pmatrix} 0 & 0 & 0 & v^* \\ 0 & 0 & u^* & 0 \\ 0 & u & 0 & 0 \\ v & 0 & 0 & 0 \end{pmatrix}, \tag{14}$$

where the basis is ordered as $\{|N_\alpha = 0; N_\beta = 0\rangle,\ |N_\alpha = 0; N_\beta = 1\rangle,\ |N_\alpha = 1; N_\beta = 0\rangle,\ |N_\alpha = 1; N_\beta = 1\rangle\}$, and $v = c_{\alpha\langle\alpha\beta\rangle} + c_{\beta\langle\alpha\beta\rangle}$ and $u = 1 + c_{\alpha\langle\alpha\beta\rangle} c^*_{\beta\langle\alpha\beta\rangle}$. The physical processes are proportional to u while the unphysical ones are proportional to v. In contrast to prior slave-particle methods where the gauge c_α is a fixed number for each local slave mode α, our approach provides the additional bond index which allows us to require the additional constraint $v = c_{\alpha\langle\alpha\beta\rangle} + c_{\beta\langle\alpha\beta\rangle} = 0$ or $c_{\alpha\langle\alpha\beta\rangle} = -c_{\beta\langle\alpha\beta\rangle}$. And the remaining gauge freedom for bond $\langle\alpha\beta\rangle$ (i.e., the value of $c_{\alpha\langle\alpha\beta\rangle}$) is chosen to ensure the correct noninteracting limit for $\hat{H}_f$, namely, that $\langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle_s = 1$ at zero interaction strengths. We take the gauge numbers $c_{\alpha\langle\alpha\beta\rangle}$ to be real so that the number of constraints (two) matches the number of gauge variables on each bond. Detailed analytical formulas are summarized in Appendix C.

For a noninteracting problem, the slave renormalization factors are unity, so that the spinon Hamiltonian alone is sufficient to compute the total energy $\langle \hat{H} \rangle$ and match the original electron problem. In addition, slave-particle methods are analytically exact at the large interaction limit [7,8], so one can view them as interpolation methods for finite interaction strengths.

From a practical viewpoint, the spinon problem on the lattice is trivial to solve by diagonalizing the associated one-particle Hamiltonian matrix $\hat{H}^0$ given by $\hat{H}^0_{\alpha\beta} = -t_{\alpha\beta} + \delta_{\alpha\beta}(\epsilon_\alpha + h'_\alpha)$ using Bloch's theorem. However, the slave Hamiltonian is still an interacting many-body problem on a lattice and is impossible to solve exactly for a large lattice. Some approximations are required as per the next section.

IV. CLUSTER APPROXIMATION

We now describe a cluster approximation based on a density matrix expansion that maps the infinite lattice slave problem onto a set of coupled finite-sized interacting clusters. We use overlapping clusters so that each chemical bond, e.g., transition-metal–oxygen bond, is included in some clusters and will be described explicitly. A choice of overlapping clusters for a two-dimensional corner-sharing (checkerboard) metal oxide layer is illustrated by the dashed black ellipses in Fig. 1 where each cluster consists of three sites: Two interacting d sites and the p site between them.

Each cluster $\mathcal{C}$ is small enough that one can, in principle, solve the many-body interacting slave problem for that cluster directly. For describing ground states, this would provide us with the cluster density matrix $\hat{\rho}_\mathcal{C}$. The question we address is how to take the set of $\{\hat{\rho}_\mathcal{C}\}$ over all the overlapping clusters and create a global quantum state (density matrix) for the entire lattice which we can then use to compute observables like the total energy.

The density matrix $\hat{\rho}$ of the entire interacting slave problem will be approximated via the following real-space, site-based cluster expansion:

$$\hat{\rho} = \bigotimes_i \hat{\rho}_i + \sum_\mathcal{C} \left(\hat{\rho}_\mathcal{C} - \bigotimes_{i\in\mathcal{C}} \hat{\rho}_i \right) \bigotimes_{k\notin\mathcal{C}} \hat{\rho}_k. \tag{15}$$

The indices i and k refer to sites (atoms) in the system, and $\hat{\rho}_i$ is the single-site density matrix for site i obtained by tracing out the degrees of freedom at all other sites

$$\hat{\rho}_i \equiv \mathrm{Tr}_{j\neq i}(\hat{\rho}). \tag{16}$$

The first term in $\hat{\rho}$ of Eq. (15) approximates the density matrix of the entire system as the tensor product over the single-site density matrices, and this represents the complexity of almost all current slave-particle approaches: Each site i is solved separately from the rest (albeit self-consistently via some type of bath linking the sites). The second term in Eq. (15) improves by incorporating the additional correlations described by the cluster compared to the single-site approximation. Figure 2 illustrates the density matrix expansion of Eq. (15). One begins with the collection of single-site density matrices $\hat{\rho}_i$ creating an approximation which is then improved by adding the cluster-wide density matrix contributions beyond the single-site approximation. This expansion is trivially exact for an infinitely large cluster, while for finite-sized clusters, spatial correlations up to the cluster size are explicitly included. These clusters connect or handshake with each other via the single-site density matrices $\hat{\rho}_i$ on the shared sites as explained below.

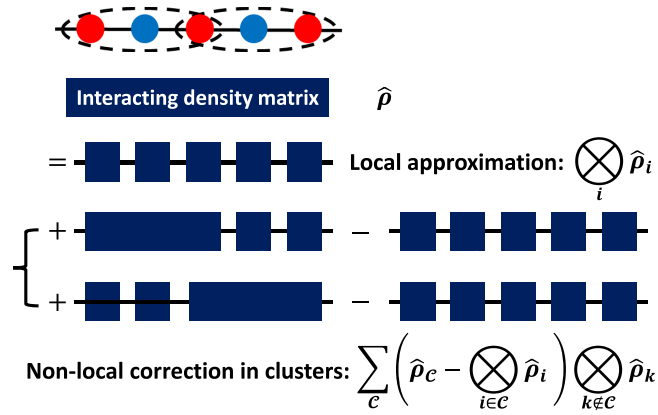

FIG. 2. An illustration of density matrix expansion for a one-dimensional system with red and blue atoms, where black dashed clusters overlap at red atom sites. The interacting density matrix and its reduced density matrices on sites are shown as dark blue rectangles. The direct product of density matrices is represented by black solid lines connecting rectangles.

Since the $\hat{\rho}_\mathcal{C}$ are the basic variables describing the full system $\hat{\rho}$, the single site $\hat{\rho}_i$ of Eq. (16) must also be derivable from the $\hat{\rho}_\mathcal{C}$. Inserting Eq. (15) into (16) yields the consistency condition

$$\hat{\rho}_i = \frac{1}{M_i} \sum_{\mathcal{C}|i\in\mathcal{C}} \mathrm{Tr}_{\mathcal{C}-i}(\hat{\rho}_\mathcal{C}), \tag{17}$$

where $\mathrm{Tr}_{\mathcal{C}-i}(\hat{\rho}_\mathcal{C})$ is shorthand for

$$\mathrm{Tr}_{\mathcal{C}-i}(\hat{\rho}_\mathcal{C}) \equiv \mathrm{Tr}_{k\in\mathcal{C}|k\neq i}(\hat{\rho}_\mathcal{C}) \tag{18}$$

which is the trace over all sites in cluster $\mathcal{C}$ excluding site i. The number M_i is the number of clusters that include site i: e.g., a d site in Fig. 1 has $M_i = 4$ whereas a p site belongs to a single cluster so it has $M_i = 1$. We will also use the shorthand

$$\hat{\rho}_i^{(\mathcal{C})} = \mathrm{Tr}_{\mathcal{C}-i}(\hat{\rho}_\mathcal{C}) \tag{19}$$

for the single-site density matrix at site i coming from the cluster density matrix of cluster $\mathcal{C}$.

Equation (17) states the sensible condition that $\hat{\rho}_i$ is the average over all single-site density matrices coming from the cluster that overlap at site i. However, to describe a consistent quantum state specified by $\hat{\rho}$ for all sites, we require a stronger consistency condition where the density matrix at each site is well defined and unique. Namely, we insist on the additional constraints that

$$\hat{\rho}_i = \hat{\rho}_i^{(\mathcal{C})} \quad \forall\, \mathcal{C}|i \in \mathcal{C} \tag{20}$$

separately for all clusters $\mathcal{C}$ containing i: i.e., all the M_i contributions in Eq. (17) are the same. Mathematically, for site i in cluster $\mathcal{C}$ we employ a matrix of Lagrange multipliers $\hat{\Lambda}_i^{(\mathcal{C})}$ with associated Lagrange multiplier term $\mathrm{Tr}(\hat{\Lambda}_i^{(\mathcal{C})}[\hat{\rho}_i - \hat{\rho}_i^{(\mathcal{C})}])$ to enforce the constraint.

Two additional properties of this constrained density matrix expansion are (i) it has a consistent description of short-ranged density matrices from the total density matrix $\hat{\rho}$, so $\mathrm{Tr}_{k\notin\mathcal{C}}(\hat{\rho}) = \hat{\rho}_{\mathcal{C}}$ and $\mathrm{Tr}_{k\neq i}(\hat{\rho}) = \hat{\rho}_i$; and (ii) it approximates the long-range behavior of the true density by single-site products: When sites i and j are far enough apart that they are not both in a single cluster, then $\mathrm{Tr}_{k\neq i,j}(\hat{\rho}) = \hat{\rho}_i \otimes \hat{\rho}_j$.

Using the density matrix of Eq. (15), the total energy $E = \mathrm{Tr}(\hat{H}[\hat{\rho}_f \otimes \hat{\rho}_s])$ turns into

$$\begin{aligned} E = &\sum_{\alpha} \epsilon_\alpha \langle \hat{n}_\alpha \rangle_{\rho_f} + \sum_i \langle \hat{H}_i^{\mathrm{int}} \rangle_{\rho_i} \\ &- \sum_{\alpha\beta|\langle\alpha\beta\rangle\in\exists\mathcal{C}} t_{\alpha\beta} \langle \hat{f}_\alpha^\dagger \hat{f}_\beta \rangle_{\rho_f} \langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle_{\rho_{\mathcal{C}}} \\ &- \sum_{\alpha\beta|\langle\alpha\beta\rangle\notin\forall\mathcal{C}} t_{\alpha\beta} \langle \hat{f}_\alpha^\dagger \hat{f}_\beta \rangle_{\rho_f} \langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \rangle_{\rho_{i|\alpha\in i}} \langle \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle_{\rho_{j|\beta\in j}}. \end{aligned} \tag{21}$$

In this formula, the hopping energy has two parts. The first is an intracluster hopping term describing hopping between two states α, β that are both inside of a cluster $\mathcal{C}$ (notation $\langle\alpha\beta\rangle \in \exists\mathcal{C}$): We can compute the associated slave-hopping expectation directly using the cluster density matrix without any factorization approximation. The second intercluster hopping term is for long-ranged hoppings between α, β when both are not in a single cluster: Here the slave-hopping expectation factorizes into the product of two single-site averages, and the consistency condition (20) ensures that the single-site averages of the hopping operators are well defined.

To minimize E with the required constraints, we use the Lagrange multiplier approach and consider the minimization of the function F:

$$\begin{aligned} F = E &- \sum_i \sum_{\mathcal{C}|i\in\mathcal{C}} \mathrm{Tr}_i \left(\hat{\Lambda}_i^{(\mathcal{C})} \left\{ \left[\frac{1}{M_i} \sum_{\mathcal{C}'|i\in\mathcal{C}'} \hat{\rho}_i^{(\mathcal{C}')} \right] - \hat{\rho}_i^{(\mathcal{C})} \right\} \right) \\ &- \sum_\alpha \sum_{\mathcal{C}|\alpha\in\mathcal{C}} h_\alpha^{(\mathcal{C})} (\nu_\alpha - \langle \hat{N}_\alpha \rangle_{\rho_{\mathcal{C}}}) - \sum_\alpha h'_\alpha (\nu_\alpha - \langle \hat{n}_\alpha \rangle_{\rho_f}) \\ &- \epsilon_f (\langle \hat{I} \rangle_{\rho_f} - 1 - \sum_{\mathcal{C}} \epsilon_{\mathcal{C}} (\langle \hat{I} \rangle_{\rho_{\mathcal{C}}} - 1). \end{aligned} \tag{22}$$

Each cluster has its own Lagrange multipliers $h_\alpha^{(\mathcal{C})}$ to enforce mean occupancy matching with the spinons $\langle \hat{n}_\alpha \rangle_f = \langle \hat{N}_\alpha \rangle_{\rho_{\mathcal{C}}}$, and $\epsilon_{\mathcal{C}}$ enforce that the trace of the cluster density matrices are unity. Interestingly, as proved in Appendix D, the additional constraints introduced by $\hat{\Lambda}_i^{(\mathcal{C})}$ turn out to be redundant for the Hamiltonians of interest here given that the mean particle numbers are already matched to the spinon occupancies $\langle \hat{n}_\alpha \rangle_f$ for a slave mode α at site i. So we can set all the $\hat{\Lambda}_i^{(\mathcal{C})} = 0$ which is a significant simplification.

