Qubit parametrization of the variational discrete action theory for the multiorbital Hubbard model

Zhengqian Cheng and Chris A. Marianetti

Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA

The variational discrete action theory (VDAT) at $\mathcal{N} = 3$ is a potent tool for accurately capturing Mott and Hund physics at zero temperature in $d = \infty$ at a cost comparable to the Gutzwiller approximation, which is recovered by VDAT at $\mathcal{N} = 2$. Here we develop a qubit parametrization of the gauge constrained algorithm of VDAT at $\mathcal{N} = 3$ for the multiorbital Hubbard model with general density-density interactions. The qubit parametrization yields an explicit variational trial energy, and the variational parameters consist of the momentum density distribution, the shape of a reference fermi surface, and the pure state of a qubit system with dimension of the local Hilbert space. To illustrate the power of the qubit parametrization, we solve for the ground-state properties of the multiorbital Hubbard model with Hund coupling for local orbital number $N_{orb} = 2$–7. A Taylor series expansion of the partially optimized trial energy is used to explain how the Hund's coupling changes the order of the Mott transition. For the case of the $SU(2N_{orb})$ Hubbard model, an explicit approach for computing the critical U_c for the Mott transition is provided, yielding an analytical expression for U_c in the large N_{orb} limit. Additionally, we provide an analytical solution for the ground-state properties of the single-band Hubbard model with a special density of states. Finally, we demonstrate that the qubit parametrization can also be applied to $\mathcal{N} = 2$, for both G-type and B-type variants, where the G-type yields an identical expression to the slave spin mean-field theory. The qubit parametrization not only improves the efficiency and transparency of VDAT at $\mathcal{N} = 3$, but also provides the key advances for the construction of a one-body reduced density matrix functional capable of capturing Mott and Hund physics.

I. INTRODUCTION

The multiorbital Hubbard model is a minimal model of interacting electrons which accounts for key elements of typical strongly correlated electron systems [1,2], and the solution in infinite dimensions provides a natural starting point for understanding the solution in finite dimensions [3,4]. The de facto standard for accurately solving the multiorbital Hubbard model in $d = \infty$ is the dynamical mean-field theory (DMFT) [3–6], which requires self-consistently solving an effective Anderson impurity model. Therefore, the proficiency of DMFT is determined by the state-of-the-art for quantum impurity solvers [7], which are still extremely limited for the multiorbital case at zero temperature (see introduction of Ref. [8]). Alternatively, the standard for efficiently solving the multiorbital Hubbard model is the Gutzwiller approximation (GA), which semiquantitatively describes the Fermi liquid phase and the metal-insulator transition, but only produces a crude description of the insulating phase [9–14]. A long standing goal has been to improve upon GA while not incurring the computational cost of DMFT. Fortunately, the variational discrete action theory (VDAT) provides an alternate route to the exact ground-state properties in $d = \infty$ [8,15–17]. The variational ansatz of VDAT is the sequential product density matrix (SPD), which is characterized by an integer $\mathcal{N}$ and as either G-type or B-type, where G-type $\mathcal{N} = 1$ recovers the Hartree-Fock wave function, G-type $\mathcal{N} = 2$ recovers the Gutzwiller wave function, and $\mathcal{N} \to \infty$ recovers the exact solution. A remarkable fact is that the G-type $\mathcal{N} = 3$ SPD accurately captures Mott and Hund physics in $d = \infty$ [8,17],

while maintaining a computational cost comparable to the Gutzwiller wave function.

An SPD with arbitrary $\mathcal{N}$ can be exactly evaluated in $d = \infty$ using the self-consistent canonical discrete action theory (SCDA) [15,16]. Aside from the present work, there are three algorithms for executing VDAT within the SCDA, with the first two being general and the third one having some restrictions (see Table I for a summary). The first approach is the most straightforward, requiring iterations over two steps [16]. In the first step, given the variational parameters, the SCDA equations can be self-consistently solved, yielding the energy for the corresponding SPD. In the second step, the variational parameters must be updated to minimize the energy. This straightforward approach was successfully executed on the single-band Hubbard model in $d = \infty$. The second approach is the decoupled minimization algorithm [8], where one simultaneously minimizes the variational parameters and updates the SCDA equations toward self-consistency. This decoupled minimization algorithm was executed for the G-type $\mathcal{N} = 2$–4 SPDs in the two orbital Hubbard model with the full rotationally invariant local interactions, demonstrating that $\mathcal{N} = 3$ accurately describes Mott and Hund physics over all parameter space. The third approach is the gauge constrained algorithm [17], which was originally applied to the case of a G-type $\mathcal{N} = 2$ and $\mathcal{N} = 3$ SPD with kinetic projectors that are diagonal in momentum space and interacting projectors that do not introduce off-diagonal terms at the single-particle level. The key feature of the gauge constrained algorithm is that the SCDA self-consistency can be automatically satisfied, greatly simplifying the task of minimizing over the

TABLE I. Summary of the different algorithms for VDAT within the SCDA, including the standard [15,16], decoupled minimization [8], original gauge constrained [17], and the qubit parametrization of the gauge constrained algorithm. For each algorithm, the SPD type, requirement of numerical self-consistency (SC), the number of constraints per spin-orbital (Const.), type of kinetic projector (Kin. Proj.), type of interacting projector (Int. Proj.), and the optimization scheme are listed. The projectors may be general or diagonal (diag.), and the optimization scheme may be arbitrary or gradient (grad.) based.

Algo.	SPD	SC	Const.	Kin. Proj.	Int. Proj.	Opt.
Standard	any	yes	0	any	any	any
Decoupled Min.	any	yes	0	any	any	grad.
Gauge Const.	$\hat{\rho}_{G3}, \hat{\rho}_{G2}\hat{\rho}_{B2}$	no	3	diag.	diag.	any
Qubit Param.	$\hat{\rho}_{G3}, \hat{\rho}_{G2}\hat{\rho}_{B2}$	no	2	diag.	diag.	any

variational parameters. Furthermore, the momentum density distribution is used to reparametrize the variational parameters in the kinetic projectors, and the trial energy can be straightforwardly optimized over the momentum density distribution. It is useful to note that the GA can only evaluate a G-type $\mathcal{N} = 2$ SPD under a Gutzwiller gauge [17]. The SCDA has no such restrictions, which is formally appealing, but imposing certain gauge constraints allows the SCDA self-consistency to be automatically satisfied, while not reducing the variational capacity.

One shortcoming of the original gauge constrained algorithm for the G-type $\mathcal{N} = 3$ SPD is that some of the variational parameters are unintuitive and there are two linear constraints on the momentum density distribution per spin-orbital [17]. In this paper, we introduce a new parametrization of the gauge constrained algorithm, referred to as the qubit parametrization, which is mathematically equivalent to the original gauge constrained algorithm and offers several important advantages. First, there is only one linear constraint per spin-orbital on the momentum density distribution, and therefore the number of variational parameters is reduced by one per spin-orbital. Second, the variational parameters are physically intuitive and facilitate a deeper understanding of how the SPD captures Mott and Hund physics. Therefore, the qubit parametrization achieves the long sought goal of resolving the shortcomings of the Gutzwiller approximation while maintaining the computational simplicity and physical appeal. The convenience of the qubit parametrization has allowed for the construction of a one-body reduced density matrix functional for the multiorbital Hubbard model which exactly encapsulates the VDAT result at $\mathcal{N} = 3$, and this is presented in a companion manuscript [18].

It is useful to compare to existing approaches in the literature which seek to achieve our same goal of efficiently and accurately solving the ground-state properties of multiorbital Hubbard models. One approach is the ghost Gutzwiller approximation (gGA) [19], which introduces the Gutzwiller wave function in an extended Hilbert space where the GA is then applied, though this comes with a substantial increase in computational cost over the usual GA given that ghost

orbitals are introduced. In $d = \infty$, the gGA exactly evaluates the Gutzwiller wave function in the extended Hilbert space, and therefore provides an upper bound on the exact energy with increasing accuracy as the extended Hilbert space grows. The gGA has not been proven to converge to the exact solution of the multiorbital Hubbard model with increasing number of ghost orbitals, though numerical evidence suggests that this might be the case [20,21]. Empirically, it has been demonstrated that at least two ghost orbitals per physical orbital are needed for a global improvement over the GA [19–21], which substantially increases the computational cost of the gGA relative to GA. Therefore, the important practical comparison between VDAT and gGA contrasts VDAT at $\mathcal{N} = 3$ to gGA with two ghost orbitals per physical orbital. VDAT at $\mathcal{N} = 3$ has been straightforwardly applied up to the eight orbital Hubbard model [17], while the gGA has been applied up to the three orbital model [21–23] using exact diagonalization and the five orbital model [24] using density matrix renormalization group (DMRG). The success of the gGA in the multiorbital Hubbard model critically depends on the ability of DMRG or related techniques to approximately obtain a ground state in an expanded local Hilbert space. In terms of accuracy, both approaches appear to be excellent over all parameter space, though it seems possible that VDAT is more adept at capturing subtle differences between competing phases (e.g., compare Fig. 4 in Ref. [8] with Fig. 2 in Ref. [23]).

Another attempt to resolve the limitations of the GA is the slave boson approach [25], which formally recasts the Hubbard model as a Hamiltonian of constrained fermions and auxiliary bosons. The simplest saddle-point approximation is typically referred to as slave boson mean-field theory (SBMF), which exactly recovers the GA in the multiorbital Hubbard model [10] and has extensive applications [26–32]. Given that the slave boson approach is an exact formalism, there have been various results including fluctuations beyond the mean-field solution [33–39], though none of these studies address corrections to the ground-state properties in the thermodynamic limit. A related idea is the time dependent Gutzwiller approximation [40–47], which has mainly focused on computing response functions. One study did demonstrate that the ground-state properties of the single orbital Hubbard model can be improved by updating the double occupancy based on the density-density response function [47], but this would be impractical in the multiorbital Hubbard model, and we are not aware of any such applications. Therefore, neither slave bosons nor time dependent GA has provided a path beyond the GA for ground-state properties, though the idea of ghost orbitals has recently been adopted into the slave boson methodology [48], where the saddle-point solution recovers the gGA. Another approach related to SBMF is the slave spin approach [49], which recasts the Hubbard model with density-density interactions as a Hamiltonian of constrained fermions and auxiliary spins, and the saddle-point approximation also recovers the GA. As in the case of slave bosons, the slave spin approach has produced many interesting results on the multiorbital Hubbard model [28,50–55], but has not offered a path beyond the mean-field solution. In the present work, we demonstrate that the qubit parametrization for a G-type

$\mathcal{N} = 2$ SPD is identical to the slave spin mean field theory (see Appendix B).

Finally, it is useful to discuss the off-shell effective energy theory (OET) [56], which can be reinterpreted through the G-type and B-type SPD for $\mathcal{N} = 2$. OET introduces an approximation to evaluate the energy under the SPD, known as the central point expansion (CPE), and this is applied to both the G-type and B-type SPD for $\mathcal{N} = 2$. Though the CPE is intrinsically different than the SCDA, it also exactly evaluates the $\mathcal{N} = 2$ SPD in $d = \infty$, which is proven in the present work (see Appendix C). OET introduces corrections to the CPE for the G-type and B-type SPD, guaranteeing the limiting behaviors for the weak and strong coupling limits, respectively. Subsequently, one evaluates the total energy in both cases and selects the solution with lower energy. OET has yielded accurate results for the single-band Hubbard model in $d = 1$, $d = 2$, and $d = \infty$. The correction to the CPE becomes far more challenging in the multiorbital Hubbard model, which has not been pursued further. However, a similar idea of empirically correcting the Gutzwiller approximation has been developed in the context of the correlation matrix renormalization approximation [57–60].

The paper is organized in the following manner. Section II provides an overview of the SPD ansatz, the gauge symmetry of the SPD, the SCDA, and the tensor product representation for evaluating expectation values. Additionally, several methodological generalizations are provided. Section III provides a high-level overview of the qubit energy form for the G-type and B-type SPD at $\mathcal{N} = 2$ and the G-type SPD at $\mathcal{N} = 3$. Detailed derivations of the results from Sec. III are presented in Secs. IV, V, and VI. Finally, the qubit energy form is used to examine the multiorbital Hubbard model at half filling with $J/U \geqslant 0$ in Sec. VII.

II. REVIEW OF THE VARIATIONAL DISCRETE ACTION THEORY

Here we review several key ingredients of VDAT, including the SPD, the SCDA, and the tensor product representation for evaluating local expectation values within the SCDA. In addition to reviewing these concepts, we generalize the gauge symmetry of the SPD beyond the case of a diagonal transformation, and generalize the tensor product representation to arbitrary $\mathcal{N}$. These developments will facilitate the subsequent analysis in this work.

A. Sequential product density matrix

The SPD is the variational ansatz used within VDAT [15,16], and the SPD is a straightforward generalization of a variational wave function ansatz first articulated in Ref. [61]. In this work, we focus on the multiorbital Hubbard model, given as

$$\hat{H} = \hat{H}_K + \hat{H}_{\mathrm{loc}} = \sum_{k\ell} \epsilon_{k\ell} \hat{n}_{k\ell} + \sum_{i,\ell<\ell'} U_{\ell\ell'} \hat{n}_{i\ell} \hat{n}_{i\ell'}, \quad (1)$$

where k and i denote momentum and real-space site indices, respectively, $\ell = 1, \ldots, 2N_{\mathrm{orb}}$ is the spin-orbital index within a local site, and $U_{\ell\ell'}$ denotes the Coulomb interaction. In the context of the multiorbital Hubbard model, the SPD with an

integer time step $\mathcal{N} \geqslant 1$ is given as

$$\hat{\rho} = \hat{K}_1 \hat{P}_1 \ldots \hat{K}_{\mathcal{N}} \hat{P}_{\mathcal{N}}, \quad (2)$$

where $\hat{K}_\tau = \exp(\sum_{k\ell} \gamma_{k\ell\tau} \hat{n}_{k\ell})$ is the kinetic projector, $\hat{P}_\tau = \exp(\sum_{i\Gamma} \upsilon_{\Gamma\tau} \hat{X}_{i\Gamma})$ is the interacting projector, $\tau = 1, \ldots, \mathcal{N}$ is the integer time index, the diagonal Hubbard operator $\hat{X}_{i\Gamma}$ is defined as

$$\hat{X}_{i\Gamma} = \prod_{\ell=1}^{2N_{\mathrm{orb}}} \hat{X}_{i\Gamma;\ell}, \quad (3)$$

$$\hat{X}_{i\Gamma;\ell} = (1 - \hat{n}_{i\ell})\delta_{0,\Gamma(\ell)} + \hat{n}_{i\ell}\delta_{1,\Gamma(\ell)}, \quad (4)$$

and $\Gamma(\ell)$ is ℓth bit in the binary representation of $\Gamma - 1$, given as $(\Gamma(1) \ldots \Gamma(2N_{\mathrm{orb}}))_2 = \Gamma - 1$, and $\Gamma = 1, \ldots, 2^{2N_{\mathrm{orb}}}$. It should be noted that in general, $\hat{P}_\tau$ is a direct product of local projectors from all sites while $\hat{K}_\tau$ is a general noninteracting projector. There are two schemes to ensure that the SPD is Hermitian and positive semidefinite, denoted as G-type or B-type [15]. In the following, we will examine both types for $\mathcal{N} \leqslant 3$. For simplicity, we restrict $\gamma_{k\ell\tau}$ and $\upsilon_{\Gamma\tau}$ to be real numbers. It is also useful to reparametrize, up to a constant, the $\hat{K}_\tau$ and $\hat{P}_\tau$ as

$$\hat{K}_\tau = \prod_{k\ell}((1 - \lambda_{k\ell\tau})(1 - \hat{n}_{k\ell}) + \lambda_{k\ell\tau}\hat{n}_{k\ell}), \quad (5)$$

$$\hat{P}_\tau = \prod_i \left(\sum_\Gamma u_{\Gamma\tau} \hat{X}_{i\Gamma} \right), \quad (6)$$

where the $\lambda_{k\ell\tau}$ and $u_{\Gamma\tau}$ are a reparametrization of $\gamma_{k\ell\tau}$ and $\upsilon_{\Gamma\tau}$, and this form is used to derive the qubit parametrization (see Sec. III).

We first examine the $\mathcal{N} = 1$ SPD. The G-type SPD for $\mathcal{N} = 1$ is defined by $\hat{\rho}_{G1} = \hat{K}_1$, which is a noninteracting many-body density matrix. If $\hat{K}_1 = |\Psi_0\rangle\langle\Psi_0|$ where $|\Psi_0\rangle$ is a single Slater determinant, then $\hat{\rho}_{G1}$ corresponds with the Hartree-Fock wave function. Conversely, the B-type SPD for $\mathcal{N} = 1$ is defined as $\hat{\rho}_{B1} = \hat{P}_1$. Normally $\hat{\rho}_{B1}$ corresponds to a mixed state, but if $\hat{P}_1 = |\Psi_{\mathrm{at}}\rangle\langle\Psi_{\mathrm{at}}|$ with $|\Psi_{\mathrm{at}}\rangle$ defined as a direct product of atomic states from all sites, then $\hat{\rho}_{B1}$ corresponds to a pure state.

The G-type SPD with $\mathcal{N} = 2$ is defined as

$$\hat{\rho}_{G2} = \hat{P}_1 \hat{K}_2 \hat{P}_1^\dagger. \quad (7)$$

If the center projector is taken as $\hat{K}_2 = |\Psi_0\rangle\langle\Psi_0|$ with $|\Psi_0\rangle$ being a single Slater determinant, then $\hat{\rho}_{G2} = |\Psi_G\rangle\langle\Psi_G|$, where $|\Psi_G\rangle = \hat{P}_1|\Psi_0\rangle$ is the Gutzwiller wave function (GWF). The B-type SPD for $\mathcal{N} = 2$ is defined as

$$\hat{\rho}_{B2} = \hat{K}_1 \hat{P}_1 \hat{K}_1^\dagger. \quad (8)$$

If the center projector is taken as $\hat{P}_1 = |\Psi_{\mathrm{at}}\rangle\langle\Psi_{\mathrm{at}}|$ with $|\Psi_{\mathrm{at}}\rangle$ defined as a direct product of atomic states from all sites, then $\hat{\rho}_{B2} = |\Psi_B\rangle\langle\Psi_B|$, where $|\Psi_B\rangle = \exp(\sum_{k\ell} \gamma_{k\ell} \hat{n}_{k\ell})|\Psi_{\mathrm{at}}\rangle$ and we have previously referred to $|\Psi_B\rangle$ as the Bearyswil wave-function [15,16]. However, it should be noted that $|\Psi_B\rangle$ is distinct from the original Bearyswil wave-function $|\Psi'_B\rangle$ [62], which is defined as $|\Psi'_B\rangle = \exp(\alpha \sum_{k\ell} \epsilon_{k\ell} \hat{n}_{k\ell})|\Psi_\infty\rangle$, where α is a variational parameter and $|\Psi_\infty\rangle$ is the fully projected GWF. Though $|\Psi'_B\rangle$ is technically a special case of a G-type $\mathcal{N} = 3$ SPD, which can also be evaluated using the SCDA, it yields the same energy as $|\Psi_B\rangle$ for the insulating phase in

$d = \infty$ for the multiorbital Hubbard model (see Sec. VI D 3 and Fig. 3).

The G-type SPD for $\mathcal{N} = 3$ combines the variational power of $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$, and is defined as

$$\hat{\rho}_{G3} = \hat{K}_1 \hat{P}_1 \hat{K}_2 \hat{P}_1^\dagger \hat{K}_1^\dagger. \tag{9}$$

If $\hat{K}_2$ corresponds to a single Slater determinant $|\Psi_0\rangle$, then $\hat{\rho}_{G3} = |\Psi_{GB}\rangle\langle\Psi_{GB}|$, where $|\Psi_{GB}\rangle = \exp(\sum_{k\ell} \gamma_{k\ell} \hat{n}_{k\ell})\hat{P}_1|\Psi_0\rangle$ is a mild generalization of the original Gutzwiller-Baeriswyl wave function $|\Psi'_{GB}\rangle$ introduced by Otsuka (see Eq. (2.3) in Ref. [63]), defined as $|\Psi'_{GB}\rangle = \exp(\alpha \sum_{k\ell} \epsilon_{k\ell} \hat{n}_{k\ell})\hat{P}_1|\Psi_0\rangle$, where α is a variational parameter. Finally, the $\mathcal{N} = 3$ B-type SPD is defined as $\hat{\rho}_{B3} = \hat{P}_1 \hat{K}_2 \hat{P}_2 \hat{K}_2^\dagger \hat{P}_1^\dagger$. If $\hat{P}_2 = |\Psi_{at}\rangle\langle\Psi_{at}|$, then $\hat{\rho}_{B3} = |\Psi_{BG}\rangle\langle\Psi_{BG}|$ where $|\Psi_{BG}\rangle = \hat{P}_1 \exp(\sum_{k\ell} \gamma_{k\ell} \hat{n}_{k\ell})|\Psi_{at}\rangle$ and we have previously referred to $|\Psi_{BG}\rangle$ as the Baeriswyl-Gutzwiller wave-function. However, it should be noted that $|\Psi_{BG}\rangle$ is distinct from the original Baeriswyl-Gutzwiller wave-function $|\Psi'_{BG}\rangle$ [61], which is defined as $|\Psi'_{BG}\rangle = \hat{P}_1 \exp(\alpha \sum_{k\ell} \epsilon_{k\ell} \hat{n}_{k\ell})|\Psi_\infty\rangle$, where α is a variational parameter and $|\Psi_\infty\rangle$ is the fully projected GWF. The $|\Psi'_{BG}\rangle$ is technically a special case of a G-type $\mathcal{N} = 4$ SPD, which can also be evaluated using the SCDA.

It is useful to understand how the SPD recovers the exact solution with increasing $\mathcal{N}$ [15]. The Trotter-Suzuki decomposition [64], given as

$$e^{-\beta H} \approx \left(e^{-\frac{\beta}{\mathcal{N}}(\sum_{k\ell} \epsilon_{k\ell} \hat{n}_{k\ell})} e^{-\frac{\beta}{\mathcal{N}}(\sum_{i,\ell<\ell'} U_{\ell\ell'} \hat{n}_{i\ell} \hat{n}_{i\ell'})}\right)^{\mathcal{N}}, \tag{10}$$

can be considered as a special case of the SPD where the variational parameters are set by the parameters of the Hamiltonian and the temperature. Given that the large $\mathcal{N}$ Trotter-Suzuki decomposition recovers the exact finite temperature density matrix, the SPD in the large $\mathcal{N}$ limit possesses sufficient variational power to solve the Hamiltonian exactly. However, for a finite $\mathcal{N}$, the SPD leverages the full variational freedom while maintaining the integer time structure, leading to an observed exponential decrease of the energy error with increasing $\mathcal{N}$ [8,16]. In contrast, the energy error for the Trotter-Suzuki decomposition diminishes polynomially with increasing $\mathcal{N}$ [64]. This key difference is the basis for the accuracy of the SPD for small $\mathcal{N}$.

Our initial efforts using VDAT have been exactly evaluating the SPD in $d = \infty$ via the SCDA, and this same approach can be used as a local approximation in finite dimensions. However, there is an existing body of literature which used variational quantum Monte Carlo (VMC) to evaluate a simplified SPD in the Hubbard model. In the case of $\hat{\rho}_{G2}$, it is most convenient to execute VMC in real space, which has been used to evaluate the Gutzwiller wave function in the Hubbard model [65–67]. For $\mathcal{N} > 2$ in the one and two-dimensional Hubbard model, a restricted form of the SPD has been evaluated using quantum Monte Carlo to solve the single-orbital Hubbard model in two dimensions [63,68–78], the p-d model [79,80], and selected molecules [81].

B. Gauge symmetry of the SPD

A important aspect of the SPD is the corresponding gauge symmetry [8,17], which will prove to be an important tool for reducing the redundancy of the parametrization of the SPD. Consider the general SPD $\hat{\rho} = \hat{K}_1 \hat{P}_1 \ldots \hat{K}_\mathcal{N} \hat{P}_\mathcal{N}$, where $\hat{K}_\tau =$

$\exp(\boldsymbol{\gamma}_\tau \cdot \hat{\boldsymbol{n}})$ is a kinetic projector and $\boldsymbol{\gamma} \cdot \hat{\boldsymbol{n}} \equiv \sum_{\ell\ell'} [\boldsymbol{\gamma}]_{\ell\ell'} [\hat{\boldsymbol{n}}]_{\ell\ell'}$ and $[\hat{\boldsymbol{n}}]_{\ell\ell'} = \hat{a}_\ell^\dagger \hat{a}_{\ell'}$, and $\hat{P}_\tau$ is a general interacting projector. The essence of the gauge symmetry lies in the fact that only the total product $\hat{\rho}$ determines physical observables, while there is considerable flexibility in the decomposition of $\hat{\rho}$ into different time steps and the separation between kinetic and interacting projectors. To analyze the gauge symmetry, we utilize the discrete action theory [15], which is a formalism for studying integer time correlation functions of the SPD. The discrete action theory can be seen as a generalization of the many-body Green's function formalism to the case of integer time, including an integer time Dyson equation given as

$$\boldsymbol{g}^{-1} - \mathbf{1} = (\boldsymbol{g}_0^{-1} - \mathbf{1})\boldsymbol{S}, \tag{11}$$

where $\boldsymbol{g}_0$ is the noninteracting integer time Green's functions, $\boldsymbol{g}$ is the interacting integer time Green's functions, and $\boldsymbol{S} = \exp(-\boldsymbol{\Sigma}^T)$ is the exponential form of the integer time self-energy, where $\boldsymbol{\Sigma}$ is the integer time self-energy. The integer time Dyson equation recovers the usual Dyson when the SPD is chosen as the Trotter-Suzuki decomposition in the large $\mathcal{N}$ limit [15]. The $\boldsymbol{g}_0$ and $\boldsymbol{g}$ are naturally defined in the compound space [15] as $\boldsymbol{g}_0 = \langle \hat{\boldsymbol{n}} \rangle_{\hat{\varrho}_0}$ and $\boldsymbol{g} = \langle \hat{\boldsymbol{n}} \rangle_{\hat{\varrho}}$, where $[\hat{\boldsymbol{n}}]_{\ell\tau,\ell'\tau'} = \underline{\hat{a}}_\ell^{\dagger(\tau)} \underline{\hat{a}}_{\ell'}^{(\tau')}$, the underbar indicates an operator in the compound space, $\ell = 1, \ldots, L$ enumerates the spin-orbitals for the entire system, and $\tau = 1, \ldots, \mathcal{N}$ labels the integer time step. The discrete Dyson equation is a matrix equation of dimension of $L\mathcal{N} \times L\mathcal{N}$, and it exactly relates the interacting and noninteracting integer time Green's function via $\boldsymbol{S}$. Given that $\boldsymbol{S}$ is more convenient to work with, we refer to $\boldsymbol{S}$ as the integer time self-energy for brevity. For the convenience of discussing the gauge transformation of the SPD, it is useful to define several additional quantities

$$\boldsymbol{S}_0 = \boldsymbol{S}_Q \boldsymbol{S}_K, \tag{12}$$

$$\boldsymbol{S}_F = \boldsymbol{S}_Q \boldsymbol{S}_K \boldsymbol{S}, \tag{13}$$

where $[\boldsymbol{S}_Q]_{\tau\tau'} = (-\delta_{\tau+1,\tau'} + \delta_{\tau-\mathcal{N}+1,\tau'})\mathbf{1}$ and $[\boldsymbol{S}_K]_{\tau\tau'} = \delta_{\tau\tau'} \exp(-\boldsymbol{\gamma}_\tau^T)$ are $L \times L$ matrices. Using these definitions, $\boldsymbol{g}_0$ and $\boldsymbol{g}$ can be defined succinctly as

$$\boldsymbol{g}_0 = (\mathbf{1} + \boldsymbol{S}_0)^{-1}, \tag{14}$$

$$\boldsymbol{g} = (\mathbf{1} + \boldsymbol{S}_F)^{-1}. \tag{15}$$

Generally, there are two types of gauge transformations of the SPD: intra-time-step and inter-time-step. The intra-time-step transformation occurs at the boundary of the kinetic and interacting projector for time step τ in the following way: $\hat{K}_\tau \rightarrow \hat{K}_\tau \exp(\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}})$ and $\hat{P}_\tau \rightarrow \exp(-\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}})\hat{P}_\tau$. Since the intra-time-step transformation does not alter the measurement at integer times, $\boldsymbol{g}$ is invariant to this transformation. However, given that $\hat{K}_\tau$ changes after the intra-time-step transformation, we have $\boldsymbol{S}'_K = \boldsymbol{S}_K \boldsymbol{N}_a$, where $[\boldsymbol{N}_a]_{\tau\tau'} = \delta_{\tau\tau'} \exp(-\boldsymbol{\mu}_\tau^T)$, while $\boldsymbol{S}'_F = \boldsymbol{S}_F$. Conversely, the inter-time-step transformation is defined at the boundary of the interacting projector at τ and the kinetic projector at $\tau + 1$ in the following way: $\hat{P}_\tau \rightarrow \hat{P}_\tau \exp(-\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}})$ and $\hat{K}_{\tau+1} \rightarrow \exp(\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}})\hat{K}_{\tau+1}$, where $\boldsymbol{\mu}_\mathcal{N} \equiv 0$. We first discuss how $\boldsymbol{g}$ changes under this inter-time-step transformation, and it is useful to note the following

identities (for details, see Eqs. (A22) and (A23) in Ref. [15] and Eqs. (A1)–(A4) in Ref. [8]):

$$e^{-\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}}} \hat{a}_\ell^\dagger e^{\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}}} = \sum_{\ell'} \left[e^{-\boldsymbol{\mu}_\tau^T} \right]_{\ell\ell'} \hat{a}_{\ell'}^\dagger, \tag{16}$$

$$e^{-\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}}} \hat{a}_\ell e^{\boldsymbol{\mu}_\tau \cdot \hat{\boldsymbol{n}}} = \sum_{\ell'} \hat{a}_{\ell'} \left[e^{\boldsymbol{\mu}_\tau^T} \right]_{\ell'\ell}. \tag{17}$$

Using the above two equations, we can determine the transformation $\boldsymbol{g}' = N_b \boldsymbol{g} N_b^{-1}$, where $[N_b]_{\tau\tau'} = \delta_{\tau\tau'} \exp(-\boldsymbol{\mu}_\tau^T)$, and correspondingly $\boldsymbol{S}'_F = N_b \boldsymbol{S}_F N_b^{-1}$. Using the transformation for the kinetic projector, we have $\boldsymbol{S}'_K = \tilde{N}_b \boldsymbol{S}_K$, where $[\tilde{N}_b]_{\tau\tau'} = \delta_{\tau\tau'} \exp(-\boldsymbol{\mu}_{\tau-1}^T)$ and $\boldsymbol{\mu}_0 \equiv 0$, and $\tilde{N}_b$ can be rewritten as $\tilde{N}_b = S_Q^{-1} N_b S_Q$. It should be noted that a general gauge transformation includes both intra-time-step and inter-time-step transformations, which can be parametrized by N_a and N_b, and the above results can be synthesized as

$$\boldsymbol{g}' = N_b \boldsymbol{g} N_b^{-1}, \tag{18}$$

$$\boldsymbol{S}'_F = (\boldsymbol{g}')^{-1} - 1 = N_b \boldsymbol{S}_F N_b^{-1}, \tag{19}$$

$$\boldsymbol{S}'_K = \tilde{N}_b \boldsymbol{S}_K N_a = S_Q^{-1} N_b S_Q \boldsymbol{S}_K N_a. \tag{20}$$

Moreover, the form of the integer time Dyson equation is invariant after the transformation, which requires

$$\boldsymbol{S}_0 \boldsymbol{S} = S_Q \boldsymbol{S}_K \boldsymbol{S} = \boldsymbol{S}_F, \tag{21}$$

$$\boldsymbol{S}'_0 \boldsymbol{S}' = S_Q \boldsymbol{S}'_K \boldsymbol{S}' = \boldsymbol{S}'_F. \tag{22}$$

Using Eq. (20), we have

$$\boldsymbol{S}'_0 = S_Q \boldsymbol{S}'_K = N_b S_Q \boldsymbol{S}_K N_a = N_b \boldsymbol{S}_0 N_a, \tag{23}$$

while using Eqs. (23), (19), (21), and (22), we have

$$\boldsymbol{S}' = (\boldsymbol{S}'_0)^{-1} \boldsymbol{S}'_F = (N_b \boldsymbol{S}_0 N_a)^{-1} N_b \boldsymbol{S}_F N_b^{-1} = N_a^{-1} \boldsymbol{S} N_b^{-1}, \tag{24}$$

which can also be directly verified by assuming $\hat{P}_\tau$ is noninteracting. Correspondingly, the noninteracting integer time Green's function changes as

$$\boldsymbol{g}'_0 = (1 + \boldsymbol{S}'_0)^{-1} = \left(1 + N_b (\boldsymbol{g}_0^{-1} - 1) N_a \right)^{-1}. \tag{25}$$

Notice that Eqs. (18)–(25) are exact relations. Within the SCDA [15], $\mathcal{G}$ will transform similarly to $\boldsymbol{g}_0$, where N_a and N_b are local. These gauge transformations provide a rigorous theoretical foundation for simplifying the task of achieving self-consistency within the SCDA. Moreover, choosing the appropriate gauge condition will also help to improve the numerical stability when minimizing over the variational parameters.

