

# Benchmarks and results of the two-band Hubbard model from the Gutzwiller conjugate gradient minimization theory

Zhuo Ye <sup>1,\*</sup> Feng Zhang <sup>1,2,†</sup> Yong-Xin Yao,<sup>1,2</sup> Cai-Zhuang Wang,<sup>1,2</sup> and Kai-Ming Ho<sup>2</sup>

<sup>1</sup>Ames National Laboratory, U.S. Department of Energy, Ames, Iowa 50011, USA

<sup>2</sup>Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA



(Received 4 April 2023; revised 31 July 2023; accepted 5 September 2023; published 22 September 2023)

Ground-state properties, such as energies and double occupancies, of a one-dimensional two-band Hubbard model are calculated using a first-principles Gutzwiller conjugate gradient minimization theory. The favorable agreement with the results from the density matrix renormalization group theory demonstrates the accuracy of our method. A rotationally invariant approach is further incorporated into the method to greatly reduce the computational complexity with a speedup of approximately 50 times. Moreover, we investigate the Mott transition between a metal and a Mott insulator by evaluating the charge gap. With greatly reduced computational effort, our method reproduces the phase diagram in reasonable agreement with the density matrix renormalization group theory.

DOI: [10.1103/PhysRevB.108.115145](https://doi.org/10.1103/PhysRevB.108.115145)

## I. INTRODUCTION

Accurate description of a large number of interacting electrons remains one of the grand scientific challenges of the present day even with the computer industry's explosive technology growth. The main source of difficulties lies in the fact that the dimension of the Hilbert space required to describe a correlated-electron system grows exponentially in the system size and that the minus sign problem arising from strongly interacting fermions leads to very slow convergence of straightforward Monte Carlo approaches. Numerical studies of most large, realistic systems are affordable only via the Kohn-Sham density function theory (DFT) [1,2]. However, DFT often fails in describing strongly correlated-electron materials, where the behavior of electrons cannot be described effectively in terms of noninteracting entities, because comparatively simple density functional approximations are not sufficiently accurate to describe the strong correlation effect. On the other hand, quantum chemistry approaches that are capable of accurately describing smaller systems are often too expensive for many large, realistic systems. Hybrid approaches that merge DFT with many-body techniques [3–5] have been demonstrated to be powerful and efficient in describing the properties of real correlated-electron materials. However, the use of adjustable screened Coulomb parameters restricts the predictive power of these methods. In order to study realistic condensed matter systems and predict unusual properties that emerge from the correlated behavior of electrons, it is highly desirable to develop *ab initio* methods that are both affordable and reasonably accurate.

We have been developing such a many-body approach, namely, the Gutzwiller conjugate gradient minimization (GCGM) method [6–12]. The GCGM method is based on

the Gutzwiller wave function (GWF) that was proposed by Gutzwiller in the 1960s [13–15]. The GWF is a simple variational wave function that can be defined as a correlation projector acting on a one-particle product state. In GWF-based theories, the Gutzwiller approximation (GA; also proposed by Gutzwiller) is commonly adopted to evaluate the expectation value of an observable with respect to the GWF. However, this approximation “decouples” the two correlated sites by using a site-site factorization and overlooks the correlation between electrons of parallel spins (the exchange hole) [16]. It recovers the exact GWF results only in infinite dimensions [17]. In finite dimensions it could be a major source of the inaccuracy although the use of the approximation greatly improves the computational efficiency [18]. As the strong-coupling counterpart of the GWF, the Baeriswyl wave function (BWF) [19,20] was proposed to describe the insulating states in the limit of large interactions. In the BWF, a projector solely dependent on the kinetic energy term is applied onto the wave function in the large Hubbard  $U$  limit:  $|\Psi_{BWF}\rangle = P_K|\Psi_\infty\rangle$ .  $|\Psi_\infty\rangle$  can be chosen as the ground state of the Heisenberg Hamiltonian or as  $|\Psi_{GWF}\rangle$  obtained in the large  $U$  limit [20,21]. Approximate combinatorial solutions similar to the GA were also developed for the BWF and the hybrid Baeriswyl-Gutzwiller wave function [21]. A comparative study demonstrated that the BWF performs better qualitatively than the GWF in describing the insulating states at large  $U$  [22]. In the GCGM theory, we use the GWF without using the GA and evaluate the expectation value of an observable in a more rigorous way. GCGM also overcomes certain limitations of the GWF itself, which introduces correlations into  $|\Psi_0\rangle$ , the trial wave function, only via an on-site correlation factor. In the case that some intersite correlation is absent in  $|\Psi_0\rangle$ , it will also be missing in the GWF. The GCGM method reinforces a site-site constraint in such situations and has been demonstrated to be successful in the case of Hubbard models [10]. Our previous work has demonstrated the improved accuracy of the GCGM method with benchmark

\*zye@iastate.edu

†fzhang@ameslab.gov

tests of molecules, such as energy calculations of both the ground and the excited states of dimers [7–9]. The benchmark tests on periodic bulk systems have focused mainly on the Hubbard model, both one- and two-dimensional, but with only one band [10]. Since our ultimate goal is to develop a numerical tool to access realistic condensed matter systems, the next benchmark test is naturally the multiband Hubbard model. In this work, we focus on a general two-band Hubbard model and compare our results with those given by density matrix renormalization group theory (DMRG) [23,24], a well-established and powerful many-body approach, especially for one-dimensional (1D) systems. We examine the ground-state properties, including the energy and the double occupancy.

When dealing with multiband/multiorbital systems, a grand challenge is that the computational complexity grows exponentially with the number of orbitals. In the second part, we introduce a rotationally invariant (RI) scheme to greatly reduce the computational cost. In the RI scheme, orbitals that satisfy certain symmetries can be grouped together and the Gutzwiller projector is parametrized based on the number of electrons occupying the group of orbitals instead of the on-site configurations. For example, for a cubic system of which the valence orbitals contain  $p$  orbitals, we may group the  $p_x, p_y, p_z$  orbitals (if they overlap with each other after rotation by  $90^\circ$  about one of the  $x, y, z$  axes) and use the number of electrons that occupy any of the  $p$  orbitals to parametrize the system. As demonstrated with benchmark tests of dimers [7,12], the scheme efficiently groups the on-site orbitals according to their symmetry and greatly reduces the computational complexity without sacrificing accuracy. In the second part of this work, we apply the RI scheme to the same two-band Hubbard model and repeat the calculations of the first part. A speedup of approximately 300 times associated with a favorable agreement with the results given by the original GCGM without applying RI demonstrates the efficiency and the accuracy of the RI scheme. We summarize this work by investigating a Mott transition in this system with GCGM, GCGM+RI, and comparing our results with DMRG.

## II. MODELS AND METHODS

### A. The two-band Hubbard model

The Hubbard model is defined by the Hamiltonian that contains a kinetic energy term and a Coulomb interaction term,

$$\hat{H} = \hat{H}_K + \hat{H}_{\text{Coul}}, \quad (1)$$

where the kinetic term  $\hat{H}_K$  is defined as

$$\hat{H}_K = -t \sum_{\langle IJ \rangle \alpha \sigma} c_{I\alpha\sigma}^\dagger c_{J\alpha\sigma} - \left( t' \sum_{I\sigma} c_{I1\sigma}^\dagger c_{I2\sigma} + \text{H.c.} \right), \quad (2)$$

and the Coulomb interaction term  $\hat{H}_{\text{Coul}}$  is defined as

$$\begin{aligned} \hat{H}_{\text{Coul}} = & U \sum_{I\alpha} n_{I\alpha\uparrow} n_{I\alpha\downarrow} + U' \sum_{I\sigma} n_{I1\sigma} n_{I2\sigma} \\ & + (U' - J) \sum_{I\sigma} n_{I1\sigma} n_{I2\sigma} + J \sum_I (c_{I1\uparrow}^\dagger c_{I2\downarrow}^\dagger c_{I1\downarrow} c_{I2\uparrow} \\ & + c_{I1\uparrow}^\dagger c_{I1\downarrow}^\dagger c_{I2\downarrow} c_{I2\uparrow} + \text{H.c.}). \end{aligned} \quad (3)$$

The operator  $c_{I\alpha\sigma}^\dagger (c_{I\alpha\sigma})$  creates (annihilates) an electron with spin  $\sigma (= \uparrow, \downarrow)$  and orbital index  $\alpha (= 1, 2)$  at the  $I$ th site (or unit cell) and  $n_{I\alpha\sigma} = c_{I\alpha\sigma}^\dagger c_{I\alpha\sigma}$ .  $t$  represents the orbital-independent nearest-neighbor hopping and  $t'$  the hybridization between the two orbitals.  $U$  ( $U'$ ) represents the intraband (interband) Coulomb interaction and  $J$  the Hund coupling.  $\bar{\sigma}$  denotes the opposite spin of  $\sigma$ , and  $\langle IJ \rangle$  a pair of nearest neighbors. The standard constraint  $U' = U - 2J$  is used given by the cubic symmetry. We only consider the regime  $0 \leq J \leq \frac{U}{3}$  throughout this study to ensure that all the Coulomb interaction coefficients in Eq. (3) are non-negative.