Redoing the logic of the minimization problem for this function F yields the spinon Hamiltonian $\hat{H}_f$ and slave Hamiltonian $\hat{H}_{\mathcal{C}}$ governing each cluster $\mathcal{C}$:

$$\begin{aligned} \hat{H}_f &= -\sum_{\alpha\beta} t_{\alpha\beta} \langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle_\rho \hat{f}_\alpha^\dagger \hat{f}_\beta + \sum_\alpha (\epsilon_\alpha + h'_\alpha) \hat{n}_\alpha, \\ \hat{H}_{\mathcal{C}} &= \sum_{i|i\in\mathcal{C}} \frac{1}{M_i} \hat{H}_i^{\mathrm{int}} + \sum_{\alpha|\alpha\in\mathcal{C}} h_\alpha^{(\mathcal{C})} \hat{N}_\alpha - \sum_{\alpha\beta|\langle\alpha\beta\rangle\in\mathcal{C}} t_{\alpha\beta} \langle \hat{f}_\alpha^\dagger \hat{f}_\beta \rangle_{\rho_f} \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} - \sum_{\beta|\beta\in\mathcal{C}\alpha|\langle\alpha\beta\rangle\notin\forall\mathcal{C}'} \frac{t_{\alpha\beta}}{M_\beta} \langle \hat{f}_\alpha^\dagger \hat{f}_\beta \rangle_{\rho_f} [\langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \rangle_{\rho_{i|\alpha\in i}} \hat{O}_{\beta\langle\alpha\beta\rangle} + \mathrm{H.c.}]. \end{aligned} \tag{23}$$

Here M_α is a shorthand that equals the number of clusters to which the state α belongs and is equivalent to M_i for any $\alpha \in i$. The scaling factors of $1/M_i$ and $1/M_\alpha$ in $\hat{H}_{\mathcal{C}}$ originate from the relations (16) and (17) for the single-site density matrix $\hat{\rho}_i$. The cluster Hamiltonian $\hat{H}_{\mathcal{C}}$ contains intracluster hoppings (first hopping term) as well as long-range intercluster hoppings ($\langle\alpha\beta\rangle$ is not in any single cluster) going outside the cluster $\mathcal{C}$ whose renormalization factors depend on the spinons and also the other slave clusters.

Using the cluster approximation, the lattice slave-bond Hamiltonian $\hat{H}_s$ of Eq. (13) is mapped into a set of clusters, each with its own Hamiltonian $\hat{H}_{\mathcal{C}}$. The density matrix of the interacting slave-bond lattice problem is then described by the density matrices of the clusters from Eq. (15). The cluster Hamiltonians $\hat{H}_{\mathcal{C}}$ in Eq. (23) are, in general, not particle conserving due to the intercluster hopping terms that involve single raising and lowering $\hat{O}$ operators. In addition, the mean number of particles on any site or over any cluster is, in general, not an integer for a *pd* type model due to *d*-*p* hybridization. Hence, the cluster density matrix $\hat{\rho}_{\mathcal{C}}$ will describe a mixed state, automatically involving nearly degenerate low-energy states. In practice, we describe these mixed states using a Boltzmann distribution for each cluster using a very small temperature. In this work, we use exact diagonalization to solve each interacting cluster problem and then use the eigenstates to compute the Boltzmann distribution: This is not prohibitively difficult for the small three-site clusters we consider here. For larger clusters or more complex systems, more efficient methods of finding the Boltzmann distributed density matrix not involving naive diagonalization can be explored in the future.

V. TESTS ON THE HUBBARD DIMER

We begin with a simple system where the analytical solution of the electron problem as well as the single- and cluster-slave problems is possible. We consider the half-filled two-site Hubbard model (Hubbard dimer), whose Hamiltonian is

$$\hat{H} = -t \sum_\sigma (\hat{c}_{1\sigma}^\dagger \hat{c}_{2\sigma} + \hat{c}_{2\sigma}^\dagger \hat{c}_{1\sigma}) + U \sum_{i=1}^2 \hat{N}_{i\uparrow} \hat{N}_{i\downarrow}. \tag{24}$$

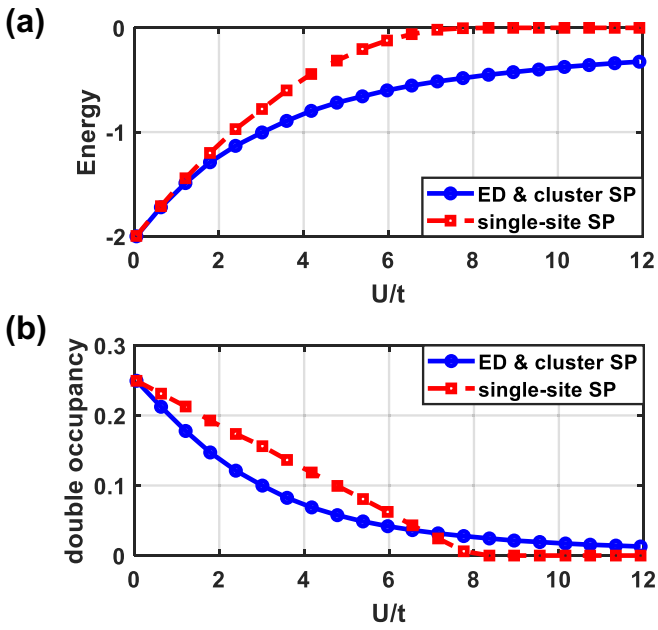

FIG. 3. (a) Total energy and (b) (mean) double occupancy of each site for the half-filled Hubbard dimer as a function of the interaction strength U/t, where the nearest-neighbor hopping strength t is the energy unit. Cluster slave-particle (SP) results are identical to the exact diagonalization (ED) results marked by blue circles, while single-site SP results are represented by red squares.

For such a small system, a single two-site slave cluster encompasses the whole system, so it is not surprising that the cluster-based method will provide excellent results. In fact, as detailed in Appendix B, exact diagonalization of the original electron Hamiltonian and solution of our cluster slave-particle method result in the same ground-state energy $E = (U - \sqrt{U^2 + 16t^2})/2$ as well as the same double occupancy

$$D_i \equiv \langle \hat{N}_{i\uparrow}\hat{N}_{i\downarrow}\rangle = \left(1 - \frac{U}{\sqrt{U^2+16t^2}}\right)\Big/4. \qquad (25)$$

This double occupancy is always positive for any real-valued U and t, direct evidence of the absence of a Mott transition in this system (also true of the one-dimensional half-filled one-band Hubbard model [2]). Figure 3 shows the exact E and D_i in blue.

However, analytical solution (see Appendix B) of the associated single-site slave-particle problem results in $E = -(U-8t)^2/32t$ for $U/t < 8$ and $E = 0$ when $U/t \geqslant 8$ and double occupancy $D_i = (8t-U)/32t$ when $U/t < 8$ and $D_i = 0$ when $U/t \geqslant 8$. Both energy and double occupancy indicate a Mott transition at $U/t = 8$ within this (erroneous) single-site approximation. Figure 3 shows the single sites E and D_i in red. This (false) Mott transition at $U/t = 8$ disagrees with the exact solution but is also inevitable within a single-site approach: A single site connected to a bath described by a single expectation $\langle O\rangle_s$ cannot know if the bath is meant to describe a zero or high-dimensional material problem; since a Mott transition can happen in higher-dimensional systems and is achievable in the single-site theoretical description when $\langle O\rangle_s = 0$, it occurs for some sufficiently large U/t in the single-site picture.

VI. TESTS ON d-p MODELS

Moving beyond the analytically solvable Hubbard dimer, we numerically test our theory on larger and more realistic models. In terms of modeling transition metal oxides, basic chemical considerations show that a minimal model should include both the d orbitals of the transition-metal atoms and the p orbitals of the oxygen atoms (a "d-p" model). For example, the single-orbital-per-site d-p model, also known as the Emery model [45–47], is intensively studied as a potential framework of the high-T_c copper-based superconductors. Due to the complexity caused by the explicit inclusion of the p states, the d-p model is numerically more challenging to solve but is also more realistic [48–50] than further simplified models such as the one-band Hubbard model [51] or the t-J model [52]. Hence, our numerical tests in this section will focus on the d-p model, while complementary one-band Hubbard model tests can be found in Appendix E.

For the systems studied below, the hopping renormalization factors and occupancies are determined self-consistently in a numerical fashion. In Appendix G, we describe the workflow of the self-consistent calculations involved in our numerical studies. As we will see below, our slave-bond method produces very accurate results compared to high-quality benchmark results in both 1D and 2D d-p systems.

We begin our tests with 1D d-p systems which have the lattice illustrated in Fig. 4(a). In all the following results, the nearest-neighbor d-p hopping strength is set to be $-t$ as illustrated in Fig. 4(a), where t is real positive and treated as the energy unit; all other hoppings are set to zero. The onsite energy of d sites is set to be $\epsilon_d = +2t$, while the onsite energy of p sites is $\epsilon_p = 0$. There is only one orbital for each site. In addition, a small temperature of $k_BT = 5 \times 10^{-3}t$ is introduced to allow the Boltzmann distribution to create a mixed state from multiple eigenfunctions of the Hamiltonian which is necessary as discussed in Sec. IV. The finite temperature does create errors in the calculation of ground-state properties, but as shown in Appendix F, these errors are controllable by reducing the temperature. Other convergence thresholds are set low enough to give accurate results: e.g., mean d-site occupancies are converged so that they differ between physically identical d sites by less than 10^{-7}.

We employ two types of benchmarks. For small systems, we use exact diagonalization (ED) of the starting electron Hamiltonian with a Boltzmann distribution at the same k_BT listed above. For larger problems, ED is infeasible, so we turn to DMRG. All DMRG calculations in this work use the ITENSOR software package (Julia version) [53,54], where the energy cutoff is set to be $10^{-8}t$; the maximum bond dimension increases by system size up to 1600 for a 96-site system, and the maximum number of sweeps is 1600. For the single-site slave-particle results, we use the boson subsidiary-solver (BOSS) software [6] with orbital- and spin-resolved slave particles using the same finite temperature.

A. Four-site 1D model

We begin with a small four-site model (i.e., a d-p-d-p chain) with periodic boundary conditions (PBC). The upper panel of Fig. 4(b) shows the total energy versus interaction

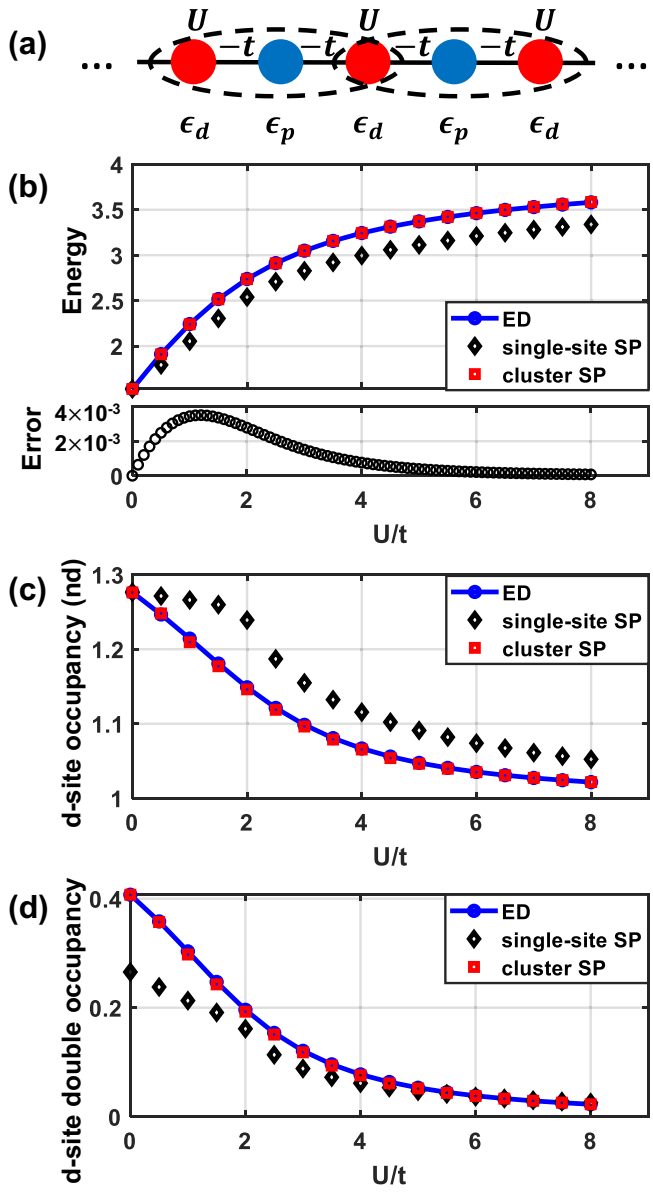

FIG. 4. (a) Illustrates the one-dimensional d-p chain with periodic boundary conditions, where red atoms represent d sites while blue atoms represent p sites. The black dashed ellipses are the clusters used in the slave-bond calculation. Upper panel of (b), and (c) and (d) show the total energy per unit cell in units of t, d-site occupancy, and d-site double occupancy versus the interaction strength U, respectively, for a four-site unit-cell linear chain with periodic boundary conditions. The results of exact diagonalization (ED), single-site slave-particle (SP), and cluster SP are marked by blue circles, black diamonds, and red squares, respectively. The lower panel of (b) shows the total energy error of the cluster SP method in units of t.

strength U. Both slave-particle (SP) methods reproduce exact total energy at $U = 0$ and $U \to \infty$ by construction. However, while the single-site method shows qualitatively correct behavior versus U, the cluster method shows quantitative accuracy. The same is true for the mean d-site occupancy $n_d = \langle \hat{N}_{d\uparrow} + \hat{N}_{d\downarrow} \rangle$ in Fig. 4(c). The double occupancy $D = \langle \hat{N}_{d\uparrow} \hat{N}_{d\downarrow} \rangle$ in Fig. 4(d) is computed in the slave-particle sector of the problem which has no reason to match the exact answer at $U = 0$, and the two SP calculations will match the exact answer only at $U \to \infty$. However, the cluster SP method remains very accurate for all values of U.