C. Review of the SCDA

The discrete action theory is a formalism designed to study integer time correlation functions of the SPD, and the SCDA is the integer time analog of the dynamical mean-field theory [15,16]. The SCDA exactly evaluates the SPD in infinite dimensions, and can be used as a local approximation for the SPD in finite dimensions. To appreciate the idea of the SCDA, we define the discrete action which encodes the integer time

correlation of the SPD as

$$\underline{\varrho} = \underline{\varrho}_0 \hat{\underline{P}}, \tag{26}$$

where $\underline{\varrho}_0 = \hat{\underline{Q}} \prod_{\tau=1}^{\mathcal{N}} \hat{\underline{K}}_\tau^{(\tau)}$ and $\hat{\underline{P}} = \prod_{\tau=1}^{\mathcal{N}} \hat{\underline{P}}_\tau^{(\tau)}$ are the noninteracting and interacting parts of the discrete action $\underline{\varrho}$, and the underbar denotes operators in the compound space.

The SCDA can be formulated as a scheme to evaluate the kinetic energy and local interacting energy through two self-consistent approximations of $\underline{\varrho}$ [8], denoted as $\hat{\rho}_K$ and $\underline{\hat{\rho}}_{\text{loc}}$, respectively. Within the SCDA, the total energy is given as

$$\langle \underline{\hat{H}}^{(\mathcal{N})} \rangle_{\underline{\varrho}} = \langle \hat{\underline{H}}_K^{(\mathcal{N})} \rangle_{\hat{\rho}_K} + \langle \hat{\underline{H}}_{\text{loc}}^{(\mathcal{N})} \rangle_{\hat{\rho}_{\text{loc}}}, \tag{27}$$

where $\hat{\rho}_K$ and $\hat{\rho}_{\text{loc}}$ are defined as

$$\underline{\hat{\rho}}_K = \underline{\hat{\varrho}}_0 \exp(-\ln \boldsymbol{S}^T \cdot \hat{\underline{\boldsymbol{n}}}), \tag{28}$$

$$\underline{\hat{\rho}}_{\text{loc}} = \exp(-\ln(\mathcal{G}^{-1} - 1)^T \cdot \hat{\underline{\boldsymbol{n}}}) \hat{\underline{P}}, \tag{29}$$

where $\mathcal{G}$ is the noninteracting integer time Green's function for the local canonical discrete action $\hat{\rho}_{\text{loc}}$. Both $\boldsymbol{S}$ and $\mathcal{G}$ are $L\mathcal{N} \times L\mathcal{N}$ matrices, where L is the number of spin-orbitals of the entire system. In the SCDA, the key assumption is that $\boldsymbol{S}$ and $\mathcal{G}$ are block diagonal in real space clusters, given by $[\boldsymbol{S}]_{ij} = \boldsymbol{S}_i \delta_{ij}$ and $[\mathcal{G}]_{ij} = \mathcal{G}_i \delta_{ij}$ where i and j are real space cluster indices. To determine $\boldsymbol{S}_i$ and $\mathcal{G}_i$, we have the following two sets of conditions for each cluster index i,

(1) The integer time Dyson equation, given as

$$\left(\boldsymbol{g}_i^{-1} - 1 \right) = \left(\mathcal{G}_i^{-1} - 1 \right) \boldsymbol{S}_i, \tag{30}$$

provides a way to compute $\boldsymbol{S}_i$ by evaluating $\hat{\rho}_{\text{loc};i} = \text{Tr}_{/i} \hat{\rho}_{\text{loc}}$, which can be expressed as

$$\hat{\rho}_{\text{loc};i} = \exp\left(-\ln \left(\mathcal{G}_i^{-1} - 1 \right)^T \cdot \hat{\underline{\boldsymbol{n}}}_i \right) \hat{\underline{P}}_i. \tag{31}$$

(2) The self-consistency condition for the local integer time Green's function, given as

$$\boldsymbol{g}_i = \boldsymbol{g}'_i, \tag{32}$$

where $\boldsymbol{g}_i = \langle \hat{\underline{\boldsymbol{n}}}_i \rangle_{\hat{\rho}_{\text{loc}}}$ and $\boldsymbol{g}'_i = \langle \hat{\underline{\boldsymbol{n}}}_i \rangle_{\hat{\rho}_K}$.

It is interesting to examine the sufficiency of Eqs. (30) and (32) to determine both $\mathcal{G}$ and $\boldsymbol{S}$ by counting the number of unknown entries and equations. We denote the number of sites in the lattice as N_{site} and the number of spin orbitals on site i as N_i. The number of unknowns in $\mathcal{G}$ and $\boldsymbol{S}$ is $2 \sum_{i=1}^{N_{\text{site}}} (N_i \mathcal{N})^2$. For a given i, the two matrix equations given by Eqs. (30) and (32) provide $2(N_i \mathcal{N})^2$ entries. Therefore, we have a sufficient number of equations to determine $\mathcal{G}$ and $\boldsymbol{S}$.

D. Tensor product representation for expectation values under $\hat{\rho}_{\text{loc};i}$

The key computational cost of implementing the SCDA is to compute local observables under $\hat{\rho}_{\text{loc};i}$, and a convenient approach was devised for $\mathcal{N} = 3$ using a tensor product representation [17]. Here we generalize the tensor product representation to arbitrary $\mathcal{N}$. We assume that the interacting projector has the form $\hat{P}_{\tau;i} = \sum_\Gamma u_{\tau i\Gamma} \hat{X}_{i\Gamma}$ and $\mathcal{G}_i$ is diagonal in the spin orbital index ℓ, given as $[\mathcal{G}_i]_{\ell\ell} = \delta_{\ell\ell'} \mathcal{G}_{i,\ell}$, where $\mathcal{G}_{i,\ell}$ is an $\mathcal{N} \times \mathcal{N}$ matrix. For a local operator $\hat{\underline{O}}$ at site i, the

expectation value can be written as

$$\langle \hat{\underline{O}} \rangle_{\underline{\rho}_{\text{loc},i}} = \frac{\langle \hat{\underline{P}}_i \hat{\underline{O}} \rangle_{\underline{\rho}_{\text{loc},i,0}}}{\langle \hat{\underline{P}}_i \rangle_{\underline{\rho}_{\text{loc},i,0}}} = \frac{(\hat{\underline{O}})_u \cdot (\otimes_\tau u_\tau)}{(\hat{\underline{1}})_u \cdot (\otimes_\tau u_\tau)}, \tag{33}$$

where the noninteracting discrete action for $\underline{\rho}_{\text{loc},i,0}$ is given by

$$\underline{\rho}_{\text{loc},i,0} = \exp\left(- \ln\left(\mathcal{G}_i^{-1} - \mathbf{1} \right)^T \cdot \hat{\underline{n}}_i \right), \tag{34}$$

where the vector $u_\tau = (u_{\tau i1}, \ldots, u_{\tau i N_\Gamma})$ enumerates over the index Γ with N_Γ as the number of local diagonal Hubbard operators at site i, the $\mathcal{N}$-dimensional tensor $(\hat{\underline{O}})_u$ is defined as

$$[(\hat{\underline{O}})_u]_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}} = \left\langle \left(\prod_{\tau=1}^{\mathcal{N}} \hat{\underline{X}}_{i\Gamma_\tau}^{(\tau)} \right) \hat{\underline{O}} \right\rangle_{\underline{\rho}_{\text{loc},i,0}}, \tag{35}$$

the direct product is defined as $[\otimes_{\tau=1}^{\mathcal{N}} u_\tau]_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}} = \prod_{\tau=1}^{\mathcal{N}} u_{\tau i \Gamma_\tau}$, and $\mathbf{A} \cdot \mathbf{B}$ is the contraction between two $\mathcal{N}$-dimensional tensors as

$$\mathbf{A} \cdot \mathbf{B} = \sum_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}} [\mathbf{A}]_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}} [\mathbf{B}]_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}}. \tag{36}$$

We now consider operators that can be written as a product over the spin orbitals, given as $\hat{\underline{O}} = \prod_\ell \hat{\underline{O}}_\ell$. Assuming that $\mathcal{G}_i$ is diagonal in the spin orbital index ℓ, we have

$$[(\hat{\underline{O}})_u]_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}} = \prod_{\ell=1}^{2N_{\text{orb}}} \left\langle \left(\prod_{\tau=1}^{\mathcal{N}} \hat{\underline{X}}_{i\Gamma_\tau;\ell}^{(\tau)} \right) \hat{\underline{O}}_\ell \right\rangle_{\underline{\rho}_{\text{loc},i}}, \tag{37}$$

where $\hat{X}_{i\Gamma;\ell}$ is defined in Eq. (4). Therefore, we can write $(\hat{\underline{O}})_u$ as a direct product

$$(\hat{\underline{O}})_u = \otimes_{\ell=1}^{2N_{\text{orb}}} (\hat{\underline{O}}_\ell)_{u;\ell}, \tag{38}$$

where the direct product is defined as

$$\left[\otimes_{\ell=1}^{2N_{\text{orb}}} \mathbf{M}_\ell \right]_{\Gamma_1, \ldots, \Gamma_{\mathcal{N}}} = \prod_{\ell=1}^{2N_{\text{orb}}} [\mathbf{M}_\ell]_{\Gamma_1(\ell), \ldots, \Gamma_{\mathcal{N}}(\ell)}, \tag{39}$$

with $(\hat{\underline{O}}_\ell)_{u;\ell}$ defined as

$$[(\hat{\underline{O}}_\ell)_{u;\ell}]_{\Gamma_1(\ell), \ldots, \Gamma_{\mathcal{N}}(\ell)} \equiv \left\langle \left(\prod_{\tau=1}^{\mathcal{N}} \hat{\underline{X}}_{i\Gamma_\tau;\ell}^{(\tau)} \right) \hat{\underline{O}}_\ell \right\rangle_{\underline{\rho}_{\text{loc},i,0}}, \tag{40}$$

which only depends $[\mathcal{G}_i]_{\ell\ell}$. For some cases, there is no interacting projector on a given time step, and the corresponding dimension of $(\hat{\underline{O}}_\ell)_{u;\ell}$ can be traced out. Moreover, Eq. (33) can be conveniently rewritten into a matrix product for all cases we study, which will be further discussed in Secs. IV, V, and VI.

III. OVERVIEW OF THE QUBIT ENERGY FORM

The SCDA can exactly evaluate an SPD in $d = \infty$ with arbitrary $\mathcal{N}$ via a self-consistent solution, as outlined in Sec. II, and therefore the total energy can only be numerically evaluated for a given set of variational parameters in general. However, the recently developed gauge constrained SCDA algorithm [17] allows for an explicit evaluation of a G-type SPD with $\mathcal{N} \leqslant 3$, circumventing the need for self-consistent

solution (see Sec. VI B for a review). In the present work, we offer a further refinement of the gauge constrained SCDA by transforming all variational parameters into physically intuitive variables, referred to as the qubit parametrization. Moreover, the qubit parametrization analytically resolves one constraint, reducing the number of variational parameters by one per spin orbital. Given the complexity of deriving the qubit parametrization, we collect all key results in this section, providing a self-contained presentation of all details needed to implement a practical calculation. A G-type and B-type SPD at $\mathcal{N}$ will be denoted as $\hat{\rho}_{G\mathcal{N}}$ and $\hat{\rho}_{B\mathcal{N}}$, respectively. For pedagogical purposes, we first present results for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$ before finally considering $\hat{\rho}_{G3}$, illustrating that $\hat{\rho}_{G3}$ can be seen as a unification of $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$. Detailed derivations of all results for $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, and $\hat{\rho}_{G3}$ are provided in Secs. IV, V, and VI, respectively, while explicit applications are presented in Sec. VII. It should be emphasized that all qubit parametrization results are mathematically identical to corresponding VDAT results.

For the sake of generality, we consider the multiorbital Hubbard Hamiltonian given as

$$\hat{H} = \sum_{k\ell} \epsilon_{k\ell} \hat{n}_{k\ell} + \sum_i H_{\text{loc}}(\{\hat{n}_{i\ell}\}), \tag{41}$$

where H_{loc} is a polynomial function of the local density operators, given as

$$H_{\text{loc}}(\{\hat{n}_{i\ell}\}) = \sum_\ell c_\ell \hat{n}_{i\ell} + \sum_{\ell_1 \ell_2} c_{\ell_1 \ell_2} \hat{n}_{i\ell_1} \hat{n}_{i\ell_2} \\ + \sum_{\ell_1 \ell_2 \ell_3} c_{\ell_1 \ell_2 \ell_3} \hat{n}_{i\ell_1} \hat{n}_{i\ell_2} \hat{n}_{i\ell_3} + \cdots, \tag{42}$$

where $c_{\ell_1 \ldots \ell_n}$ are parameters that define the n-body interactions. For a typical Hubbard model, only two-body interactions will be needed.

We begin by defining the local Hilbert space and the corresponding fermionic operators, and the corresponding map to qubit operators via the Jordan-Wigner transformation. For a given site i, the atomic configurations $\{|\Gamma\rangle\}$ form a complete basis for the local Hilbert space with dimension $2^{2N_{\text{orb}}}$, and we use $\hat{a}_\ell^\dagger$ and $\hat{a}_\ell$ to represent the creation and annihilation operators for spin-orbital ℓ. Then we express the fermionic operators in terms of spin operators via the Jordan-Wigner transformation defined as

$$\hat{a}_\ell^\dagger = \left(\prod_{\ell' < \ell} \hat{\sigma}_{\ell'}^z \right) \hat{\sigma}_\ell^-, \tag{43}$$

$$\hat{a}_\ell = \left(\prod_{\ell' < \ell} \hat{\sigma}_{\ell'}^z \right) \hat{\sigma}_\ell^+, \tag{44}$$

$$\hat{a}_\ell^\dagger \hat{a}_\ell = \hat{n}_\ell = \frac{1}{2}\left(1 - \hat{\sigma}_\ell^z \right), \tag{45}$$

where $\hat{\sigma}_\ell^\pm = (\hat{\sigma}_\ell^x \pm i\hat{\sigma}_\ell^y)/2$ and $\hat{\sigma}_\ell^\mu$ is defined as

$$\hat{\sigma}_\ell^\mu = \hat{I}_{2^{\ell-1}} \otimes \hat{\sigma}^\mu \otimes \hat{I}_{2^{N_{\text{orb}}-\ell}}, \tag{46}$$

where $\mu \in \{x, y, z\}$ and $\hat{\sigma}^\mu$ is a standard 2×2 Pauli matrix. It should be noted that we choose a convention for the Jordan-Wigner transformation such that $|0\rangle$ is associated with $|\uparrow\rangle$ and

$|1\rangle$ is associated with $|\downarrow\rangle$. This local Jordan-Wigner transformation can be used to transform the usual Gutzwiller energy form into the qubit energy form, though spin operators will naturally emerge when using the tensor representation of local observables within the SCDA.

We proceed by evaluating the ground-state energy under $\hat{\rho}_{G2}$, where $\hat{\rho}_{G2}$ is parametrized as

$$\hat{\rho}_{G2} = \hat{P}(\{u_\Gamma\})\hat{K}(\{n_{k\ell,0}\})\hat{P}(\{u_\Gamma\}), \quad (47)$$

$$\hat{P}(\{u_\Gamma\}) = \prod_i \left(\sum_\Gamma u_\Gamma \hat{X}_{i\Gamma} \right), \quad (48)$$

$$\hat{K}(\{n_{k\ell,0}\}) = \prod_{k\ell}((1 - n_{k\ell,0})(1 - \hat{n}_{k\ell}) + n_{k\ell,0}\hat{n}_{k\ell}), \quad (49)$$

where the variational parameters $\{u_\Gamma\}$ and $\{n_{k\ell,0}\}$ are real and $n_{k\ell,0} \in [0, 1]$. The variational parameters $\{u_\Gamma\}$ can be reparametrized as a pure state of an effective $2N_{\text{orb}}$ qubit system (i.e., a $2N_{\text{orb}}$ spin $1/2$ system) characterized by a many-body density matrix $\boldsymbol{\rho}$. The total trial energy $E(\boldsymbol{\rho}, \{n_{k\ell,0}\})$ for $\hat{\rho}_{G2}$ can then be written as

$$E(\boldsymbol{\rho}, \{n_{k\ell,0}\}) = \sum_\ell \int dk \epsilon_{k\ell} n_{k\ell} + \langle H_{\text{loc}}(\{\hat{n}_\ell\})\rangle_{\boldsymbol{\rho}}, \quad (50)$$

$$n_{k\ell} = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}} + \frac{\xi_\ell^2}{\xi_{\ell,0}^2}(n_{k\ell,0} - \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}), \quad (51)$$

$$\xi_\ell = \frac{1}{2}\langle \hat{\sigma}_\ell^x \rangle_{\boldsymbol{\rho}}, \quad (52)$$

$$\xi_{\ell,0} = \sqrt{\langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}\langle 1 - \hat{n}_\ell \rangle_{\boldsymbol{\rho}}}, \quad (53)$$

with a constraint for each ℓ,

$$\int dk n_{k\ell;0} = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}. \quad (54)$$

Here we have taken the continuum limit of the discretized $n_{k\ell}$ and choose the convention $\int dk = 1$. Equations (50)–(54) give an *explicit* functional form for the trial energy as a function of the variational parameters. The qubit energy form in Eqs. (50)–(54) is equivalent to a previous result obtained using the slave spin mean-field method [49] (see Appendix B for a detailed comparison). It should be emphasized that Eqs. (50)–(54) are simply a transformation of the SCDA algorithm, and therefore are completely equivalent to the usual Gutzwiller approximation.

We now proceed to study $\hat{\rho}_{B2}$, which is the dual of $\hat{\rho}_{G2}$, given as

$$\hat{\rho}_{B2} = \hat{K}(\{\lambda_{k\ell}\})\hat{P}(\{u_\Gamma\})\hat{K}(\{\lambda_{k\ell}\}), \quad (55)$$

where the variational parameters $\{u_\Gamma\}$ and $\{\lambda_{k\ell}\}$ are real and $u_\Gamma \geqslant 0$. The qubit parametrization reparametrizes the $\{u_\Gamma\}$ and $\{\lambda_{k\ell}\}$ into $\boldsymbol{\rho}$ and $n_{k\ell} \in [0, 1]$, where $\boldsymbol{\rho}$ is a diagonal, positive semidefinite $2^{N_{\text{orb}}} \times 2^{N_{\text{orb}}}$ matrix and $n_{k\ell}$ is the physical single particle density matrix, yielding the following total energy:

$$E(\boldsymbol{\rho}, \{n_{k\ell}\}) = \sum_\ell \int dk \epsilon_{k\ell} n_{k\ell} + \langle H_{\text{loc}}(\{\hat{n}_{\text{eff},\ell}\})\rangle_{\boldsymbol{\rho}}, \quad (56)$$

$$\hat{n}_{\text{eff},\ell} = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}} + \frac{A_\ell^2}{\xi_{\ell,0}^2}(\hat{n}_\ell - \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}), \quad (57)$$

$$A_\ell = \int dk\sqrt{n_{k\ell}(1 - n_{k\ell})}, \quad (58)$$

$$\xi_{\ell,0} = \sqrt{\langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}\langle 1 - \hat{n}_\ell \rangle_{\boldsymbol{\rho}}}, \quad (59)$$

with a constraint for each ℓ

$$\int dk n_{k\ell} = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}. \quad (60)$$

It should be noted that the operators $\hat{n}_{\text{eff},\ell}$ and $\hat{n}_\ell$ have the same expectation values under $\boldsymbol{\rho}$, which is evident from Eq. (57). This energy form provides a minimal description for the Mott insulating state in the multiorbital Hubbard model in $d = \infty$, much like the result of the Gutzwiller approximation for the metallic phase.

We now proceed to study $\hat{\rho}_{G3}$, which combines the variational capacity of both $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$. The key result of this paper is to recast the previously obtained explicit evaluation of the energy under $\hat{\rho}_{G3}$ into a physically intuitive form which can be viewed as a combination of the variational parameters of the qubit energy form for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$. The $\hat{\rho}_{G3}$ is given as

$$\hat{\rho}_{G3} = \hat{K}(\{\lambda_{k\ell}\})\hat{P}(\{u_\Gamma\})\hat{K}(\{n_{k\ell,0}\})\hat{P}(\{u_\Gamma\})\hat{K}(\{\lambda_{k\ell}\}), \quad (61)$$

where the variational parameters $\{\lambda_{k\ell}\}$, $\{u_\Gamma\}$, and $\{n_{k\ell,0}\}$ are real and $n_{k\ell;0} \in [0, 1]$. As demonstrated in Ref. [17], when $\hat{K}(\{n_{k\ell,0}\})$ corresponds to a Slater determinant (i.e., $n_{k\ell;0} = 0$ or $n_{k\ell;0} = 1$), the total energy can be explicitly evaluated, and therefore we follow this condition. Under this restricted form of $\{n_{k\ell,0}\}$, it is useful to introduce the concept of a reference Fermi surface, which delineates the boundary between $n_{k\ell;0} = 0$ and $n_{k\ell;0} = 1$ throughout the Brillouin zone. We can reparametrize $\{\lambda_{k\ell}\}$ and $\{u_\Gamma\}$ in terms of $\{n_{k\ell}\}$ and $\boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is $2^{N_{\text{orb}}} \times 2^{N_{\text{orb}}}$ matrix corresponding to a pure state of a $2N_{\text{orb}}$ qubit system and $n_{k\ell} \in [0, 1]$ is the physical momentum density distribution. The total trial energy is given as

$$E(\boldsymbol{\rho}, \{n_{k\ell}\}, \{n_{k\ell,0}\}) = \sum_\ell \int dk \epsilon_{k\ell} n_{k\ell} + \langle H_{\text{loc}}(\{\hat{n}_{\text{eff},\ell}\})\rangle_\rho, \quad (62)$$

$$\hat{n}_{\text{eff},\ell} = f_{\ell,0} + f_{\ell,x}\hat{\sigma}_\ell^x + f_{\ell,z}\hat{\sigma}_\ell^z, \quad (63)$$

and there are two constraints for each ℓ given as

$$\int_< dk = \int dk n_{k\ell} = \langle \hat{n}_\ell \rangle_\rho, \quad (64)$$

where the symbol $\langle$ indicates that the integration is over the region where $n_{k\ell,0} = 1$ for a given ℓ, and the symbol $\rangle$ indicates the region where $n_{k\ell,0} = 0$. For a given ℓ, the quantities $f_{\ell,0}$, $f_{\ell,x}$ and $f_{\ell,z}$ are nontrivial analytical functions of the following five variables

$$n_\ell = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}, \quad (65)$$

$$\xi_\ell = \frac{1}{2}\langle \hat{\sigma}_\ell^x \rangle_{\boldsymbol{\rho}}, \quad (66)$$

$$\Delta_\ell = \int_> dk n_{k\ell} = \int_< dk(1 - n_{k\ell}), \quad (67)$$

$$\mathcal{A}_{<\ell} = \int_< dk\sqrt{n_{k\ell}(1 - n_{k\ell})}, \quad (68)$$

$$\mathcal{A}_{>\ell} = \int_> dk\sqrt{n_{k\ell}(1 - n_{k\ell})}, \quad (69)$$

where $\boldsymbol{\rho}$ and $n_{k\ell}$ are constrained such that

$$|\xi_\ell| \leqslant \sqrt{(1-n_\ell)n_\ell - (1-\Delta_\ell)\Delta_\ell}, \tag{70}$$

which ensures that $f_{\ell,0}$, $f_{\ell,x}$, and $f_{\ell,z}$ are real numbers [see Eq. (73)]. The explicit functional dependence is given by the following list of equations

$$\xi_{\ell,0} = \sqrt{(1-n_\ell)n_\ell}, \tag{71}$$

$$\delta n_\ell = n_\ell - \frac{1}{2}, \tag{72}$$

$$\theta_\ell = \cos^{-1}\left(\frac{\xi_\ell}{\sqrt{\xi_{\ell,0}^2 + (\Delta_\ell - 1)\Delta_\ell}}\right), \tag{73}$$

$$g_{\ell,12} = \Delta_\ell \cot(\theta_\ell), \tag{74}$$

$$c_\ell = \sqrt{1 - \frac{2\delta n_\ell}{g_{\ell,12}^2 + n_\ell^2}}, \tag{75}$$

$$\phi_\ell = \tan^{-1}\left(\frac{g_{\ell,12}}{g_{\ell,12}^2 - \xi_{\ell,0}^2}\right), \tag{76}$$

$$\phi_{\ell,0} = \phi_\ell - \theta_\ell, \tag{77}$$

$$\mathcal{G}_{\ell,12} = -\frac{1}{2}\tan\left(\frac{\phi_{\ell,0}}{2}\right), \tag{78}$$

$$\mathcal{I}_\ell = g_{\ell,12}\mathcal{G}_{\ell,12} + \frac{n_\ell}{2}, \tag{79}$$

$$\mathcal{J}_\ell = \frac{g_{\ell,12}}{2} - n_\ell\mathcal{G}_{\ell,12}, \tag{80}$$

$$\mathcal{A}'_{<,\ell} = \mathcal{A}_{<,\ell} + \sin(\theta_\ell)\mathcal{A}_{>,\ell}, \tag{81}$$

$$\mathcal{A}'_{>,\ell} = \cos(\theta_\ell)\mathcal{A}_{>,\ell}, \tag{82}$$

$$\mathcal{G}_{\ell,13} = -\mathcal{G}_{\ell,32} = \frac{\mathcal{I}_\ell\mathcal{A}'_{>,\ell} - \mathcal{J}_\ell\mathcal{A}'_{<,\ell}}{(g_{\ell,12}^2 + n_\ell^2)\sqrt{c_\ell}\sin(\theta_\ell)}, \tag{83}$$

$$\mathcal{G}_{\ell,23} = -\mathcal{G}_{\ell,31} = \frac{\mathcal{I}_\ell\mathcal{A}'_{<,\ell} + \mathcal{J}_\ell\mathcal{A}'_{>,\ell}}{(g_{\ell,12}^2 + n_\ell^2)\sqrt{c_\ell}\sin(\theta_\ell)}, \tag{84}$$

$$i_\ell = \frac{8\mathcal{G}_{\ell,13}\mathcal{G}_{\ell,23}}{4\mathcal{G}_{\ell,12}^2 + 1}, \tag{85}$$

$$j_\ell = \frac{4(\mathcal{G}_{\ell,13}^2 - \mathcal{G}_{\ell,23}^2)}{4\mathcal{G}_{\ell,12}^2 + 1}, \tag{86}$$

$$\mathcal{G}_{\ell,33} = n_\ell + j_\ell(g_{\ell,12} - \mathcal{G}_{\ell,12}) - i_\ell\delta n_\ell, \tag{87}$$

$$p_\ell = \frac{\xi_\ell(\cot(\phi_{\ell,0}) - 2g_{\ell,12}) - 2\delta n_\ell^2}{2(\xi_\ell^2 + \delta n_\ell^2)}, \tag{88}$$

$$q_\ell = \frac{\delta n_\ell(\cot(\phi_{\ell,0}) - 2g_{\ell,12} + 2\xi_\ell)}{2(\xi_\ell^2 + \delta n_\ell^2)}, \tag{89}$$

$$f_{\ell,0} = \mathcal{G}_{\ell,33} - \frac{1}{2}j_\ell\csc(\phi_{\ell,0}), \tag{90}$$

$$f_{\ell,z} = \frac{1}{2}(i_\ell p_\ell - j_\ell q_\ell), \tag{91}$$

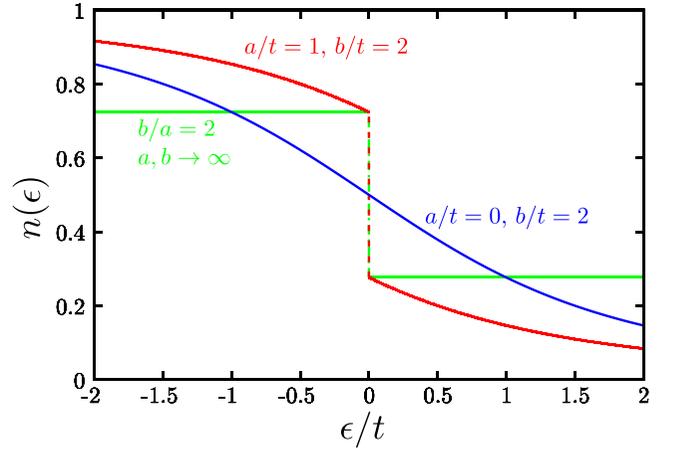$$f_{\ell,x} = \frac{1}{2}(i_\ell q_\ell + j_\ell p_\ell). \tag{92}$$

FIG. 1. Plot of the partially optimized momentum density distribution $n(\epsilon, a, b)$ for various values of a and b.