Under the linear transformation  $\tilde{c}_{I1\sigma} = \frac{1}{\sqrt{2}}(c_{I1\sigma} + c_{I2\sigma})$  and  $\tilde{c}_{I2\sigma} = \frac{1}{\sqrt{2}}(c_{I1\sigma} - c_{I2\sigma})$ , where  $\tilde{c}_{I1\sigma} / \tilde{c}_{I2\sigma}$  annihilates an electron from the bonding/antibonding band, respectively, the kinetic Hamiltonian term can be written in a more familiar form,

$$\hat{H}_K = \sum_{I\alpha\sigma} \Delta_\alpha \tilde{c}_{I\alpha\sigma}^\dagger \tilde{c}_{I\alpha\sigma} - t \sum_{\langle IJ \rangle \alpha \sigma} \tilde{c}_{I\alpha\sigma}^\dagger \tilde{c}_{J\alpha\sigma}, \quad (4)$$

where  $\Delta_\alpha$  is the orbital-dependent crystal-field splitting and  $\Delta_1 = -t'$ ,  $\Delta_2 = t'$ , and the Coulomb interaction term

$$\begin{aligned} \hat{H}_{\text{Coul}} = & U \sum_{I\alpha} \tilde{n}_{I\alpha\uparrow} \tilde{n}_{I\alpha\downarrow} + U' \sum_{I\sigma} \tilde{n}_{I1\sigma} \tilde{n}_{I2\sigma} \\ & + (U' - J) \sum_{I\sigma} \tilde{n}_{I1\sigma} \tilde{n}_{I2\sigma} + J \sum_I (\tilde{c}_{I1\uparrow}^\dagger \tilde{c}_{I2\downarrow}^\dagger \tilde{c}_{I1\downarrow} \tilde{c}_{I2\uparrow} \\ & + \tilde{c}_{I1\uparrow}^\dagger \tilde{c}_{I1\downarrow}^\dagger \tilde{c}_{I2\downarrow} \tilde{c}_{I2\uparrow} + \text{H.c.}), \end{aligned} \quad (5)$$

where  $\tilde{n}_{I\alpha\sigma} = \tilde{c}_{I\alpha\sigma}^\dagger \tilde{c}_{I\alpha\sigma}$ . It can be seen that the Kanamori Coulomb interaction term  $\hat{H}_{\text{Coul}}$  keeps the same form as in Eq. (3) under the linear transformation, i.e.,  $\hat{H}_{\text{Coul}}$  is rotationally invariant.

### B. The GCGM method

In a recent review work on the GCGM approach [7], we outlined the complete formalism of GCGM starting from a two-atom system. The formalism is easier to follow for this simplest system, and its extension to periodic bulk systems is straightforward. The formalism of the GCGM method for periodic systems has been presented elsewhere [6,7,10]. However, to facilitate a better understanding of the method, we still outline the essential formalism here. The formalism presented here has a general form and can be applied to any periodic system other than the Hubbard model. We start from a general nonrelativistic and periodic Hamiltonian,

$$\begin{aligned} \hat{H} = & \sum_{Ii\alpha, Jj\beta, \sigma} t_{Ii\alpha, Jj\beta, \sigma} c_{Ii\alpha\sigma}^\dagger c_{Jj\beta\sigma} \\ & + \frac{1}{2} \sum_{\substack{Ii\alpha, Jj\beta \\ Kk\gamma, Ll\delta, \sigma\sigma'}} u(Ii\alpha, Jj\beta; Kk\gamma, Ll\delta) \\ & \times c_{Ii\alpha\sigma}^\dagger c_{Jj\beta\sigma'}^\dagger c_{Ll\delta\sigma} c_{Kk\gamma\sigma}, \end{aligned} \quad (6)$$

where  $I, J, K, L$  are the unit cell indices,  $i, j, k, l$  the atomic site indices,  $\alpha, \beta, \gamma, \delta$  the orbital indices, and  $\sigma, \sigma'$  the spin indices. Here,  $t$  and  $u$  are the one-electron hopping integral and the two-electron Coulomb integral, respectively. We note that in the Hubbard model described in Eqs. (1)–(5), there is

only one atom per unit cell, and that the atomic site indices are therefore removed from these equations. We start from a trial wave function  $|\Psi_0\rangle$  that is often noninteracting, i.e., a single Slater determinant, which can be written as the projection onto on-site configurations,

$$|\Psi_0\rangle = \sum_{\{\Gamma_{Ii}\}} |\{\Gamma_{Ii}\}\rangle \langle \{\Gamma_{Ii}\} | \Psi_0 \rangle = \sum_{\{\Gamma_{Ii}\}} \lambda_{\{\Gamma_{Ii}\}} |\{\Gamma_{Ii}\}\rangle, \quad (7)$$

where the summation runs through all possible single-atom configurations. The coefficient  $\lambda_{\{\Gamma_{Ii}\}}$  is defined as  $\lambda_{\{\Gamma_{Ii}\}} = \langle \{\Gamma_{Ii}\} | \Psi_0 \rangle$  and  $\Gamma_{Ii}$  is the on-site configuration at atomic site  $Ii$ , which is defined as a Fock state  $|\Gamma_{Ii}\rangle \equiv \prod_{\alpha\sigma \in \Gamma_{Ii}} c_{\alpha\sigma}^\dagger |\emptyset\rangle$ . Here the creation operator  $c_{\alpha\sigma}^\dagger$  creates an electron at the orbital  $\alpha$  with spin  $\sigma$  in the vacuum state  $|\emptyset\rangle$ . The GWF is constructed from  $|\Psi_0\rangle$  and can be written as

$$\begin{aligned} |\Psi_{GWF}\rangle &= \prod_{Ii} \left( \sum_{\Gamma} g(\Gamma_{Ii}) |\Gamma_{Ii}\rangle \langle \Gamma_{Ii}| \right) |\Psi_0\rangle \\ &= \sum_{\{\Gamma_{Ii}\}} \left( \prod_{Ii} g(\Gamma_{Ii}) \right) \lambda_{\{\Gamma_{Ii}\}} |\{\Gamma_{Ii}\}\rangle, \end{aligned} \quad (8)$$

where  $g(\Gamma_{Ii})$  are Gutzwiller variational parameters introduced to control the weight of the atomic configurations  $\Gamma_{Ii}$  in

$|\Psi_{GWF}\rangle$ . Since all unit cells are identical,  $g(\Gamma_{Ii})$  is independent of a specific unit cell and so  $g(\Gamma_{Ii}) = g(\Gamma_i)$ . The total energy of the system can be expressed as

$$\begin{aligned} E_{GWF} &= \sum_{Ii\alpha, Jj\beta, \sigma} t_{Ii\alpha, Jj\beta} \langle c_{Ii\alpha\sigma}^\dagger c_{Jj\beta\sigma} \rangle_{GWF} \\ &+ \frac{1}{2} \sum_{\substack{Ii\alpha, Jj\beta \\ Kk\gamma, Ll\delta, \sigma\sigma'}} u(Ii\alpha, Jj\beta; Kk\gamma, Ll\delta) \\ &\times \langle c_{Ii\alpha\sigma}^\dagger c_{Jj\beta\sigma'}^\dagger c_{Ll\delta\sigma'} c_{Kk\gamma\sigma} \rangle_{GWF}, \end{aligned} \quad (9)$$

where  $\langle \hat{O} \rangle_{GWF}$  is a shorthand notation for  $\frac{\langle \Psi_{GWF} | \hat{O} | \Psi_{GWF} \rangle}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle}$ . Then we evaluate rigorously the on-site two-particle correlation matrix (2PCM) where  $Ii = Jj = Kk = Ll$  and use Wick's theorem [25] to evaluate the intersite 2PCM approximately,

$$\begin{aligned} &\langle c_{Ii\alpha\sigma}^\dagger c_{Jj\beta\sigma'}^\dagger c_{Ll\delta\sigma'} c_{Kk\gamma\sigma} \rangle_{GWF} \\ &\approx \langle c_{Ii\alpha\sigma}^\dagger c_{Kk\gamma\sigma} \rangle_{GWF} \langle c_{Jj\beta\sigma'}^\dagger c_{Ll\delta\sigma'} \rangle_{GWF} \\ &- \delta_{\sigma\sigma'} \langle c_{Ii\alpha\sigma}^\dagger c_{Ll\delta\sigma} \rangle_{GWF} \langle c_{Jj\beta\sigma}^\dagger c_{Kk\gamma\sigma} \rangle_{GWF}. \end{aligned} \quad (10)$$

The total energy can be expressed as

$$\begin{aligned} E_{GWF} &\approx \sum_{Ii\alpha, Jj\beta, \sigma} t_{Ii\alpha, Jj\beta} \langle c_{Ii\alpha\sigma}^\dagger c_{Jj\beta\sigma} \rangle_{GWF} + \frac{1}{2} \sum_{\substack{Ii, \alpha\beta\gamma\delta \\ \sigma\sigma'}} u(Ii\alpha, Ii\beta; Ii\gamma, Ii\delta) \langle c_{Ii\alpha\sigma}^\dagger c_{Ii\beta\sigma'}^\dagger c_{Ii\delta\sigma'} c_{Ii\gamma\sigma} \rangle_{GWF} \\ &+ \frac{1}{2} \sum_{\substack{Ii\alpha, Jj\beta \\ Kk\gamma, Ll\delta, \sigma\sigma'}} [u(Ii\alpha, Jj\beta; Kk\gamma, Ll\delta) - \delta_{\sigma\sigma'} u(Ii\alpha, Jj\beta; Ll\delta, Kk\gamma)] \langle c_{Ii\alpha\sigma}^\dagger c_{Kk\gamma\sigma} \rangle_{GWF} \langle c_{Jj\beta\sigma'}^\dagger c_{Ll\delta\sigma'} \rangle_{GWF}, \end{aligned} \quad (11)$$

where  $\sum'$  indicates that the pure on-site terms are excluded from the summation. To reduce the Hartree-Fock (HF)-type factorization error resulting from the use of Wick's theorem, the sum-rule correction [26] that we developed earlier will be added to the original Hamiltonian. In this work, we focus on the Hubbard model only, where the intersite Coulomb terms are excluded in the Hamiltonian. In this specific case, the HF-type factorized approximation is not needed. So, we do not present the expression of the sum-rule term for conciseness.