We note that for both energy and d-site occupancy, both cluster SP results agree with ED for both small and large U which is as expected. In the noninteracting limit, the c gauges enforce the spinon Hamiltonian alone to reproduce the noninteracting electron Hamiltonian: With $\langle \hat{O}^{\dagger}_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle = 1$ and no interaction terms, the total energy of Eq. (21) is the noninteracting energy. In the large U or atomic limit, the p sites are filled while the d sites are half-filled in both SP or ED calculations. Thus, the cluster SP reproduces the exact energy and occupation numbers in both limits. The single-site SP method has the same properties in both limits but has larger errors at finite U.

The lower panel of Fig. 4(b) shows the error of the total energy of the cluster SP as a function of interaction strength (U/t) calculated by subtracting the ED energy from the cluster SP energy. To further reduce the finite-temperature effect, the temperature is decreased to $k_B T = 1 \times 10^{-4} t$ in this particular calculation. Among all the 81 different interaction strengths U/t sampled from 0 to 8, the maximum energy error in cluster SP is $3.5 \times 10^{-3} t$ at $U/t = 1.2$, while such error is about 50 times larger in the single-site SP method. The maximum errors for occupancy and double occupancy in cluster SP are 1.1×10^{-3} and 1.6×10^{-3}, respectively.

B. System-size dependence in 1D

Beyond the four-site d-p system where exact diagonalization is possible, we have tested longer 1D chains with 8, 16, 32, 64, 96, and 128 sites at $U/t = 0$ and 8. Here we assume translational symmetry with the four-site unit cell which is then replicated: The objective is to gauge convergence versus system size. Benchmark results on the different-sized systems are generated by either exact diagonalization at $U = 0$ or DMRG [3] with open boundary condition (OBC) at finite U. We note that PBC is replaced by OBC in the DMRG calculations because PBC is computationally much more expensive than OBC according to the area law [55]. Correspondingly, we calculate and report the total energy per four-site unit cell in Fig. 5(a) for OBC DMRG calculations to compare it with cluster SP and exact diagonalization results. The d-site occupancy and double occupancy in Figs. 5(b) and 5(c) are computed around the middle point of the chain for DMRG so as to be the farthest away from the open boundaries. For the cluster SP calculations with PBC, we can use the translational symmetry to work with the four-site unit cell together with k-point sampling of the spinon problem: Larger systems correspond to denser k sampling of the noninteracting spinon problem. This makes for a very cheap computational scaling versus system size.

The system-size dependencies are shown in Fig. 5, where the total energy is quoted per four-site unit cell. Unsurprisingly, the cluster SP method is equivalent to exact diagonalization at $U = 0$ as discussed at the end of Sec. III. We also notice that both cluster SP and DMRG converge faster versus system size at $U/t = 8$ case than at $U/t = 0$. This comes from the intensively studied "band narrowing" effect

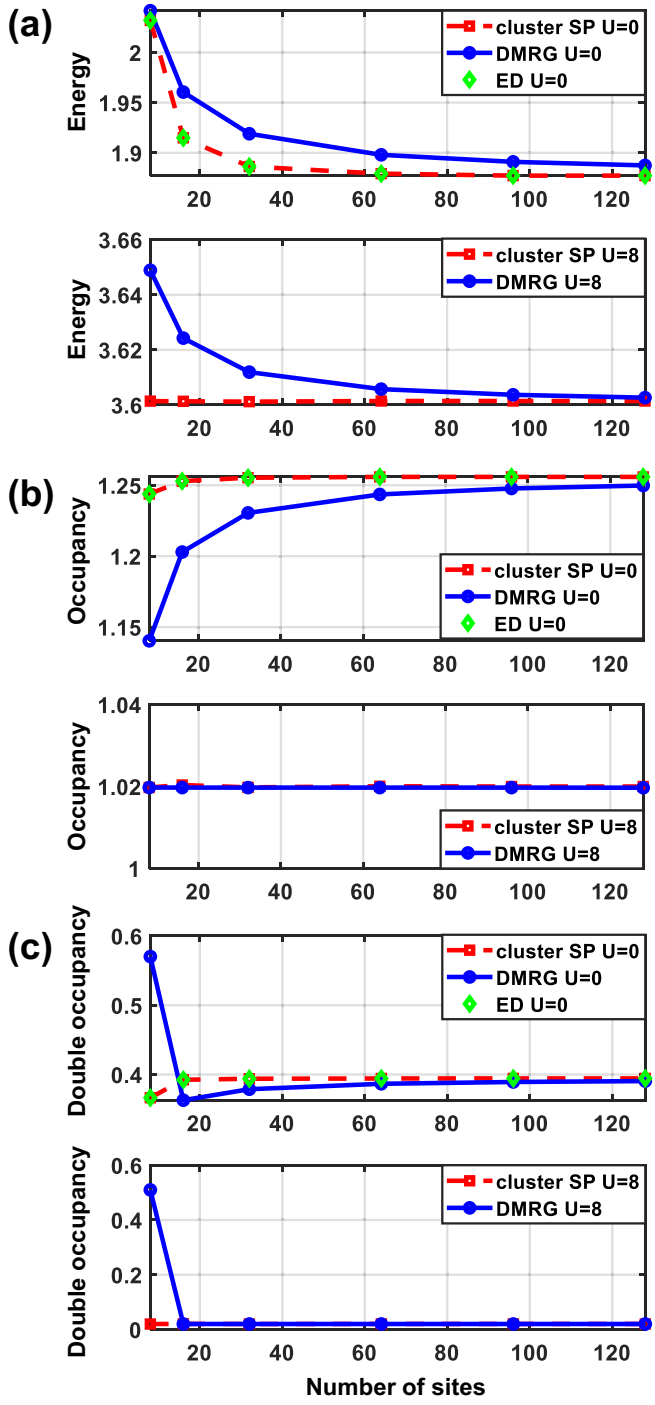

FIG. 5. Convergence of observables versus system size for the 1D half-filled d-p Hubbard chain: System sizes used are 8, 16, 32, 64, 96, and 128 sites. (a)–(c) Show the total energy of each four-site unit cell in units of t, the d-site occupancy, and the d-site double occupancy versus the number of sites, respectively. For each panel, the interaction strength used is $U = 0$ for the upper subpanel and $U = 8t$ for the lower one. The results of DMRG with OBC, cluster SP with PBC, and ED with PBC are marked by blue circles, red squares, and green diamonds, respectively.

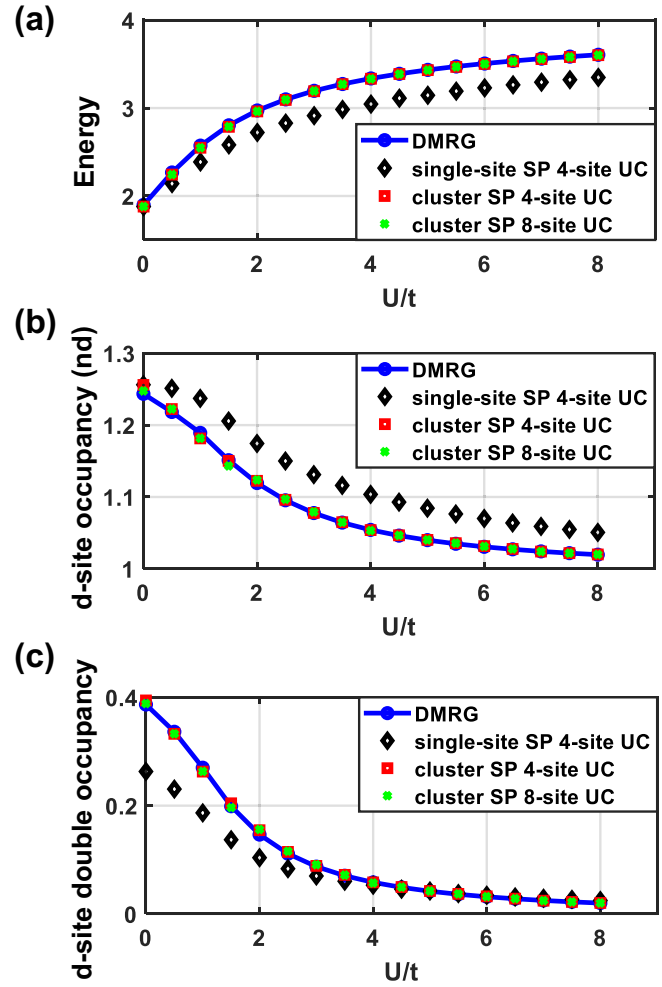

FIG. 6. (a)–(c) Show the total energy of each four-site unit cell in units of t, d-site occupancy, and d-site double occupancy versus the interaction strength U, respectively. All results come from 64-site 1D d-p system, blue circles represent DMRG results with OBC; black diamonds represent single-site SP results assuming a translational period of four sites; red squares and green crosses almost completely overlap with each other and represent cluster SP assuming a translational period of four sites and eight sites.

[18,35,56] due to finite U/t which makes the quasiparticle bands less dispersive and allows sparser k sampling for the same accuracy. In addition, cluster SP is well converged for most of the observables when the system size is larger than 16 sites, while DMRG needs more than 100 sites for the same level of convergence. The slower convergence of DMRG is due to edge effects induced by the open boundary condition needed to reduce bond dimension [53,54].

From a computational vantage point, to converge the total energy to $10^{-7}t$, each data point takes less than 1 CPU minute on a laptop for the slave SP calculations regardless of system size. On the other hand, for the OBC DMRG, we require about 6 CPU hours for a 64-site calculation and takes longer than one day for a 128-site calculation on a standard, contemporary Linux cluster.

We further examine the 64-site system as a function of U/t in Fig. 6. The cluster SP results marked by red squares assume a translational period of four sites and match well with the DMRG results marked by blue circles. As a comparison, the single-site SP method assuming the same translational

period is marked by black diamonds, which shows much larger errors in all three observables. Moreover, an additional cluster SP calculation assuming a translation period of eight sites is performed, whose results are marked by green crosses. This calculation with a double-sized unit cell behaves almost exactly the same as the four-site unit-cell calculation, which indicates the convergence of unit-cell size in our cluster SP results.

C. Doped 1D chains

We perform further tests on the 64-site 1D d-p system by hole doping, where the average doping density ranges from 0 up to 0.5 holes for each d-p pair. The local interaction strength on d site is fixed to be $U = 2t$, while other parameters such as the onsite energies, hopping strengths, and the temperature are unchanged from the no-doping calculations of the previous sections. We choose $U/t = 2$ because the errors caused by the cluster SP method are relatively large around this choice, as shown in the lower panel of Fig. 4(b), providing a stringent test of the cluster SP.

Our results of the doped systems are summarized in Fig. 7, where the two curves are results from cluster SP and DMRG. In Figs. 7(a) and 7(b), the two curves are almost overlapping with each other, indicating that the cluster SP method reproduces the energy per unit cell and the d-site occupancy of the ground state extremely precisely. In Fig. 7(c), the cluster SP reproduces good d-site double occupancy D with some small variations. While both DMRG and cluster SP methods show a nice linear relation between the total energy and doping level in Fig. 7(a), they both show small discrepancies away from linear relations in Figs. 7(b) and 7(c). This is because both DMRG and the cluster SP methods are variational approaches for the ground-state total energy, so a high level of energy convergence is guaranteed, but the convergence of other non-variational observables such as the site occupancy or double occupancy is much poorer. Note that the double-occupancy deviations are below 0.01, which is adequately small and about the same order of magnitude as other errors caused by the cluster approximation and finite-temperature effect and is almost invisible in Fig. 6(c). It is one order of magnitude smaller than the error caused by single-site SP approximation.