Several important points should be noted. First, the $\{n_{k\ell}\}$ only enter the local interaction energy through Δ_ℓ, $\mathcal{A}_{<\ell}$, $\mathcal{A}_{>\ell}$, and the constraint $\int dk\, n_{k\ell} = n_\ell$. Second, the $\{n_{k\ell,0}\}$ only enter the local interaction energy through the regions of integration (i.e., $>$ and $<$) and the constraint $\int dk\, n_{k\ell,0} = n_\ell$. Third, the operators $\hat{n}_{\mathrm{eff},\ell}$ and $\hat{n}_\ell$ have the same expectation values under $\boldsymbol{\rho}$, as in the case of the $\hat{\rho}_{B2}$. Finally, it should be clear that evaluating the $\hat{\rho}_{G3}$ has a similar computational cost as compared to $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$, where the largest computational cost is associated with evaluating the local interaction energy.

In the following, we briefly discuss how to numerically minimize the qubit energy form for $\hat{\rho}_{G3}$. For a multiorbital Hubbard model with $2N_{\mathrm{orb}}$ spin-orbitals, the formal variational parameters are $\{n_{k\ell}\}$, $\{n_{k\ell;0}\}$, and $\boldsymbol{\rho}$ with $\ell = 1, \ldots, 2N_{\mathrm{orb}}$, and there are two local density constraints per spin orbital [see Eq. (64)]. We choose $\{n_{k\ell;0}\}$ corresponding to the momentum density distribution for the noninteracting Hamiltonian with local density $\{n_\ell\}$, and therefore $\{n_{k\ell;0}\}$ is determined from $\{n_\ell\}$. It is important to realize that the local interaction energy $\langle H_{\mathrm{loc}}(\{\hat{n}_{\mathrm{eff},\ell}\})\rangle_{\boldsymbol{\rho}}$ does not depend on the full details of $\{n_{k\ell}\}$, but instead only on $\{n_\ell\}$, $\{\Delta_\ell\}$, $\{\mathcal{A}_{<\ell}\}$, and $\{\mathcal{A}_{>\ell}\}$. Therefore, for a given spin orbital ℓ, four Lagrange multipliers, $a_{<\ell}$, $b_{<\ell}$, $a_{>\ell}$, and $b_{>\ell}$, can be used to obtain the partially optimized $n_{k\ell}$ as [17]

$$n_{k\ell}\big|_{k\in X} = \frac{1}{2}\left(\frac{a_{X\ell} - \epsilon_{k\ell}}{\sqrt{(a_{X\ell} - \epsilon_{k\ell})^2 + b_{X\ell}^2}} + 1\right), \tag{93}$$

where X is either $\langle$ or $\rangle$. Interestingly, the optimized $n_{k\ell}$ yields a nontrivial dependence on $\epsilon_{k\ell}$, in contrast to the Gutzwiller approximation (see Fig. 1). There is a local density constraint between $\{n_{k\ell}\}$ and $\boldsymbol{\rho}$ given by Eq. (64), and there are two strategies to enforce this. The first strategy is to start from $\{n_{k\ell}\}$, which are parametrized by $5 \times 2N_{\mathrm{orb}}$ variational parameters $\{n_\ell\}$, $\{a_{<\ell}\}$, $\{a_{>\ell}\}$, $\{b_{<\ell}\}$, and $\{b_{>\ell}\}$, though only $4 \times 2N_{\mathrm{orb}}$ are independent given Eq. (64). Subsequently, the $\boldsymbol{\rho}$ can be parametrized by $2^{2N_{\mathrm{orb}}} - 2N_{\mathrm{orb}} - 1$ independent variational parameters [17]. Therefore, there are a total of $2^{2N_{\mathrm{orb}}} + 3 \times 2N_{\mathrm{orb}} - 1$ independent variational parameters. This strategy motivates the construction of a one-body reduced density

matrix functional, as presented in our companion paper [18]. The second strategy is to start from $\boldsymbol{\rho}$, which can be parametrized by $2^{2N_{orb}} - 1$ independent variational parameters, which determines $\{n_\ell\}$. Subsequently, the $\{\Delta_\ell\}$, $\{b_{<\ell}\}$, and $\{b_{>\ell}\}$ are chosen as independent variational parameters, which determines $\{a_{X\ell}\}$, yielding $2^{2N_{orb}} + 3 \times 2N_{orb} - 1$ independent variational parameters. The second strategy allows for a trivial implementation of the inequality constraint among Δ_ℓ, n_ℓ, and ξ_ℓ [see Eq. (303)], for a given ℓ, which is very important in practice. In either case, one is left with minimizing over $2^{2N_{orb}} + 3 \times 2N_{orb} - 1$ variational parameters, which may be achieved using a variety of standard approaches.

Finally, we summarize the applications that are studied in Sec. VII B. Consider the multiorbital Hubbard model with density-density interactions given as

$$\hat{H}_{loc} = U\hat{O}_1 + (U - 2J)\hat{O}_2 + (U - 3J)\hat{O}_3, \qquad (94)$$

where $\hat{O}_1 = \sum_\alpha \delta\hat{n}_{\alpha\uparrow}\delta\hat{n}_{\alpha\downarrow}$, $\hat{O}_2 = \sum_{\alpha<\beta,\sigma} \delta\hat{n}_{\alpha\sigma}\delta\hat{n}_{\beta\bar\sigma}$, $\hat{O}_3 = \sum_{\alpha<\beta,\sigma} \delta\hat{n}_{\alpha\sigma}\delta\hat{n}_{\beta\sigma}$, and $\delta\hat{n}_{\alpha\sigma} = \hat{n}_{\alpha\sigma} - \frac{1}{2}$, with the orbital indices α, β taking values of $1, \ldots, N_{orb}$ and $\sigma \in \{\uparrow, \downarrow\}$. We consider the special case where $\epsilon_{k\ell}$ is independent of ℓ and has particle-hole symmetry. In this case, we show that Eq. (62) can be optimized over $\boldsymbol{\rho}$ and $\{n_{k\ell;0}\}$, with the assumption that the optimized $n_{k\ell;0}$ is given as $\theta(n_{k\ell} - n_\ell^\star)$ where θ is the Heaviside function and $n_\ell^\star$ is chosen such that $\int dk n_{k\ell;0} = n_\ell$, yielding a trial energy purely as a functional of $\{n_{k\ell}\}$ as

$$E(\{n_{k\ell}\}) = \sum_\ell \int dk \epsilon_{k\ell} n_{k\ell} + UA^4 \tilde{O}(\Delta), \qquad (95)$$

where $\tilde{O}(\Delta)$ depends on N_{orb} and J/U, and $n_{k\ell}$ is independent of ℓ and has particle-hole symmetry. Here we have dropped the spin-orbital index ℓ (e.g. $\Delta_\ell \to \Delta$) and define $A = \mathcal{A}_< + \mathcal{A}_>$. An important result is that $\tilde{O}(\Delta)$ is nonanalytic at $\Delta = \Delta_c$, which can be used to infer the nature of the Mott transition. For $N_{orb} > 1$, the $\tilde{O}(\Delta)$ can be numerically represented by a one-dimensional spline function [18], while $N_{orb} = 1$ has an explicit expression which can be derived as

$$\tilde{O}(\Delta) = \begin{cases} -\frac{1-4d_1(\Delta)}{4(2-4\Delta)^2\Delta^2}, & \Delta < \Delta_c \\ -4, & \Delta \geq \Delta_c \end{cases}, \qquad (96)$$

$$d_1(\Delta) = \frac{1}{4} - \frac{2\sqrt{\frac{2}{3}}\Delta^2(-4\Delta + h_2 + 2)^2}{3(4(\Delta-1)\Delta + (2\Delta-1)h_1 + 3)^{3/2}}, \qquad (97)$$

$$h_1 = \sqrt{4(\Delta-1)\Delta + 9}, \qquad (98)$$

$$h_2 = \sqrt{4\Delta(10\Delta + 3h_1 - 10) - 6h_1 + 18}, \qquad (99)$$

$$\Delta_c = \frac{1}{6}(3 - \sqrt{3}). \qquad (100)$$

The above equations finally deliver a closed form expression for a trial energy which qualitatively and quantitatively captures the Mott transition in the single-band Hubbard model at half-filling in $d = \infty$, exactly reproducing the VDAT results for a G-type SPD at $\mathcal{N} = 3$.

IV. DERIVATION OF THE QUBIT ENERGY FORM FOR $\hat{\rho}_{G2}$

In this section, we derive Eq. (50) using several different approaches. In Sec. IV A, we review the GA using both

a heuristic derivation and the CPE, and convert the standard energy form of the GA to the qubit energy form using the Jordan-Wigner transformation. In Sec. IV B, we use the SCDA within the Gutzwiller gauge to evaluate $\hat{\rho}_{G2}$, and the tensor product representation is used to obtain the qubit energy form.

A. Derivation of the qubit energy form from the GA

In this section, we provide an elementary derivation of the qubit energy form, which consists of two steps. In the first step, we derive the standard form of the GA, using both a heuristic argument and the CPE [56], where the energy is parametrized by $\{n_{k\ell;0}\}$ and ρ_{loc}, where ρ_{loc} is the local reduced density matrix of the ansatz, which is a diagonal matrix in the basis of $|\Gamma\rangle$. The quasiparticle weight is constructed from ρ_{loc} and the fermionic creation and annihilation operators. In the second step, we convert the standard form of the GA to the qubit energy form which is parametrized by $\{n_{k\ell;0}\}$ and $\boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is a pure state in a $2N_{orb}$-qubit space, and the quasiparticle weight is constructed from $\boldsymbol{\rho}$ and the Pauli spin operators.

1. Derivation of the GA: A heuristic argument and the CPE

While the Gutzwiller approximation is well known, it is normally applied to the special case where $\hat{\rho}_{G2}$ is a pure state. Here we provide a heuristic derivation for a general $\hat{\rho}_{G2}$. Additionally, we use the CPE to derive the same result from a different perspective. We begin by presenting the heuristic derivation of the GA. First consider the expectation values for the diagonal Hubbard operator $\hat{X}_{i\Gamma}$ measured under $\hat{K}_2 = \hat{K}(\{n_{k\ell;0}\})$, given as

$$\langle \hat{X}_{i\Gamma}\rangle_{\hat{K}_2} = \prod_{\ell=1}^N (\delta_{\Gamma_\ell,0}(1-n_\ell) + \delta_{\Gamma_\ell,1}n_\ell), \qquad (101)$$

where $n_\ell = \langle\hat{n}_{i\ell}\rangle_{\hat{K}_2}$ denotes the local density at a given site i for the spin orbital ℓ under $\hat{K}_2$. The first assumption of the GA is that the atomic configuration distribution under $\hat{\rho}_{G2}$ can be approximated as

$$\langle \hat{X}_{i\Gamma}\rangle_{\hat{\rho}_{G2}} = u_\Gamma^2 \langle\hat{X}_{i\Gamma}\rangle_{\hat{K}_2} \Big/ \sum_\Gamma u_\Gamma^2 \langle\hat{X}_{i\Gamma}\rangle_{\hat{K}_2}, \qquad (102)$$

which ignores off-site contributions in $\hat{K}_2$. Additionally, a constraint is applied to $\hat{P}_1$ such that the local density is invariant, given as

$$\langle\hat{n}_{i\ell}\rangle_{\hat{\rho}_{G2}} = \langle\hat{n}_{i\ell}\rangle_{\hat{K}_2}, \qquad (103)$$

which we refer to as the Gutzwiller constraint. The denominator in Eq. (102) ensures that the sum of all expectation values for the diagonal Hubbard operators are normalized. Given that the off-diagonal Hubbard operators have zero expectation value due to the restriction of the SPD used in this study, the expectation values for all diagonal Hubbard operators allow the evaluation of any local observable. The second assumption of the GA lies within the evaluation for the hopping term between two distinct sites. The idea is to break a hopping term like $\hat{a}_{i\ell}^\dagger \hat{a}_{j\ell}$ into two processes: first annihilating an electron at site j and then creating an electron at site i, and the GA

assumes the probabilities for the two steps are independent and only depend on the atomic distributions on site i and j. The probability of creating or destroying an electron is renormalized by the interacting projector, and the GA assumes this is obtained by counting all relevant one-particle excitation processes as

$$\mathcal{R}_\ell = \frac{\sum_{\Gamma\Gamma'} \sqrt{\langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}} \langle \hat{X}_{i\Gamma'} \rangle_{\hat{\rho}_{G2}}} |\langle \Gamma' | \hat{a}_{i\ell}^\dagger | \Gamma \rangle|^2}{\sum_{\Gamma\Gamma'} \sqrt{\langle \hat{X}_{i\Gamma} \rangle_{\hat{K}_2} \langle \hat{X}_{i\Gamma'} \rangle_{\hat{K}_2}} |\langle \Gamma' | \hat{a}_{i\ell}^\dagger | \Gamma \rangle|^2}. \quad (104)$$

To simplify this expression, we introduce ρ_{loc} and $\rho_{\text{loc};0}$, which are defined within the local Hilbert space as

$$[\rho_{\text{loc}}]_{\Gamma\Gamma'} = \delta_{\Gamma\Gamma'} \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}}, \quad (105)$$

$$[\rho_{\text{loc};0}]_{\Gamma\Gamma'} = \delta_{\Gamma\Gamma'} \langle \hat{X}_{i\Gamma} \rangle_{\hat{K}_2}, \quad (106)$$

which are matrices of dimension $2^{2N_{\text{orb}}} \times 2^{2N_{\text{orb}}}$, yielding

$$\mathcal{R}_\ell = \frac{\text{Tr}(\sqrt{\rho_{\text{loc}}} \hat{a}_\ell^\dagger \sqrt{\rho_{\text{loc}}} \hat{a}_\ell)}{\text{Tr}(\sqrt{\rho_{\text{loc};0}} \hat{a}_\ell^\dagger \sqrt{\rho_{\text{loc};0}} \hat{a}_\ell)}. \quad (107)$$

The Gutzwiller constraint, Eq. (103), can be rewritten as $\langle \hat{n}_\ell \rangle_{\rho_{\text{loc}}} = \langle \hat{n}_\ell \rangle_{\rho_{\text{loc};0}}$. Given the second assumption of the GA, the single particle density matrix between two different sites j and j' is renormalized as

$$\mathcal{Z}_\ell \equiv \frac{\langle \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{\rho}_{G2}}}{\langle \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{K}_2}} = \mathcal{R}_\ell^2. \quad (108)$$

Combining Eq. (108) with the Gutzwiller constraint, the single particle density matrix for arbitrary j and j' is given as

$$\langle \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{\rho}_{G2}} = \mathcal{Z}_\ell \langle \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{K}_2} + \delta_{jj'}(1 - \mathcal{Z}_\ell) n_\ell. \quad (109)$$

The momentum density distribution $n_{k\ell} = \langle \hat{a}_{k\ell}^\dagger \hat{a}_{k\ell} \rangle_{\hat{\rho}_{G2}}$ can then be computed as

$$n_{k\ell} = \frac{1}{N_{\text{site}}} \sum_{jj'} e^{ik(j-j')} \langle \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{\rho}_{G2}} \quad (110)$$

$$= \mathcal{Z}_\ell n_{k\ell;0} + (1 - \mathcal{Z}_\ell) n_\ell, \quad (111)$$

where N_{site} is the number of k points, $\hat{a}_k^\dagger = (1/\sqrt{N_{\text{site}}}) \sum_j e^{ik \cdot j} \hat{a}_j^\dagger$, and $n_{k\ell;0} = \langle a_{k\ell}^\dagger a_{k\ell} \rangle_{\hat{K}_2}$. Therefore, the standard form for the total energy per site of the GA is given by

$$E(\{n_{k\ell,0}\}, \rho_{\text{loc}}) = \sum_\ell \int dk \epsilon_{k\ell} n_{k\ell} + \langle H_{\text{loc}}(\{\hat{n}_\ell\}) \rangle_{\rho_{\text{loc}}}, \quad (112)$$

$$n_{k\ell} = \langle \hat{n}_\ell \rangle_{\rho_{\text{loc}}} + \mathcal{R}_\ell^2 (n_{k\ell,0} - \langle \hat{n}_\ell \rangle_{\rho_{\text{loc}}}), \quad (113)$$

$$\mathcal{R}_\ell = \frac{\text{Tr}(\sqrt{\rho_{\text{loc}}} \hat{a}_\ell^\dagger \sqrt{\rho_{\text{loc}}} \hat{a}_\ell)}{\sqrt{\langle \hat{n}_\ell \rangle_{\rho_{\text{loc}}} \langle 1 - \hat{n}_\ell \rangle_{\rho_{\text{loc}}}}}, \quad (114)$$

with the density constraint

$$\int dk n_{k\ell;0} = \langle \hat{n}_\ell \rangle_{\rho_{\text{loc}}}. \quad (115)$$

The variational parameters satisfy $n_{kl;0} \in [0, 1]$ and ρ_{loc} is a diagonal positive semidefinite matrix.

We now provide an overview of how to derive Eq. (112) using the CPE (see Appendix C 1). We begin by making

several observations about the GA. First, when $n_{k\ell;0} = n_\ell$, the $\hat{\rho}_{G2}$ describes a collection of atoms, and in this case Eq. (112) yields an exact evaluation in any dimension. Second, there is a linear relation between $n_{k\ell}$ and $n_{k\ell;0}$ given by

$$n_{k\ell} - n_\ell = \mathcal{Z}_\ell (n_{k\ell;0} - n_\ell). \quad (116)$$

Third, ρ_{loc} is directly determined by n_ℓ and $\{u_\Gamma\}$, and is independent of $n_{k\ell;0}$. Recall that the $\hat{\rho}_{G2}$ is defined as $\hat{\rho}_{G2} = \hat{P}_1 \hat{K}_2 \hat{P}_1$, where $\hat{K}_2 = \hat{K}(\{n_{k\ell;0}\})$ and $\hat{P}_1 = \hat{P}(\{u_\Gamma\})$. We now introduce $\hat{K}_2^\star = \hat{K}(\{n_{k\ell;0} = n_\ell\})$, which has the same local density matrix as $\hat{K}_2$, and $\rho_{G2}^\star = \hat{P}_1 \hat{K}_2^\star \hat{P}_1$, which describes a collection of atoms. The $\{u_\Gamma\}$ can be reparametrized by the local reduced density matrix $\rho_{\text{loc}}^\star$ of $\rho_{G2}^\star$, where $[\rho_{\text{loc}}^\star]_{\Gamma\Gamma'} = \delta_{\Gamma\Gamma'} \langle \hat{X}_{i\Gamma} \rangle_{\rho_{G2}^\star}$ and is constrained by $n_\ell = \langle \hat{a}_\ell^\dagger \hat{a}_\ell \rangle_{\rho_{\text{loc}}^\star}$. Therefore, $n_{k\ell}$ and ρ_{loc} are functionals of $\{n_{k\ell;0}\}$ and $\rho_{\text{loc}}^\star$. The CPE amounts to the expansion of observables in terms of $\{n_{k\ell;0}\}$ about $\{n_{k\ell;0} = n_\ell\}$, and up to the first-order, one recovers Eq. (116) and $\rho_{\text{loc}} = \rho_{\text{loc}}^\star$, proving that the first-order CPE recovers the GA.

2. Converting the GA energy into the qubit form using the Jordan-Wigner transformation

Here we discuss how to convert the standard form of the GA energy into the qubit form, which is mathematically equivalent. The qubit form provides a unified view of the energy evaluated using $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, and $\hat{\rho}_{G3}$. The qubit parametrization consists of two steps. First, we introduce a purified many-body density matrix $\boldsymbol{\rho}$ from ρ_{loc}. Second, we perform the Jordan-Wigner transformation, which converts $\mathcal{R}_\ell$ [Eq. (114)] from an expression involving fermionic operators $\hat{a}_\ell^\dagger$ and $\hat{a}_\ell$ into an expression involving spin operators $\hat{\sigma}_\ell^x$. We begin by defining $\boldsymbol{\rho} = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle = \sum_\Gamma \sqrt{[\rho_{\text{loc}}]_{\Gamma\Gamma}} |\Gamma\rangle$, which yields

$$[\boldsymbol{\rho}]_{\Gamma\Gamma'} = \sqrt{[\rho_{\text{loc}}]_{\Gamma\Gamma} [\rho_{\text{loc}}]_{\Gamma'\Gamma'}}. \quad (117)$$

Using the Jordan-Wigner transformation [see Eqs. (43) and (44)], we obtain

$$|\langle \Gamma' | \hat{a}_\ell^\dagger | \Gamma \rangle|^2 = \langle \Gamma' | \hat{a}_\ell^\dagger | \Gamma \rangle \langle \Gamma | \hat{a}_\ell | \Gamma' \rangle \quad (118)$$

$$= \langle \Gamma' | \left(\prod_{\ell' < \ell} \hat{\sigma}_{\ell'}^z \right) \hat{\sigma}_\ell^+ | \Gamma \rangle \langle \Gamma | \left(\prod_{\ell' < \ell} \hat{\sigma}_{\ell'}^z \right) \hat{\sigma}_\ell^- | \Gamma' \rangle \quad (119)$$

$$= \langle \Gamma' | \hat{\sigma}_\ell^+ | \Gamma \rangle \langle \Gamma | \hat{\sigma}_\ell^- | \Gamma' \rangle = |\langle \Gamma' | \hat{\sigma}_\ell^+ | \Gamma \rangle|^2. \quad (120)$$

Therefore, we can rewrite the numerator of $\mathcal{R}_\ell$ [see Eq. (107)] as

$$\sum_{\Gamma\Gamma'} \sqrt{[\rho_{\text{loc}}]_{\Gamma\Gamma} [\rho_{\text{loc}}]_{\Gamma'\Gamma'}} |\langle \Gamma' | \hat{a}_\ell^\dagger | \Gamma \rangle|^2 \quad (121)$$

$$= \sum_{\Gamma\Gamma'} [\boldsymbol{\rho}]_{\Gamma\Gamma'} |\langle \Gamma' | \sigma_\ell^+ | \Gamma \rangle|^2 \quad (122)$$

$$= \frac{1}{2} \sum_{\Gamma\Gamma'} [\boldsymbol{\rho}]_{\Gamma\Gamma'} \langle \Gamma' | \hat{\sigma}_\ell^x | \Gamma \rangle = \frac{1}{2} \langle \hat{\sigma}_\ell^x \rangle_{\boldsymbol{\rho}} = \xi_\ell, \quad (123)$$

where we used the fact that $\boldsymbol{\rho}$ is a real symmetric matrix and the following relation

$$\langle \Gamma' | \hat{\sigma}_\ell^x | \Gamma \rangle = |\langle \Gamma' | \hat{\sigma}_\ell^+ | \Gamma \rangle|^2 + |\langle \Gamma | \hat{\sigma}_\ell^+ | \Gamma' \rangle|^2. \quad (124)$$

Moreover, the denominator of $\mathcal{R}_\ell$ [see Eq. (107)] is given as

$$\text{Tr}\left(\sqrt{\rho_{\text{loc};0}}\,\hat{a}_\ell^\dagger \sqrt{\rho_{\text{loc};0}}\,\hat{a}_\ell\right) = \sqrt{n_\ell(1-n_\ell)} = \xi_{\ell;0}. \quad (125)$$

Therefore, we have demonstrated that $\mathcal{R}_\ell = \xi_\ell/\xi_{\ell;0}$, and the local interaction can be written as

$$\langle H_{\text{loc}}(\{\hat{n}_\ell\})\rangle_{\rho_{\text{loc}}} = \langle H_{\text{loc}}(\{\hat{n}_\ell\})\rangle_\rho. \quad (126)$$

Using the expression for $\mathcal{R}_\ell$ and the local energy, we arrive at the energy expressions given in Eqs. (50)–(53). This proof demonstrates that the usual Gutzwiller approximation can be straightforwardly transformed into the qubit energy form, which is equivalent to the result of the slave spin mean-field theory [49] (see Appendix B for additional details).

B. Derivation of the qubit energy form using the SCDA

In this section, we use the gauge constrained SCDA to derive the qubit energy form. The derivation consists of two steps. First, in Sec. IV B 1, the Gutzwiller gauge is used to rederive the standard form of the GA energy. Second, in Sec. IV B 2, we derive the qubit energy form using the tensor representation, which does not rely on the Jordan-Wigner transformation. Additionally, in Sec. IV B 3, we discuss how the quantities $\mathcal{G}$, S, and g change under a general gauge transformation.