Since all operators in Eq. (11) for the observables of interest act on at most two atomic sites, we can define a prefactor  $\xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}}$  that depends only on the atomic configurations on two sites  $Ii$  and  $Jj$  by summing over all the other sites:

$$\begin{aligned} \xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}} &= \sum_{\{\Gamma_{Kk}, Kk \neq Ii, Jj\}} \left( \prod_{Kk} g(\Gamma_{Kk})^2 \right) \langle \Psi_0 | \Gamma_{Ii}, \Gamma_{Jj}, \{\Gamma_{Kk}\} \rangle \langle \Gamma'_{Ii}, \Gamma'_{Jj}, \{\Gamma_{Kk}\} | \Psi_0 \rangle \\ &= \sum_{\{\Gamma_{Kk}, Kk \neq Ii, Jj\}} \left( \prod_{Kk} g(\Gamma_{Kk})^2 \right) \lambda_{\Gamma_{Ii}, \Gamma_{Jj}, \{\Gamma_{Kk}\}} \lambda_{\Gamma'_{Ii}, \Gamma'_{Jj}, \{\Gamma_{Kk}\}}. \end{aligned} \quad (12)$$

Here,  $|\Gamma_{Ii}, \Gamma_{Jj}, \{\Gamma_{Kk}\}\rangle$  denotes a basis configuration in the many-body Hilbert space with  $\Gamma_{Ii}$  and  $\Gamma'_{Ii}$  being the atomic configurations at site  $Ii$ ,  $\Gamma_{Jj}$  and  $\Gamma'_{Jj}$  the configurations at site  $Jj$ , and  $\{\Gamma_{Kk}\}$  the set of configurations on sites other than  $Ii, Jj$ . Then, each component of the total energy can be straightforwardly expressed in terms of  $\xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}}$  and the expectation values of the corresponding operator in the subspace spanned by sites  $Ii$  and  $Jj$ . However, as  $\xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}}$  is determined from both  $|\Psi_0\rangle$  and the Gutzwiller variational parameters of all the rest of the sites other than  $(Ii, Jj)$ , it is a grand challenge to rigorously evaluate the coefficient  $\xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}}$  because the computational complexity grows exponentially with respect to the number of atomic sites. On the other hand, the noncorrelated  $\xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}}^0$  obtained by setting  $g(\Gamma_k) = 1$  in Eq. (12) can be

expressed as

$$\begin{aligned}\xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma'_{jj}}^0 &= \sum_{\{\Gamma_{kk},Kk\neq Ii,Jj\}} \langle \Psi_0 | \Gamma_{ii}, \Gamma_{jj}, \{\Gamma_{kk}\} \rangle \langle \Gamma'_{ii}, \Gamma'_{jj}, \{\Gamma_{kk}\} | \Psi_0 \rangle \\ &= \sum_{\{\Gamma_{kk},Kk\neq Ii,Jj\}} \lambda_{\Gamma_{ii},\Gamma_{jj},\{\Gamma_{kk}\}} \lambda_{\Gamma'_{ii},\Gamma'_{jj},\{\Gamma_{kk}\}},\end{aligned}\quad (13)$$

and can be efficiently evaluated using the Wick's theorem. Then,  $\xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma'_{jj}}$  can be estimated using the following expression:

$$\xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma'_{jj}} \approx \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma'_{jj}}^0 F_{\uparrow}(n_{Ii\uparrow}, n_{Jj\uparrow}) F_{\downarrow}(n_{Ii\downarrow}, n_{Jj\downarrow}),\quad (14)$$

where

$$F_{\sigma}(n_{Ii\sigma}, n_{Jj\sigma}) = \begin{cases} \left( \sum_{\Gamma_k} p(\Gamma_k) \frac{g(\Gamma_k^{\sigma+})^2}{g(\Gamma_k)^2} \right)^{-(n_{Ii\sigma} + n_{Jj\sigma} - n_{0\sigma})}, & n_{Ii\sigma} + n_{Jj\sigma} < n_{0\sigma} \\ 1, & n_{Ii\sigma} + n_{Jj\sigma} = n_{0\sigma} \\ \left( \sum_{\Gamma_k} p(\Gamma_k) \frac{g(\Gamma_k^{\sigma-})^2}{g(\Gamma_k)^2} \right)^{n_{Ii\sigma} + n_{Jj\sigma} - n_{0\sigma}}, & n_{Ii\sigma} + n_{Jj\sigma} > n_{0\sigma}. \end{cases}\quad (15)$$

Here,  $n_{Ii\sigma}$  and  $n_{Jj\sigma}$  are the number of electrons with spin  $\sigma$  occupying site  $Ii$  and  $Jj$ , respectively, and  $n_0$  is the expected total number of electrons with spin  $\sigma$  occupying sites  $Ii$  and  $Jj$ .  $p(\Gamma_k)$  is the probability of finding the configuration  $\Gamma_k$  at site  $k$ , and  $\Gamma_k^{\sigma+}$  ( $\Gamma_k^{\sigma-}$ ) is the new configuration when one electron with spin  $\sigma$  is added to (removed from) the configuration  $\Gamma_k$ . For a half-filling system,  $p(\Gamma_k)$  can be given by<sup>1</sup>

$$p(\Gamma_i) = \frac{\sum_{\Gamma_{jj}} \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma_{ii},\Gamma_{jj}}^0 g(\Gamma_i)^2 g(\Gamma_j)^2}{\sum_{\Gamma_{ii},\Gamma_{jj}} \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma_{ii},\Gamma_{jj}}^0 g(\Gamma_i)^2 g(\Gamma_j)^2} \Bigg|_{\substack{n_{Ii\uparrow} + n_{Jj\uparrow} = n_{0\uparrow}, \\ n_{Ii\downarrow} + n_{Jj\downarrow} = n_{0\downarrow}}}\quad (16)$$

where site  $Jj$  is defined as the nearest-neighbor of site  $Ii$ . The physical meaning underlying Eqs. (14) and (15) is discussed in Ref. [10] and is not presented here for conciseness. Equation (14) is the key approximation to efficiently evaluate the energy. It has been demonstrated to be more accurate than the commonly used GA for the single-band Hubbard model, in both one and two dimensions [10]. In this work, we use a two-band Hubbard model with an extensive parameter space as a testbed to demonstrate the accuracy of Eq. (14) for multiband/multiorbital systems.

Finally, we present the expressions for the quantities necessary to evaluate the total energy as defined in Eq. (11). The one-particle density matrix (1PDM) can be expressed as

$$\langle c_{Ii\alpha\sigma}^{\dagger} c_{Ii\beta\sigma} \rangle_{GWF} = \frac{1}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{Ii,Jj}} \sum_{\Gamma_{ii},\Gamma'_{ii},\Gamma_{jj}} \langle \Gamma_{ii} | c_{Ii\alpha\sigma}^{\dagger} c_{Ii\beta\sigma} | \Gamma'_{ii} \rangle g(\Gamma_i) g(\Gamma'_i) g(\Gamma_j)^2 \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma_{jj}},\quad (17)$$

$$\begin{aligned}\langle c_{Ii\alpha\sigma}^{\dagger} c_{Jj\beta\sigma} \rangle_{GWF} &= \frac{1}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{Ii,Jj}} \sum_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma'_{jj}} \langle \Gamma_{ii} | c_{Ii\alpha\sigma}^{\dagger} | \Gamma'_{ii} \rangle \langle \Gamma_{jj} | c_{Jj\beta\sigma} | \Gamma'_{jj} \rangle \\ &\times g(\Gamma_i) g(\Gamma_j) g(\Gamma'_i) g(\Gamma'_j) \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma'_{jj}}, \quad \text{for } (I, i) \neq (J, j),\end{aligned}\quad (18)$$

where the atomic site  $Jj$  is chosen to be the nearest neighbor of site  $Ii$  in Eq. (17) and

$$\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{Ii,Jj} = \sum_{\Gamma_{ii},\Gamma_{jj}} \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma_{ii},\Gamma_{jj}} g(\Gamma_i)^2 g(\Gamma_j)^2,\quad (19)$$

where, the subscript  $(Ii, Jj)$  on  $\langle \Psi_{GWF} | \Psi_{GWF} \rangle$  serves as a reminder that it is an approximate expression based on the approximate values of  $\xi_{\Gamma_{ii},\Gamma_{jj},\Gamma_{ii},\Gamma_{jj}}$  defined on sites  $(Ii, Jj)$ . The on-site 2PCM can be expressed as

$$\langle c_{Ii\alpha\sigma}^{\dagger} c_{Ii\beta\sigma}^{\dagger} c_{Ii\delta\sigma'} c_{Ii\gamma\sigma} \rangle_{GWF} = \frac{1}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{Ii,Jj}} \sum_{\Gamma_{ii},\Gamma'_{ii},\Gamma_{jj}} \langle \Gamma_{ii} | c_{Ii\alpha\sigma}^{\dagger} c_{Ii\beta\sigma}^{\dagger} c_{Ii\delta\sigma'} c_{Ii\gamma\sigma} | \Gamma'_{ii} \rangle g(\Gamma_i) g(\Gamma'_i) g(\Gamma_j)^2 \xi_{\Gamma_{ii},\Gamma_{jj},\Gamma'_{ii},\Gamma_{jj}}.\quad (20)$$

Again, the atomic site  $Jj$  is chosen to be the nearest neighbor of site  $Ii$ . For the Hubbard model discussed in this work, only the on-site 2PCM  $\langle c_{Ii\alpha\sigma}^{\dagger} c_{Ii\beta\sigma}^{\dagger} c_{Ii\delta\sigma'} c_{Ii\gamma\sigma} \rangle_{GWF}$  are included, which can be evaluated using Eq. (20).