In short, based on all these tests, our cluster SP method reproduces overall very accurate results hundreds of times faster than DMRG for the one-dimensional d-p model. It also represents a significant quantitative improvement over the single-site SP method.

D. 2D d-p systems

The generalization to two-dimensional systems is straightforward in our cluster slave-particle theory. We choose a checkerboard lattice where the lattice structure and the clusters used in the calculations are illustrated in Fig. 1. The red and blue circles represent d and p sites, where each site contains only one orbital. This is a frequently studied model for cuprates known as the "three-band model" [48,49].

Figure 8(a) illustrates a primitive cell used in the calculation which is a parallelogram and contains two d sites and four

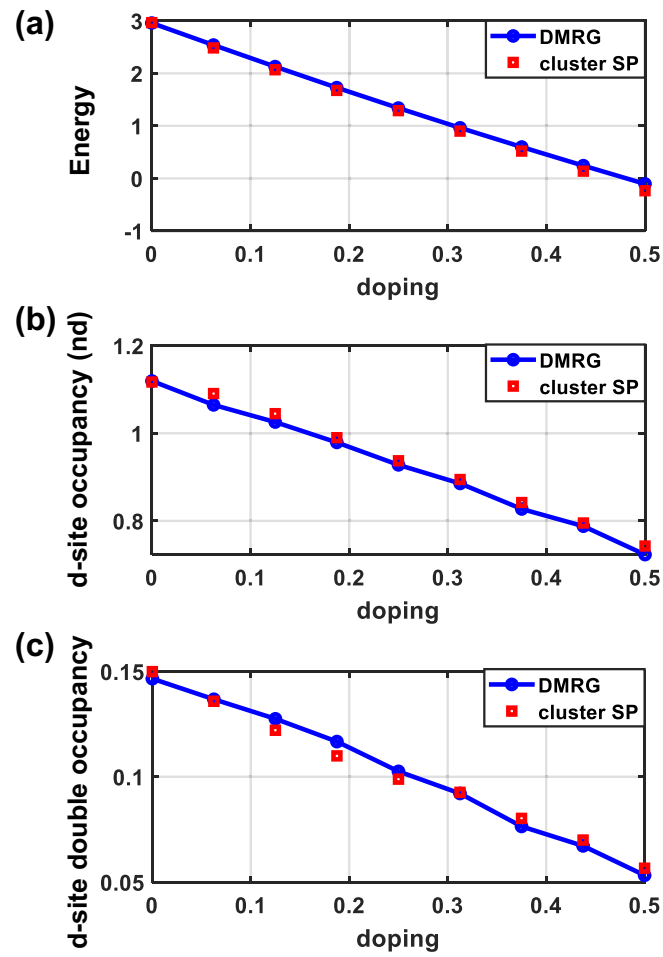

FIG. 7. (a)–(c) Show the total energy of each four-site unit cell (units of t), d-site occupancy, and d-site double occupancy, respectively. They are plotted as functions of the average hole doping level on each d-p pair. Blue circles represent the DMRG results as benchmarks, while the red squares stand for the cluster SP results assuming a translational period of four sites.

p sites. This is because we are permitting for Néel checkerboard antiferromagnetic (AFM) correlation (if we assumed a stripe pattern, the unit cells would be chosen differently). Similar to the 1D calculations discussed above, the nearest-neighbor hopping strength t is treated as the energy unit. All the onsite energies of d sites are set to be $\epsilon_d = +2t$, while the onsite energies of p sites are $\epsilon_p = 0$, and the small temperature applied is $k_B T = 5 \times 10^{-3} t$. Only nearest-neighbor d-p hopping is included.

Figures 8(b)–8(d) show our results for a small system with only one primitive cell, where ED is feasible. As before, the cluster SP results match well with the ED ones for all three observables studied in this work, while the single-site SP method causes much larger errors. We note that while a larger system is very straightforward to treat with cluster SP simply by using translational symmetry and k-point sampling, it is much more difficult to find a good exact benchmark for a large 2D system, especially for the three-band model. Such a comparison will be a topic of future work.

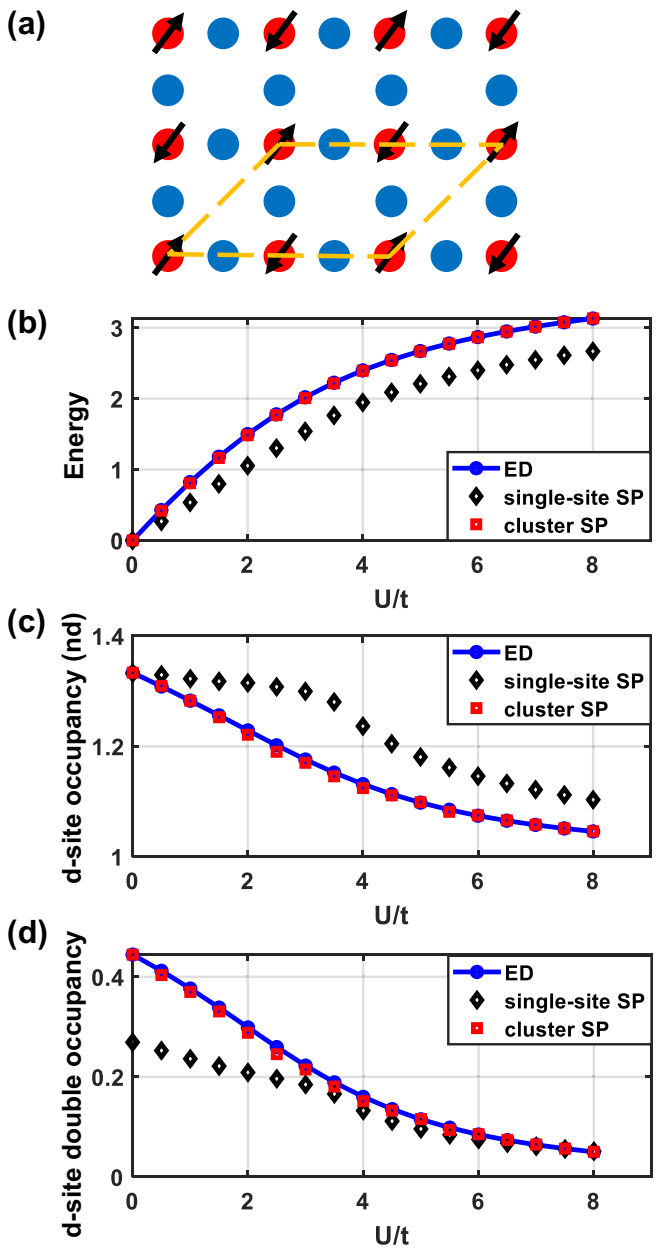

FIG. 8. (a) Illustrates the purposed spin structure in a checkerboard lattice, where the nearly half-filled d sites marked by red circles are spin polarized, while the nearly full-filled p sites marked by blue circles are not spin polarized. The orange dashed parallelogram represents the primitive cell under the Néel AFM correlation. (b)–(d) Show the total energy for each primitive cell, d-site occupancy, and d-site double occupancy versus the interaction strength U, respectively. The results of ED, single-site SP, and cluster SP methods are marked by blue circles, black diamonds, and red squares, respectively.

VII. CONCLUSION

We have introduced a nonlocal slave-particle representation defined on bonds, which improves the slave-particle decomposition by enforcing particle-conserving hoppings on each bond, in addition to the site-centered spinon-slave number matching constraints from previous slave-particle methods. We have further developed a cluster approximation for the interacting slave-particle problem based on the expansion and matching of density matrices which maps the slave-particle problem to a set of small overlapping clusters that can be solved self-consistently.

As a significant improvement to the previous single-site slave-particle theory, our theory correctly predicts the absence of Mott transition in the 1D half-filled Hubbard model (single-site slave-particle methods predict a false Mott transition). The method also shows remarkably high accuracy for a wide range of interaction strengths, unit-cell sizes, doping levels, and in both one and two dimensions when compared to exact or high-accuracy benchmark methods. Computationally, the method is very efficient and requires only a few minutes of CPU time on a serial laptop to find the ground state of the coupled spinon and slave problems. Future work will benchmark this method more extensively in 2D as well as on more complex multiorbital realistic material systems.

ACKNOWLEDGMENT

We thank Alexandru Georgescu for the helpful discussions. This work was supported by NSF DMR 2237469, and NSF XSEDE supercomputing resources TG-MCA08X007.

APPENDIX A: MAPPING AND COMMUTATION RELATIONS

Electrons obey the anticommutation relations

$$\{\hat{c}_\alpha,\ \hat{c}^\dagger_{\alpha'}\} = \delta_{\alpha\alpha'}. \tag{A1}$$

The spinons ($\hat{f}_\alpha$) obey the same commutation relations as well, as they are fermions. The slave-particle lowering operator $\hat{O}$ is then defined to obey the mapping from the original electron Hilbert space to the number-matching states in the enlarged spinon+slave Hilbert space. The mappings in Eqs. (4) and (7) and (4) are made to match all matrix elements. That is, for each bond $\langle\alpha\beta\rangle$, we require

$$\langle n'|\hat{c}_\alpha|n\rangle = \langle n^{f\prime} = n'; N' = n'|\hat{f}_\alpha\hat{O}_{\alpha\langle\alpha\beta\rangle}|n^f = n; N = n\rangle. \tag{A2}$$

Also, the operator $\hat{f}_\alpha\hat{O}_{\alpha\langle\alpha\beta\rangle}$ should still obey anticommutation relations like Eq. (A1):

$$\langle n^{f\prime} = n'; N' = n'|\{\hat{f}_\alpha\hat{O}_{\alpha\langle\alpha\beta\rangle},\ \hat{f}^\dagger_{\alpha'}\hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle}\} \times |n^f = n; N = n\rangle = \delta_{\alpha\alpha'}. \tag{A3}$$

In the following, we are going to prove that the most general form of $\hat{O}_{\alpha\langle\alpha\beta\rangle}$ shown in Eq. (5) always obeys the two requirements in Eqs. (A2) and (A3).

For the right side of Eq. (A2), the matrix element can be rewritten as the product of two matrix elements from spinon and slave, separately.

$$\langle n^{f\prime} = n'; N' = n'|\hat{f}_\alpha\hat{O}_{\alpha\langle\alpha\beta\rangle}|n^f = n; N = n\rangle = \langle n^{f\prime} = n'|\hat{f}_\alpha|n^f = n\rangle\langle N' = n'|\hat{O}_{\alpha\langle\alpha\beta\rangle}|N = n\rangle. \tag{A4}$$

Now we consider all the cases. When $n_\alpha = 0$, both $\hat{f}_\alpha$ and $\hat{c}_\alpha$ kill (zero) the matrix element on both sides of Eq. (A2). When $n_\alpha = 1$, since $\hat{O}_{\alpha\langle\alpha\beta\rangle}|N_\alpha = 1\rangle = |N_\alpha = 0\rangle$, the only nonzero

spinon and slave matrix elements come from $n'_\alpha = 0$ and $n'_\gamma = n_\gamma$ for other modes $\gamma \neq \alpha$. We end up with $\langle n^{f\prime} = n'|\hat{f}_\alpha|n^f = n\rangle$ for the right side of Eq. (A2). Realizing that $\hat{f}_\alpha$ and $\hat{c}_\alpha$ act identically as same fermionic annihilation operator on identical Hilbert spaces, their matrix elements are the same, so Eq. (A2) always holds.

For Eq. (A3), using the anticommutation relations of the spinons, we get

$$\{\hat{f}_\alpha \hat{O}_{\alpha\langle\alpha\beta\rangle}, \hat{f}^\dagger_{\alpha'} \hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle}\} = \hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle} \hat{O}_{\alpha\langle\alpha\beta\rangle} \delta_{\alpha\alpha'} + \hat{f}_\alpha \hat{f}^\dagger_{\alpha'} [\hat{O}_{\alpha\langle\alpha\beta\rangle}, \hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle}]. \quad \text{(A5)}$$

The two terms on the right side of Eq. (A5) are discussed case by case.