1. SCDA within the Gutzwiller gauge

In this section, we demonstrate how the gauge of the SPD can be used to automatically satisfy the SCDA self-consistency condition [15], and we utilize some notation from the CPE. Interestingly, the Gutzwiller constraint of the GA can be used to define an appropriate gauge for the SPD, which we refer to as the Gutzwiller gauge. Starting from $\hat{\rho}_{G2} = \hat{P}_1 \hat{K}_2 \hat{P}_1$ with an arbitrary gauge, a gauge transformation $\hat{P}_1 \to \hat{P}_1 \hat{N}^{-1}$ and $\hat{K}_2 \to \hat{N} \hat{K}_2 \hat{N}$ can always be performed to ensure that $\langle \hat{n}_{i\ell}\rangle_{\hat{\rho}_{G2}} = \langle \hat{n}_{i\ell}\rangle_{\hat{K}_2} = n_\ell$, where $\hat{N} = \exp(\sum_{i\ell} \mu_\ell \hat{n}_{i\ell})$ is chosen to satisfy the Gutzwiller constraint. Within this Gutzwiller gauge, we can choose $[\mathcal{G}_i]_{\ell\ell'} = \delta_{\ell\ell'}\mathcal{G}_\ell$, where the component for the spin-orbital ℓ is given by

$$\mathcal{G}_\ell = \begin{pmatrix} n_\ell & 1-n_\ell \\ -n_\ell & n_\ell \end{pmatrix}. \quad (127)$$

We now prove that this $\mathcal{G}_\ell$ ensures that the SCDA self-consistency condition is automatically satisfied. The first step of the proof relies on the fact that under the Gutzwiller gauge $\hat{\rho}_{\text{loc}}$ within the SCDA is the discrete action of the central point of the SPD $\rho_{G2}^\star = \hat{P}_1 \hat{K}_2^\star \hat{P}_1$, which can be shown as follows. The local reduced density matrix of $\rho_{G2}^\star$ at site i is $\hat{\rho}_{G2;i}^\star = \hat{P}_{1;i} \hat{K}_{2;i}^\star \hat{P}_{1;i}$, which yields a discrete action $\underline{\hat{\rho}}_{G2;i}^\star = \underline{\hat{\rho}}_{G2;i;0}^\star \underline{\hat{P}}_{1;i}^{(1)} \underline{\hat{P}}_{1;i}^{(2)}$, where $\underline{\hat{\rho}}_{G2;i;0}^\star = \underline{\hat{Q}} \underline{\hat{K}}_{2;i}^\star$. Given that $\langle \hat{n}_{i\ell}\rangle_{\hat{K}_{2;i}^\star} = n_\ell$, the integer time Green's function for $\underline{\hat{\rho}}_{G2;i;0}^\star$ is given by Eq. (127), and correspondingly $\underline{\hat{\rho}}_{G2;i}^\star$ is equivalent to the local discrete action $\hat{\rho}_{\text{loc};i}$ for the SCDA, proving $\hat{\rho}_{G2}^\star = \hat{\rho}_{\text{loc}}$. Local observables within the SCDA can then be evaluated under the central point of the SPD as

$$\langle \hat{X}_{i\Gamma}\rangle_{\hat{\rho}_{G2;i}} = \langle \underline{\hat{X}}_{i\Gamma}^{(2)}\rangle_{\underline{\hat{\rho}}_{\text{loc};i}} = \langle \hat{X}_{i\Gamma}\rangle_{\hat{\rho}_{G2;i}^\star}, \quad (128)$$

consistent with the relation derived from the CPE [see Eq. (C21)]. Similarly, the local interacting integer time Green's function $[g_i]_{\ell\ell'} = \delta_{\ell\ell'}g_\ell$ can be evaluated as

$$g_\ell = \begin{pmatrix} n_\ell & (1-n_\ell)\mathcal{R}_{\ell,12} \\ -n_\ell \mathcal{R}_{\ell,21} & n_\ell \end{pmatrix}, \quad (129)$$

where

$$\mathcal{R}_{\ell,12} \equiv \frac{\text{Tr}(\hat{P}_{1;i}\hat{a}_{i\ell}^\dagger \hat{K}_{2;i}^\star \hat{P}_{1;i}\hat{a}_{i\ell})/\text{Tr}(\hat{\rho}_{G2;i}^\star)}{\text{Tr}(\hat{a}_{i\ell}^\dagger \hat{K}_{2;i}^\star \hat{a}_{i\ell})/\text{Tr}(\hat{K}_{2;i}^\star)}, \quad (130)$$

and

$$\mathcal{R}_{\ell,21} \equiv \frac{\text{Tr}(\hat{P}_{1;i}\hat{a}_{i\ell}\hat{K}_{2;i}^\star \hat{P}_{1;i}\hat{a}_{i\ell}^\dagger)/\text{Tr}(\hat{\rho}_{G2;i}^\star)}{\text{Tr}(\hat{a}_{i\ell}\hat{K}_{2;i}^\star \hat{a}_{i\ell}^\dagger)/\text{Tr}(\hat{K}_{2;i}^\star)}, \quad (131)$$

and the Gutzwiller gauge ensures that the diagonal part of g_ℓ is n_ℓ. To connect $\mathcal{R}_{\ell,12}$ and $\mathcal{R}_{\ell,21}$ with $\mathcal{R}_\ell$, we use the following relations:

$$\hat{a}_{i\ell}^\dagger \sqrt{\hat{K}_{2;i}^\star} = \sqrt{(1-n_\ell)/n_\ell}\sqrt{\hat{K}_{2;i}^\star}\,\hat{a}_{i\ell}^\dagger, \quad (132)$$

$$\hat{a}_{i\ell}\sqrt{\hat{K}_{2;i}^\star} = \sqrt{n_\ell/(1-n_\ell)}\sqrt{\hat{K}_{2;i}^\star}\,\hat{a}_{i\ell}, \quad (133)$$

which yields $\mathcal{R}_{\ell,12} = \mathcal{R}_{\ell,21} = \mathcal{R}_\ell$. The integer time self-energy can then be computed as $[S_i]_{\ell\ell'} = \delta_{\ell\ell'}S_\ell$, using $S_\ell = (\mathcal{G}_\ell^{-1}-1)^{-1}(g_\ell^{-1}-1)$, which yields

$$S_\ell = \frac{1}{n_\ell - (n_\ell-1)\mathcal{R}_\ell^2}$$
$$\times \begin{pmatrix} \mathcal{R}_\ell & (n_\ell-1)(\mathcal{R}_\ell^2-1) \\ n_\ell(\mathcal{R}_\ell^2-1) & \mathcal{R}_\ell \end{pmatrix}. \quad (134)$$

Therefore, the interacting integer time Green's function for a given k point is given by $[g_k]_{\ell\ell'} = \delta_{\ell\ell'}g_{k\ell}$, where $g_{k\ell} = (1+(g_{k\ell;0}^{-1}-1)S_\ell)^{-1}$, which yields

$$g_{k\ell} = \begin{pmatrix} n_{k\ell,0} & \mathcal{R}_\ell(1-n_{k\ell,0}) \\ -\mathcal{R}_\ell n_{k\ell,0} & \mathcal{R}_\ell^2(n_{k\ell,0}-n_\ell)+n_\ell \end{pmatrix}, \quad (135)$$

and the momentum density distribution $n_{k\ell} = [g_{k\ell}]_{22}$ is in agreement with Eq. (111). Finally, the SCDA self-consistency can be verified as $\frac{1}{N_{\text{site}}}\sum_k g_{k\ell} = g_\ell$.

2. Derivation of the qubit energy form

Here we show how to derive the qubit energy form, given in Eqs. (50)–(54), using the tensor product representation. We begin by evaluating the relevant observables under $\hat{\rho}_{\text{loc}}$, where the tensor product representation in Eq. (33) simplifies to

$$(\hat{\underline{O}})_u \cdot (u_1 \otimes u_2) = u^T (\hat{\underline{O}})_u u, \quad (136)$$

given that $u^T = u_1 = u_2$. The relevant components needed to evaluate the local integer time Green's function in the Gutzwiller gauge are given by

$$(\hat{\underline{1}})_{u;\ell} = \begin{pmatrix} 1-n_\ell & 0 \\ 0 & n_\ell \end{pmatrix}, \quad (137)$$

$$(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)})_{u;\ell} = \begin{pmatrix} 0 & 0 \\ 0 & n_\ell \end{pmatrix}, \quad (138)$$

$$(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)})_{u;\ell} = \begin{pmatrix} 0 & 0 \\ 1-n_\ell & 0 \end{pmatrix}, \quad (139)$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(2)}\hat{\underline{a}}_\ell^{(1)}\right)_{u;\ell} = \begin{pmatrix} 0 & -n_\ell \\ 0 & 0 \end{pmatrix}, \tag{140}$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(2)}\hat{\underline{a}}_\ell^{(2)}\right)_{u;\ell} = \begin{pmatrix} 0 & 0 \\ 0 & n_\ell \end{pmatrix}. \tag{141}$$

A linear transformation $u = Vw$ can be introduced such that

$$w^T(\hat{\underline{O}})_w w = u^T(\hat{\underline{O}})_u u, \tag{142}$$

where

$$(\hat{\underline{O}})_w = (\hat{\underline{O}}_1)_{w;1} \otimes \cdots \otimes (\hat{\underline{O}}_M)_{w;2N_{\rm orb}}, \tag{143}$$

$$(\hat{\underline{O}}_\ell)_{w;\ell} \equiv V_\ell^T (\hat{\underline{O}}_\ell)_{u;\ell} V_\ell, \tag{144}$$

$$V = V_1 \otimes \cdots \otimes V_{2N_{\rm orb}}. \tag{145}$$

We choose V_ℓ such that the identity operator in the w representation is an identity matrix, which is given by

$$V_\ell = \begin{pmatrix} \frac{1}{\sqrt{1-n_\ell}} & 0 \\ 0 & \frac{1}{\sqrt{n_\ell}} \end{pmatrix}, \tag{146}$$

and correspondingly, the components in the w representation are given by

$$(\hat{\underline{1}})_{w;\ell} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{147}$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{w;\ell} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{148}$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{w;\ell} = \begin{pmatrix} 0 & 0 \\ \frac{\sqrt{1-n_\ell}}{\sqrt{n_\ell}} & 0 \end{pmatrix}, \tag{149}$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(2)}\hat{\underline{a}}_\ell^{(1)}\right)_{w;\ell} = \begin{pmatrix} 0 & -\frac{\sqrt{n_\ell}}{\sqrt{1-n_\ell}} \\ 0 & 0 \end{pmatrix}, \tag{150}$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(2)}\hat{\underline{a}}_\ell^{(2)}\right)_{w;\ell} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{151}$$

To connect with the qubit energy form, we define the many-body density matrix corresponding to a pure state of the qubit system as

$$\boldsymbol{\rho} = ww^T, \tag{152}$$

where the renormalization factor $\mathcal{R}_\ell$ can now be rewritten as

$$\mathcal{R}_\ell = \frac{1}{2}\left\langle \frac{\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_w}{1-n_\ell} - \frac{\left(\hat{\underline{a}}_\ell^{\dagger(2)}\hat{\underline{a}}_\ell^{(1)}\right)_w}{n_\ell} \right\rangle_{\boldsymbol{\rho}} \tag{153}$$

$$= \frac{\frac{1}{2}\langle\hat{\sigma}_\ell^x\rangle_{\boldsymbol{\rho}}}{\sqrt{(1-n_\ell)n_\ell}} = \frac{\xi_\ell}{\xi_{\ell;0}}, \tag{154}$$

using Eqs. (149), (150), and (129). It should be emphasized that $\hat{\sigma}_\ell^x$ is a $2^{2N_{\rm orb}} \times 2^{2N_{\rm orb}}$ matrix, defined in Eq. (46). Additionally, the local expectation value of the interaction energy can be rewritten as

$$\left\langle H_{\rm loc}\left(\left\{\hat{\underline{n}}_\ell^{(2)}\right\}\right)\right\rangle_{\hat{\rho}_{\rm loc}} = \frac{{\rm Tr}\left(\boldsymbol{\rho} H_{\rm loc}\left(\left\{\left(\hat{\underline{n}}_\ell^{(2)}\right)_w\right\}\right)\right)}{{\rm Tr}\left(\boldsymbol{\rho}(\hat{\underline{1}})_w\right)} \tag{155}$$

$$= \langle H_{\rm loc}(\{\hat{n}_\ell\})\rangle_{\boldsymbol{\rho}}, \tag{156}$$

using Eqs. (151) and (147), where $\hat{n}_\ell$ is defined in Eq. (45).

3. SCDA under a general gauge transformation

Here we discuss the SCDA within an arbitrary gauge, where $\hat{\rho}_{G2} = \hat{P}_1\hat{K}_2\hat{P}_1^\dagger$, as this will be important to understanding the gauge constrained algorithm for $\mathcal{N} = 3$. A general gauge transformation which maintains a G-type form is defined as $\hat{P}_1 \rightarrow \hat{P}_1\hat{N}^{-1}$, $\hat{K}_2 \rightarrow \hat{N}\hat{K}_2\hat{N}^\dagger$, $\hat{P}_1^\dagger \rightarrow (\hat{N}^\dagger)^{-1}\hat{P}_1^\dagger$, where $\hat{N} = \exp(\boldsymbol{\mu} \cdot \hat{\boldsymbol{n}})$. Using the results of Sec. II B, this gauge transformation can be decomposed in terms of an intra-time-step transformation given by $N_a = {\rm diag}(\mathbf{1}, \exp(-\boldsymbol{\mu}^*))$, and an inter-time-step transformation given by $N_b = {\rm diag}(\exp(-\boldsymbol{\mu}^T), \mathbf{1})$. The corresponding transformations for the integer time Green's functions can be found in Sec. II B. In the following, we focus on the transformation with the form $[\boldsymbol{\mu}]_{i\ell,i'\ell} = \delta_{ii'}\delta_{\ell\ell'}\mu_\ell$ where μ_ℓ is a real number. Using Eq. (18), the transformation of $[\boldsymbol{g}_i]_{\ell\ell'} = \delta_{\ell\ell'}g_\ell$ with the component g_ℓ is given as

$$g_\ell' = \begin{pmatrix} n_\ell & e^{-\mu_\ell}(1-n_\ell)\mathcal{R}_\ell \\ -e^{\mu_\ell}n_\ell\mathcal{R}_\ell & n_\ell \end{pmatrix}. \tag{157}$$

Using Eq. (24), the transformation for $\boldsymbol{S}_i$ is given by

$$S_\ell' = \frac{1}{n_\ell + (1-n_\ell)\mathcal{R}_\ell^2}$$
$$\times \begin{pmatrix} e^{\mu_\ell}\mathcal{R}_\ell & (n_\ell-1)(\mathcal{R}_\ell^2-1) \\ e^{2\mu_\ell}n_\ell(\mathcal{R}_\ell^2-1) & e^{\mu_\ell}\mathcal{R}_\ell \end{pmatrix}. \tag{158}$$

Finally, the transformation for $\mathcal{G}_i$ is given by

$$\mathcal{G}_\ell' = \frac{1}{n_\ell + e^{-2\mu_\ell}(1-n_\ell)} \begin{pmatrix} n_\ell & e^{-2\mu_\ell}(1-n_\ell) \\ -n_\ell & n_\ell \end{pmatrix}. \tag{159}$$

We proceed by exploring a particular choice of gauge for the SPD, referred to as the antisymmetric gauge, which will motivate the gauge choice for the case of $\mathcal{N} = 3$. The essence of the antisymmetric gauge is to choose μ_ℓ such that $[g_\ell']_{12} = -[g_\ell']_{21}$, which can be accomplished as

$$\mu_\ell = \frac{1}{2}\ln\frac{1-n_\ell}{n_\ell}. \tag{160}$$

Under the antisymmetric gauge, we have

$$\mathcal{G}_\ell' = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}, \tag{161}$$

$$S_\ell' = \frac{1}{n_\ell + (1-n_\ell)\mathcal{R}_\ell^2} \tag{162}$$

$$\times \begin{pmatrix} \sqrt{\frac{1}{n_\ell}-1}\mathcal{R}_\ell & (n_\ell-1)(\mathcal{R}_\ell^2-1) \\ -(n_\ell-1)(\mathcal{R}_\ell^2-1) & \sqrt{\frac{1}{n_\ell}-1}\mathcal{R}_\ell \end{pmatrix}, \tag{163}$$

$$g_\ell' = \begin{pmatrix} n_\ell & \sqrt{(1-n_\ell)n_\ell}\mathcal{R}_\ell \\ -\sqrt{(1-n_\ell)n_\ell}\mathcal{R}_\ell & n_\ell \end{pmatrix}. \tag{164}$$

The antisymmetric gauge can also automatically satisfy the SCDA self-consistency condition, and be used to derive the qubit energy form.

V. DERIVATION OF THE QUBIT ENERGY FORM FOR $\hat{\rho}_{B2}$

In this section, we derive Eq. (56) using several different approaches. In Sec. V A, we present two derivations: a heuris-

tic approach and the CPE. In Sec. V B, we use the SCDA to evaluate $\hat{\rho}_{B2}$, and the tensor product representation is used to obtain the qubit energy form.

A. Derivation of the qubit energy form: a heuristic approach and the CPE

To begin, it is useful to rewrite the local interaction Hamiltonian given in Eq. (42) as

$$H_{\text{loc}}(\{\hat{n}_{i\ell}\}) = \sum_I E_I \delta \hat{D}_{iI}, \tag{165}$$

where $\delta \hat{D}_{iI} = \prod_{\ell \in I} \delta \hat{n}_{i\ell}$, the density fluctuation is defined as $\delta \hat{n}_{i\ell} = \hat{n}_{i\ell} - n_\ell$, the index I enumerates all possible subsets of the local spin orbitals, with $\delta \hat{D}_{iI} = 1$ when $I = \{\}$, and the parameters E_I reparametrize the coefficients in Eq. (42). Using the alternative form of the local interaction Hamiltonian given in Eq. (165), the corresponding qubit form of the trial energy is given as

$$E(\{n_{k\ell}\}, \boldsymbol{\rho}) = \sum_\ell \int dk \epsilon_{k\ell} n_{k\ell} + \sum_I E_I \delta D_I, \tag{166}$$

$$\delta D_I = \left(\prod_{\ell \in I} \mathcal{F}_\ell \right) \langle \delta \hat{D}_I \rangle_{\boldsymbol{\rho}}, \tag{167}$$

$$\mathcal{F}_\ell = \frac{A_\ell^2}{\xi_{\ell,0}^2} = \left(\frac{\int dk \sqrt{n_{k\ell}(1 - n_{k\ell})}}{\sqrt{n_\ell(1 - n_\ell)}} \right)^2, \tag{168}$$

where the variational parameters $n_{k\ell} \in [0, 1]$, the $\boldsymbol{\rho}$ is a many-body density matrix for a $2N_{\text{orb}}$ qubit system which is diagonal in the Pauli-Z basis, $\delta \hat{D}_I = \prod_{\ell \in I}(\hat{n}_\ell - n_\ell)$, and $n_\ell = \int dk n_{k\ell}$. For a given ℓ, there is constraint given by

$$\int dk n_{k\ell} = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}. \tag{169}$$

Below, we present two approaches for deriving Eq. (166).

A heuristic approach for deriving Eq. (166) is via the formal duality between $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$. First, in $\hat{\rho}_{G2}$, the center projector $\hat{K}_2$ is constrained to have the same local density as $\hat{\rho}_{G2}$, i.e., $\langle \hat{n}_{i\ell} \rangle_{\hat{K}_2} = \langle \hat{n}_{i\ell} \rangle_{\hat{\rho}_{G2}}$. Similarly, the center projector $\hat{P}_1$ in $\hat{\rho}_{B2}$ is constrained to have the same local density as $\hat{\rho}_{B2}$, i.e., $\langle \hat{n}_{i\ell} \rangle_{\hat{P}_1} = \langle \hat{n}_{i\ell} \rangle_{\hat{\rho}_{B2}}$. Second, in $\hat{\rho}_{G2}$ [see Eq. (50)], the momentum density fluctuation $\delta n_{k\ell} = n_{k\ell} - n_\ell$ is renormalized from the bare momentum density fluctuation $\delta n_{k\ell;0} = n_{k\ell;0} - n_\ell$ with a factor $\mathcal{Z} = \xi_\ell^2 / \xi_{\ell,0}^2$, which depends on $\boldsymbol{\rho}$. To the contrary, $\langle \delta \hat{D}_{iI} \rangle_{\hat{\rho}_{B2}}$ is renormalized from the reference value $\langle \delta \hat{D}_{iI} \rangle_{\hat{P}_1}$ with a factor $\prod_{\ell \in I} \mathcal{F}_\ell$, where $\mathcal{F}_\ell$ depends on $n_{k\ell}$. The expression for $\mathcal{F}_\ell$ can be determined using a counting scheme similar to the one used within the GA. We begin by transforming $\delta \hat{D}_{iI}$ into momentum space, which requires a summation over terms consisting of N_I creation and N_I annihilation operators, where N_I is the number of spin orbitals in set I. Each creation or annihilation process for $k\ell$ will be scaled by

$$\frac{\sum_{\Gamma\Gamma'} \sqrt{\langle \hat{X}_{k\ell\Gamma} \rangle_{\hat{\rho}_{B2}} \langle \hat{X}_{k\ell\Gamma'} \rangle_{\hat{\rho}_{B2}}} |\langle \Gamma' | \hat{a}_{k\ell}^\dagger | \Gamma \rangle|^2}{\sum_{\Gamma\Gamma'} \sqrt{\langle \hat{X}_{k\ell\Gamma} \rangle_{\hat{P}_1} \langle \hat{X}_{k\ell\Gamma'} \rangle_{\hat{P}_1}} |\langle \Gamma' | \hat{a}_{k\ell}^\dagger | \Gamma \rangle|^2} = \frac{\sqrt{n_{k\ell}(1 - n_{k\ell})}}{\sqrt{n_\ell(1 - n_\ell)}}, \tag{170}$$

where $\Gamma \in \{1, 2\}$ enumerates the empty and occupied states for a given $k\ell$ and

$$\hat{X}_{k\ell\Gamma} = \delta_{\Gamma-1,0}(1 - \hat{n}_{k\ell}) + \delta_{\Gamma-1,1}\hat{n}_{k\ell}. \tag{171}$$

Using the infinite-dimensional approximation where momentum conservation can be neglected, we obtain $\langle \delta \hat{D}_{iI} \rangle_{\hat{\rho}_{B2}} = (\prod_{\ell \in I} \mathcal{F}_\ell) \langle \delta \hat{D}_{iI} \rangle_{\hat{P}_1}$. By identifying $\boldsymbol{\rho}$ as the local reduced density matrix of $\hat{P}_1$, we obtain the form given in Eq. (167).

A more rigorous way to derive the factor in Eq. (167) is to use the CPE. We outline the derivation here, and the details can be found in Appendix C 2. The CPE for $\hat{\rho}_{B2}$ is a dual version to the CPE of $\hat{\rho}_{G2}$, where the latter is described in Sec. IV A 1. We start by defining the central point as $\hat{\rho}_{B2}^\star = \hat{K}_1 \hat{P}_1^\star \hat{K}_1$, where $\hat{P}_1^\star$ is a noninteracting projector chosen such that $n_\ell \equiv \langle \hat{n}_{i\ell} \rangle_{\hat{P}_1} = \langle \hat{n}_{i\ell} \rangle_{\hat{P}_1^\star}$. The kinetic projector $\hat{K}_1$ can be parametrized using $n_{k\ell}^\star = \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{B2}^\star}$ with the constraint $\int dk n_{k\ell}^\star = n_\ell$, while $\hat{P}_1$ can be parametrized using $\delta D_{iI;0} = \langle \delta \hat{D}_{iI} \rangle_{\hat{P}_1}$ and $\{n_\ell\}$. Therefore, the observables under $\hat{\rho}_{B2}$ are functionals of $n_{k\ell}^\star$ and $\{\delta D_{iI;0}\}$. Performing a first-order expansion in terms of $\{\delta D_{iI;0}\}$ about $\{\delta D_{iI;0} = 0\}$, we obtain $n_{k\ell} = n_{k\ell}^\star$ and

$$\delta D_{iI} = \left(\prod_{\ell \in I} \mathcal{F}_\ell \right) \delta D_{iI;0}. \tag{172}$$

Finally, we explain how to obtain Eqs. (56)–(58) from Eq. (172). We begin by defining the effective density operator $\hat{n}_{\text{eff},\ell} = n_\ell + \mathcal{F}_\ell \delta \hat{n}_\ell$ where $\delta \hat{n}_\ell = \hat{n}_\ell - n_\ell$, the corresponding fluctuation form $\delta \hat{n}_{\text{eff},\ell} \equiv \hat{n}_{\text{eff},\ell} - n_\ell = \mathcal{F}_\ell \delta \hat{n}_\ell$, and a diagonal many-body density matrix $\boldsymbol{\rho}$ for a $2N_{\text{orb}}$ qubit system $[\boldsymbol{\rho}]_{\Gamma\Gamma'} = \delta_{\Gamma\Gamma'} \langle \hat{X}_{i\Gamma} \rangle_{\hat{P}_1}$. Using $\delta D_{iI,0} = \langle \prod_{i\in I} \delta \hat{n}_\ell \rangle_{\boldsymbol{\rho}}$ and Eq. (172), we have $\delta D_{iI} = \langle \prod_{i\in I} \delta \hat{n}_{\text{eff},\ell} \rangle_{\boldsymbol{\rho}}$. Furthermore, $n_\ell = \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}}$ implies that $\langle \delta \hat{n}_\ell \rangle_{\boldsymbol{\rho}} = \langle \delta \hat{n}_{\text{eff},\ell} \rangle_{\boldsymbol{\rho}} = 0$. Therefore, the expectation value of $H_{\text{loc}}(\{\hat{n}_{i\ell}\})$ is given by $\langle H_{\text{loc}}(\{\hat{n}_{\text{eff},\ell}\}) \rangle_{\boldsymbol{\rho}}$, providing the connection to Eqs. (56)–(58).

B. Derivation of the qubit energy form via the SCDA

In this section, we use the gauge constrained SCDA to derive the qubit energy form. The derivation consists of two steps. First, in Sec. V B 1, we use the Gutzwiller gauge to automatically satisfy the SCDA self-consistency condition. Second, in Sec. V B 2, we derive the qubit energy form using the tensor representation. Additionally, in Sec. V B 3, we discuss how the quantities $\boldsymbol{\mathcal{G}}$, $\boldsymbol{S}$, and $\boldsymbol{g}$ change under a general gauge transformation.

1. SCDA within the Gutzwiller gauge

In this section, we use the gauge constrained SCDA to evaluate $\hat{\rho}_{B2} = \hat{K}_1 \hat{P}_1 \hat{K}_1$, where $\hat{P}_1 = \exp(\sum_{i\Gamma} \upsilon_\Gamma \hat{X}_{i\Gamma})$ and $\hat{K}_1 = \exp(\sum_{k\ell} \gamma_{k\ell} \hat{n}_{k\ell})$, recovering the results from Sec. V A. The key observation is that for $\hat{\rho}_{B2}$, the interacting projectors only act on the first integer time step. Consider a gauge transformation for $\hat{\rho}_{B2}$ given as $\hat{K}_1 \rightarrow \hat{K}_1 \hat{N}$ and $\hat{P}_1 \rightarrow \hat{N}^{-1} \hat{P}_1 \hat{N}^{-1}$ where $\hat{N} = \exp(\sum_{i\ell} \mu_\ell \hat{n}_{i\ell})$. We choose the gauge transformation such that $\boldsymbol{\mathcal{G}}_i$ is given as $[\boldsymbol{\mathcal{G}}_i]_{\ell\ell'} = \delta_{\ell\ell'} \mathcal{G}_\ell$, where

$$\mathcal{G}_\ell = \begin{pmatrix} n_\ell & \mathcal{G}_{\ell,1,2} \\ \mathcal{G}_{\ell,2,1} & \mathcal{G}_{\ell,2,2} \end{pmatrix}, \tag{173}$$

and $n_\ell = \langle \hat{n}_{i\ell} \rangle_{\hat{\rho}_{B2;0}}$ is the local density for the noninteracting SPD $\hat{\rho}_{B2;0} = \hat{K}_1 \hat{K}_1$. Under this gauge choice, $\hat{P}_1$ will ensure that $[\boldsymbol{g}_i]_{\ell\ell'} = \delta_{\ell\ell'} g_\ell$ with $[g_\ell]_{11} = [\mathcal{G}_\ell]_{11} = n_\ell$. The integer time self-energy $[\boldsymbol{S}_i]_{\ell\ell'} = \delta_{\ell\ell'} S_\ell$ can then be determined as

$$S_\ell = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{174}$$

The remaining entries of Eq. (173) should be determined from the self-consistency condition in Eq. (32). Given that S_ℓ is the identity matrix, the noninteracting and interacting integer time Green's functions are given by $[\boldsymbol{g}_{k;0}]_{\ell\ell'} = \delta_{\ell\ell'} g_{k\ell;0}$ and $[\boldsymbol{g}_k]_{\ell\ell'} = \delta_{\ell\ell'} g_{k\ell}$ where

$$g_{k\ell;0} = g_{k\ell} = \begin{pmatrix} n_{k\ell} & \sqrt{(1-n_{k\ell})n_{k\ell}} \\ -\sqrt{(1-n_{k\ell})n_{k\ell}} & n_{k\ell} \end{pmatrix}, \tag{175}$$

and $n_{k\ell}$ is the momentum density distribution. We can then verify that the self-consistency condition $\boldsymbol{g}_i = (1/N_{\text{site}})\sum_k \boldsymbol{g}_k$ is fulfilled, given that

$$g_\ell = \mathcal{G}_\ell = \begin{pmatrix} n_\ell & A_\ell \\ -A_\ell & n_\ell \end{pmatrix}, \tag{176}$$

where $n_\ell = (1/N_{\text{site}})\sum_k n_{k\ell}$ and the A_ℓ is defined as

$$A_\ell = \frac{1}{N_{\text{site}}} \sum_k \sqrt{(1-n_{k\ell})n_{k\ell}}. \tag{177}$$

We analogously refer to this gauge as the Gutzwiller gauge given that $g_\ell = \mathcal{G}_\ell$, and therefore the diagonal elements of the two matrices are the same.