<sup>1</sup>For systems with other electron densities, the expression of  $p(\Gamma_i)$  can be found in Ref. [10].

### C. The RI scheme

In the original GWF, the Gutzwiller projector in GCGM is parametrized based on the on-site configurations  $\Gamma$ . As a result, the computational cost scales exponentially with the increasing number of electrons and orbitals, which can be a major bottleneck in the GCGM method. To reduce the computational complexity, a RI scheme is incorporated into the GCGM method to group certain orbitals and describe the system with the number of electrons occupying the orbitals instead of the configuration  $\Gamma_i$ . The Gutzwiller projector  $\hat{G} = \prod_{li} (\sum_{\Gamma_{li}} g(\Gamma_i) |\Gamma_{li}\rangle \langle \Gamma_{li}|)$  in Eq. (7) has a diagonal form. If we use another orthonormal set of basis  $\{|\tilde{\Gamma}_{li}\rangle\}$  instead of  $\{|\Gamma_{li}\rangle\}$ , the Gutzwiller projector becomes

$$\begin{aligned} \hat{G} &= \prod_{li} \left( \sum_{\Gamma_{li}} g(\Gamma_i) |\Gamma_{li}\rangle \langle \Gamma_{li}| \right) \\ &= \prod_{li} \left( \sum_{\Gamma_i} g(\Gamma_i) \sum_{\tilde{\Gamma}_{li}, \tilde{\Gamma}'_{li}} |\tilde{\Gamma}_{li}\rangle \langle \tilde{\Gamma}_{li}| \langle \Gamma_{li}| \tilde{\Gamma}'_{li}\rangle \langle \tilde{\Gamma}'_{li}| \right) \\ &= \prod_{li} \sum_{\tilde{\Gamma}_{li}, \tilde{\Gamma}'_{li}} \left( \sum_{\Gamma_i} \langle \tilde{\Gamma}_{li}| \Gamma_{li}\rangle g(\Gamma_i) \langle \Gamma_{li}| \tilde{\Gamma}'_{li}\rangle \right) |\tilde{\Gamma}_{li}\rangle \langle \tilde{\Gamma}'_{li}|. \quad (21) \end{aligned}$$

If we want to enforce the diagonal form of  $\hat{G}$  during the rotational transformation of the basis, we need the term

$\sum_{\Gamma_{li}} \langle \tilde{\Gamma}_{li}| \Gamma_{li}\rangle g(\Gamma_i) \langle \Gamma_{li}| \tilde{\Gamma}'_{li}\rangle$  to be  $g(\tilde{\Gamma}_i) \delta_{\tilde{\Gamma}_{li}, \tilde{\Gamma}'_{li}}$ . It is apparent that  $\langle \tilde{\Gamma}_{li}| \Gamma_{li}\rangle$  and  $\langle \Gamma_{li}| \tilde{\Gamma}'_{li}\rangle$  can be nonzero only when the electron number is conserved, i.e.,  $n(\tilde{\Gamma}_i, \sigma) = n(\Gamma_i, \sigma)$  and  $n(\Gamma_{li}, \sigma) = n(\tilde{\Gamma}'_{li}, \sigma)$ , where  $n(\Gamma_{li}, \sigma)$  denotes the number of spin- $\sigma$  electrons of the configuration  $\Gamma_{li}$ . So, if  $g$  is dependent only on  $n(\Gamma_i, \sigma)$  instead of the configuration  $\Gamma_i$  on site  $i$ , i.e.,  $g(\Gamma_i) = g[n(\Gamma_i, \sigma)] = g[n(\Gamma_i, \uparrow), n(\Gamma_i, \downarrow)]$ , the term in Eq. (21) becomes  $\sum_{\Gamma_{li}} \langle \tilde{\Gamma}_{li}| \Gamma_{li}\rangle g(\Gamma_i) \langle \Gamma_{li}| \tilde{\Gamma}'_{li}\rangle = g(n(\Gamma_i), \sigma) \sum_{\Gamma_{li}} \langle \tilde{\Gamma}_{li}| \Gamma_{li}\rangle \langle \Gamma_{li}| \tilde{\Gamma}'_{li}\rangle = g(n(\Gamma_i), \sigma) \delta_{\tilde{\Gamma}_{li}, \tilde{\Gamma}'_{li}}$ . So, Eq. (21) becomes

$$\hat{G} = \prod_{li} \sum_{\tilde{\Gamma}_i} g[n(\tilde{\Gamma}_i, \sigma)] |\tilde{\Gamma}_{li}\rangle \langle \tilde{\Gamma}_{li}|, \quad (22)$$

which satisfies the diagonal form. Equation (22) indicates that the system should be parametrized based on  $n(\Gamma_i, \sigma)$  instead of  $\Gamma_i$  if we want to keep the off-diagonal term of  $\hat{G}$  to be zero during the rotational transformation of the basis. So, we can group the orbitals if they satisfy certain symmetry and use the number of electrons occupying the grouped orbitals to describe the system. The advantage of doing this is that the number of possible configurations grows exponentially with orbitals while the possible number of electrons occupying the grouped orbitals grows only linearly.

In the RI scheme, the 1PDM in Eq. (17) can be rewritten as

$$\begin{aligned} \langle c_{li\alpha\sigma}^\dagger c_{li\beta\sigma} \rangle_{GWF} &= \frac{1}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{li, jj}} \sum_{\substack{\{n_{li\uparrow}, n_{li\downarrow}\} \\ \{n_{jj\uparrow}, n_{jj\downarrow}\}}} g(\{\dots, n_i(A, \sigma), n_i(B, \sigma), \dots\}, \{n_i(-\sigma)\}) \\ &\quad \times g(\{\dots, n_i(A, \sigma) - 1, n_i(B, \sigma) + 1, \dots\}, \{n_i(-\sigma)\}) g(\{n_j(\uparrow)\}, \{n_j(\downarrow)\})^2 \\ &\quad \times F_\uparrow(n_{li\uparrow}, n_{jj\uparrow}) F_\downarrow(n_{li\downarrow}, n_{jj\downarrow}) \sum_{\Gamma_{li}, \Gamma'_{li}, \Gamma_{jj}} \langle \Gamma_{li} | c_{li\alpha\sigma}^\dagger c_{li\beta\sigma} | \Gamma'_{li} \rangle \xi_{\Gamma_{li}, \Gamma_{jj}, \Gamma'_{li}, \Gamma_{jj}}^0, \quad (23) \end{aligned}$$

when orbital  $\alpha$  and  $\beta$  cannot be grouped together, i.e.,  $\alpha$  belongs to group  $A$  of orbitals and  $\beta$  belongs to a different group  $B$ .  $-\sigma$  denotes the opposite spin of  $\sigma$ . Here,  $g(\Gamma_j)$  in Eq. (14) is now rewritten as  $g(\{n_j(\uparrow)\}, \{n_j(\downarrow)\})$ ,  $g(\Gamma_i)$  is rewritten as  $g(\{\dots, n_i(A, \sigma), n_i(B, \sigma), \dots\}, \{n_i(-\sigma)\})$ , and  $g(\Gamma'_i)$  is rewritten as  $g(\{\dots, n_i(A, \sigma) - 1, n_i(B, \sigma) + 1, \dots\}, \{n_i(-\sigma)\})$ . Here,  $\Gamma_{li}$  and  $\Gamma'_{li}$  are correlated by the selection rule:  $\langle \Gamma_{li} | c_{li\alpha\sigma}^\dagger c_{li\beta\sigma} | \Gamma'_{li} \rangle$  in Eq. (20) is nonzero only if  $\Gamma_{li}$  and  $\Gamma'_{li}$  differ by one spin orbital (one electron hops from orbital  $\alpha$  to  $\beta$ ). For  $\Gamma_{li}$ , if the number of electrons occupying orbital groups  $(A, \sigma)$  and  $(B, \sigma)$  are  $n_i(A, \sigma)$  and  $n_i(B, \sigma)$ , then for  $\Gamma'_{li}$  the number of electrons occupying orbital groups  $(A, \sigma)$  and  $(B, \sigma)$  are  $n_i(A, \sigma) - 1$  and  $n_i(B, \sigma) + 1$ , respectively. After orbitals are grouped under the RI principle, the  $g$  factors can also be grouped, and those common  $g$  factors can be factored out from the summation in Eq. (14), which then becomes Eq. (23). When orbitals  $\alpha$  and  $\beta$  can be grouped in the same group  $A$ , Eq. (17) can be rewritten as