When $\alpha = \alpha'$, using Eq. (5) we have

$$\hat{O}^\dagger_{\alpha\langle\alpha\beta'\rangle} \hat{O}_{\alpha\langle\alpha\beta\rangle} = \begin{pmatrix} c^*_{\alpha\langle\alpha\beta'\rangle} c_{\alpha\langle\alpha\beta\rangle} & 0 \\ 0 & 1 \end{pmatrix} \quad \text{(A6)}$$

and

$$[\hat{O}_{\alpha\langle\alpha\beta\rangle}, \hat{O}^\dagger_{\alpha\langle\alpha\beta'\rangle}] = (1 - c^*_{\alpha\langle\alpha\beta'\rangle} c_{\alpha\langle\alpha\beta\rangle}) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad \text{(A7)}$$

We substitute this into the matrix element on the left side of Eq. (A3), and we discuss how the anticommutator acts on the physical ket state $|n^f = n; N = n\rangle$. If $n_\alpha = 0$ in the physical state, then the action $\hat{f}_\alpha \hat{f}^\dagger_\alpha$ is the identity operation on this state. The two terms in Eq. (A5) add to $\hat{O}_{\alpha\langle\alpha\beta\rangle} \hat{O}^\dagger_{\alpha\langle\alpha\beta'\rangle}$ which acts as identity on for this ket state. If $n_\alpha = 1$ in the physical state, the second term of Eq. (A5) is zero due to the zeroing action of $\hat{f}_\alpha \hat{f}^\dagger_\alpha$, and the remaining $\hat{O}^\dagger_{\alpha\langle\alpha\beta'\rangle} \hat{O}_{\alpha\langle\alpha\beta\rangle}$ acts as identity on this state. Thus, Eq. (A3) holds for $\alpha = \alpha'$.

When $\alpha \neq \alpha'$, the first term in Eq. (A5) is zero. For the second term, with basis ordered as $\{|N_\alpha = 0; N_{\alpha'} = 0\rangle, |N_\alpha = 0; N_{\alpha'} = 1\rangle, |N_\alpha = 1; N_{\alpha'} = 0\rangle, |N_\alpha = 1; N_{\alpha'} = 1\rangle\}$, we have

$$\hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle} \hat{O}_{\alpha\langle\alpha\beta\rangle} = \begin{pmatrix} 0 & 0 & 0 & c^*_{\alpha'\langle\alpha'\beta'\rangle} \\ 0 & 0 & c^*_{\alpha'\langle\alpha'\beta'\rangle} c_{\alpha\langle\alpha\beta\rangle} & 0 \\ 0 & 1 & 0 & 0 \\ c_{\alpha\langle\alpha\beta\rangle} & 0 & 0 & 0 \end{pmatrix} = \hat{O}_{\alpha\langle\alpha\beta\rangle} \hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle}. \quad \text{(A8)}$$

Thus, the commutator vanishes, Eq. (A5) is zero in this case, and Eq. (A3) holds for $\alpha \neq \alpha'$ too.

Based on the above, the commutation relations of the slave-bond operators are

$$[\hat{O}_{\alpha\langle\alpha\beta\rangle}, \hat{O}^\dagger_{\alpha'\langle\alpha'\beta'\rangle}] = \delta_{\alpha\alpha'} (1 - c_{\alpha\langle\alpha\beta'\rangle} c_{\alpha\langle\alpha\beta\rangle}) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad \text{(A9)}$$

This commutation relation means that the "lowering" operators $\hat{O}$ are similar to but are *not* the same as bosonic field lowering operators. They are simply defined to obey the mappings of Eq. (A2) and the anticommutation relations of Eq. (A3). Hence, we call them slave-particle (as opposed to slave-boson) operators.

APPENDIX B: TWO-SITE HUBBARD MODEL

In this Appendix, we derive analytical results for the two-site half-filled Hubbard model using three different methods: Exact diagonalization of the original fermion Hamiltonian, the single-site slave-particle method, and our cluster slave-bond method.

1. Exact diagonalization

The ground state of the half-filled two-site Hubbard model will be a linear combination of the doubly occupied states and the singlet state, so the basis of the subspace is $|\uparrow\downarrow, 0\rangle, |0, \uparrow\downarrow\rangle, (|\uparrow, \downarrow\rangle - |\downarrow, \uparrow\rangle)/\sqrt{2}$, where the Hamiltonian is represented in this basis as

$$\hat{H} = \begin{pmatrix} U & 0 & -\sqrt{2}t \\ 0 & U & -\sqrt{2}t \\ -\sqrt{2}t & -\sqrt{2}t & 0 \end{pmatrix}. \quad \text{(B1)}$$

The ground state of this Hamiltonian is

$$|G\rangle = A[4t(|\uparrow\downarrow, 0\rangle + |0, \uparrow\downarrow\rangle) + (U + \sqrt{U^2 + 16t^2})(|\uparrow, \downarrow\rangle - |\downarrow, \uparrow\rangle)], \quad \text{(B2)}$$

where A is the normalization factor. The corresponding ground-state energy is $E = (U - \sqrt{U^2 + 16t^2})/2$ and the double occupancy on each site is

$$D_i \equiv \langle \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rangle = \left(1 - \frac{U}{\sqrt{U^2 + 16t^2}}\right) \Big/ 4. \quad \text{(B3)}$$

Hence, the energy is $-2t$ in the noninteracting limit, and becomes $-4t^2/U$ in the large interaction limit. The double occupancy is 0.25 and $-2t^2/U^2$ in the same two limits. So there is not a Mott transition for any finite U based on these exact results. The analytical results of ground-state energy and double occupancy derived above match with the numerical results in Fig. 3.

2. Single-site slave particle

In the single-site slave-particle method, the total energy is defined as

$$E = -t \sum_\sigma (\langle \hat{f}^\dagger_{1\sigma} \hat{f}_{2\sigma} \rangle \langle \hat{O}^\dagger_{1\sigma} \rangle \langle \hat{O}_{2\sigma} \rangle + \text{c.c.}) + U \sum_i \langle \hat{N}_{i\uparrow} \hat{N}_{i\downarrow} \rangle. \quad \text{(B4)}$$

To minimize the total energy, the corresponding spinon and the first-site slave-particle Hamiltonian is

$$\begin{aligned} \hat{H}_f &= -t \sum_\sigma (\langle \hat{O}^\dagger_{1\sigma} \rangle \langle \hat{O}_{2\sigma} \rangle \hat{f}^\dagger_{1\sigma} \hat{f}_{2\sigma} + \text{H.c.}) + \sum_{i\sigma} h'_{i\sigma} \hat{N}_{i\sigma}, \\ \hat{H}_{s1} &= -t \sum_\sigma (\langle \hat{f}^\dagger_{1\sigma} \hat{f}_{2\sigma} \rangle \langle \hat{O}_{2\sigma} \rangle \hat{O}^\dagger_{1\sigma} + \text{H.c.}) \\ &\quad + U \hat{N}_{1\uparrow} \hat{N}_{1\downarrow} + \sum_\sigma h_{1\sigma} \hat{N}_{1\sigma}, \end{aligned} \quad \text{(B5)}$$

where h are Lagrange multipliers, while h' combine the Lagrange multipliers and the symmetry-breaking fields to be determined variationally. (The second slave-site Hamiltonian $\hat{H}_{s2}$ is identical to the first-site one other than relabeling.) According to the spin symmetry and inversion symmetry in this system, we conclude that $h'_{i\sigma}$ should be the same for all sites and spins, and the same statement also holds for $h_{i\sigma}$. Based on this symmetry analysis, the spinon Hamiltonian is

solved easily by diagonalizing 2×2 matrices for each spin channel. As a result, the occupancy is $\langle n_{i\sigma}\rangle = \frac{1}{2}$ for every site and spin, and the spinon density matrix element of interest is $\langle \hat{f}^\dagger_{1\sigma}\hat{f}_{2\sigma}\rangle = \frac{1}{2}$.

Next, in the basis $|0\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle$, we have

$$\hat{H}_{s1} = \begin{pmatrix} 0 & -y & -y & 0 \\ -y & h & 0 & -y \\ -y & 0 & h & -y \\ 0 & -y & -y & 2h+U \end{pmatrix}, \tag{B6}$$

where we set $h_{i\sigma} = h$ based on the symmetry discussed above, and $y \equiv 0.5t(1+c)\langle O_{2\sigma}\rangle$, in which c is the real c-gauge number determined to make $\langle O_{i\sigma}\rangle = 1$ in the noninteracting limit.

When $U = 0$, $h = 0$ and the ground state of the slave-particle Hamiltonian is

$$|G_0\rangle = (|0\rangle + |\uparrow\rangle + |\downarrow\rangle + |\uparrow\downarrow\rangle)/2.$$

In order to have the spinon hopping renormalization factor $\langle O_{i\sigma}\rangle = 1$ with this state, the c-gauge number is $c = 1$, and thus $y = t\langle O_{2\sigma}\rangle$.

For a finite interaction $U/t > 0$, to obey the particle-number-matching constraint in Eq. (9), we find that $h = -U/2$. So the ground state with finite interaction is

$$|G\rangle = (|0\rangle + a|\uparrow\rangle + a|\downarrow\rangle + |\uparrow\downarrow\rangle)/\sqrt{2(1+a^2)},$$

where $a \equiv \sqrt{(8t+U)/(8t-U)}$ when $U < 8t$, while $a \to \infty$ when $U \geqslant 8t$. Finally, using the single-site slave-particle method, when $U < 8t$, the ground-state energy is $E = -(U-8t)^2/32t$, and the double occupancy is $D_i = (8t-U)/32t$, while when $U \geqslant 8t$, both ground-state energy and double occupancy are zero, which match with the numerical results in Fig. 3. This result clearly indicates a false Mott transition at $U = 8t$ in contrast to the exact diagonalization method.

3. Cluster slave-particle method

In the cluster slave-particle method, the total energy is defined as

$$E = -t\sum_\sigma (\langle \hat{f}^\dagger_{1\sigma}\hat{f}_{2\sigma}\rangle\langle \hat{O}^\dagger_{1\sigma}\hat{O}_{2\sigma}\rangle + \text{c.c.}) + U\sum_i \langle \hat{N}_{i\uparrow}\hat{N}_{i\downarrow}\rangle, \tag{B7}$$

where we hide the bond index for $\hat{O}$ operators because there is only one bond, and the c gauge for spin up and spin down are the same by symmetry in this system. To minimize the total energy, using a two-site cluster, the corresponding spinon and the cluster slave-particle Hamiltonian are

$$\begin{aligned} \hat{H}_f =& -t\sum_\sigma \left(\langle \hat{O}^\dagger_{1\sigma}\hat{O}_{2\sigma}\rangle \hat{f}^\dagger_{1\sigma}\hat{f}_{2\sigma} + \text{H.c.}\right) + \sum_{i\sigma} h'_{i\sigma}\hat{N}_{i\sigma}, \\ \hat{H}_C =& -t\sum_\sigma \left(\langle \hat{f}^\dagger_{1\sigma}\hat{f}_{2\sigma}\rangle \hat{O}^\dagger_{1\sigma}\hat{O}_{2\sigma} + \text{H.c.}\right) \\ &+ \sum_i \left(U\hat{N}_{i\uparrow}\hat{N}_{i\downarrow} + \sum_\sigma h_{i\sigma}\hat{N}_{i\sigma}\right). \end{aligned} \tag{B8}$$

Similarly, h are Lagrange multipliers, while h' combine the Lagrange multipliers and the symmetry-breaking fields to be determined variationally. The spin symmetry and inversion symmetry also hold here, and the spinon Hamiltonian can be solved easily as before: We find $\langle n_{i\sigma}\rangle = \frac{1}{2}$ and $\langle \hat{f}^\dagger_{1\sigma}\hat{f}_{2\sigma}\rangle = \frac{1}{2}$.

FIG. 9. An illustration of the one-dimensional d-p chain, where red atoms represent d sites while blue atoms represent p sites, and each atom is labeled by white letters. The black dashed ellipses are the clusters used in the cluster slave-bond calculation.

The noninteracting slave Hamiltonian can be separated by spin channels, where the spin σ Hamiltonian is represented in the basis ordered as $|0,0\rangle, |\sigma,0\rangle, |0,\sigma\rangle, |\sigma,\sigma\rangle$:

$$\hat{H}_{C\sigma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h & -y & 0 \\ 0 & -y & h & 0 \\ 0 & 0 & 0 & 2h \end{pmatrix}, \tag{B9}$$

where $y = 0.5t(1-C^2)$, and C stands for the gauge to be determined and is assumed to be real. We find that, in order to have the noninteracting spinon hopping renormalization factor $\langle \hat{O}^\dagger_{1\sigma}\hat{O}_{2\sigma}\rangle = 1$ and the particle number match with the spinon result, the c-gauge number must be $C = \sqrt{3}$.