2. Derivation of the qubit energy form

In this section, we use the tensor product representation to derive the qubit energy form, paralleling the procedure in Sec. IV B 2. Given that we only have interacting projectors at the first integer time step, operators in the u representation $(\hat{\underline{O}})_u$ can be reduced from a matrix to a vector, given as

$$[(\hat{\underline{O}})_u]_\Gamma = \langle \underline{\hat{X}}_{i\Gamma}^{(1)} \hat{\underline{O}} \rangle_{\hat{\rho}_{\text{loc},i;0}}, \tag{178}$$

and the expectation value of $\hat{\underline{O}}$ at site i is given as

$$\langle \hat{\underline{O}} \rangle_{\hat{\rho}_{\text{loc};i}} = \frac{(\hat{\underline{O}})_u \cdot u}{(\hat{\underline{1}})_u \cdot u}, \tag{179}$$

where $u = u_1$ and the dot denotes the normal dot product between two vectors. Given the direct product structure of $(\hat{\underline{O}})_u$, we only need to compute the component for a given spin-orbital. Here, we only list the relevant matrix elements needed to derive the w representation, given as

$$(\hat{\underline{1}})_{u;\ell} = (1 - n_\ell \quad n_\ell), \tag{180}$$

$$(\hat{\underline{a}}_\ell^{\dagger(1)} \hat{\underline{a}}_\ell^{(1)})_{u;\ell} = (0 \quad n_\ell), \tag{181}$$

$$(\hat{\underline{a}}_\ell^{\dagger(2)} \hat{\underline{a}}_\ell^{(2)})_{u;\ell} = (-A_\ell^2 - n_\ell^2 + n_\ell \quad A_\ell^2 + n_\ell^2). \tag{182}$$

The u representation and w representation are related by $u = Vw$ such that

$$(\hat{\underline{O}})_w w = (\hat{\underline{O}})_u u, \tag{183}$$

and therefore

$$(\hat{\underline{O}})_w = (\hat{\underline{O}})_u V. \tag{184}$$

When V has a direct product form $V = V_1 \otimes \cdots \otimes V_{2N_{\text{orb}}}$, then

$$(\hat{\underline{O}})_{w;\ell} = (\hat{\underline{O}})_{u;\ell} V_\ell, \tag{185}$$

and V_ℓ is chosen to obtain $(\hat{\underline{1}})_{w;\ell}$ as a vector of ones. The matrix elements are given as

$$V_\ell = \begin{pmatrix} \frac{1}{1-n_\ell} & 0 \\ 0 & \frac{1}{n_\ell} \end{pmatrix}, \tag{186}$$

$$(\hat{\underline{1}})_{w;\ell} = (1 \quad 1), \tag{187}$$

$$(\hat{\underline{a}}_\ell^{\dagger(1)} \hat{\underline{a}}_\ell^{(1)})_{w;\ell} = (0 \quad 1), \tag{188}$$

$$(\hat{\underline{a}}_\ell^{\dagger(2)} \hat{\underline{a}}_\ell^{(2)})_{w;\ell} = \begin{pmatrix} \frac{A_\ell^2}{n_\ell - 1} + n_\ell & \frac{A_\ell^2}{n_\ell} + n_\ell \end{pmatrix}. \tag{189}$$

From the preceding equations, we have

$$(\hat{\underline{a}}_\ell^{\dagger(2)} \hat{\underline{a}}_\ell^{(2)} - n_\ell \hat{\underline{1}})_{w;\ell} = (\hat{\underline{a}}_\ell^{\dagger(1)} \hat{\underline{a}}_\ell^{(1)} - n_\ell \hat{\underline{1}})_{w;\ell} \mathcal{F}_\ell, \tag{190}$$

where

$$\mathcal{F}_\ell = \frac{A_\ell^2}{(1-n_\ell)n_\ell}, \tag{191}$$

and therefore we have

$$\langle \delta \hat{\underline{D}}_{iI}^{(2)} \rangle_{\hat{\rho}_{\text{loc};i}} = \left(\prod_{\ell \in I} \mathcal{F}_\ell \right) \langle \delta \hat{\underline{D}}_{iI}^{(1)} \rangle_{\hat{\rho}_{\text{loc};i}}. \tag{192}$$

We now proceed to reinterpret the local energy in terms of the qubit representation, identifying $(\hat{\underline{a}}_\ell^{\dagger(1)} \hat{\underline{a}}_\ell^{(1)})_w = \text{diag}(\hat{n}_\ell)$ and $(\hat{\underline{a}}_\ell^{\dagger(2)} \hat{\underline{a}}_\ell^{(2)})_w = \text{diag}(\hat{n}_{\text{eff},\ell})$, Eq. (190) becomes

$$\hat{n}_{\text{eff},\ell} - n_\ell = (\hat{n}_\ell - n_\ell) \mathcal{F}_\ell. \tag{193}$$

Furthermore, we define a diagonal many-body density matrix $\boldsymbol{\rho}$ with $\text{diag}(\boldsymbol{\rho}) = w$, resulting in

$$\langle H_{\text{loc}}(\{\hat{\underline{n}}_\ell^{(2)}\}) \rangle_{\hat{\rho}_{\text{loc};i}} = \frac{(H_{\text{loc}}(\{\hat{\underline{n}}_\ell^{(2)}\}))_w \cdot w}{(\hat{\underline{1}})_w \cdot w} \tag{194}$$

$$= \langle H_{\text{loc}}(\{n_{\text{eff},\ell}\}) \rangle_{\boldsymbol{\rho}}. \tag{195}$$

Therefore, the qubit energy form has been recovered.

3. SCDA under a general gauge transformation

Here we discuss the SCDA within an arbitrary gauge, analogous to Sec. IV B 3. Given $\hat{\rho}_{B2} = \hat{K}_1 \hat{P}_1 \hat{K}_1^\dagger$, a general gauge transformation is given as: $\hat{K}_1 \to \hat{K}_1 \hat{N}$, $\hat{P}_1 \to \hat{N}^{-1} \hat{P}_1 (\hat{N}^\dagger)^{-1}$, $\hat{K}_1^\dagger \to \hat{N}^\dagger \hat{K}_1^\dagger$, where $\hat{N} = \exp(\boldsymbol{\mu} \cdot \hat{\boldsymbol{n}})$. Therefore, we have $N_a = \text{diag}(\exp(-\boldsymbol{\mu}^T), \boldsymbol{1})$ and $N_b = \text{diag}(\exp(-\boldsymbol{\mu}^*), \boldsymbol{1})$. Assuming $[\boldsymbol{\mu}]_{\ell\ell'} = \delta_{\ell\ell'}\mu_\ell$, where μ_ℓ is real, we can obtain the transformation for g_ℓ, S_ℓ, and $\mathcal{G}_\ell$ as

$$g'_\ell = \begin{pmatrix} n_\ell & A_\ell e^{-\mu_\ell} \\ -A_\ell e^{\mu_\ell} & n_\ell \end{pmatrix}, \tag{196}$$

$$S'_\ell = \begin{pmatrix} e^{2\mu_\ell} & 0 \\ 0 & 1 \end{pmatrix}, \tag{197}$$

$$\mathcal{G}'_\ell = \frac{1}{(e^{2\mu_\ell} - 1)n_\ell + 1} \tag{198}$$

$$\times \begin{pmatrix} e^{2\mu_\ell} n_\ell & A_\ell e^{\mu_\ell} \\ A_\ell(-e^{\mu_\ell}) & (e^{2\mu_\ell} - 1)(A_\ell^2 + n_\ell^2) + n_\ell \end{pmatrix}. \tag{199}$$

We can define an antisymmetric gauge, similar to the $\hat{\rho}_{G2}$ case, where we choose $[\mathcal{G}'_\ell]_{11} = 1/2$, leading to $\mu_\ell = \frac{1}{2}\ln(\frac{1-n_\ell}{n_\ell})$, yielding

$$g'_\ell = \begin{pmatrix} n_\ell & A_\ell/\sqrt{\frac{1}{n_\ell}-1} \\ -A_\ell\sqrt{\frac{1}{n_\ell}-1} & n_\ell \end{pmatrix}, \quad (200)$$

$$S'_\ell = \begin{pmatrix} \frac{1-n_\ell}{n_\ell} & 0 \\ 0 & 1 \end{pmatrix}, \quad (201)$$

$$\mathcal{G}'_\ell = \begin{pmatrix} \frac{1}{2} & \frac{A_\ell}{2\sqrt{-(n_\ell-1)n_\ell}} \\ -\frac{A_\ell}{2\sqrt{-(n_\ell-1)n_\ell}} & \frac{A_\ell^2(2n_\ell-1)}{2(n_\ell-1)n_\ell}+n_\ell \end{pmatrix}. \quad (202)$$

The antisymmetric gauge can also automatically satisfy the SCDA self-consistency condition, and be used to derive the qubit energy form.

VI. DERIVATION OF THE QUBIT ENERGY FORM FOR $\hat{\rho}_{G3}$

In this section, we derive Eq. (62) using the gauge constrained SCDA. In Sec. VI A, we provide a high-level comparison of the original gauge constrained algorithm and the qubit parametrization. In Sec. VI B, we provide a review of the original gauge constrained algorithm, which is necessary to understand the qubit parametrization in this work. In Sec. VI C, we propose the qubit parametrization which yields the qubit energy form. In Sec. VI D, we examine the case of half-filling, and explore how the qubit energy form for $\hat{\rho}_{G3}$ can recover the cases of $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$.

A. Comparing the original gauge constrained trial energy to the qubit energy form

In this section, we outline how the qubit parametrization improves the original gauge constrained algorithm [17]. In the original gauge constrained algorithm, the trial energy under $\hat{\rho}_{G3}$ is given as

$$\sum_\ell \int dk \,\epsilon_{k\ell} n_{k\ell} + E_{\text{loc}}(\{\mathcal{G}_{12,\ell}\}, \{u_\Gamma\}, \{\mathcal{A}_{<\ell}\}, \{\mathcal{A}_{>\ell}\}), \quad (203)$$

which is a function of $\{n_{k\ell}\}$, $\{\mathcal{G}_{12,\ell}\}$, $\{u_\Gamma\}$, and $\{n_{k\ell;0}\}$, given that $\{\mathcal{A}_{<\ell}\}$ and $\{\mathcal{A}_{>\ell}\}$ are functions of $\{n_{k\ell}\}$ and $\{n_{k\ell;0}\}$. It should be noted that $n_{k\ell;0} \in \{0,1\}$ and $n_{k\ell} \in [0,1]$, and there are three constraints for a given ℓ

$$\int_< dk = n_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\}), \quad (204)$$

$$\int n_{k\ell} dk = n_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\}), \quad (205)$$

$$\Delta_\ell \equiv \int_> n_{k\ell} dk = \Delta_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\}), \quad (206)$$

where the functions $n_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\})$ and $\Delta_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\})$ are explicitly defined in Ref. [17]. While this parametrization allows for an explicit evaluation of the total energy, there are several shortcomings of this parametrization. First, the function $n_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\})$ is highly nontrivial and thus the minimization under a fixed density is cumbersome. Previously, this problem was addressed by introducing a linear

transformation over $\{u_\Gamma\}$, known as the w representation. Second, the function $\Delta_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\})$ may yield a value outside the allowed bounds for Δ_ℓ. This problem can be addressed by imposing appropriate restrictions on $\{\mathcal{G}_{12;\ell}\}$. Finally, $\{\mathcal{G}_{12;\ell}\}$ makes the physical interpretation of the total energy expression somewhat obscure.

In this paper, the aforementioned shortcomings are resolved using the qubit parametrization (see Sec. III). The qubit parametrization has two important differences. First, the qubit parametrization employs an effective many-body density matrix $\boldsymbol{\rho}$, having dimension $2^{2N_{\text{orb}}} \times 2^{2N_{\text{orb}}}$, corresponding to a pure state of a $2N_{\text{orb}}$ qubit system. The $\boldsymbol{\rho}$ is constructed such that the density of $\boldsymbol{\rho}$ is the same as the physical local density, and it can be viewed as a function of $\{u_\Gamma\}$ and $\{\mathcal{G}_{12;\ell}\}$. Second, the qubit parametrization uses $\Delta_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\})$ to solve $\{\mathcal{G}_{12,\ell}\}$ as a function of $\boldsymbol{\rho}$ and $\{\Delta_\ell\}$, reducing the total number of constraints per spin orbital from three to two.

B. Review of the gauge constrained SCDA algorithm

In this section, we use original gauge constrained SCDA to evaluate $\hat{\rho}_{G3}$. It should be noted that there are several restrictions on the variational freedom of the SPD when using the gauge constrained SCDA. First, the kinetic projector must be diagonal in momentum space. Second, the interacting projector may not introduce off-diagonal terms at the single-particle level. These two restrictions guarantee that the local integer time self-energy and Green's function are diagonal in the original basis. For Hamiltonians with density-density interactions and hopping parameters that are diagonal in the orbital index, such as the ones treated in this paper, the aforementioned variational restrictions do not limit the variational power of the SPD. There are two critical insights in the gauge constrained algorithm. First, the integer time self-energy only has nontrivial values within the time steps containing the interacting projector, and therefore $\mathcal{G}$ only needs to be specified in the corresponding regions. Second, the gauge symmetry can be used to restrict the form of $\mathcal{G}$.

We start by examining the gauge symmetry of $\hat{\rho}_{G3} = \hat{K}_1\hat{P}_1\hat{K}_2\hat{P}_1^\dagger\hat{K}_1^\dagger$, where the gauge transformation is given by $\hat{K}_1 \to \hat{K}_1\hat{N}_1$, $\hat{P}_1 \to \hat{N}_1^{-1}\hat{P}_1\hat{N}_1^{-1}$, and $\hat{K}_2 \to \hat{N}_2\hat{K}_2\hat{N}_2^\dagger$, where $\hat{N}_1 = \exp(\boldsymbol{\mu}_1 \cdot \hat{\boldsymbol{n}})$ and $\hat{N}_2 = \exp(\boldsymbol{\mu}_2 \cdot \hat{\boldsymbol{n}})$. The gauge transformation can be parametrized by

$$N_a = \text{diag}\big(\exp(-\boldsymbol{\mu}_1^T), \exp(-\boldsymbol{\mu}_2^*), 1\big), \quad (207)$$

$$N_b = \text{diag}\big(\exp(-\boldsymbol{\mu}_2^T), \exp(-\boldsymbol{\mu}_1^\star), 1\big), \quad (208)$$

as explained in Sec. II B. In the following, we assume $[\boldsymbol{\mu}_i]_{\ell\ell'} = \delta_{\ell\ell'}\mu_{i,\ell}$ where $\mu_{i,\ell}$ is a real number, yielding

$$g'_\ell = \text{diag}(e^{-\mu_{2,\ell}}, e^{-\mu_{1,\ell}}, 1)g_\ell\text{diag}(e^{\mu_{2,\ell}}, e^{\mu_{1,\ell}}, 1), \quad (209)$$

$$S'_\ell = \text{diag}(e^{\mu_{1,\ell}}, e^{\mu_{2,\ell}}, 1)S_\ell\text{diag}(e^{\mu_{2,\ell}}, e^{\mu_{1,\ell}}, 1), \quad (210)$$

$$(\mathcal{G}'_\ell)^{-1} - 1 = \text{diag}(e^{-\mu_{2,\ell}}, e^{-\mu_{1,\ell}}, 1)$$
$$\times (\mathcal{G}_\ell^{-1} - 1)\text{diag}(e^{-\mu_{1,\ell}}, e^{-\mu_{2,\ell}}, 1). \quad (211)$$

Notice that the interacting projectors only act on the first and second time step, and therefore it is useful to split $\mathcal{G}_\ell$ into the

following block structure:

$$
\mathcal{G}_\ell = \left(\begin{array}{cc|c} [\mathcal{G}_\ell]_{11} & [\mathcal{G}_\ell]_{12} & [\mathcal{G}_\ell]_{13} \\ [\mathcal{G}_\ell]_{21} & [\mathcal{G}_\ell]_{22} & [\mathcal{G}_\ell]_{23} \\ \hline [\mathcal{G}_\ell]_{31} & [\mathcal{G}_\ell]_{32} & [\mathcal{G}_\ell]_{33} \end{array} \right) = \begin{pmatrix} \mathcal{G}_{\ell;A} & \mathcal{G}_{\ell;B} \\ \mathcal{G}_{\ell;C} & \mathcal{G}_{\ell;D} \end{pmatrix}.
$$

$$(212)$$

A similar block structure is adopted for g_ℓ and S_ℓ.

The first step is to focus on the A block, which is sufficient to determine the integer time self-energy. Similar to the case of $\hat{\rho}_{B2}$, we only need to specify $\mathcal{G}_{\ell;A}$, which is sufficient to determine $g_{\ell;A}$ and $S_{\ell;A}$ and therefore S_ℓ. Moreover, similar to the derivation of Eq. (211), we have

$$
\left(\mathcal{G}'_{\ell;A}\right)^{-1} - 1 = \mathrm{diag}(e^{-\mu_{2,\ell}}, e^{-\mu_{1,\ell}})\left(\mathcal{G}^{-1}_{\ell;A} - 1\right)
$$
$$
\times \mathrm{diag}(e^{-\mu_{1,\ell}}, e^{-\mu_{2,\ell}}). \quad (213)
$$

The gauge transformation can be used to further restrict the form of $\mathcal{G}_{\ell;A}$. Given that Eq. (213) involves the inverse of $\mathcal{G}_{\ell;A}$, it is more convenient to first use Eq. (209). Notice that $[g_\ell]_{11} = [g_\ell]_{22}$ given translation symmetry and the fact that the total particle number of a given spin-orbital ℓ commutes with $\hat{P}_1$. Therefore, Eq. (209) indicates that the diagonal elements of g_ℓ are invariant, while $\mu_{1;\ell} - \mu_{2;\ell}$ can be chosen such that $[g'_\ell]_{12} = -[g'_\ell]_{21}$. Notice that the interacting projectors are the same for the first and second time step, and therefore $[\mathcal{G}'_\ell]_{11} = [\mathcal{G}'_\ell]_{22}$ and $[\mathcal{G}'_\ell]_{12} = -[\mathcal{G}'_\ell]_{21}$. Now consider a gauge transformation with $\mu_{1;\ell} = \mu_{2;\ell} = \mu_\ell$, which still preserves $[\mathcal{G}''_\ell]_{11} = [\mathcal{G}''_\ell]_{22}$ and $[\mathcal{G}''_\ell]_{12} = -[\mathcal{G}''_\ell]_{21}$, and we can choose μ_ℓ such that $[\mathcal{G}''_\ell]_{11} = 1/2$. Therefore, after fully exploring the gauge symmetry of the SPD, it is sufficient to use just one parameter $\mathcal{G}_{\ell,12}$ to parametrize $\mathcal{G}_{\ell;A}$ as

$$
\mathcal{G}_{\ell;A} = \begin{pmatrix} \frac{1}{2} & \mathcal{G}_{\ell,12} \\ -\mathcal{G}_{\ell,12} & \frac{1}{2} \end{pmatrix}, \quad (214)
$$

and correspondingly

$$
g_{\ell;A} = \begin{pmatrix} n_\ell & g_{\ell,12} \\ -g_{\ell,12} & n_\ell \end{pmatrix}. \quad (215)
$$

It is useful to define a new quantity

$$
S^{(A)}_{0,\ell} \equiv \mathcal{G}^{-1}_{\ell;A} - 1 = \begin{pmatrix} \frac{2}{4\mathcal{G}^2_{\ell,12}+1} - 1 & -\frac{4\mathcal{G}_{\ell,12}}{4\mathcal{G}^2_{\ell,12}+1} \\ \frac{\mathcal{G}_{\ell,12}}{\mathcal{G}^2_{\ell,12}+\frac{1}{4}} & \frac{2}{4\mathcal{G}^2_{\ell,12}+1} - 1 \end{pmatrix}, \quad (216)
$$

which is different from the A block of $S_{0,\ell} = \mathcal{G}^{-1}_\ell - 1$, denoted as $S_{0,\ell;A}$. A similar quantity $S^{(A)}_{F,\ell}$ can also be defined as

$$
S^{(A)}_{F,\ell} = g^{-1}_{\ell;A} - 1 = \begin{pmatrix} \frac{n_\ell}{g^2_{\ell,12}+n^2_\ell} - 1 & -\frac{g_{\ell,12}}{g^2_{\ell,12}+n^2_\ell} \\ \frac{g_{\ell,12}}{g^2_{\ell,12}+n^2_\ell} & \frac{n_\ell}{g^2_{\ell,12}+n^2_\ell} - 1 \end{pmatrix}, \quad (217)
$$

which is different from the A block of $S_{F,\ell} = g^{-1}_\ell - 1$, denoted as $S_{F,\ell;A}$. Using the results of Sec. VI B 2, the integer time Dyson equation within the A block may be written as

$$
S^{(A)}_{F,\ell} = S^{(A)}_{0,\ell}S_{\ell;A}, \quad (218)
$$

and $S_{\ell;A}$ can be determined as

$$
S_{\ell;A} = \begin{pmatrix} S_{\ell,11} & S_{\ell,12} \\ -S_{\ell,12} & S_{\ell,11} \end{pmatrix},
$$

where Eqs. (214) and (215) can be used to obtain

$$
S_{\ell,11} = \left(4\mathcal{G}^2_{\ell,12}+1\right)^{-1}\left(g^2_{\ell,12}+n^2_\ell\right)^{-1}\left(4g_{\ell,12}\mathcal{G}_{\ell,12} \right.
$$
$$
\left. + 4\mathcal{G}^2_{\ell,12}\left(g^2_{\ell,12}+(n_\ell-1)n_\ell\right) - g^2_{\ell,12} - n^2_\ell + n_\ell\right), \quad (219)
$$

$$
S_{\ell,12} = \left(4\mathcal{G}^2_{\ell,12}+1\right)^{-1}\left(g^2_{\ell,12}+n^2_\ell\right)^{-1}(4(1-n_\ell)n_\ell\mathcal{G}_{\ell,12}
$$
$$
+ g_{\ell,12}(4\mathcal{G}_{\ell,12}(\mathcal{G}_{\ell,12}-g_{\ell,12})-1)). \quad (220)
$$

The S_ℓ is a 3×3 matrix obtained from $S_{F,\ell} = S_{0,\ell}S_\ell$, yielding

$$
S_\ell = \begin{pmatrix} S_{\ell;A} & 0 \\ 0 & 1 \end{pmatrix}. \quad (221)
$$

The second step is to examine the lattice integer time Green's function and use the self-consistency condition to resolve the B, C, and D blocks. The lattice integer time Green's function can be parametrized by the physical momentum density distribution $n_{k\ell} \equiv \langle \hat{\underline{n}}^{(3)}_{k\ell} \rangle_{\hat{\rho}_K}$ and the integer time self-energy S_ℓ by assuming $\hat{K}_2$ is a single Slater determinant with $\langle \hat{n}_{k\ell} \rangle_{\hat{K}_2} = 1$ for $k \in \langle$ and $\langle \hat{n}_{k\ell} \rangle_{\hat{K}_2} = 0$ for $k \in \rangle$, where the symbols $\langle$ or $\rangle$ denote the occupied and unoccupied regions of $\hat{K}_2$. The local integer time Green's function for the lattice is denoted as $[\boldsymbol{g}'_i]_{\ell\ell'} = \delta_{\ell\ell'}g'_\ell$, where

$$
g'_\ell = \begin{pmatrix} g'_{\ell,A} & g'_{\ell,B} \\ g'_{\ell,C} & n_\ell \end{pmatrix}, \quad (222)
$$

where $g'_{\ell,A}$, $g'_{\ell,B}$, and $g'_{\ell,C}$ are 2×2, 2×1, and 1×2 matrices, respectively, defined as

$$
g'_{\ell,A} = \begin{pmatrix} n_\ell & \frac{S_{\ell,11}}{S_{\ell,12}}\Delta_\ell \\ -\frac{S_{\ell,11}}{S_{\ell,12}}\Delta_\ell & n_\ell \end{pmatrix}, \quad (223)
$$

$$
g'_{\ell,B} = \begin{pmatrix} \frac{S_{\ell,11}}{\sqrt{S_{\ell,12}(S^2_{\ell,11}+S^2_{\ell,12})}}\mathcal{A}_{>,\ell} \\ \frac{1}{\sqrt{S_{\ell,12}}}\mathcal{A}_{<,\ell} + \sqrt{\frac{S_{\ell,12}}{S^2_{\ell,11}+S^2_{\ell,12}}}\mathcal{A}_{>,\ell} \end{pmatrix}, \quad (224)
$$

$$
g'_{\ell,C} = \begin{pmatrix} -\frac{S_{\ell,11}}{\sqrt{S_{\ell,12}}}\mathcal{A}_{<,\ell} \\ -\sqrt{S_{\ell,12}}\mathcal{A}_{<,\ell} - \sqrt{\frac{S^2_{\ell,11}+S^2_{\ell,12}}{S_{\ell,12}}}\mathcal{A}_{>,\ell} \end{pmatrix}^T, \quad (225)
$$

where $\mathcal{A}_{<,\ell}$, $\mathcal{A}_{>,\ell}$, and Δ_ℓ are defined in Eqs. (68), (69), and (67).

Using Eqs. (222), (260), (261), and (262), the B, C, and D blocks of Eq. (212) can be fully determined, completing the algorithm. Finally, the local interaction energy can then be written using the tensor product representation as

$$
E_{\mathrm{loc}}(\{\mathcal{G}_{12,\ell}\}, \{u_\Gamma\}, \{\mathcal{A}_{<,\ell}\}, \{\mathcal{A}_{>,\ell}\}) = \frac{u^T\left(\hat{\underline{\boldsymbol{H}}}^{(3)}_{\mathrm{loc};i}\right)_u u}{u^T\left(\hat{\underline{1}}\right)_u u}, \quad (226)
$$

while the kinetic energy is $\sum_\ell \int dk \epsilon_{k\ell}n_{k\ell}$. The $n_{k\ell}$ are subject to the two linear constraints given in Eqs. (205) and (206), which can be implemented in the u representation using

$$
n_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\}) = \frac{u^T\left(\hat{\underline{\boldsymbol{a}}}^{\dagger(1)}_\ell\hat{\underline{\boldsymbol{a}}}^{(1)}_\ell\right)_u u}{u^T\left(\hat{\underline{1}}\right)_u u}, \quad (227)
$$

$$
\Delta_\ell(\{\mathcal{G}_{12;\ell}\}, \{u_\Gamma\}) = \frac{S_{\ell,12}}{S_{\ell,11}}\frac{u^T\left(\hat{\underline{\boldsymbol{a}}}^{\dagger(1)}_\ell\hat{\underline{\boldsymbol{a}}}^{(2)}_\ell\right)_u u}{u^T\left(\hat{\underline{1}}\right)_u u}, \quad (228)
$$

while the $n_{k\ell;0}$ are subject to $\int_< dk = n_\ell$.

It is useful to appreciate what can be gleaned from the form of the preceding equations. The $g'_{\ell,A}$ block is reminiscent of the $\hat{\rho}_{G2}$ case where Δ_ℓ captures the quasiparticle renormalization, while $g'_{\ell,B}$ and $g'_{\ell,C}$ are reminiscent of the $\hat{\rho}_{B2}$ case where $\mathcal{A}_{<\ell}$ and $\mathcal{A}_{>\ell}$ capture the super-exchange effects. This observation helps illustrate how the $\hat{\rho}_{G3}$ simultaneously captures the physics of both $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$.

In the following, we further elaborate on two points which were not fully elucidated in Ref. [17]. First, we provide explicit expressions for the tensor product representation in the case of $\mathcal{N} = 3$. Second, we explore various relations derived using the block structure of the integer time Dyson equation.