$$\begin{aligned} \langle c_{li\alpha\sigma}^\dagger c_{li\beta\sigma} \rangle_{GWF} &= \frac{1}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{li, jj}} \sum_{\substack{\{n_{li\uparrow}, n_{li\downarrow}\} \\ \{n_{jj\uparrow}, n_{jj\downarrow}\}}} g(\{n_i(\uparrow)\}, \{n_i(\downarrow)\})^2 g(\{n_j(\uparrow)\}, \{n_j(\downarrow)\})^2 \\ &\quad \times F_\uparrow(n_{li\uparrow}, n_{jj\uparrow}) F_\downarrow(n_{li\downarrow}, n_{jj\downarrow}) \sum_{\Gamma_{li}, \Gamma'_{li}, \Gamma_{jj}} \langle \Gamma_{li} | c_{li\alpha\sigma}^\dagger c_{li\beta\sigma} | \Gamma'_{li} \rangle \xi_{\Gamma_{li}, \Gamma_{jj}, \Gamma'_{li}, \Gamma_{jj}}^0, \quad (24) \end{aligned}$$

where both  $g(\Gamma_i)$  and  $g(\Gamma'_i)$  in Eq. (17) can be now rewritten as  $g(\{\dots, n_i(A, \sigma), \dots\}, \{n_i(-\sigma)\})$ , i.e.,  $g(\{n_i(\uparrow)\}, \{n_i(\downarrow)\})$ , since the number of electrons occupying the orbitals of group  $A$  are the same for  $\Gamma_i$  and  $\Gamma'_i$ . Equation (18) can be rewritten as

$$\begin{aligned} \langle c_{li\alpha\sigma}^\dagger c_{lj\beta\sigma} \rangle_{GWF} &= \frac{1}{\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{li, jj}} \sum_{\substack{\{n_{li\uparrow}, n_{li\downarrow}\} \\ \{n_{lj\uparrow}, n_{lj\downarrow}\}}} g(\{\dots, n_i(A, \sigma), \dots\}, \{n_i(-\sigma)\}) \\ &\quad \times g(\{\dots, n_j(B, \sigma), \dots\}, \{n_j(-\sigma)\}) g(\{\dots, n_i(A, \sigma) - 1, \dots\}, \{n_i(-\sigma)\}) \end{aligned}$$

$$\begin{aligned} & \times g(\{\dots, n_j(B, \sigma) + 1, \dots\}, \{n_j(-\sigma)\}) F_{\uparrow}(n_{Ii\uparrow}, n_{Jj\uparrow}) F_{\downarrow}(n_{Ii\downarrow}, n_{Jj\downarrow}) \\ & \times \sum_{\Gamma_{Ii}, \Gamma'_{Ii}, \Gamma_{Jj}, \Gamma'_{Jj}} \langle \Gamma_{Ii} | c_{Ii\alpha\sigma}^{\dagger} | \Gamma'_{Ii} \rangle \langle \Gamma_{Jj} | c_{Jj\beta\sigma} | \Gamma'_{Jj} \rangle \xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma'_{Ii}, \Gamma'_{Jj}}^0, \quad \text{for } (I, i) \neq (J, j). \end{aligned} \quad (25)$$

The expression of the on-site 2PCM  $\langle c_{Ii\alpha\sigma}^{\dagger} c_{Ii\beta\sigma'}^{\dagger} c_{Ii\delta\sigma'} c_{Ii\gamma\sigma} \rangle_{GWF}$  has a similar form with Eqs. (24) and (25) and is not presented here for conciseness. Equation (19) becomes

$$\langle \Psi_{GWF} | \Psi_{GWF} \rangle_{Ii, Jj} = \sum_{\substack{\{n_{Ii\uparrow}\}, \{n_{Ii\downarrow}\} \\ \{n_{Jj\uparrow}\}, \{n_{Jj\downarrow}\}}} g(\{n_{Ii}(\uparrow)\}, \{n_{Ii}(\downarrow)\})^2 g(\{n_{Jj}(\uparrow)\}, \{n_{Jj}(\downarrow)\})^2 F_{\uparrow}(n_{Ii\uparrow}, n_{Jj\uparrow}) F_{\downarrow}(n_{Ii\downarrow}, n_{Jj\downarrow}) \sum_{\Gamma_{Ii}, \Gamma_{Jj}} \xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma_{Ii}, \Gamma_{Jj}}^0. \quad (26)$$

By using the RI approach, the computational cost will be significantly reduced. If we compare, for example, Eqs. (17) and (24), we can find that the summation in Eq. (17) goes over all the possible configurations  $\{\Gamma_{Ii}, \Gamma_{Jj}\}$  ( $\Gamma'_{Ii}$  are determined by  $\Gamma_{Ii}$  with the selection rule) while the summation in Eq. (24) breaks down into two parts: the first summation goes over all the possible electron numbers occupying the grouped orbitals  $\{n_{Ii}(\uparrow), n_{Ii}(\downarrow), n_{Jj}(\uparrow), n_{Jj}(\downarrow)\}$  and the second summation goes over all the possible configurations  $\{\Gamma_{Ii}, \Gamma_{Jj}\}$ , of which the electron numbers occupying the grouped orbitals satisfy  $\{n_{Ii}(\uparrow), n_{Ii}(\downarrow), n_{Jj}(\uparrow), n_{Jj}(\downarrow)\}$ . Since all that is included in the second summation is the selection rule and the predetermined HF coefficients  $\xi_{\Gamma_{Ii}, \Gamma_{Jj}, \Gamma_{Ii}, \Gamma_{Jj}}^0$ , the summations can be evaluated only once and can be stored in the memory as invariants that are not dependent on  $g$  factors. During the minimization of energy, we only need to evaluate the first summation over  $\{n_{Ii}(\uparrow), n_{Ii}(\downarrow), n_{Jj}(\uparrow), n_{Jj}(\downarrow)\}$ , the possible choices of which scale only linearly with the number of orbitals. Therefore, the computational cost is greatly reduced compared to the cost of evaluating the summation over  $\{\Gamma_{Ii}, \Gamma_{Jj}\}$  in Eq. (17), the possible choices of which scale exponentially with the number of orbitals.

### III. RESULTS

In this section we test our GCGM approach with the 1D two-band Hubbard model described by Eqs. (1)–(5) and compare our numerical results with those given by the DMRG method as implemented in the ITensor library [27]. In DMRG calculations, the system contains 100 sites with open boundaries to ensure that the bulk properties are approached at the center. For the two-band model at half-filling, each site contains two electrons with opposite spins in the initial matrix product state. Observables such as the double occupancy are averaged over the central 20 sites.

Firstly, we focus on the original GCGM method without the RI add-on to validate the GCGM method itself. We systematically compute the ground-state energy per site and the double occupancy. We then continue to test the accuracy of the RI approach, which is integrated into the GCGM method to largely improve its efficiency. The electron density is two electrons per site, i.e., half-filling throughout this work. All the calculations are at zero temperature. To conclude our benchmark test, we study a Mott transition between a metal and a Mott insulator by evaluating the charge gap. The phase diagram given by GCGM with RI agrees reasonably well with that given by DMRG.

Figure 1 plots the ground-state energy as a function of  $U/t$  given by the original GCGM method and the DMRG method. In the GCGM calculation 32 atomic sites, or equivalently, 32  $k$  points are used; while in the DMRG calculation 100 atomic sites are used. The periodic boundary condition is not enforced in the DMRG calculation. Thus, more atomic sites are needed to imitate a periodic system. We perform the test over a wide range of  $J$ , the Hund coupling, and  $\Delta (= \Delta_2 = -\Delta_1 = t')$ , the crystal-field splitting. We present the results with selected values of  $J = 0, 0.1U, 0.2U, 0.3U$  and  $\Delta = 0.4t, t, 2t$ . For smaller  $J$  of 0 or  $0.1U$ , the GCGM gives results in satisfying agreement with the DMRG method. For larger  $J$  of  $0.2U$  or  $0.3U$ , the GCGM gives energies that are slightly higher than the DMRG energies at intermediate to strong interaction strengths. Other than that, the GCGM method in general estimates the energy accurately. The two methods give an energy difference  $\varepsilon_{\text{Energy}} \leq 0.21$ . Figure 2 plots the double occupancy (defined as  $\langle n_{I\uparrow} n_{I\downarrow} \rangle_{GWF} = \langle c_{I\uparrow}^{\dagger} c_{I\uparrow} c_{I\downarrow}^{\dagger} c_{I\downarrow} \rangle_{GWF}$ ) of the ground state as a function of  $U/t$  given by the original GCGM method and the DMRG method. The GCGM method, in general, gives a qualitatively correct estimate of the double occupancy. It sometimes moderately overestimates the double occupancy compared to the DMRG

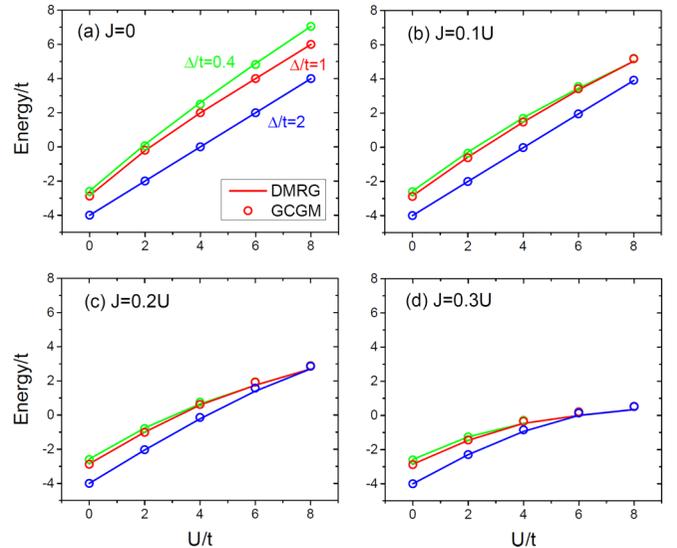


FIG. 1. The ground-state energy of the 1D two-band Hubbard model given by the GCGM and the DMRG method as a function of  $U/t$  with (a)  $J = 0$ , (b)  $J = 0.1U$ , (c)  $J = 0.2U$ , and (d)  $J = 0.3U$ ; and with the crystal-field splitting  $\Delta = 0.4t, t, 2t$ . The constraint  $U = U' + 2J$  is imposed.