With a finite interaction U, matching particle numbers give the final slave-particle ground state

$$\begin{aligned} |G\rangle =& A[-4t(|\uparrow\downarrow,0\rangle + |0,\uparrow\downarrow\rangle) \\ &+ (U + \sqrt{U^2+16t^2})(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)], \end{aligned} \tag{B10}$$

where A is the normalization factor. The corresponding renormalization factor is $\langle \hat{O}^\dagger_{1\sigma}\hat{O}_{2\sigma}\rangle = 4t/\sqrt{U^2+16t^2}$. In other words, the intersite hopping (and thus renormalization factors) of the slave bond is always nonzero in this system which indicates an absence of a Mott transition. Finally, observables can be calculated using the slave-particle ground-state expression (B10). The ground-state energy given by Eq. (B7) is $E = (U - \sqrt{U^2+16t^2})/2$. The double occupancy on each site is

$$D_i \equiv \langle \hat{N}_{i\uparrow}\hat{N}_{i\downarrow}\rangle = \left(1 - \frac{U}{\sqrt{U^2+16t^2}}\right)\Big/4, \tag{B11}$$

where the double occupancy is always nonzero for arbitrary U/t. Finally, note that the expressions of ground-state energy and double occupancy are exactly the same as the exact diagonalization results derived above in Appendix B 1.

APPENDIX C: FORMALISM FOR c GAUGE

In this Appendix, we derive the analytical formulas of Eqs. (C1) and (C2) for the c-gauge numbers. As we introduced in Sec. III, the c gauges are chosen (i) to forbid all unphysical intracluster hoppings in slave problem that change particle numbers, and (ii) to ensure that the spinon problem recovers the original fermion problem in the noninteracting limit.

1. Intracluster hoppings

By definition, both sites of an intracluster hopping can be found in one same cluster. Note that even though the nearest-neighbor hopping such as the d_2-p_2 hopping in Fig. 9

can appear as an outward hopping with respect to the cluster $[d_1p_1d_2]$, it is still treated as a short-ranged intracluster hopping, whose c gauge is computed in cluster $[d_2p_2d_3]$. For intracluster hoppings, the first constraint (i) requires $c_{\alpha\langle\alpha\beta\rangle} + c_{\beta\langle\alpha\beta\rangle} = 0$, and the second (ii) requires $\langle\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\hat{O}_{\beta\langle\alpha\beta\rangle}\rangle = 1$ in the noninteracting calculation. For each bond $\langle\alpha\beta\rangle$, there are two c-gauge variables to be determined, obeying the above two equations. By solving the two equations, we find

$$c_{\alpha\langle\alpha\beta\rangle} = \sqrt{1 - \frac{1}{\langle 0,1|\hat{\rho}^0_{\alpha\beta}|1,0\rangle}}, \tag{C1}$$

and $\hat{\rho}^0_{\alpha\beta}$ is a noninteracting bond density matrix, where the 0 and 1 in the bras and kets are the occupation numbers on the slave modes α and β, respectively. The superscript 0 represents the operators in noninteracting slave-particle calculations. The bond density matrix is defined by tracing out other degrees of freedom in the cluster density matrix $\hat{\rho}^0_{\alpha\beta} \equiv \mathrm{Tr}_{\gamma\in\mathcal{C}|\gamma\neq\alpha,\beta}(\hat{\rho}^0_{\mathcal{C}})$.

2. Intercluster hoppings

Aside from the intracluster hoppings, the other hoppings are long-ranged intercluster hoppings, such as the p_1-p_2 hopping in Fig. 9. Based on the density matrix expansion in Eq. (15), an intercluster hopping $\langle\alpha\beta\rangle$ in our slave-particle approach is decoupled as $\langle\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\hat{O}_{\beta\langle\alpha\beta\rangle}\rangle = \langle\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\rangle\langle\hat{O}_{\beta\langle\alpha\beta\rangle}\rangle$. The constraints on the bond $\langle\alpha\beta\rangle$ are $\langle\hat{O}_{\alpha\langle\alpha\beta\rangle}\rangle_{\rho^0} = \langle\hat{O}_{\beta\langle\alpha\beta\rangle}\rangle_{\rho^0} = 1$ in order to recover the noninteracting Hamiltonian with the spinons alone (ρ^0 is the noninteracting slave density matrix). Each of these constraints is easily solved to get

$$c_{\alpha\langle\alpha\beta\rangle} = \frac{1}{\langle 0|\hat{\rho}^0_\alpha|1\rangle} - 1, \tag{C2}$$

where 0 and 1 are the occupation numbers of the slave mode α. The single-orbital density matrix is determined as

$$\hat{\rho}^0_\alpha \equiv \frac{1}{M_\alpha}\sum_{\mathcal{C}|\alpha\in\mathcal{C}} \mathrm{Tr}_{\gamma\in\mathcal{C}|\gamma\neq\alpha}(\hat{\rho}_{\mathcal{C}}), \tag{C3}$$

where M_α is the number of clusters overlapping on the slave mode α.

APPENDIX D: REDUNDANCY OF SINGLE-SITE DENSITY MATRIX CONSTRAINTS

In this Appendix, we describe in detail the constraints ensuring that the single-site density matrix $\hat{\rho}_i$ at a shared site i among multiple clusters is described consistently among them. As explained in the main text, these constraints are imposed by the Lagrange multipliers $\hat{\Lambda}^{(\mathcal{C})}_i$ and associated energy terms in the function F defined in Eq. (22) that are minimized. The key result of this Appendix is that all these Lagrange multipliers $\hat{\Lambda}^{(\mathcal{C})}_i$ can be set to zero.

First, there are some redundant Lagrange multipliers in the above set of constraints. For a site i, there are M_i Lagrange multiplier matrices $\hat{\Lambda}^{(\mathcal{C})}_i$ since there are M_i distinct clusters overlapping on the site i. The task of the $\hat{\Lambda}^{(\mathcal{C})}_i$ is to make all the site density matrices equal to each other (and equal to $\hat{\rho}_i$), $\hat{\rho}^{(\mathcal{C})}_i = \hat{\rho}_i$. This means that we only require $M_i - 1$ distinct Lagrange multiplier matrices at site i, and we can pick one of the clusters to be a reference cluster $\mathcal{C}^*_i$ whose single-site density matrix the other clusters must match. For example, if a site i is not shared by multiple clusters (so $M_i = 1$), then no Lagrange multipliers should be added to match the single-site density matrix to itself; if $M_i = 2$, then we need only match the density matrix of the second cluster with the first. We can safely set the Lagrange multiplier matrix for our reference cluster to zero, $\hat{\Lambda}^{(\mathcal{C}^*_i)}_i = 0$.

Second, the density matrix matching constraints are obeyed throughout the entire minimization process. This means that not only $\hat{\rho}^{(\mathcal{C})}_i = \hat{\rho}_i$ at the beginning of the minimization, but also that $\delta\hat{\rho}^{(\mathcal{C})}_i = \delta\hat{\rho}_i$ when minimizing the function F along the gradient. The gradient of F along $\hat{\rho}_{\mathcal{C}}$ is

$$\begin{aligned}\frac{\partial F}{\partial\hat{\rho}_{\mathcal{C}}} = &-\sum_{\alpha\beta|\langle\alpha\beta\rangle\in\mathcal{C}} t_{\alpha\beta}\langle\hat{f}^\dagger_\alpha\hat{f}_\beta\rangle_{\rho_f}\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\hat{O}_{\beta\langle\alpha\beta\rangle} \\ &-\sum_{\substack{\beta|\beta\in\mathcal{C}\\ \alpha|\langle\alpha\beta\rangle\notin\forall\,\mathcal{C}'}}\frac{t_{\alpha\beta}}{M_\beta}\langle\hat{f}^\dagger_\alpha\hat{f}_\beta\rangle_{\rho_f}[\langle\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\rangle_{\rho_{i|\alpha\in i}}\hat{O}_{\beta\langle\alpha\beta\rangle} + \mathrm{H.c.}] \\ &+\sum_{i|i\in\mathcal{C}}\frac{1}{M_i}\hat{H}^{\mathrm{int}}_i + \sum_{\alpha|\alpha\in\mathcal{C}} h^{(\mathcal{C})}_\alpha\hat{N}_\alpha - \epsilon_{\mathcal{C}}\hat{I}_{\mathcal{C}} \\ &+\sum_{i|i\in\mathcal{C}}(1-1/M_i)\hat{\Lambda}^{(\mathcal{C})}_i.\end{aligned} \tag{D1}$$

We collect the derivatives versus all N clusters into the gradient $\nabla F = (\partial F/\partial\hat{\rho}_{\mathcal{C}_1}, \partial F/\partial\hat{\rho}_{\mathcal{C}_2}, \ldots, \partial F/\partial\hat{\rho}_{\mathcal{C}_N})$. As is standard in the Lagrange multiplier method, we will project out the part of ∇F that breaks the constraints to determine the value of the Lagrange multipliers. This is most easily done by considering a step of size η along the gradient $(\delta\hat{\rho}_{\mathcal{C}_1}, \delta\hat{\rho}_{\mathcal{C}_2}, ..., \delta\hat{\rho}_{\mathcal{C}_N}) = \eta\nabla F$, so the change $\delta\hat{\rho}^{(\mathcal{C})}_i$ is given by

$$\delta\hat{\rho}^{(\mathcal{C})}_i = \mathrm{Tr}_{\mathcal{C}-i}(\delta\hat{\rho}_{\mathcal{C}}) = \eta\mathrm{Tr}_{\mathcal{C}-i}\left(\frac{\partial F}{\partial\hat{\rho}_{\mathcal{C}}}\right), \tag{D2}$$

where, again, $\mathrm{Tr}_{\mathcal{C}-i}$ denotes a trace over the cluster $\mathcal{C}$ but excluding the degrees of freedom on site i. The key enabling observation allowing us to move forward is that the hopping operators $\hat{O}$, $\hat{O}^\dagger$ are traceless. Hence, the intracluster hopping terms in Eq. (D1) involve hoppings between different sites so they have zero traces, the intercluster hopping terms are zero unless the state β is on the site i, and the remaining terms are local single-site operators. At this stage, we have

$$\begin{aligned}\delta\hat{\rho}^{(\mathcal{C})}_i = \eta\Bigg(&-\sum_{\substack{\beta|\beta\in i\\ \alpha|\langle\alpha\beta\rangle\notin\forall\,\mathcal{C}'}}\frac{t_{\alpha\beta}}{M_\beta}\langle\hat{f}^\dagger_\alpha\hat{f}_\beta\rangle_{\rho_f}[\langle\hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle}\rangle_{\rho_{k|\alpha\in k}}\hat{O}_{\beta\langle\alpha\beta\rangle} \\ &+\mathrm{H.c.}] + \frac{\hat{H}^{\mathrm{int}}_i}{M_i} + \sum_{\alpha|\alpha\in i} h^{(\mathcal{C})}_\alpha\hat{N}_\alpha - \epsilon_{\mathcal{C}}\hat{I}_i \\ &+(1-1/M_i)\hat{\Lambda}^{(\mathcal{C})}_i + q^{(\mathcal{C})}_i\hat{I}_i\Bigg)\prod_{j\in\mathcal{C}|j\neq i} S_j,\end{aligned} \tag{D3}$$

where $q_i^{(\mathcal{C})}$ is a potentially cluster-dependent number coming from traces of local operators not on site i, and S_j is the dimension of the single-site density matrix $S_j \equiv \mathrm{Tr}_j(\hat{I}_j)$ (i.e., dimension of the Hilbert space on site j). A moment's reflection shows that the first two terms in Eq. (D3) are independent of the cluster $\mathcal{C}$, and we combine them into an operator $\hat{X}_i$. Upon creating constants $r_i^{(\mathcal{C})} = \epsilon_C + q_i^{(\mathcal{C})}$ and $\zeta = \eta \prod_{j\in\mathcal{C}|j\neq i} S_j$, we have

$$\delta\hat{\rho}_i^{(\mathcal{C})} = \zeta\left(\hat{X}_i + \sum_{\alpha|\alpha\in i} h_\alpha^{(\mathcal{C})}\hat{N}_\alpha + (1-1/M_i)\hat{\Lambda}_i^{(\mathcal{C})} + r_i^{(\mathcal{C})}\hat{I}_i\right). \tag{D4}$$

As per Eq. (17), we sum this $\delta\hat{\rho}_i^{(\mathcal{C})}$ over all clusters $\mathcal{C}$ containing site i and divide by M_i to find an analogous expression for $\delta\hat{\rho}_i$:

$$\delta\hat{\rho}_i = \zeta\left(\hat{X}_i + \sum_{\alpha|\alpha\in i} \bar{h}_\alpha\hat{N}_\alpha + (1-1/M_i)\overline{\hat{\Lambda}_i} + \bar{r}_i\hat{I}_i\right), \tag{D5}$$

where overbars mean averaging over the clusters. Equating Eqs. (D4) and (D5) yields

$$\hat{\Lambda}_i^{(\mathcal{C})} = \overline{\hat{\Lambda}_i} + (1-1/M_i)^{-1}\sum_{\alpha|\alpha\in i}\left(\bar{h}_\alpha - h_\alpha^{(\mathcal{C})}\right)\hat{N}_\alpha + (1-1/M_i)^{-1}\left(\bar{r}_i - r_i^{(\mathcal{C})}\right)\hat{I}_i. \tag{D6}$$

The second and third terms above are either redundant or irrelevant: The second term involves shifts of onsite state energies on site i which can be absorbed into the $h_\alpha^{(\mathcal{C})}$ Lagrange multipliers that enforce spinon and slave-particle number matching, while the third term represents a shift of the cluster Hamiltonian $\hat{H}_\mathcal{C}$ by a constant which does not change its eigenfunctions, shifts all eigenenergies by the same amount, and thus does not change the thermal density matrix $\hat{\rho}_\mathcal{C}$ computed from the Boltzmann distribution. Thus, the only remaining meaningful term is $\hat{\Lambda}_i^{(\mathcal{C})} = \overline{\hat{\Lambda}_i}$ which is cluster independent: Since we have a reference cluster for which $\Lambda_i^{\mathcal{C}_i^*} = 0$, all the $\hat{\Lambda}_i^{(\mathcal{C})} = 0$.