1. Evaluating observables under $\hat{\rho}_{\mathrm{loc}}$ using the tensor product representation

Consider the tensor product representation of $\hat{\underline{O}}$ evaluated under $\hat{\underline{\rho}}_{\mathrm{loc}}$. Given that the interacting projector only acts on the first and second time step, the three-dimensional tensor representation may be reduced to a two-dimensional tensor representation as

$$[(\hat{\underline{O}})_u]_{\Gamma_1\Gamma_2} = \langle \hat{\underline{X}}^{(1)}_{i\Gamma_1} \hat{\underline{X}}^{(2)}_{i\Gamma_2} \hat{\underline{O}} \rangle_{\hat{\underline{\rho}}_{\mathrm{loc},i;0}}. \tag{229}$$

Furthermore, the tensor contraction can be simplified as $(\hat{\underline{O}})_u \cdot (u_1 \otimes u_2) = u^T (\hat{\underline{O}})_u u$ with $u^T = u_1 = u_2$. Therefore, even though the compound space of $\mathcal{N} = 3$ is larger than $\mathcal{N} = 2$ for a given original Hilbert space, the computational cost in the tensor representation for $\hat{\rho}_{G3}$ is similar to that of $\hat{\rho}_{G2}$ due to this dimension reduction. Similar to the case of the $\hat{\rho}_{G2}$, $(\hat{\underline{O}})_u$ has a direct product structure given by Eq. (38). We first specify the components of the A block quantities at a given spin orbital ℓ

$$(\hat{\underline{1}})_{u;\ell} = \begin{pmatrix} \mathcal{G}^2_{\ell,12} + \frac{1}{4} & \frac{1}{4} - \mathcal{G}^2_{\ell,12} \\ \frac{1}{4} - \mathcal{G}^2_{\ell,12} & \mathcal{G}^2_{\ell,12} + \frac{1}{4} \end{pmatrix}, \tag{230}$$

$$(\hat{\underline{a}}^{\dagger(1)}_\ell \hat{\underline{a}}^{(1)}_\ell)_{u;\ell,S} = (\hat{\underline{a}}^{\dagger(2)}_\ell \hat{\underline{a}}^{(2)}_\ell)_{u;\ell,S} \tag{231}$$

$$= \begin{pmatrix} 0 & \frac{1}{8}(1 - 4\mathcal{G}^2_{\ell,12}) \\ \frac{1}{8}(1 - 4\mathcal{G}^2_{\ell,12}) & \mathcal{G}^2_{\ell,12} + \frac{1}{4} \end{pmatrix}, \tag{232}$$

$$(\hat{\underline{a}}^{\dagger(1)}_\ell \hat{\underline{a}}^{(2)}_\ell)_{u;\ell,S} = -(\hat{\underline{a}}^{\dagger(2)}_\ell \hat{\underline{a}}^{(1)}_\ell)_{u;\ell,S} \tag{233}$$

$$= \begin{pmatrix} 0 & \frac{\mathcal{G}_{\ell,12}}{2} \\ \frac{\mathcal{G}_{\ell,12}}{2} & 0 \end{pmatrix}, \tag{234}$$

where S indicates the symmetric part of the matrix, which is defined as

$$(\hat{\underline{O}})_{u,S} = \tfrac{1}{2}((\hat{\underline{O}})_u + (\hat{\underline{O}})^T_u), \tag{235}$$

which is useful given that $u^T (\hat{\underline{O}})_u u = u^T (\hat{\underline{O}})_{u,S} u$. Similarly, the w representation can be defined using $u = Vw$, where $V = \otimes_\ell V_\ell$ with

$$V_\ell = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{2\mathcal{G}_{\ell,12}} + 1 & 1 - \frac{1}{2\mathcal{G}_{\ell,12}} \\ 1 - \frac{1}{2\mathcal{G}_{\ell,12}} & \frac{1}{2\mathcal{G}_{\ell,12}} + 1 \end{pmatrix}, \tag{236}$$

such that $(\hat{\underline{1}})_{w;\ell}$ is the identity matrix and $(\hat{\underline{a}}^{\dagger(1)}_\ell \hat{\underline{a}}^{(1)}_\ell)_{w;\ell,S}$ is diagonal, resulting in

$$(\hat{\underline{1}})_{w;\ell} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{237}$$

$$(\hat{\underline{a}}^{\dagger(1)}_\ell \hat{\underline{a}}^{(1)}_\ell)_{w;\ell,S} = (\hat{\underline{a}}^{\dagger(2)}_\ell \hat{\underline{a}}^{(2)}_\ell)_{w;\ell,S}, \tag{238}$$

$$= \begin{pmatrix} -\frac{(1-2\mathcal{G}_{\ell,12})^2}{8\mathcal{G}_{\ell,12}} & 0 \\ 0 & \frac{(2\mathcal{G}_{\ell,12}+1)^2}{8\mathcal{G}_{\ell,12}} \end{pmatrix}, \tag{239}$$

and correspondingly, we have

$$(\hat{\underline{a}}^{\dagger(1)}_\ell \hat{\underline{a}}^{(2)}_\ell)_{w;\ell,S} = -(\hat{\underline{a}}^{\dagger(2)}_\ell \hat{\underline{a}}^{(1)}_\ell)_{w;\ell,S} \tag{240}$$

$$= \begin{pmatrix} \frac{1}{2}\left(\mathcal{G}_{\ell,12} - \frac{1}{4\mathcal{G}_{\ell,12}}\right) & \frac{\mathcal{G}_{\ell,12}}{2} + \frac{1}{8\mathcal{G}_{\ell,12}} \\ \frac{\mathcal{G}_{\ell,12}}{2} + \frac{1}{8\mathcal{G}_{\ell,12}} & \frac{1}{2}\left(\mathcal{G}_{\ell,12} - \frac{1}{4\mathcal{G}_{\ell,12}}\right) \end{pmatrix}. \tag{241}$$

Finally, we provide explicit expressions for $(\hat{\underline{a}}^{\dagger(3)}_\ell \hat{\underline{a}}^{(3)}_\ell)_{w;\ell}$, with the components given by

$$\left[(\hat{\underline{a}}^{\dagger(3)}_\ell \hat{\underline{a}}^{(3)}_\ell)_{w;\ell}\right]_{11} = \frac{1}{2\mathcal{G}_{\ell,12}}(\mathcal{G}_{\ell,13}(\mathcal{G}_{\ell,31} - \mathcal{G}_{\ell,32})$$
$$+ \mathcal{G}_{\ell,23}(\mathcal{G}_{\ell,31} + \mathcal{G}_{\ell,32}) + 2\mathcal{G}_{\ell,12}\mathcal{G}_{\ell,33}), \tag{242}$$

$$\left[(\hat{\underline{a}}^{\dagger(3)}_\ell \hat{\underline{a}}^{(3)}_\ell)_{w;\ell}\right]_{12} = \frac{(\mathcal{G}_{\ell,13} - \mathcal{G}_{\ell,23})(\mathcal{G}_{\ell,31} + \mathcal{G}_{\ell,32})}{2\mathcal{G}_{\ell,12}}, \tag{243}$$

$$\left[(\hat{\underline{a}}^{\dagger(3)}_\ell \hat{\underline{a}}^{(3)}_\ell)_{w;\ell}\right]_{21} = -\frac{(\mathcal{G}_{\ell,13} + \mathcal{G}_{\ell,23})(\mathcal{G}_{\ell,31} - \mathcal{G}_{\ell,32})}{2\mathcal{G}_{\ell,12}}, \tag{244}$$

$$\left[(\hat{\underline{a}}^{\dagger(3)}_\ell \hat{\underline{a}}^{(3)}_\ell)_{w;\ell}\right]_{22} = \frac{1}{2\mathcal{G}_{\ell,12}}(\mathcal{G}_{\ell,23}(\mathcal{G}_{\ell,31} - \mathcal{G}_{\ell,32})$$
$$- \mathcal{G}_{\ell,13}(\mathcal{G}_{\ell,31} + \mathcal{G}_{\ell,32}) + 2\mathcal{G}_{\ell,12}\mathcal{G}_{\ell,33}), \tag{245}$$

which can be used to compute the local interaction energy.

2. Block structure of the integer time Dyson equation

Here we derive various useful equations using the block form of the integer time Dyson equation. To make our discussion general, we assume that the integer time Green's functions are not diagonal in the orbital index, resulting in the following block matrix equation

$$\begin{pmatrix} S_{F,A} & S_{F,B} \\ S_{F,C} & S_{F,D} \end{pmatrix} = \begin{pmatrix} S_{0,A} & S_{0,B} \\ S_{0,C} & S_{0,D} \end{pmatrix}\begin{pmatrix} S_A & 0 \\ 0 & 1 \end{pmatrix}, \tag{246}$$

where the blocks for S_F and S_0 can be explicitly expressed in terms of the blocks of g and $\mathcal{G}$ using the inverse formula for a 2×2 block matrix, resulting in the following relations:

$$1 + S_{F,D} = \frac{1}{g_D - g_C g_A^{-1} g_B} \tag{247}$$

$$= 1 + S_{0,D} = \frac{1}{\mathcal{G}_D - \mathcal{G}_C \mathcal{G}_A^{-1} \mathcal{G}_B}, \tag{248}$$

$$S_{F,B} = -g_A^{-1} g_B(1 + S_{F,D}) \tag{249}$$

$$= S_{0,B} = -\mathcal{G}_A^{-1}\mathcal{G}_B(1 + S_{0,D}), \tag{250}$$

$$S_{F,C} = -(1 + S_{F,D})g_C g_A^{-1} \tag{251}$$

$$= S_{0,C}S_A = (1 + S_{0,D})\mathcal{G}_C\mathcal{G}_A^{-1}S_A, \tag{252}$$

$$S_{F,A} = S_F^{(A)} + g_A^{-1}g_B(1 + S_{F,D})g_Cg_A^{-1} \tag{253}$$

$$= S_{0,A}S_A = S_0^{(A)}S_A + \mathcal{G}_A^{-1}\mathcal{G}_B(1 + S_{0,D})\mathcal{G}_C\mathcal{G}_A^{-1}S_A. \tag{254}$$

Using Eqs. (247), (248), (249), and (250), we obtain

$$g_A^{-1}g_B = \mathcal{G}_A^{-1}\mathcal{G}_B. \tag{255}$$

Similarly, using Eqs. (247), (248), and (251), we obtain

$$g_Cg_A^{-1} = \mathcal{G}_C\mathcal{G}_A^{-1}S_A. \tag{256}$$

Using Eqs. (255), (256), (253), and (254), we verify that $\mathcal{G}_A$ and g_A can be used to determine S_A as

$$S_F^{(A)} = g_A^{-1} - 1 = S_0^{(A)}S_A = (\mathcal{G}_A^{-1} - 1)S_A. \tag{257}$$

Using Eqs. (256) and (257), we have

$$g_C(1 - g_A)^{-1} = \mathcal{G}_C(1 - \mathcal{G}_A)^{-1}. \tag{258}$$

Finally, using Eq. (257), we have

$$g_D - g_Cg_A^{-1}g_B = \mathcal{G}_D - \mathcal{G}_C\mathcal{G}_A^{-1}\mathcal{G}_B, \tag{259}$$

which can be used to solve for $\mathcal{G}_A$. Additionally, it is useful to explicitly write expressions for $\mathcal{G}_B$, $\mathcal{G}_C$, and $\mathcal{G}_D$ as

$$\mathcal{G}_B = \mathcal{G}_Ag_A^{-1}g_B, \tag{260}$$

$$\mathcal{G}_C = g_C(1 - g_A)^{-1}(1 - \mathcal{G}_A), \tag{261}$$

$$\mathcal{G}_D = g_D + g_C(1 - g_A)^{-1}(g_A - \mathcal{G}_A)g_A^{-1}g_B. \tag{262}$$

C. Derivation of the qubit energy form

In this section, we derive the qubit energy form, given in Eq. (62), in several steps. First, in Sec. VI C 1, we introduce a polar representation for the A block. Second, in Sec. VI C 2, we introduce $\boldsymbol{\rho}$ as a unitary transformation of ww^T, and we solve the self-consistency condition for the A block using n_ℓ, ξ_ℓ, and Δ_ℓ. Third, in Sec. VI C 3, we resolve the self-consistency of the B, C, and D blocks using the extra information provided by $\mathcal{A}_{<\ell}$ and $\mathcal{A}_{>\ell}$.

1. Polar Representation of the A block

We first review some mathematical properties of the matrix group with following form:

$$\mathcal{S}(c, \phi) = c\begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix}, \tag{263}$$

where $c > 0$ and $\phi \in [0, 2\pi]$ and the group multiplication is

$$\mathcal{S}(c_1, \phi_1)\mathcal{S}(c_2, \phi_2) = \mathcal{S}(c, \phi). \tag{264}$$

There is an isomorphism for this matrix group to $\mathbb{R}^+ \times S^1$ with $\mathcal{S}(c, \phi) \to (c, \phi)$, with the group product

taken as

$$c_1c_2 = c, \quad \phi_1 + \phi_2 = \phi. \tag{265}$$

Interpreting $\mathcal{S}(c, \phi)$ as the integer time self-energy, the corresponding integer time Green's function is given as

$$g(c, \phi) = \frac{1}{1 + \mathcal{S}} = \begin{pmatrix} \frac{c\cos(\phi)+1}{c^2+2c\cos(\phi)+1} & -\frac{c\sin(\phi)}{c^2+2c\cos(\phi)+1} \\ \frac{c\sin(\phi)}{c^2+2c\cos(\phi)+1} & \frac{c\cos(\phi)+1}{c^2+2c\cos(\phi)+1} \end{pmatrix}, \tag{266}$$

and it can be seen that

$$g_{11} \equiv [g(c, \phi)]_{11} = [g(c, \phi)]_{22}, \tag{267}$$

$$g_{12} \equiv [g(c, \phi)]_{12} = -[g(c, \phi)]_{21}. \tag{268}$$

It is also useful to express c and ϕ in terms of g_{11} and g_{12} as

$$c = \sqrt{\frac{1 - 2g_{11}}{g_{11}^2 + g_{12}^2} + 1}, \tag{269}$$

$$\tan(\phi) = \frac{g_{12}}{g_{12}^2 + (g_{11} - 1)g_{11}}. \tag{270}$$

We now use the preceding results to study the A block of the integer time Green's function. Given that $g_{\ell,A}$ and $\mathcal{G}_{\ell,A}$ have the form given by Eq. (266), the $S_{F,\ell}^{(A)}$ and $S_{0,\ell}^{(A)}$ also have the form of Eq. (263), and therefore $S_{\ell,A}$ also has the form of Eq. (263). We can then use the polar representation to express the following quantities

$$S_{0,\ell}^{(A)} = \mathcal{G}_{\ell;A}^{-1} - 1 = \mathcal{S}(1, \phi_{\ell,0}), \tag{271}$$

$$S_{F,\ell}^{(A)} = g_{\ell;A}^{-1} - 1 = \mathcal{S}(c_\ell, \phi_\ell), \tag{272}$$

$$S_{\ell,A} = (S_{0,\ell}^{(A)})^{-1}S_{F,\ell}^{(A)} = \mathcal{S}(c_\ell, \theta_\ell), \tag{273}$$

allowing for the integer time Dyson equation to be recast as

$$\theta_\ell = \phi_\ell - \phi_{\ell,0}, \tag{274}$$

which yields Eq. (77). Using $g_{11} \to n_\ell$, $g_{12} \to g_{\ell,12}$, $c \to c_\ell$, and $\phi \to \phi_\ell$ in Eqs. (269) and (270), we obtain Eqs. (75) and (76). Using $g_{11} \to 1/2$, $g_{12} \to \mathcal{G}_{\ell,12}$, $c \to 1$, and $\phi \to \phi_{\ell,0}$, we obtain

$$\tan(\phi_{\ell,0}) = \frac{\mathcal{G}_{\ell,12}}{\mathcal{G}_{\ell,12}^2 - \frac{1}{4}}, \tag{275}$$

which yields Eq. (78) by assuming $\phi_{\ell,0} \in [-\pi/2, 0]$. Using Eq. (273), we can rewrite Eq. (222) in terms of c_ℓ, θ_ℓ, and $n_{k\ell}$ as

$$g'_{\ell,A} = \begin{pmatrix} n_\ell & \cot(\theta_\ell)\Delta_\ell \\ -\cot(\theta_\ell)\Delta_\ell & n_\ell \end{pmatrix}, \tag{276}$$

$$g'_{\ell,B} = \sqrt{\frac{1}{c_\ell \sin(\theta_\ell)}} \begin{pmatrix} \cos(\theta_\ell)\mathcal{A}_{>,\ell} \\ \mathcal{A}_{<,\ell} + \sin(\theta_\ell)\mathcal{A}_{>,\ell} \end{pmatrix}, \tag{277}$$

$$g'_{\ell,C} = -\sqrt{\frac{c_\ell}{\sin(\theta_\ell)}} \begin{pmatrix} \cos(\theta_\ell)\mathcal{A}_{<,\ell} \\ \sin(\theta_\ell)\mathcal{A}_{<,\ell} + \mathcal{A}_{>,\ell} \end{pmatrix}^T. \tag{278}$$

The matrix elements in w representation are given by

$$\left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{w;\ell,S}\right]_{11} = \tfrac{1}{2}(\csc(\phi_{\ell,0}) + 1), \qquad (279)$$

$$\left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{w;\ell,S}\right]_{22} = \tfrac{1}{2}(1 - \csc(\phi_{\ell,0})), \qquad (280)$$

$$\left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{w;\ell,S}\right]_{12} = \left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{w;\ell,S}\right]_{21} = 0, \qquad (281)$$

$$\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{w;\ell,S} = \begin{pmatrix} \tfrac{1}{2}\cot(\phi_{\ell,0}) & -\tfrac{1}{2}\csc(\phi_{\ell,0}) \\ -\tfrac{1}{2}\csc(\phi_{\ell,0}) & \tfrac{1}{2}\cot(\phi_{\ell,0}) \end{pmatrix}. \qquad (282)$$

2. Resolving the self-consistency in the A block

A key ingredient of the qubit parametrization is the introduction of the qubit representation, defined by

$$(\hat{\underline{O}})_q = \mathcal{U}^\dagger(\hat{\underline{O}})_w\mathcal{U}, \qquad (283)$$

$$\boldsymbol{\rho} = \mathcal{U}^\dagger ww^T\mathcal{U}, \qquad (284)$$

where $\mathcal{U}$ is a unitary matrix defined as $\mathcal{U} = \mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_{2N_{\mathrm{orb}}}$ and

$$\mathcal{U}_\ell = \begin{pmatrix} \cos(\psi_\ell) & \sin(\psi_\ell) \\ -\sin(\psi_\ell) & \cos(\psi_\ell) \end{pmatrix}, \qquad (285)$$

where ψ_ℓ is a parameter that will be determined by demanding that $\boldsymbol{\rho}$ yields the physical density. This should be contrasted to the w representation, where the local density of ww^T is generally different from the physical density.

We begin by deriving Eq. (73), which determines θ_ℓ for a given n_ℓ, ξ_ℓ, and Δ_ℓ. Using Eq. (283), Eqs. (279)–(282) are transformed to

$$\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{q;\ell,S} = \begin{pmatrix} \tfrac{1}{2}(p_\ell + 1) & \tfrac{q_\ell}{2} \\ \tfrac{q_\ell}{2} & \tfrac{1}{2}(1 - p_\ell) \end{pmatrix}, \qquad (286)$$

$$\left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{q;\ell,S}\right]_{11} = \frac{1}{2}(\cot(\phi_{\ell,0}) + q_\ell), \qquad (287)$$

$$\left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{q;\ell,S}\right]_{22} = \frac{1}{2}(\cot(\phi_{\ell,0}) - q_\ell), \qquad (288)$$

$$\left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{q;\ell,S}\right]_{12} = \left[\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{q;\ell,S}\right]_{21} = -\frac{p_\ell}{2}, \qquad (289)$$

where

$$p_\ell = \cos(2\psi_\ell)\csc(\phi_{\ell,0}), \qquad (290)$$

$$q_\ell = \sin(2\psi_\ell)\csc(\phi_{\ell,0}). \qquad (291)$$

Using the self-consistency condition for the 1,1 and 1,2 entries, we have

$$\mathrm{Tr}\left(\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(1)}\right)_{q;\ell,S}\boldsymbol{\rho}_\ell\right) = n_\ell, \qquad (292)$$

$$\mathrm{Tr}\left(\left(\hat{\underline{a}}_\ell^{\dagger(1)}\hat{\underline{a}}_\ell^{(2)}\right)_{q;\ell,S}\boldsymbol{\rho}_\ell\right) = g_{\ell,12}, \qquad (293)$$

$$\boldsymbol{\rho}_\ell = \mathrm{Tr}_{/\ell}\boldsymbol{\rho} = \begin{pmatrix} 1 - n_\ell & \xi_\ell \\ \xi_\ell & n_\ell \end{pmatrix}. \qquad (294)$$

Equations (292) and (293) reduce to two linear equations in p_ℓ and q_ℓ as

$$\xi_\ell q_\ell - \delta n_\ell p_\ell = \delta n_\ell, \qquad (295)$$

$$\xi_\ell p_\ell + \delta n_\ell q_\ell = \frac{\cot(\phi_{\ell,0})}{2} - g_{\ell,12}, \qquad (296)$$

and p_ℓ and q_ℓ can be determined from Eqs. (88) and (89). Using Eqs. (290) and (291), we obtain

$$p_\ell^2 + q_\ell^2 = \csc^2(\phi_{\ell,0}) = \cot^2(\phi_{\ell,0}) + 1. \qquad (297)$$

Both $\cot(\phi_\ell)$ and $\cot(\phi_{\ell,0})$ can be expressed in terms of $\cot(\theta_\ell)$ as

$$\cot(\phi_\ell) = \Delta_\ell\cot(\theta_\ell) - \frac{\xi_{\ell,0}^2}{\Delta_\ell\cot(\theta_\ell)}, \qquad (298)$$

$$\cot(\phi_{\ell;0}) = \cot(\phi_\ell - \theta_\ell) = \frac{\cot(\phi_\ell)\cot(\theta_\ell) + 1}{\cot(\theta_\ell) - \cot(\phi_\ell)} \qquad (299)$$

$$= -\frac{\cot(\theta_\ell)\left(-\xi_{\ell,0}^2 + \Delta_\ell^2\cot^2(\theta_\ell) + \Delta_\ell\right)}{(\Delta_\ell - 1)\Delta_\ell\cot^2(\theta_\ell) - \xi_{\ell,0}^2}. \qquad (300)$$

Substituting Eqs. (88) and (89) into Eq. (297) and using Eq. (300), we obtain a sixth order equation in $\cot(\theta_\ell)$ which can be factored into the following form

$$\left(\Delta_\ell^2\left(1 - 2\xi_{\ell,0}^2\right)\cot^2(\theta_\ell) + \xi_{\ell,0}^4 + \Delta_\ell^4\cot^4(\theta_\ell)\right)$$
$$\times \left(\cot^2(\theta_\ell)\left(\xi_{\ell,0}^2 + \Delta_\ell^2 - \Delta_\ell\right) - \xi_\ell^2\csc^2(\theta_\ell)\right) = 0. \qquad (301)$$

Notice that the first factor in Eq. (301) is positive, implying that the second factor is zero, and θ_ℓ may be obtained as

$$\cos(\theta_\ell) = \frac{\xi_\ell}{\sqrt{\xi_{\ell,0}^2 + (\Delta_\ell - 1)\Delta_\ell}}, \qquad (302)$$

which yields Eq. (73). Notice that $|\cos(\theta_\ell)| \leqslant 1$, yielding a constraint on ξ_ℓ as

$$|\xi_\ell| \leqslant \sqrt{\xi_{\ell,0}^2 + (\Delta_\ell - 1)\Delta_\ell}. \qquad (303)$$

3. Resolving the B, C, and D blocks

We begin by deriving a subtle symmetry between $\mathcal{G}_B$ and $\mathcal{G}_C$, described by Eq. (83) and Eq. (84), which can be rewritten as

$$\widetilde{\mathcal{G}_C^T} = -\mathcal{G}_B, \qquad (304)$$

where the tilde is defined by the following rules. For a 2×1 matrix m, we have $[\tilde{m}]_{21} = [m]_{11}$ and $[\tilde{m}]_{11} = [m]_{21}$. For a 1×2 matrix m, we have $[\tilde{m}]_{12} = [m]_{11}$ and $[\tilde{m}]_{11} = [m]_{12}$. For a 2×2 matrix m, we have $[\tilde{m}]_{11} = [m]_{22}$, $[\tilde{m}]_{22} = [m]_{11}$, $[\tilde{m}]_{12} = [m]_{21}$, and $[\tilde{m}]_{21} = [m]_{12}$. To prove Eq. (304), we use Eq. (256) to obtain

$$\mathcal{G}_C = g_C g_A^{-1} S_A^{-1}\mathcal{G}_A, \qquad (305)$$

$$\widetilde{\mathcal{G}_C^T} = \widetilde{\mathcal{G}_A^T}\widetilde{\left(S_A^{-1}\right)^T}\widetilde{\left(g_A^{-1}\right)^T}\widetilde{g_C^T} = \mathcal{G}_A S_A^{-1} g_A^{-1}\widetilde{g_C^T}. \qquad (306)$$

Given that $\mathcal{G}_B = \mathcal{G}_A g_A^{-1} g_B$ [see Eq. (260)] and that S_A commutes with g_A (see Sec. VI C 2), Eq. (304) can be proven by verifying that

$$-S_A g_B = \widetilde{g_C^T} \qquad (307)$$

$$= -\sqrt{\frac{c_\ell}{\sin(\theta_\ell)}}\begin{pmatrix} \sin(\theta_\ell)\mathcal{A}_{<,\ell} + \mathcal{A}_{>,\ell} \\ \cos(\theta_\ell)\mathcal{A}_{<,\ell} \end{pmatrix}. \qquad (308)$$

Using Eq. (260), we obtain

$$\mathcal{G}_{\ell,13} = -\mathcal{G}_{\ell,32} = \frac{1}{2(g_{\ell,12}^2 + n_\ell^2)\sqrt{c_\ell}\sin(\theta_\ell)}$$
$$\times ((2g_{\ell,12}\mathcal{G}_{\ell,12} + n_\ell)\cos(\theta_\ell)\mathcal{A}_{>,\ell}$$
$$- (g_{\ell,12} - 2n_\ell\mathcal{G}_{\ell,12})(\mathcal{A}_{<,\ell} + \sin(\theta_\ell)\mathcal{A}_{>,\ell})), \quad (309)$$

$$\mathcal{G}_{\ell,23} = -\mathcal{G}_{\ell,31} = \frac{1}{2(g_{\ell,12}^2 + n_\ell^2)\sqrt{c_\ell}\sin(\theta_\ell)}$$
$$\times ((2g_{\ell,12}\mathcal{G}_{\ell,12} + n_\ell)(\mathcal{A}_{<,\ell} + \sin(\theta_\ell)\mathcal{A}_{>,\ell})$$
$$+ (g_{\ell,12} - 2n_\ell\mathcal{G}_{\ell,12})\cos(\theta_\ell)\mathcal{A}_{>,\ell}). \quad (310)$$

To simplify these two equations, we introduce $\mathcal{I}_\ell$, $\mathcal{J}_\ell$, $\mathcal{A}'_{<,\ell}$, and $\mathcal{A}'_{>,\ell}$ [defined in Eqs. (79)–(82)], which yields Eqs. (83) and (84). To compute $\mathcal{G}_{\ell,33}$, Eqs. (262), (255), and (258) can be used to obtain

$$\mathcal{G}_D = g_D + \mathcal{G}_C(1 - \mathcal{G}_A)^{-1}(g_A - \mathcal{G}_A)\mathcal{G}_A^{-1}\mathcal{G}_B \quad (311)$$
$$= n_\ell + \frac{4}{4\mathcal{G}_{\ell,12}^2 + 1}((\mathcal{G}_{\ell,13}^2 - \mathcal{G}_{\ell,23}^2)(g_{\ell,12} - \mathcal{G}_{\ell,12})$$
$$- 2\delta n_\ell \mathcal{G}_{\ell,13}\mathcal{G}_{\ell,23}). \quad (312)$$

Introducing i_ℓ and j_ℓ, defined in Eqs. (85) and (86), we can simplify Eq. (312) to (87). The matrix elements of $(\underline{\hat{a}}_\ell^{\dagger(3)}\underline{\hat{a}}_\ell^{(3)})_{w,\ell}$ can then be determined as

$$(\underline{\hat{a}}_\ell^{\dagger(3)}\underline{\hat{a}}_\ell^{(3)})_{w,\ell} = \begin{pmatrix} h_z + h_0 & h_x \\ h_x & h_0 - h_z \end{pmatrix}, \quad (313)$$

$$h_0 = \mathcal{G}_{\ell,33} - \tfrac{1}{2}j_\ell\csc(\phi_{\ell,0}), \quad (314)$$

$$h_z = \tfrac{1}{2}i_\ell\csc(\phi_{\ell,0}), \quad (315)$$

$$h_x = \tfrac{1}{2}j_\ell\csc(\phi_{\ell,0}). \quad (316)$$

Equations (283) and (285) may be used to evaluate

$$(\underline{\hat{a}}_\ell^{\dagger(3)}\underline{\hat{a}}_\ell^{(3)})_{q,\ell} = \begin{pmatrix} f_{\ell,z} + f_{\ell,0} & f_{\ell,x} \\ f_{\ell,x} & f_{\ell,0} - f_{\ell,z} \end{pmatrix} \quad (317)$$
$$= f_{\ell,0} + f_{\ell,x}\hat{\sigma}^x + f_{\ell,z}\hat{\sigma}^z, \quad (318)$$

and we define

$$\hat{n}_{\text{eff},\ell} \equiv (\underline{\hat{a}}_\ell^{\dagger(3)}\underline{\hat{a}}_\ell^{(3)})_q \quad (319)$$
$$= I_{2^{\ell-1}} \otimes (\underline{\hat{a}}_\ell^{\dagger(3)}\underline{\hat{a}}_\ell^{(3)})_{q,\ell} \otimes I_{2^{N_{\text{orb}}-\ell}} \quad (320)$$
$$= f_{\ell,0} + f_{\ell,x}\hat{\sigma}_\ell^x + f_{\ell,z}\hat{\sigma}_\ell^z, \quad (321)$$

where $f_{\ell,0}$, $f_{\ell,x}$, and $f_{\ell,z}$ are given in Eqs. (90), (92), and (91). The derivations of Eqs. (71)–(92) are now complete.

D. Examining the qubit energy form in special cases

In this section, we showcase the qubit energy form for the special case of half-filled orbitals. Additionally, we examine how the qubit energy form for $\hat{\rho}_{G3}$ recovers the qubit energy forms for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$.