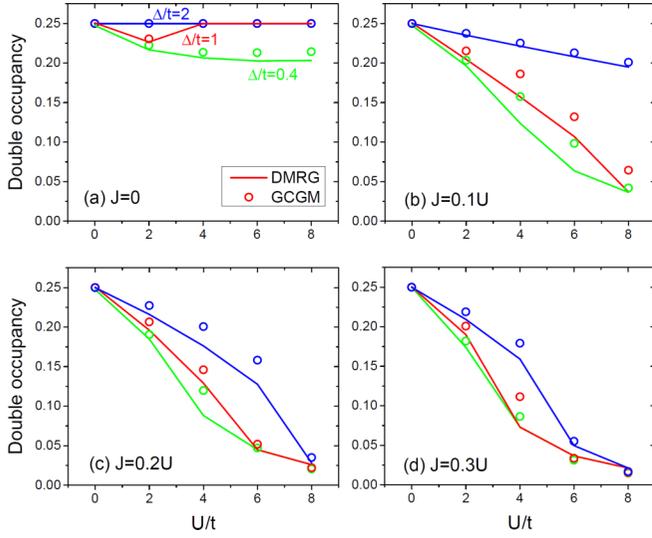


FIG. 2. The double occupancy of the 1D two-band Hubbard model given by the GCGM and the DMRG method as a function of  $U/t$  with (a)  $J = 0$ , (b)  $J = 0.1U$ , (c)  $J = 0.2U$ , and (d)  $J = 0.3U$ ; and with the crystal-field splitting  $\Delta = 0.4t$ ,  $t$ ,  $2t$ .

results. The most obvious discrepancy happens at the intermediate to strong correlation strength ( $4 \leq U/t \leq 6$ ). The two methods give a difference of double occupancy  $\varepsilon_D \leq 0.038$ . We note that the evaluation of the double occupancy is not as accurate as that in the single-band case [10].

Figures 3 and 4 plot the ground-state energy and the double occupancy, respectively, as a function of  $U/t$  given by the original GCGM method and the GCGM within the RI scheme. The RI approximation is accurate in the limit of  $J = 0$ , as verified in the calculations of the energy [Fig. 3(a)] and of the double occupancy [Fig. 4(a)]. When  $J \neq 0$ , using the RI approach still nearly perfectly reproduces the energy.

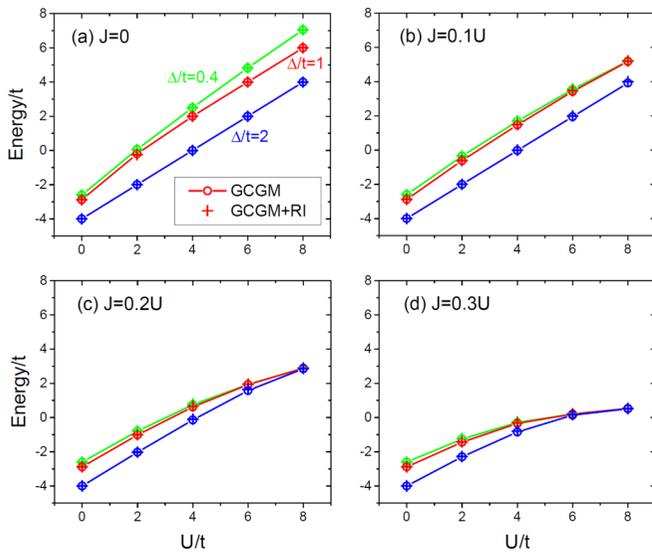


FIG. 3. The ground-state energy of the 1D two-band Hubbard model given by the original GCGM method and the GCGM within the RI approach as a function of  $U/t$  with (a)  $J = 0$ , (b)  $J = 0.1U$ , (c)  $J = 0.2U$ , and (d)  $J = 0.3U$ ; and with the crystal-field splitting  $\Delta = 0.4t$ ,  $t$ ,  $2t$ .

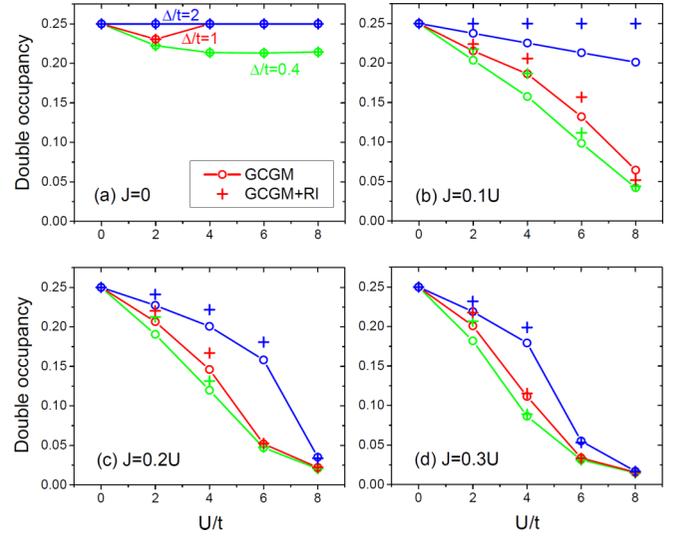


FIG. 4. The double occupancy of the 1D two-band Hubbard model given by the original GCGM method and the GCGM with the RI approach as a function of  $U/t$  with (a)  $J = 0$ , (b)  $J = 0.1U$ , (c)  $J = 0.2U$ , and (d)  $J = 0.3U$ ; and with the crystal-field splitting  $\Delta = 0.4t$ ,  $t$ ,  $2t$ .

The discrepancy in the energies given by the two methods is small with  $\varepsilon_{\text{Energy}} \leq 0.08$ , demonstrating that the RI approach is a good approximation in describing the energy. However, the RI approach introduces some inaccuracy when it is used to estimate the double occupancy, as shown in Figs. 4(b)–4(d). It often overestimates the double occupancy, particularly at the weak to intermediate correlation strength. The most obvious discrepancy ( $\varepsilon_D \sim 0.049$ ) happens when  $J = 0.1U$ ,  $\Delta = 2t$ , as shown in Fig. 4(b). The minimized energy favors a pair of electrons of opposite spin occupying both orbitals (e.g., the spin-up electron occupies orbital No. 1 and the spin-down electron occupies orbital No. 2) instead of only one orbital (e.g., both electrons occupy orbital No. 1). So, the  $g$  factor of the first case is larger than that of the second case. However, in the rotationally invariant approximation, the two scenarios are treated equally and the  $g$  factors are enforced to be the same for the two scenarios, resulting in the discrepancy as seen in Fig. 4(b). Nevertheless, although the estimation of the double occupancy is not very accurate with the rotationally invariant approach, the energy is well reproduced, as shown in Fig. 3(b).

An original motivation for the Hubbard model is to use it as a prototype to achieve fundamental understanding of the Mott transition between a metal and a Mott insulator induced by the repulsive interactions among electrons [28–32]. For the single-band 1D Hubbard model, Lieb and Wu established that Mott transition does not exist and the system is insulating at any finite  $U$  [33]. In contrast, the two-band Hubbard model displays a rich behavior of Mott transitions, thus providing another testbed to benchmark the GCGM method [34,35].