Hence, we conclude that the mean particle-number-matching constraints between spinon and slave particles are sufficient for describing overlapping clusters. Additional constraints, beyond the automatic mean particle-number matching, are redundant.

APPENDIX E: TESTS ON SINGLE-BAND HUBBARD MODELS

In this Appendix, we show our benchmarking results for 1D and 2D half-filled single-band Hubbard models, i.e., when there is one correlated orbital at each site. This type of model is a "d"-only model when describing a transition-metal oxide since the oxygen orbitals are removed from the model. *A priori* we expect this type of model to be less accurate than the d-p models studied in the main text for at least two reasons: (i) the removal of degrees of freedom can be accommodated by renormalizing the remaining parameters of the model (e.g., nearest-neighbor hopping element t), but this type of process is always approximate, and (ii) removal of the oxygen p orbitals gives rise to effective orbitals (e.g., Wannier states) on the d sites that are much less localized than the starting orbitals of the d-p model, so that using an onsite only form of the interaction is much less justified for the d-only case than for the d-p case (i.e., the form of the Hamiltonian is less accurate in the d only case with only onsite interactions). However, the d-only model is a standard benchmark in the field so comparisons to it are still helpful for connecting to prior literature.

As is standard for one-band Hubbard models [57], the nearest-neighbor hopping strength is set to be $-t$, while all other hoppings are set to zero. The onsite energy of each d site is set to zero, and the local onsite interaction strength is set to U. In addition, a small temperature of $k_BT = 5\times10^{-3}t$ is introduced in the cluster slave-particle calculations in an identical manner to that for the d-p models in Sec. VI. (From a practical viewpoint, to simulate the d-only systems without having to modify our d-p software implementation, we choose a d-p model where there are no hoppings involving the p orbitals which have a very low energy and are filled with electrons and thus are completely inert, while direct d-d hoppings $-t$ are introduced for the remaining electrons in the d manifold.)

The first set of tests is on a 1D 64-site single-band Hubbard model with periodic boundary conditions, whose lattice is illustrated in Fig. 10(a). The clusters used in the cluster SP calculation are marked by black dashed ellipses and consist of two neighboring d sites and neighboring clusters overlap with each other. For cluster SP calculations with PBC, we assume a translational period of two sites (i.e., a two-site unit cell), while for DMRG benchmarks, we use OBC as discussed in Sec. VI B. Each data point requires only about 10 CPU seconds for the cluster slave-particle calculations but needs about 7 hours for the DMRG calculations.

Although the two-site SP clusters in these tests are smaller than the three-site SP clusters used in d-p model testing in Sec. VI, the two-site clusters are capable of reproducing high-quality results as shown in Fig. 10. Noticing that the double occupancy of both DMRG and cluster SP are finite regardless of interaction strength, we correctly predict the absence of a Mott transition in this 1D model in contrast to the false Mott transition predicted in single-site slave-particle theory and DMFT [19,27–29].

Next, we examine a two-dimensional (64×64)-site one-band Hubbard model at half-filling with PBC whose lattice is illustrated in Fig. 11(a). The two-site clusters marked by black dashed ellipses and the two-site Néel-ordered unit cell marked by the orange dashed parallelogram are chosen in the same manner as for the 2D d-p model of Sec. VI D. Figures 11(b) and 11(c) show the total-energy and double-occupancy results of the cluster SP method. As a comparison, we also show thermodynamic limit results from auxiliary-field quantum Monte Carlo (AFQMC) calculations [58]. Using SP clusters containing only two sites, the cluster SP method shows remarkable quantitative accuracy with a modest computational cost of only about two serial CPU minutes for each data point.

APPENDIX F: FINITE-TEMPERATURE ERROR

Generally, the partial trace of a pure state will be a mixed state. Thus, even if our current cluster SP method only aims

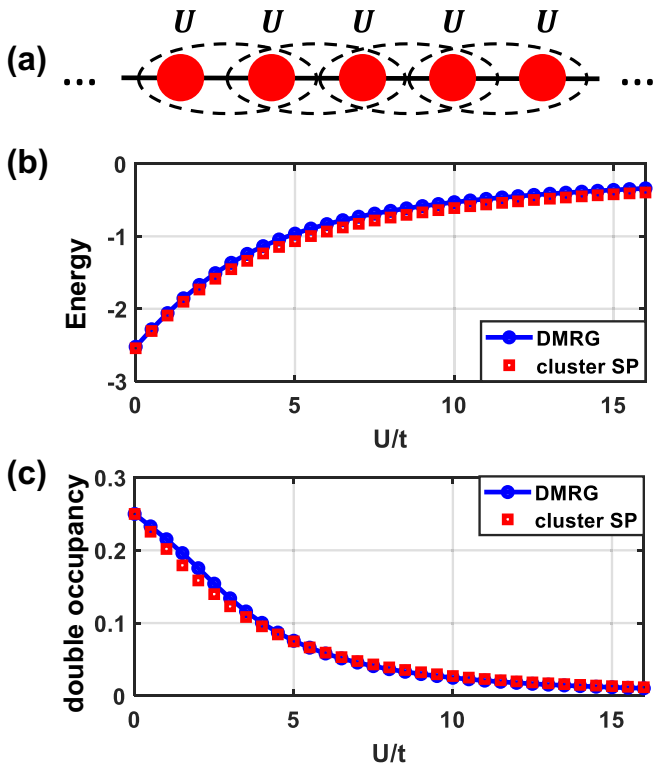

FIG. 10. (a) Illustrates the one-dimensional single-band Hubbard model with periodic boundary conditions. The black dashed ellipses are the clusters used in the cluster slave-particle calculation. (b), (c) Show the total energy per two-site unit cell in units of t, and the double occupancy versus the interaction strength U, respectively. All results are for a 64-site 1D single-band Hubbard system at half-filling: Blue circles represent DMRG results with OBC, while red squares represent cluster SP results assuming a translational period of two sites (PBC).

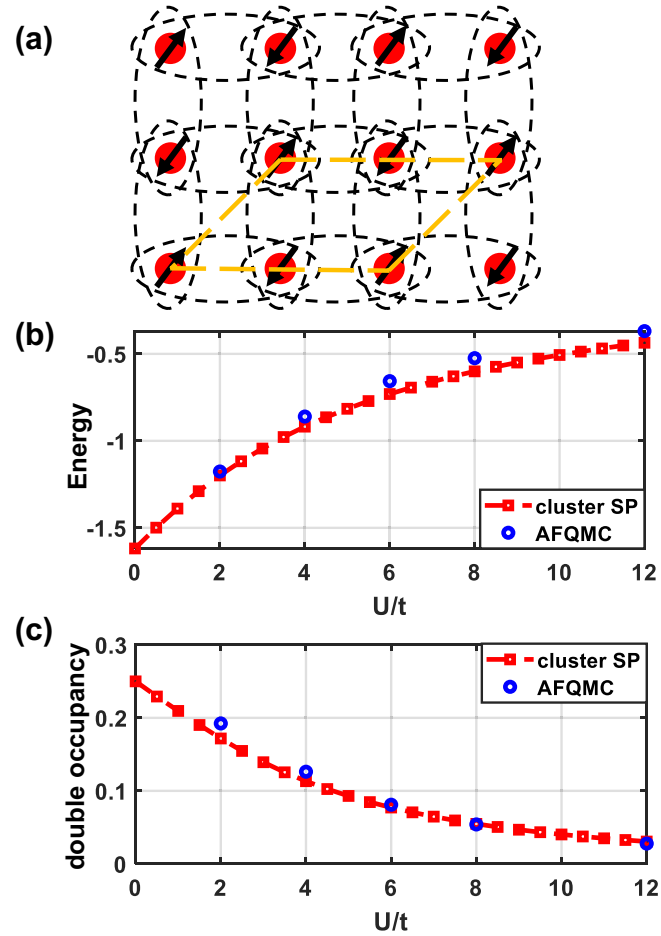

FIG. 11. (a) Illustrates the two-dimensional Hubbard model with periodic boundary conditions. The black dashed ellipses are the clusters used in the cluster slave-particle calculation, and the orange dashed parallelogram represents the two-site primitive cell under the Néel AFM correlation. (b), (c) Show the total energy per two-site unit cell in units of t, and the double occupancy versus the interaction strength U, respectively. The red squares represent cluster SP results generated by a (64×64)-site system at half-filling with PBC, while the blue circles are auxiliary-field quantum Monte Carlo (AFQMC) benchmarks [58] in the thermodynamic limit.

to predict the ground-state (pure-state) properties of the entire (lattice) system, in practice we use a finite temperature in all the slave-particle cluster calculations. This is because the mixed state introduced by the Boltzmann distribution can better capture the cluster's statistical properties.

The use of a small but finite-temperature Boltzmann distribution is a straightforward way to generate a mixed state made from a number of low-energy eigenstates of the cluster Hamiltonian, but the finite temperature itself introduces errors due to the inclusion of higher-energy excited states. Here we show that the errors are quite small in magnitude and easily controllable.

Numerically, the error in the total energy per unit cell has the same order of magnitude as the temperature itself as shown in Fig. 12(a). For example, compared to the benchmark $T = 0$ DMRG results in Fig. 6, the largest error caused by $k_BT = 0.32t$ is $0.28t$, while the largest error caused by $k_BT = 0.16t$ is $0.10t$. For this system, the finite-temperature error in energy gets almost negligible when k_BT is below $0.04t$. Similarly, the d-site occupancy and double occupancy are quite accurate at $k_BT = 0.04t$, as shown in Figs. 12(b) and 12(c). All results in the main text are based on an even lower temperature of $k_BT = 0.005t$, so it is safe to ignore the finite-temperature effect in our results.

From a practical viewpoint, starting from an arbitrary initial setup, a high-temperature calculation typically converges much faster than a low-temperature calculation. Hence, for any given system and set of parameters, we begin with a high-temperature calculation and use its self-consistent converged solution as the starting point for a lower-temperature calculation (and repeat the process for ever lower temperatures), a process that leads to significant computational efficiency.

APPENDIX G: WORKFLOW

In this section, we describe the workflow of a typical slave-particle calculation in Fig. 13. It contains the following steps:

(1) Initialize all parameters, including all the Lagrange multipliers h, c-gauge numbers, hopping renormalization factors $\langle \hat{O}^{\dagger}_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle$ and $\langle \hat{O}_{\alpha\langle\alpha\beta\rangle} \rangle$. When starting from scratch,

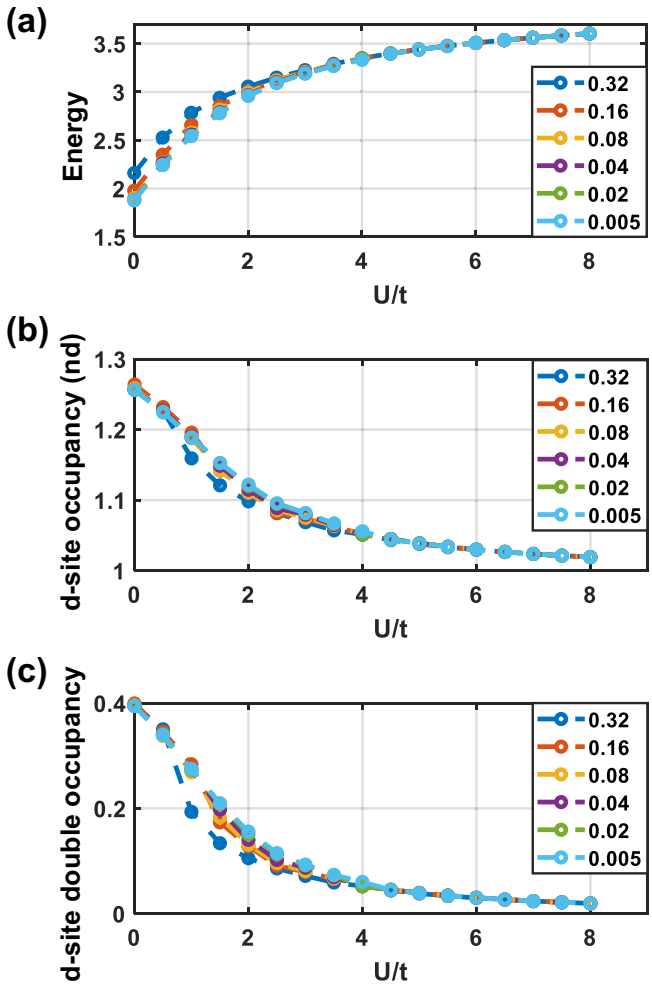

FIG. 12. (a) Total energy per unit cell (units of t), (b) d-site occupancy, and (c) d-site double occupancy versus the interaction strength U, respectively, for a 64-site 1D d-p system with cluster SP method and assuming translational unit cell of four sites. Different curves show results with different temperatures marked by the legend, where the temperatures k_BT are in units of t.