1. Case of half-filled orbitals

In this section, we examine the case of half-filling with particle-hole symmetry where $n_\ell = 1/2$ and $\mathcal{A}_{<,\ell} = \mathcal{A}_{>,\ell} =$

$\frac{1}{2}A_\ell$. Using the general algorithm given in Eqs. (71)–(92), we provide corresponding results. Starting with $\xi_{\ell,0} = \frac{1}{2}$ and $\delta n_\ell = 0$, we obtain

$$\theta_\ell = \cos^{-1}\left(\frac{2\xi_\ell}{1 - 2\Delta_\ell}\right), \quad (322)$$

$$c_\ell = 1, \quad (323)$$

$$\cot(\phi_{\ell,0}) = \frac{\cot(\theta_\ell)(1 - 4\Delta_\ell(\Delta_\ell\cot^2(\theta_\ell) + 1))}{4(\Delta_\ell - 1)\Delta_\ell\cot^2(\theta_\ell) - 1}, \quad (324)$$

$$\mathcal{G}_{\ell,12} = \frac{\cot\left(\frac{\theta_\ell}{2}\right)(-2\Delta_\ell\cos(\theta_\ell) + \cos(\theta_\ell) - 1)}{2(2\Delta_\ell - 1)\cos(\theta_\ell) - 2}, \quad (325)$$

$$\mathcal{I}_\ell = \frac{4\Delta_\ell^2\cos^2(\theta_\ell) + \sin^2(\theta_\ell)}{4(\cos(\theta_\ell) - 1)((2\Delta_\ell - 1)\cos(\theta_\ell) - 1)}, \quad (326)$$

$$\mathcal{J}_\ell = \frac{4\Delta_\ell^2\cos(\theta_\ell)\cot(\theta_\ell) + \sin(\theta_\ell)}{4(2\Delta_\ell - 1)\cos(\theta_\ell) - 4}, \quad (327)$$

$$\mathcal{A}'_{<,\ell} = \frac{1}{2}A_\ell(\sin(\theta_\ell) + 1), \quad (328)$$

$$\mathcal{A}'_{>,\ell} = \frac{1}{2}A_\ell\cos(\theta_\ell), \quad (329)$$

$$\mathcal{G}_{\ell,13} = \mathcal{G}_{\ell,23} = -\frac{A_\ell(\sin(\theta_\ell) + \cos(\theta_\ell) + 1)}{2\sqrt{\sin(\theta_\ell)}((2\Delta_\ell - 1)\cos(\theta_\ell) - 1)}, \quad (330)$$

$$i_\ell = \frac{2A_\ell^2(\csc(\theta_\ell) + 1)}{4\Delta_\ell^2\cot^2(\theta_\ell) + 1}, \quad (331)$$

and $j_\ell = 0$, $\mathcal{G}_{\ell,33} = \frac{1}{2}$, and

$$p_\ell = \frac{\cot(\theta_\ell)}{2\xi_\ell}\left(\frac{1 - 4\Delta_\ell(\Delta_\ell\cot^2(\theta_\ell) + 1)}{4(\Delta_\ell - 1)\Delta_\ell\cot^2(\theta_\ell) - 1} - 2\Delta_\ell\right), \quad (332)$$

and $q_\ell = 0$, $f_{\ell,0} = \frac{1}{2}$, $f_{\ell,x} = 0$, and

$$f_{\ell,z} = -\frac{A_\ell^2(2\Delta_\ell - 1)\cot(\theta_\ell)(\csc(\theta_\ell) + 1)}{2\xi_\ell(4(\Delta_\ell - 1)\Delta_\ell\cot^2(\theta_\ell) - 1)} \quad (333)$$
$$= \frac{A_\ell^2}{4\xi_\ell^2 - 1}\left(\sqrt{1 - \frac{4\xi_\ell^2}{(1 - 2\Delta_\ell)^2}} + 1\right). \quad (334)$$

It is also useful to define

$$\mathcal{F}_\ell \equiv -2f_{\ell,z} = \frac{2A_\ell^2}{1 - 4\xi_\ell^2}\left(\sqrt{1 - \frac{4\xi_\ell^2}{(1 - 2\Delta_\ell)^2}} + 1\right), \quad (335)$$

yielding

$$(\underline{\hat{n}}_\ell^{(3)})_{q,\ell} = \begin{pmatrix} \frac{1}{2}(1 - \mathcal{F}_\ell) & 0 \\ 0 & \frac{1}{2}(\mathcal{F}_\ell + 1) \end{pmatrix}, \quad (336)$$

$$\hat{n}_{\text{eff},\ell} = (\underline{\hat{n}}_\ell^{(3)})_q = \frac{1}{2} + (\hat{n}_\ell - \frac{1}{2})\mathcal{F}_\ell. \quad (337)$$

Equation (337) can be plugged into Eq. (62) to obtain the total energy. This result will be applied to the multiorbital Hubbard model in Sec. VII.

2. Recovering the qubit energy form for $\hat{\rho}_{G2}$

Given that $\hat{\rho}_{G2}$ is a special case of $\hat{\rho}_{G3}$, it is clear that the former can be obtained by constraining the latter. We

previously demonstrated that restricting the momentum density distribution to be flat in each region and taking $\mathcal{G}_{\ell,12} = 1/2$ within $\hat\rho_{G3}$ will recover $\hat\rho_{G2}$ in the case where $\{n_{kl;0}\}$ corresponds to a pure state [17]. Here we illustrate this fact using the qubit parametrization. We begin by enforcing $\mathcal{G}_{\ell,12} = 1/2$, and Eq. (78) yields $\phi_{\ell,0} = -\pi/2$. Using Eq. (300) and $\phi_{\ell,0} = -\pi/2$, we obtain

$$-\xi_{\ell,0}^2 + \Delta_\ell^2 \cot^2(\theta_\ell) + \Delta_\ell = 0. \quad (338)$$

Using Eqs. (73) and (338), we obtain

$$\Delta_\ell = \xi_{\ell,0}^2 - \xi_\ell^2, \quad (339)$$

$$\cot(\theta_\ell) = \frac{\xi_\ell}{\xi_{\ell,0}^2 - \xi_\ell^2}, \quad (340)$$

$$g_{\ell,12} = \xi_\ell, \quad (341)$$

$$c_\ell = \sqrt{1 - \frac{2\delta n_\ell}{n_\ell^2 + \xi_\ell^2}}, \quad (342)$$

yielding $\mathcal{I}_\ell = \frac{1}{2}(n_\ell + \xi_\ell)$ and $\mathcal{J}_\ell = \frac{1}{2}(\xi_\ell - n_\ell)$. Using the assumption of a flat momentum density distribution, given by $n_{k\ell}|_{k\in<} = \frac{\xi_\ell^2}{n_\ell} + n_\ell$ and $n_{k\ell}|_{k\in>} = \frac{\xi_\ell^2}{n_\ell - 1} + n_\ell$, which yields a quasiparticle weight of

$$n_{k\ell}|_{k\in<} - n_{k\ell}|_{k\in>} = \frac{\xi_\ell^2}{\xi_{\ell,0}^2}, \quad (343)$$

we obtain

$$\mathcal{A}_{<,\ell} = \sqrt{(\xi_{\ell,0}^2 - \xi_\ell^2)(n_\ell^2 + \xi_\ell^2)}, \quad (344)$$

$$\mathcal{A}_{>,\ell} = \sqrt{(\xi_{\ell,0}^2 - \xi_\ell^2)((n_\ell - 1)^2 + \xi_\ell^2)}, \quad (345)$$

$$\mathcal{A}'_{<,\ell} = n_\ell \sqrt{\frac{\xi_{\ell,0}^2 - \xi_\ell^2}{n_\ell^2 + \xi_\ell^2}}, \quad (346)$$

$$\mathcal{A}'_{>,\ell} = \xi_\ell \sqrt{\frac{\xi_{\ell,0}^2 - \xi_\ell^2}{n_\ell^2 + \xi_\ell^2}}, \quad (347)$$

$$\mathcal{G}_{\ell,13} = \mathcal{G}_{\ell,23} = -\mathcal{G}_{\ell,31} = -\mathcal{G}_{\ell,32} = \frac{1}{2}, \quad (348)$$

and $i_\ell = 1$, $j_\ell = 0$, $\mathcal{G}_{\ell,33} = \frac{1}{2}$, $p_\ell = -1$, $q_\ell = 0$, $f_{\ell,0} = \frac{1}{2}$, $f_{\ell,z} = -\frac{1}{2}$, $f_{\ell,x} = 0$, and

$$\left(\hat{\underline{n}}_\ell^{(3)}\right)_{q,\ell} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad (349)$$

yielding $\hat{n}_{\mathrm{eff},\ell} = \hat{n}_\ell$, recovering the qubit energy form obtained from $\hat\rho_{G2}$.

3. Recovering the qubit energy form for $\hat\rho_{B2}$

Given that $\hat\rho_{B2}$ is a special case of $\hat\rho_{G3}$, it is clear that the former can be obtained by constraining the latter. However, it should be noted that in the qubit parametrization, $\hat{K}_2$ in $\hat\rho_{G3}$ is assumed to correspond to a Slater determinant, but $\hat{K}_2$ must be the identity to recover $\hat\rho_{B2}$. Nonetheless, we demonstrate that the qubit energy form of $\hat\rho_{G3}$ can still be constrained to recover the qubit energy form of $\hat\rho_{B2}$. The solution is to restrict $\boldsymbol{\rho}$ to be a diagonal matrix, implying that $\xi_\ell = 0$, yielding $\theta_\ell = \frac{\pi}{2}$, $g_{\ell,12} = 0$, $c_\ell = \frac{1}{n_\ell} - 1$, $\phi_\ell = 0$, $\phi_{\ell,0} = -\frac{\pi}{2}$, $\mathcal{G}_{\ell,12} = \frac{1}{2}$, $\mathcal{I}_\ell =$

$\frac{n_\ell}{2}$, $\mathcal{J}_\ell = -\frac{n_\ell}{2}$, $\mathcal{A}'_{<,\ell} = \mathcal{A}_{<,\ell} + \mathcal{A}_{>,\ell} = A_\ell$, $\mathcal{A}'_{>,\ell} = 0$, $\mathcal{G}_{\ell,13} = \mathcal{G}_{\ell,23} = \frac{A_\ell}{2\xi_{\ell,0}}$, $i_\ell = \frac{A_\ell^2}{\xi_{\ell,0}^2}$, $j_\ell = 0$, $\mathcal{G}_{\ell,33} = n_\ell - \frac{A_\ell^2 \delta n_\ell}{\xi_{\ell,0}^2}$, $p_\ell = -1$, $q_\ell = 0$, $f_{\ell,0} = n_\ell - \frac{A_\ell^2 \delta n_\ell}{\xi_{\ell,0}^2}$, $f_{\ell,z} = -\frac{A_\ell^2}{2\xi_{\ell,0}^2}$, and $f_{\ell,x} = 0$. Finally, we have

$$\left(\hat{\underline{n}}_\ell^{(3)}\right)_{q,\ell} = n_\ell + \frac{A_\ell^2}{\xi_{\ell,0}^2} \begin{pmatrix} -n_\ell & 0 \\ 0 & 1 - n_\ell \end{pmatrix}, \quad (350)$$

$$\hat{n}_{\mathrm{eff},\ell} = n_\ell + \frac{A_\ell^2}{\xi_{\ell,0}^2}(\hat{n}_\ell - n_\ell), \quad (351)$$

recovering the qubit energy form of $\hat\rho_{B2}$. The preceding result demonstrates that when all orbitals have $\xi_\ell = 0$, implying that the system is a Mott insulator, the $\hat\rho_{G3}$ solution can be recovered by $\hat\rho_{B2}$, demonstrating the power of $\hat\rho_{B2}$ in the Mott phase.

VII. APPLICATIONS: MULTIORBITAL HUBBARD MODEL AT HALF FILLING WITH PARTICLE-HOLE SYMMETRY IN $d = \infty$

Here we showcase how to use the qubit energy form for $\hat\rho_{G3}$ to study the multiorbital Hubbard model in $d = \infty$ with density-density interactions. While the qubit energy form can be applied to arbitrary densities, here we study the case of half-filling where the local interaction energy takes a simple form. In Sec. VII A, we explicitly evaluate the qubit energy form for half-filling and demonstrate how to analytically minimize over the momentum density distribution, yielding a final energy form which can straightforwardly be numerically minimized. In Sec. VII B, we demonstrate that Δ_ℓ, defined in Eq. (67), is a key variable for understanding the Mott transition. For a special density-of-states, the minimization can be analytically performed, yielding an analytical relation between the ground-state energy and the Hubbard U via Δ_ℓ. In Sec. VII C, we study the effect of the Hund coupling J on the nature of the Mott transition in the multiorbital Hubbard model.

A. Numerical minimization of the qubit trial energy

1. Evaluating the qubit trial energy

To elucidate the results of the multiorbital Hubbard model, we first begin by considering the single orbital case with density of states $D(\epsilon)$ having particle-hole symmetry. Due to spin symmetry, we omit the spin-orbital index ℓ in ξ_ℓ, Δ_ℓ, and $A_\ell = \mathcal{A}_{<,\ell} + \mathcal{A}_{>\ell}$. The qubit energy form for $\hat\rho_{G3}$ is given as (see Sec. VI D 1 for derivation)

$$E(n(\epsilon), \boldsymbol{\rho}) = 2\int_{-\infty}^{\infty} d\epsilon D(\epsilon)\epsilon n(\epsilon) + U\langle \hat{n}_{\mathrm{eff},\uparrow}\hat{n}_{\mathrm{eff},\downarrow}\rangle_{\boldsymbol{\rho}}, \quad (352)$$

$$\hat{n}_{\mathrm{eff},\ell} = \frac{1}{2} + \mathcal{F}(\Delta, A, \xi)\left(\hat{n}_\ell - \frac{1}{2}\right), \quad (353)$$

$$\mathcal{F}(\Delta, A, \xi) = \frac{2A^2}{1 - 4\xi^2}\left(\sqrt{1 - \frac{4\xi^2}{(1 - 2\Delta)^2}} + 1\right), \quad (354)$$

$$A = \int_{-\infty}^{\infty} d\epsilon D(\epsilon)\sqrt{n(\epsilon)(1 - n(\epsilon))}, \quad (355)$$

$$\Delta = \int_0^\infty d\epsilon D(\epsilon) n(\epsilon), \tag{356}$$

$$\xi = \frac{1}{2} \langle \hat{\sigma}_\ell^x \rangle_\rho, \tag{357}$$

where particle-hole symmetry implies $n(\epsilon) + n(-\epsilon) = 1$, and to ensure that $\mathcal{F}(\Delta, A, \xi)$ is real, ξ must satisfy the following constraint:

$$|\xi| \leqslant \frac{1}{2} - \Delta. \tag{358}$$

The effective many-body density matrix ρ can be encoded as $\rho = |\Psi\rangle\langle\Psi|$ where $|\Psi\rangle$ can be parametrized using a single parameter $\mathcal{D} \in [0, 1/2]$ as

$$|\Psi\rangle = \left(\sqrt{\mathcal{D}} \quad \sqrt{\frac{1}{2} - \mathcal{D}} \quad \sqrt{\frac{1}{2} - \mathcal{D}} \quad \sqrt{\mathcal{D}} \right)^T. \tag{359}$$

This parametrization ensures that $n_\ell = 1/2$ and ξ can be determined as

$$\xi = 2\sqrt{\left(\frac{1}{2} - \mathcal{D}\right)\mathcal{D}}. \tag{360}$$

The double occupancy is given as

$$d(\Delta, A, \mathcal{D}) = \langle \hat{n}_{\text{eff},\uparrow} \hat{n}_{\text{eff},\downarrow} \rangle_\rho = \left(\mathcal{D} - \frac{1}{4}\right)\mathcal{F}^2 + \frac{1}{4} \tag{361}$$

$$= \frac{1}{4}\left(\frac{4A^4(1 - 2\Delta + h(\Delta, \mathcal{D}))^2}{(1 - 2\Delta)^2(4\mathcal{D} - 1)^3} + 1\right), \tag{362}$$

$$h(\Delta, \mathcal{D}) = \sqrt{8\mathcal{D}(2\mathcal{D} - 1) + (1 - 2\Delta)^2} \tag{363}$$

and to have a real $d(\Delta, A, \mathcal{D})$, the $h(\Delta, \mathcal{D})$ must be real which requires $\mathcal{D}$ to satisfy the following constraint:

$$\mathcal{D} \in [0, \mathcal{D}_-] \cup \left[\mathcal{D}_+, \frac{1}{2}\right], \tag{364}$$

$$\mathcal{D}_\pm = \frac{1}{4} \pm \frac{1}{2}\sqrt{(1 - \Delta)\Delta}. \tag{365}$$

Finally, the trial energy for the single orbital case is a function only of $n(\epsilon)$ and $\mathcal{D}$, given as

$$E(n(\epsilon), \mathcal{D}) = 2\int_{-\infty}^\infty d\epsilon D(\epsilon) n(\epsilon) + U d(\Delta, A, \mathcal{D}). \tag{366}$$

We now proceed to the multiorbital case, with a local interaction given as

$$\hat{H}_{\text{loc}} = \sum_{\ell < \ell'} U_{\ell\ell'} \delta\hat{n}_\ell \delta\hat{n}_{\ell'}, \tag{367}$$

where $\delta\hat{n}_\ell = \hat{n}_\ell - \frac{1}{2}$. The qubit trial energy is given as

$$E(\{n_\ell(\epsilon)\}, \rho) = \sum_\ell \int_{-\infty}^\infty d\epsilon D_\ell(\epsilon) n_\ell(\epsilon) + \sum_{\ell < \ell'} \mathcal{F}(\Delta_\ell, A_\ell, \xi_\ell)$$
$$\times \mathcal{F}(\Delta_{\ell'}, A_{\ell'}, \xi_{\ell'}) U_{\ell\ell'} \langle \delta\hat{n}_\ell \delta\hat{n}_{\ell'} \rangle_\rho, \tag{368}$$

where ρ is a pure state that is restricted to $\langle \hat{n}_\ell \rangle_\rho = \frac{1}{2}$, particle-hole symmetry implies $n_\ell(\epsilon) + n_\ell(-\epsilon) = 1$, and ξ_ℓ must satisfy the following constraint:

$$|\xi_\ell| \leqslant \frac{1}{2} - \Delta_\ell. \tag{369}$$

The expression for $\mathcal{F}(\Delta_\ell, A_\ell, \xi_\ell)$ is given in Eq. (354). The most straightforward approach to satisfying Eq. (369) is to

first choose ρ, and then Eq. (369) becomes a linear constraint on $n_\ell(\epsilon)$ (see Sec. III for further discussion).

2. Numerical minimization of the qubit trial energy

We now proceed to minimizing the qubit trial energy. As before, we first focus on the single orbital model for clarity, and then consider the multiorbital case. The general numerical minimization has been described in Sec. III, which consists of two steps: First, the momentum density distribution is partially optimized under the constraint of n_ℓ, Δ_ℓ, $\mathcal{A}_{<\ell}$, and $\mathcal{A}_{>\ell}$, or through four Lagrange multiplier $a_{<\ell}$, $a_{>\ell}$, $b_{<\ell}$, and $b_{>\ell}$. Second, one needs to minimize over the remaining $2^{2N_{\text{orb}}} + 3 \times 2N_{\text{orb}} - 1$ independent variational parameters subjected to the inequality constraint given by Eq. (70). For the half-filled, particle-hole symmetric case, there are several simplifications. First, the optimized momentum density distribution now only depends on Δ_ℓ and $A_\ell = \mathcal{A}_{<\ell} + \mathcal{A}_{>\ell}$, or just two Lagrange multipliers $a = a_{<\sigma} = -a_{>\sigma}$ and $b = b_{>\sigma} = b_{<\sigma}$, and the partially optimized momentum density distribution is given as

$$n(\epsilon, a, b) = \begin{cases} \frac{1}{2}\left(1 - \frac{a+\epsilon}{\sqrt{(a+\epsilon)^2 + b^2}}\right) & \epsilon > 0 \\ \frac{1}{2}\left(1 + \frac{a-\epsilon}{\sqrt{(a-\epsilon)^2 + b^2}}\right) & \epsilon < 0 \end{cases}, \tag{370}$$

where the Lagrange multipliers a and b are determined by Δ and A via inverting the following relation

$$A(a, b) = \int_{-\infty}^\infty d\epsilon D(\epsilon)\sqrt{n(\epsilon, a, b)(1 - n(\epsilon, a, b))}, \tag{371}$$

$$\Delta(a, b) = \int_0^\infty d\epsilon D(\epsilon) n(\epsilon, a, b), \tag{372}$$

which yields $a(\Delta, A)$ and $b(\Delta, A)$. To better appreciate the flexibility of $n(\epsilon, a, b)$, it is useful to examine the behavior for various choices of a and b (see Fig. 1). For $a > 0$ and $b > 0$, the distribution corresponds to a Fermi liquid, given that the discontinuity of $n(\epsilon, a, b)$ at $\epsilon = 0$ yields a quasiparticle weight Z of

$$Z = \frac{a}{\sqrt{a^2 + b^2}}, \tag{373}$$

which will be between zero and one. For the special case where $a \to \infty$ and $b \to \infty$ with a/b remaining finite, the distribution recovers a flat distribution, which is obtained for an optimized $\hat{\rho}_{G2}$. Finally, when $a = 0$ and $b > 0$, the system is in the Mott phase where $Z = 0$.

The second simplification for half-filling and particle-hole symmetry is that the number of independent parameters is $2^{2N_{\text{orb}}} + 2N_{\text{orb}} - 1$, given that $n_\ell = 1/2$ and $\mathcal{A}_{<\ell} = \mathcal{A}_{>\ell}$. If we have further symmetry between different spin orbitals, the number of independent parameters can further be reduced. For example, for the single-orbital case with spin symmetry, we have $A_\ell = A$ and $\Delta_\ell = \Delta$, and therefore the number of independent parameters is $2^2 + 2 - 1 - 2 = 3$. As discussed in Sec. III, there are two possible strategies. The first strategy starts from a, b, which determines $n(\epsilon)$, and ρ is specified via $\mathcal{D}$, which will have a range given by Eq. (364).

Mathematically, this qubit trial energy is given by

$$E(a, b, \mathcal{D}) = 2 \int_{-\infty}^{\infty} d\epsilon D(\epsilon)\epsilon n(\epsilon, a, b)$$
$$+ U d(\Delta(a, b), A(a, b), \mathcal{D}). \quad (374)$$

This strategy allows the energy to be explicitly written in terms of the variational parameters, and thus it is straightforward to compute the derivatives via automatic differentiation. For the multiorbital case, the constraint between $\boldsymbol{\rho}$ and $\{\Delta_\ell\}$ is given by Eq. (369), which is not straightforward to implement for a given $\{\Delta_\ell\}$. Therefore, it is useful to pursue a second strategy.

We begin by illustrating the second strategy in the single-orbital case. This strategy begins by specifying $\boldsymbol{\rho}$ via $\mathcal{D}$, then the range of Δ can be determined from the constraint given by Eq. (358), and then Δ and b can be specified, which determines a via inverting Eq. (372), denoted as $a(\Delta, b)$. Mathematically, the resulting qubit trial energy is given by

$$E(\Delta, b, \mathcal{D}) = 2 \int_{-\infty}^{\infty} d\epsilon D(\epsilon)\epsilon n(\epsilon, a(\Delta, b), b)$$
$$+ U d(\Delta, A(a(\Delta, b), b), \mathcal{D}). \quad (375)$$

For the multiorbital case, the qubit trial energy is given as

$$E(\{\Delta_\ell\}, \{b_\ell\}, \boldsymbol{\rho}) = \sum_\ell \int_{-\infty}^{\infty} d\epsilon D_\ell(\epsilon)\epsilon n(\epsilon, a_\ell(\Delta_\ell, b_\ell), b_\ell)$$
$$+ \sum_{\ell < \ell'} \mathcal{F}_\ell \mathcal{F}_{\ell'} U_{\ell\ell'} \langle \delta\hat{n}_\ell \delta\hat{n}_{\ell'} \rangle_{\boldsymbol{\rho}}, \quad (376)$$

$$\mathcal{F}_\ell = \mathcal{F}\left(\Delta_\ell, A_\ell(a_\ell(\Delta_\ell, b_\ell), b_\ell), \frac{1}{2}\langle \hat{\sigma}_\ell^x \rangle_{\boldsymbol{\rho}}\right), \quad (377)$$

where $a_\ell(\Delta_\ell, b_\ell)$ and $A_\ell(a_\ell, b_\ell)$ are the multiorbital versions of $a(\Delta, b)$ and $A(a, b)$. The second strategy allows one to automatically implement all of the constraints in the multiorbital case, and the only downside is that $a_\ell(\Delta_\ell, b_\ell)$ must be numerically evaluated, though this is a trivial task.

B. Understanding the Mott transition in the single orbital model

In the preceding section, the resulting qubit trial energy for the single orbital model at half-filling is either parametrized by a, b, and $\mathcal{D}$ [i.e., Eq. (374)], or Δ, b and $\mathcal{D}$ [i.e., Eq. (375)], which both allow for straightforward numerical minimization for a given U and $D(\epsilon)$. In this section, we explore a different perspective based on the one-body reduced density matrix viewpoint (see Ref. [18]), which provides a clearer understanding of the Mott transition. The first step begins with the parametrization of the qubit trial energy using Δ, A, and $\mathcal{D}$, where the local interaction energy is $U d(\Delta, A, \mathcal{D})$, where $d(\Delta, A, \mathcal{D})$ is defined in Eq. (362). It should be noted that $d(\Delta, A, \mathcal{D})$ has the form $1/4 + A^4 f(\Delta, \mathcal{D})$, and therefore we can optimize $d(\Delta, A, \mathcal{D})$ with respect to $\mathcal{D}$ for fixed Δ and A, yielding an optimized $\mathcal{D}$ as a function of Δ, denoted as $\mathcal{D}(\Delta)$. Therefore, the interaction energy is purely a function of Δ and A, which can be viewed as a one-body reduced density matrix functional, yielding a simple picture of the Mott transition. Two key points should be appreciated. First, $\mathcal{D}(\Delta)$ has a critical value of Δ, denoted Δ_c, and $\mathcal{D}(\Delta) = 0$ for $\Delta \geq \Delta_c$, which corresponds to the Mott insulator. The second
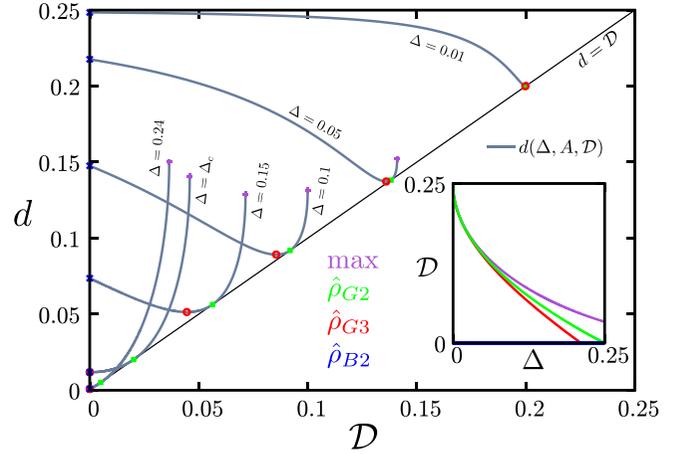

FIG. 2. Plot of the double occupancy d given in Eq. (362) as a function of $\mathcal{D}$ for various Δ and $A = \sqrt{(2 - 4\Delta)\Delta}$. The points labeled $\hat{\rho}_{G3}$ provide the optimized value for $\mathcal{D}$ given by Eq. (383), while the $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$ points provide the $\mathcal{D}$ values for the $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$ given by Eq. (386) and $\mathcal{D} = 0$, respectively. The purple points labeled max provide the maximum value for $\mathcal{D}$, given as $\mathcal{D}_-$ defined in Eq. (365). The inset plots $\mathcal{D}$ as a function of Δ for $\hat{\rho}_{G3}$, $\hat{\rho}_{G2}$, and $\hat{\rho}_{B2}$, in addition to the maximum value.

point is that Δ_c is independent of $D(\epsilon)$. While the total energy can be obtained by numerically minimizing over Δ and A, further insight can be obtained by expressing U and the total energy as functions of Δ.