We use the charge gap  $\Delta_c$  to identify a Mott transition.  $\Delta_c$  can be defined as

$$\Delta_c = \frac{1}{2}[E(N/2 + 1, N/2 + 1) + E(N/2 - 1, N/2 - 1) - 2E(N/2, N/2)], \quad (27)$$

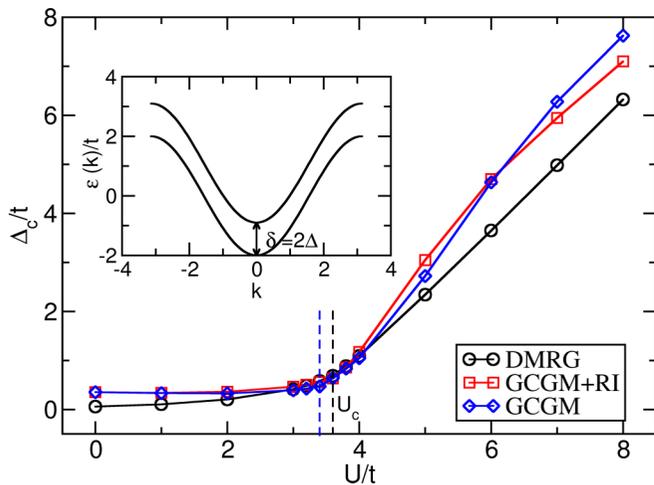


FIG. 5.  $\Delta_c$  as a function of  $U$  calculated by DMRG and GCGM with or without rotational invariance. Inset gives the band structure of the noninteracting Hamiltonian. The dashed black line gives a critical  $U_c = 3.6t$  for both DMRG and GCGM+RI methods, and the dashed blue line gives a similar  $U_c = 3.4t$  for the GCGM method without RI.

where  $E(x, y)$  is the ground-state energy of the system with  $x$  spin-up electrons and  $y$  spin-down electrons.  $\Delta_c$  should be zero in a metallic state and takes finite values when the system becomes insulating. However, it is numerically challenging to capture a vanishing  $\Delta_c$  according to Eq. (27) since it involves obtaining a small number by subtraction of large numbers. To see this, we show in Fig. 5  $\Delta_c$  as a function of  $U$  for the crystal-field splitting  $\Delta = 0.6t$  and the Hund coupling  $J = 0.2U$ , calculated by DMRG and GCGM with or without RI. The inset in Fig. 5 gives the band structure of the noninteracting Hamiltonian ( $U = 0$ ), in which the two bands are separated by  $\delta = 2\Delta$ . As long as  $\delta < 4t$  (the bandwidth), or  $\Delta < 2t$ , the noninteracting system is metallic. However, as shown in Fig. 5, both DMRG and GCGM give small (less than 1% of the total energy)  $\Delta_c$  when  $U$  is close to zero, due to numerical inaccuracy of the methods. Nevertheless, one can clearly identify a “kink” at  $U \approx 3.6t$  for DMRG and GCGM+RI, and at  $U \approx 3.4t$  for GCGM without RI, which shows a sudden change from a weak to a strong response of  $\Delta_c$  with respect to  $U$ , signaling a metal-to-insulator transition.

Similar calculations are repeated for a range of  $\Delta$  values to map out the critical  $U$  at the transition state ( $U_c$ ) as a function of  $\Delta$ , which is plotted in Fig. 6. When  $J = 0.1U$ , the GCGM method gives  $U_c$  that compares favorably with DMRG, while GCGM+RI overestimates  $U_c$  for relatively small  $\Delta$  and underestimates  $U_c$  for large  $\Delta$ . For a larger Hund’s coupling of  $J = 0.2U$ , the agreement between GCGM+RI and DMRG improves, and the results from both GCGM and GCGM+RI fluctuate around those from DMRG. Overall, both GCGM and GCGM+RI agree reasonably well with DMRG, with the agreement for GCGM being slightly better, possibly because of the extra approximations involved in the GCGM+RI formalism. We use the charge gap  $\Delta_c$  to identify a Mott transition, which relies on how accurately we could calculate the energy. As shown in Figs. 1 and 3, GCGM in general gives a relatively accurate energy. So, although the estimation of the

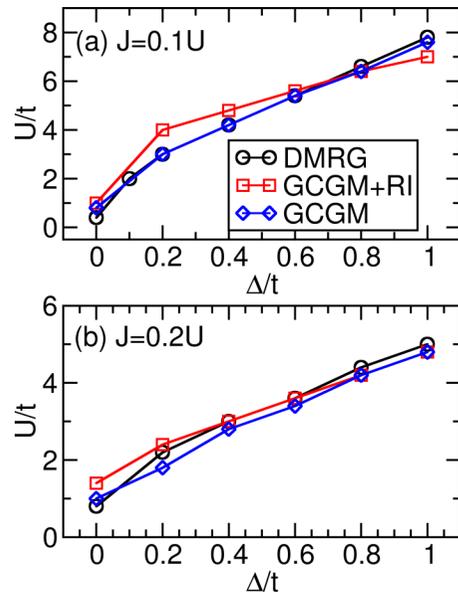


FIG. 6. The critical  $U$  for metal-to-Mott insulator transitions as a function of the crystal-field splitting  $\Delta/t$  for (a)  $J = 0.1U$  and (b)  $J = 0.2U$ .

double occupancy is not very accurate as shown in Figs. 2 and 4, GCGM is still able to identify a Mott transition with reasonable accuracy.

Now we discuss the speedup of the RI approach. Table I lists the computation time of the three methods. The original GCGM and GCGM with RI were run on a single core since they are efficient enough, while a parallel version of the DMRG code was executed on one computing node with 32 cores. In the GCGM calculation 32 atomic sites are used with periodic boundary conditions; while in the DMRG calculation 100 atomic sites are used without periodic boundary conditions. The original GCGM takes  $\sim 12$  s with one core to calculate and store the predetermined factors [ $\xi_{\Gamma_{ii}, \Gamma_{jj}, \Gamma'_{ii}, \Gamma'_{jj}}^0$  in Eq. (13)] and  $\sim 20$  s per iteration during the minimization of the energy. It usually takes up to 30 iterations for the energy to converge. The GCGM in conjunction with the RI takes  $\sim 11$  s to compute and store the prefactors [the summations that contain  $\xi_{\Gamma_{ii}, \Gamma_{jj}, \Gamma'_{ii}, \Gamma'_{jj}}^0$  in Eqs. (23)–(26)] and the energy usually converges within 1–2 s, that is, roughly 0.04–0.06 s per iteration. The predetermined factors need to be estimated only once and do not change in the process of energy minimization. The speedup is 300–500 times for each iteration, and the overall speedup is  $\sim 50$  times if it takes  $\sim 30$  iterations to converge. In addition, the GCGM with RI requires a reduced memory to store all the predetermined coefficients from  $|\Psi_0\rangle$ . For the original GCGM, there are  $\sim 157\,000$   $\xi_{\Gamma_{ii}, \Gamma_{jj}, \Gamma'_{ii}, \Gamma'_{jj}}^0$  that need to be precalculated and stored. For the GCGM in conjunction with RI, we only need to calculate and store  $\sim 13\,000$  prefactors since  $\xi_{\Gamma_{ii}, \Gamma_{jj}, \Gamma'_{ii}, \Gamma'_{jj}}^0$  can be grouped according to the electron occupancies of the configurations. That is an approximately tenfold reduction from the original GCGM. DMRG, on the other hand, is a much more expensive method. It takes about 1 h core to complete an iteration in DMRG, compared with 0.04–0.06 s core in GCGM+RI. Although DMRG appears to converge faster (less iterations), the total

TABLE I. The computation time of the three methods with a single core: the original GCGM, GCGM with RI, and DMRG.

Method	Original GCGM	GCGM+RI	DMRG
Time (s core) to compute prefactors	12	11	NA
Time (s core) for each iteration/sweep	20	0.04–0.06	3600
Number of iterations/sweeps	30	30	10
Total computation time (s core)	612	13	36 000

computation cost for DMRG is still three orders of magnitude larger than GCGM+RI.

We can roughly estimate the computational complexity of the GCGM with/without the use of RI by considering a general single atom per cell  $N$  orbital (i.e.,  $2N$  spin orbitals) system, where the  $N$  orbitals can be grouped under the RI scheme. To estimate nearest-neighbor hopping  $\langle c_{I\alpha\sigma}^\dagger c_{J\beta\sigma} \rangle_{GWF}$  or the two-particle term  $\langle c_{I\alpha\sigma}^\dagger c_{I\beta\sigma'}^\dagger c_{I\delta\sigma'} c_{I\gamma\sigma} \rangle_{GWF}$ , we need to sum over all the possible configurations  $(\Gamma_{Ii}, \Gamma_{Jj})$  without the use of RI. There are  $2^{2N}$  choices for each of  $(\Gamma_{Ii}, \Gamma_{Jj})$ ,  $2^{2N} \times 2^{2N} = 2^{4N}$  possible combinations of  $(\Gamma_{Ii}, \Gamma_{Jj})$ . So, the complexity to compute the total energy  $E_{GWF}$  is proportional to  $2^{4N}$ . The complexity to compute the derivative for each  $\frac{\partial E_{GWF}}{\partial g(\Gamma_i)}$  is also proportional to  $2^{4N}$ . Given there are altogether  $2^{2N}$   $g$  factors, the complexity to compute all the derivatives is proportional to  $2^{2N} \times 2^{4N} = 2^{6N}$ . With the use of RI, the summation would be over the number of spin-up and spin-down electrons occupying the grouped  $N$  orbitals  $\{n_{Ii}(\uparrow), n_{Ii}(\downarrow), n_{Jj}(\uparrow), n_{Jj}(\downarrow)\}$ . For example,  $n_{Ii}(\uparrow)$ , the number of spin-up electrons occupying the grouped  $N$  orbitals, could be  $0, 1, 2, \dots, N$ , i.e.,  $N + 1$  choices. Therefore, there are  $(N + 1)^4$  choices of  $\{n_{Ii}(\uparrow), n_{Ii}(\downarrow), n_{Jj}(\uparrow), n_{Jj}(\downarrow)\}$ . The complexity to compute  $E_{GWF}$  is thus proportional to  $(N + 1)^4$ . As there are  $(N + 1)^2$   $g$  factors, the complexity to compute all the derivatives  $\frac{\partial E_{GWF}}{\partial g(\Gamma_i)}$  is proportional to  $(N + 1)^4(N + 1)^2 = (N + 1)^6$ . The scaling is now reduced from exponential to polynomial with the use of RI.