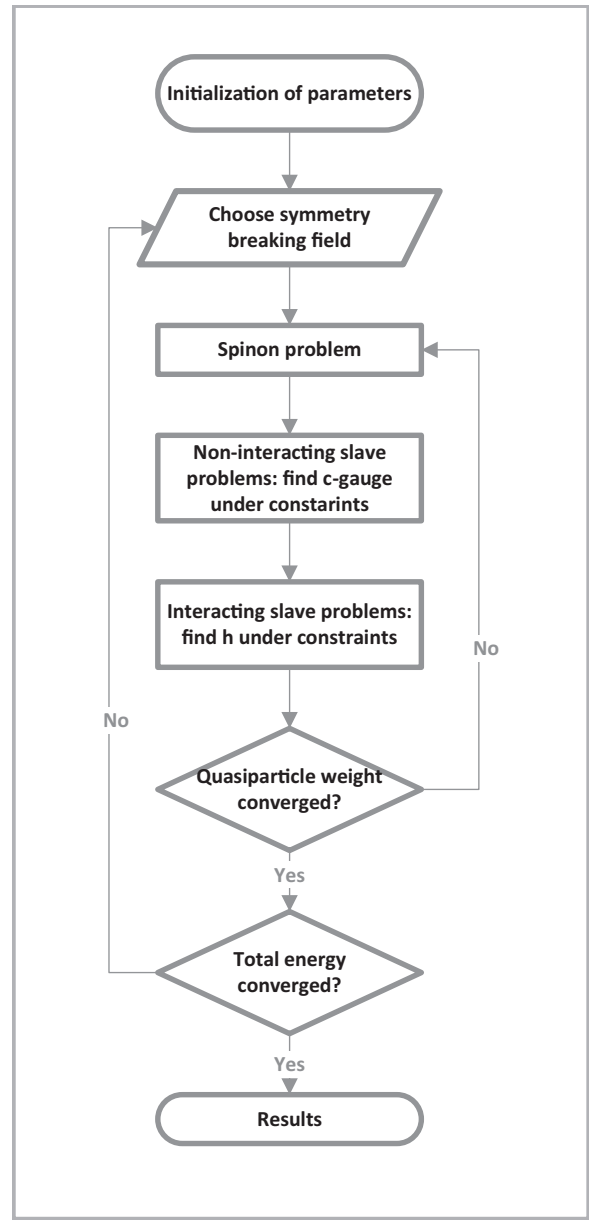

FIG. 13. The workflow of a typical calculation based on the slave-particle approach.

reasonable *a priori* choices are zero for h, and unity for c and the renormalization factors.

(2) For the first iteration, guess some symmetry-breaking field, typically small random numbers much smaller in magnitude than t or U. For the following iterations, choose the symmetry-breaking field variationally by gradient descent of the total energy. The updating of the symmetry-breaking fields is the major outer loop of energy minimization.

(3) Solve the spinon Hamiltonian in Eq. (23), then compute the hopping renormalization factors $\langle \hat{f}^\dagger_\alpha \hat{f}_\beta \rangle$ for use in the slave problem in the next step and the spinon occupation numbers $\langle \hat{n}_\alpha \rangle$.

(4) Solve the noninteracting slave Hamiltonian in Eq. (23), and search for c-gauge number to obey the corresponding constraints. This is most easily done via Newton's method.

(5) Solve the interacting slave Hamiltonian and search for the Lagrange multipliers h under the corresponding constraints. Note that these h will be generally different from the noninteracting h in the previous step, although the noninteracting ones provide a good initial guess when starting from scratch.

(6) If the renormalization factors $\langle \hat{O}^\dagger_{\alpha\langle\alpha\beta\rangle} \hat{O}_{\beta\langle\alpha\beta\rangle} \rangle$ and $\langle \hat{O}_{\alpha\langle\alpha\beta\rangle} \rangle$ differ from the prior step by more than a small convergence tolerance (labeled as Quasiparticle weight in the Figure), go back to step 3 while using the updated renormalization factors, and repeat the calculation until tolerance is achieved.

(7) If the total energy is not converged with respect to the symmetry-breaking field (i.e., the successive change of energy is too large between updates of symmetry-breaking fields), then return to step 2 and update the symmetry-breaking field to minimize the total energy.

We note that methods such as Newton's method or gradient descent require derivatives, and since they are computed numerically by finite differences (absent analytical expressions

for derivatives at present time), in practice we encapsulate inner loops (e.g., the noninteracting slave solver) as routines that are called repeatedly to compute numerical finite-difference derivatives.

Based on the workflow chart above, we can estimate the computational cost needed to complete the full calculation. First, the single-particle spinon Hamiltonian is practically very simple and a variety of methods can be unleashed to find its ground state (direct diagonalization is simplest but cubic in system size while linear-scaling methods are available if needed in the very large system limit [59]). For the more complex cluster slave-particle problem, the computational cost is $\propto S_C N_C N_L$, where S_C is the computational cost of the exact diagonalization of one cluster, N_C is the number of clusters in one unit cell, and N_L is the total number of cluster solutions needed (a loop count). N_L is the product of the number of Lagrange multipliers h, the number of self-consistent steps needed to converge the renormalization factors, and the number of gradient descent steps needed to converge the symmetry-breaking fields. Both N_C and the number of Lagrange multipliers h are linear in the volume of the unit cell while the other two loops count represent intensive quantities (e.g., energy per atom) and are expected to be (relatively) constant with unit-cell size. Hence, the overall slave part of the calculation will scale quadratically with unit-cell volume for the current approach.

[1] H. Bethe, Z. Phys. **71**, 205 (1931).
[2] E. H. Lieb and F.-Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968).
[3] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[4] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).
[5] M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K.-L. Chan, N. Chepiga, P. Helms, C. Jiménez-Hoyos, A. J. Millis, U. Ray, E. Ronca, H. Shi, S. Sorella, E. M. Stoudenmire, S. R. White, and S. Zhang, Phys. Rev. X **10**, 031058 (2020).
[6] A. B. Georgescu, M. Kim, and S. Ismail-Beigi, Comput. Phys. Commun. **265**, 107991 (2021).
[7] S. Barnes, J. Phys. F: Met. Phys. **6**, 1375 (1976).
[8] S. Barnes, J. Phys. F: Met Phys. **7**, 2637 (1977).
[9] M. C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963).
[10] M. C. Gutzwiller, Phys. Rev. **137**, A1726 (1965).
[11] G. Kotliar and A. E. Ruckenstein, Phys. Rev. Lett. **57**, 1362 (1986).
[12] M. Raczkowski, R. Frésard, and A. Oleś, Europhys. Lett. **76**, 128 (2006).
[13] R. Frésard and G. Kotliar, Phys. Rev. B **56**, 12909 (1997).
[14] S. Florens and A. Georges, Phys. Rev. B **70**, 035114 (2004).
[15] J. G. Rau and H.-Y. Kee, Phys. Rev. Lett. **106**, 056405 (2011).
[16] S. Florens and A. Georges, Phys. Rev. B **66**, 165111 (2002).
[17] L. de'Medici, A. Georges, and S. Biermann, Phys. Rev. B **72**, 205124 (2005).
[18] A. B. Georgescu and S. Ismail-Beigi, Phys. Rev. B **92**, 235117 (2015).
[19] A. B. Georgescu and S. Ismail-Beigi, Phys. Rev. B **96**, 165135 (2017).
[20] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[21] F. Lechermann, A. Georges, G. Kotliar, and O. Parcollet, Phys. Rev. B **76**, 155102 (2007).
[22] M. Ferrero, P. S. Cornaglia, L. De Leo, O. Parcollet, G. Kotliar, and A. Georges, Phys. Rev. B **80**, 064501 (2009).
[23] A. Isidori and M. Capone, Phys. Rev. B **80**, 115120 (2009).
[24] R. Yu, J.-X. Zhu, and Q. Si, Phys. Rev. Lett. **106**, 186401 (2011).
[25] W.-H. Ko and P. A. Lee, Phys. Rev. B **83**, 134515 (2011).
[26] B. Lau and A. J. Millis, Phys. Rev. Lett. **110**, 126404 (2013).
[27] T.-H. Lee, T. Ayral, Y.-X. Yao, N. Lanata, and G. Kotliar, Phys. Rev. B **99**, 115129 (2019).
[28] C. J. Bolech, S. S. Kancharla, and G. Kotliar, Phys. Rev. B **67**, 075110 (2003).
[29] D. Zgid, E. Gull, and G. K.-L. Chan, Phys. Rev. B **86**, 165128 (2012).
[30] E. Zhao and A. Paramekanti, Phys. Rev. B **76**, 195101 (2007).
[31] S. Hassan and L. de'Medici, Phys. Rev. B **81**, 035106 (2010).
[32] T. Ayral, T.-H. Lee, and G. Kotliar, Phys. Rev. B **96**, 235139 (2017).
[33] G. Knizia and G. K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).
[34] G. Knizia and G. K.-L. Chan, J. Chem. Theory Comput. **9**, 1428 (2013).
[35] R. Sakuma, P. Werner, and F. Aryasetiawan, Phys. Rev. B **88**, 235110 (2013).
[36] J. M. Tomczak, M. Casula, T. Miyake, and S. Biermann, Phys. Rev. B **90**, 165138 (2014).
[37] M. Kim and B. I. Min, Phys. Rev. B **91**, 205116 (2015).
[38] L. Craco and S. Leoni, Phys. Rev. B **100**, 115156 (2019).
[39] P. Staar, T. Maier, and T. C. Schulthess, Phys. Rev. B **88**, 115101 (2013).
[40] U. R. Hähner, T. A. Maier, and T. C. Schulthess, Phys. Rev. B **101**, 195114 (2020).
[41] Y. Nagai and H. Shinaoka, J. Phys. Soc. Jpn. **88**, 064004 (2019).
[42] M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H. R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998).
[43] M. H. Hettler, M. Mukherjee, M. Jarrell, and H. R. Krishnamurthy, Phys. Rev. B **61**, 12739 (2000).
[44] N. Lanata, Y. Yao, C.-Z. Wang, K.-M. Ho, and G. Kotliar, Phys. Rev. X **5**, 011008 (2015).
[45] J. Zaanen, G. A. Sawatzky, and J. W. Allen, Phys. Rev. Lett. **55**, 418 (1985).
[46] V. Emery, Phys. Rev. Lett. **58**, 2794 (1987).
[47] C. Varma, S. Schmitt-Rink, and E. Abrahams, Solid State Commun. **62**, 681 (1987).
[48] S. R. White and D. J. Scalapino, Phys. Rev. B **92**, 205112 (2015).
[49] Z.-H. Cui, C. Sun, U. Ray, B.-X. Zheng, Q. Sun, and G. K.-L. Chan, Phys. Rev. Res. **2**, 043259 (2020).
[50] A. Chiciak, E. Vitali, and S. Zhang, Phys. Rev. B **102**, 214512 (2020).
[51] P. W. Anderson, Science **235**, 1196 (1987).

[52] F. C. Zhang and T. M. Rice, Phys. Rev. B **37**, 3759 (1988).
[53] M. Fishman, S. R. White, and E. M. Stoudenmire, SciPost Phys. Codebases, 4 (2022)
[54] M. Fishman, S. R. White, and E. M. Stoudenmire, SciPost Phys. Codebases, 4–r0.3 (2022)
[55] B. Bauer and C. Nayak, J. Stat. Mech.: Theory Exp. (2013) P09005.
[56] T. Zhu and G. K.-L. Chan, Phys. Rev. X **11**, 021006 (2021).
[57] F. H. Essler, H. Frahm, F. Göhmann, A. Klümper, and V. E. Korepin, *The One-dimensional Hubbard Model* (Cambridge University Press, Cambridge, 2005).
[58] J. P. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X. W. Liu, A. J. Millis, N. V. Prokofev, M. Qin, G. E. Scuseria, H. Shi, B. V. Svistunov, L. F. Tocchio, I. S. Tupitsyn *et al.* Phys. Rev. X **5**, 041041 (2015).
[59] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).