In the above rough estimate of complexity, we overlook some factors. For example, the terms of the selection rule [such as  $\langle \Gamma_{Ii} | c_{I\alpha\sigma}^\dagger c_{I\beta\sigma} | \Gamma'_{Ii} \rangle$  in Eq. (14)] are now packed into the predetermined coefficients with the use of RI in Eq. (24) and need to be evaluated only once, resulting in further reduced complexity. The electron density also plays a role in the evaluation of the predetermined coefficients. For example, half-filling often gives the largest number of nonzero coefficients. But since these coefficients are evaluated only once, the electron density likely has little impact on the reduction of complexity with the use of RI.

#### IV. SUMMARY

In our previous studies, we have developed the GCGM method for correlated-electron systems and benchmarked it

with molecules [8,9,12] and bulk systems, in particular, the Hubbard model [10,11]. As our previous study only focused on the single-band Hubbard model, in this work we continue our investigation and validate the GCGM method with a two-band/orbital Hubbard model with demonstrated accuracy. One fundamental challenge of solving the many-body Schrödinger equation for multiorbital systems is that the computational effort grows exponentially with the number of on-site orbitals. To overcome this bottleneck, we introduce a RI approach to efficiently group orbitals that satisfy certain symmetries and parametrize the system based on the number of electrons occupying the grouped orbitals. We validate the RI approach with ground-state energy calculations and characterizations of a Mott transition between a metal and a Mott insulator. An approximately 50 times speedup makes GCGM in conjunction with RI much faster than the original GCGM and DMRG. We also note that the GCGM method, unlike DMRG, is not restricted to dealing with 1D systems, although in this work we focus on the 1D Hubbard model only. Reference [10] gives an example where we applied GCGM to the 2D Hubbard model.

Reference [12] and this work provide a systematic validation of the RI approach on smaller molecular and bulk systems. With the promising speedup of the RI approach, we will be able to study larger systems with GCGM. Multiorbital molecules for which accurate data are available for comparison (e.g.,  $\text{Cr}_2$ ) are a good testbed where the performance of GCGM on real materials can be evaluated to help the development of the method towards treating real correlated bulk materials. Another benchmark study would be the three-band or four-band Hubbard models. After these benchmark studies of more complex molecules and bulk systems are accomplished, calculations of large, realistic bulk systems will be readily carried out.

#### ACKNOWLEDGMENTS

Work at Ames National Laboratory was supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science and Engineering Division including a grant of computer time at the National Energy Research Scientific Computing Center (NERSC) in Berkeley. Ames Laboratory is operated for the U.S. DOE by Iowa State University under Contract No. DE-AC02-07CH11358.

- [1] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.* **136**, B864 (1964).  
[2] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140**, A1133 (1965).

- [3] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Band theory and Mott insulators: Hubbard  $U$  instead of Stoner  $I$ , *Phys. Rev. B* **44**, 943 (1991).  
[4] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, First-principles calculations of the electronic structure and spectra

- of strongly correlated systems: the LDA+ $U$  method. *J. Phys.: Condens. Matter* **9**, 767 (1997).
- [5] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [6] Z. Ye, Y.-X. Yao, X. Zhao, C.-Z. Wang, and K.-M. Ho, First-principles calculation of correlated electron materials based on Gutzwiller wave function beyond Gutzwiller approximation, *J. Phys.: Condens. Matter* **31**, 335601 (2019).
- [7] Z. Ye, Y. Fang, H. Zhang, F. Zhang, S. Wu, W.-C. Lu, Y.-X. Yao, C.-Z. Wang, and K.-M. Ho, The Gutzwiller conjugate gradient minimization method for correlated electron systems, *J. Phys.: Condens. Matter*, **34**, 243001 (2022).
- [8] Z. Ye, Y.-X. Yao, C.-Z. Wang, and K.-M. Ho, First-principles calculation of excited states of diatomic molecules: a benchmark for the Gutzwiller conjugate gradient minimization method, *Mol. Phys.* **118**, e1734243 (2020).
- [9] Y. H. Dong, Z. Ye, W.-C. Lu, Y.-X. Yao, C.-Z. Wang, and K.-M. Ho, A benchmark of Gutzwiller conjugate gradient minimization method in ground state energy calculations of dimers, *Comput. Theor. Chem.* **1185**, 112877 (2020).
- [10] Z. Ye, F. Zhang, Y.-X. Yao, C.-Z. Wang and K.-M. Ho, Ground-state properties of the Hubbard model in one and two dimensions from the Gutzwiller conjugate gradient minimization theory, *Phys. Rev. B* **101**, 205122 (2020).
- [11] Z. Ye, F. Zhang, Y.-X. Yao, C.-Z. Wang, and K.-M. Ho, Ground state wave functions for single-band Hubbard models from the Gutzwiller conjugate gradient minimization theory, *Mol. Phys.* **119**, e1797917 (2021).
- [12] Z. Ye, F. Zhang, Y. Fang, H. Zhang, S. Wu, W.-C. Lu, Y.-X. Yao, C.-Z. Wang, and K.-M. Ho, A rotationally invariant approach based on Gutzwiller wave function for correlated electron systems, *J. Phys.: Condens. Matter* **34**, 495601 (2022).
- [13] M. C. Gutzwiller, Effect of Correlation on the Ferromagnetism of Transition Metals, *Phys. Rev. Lett.* **10**, 159 (1963).
- [14] M. C. Gutzwiller, Effect of correlation on the ferromagnetism of transition metals, *Phys. Rev.* **134**, A923 (1964).
- [15] M. C. Gutzwiller, Correlation of electrons in a narrow  $s$  band, *Phys. Rev.* **137**, A1726 (1965).
- [16] B. Hetényi, H. G. Evertz, and W. von der Linden, Effect of the exchange hole on the Gutzwiller approximation in one dimension, *Phys. Rev. B* **80**, 045107 (2009).
- [17] W. Metzner and D. Vollhardt, Correlated Lattice Fermions in  $d = \infty$  Dimensions, *Phys. Rev. Lett.* **62**, 324 (1989).
- [18] Y.-X. Yao, J. Liu, C.-Z. Wang, and K.-M. Ho, Correlation matrix renormalization approximation for total-energy calculations of correlated electron systems, *Phys. Rev. B* **89**, 045131 (2014).
- [19] D. Baeriswyl, Variational schemes for many-electron systems, *Proceedings of the International Conference on Nonlinearity in Condensed Matter*, Springer Series in Solid-State Sciences Vol. 69 (Springer, New York, 1986).
- [20] D. Baeriswyl, Variational scheme for the Mott transition, *Found. Phys.* **30**, 2033 (2000).
- [21] B. Hetényi, Approximate solution of variational wave functions for strongly correlated systems: Description of bound excitons in metals and insulators, *Phys. Rev. B* **82**, 115104 (2010).
- [22] M. Dzierzawa, D. Baeriswyl, and L. M. Martelo, Variational wave functions for the Mott-Hubbard transition, *Helv. Phys. Acta* **70**, 124 (1997).
- [23] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [24] S. R. White, Density-matrix algorithms for quantum renormalization groups, *Phys. Rev. B* **48**, 10345 (1993).
- [25] G. C. Wick, The evaluation of the collision matrix, *Phys. Rev.* **80**, 268 (1950).
- [26] C. Liu, J. Liu, Y. X. Yao, P. Wu, C. Z. Wang, and K. M. Ho, Correlation matrix renormalization theory: Improving accuracy with two-electron density-matrix sum rules, *J. Chem. Theory Comput.* **12**, 4806 (2016).
- [27] M. Fishman, S. White, and E. Stoudenmire, The ITensor software library for tensor network calculations, *SciPost Phys. Codebases.* **4** (2022).
- [28] P. W. Anderson, New approach to the theory of superexchange interactions, *Phys. Rev.* **115**, 2 (1959).
- [29] J. Hubbard, Electron correlations in narrow energy bands, *Proc. R. Soc. London, Ser. A* **276**, 238 (1963).
- [30] J. Hubbard, Electron correlations in narrow energy bands. II. The degenerate band case, *Proc. R. Soc. London, Ser. A* **277**, 237 (1964).
- [31] J. Hubbard, Electron correlations in narrow energy bands III. An improved solution, *Proc. R. Soc. London, Ser. A* **281**, 401 (1964).
- [32] J. Kanamori, Electron correlation and ferromagnetism of transition metals, *Prog. Theor. Phys.* **30**, 275 (1963).
- [33] E. H. Lieb and F. Y. Wu, Absence of Mott Transition in an Exact Solution of the Short-Range, One-Band Model in One Dimension, *Phys. Rev. Lett.* **20**, 1445 (1968).
- [34] Y. Ōno, R. Bulla, and A. C. Hewson, Phase diagram of the Mott transition in a two-band Hubbard model in infinite dimensions, *Eur. Phys. J. B* **19**, 375 (2001).
- [35] A. K. Maurya, M. T. Sarder, and A. Medhi, Mott transition, magnetic and orbital orders in the ground state of the two-band Hubbard model using variational slave-spin mean field formalism, *J. Phys.: Condens. Matter* **34**, 055602 (2021).