1. Qubit trial energy in terms of Δ and A

As outlined in Sec. VII A 2, the qubit trial energy can be parametrized as

$$E(\Delta, A, \mathcal{D}) = K(\Delta, A) + U d(\Delta, A, \mathcal{D}), \quad (378)$$

$$K(\Delta, A) = 2 \int_{-\infty}^{\infty} d\epsilon D(\epsilon)\epsilon n(\epsilon, a(\Delta, A), b(\Delta, A)). \quad (379)$$

For simplicity, we study the case where $U \geqslant 0$. We now consider how to minimize the total energy over $\mathcal{D}$, which amounts to minimizing $d(\Delta, A, \mathcal{D})$ over $\mathcal{D}$. To visualize the minimization, we plot d as a function of $\mathcal{D}$ for a given Δ and A. Given that A will not influence the minimization over $\mathcal{D}$, we choose $A = \sqrt{(2 - 4\Delta)\Delta}$ from a flat momentum density distribution (see Fig. 2). For a given Δ curve, there are four distinct sets of points denoted as $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, $\hat{\rho}_{G3}$, and max, where the $\hat{\rho}_{G3}$ points provide the optimized values for $\mathcal{D}$, the $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$ points provide the $\mathcal{D}$ values for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$, respectively, and a max point provides the maximum value for $\mathcal{D}$ given by $\mathcal{D}_-$ in Eq. (365). For small values of Δ, the optimized value of $\mathcal{D}$ is nonzero, and $\mathcal{D}$ monotonically decreases with increasing Δ. For Δ larger than some critical value, the optimized value for $\mathcal{D}$ is zero. Having obtained a graphical understanding of this function, we proceed to mathematically minimize the d over $\mathcal{D}$ for a given Δ and A, which is a constrained minimization given that $\mathcal{D} \in [0, \mathcal{D}_-]$. Solving $\partial d / \partial \mathcal{D} = 0$ yields

$$\mathcal{D}^\star(\Delta) = \frac{1}{4} - \frac{1}{4}\sqrt{6(\Delta - 1)\Delta + \left(3\Delta - \frac{3}{2}\right)h_1 + \frac{9}{2}}, \quad (380)$$

$$h_1 = \sqrt{4(\Delta - 1)\Delta + 9}. \quad (381)$$

When $\mathcal{D}^\star(\Delta) \in [0, \mathcal{D}_-]$, then $\mathcal{D}^\star(\Delta)$ yields the optimized value for $\mathcal{D}$, and otherwise the optimized value is given by the minimum value for the boundary points. It is useful to solve $\mathcal{D}^\star(\Delta) = 0$, yielding

$$\Delta_c = \tfrac{1}{6}(3 - \sqrt{3}) \approx 0.211325. \tag{382}$$

Therefore, the optimized value for $\mathcal{D}$ is

$$\mathcal{D}(\Delta) = \begin{cases} \mathcal{D}^\star(\Delta), & \Delta < \Delta_c, \\ 0, & \Delta \geqslant \Delta_c, \end{cases} \tag{383}$$

having two distinct regimes as a function of Δ (see inset of Fig. 2). Finally, the physical double occupancy can be written as a function of Δ and A as

$$d(\Delta, A) = \begin{cases} d(\Delta, A, \mathcal{D}^\star(\Delta)), & \Delta < \Delta_c, \\ \tfrac{1}{4} - 4A^4, & \Delta \geqslant \Delta_c. \end{cases} \tag{384}$$

Therefore, the total trial energy can be written purely in terms of Δ and A, given as

$$E(\Delta, A) = K(\Delta, A) + U d(\Delta, A). \tag{385}$$

To find the ground state for a given $D(\epsilon)$ and U, it is necessary to minimize over Δ and A, which cannot be performed analytically in general. However, it is important to appreciate that the optimized value of Δ is sufficient to determine if the system is in the Mott phase. For $\Delta = 0$, the ansatz corresponds to the Hartree-Fock wave function, while the maximum value of $\Delta = 1/4$ corresponds to a collection of isolated atoms, and the Mott transition occurs when $\Delta = \Delta_c$, before the system becomes a collection of atoms. We now verify that the system is indeed a Mott insulator for $\Delta > \Delta_c$. First, the local interaction energy is independent of Δ for $\Delta > \Delta_c$ [see Eq. (383)], dictating that $a = 0$ for the optimized Δ. Second, Eq. (373) dictates that the quasiparticle weight is zero when $a = 0$, implying a Mott insulating state. Alternatively, when $\Delta < \Delta_c$, the ansatz describes a metallic phase.

Following the results of Secs. VI D 2 and VI D 3, we illustrate how the qubit energy form for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$ can be recovered from $\hat{\rho}_{G3}$ when properly restricting the variational parameters. We begin with $\hat{\rho}_{B2}$, which can be viewed as a continuation of the Mott phase for $\hat{\rho}_{G3}$, where $\mathcal{D} = 0$ and d is given by $\tfrac{1}{4} - 4A^4$ (see Fig. 2). Alternatively, the $\hat{\rho}_{G2}$ is characterized by a flat momentum density distribution determined by Δ where d and $\mathcal{D}$ are identical, given as

$$d_{G2} = \mathcal{D}_{G2} = \tfrac{1}{4}(1 - 2\sqrt{\Delta}). \tag{386}$$

It is well known that for $\hat{\rho}_{G2}$ the Mott transition occurs when the system becomes a collection of isolated atoms, meaning that the transition happens for $\Delta = 1/4$.

2. Solution for the two-peak density of states

In Sec. VII B 1, we demonstrated that the interaction energy can be written analytically in terms of Δ and A, but the kinetic energy must be numerically determined in terms of Δ and A. While the latter is a trivial numerical problem, it still precludes a completely analytic solution for the energy in terms of U. Therefore, we introduce the two-peak density of states, where the kinetic energy can be analytically evaluated in terms of Δ and A, allowing for an analytical relation between the total energy and the Hubbard $U/|K_0|$ in all three
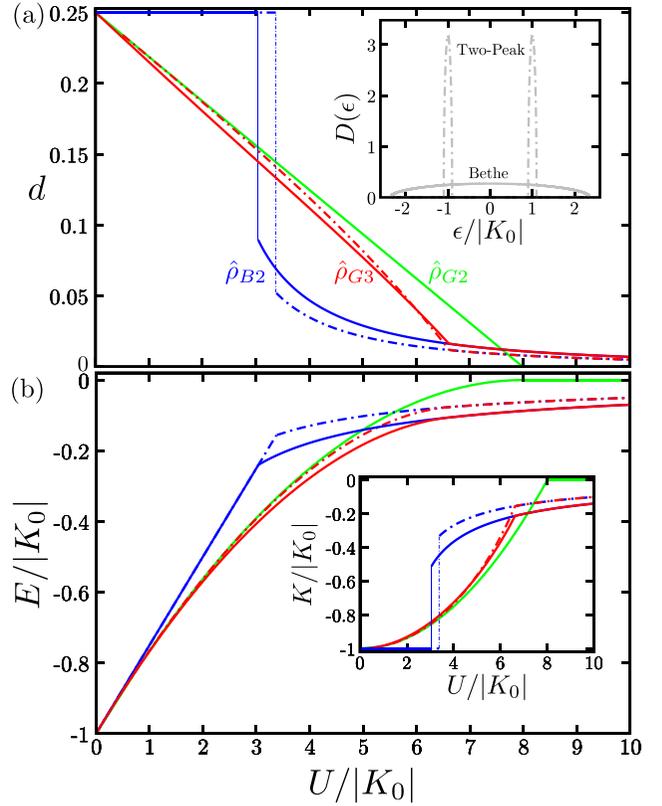

FIG. 3. Plot of the double occupancy (a) and the total energy $E/|K_0|$ (b) vs $U/|K_0|$ for the single-band Hubbard model in $d = \infty$ at half-filling for the Bethe lattice and the two-peak model [see Eq. (387) and inset of panel (a)]. Both models are solved using $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, and $\hat{\rho}_{G3}$. The inset of panel (b) plots $K/|K_0|$ vs $U/|K_0|$.

ansatz, where $K_0 < 0$ is the noninteracting kinetic energy for the lattice. Specifically, the two-peak density of states is given by

$$D(\epsilon) = \tfrac{1}{2}(\delta(\epsilon - K_0) + \delta(\epsilon + K_0)). \tag{387}$$

While this density of states is somewhat unphysical, the results retain all of the qualitative features of the Bethe lattice, in addition to being quantitatively very similar assuming the same K_0 (see Fig. 3). As expected, the $\hat{\rho}_{G2}$ solution is identical for both density of states, given that $\hat{\rho}_{G2}$ only depends on K_0, independent of the details of $D(\epsilon)$. Interestingly, $\hat{\rho}_{G3}$ has a very similar critical value of $U/|K_0|$ for both density of states, with small differences differences in the double occupancy for a given $U/|K_0|$.

We now proceed to analytically evaluate the total trial energy as a function of Δ for the two-peak model. Using Eqs. (387), (356), and (355), the kinetic energy and A are determined as $K = (1 - 4\Delta)K_0$ and $A = \sqrt{(2 - 4\Delta)\Delta}$, yielding the total trial energy in terms of the single variational parameter Δ, given as

$$E(\Delta) = (1 - 4\Delta)K_0 + U d(\Delta), \tag{388}$$

where

$$d(\Delta) = \begin{cases} d_1(\Delta), & \Delta \leqslant \Delta_c, \\ \tfrac{1}{4} - 16(1 - 2\Delta)^2\Delta^2, & \Delta > \Delta_c, \end{cases} \tag{389}$$
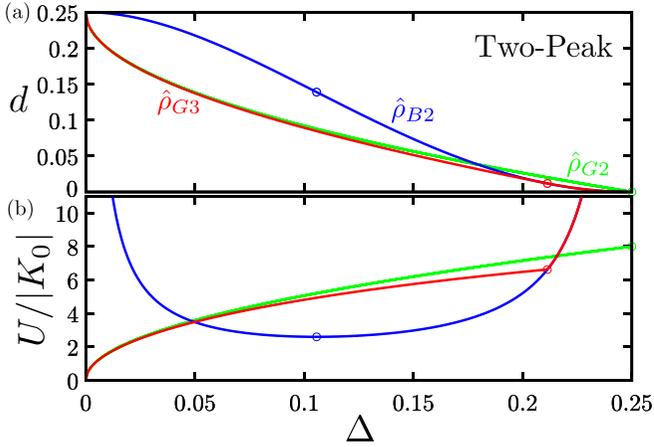
FIG. 4. Plots of the double occupancy d (a) and $U/|K_0|$ (b) as functions of Δ for the two-peak model in $d = \infty$ solved using $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, and $\hat{\rho}_{G3}$.

and $d_1(\Delta)$ is defined in Eq. (97). The $d(\Delta)$ is plotted in Fig. 4(a). While for a given U the energy cannot be analytically minimized over Δ, it is straightforward to find U for a given value of Δ that satisfies $dE(\Delta)/d\Delta = 0$, given as

$$U(\Delta) = -4|K_0|\left(\frac{d}{d\Delta}d(\Delta)\right)^{-1}, \qquad (390)$$

and $U/|K_0|$ is plotted as a function of Δ in Fig. 4(b). The critical value of U and double occupancy at the Mott transition, where $\Delta = \Delta_c$, are given as

$$U_c/|K_0| = \tfrac{3}{8}(9 + 5\sqrt{3}) \approx 6.6226, \qquad (391)$$

$$d_c = \tfrac{1}{36}(32\sqrt{3} - 55) \approx 0.0118229. \qquad (392)$$

These values are the same from both the metallic and insulating sides of the Mott transition, which can be seen in Fig. 4, confirming that the transition is continuous.

Corresponding equations for d and $U/|K_0|$ as functions of Δ can be obtained for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$ by substituting the corresponding $d(\Delta)$ relations into Eq. (390) (see Fig. 4 for plots). For $\hat{\rho}_{G2}$, we have

$$d_{G2}(\Delta) = \tfrac{1}{4}(1 - 2\sqrt{\Delta}), \qquad (393)$$

$$U_{G2}(\Delta) = 16|K_0|\sqrt{\Delta}, \qquad (394)$$

which recovers the Gutzwiller approximation, and the Mott transition occurs at $\Delta = 1/4$, where $U_{c,G2}/|K_0| = 8$ and

$d_{c,G2} = 0$. For $\hat{\rho}_{B2}$, there are some subtleties to consider. For a given U, there are several candidate values of Δ: the value given by the saddle-point Eq. (390), which yields

$$d_{B2}(\Delta) = \frac{1}{4} - 16(1 - 2\Delta)^2\Delta^2, \qquad (395)$$

$$U_{B2}(\Delta) = \frac{|K_0|}{64\Delta^3 - 48\Delta^2 + 8\Delta}, \qquad (396)$$

and the boundary values of $\Delta = 0$ and $\Delta = 1/2$. The total energy must be used to evaluate these candidate values of Δ and select the global minimum. It should be noted that $\Delta=0$ recovers the Hartree-Fock solution. There are two critical points to consider: the local stability of $\hat{\rho}_{B2}$ and a transition from a saddle-point solution to the Hartree-Fock solution. The local stability of the $\hat{\rho}_{B2}$ is determined by the minimal value of U in Eq. (396), which is given by

$$U_{c,B2}/|K_0| = \frac{3\sqrt{3}}{2} \approx 2.59808, \qquad (397)$$

$$d_{c,B2} = \frac{5}{36} \approx 0.138889, \qquad (398)$$

$$\Delta_{c,B2} = \frac{1}{12}(3 - \sqrt{3}) \approx 0.105662. \qquad (399)$$

For any $U > U_{c,B2}$, there exists a locally stable $\hat{\rho}_{B2}$ solution. However, one should compare the energy of the saddle-point solution to the Hartree-Fock solution, which yields another critical value of $U'_{c,B2}/|K_0| = 3.375$. For $U > U'_{c,B2}$, the saddle-point solution is the global minimum, while for $U < U'_{c,B2}$ the Hartree-Fock solution is the global minimum. We summarize all the critical points for $\hat{\rho}_{G3}$, $\hat{\rho}_{G2}$, and $\hat{\rho}_{B2}$ in Table II.

3. Solution for a general density of states

In Sec. VII B 2, we demonstrated that the qubit trial energy can be written solely in terms of Δ for the case of a two-peak density of states. Here we extend this strategy for a general density of states, allowing for the optimized value of A to be evaluated as a function of Δ, which is completely independent of U. Therefore, this approach is particularly useful for analyzing the Mott transition. We begin by rewriting the qubit trial energy from Eq. (385) using the result of Eq. (389), yielding

$$E(\Delta, A) = K(\Delta, A) + \tfrac{1}{4}U + UA^4\tilde{O}(\Delta), \qquad (400)$$

TABLE II. Values of Δ, U, and d at various critical points for the two-peak Hubbard model solved using $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, and $\hat{\rho}_{G3}$. For $\hat{\rho}_{B2}$, the critical point for local stability [$\hat{\rho}_{B2}$ (L)] and the transition to the Hartree-Fock solution [$\hat{\rho}_{B2}$ (HF)] are provided. For $\hat{\rho}_{G2}$ and $\hat{\rho}_{G3}$, the critical values for the Mott transition are provided.

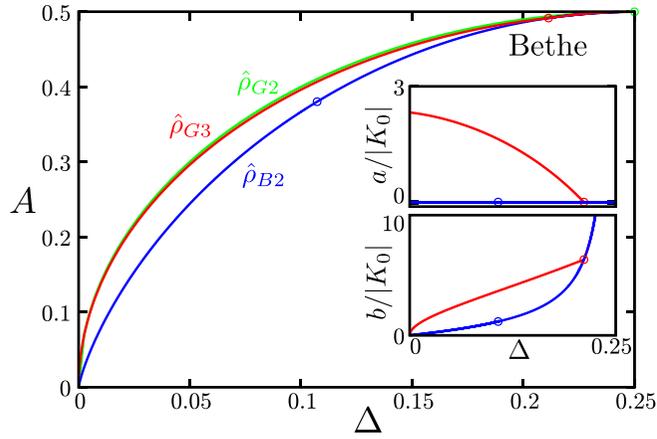|	Δ	$U/|K_0|$	d
$\hat{\rho}_{G3}$	$\frac{1}{6}(3 - \sqrt{3}) \approx 0.211325$	$\frac{3}{8}(9 + 5\sqrt{3}) \approx 6.6226$	$\frac{1}{36}(32\sqrt{3} - 55) \approx 0.0118229$
$\hat{\rho}_{G2}$	$\frac{1}{4}$	8	0
$\hat{\rho}_{B2}$ (L)	$\frac{1}{12}(3 - \sqrt{3}) \approx 0.105662$	$\frac{3\sqrt{3}}{2} \approx 2.59808$	$\frac{5}{36} \approx 0.138889$
$\hat{\rho}_{B2}$(HF)	$\frac{1}{6}$	$\frac{27}{8} = 3.375$	$\frac{17}{324} \approx 0.0524691$

FIG. 5. Parameters A, a, and b as a function of Δ for the single orbital Hubbard model on the Bethe lattice in $d = \infty$ at half-filling, where K_0 is the noninteracting kinetic energy.

where $\tilde{O}(\Delta)$ is given in Eq. (96). We proceed by constructing the saddle-point equations of Δ and A for a given U, yielding the following two equations:

$$\frac{\partial K}{\partial \Delta} = -4a = -UA^4 \frac{\partial \tilde{O}(\Delta)}{\partial \Delta}, \tag{401}$$

$$\frac{\partial K}{\partial A} = 2b = -4UA^3 \tilde{O}(\Delta), \tag{402}$$

where a and b are the Lagrange multipliers from Eq. (370). A practical approach for solving the two preceding equations for a given U is to express A and Δ in terms of a and b, denoted as $A(a, b)$ and $\Delta(a, b)$, and then solve for a and b. However, we take an alternative approach, as our goal is to determine the optimized value of A for a given Δ. Therefore, we proceed by dividing Eq. (401) by Eq. (402), which yields

$$-\frac{8a}{A(a, b)b} = \frac{d \ln |\tilde{O}(\Delta)|}{d\Delta}. \tag{403}$$

Moreover, a and b are required to yield the given value of Δ, such that

$$\Delta(a, b) = \Delta. \tag{404}$$

Simultaneously solving Eqs. (403) and (404) yields a and b as functions of Δ, and therefore all quantities that depend on $n(\epsilon)$, including A and K, are now functions purely of Δ.

Subsequently, the double occupancy and U can be determined as a function of Δ as

$$d(\Delta) = \frac{1}{4} + UA^4(\Delta)\tilde{O}(\Delta), \tag{405}$$

$$U(\Delta) = -\frac{b(\Delta)}{2A^3(\Delta)\tilde{O}(\Delta)}. \tag{406}$$

Alternatively, the U can also be expressed as

$$U(\Delta) = -\frac{dK}{d\Delta}\left(\frac{d}{d\Delta}d(\Delta)\right)^{-1}, \tag{407}$$

which will recover Eq. (390) when evaluating the two-peak model. For the case of the Bethe lattice, we plot A, a, and b as functions of Δ in Fig. 5, and the critical values of all quantities are listed in Table III.

C. Understanding the effect of Hund's coupling in the multiorbital Hubbard model

Here we generalize the treatment from Sec. VII B to the multiorbital case including the Hund coupling J. The key difference is that in the multiorbital case, one cannot analytically minimize over the local variational parameters, though these parameters can easily be numerically minimized as a function of Δ for a given J/U. The remaining procedure closely follows the single orbital case.

Consider the multiorbital Hubbard model defined in Eq. (94). The qubit trial energy can be written as

$$E(\Delta, A, \boldsymbol{\rho}) = K(\Delta, A) + E_{\text{loc}}(\Delta, A, \boldsymbol{\rho}), \tag{408}$$

$$K(\Delta, A) = 2N_{\text{orb}} \int_{-\infty}^{\infty} d\epsilon D(\epsilon)\epsilon n(\epsilon, a(\Delta, A), b(\Delta, A)), \tag{409}$$

$$E_{\text{loc}}(\Delta, A, \boldsymbol{\rho}) = U\mathcal{F}^2(\Delta, A, \xi(\boldsymbol{\rho}))\langle \hat{O}\rangle_{\boldsymbol{\rho}}, \tag{410}$$

where $\mathcal{F}(\Delta, A, \xi)$ is defined in Eq. (354), $\xi(\boldsymbol{\rho}) = \frac{1}{2}\langle \hat{\sigma}_\ell^x \rangle_{\boldsymbol{\rho}}$, and

$$\hat{O} = \hat{O}_1 + \left(1 - 2\frac{J}{U}\right)\hat{O}_2 + \left(1 - 3\frac{J}{U}\right)\hat{O}_3, \tag{411}$$

where $\hat{O}_i$ are defined in Eq. (94). Having defined the qubit trial energy, we proceed to obtain the solutions for the two-peak density of states and the Bethe lattice using the Δ parametrization. We first rewrite the local interaction energy as

$$E_{\text{loc}}(\Delta, A, \boldsymbol{\rho}) = UA^4 \tilde{\mathcal{F}}^2(\Delta, \xi(\boldsymbol{\rho}))\langle \hat{O}\rangle_{\boldsymbol{\rho}}, \tag{412}$$

TABLE III. Values of Δ, U, and d at various critical points for the single orbital Hubbard model on the Bethe lattice in $d = \infty$ solved using $\hat{\rho}_{G2}$, $\hat{\rho}_{B2}$, and $\hat{\rho}_{G3}$. For $\hat{\rho}_{B2}$, the critical point for local stability [$\hat{\rho}_{B2}$ (L)] and the transition to the Hartree-Fock solution [$\hat{\rho}_{B2}$ (HF)] are provided. For $\hat{\rho}_{G2}$ and $\hat{\rho}_{G3}$, the critical values for the Mott transition are provided.

	Δ	$U/|K_0|$	U/t	d	A
$\hat{\rho}_{G3}$	$\frac{1}{6}(3 - \sqrt{3}) \approx 0.211325$	6.61836	5.61784	0.0162523	0.491668
$\hat{\rho}_{G2}$	$\frac{1}{4}$	8	6.79061	0	$\frac{1}{2}$
$\hat{\rho}_{B2}$ (L)	0.107273	2.60206	2.2087	0.166367	0.380259
$\hat{\rho}_{B2}$ (HF)	0.153687	3.04071	2.58103	0.089871	0.447304

where

$$\tilde{\mathcal{F}}(\Delta, \xi) = \frac{2}{1 - 4\xi^2} \left(\sqrt{1 - \frac{4\xi^2}{(1 - 2\Delta)^2}} + 1 \right), \quad (413)$$

which is obtained from $\mathcal{F}(\Delta, A, \xi)/A^2$. Given the form of Eq. (412), the optimized ρ will only depend on Δ, motivating the definition of the following function

$$\tilde{O}(\Delta) = \min_{\rho} \tilde{\mathcal{F}}^2(\Delta, \xi(\rho)) \langle \hat{O} \rangle_{\rho}. \quad (414)$$

A convenient way to generate this function is to perform a two stage minimization. First, we perform a constrained minimization with the restriction $\xi(\rho) = \xi$ on ρ. Second, we minimize the expression over ξ. In the first stage, given that $\tilde{\mathcal{F}}^2(\Delta, \xi(\rho))$ is fixed, we only need to minimize $\langle \hat{O} \rangle_{\rho}$, which can be mathematically expressed as

$$\mathcal{O}(\xi) = \min_{\rho} \left\{ \langle \hat{O} \rangle_{\rho} \big| \tfrac{1}{2} \langle \hat{\sigma}_{\ell}^x \rangle_{\rho} = \xi, \langle \hat{\sigma}_{\ell}^z \rangle_{\rho} = 0 \right\}. \quad (415)$$

To efficiently generate $\mathcal{O}(\xi)$, we can introduce a Lagrange multiplier λ and determine the ground state for $\hat{\mathcal{H}} = \hat{O} - \lambda \sum_{\alpha\sigma} \hat{\xi}_{\alpha\sigma}$, yielding the optimal ρ and corresponding ξ and $\langle \hat{O} \rangle_{\rho}$ for a given λ. One can then perform such calculations over a grid of λ, and then spline the relationship between $\langle \hat{O} \rangle_{\rho}$ and ξ. A plot of $\mathcal{O}(\xi)$ is provided in Fig. 2 of Ref. [18]. Finally, the partially optimized local energy can be written as

$$E_{\text{loc}}(\Delta, A) = U A^4 \tilde{O}(\Delta), \quad (416)$$

$$\tilde{O}(\Delta) = \min_{\xi \in [0, \frac{1}{2} - \Delta]} \tilde{\mathcal{F}}^2(\Delta, \xi) \mathcal{O}(\xi). \quad (417)$$

To visualize $\tilde{O}(\Delta)$ and ξ as a function of Δ, we plot these quantities for $J/U = 0, 0.05, 0.25$ and $N_{\text{orb}} = 2, 3, 5, 7$ (see Fig. 6). For $J/U = 0$, one can clearly observe that ξ continuously goes to zero, while there is a discontinuity for $J/U = 0.05$. For $J/U = 0.25$, ξ discontinuously goes to zero for $N_{\text{orb}} = 2, 3$ and continuously goes to zero for $N_{\text{orb}} = 5, 7$. It should be noted that $\tilde{O}(\Delta)$ can be applied to solve an arbitrary particle-hole symmetric $D(\epsilon)$, and therefore captures the essence of the Mott transition for a given J/U.

1. Understanding the nonanalytic behavior of $\tilde{O}(\Delta)$ via a Taylor series

In Sec. VII B, we demonstrated that there is nonanalyticity in $\tilde{O}(\Delta)$ at $\Delta = \Delta_c$. Here we provide a Taylor series analysis to explain how the nonanalyticity emerges. Equation (417) indicates that $\tilde{O}(\Delta)$ is the minimum of $\tilde{\mathcal{F}}^2(\Delta, \xi) \mathcal{O}(\xi)$ within the range $\xi \in [0, \frac{1}{2} - \Delta]$, and it is convenient to study the quantity

$$\mathcal{L}(\Delta, \xi) = -\frac{\tilde{\mathcal{F}}^2(\Delta, \xi)}{\tilde{\mathcal{F}}^2(\Delta, 0)} \frac{\mathcal{O}(\xi)}{\mathcal{O}(0)}. \quad (418)$$

Finding the minimum of $\mathcal{L}(\Delta, \xi)$ will yield the minimum of $\tilde{\mathcal{F}}^2(\Delta, \xi) \mathcal{O}(\xi)$ given that $\tilde{\mathcal{F}}(\Delta, 0) = 4$ and $\mathcal{O}(0) = -\frac{1}{4} N_{\text{orb}} (1 + (N_{\text{orb}} - 1) J/U)$.

We begin by Taylor series expanding $\mathcal{L}(\Delta, \xi)$ to sixth order in ξ about $\xi = 0$, and the second-order coefficient in ξ is expanded in Δ about $\Delta_{c,I}$ such that it is zero for $\Delta = \Delta_{c,I}$,
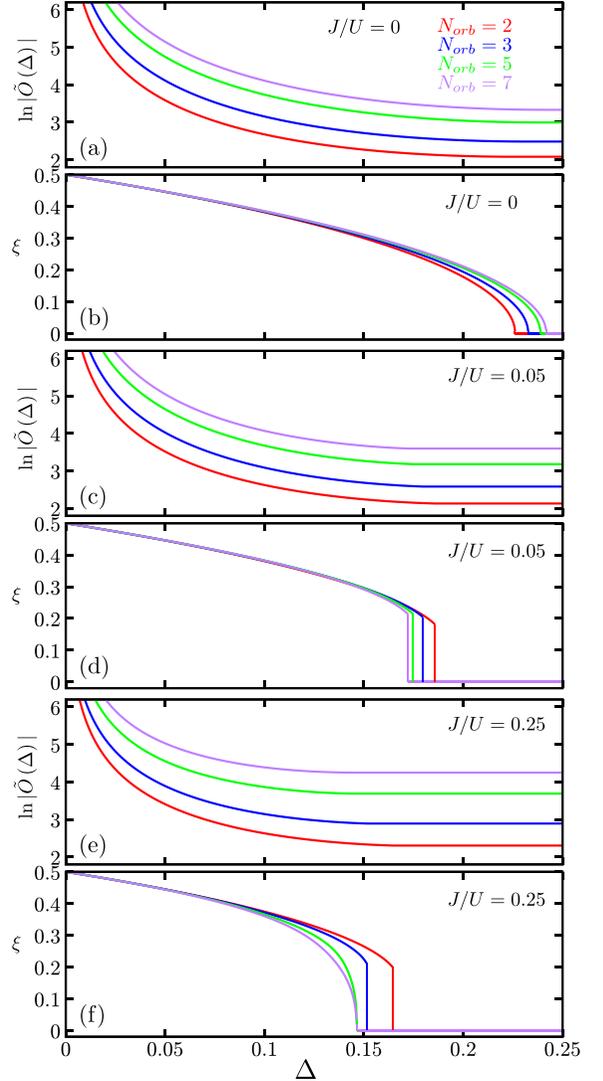

FIG. 6. Plots of $\ln |\tilde{O}(\Delta)|$ and ξ as a function of Δ for the multiorbital Hubbard model with $N_{\text{orb}} = 2, 3, 5, 7$ at half-filling for $J/U = 0, 0.05, 0.25$, which can be applied to all particle-hole symmetric $D(\epsilon)$.

yielding

$$\mathcal{L}(\Delta, \xi) \approx -1 + c_2 (\Delta - \Delta_{c,I}) \xi^2 + c_4 \xi^4 + c_6 \xi^6, \quad (419)$$

where c_2, c_4, c_6, and $\Delta_{c,I}$ are constants for a given N_{orb} and J/U. Given that the optimized ξ goes to zero with increasing Δ (see Fig. 6), this requires $c_2 > 0$. Furthermore, we take $c_6 > 0$, though there will be cases where c_6 is negative and a higher order expansion is necessary. Therefore, we only need to understand how the sign of c_4 influences the nonanalyticity in $\tilde{O}(\Delta)$. For $c_4 > 0$, when $\Delta > \Delta_{c;I}$, the minimum of $\mathcal{L}$ is given by $\mathcal{L} = -1$ with $\xi = 0$, while for $\Delta < \Delta_{c;I}$, the minimum of $\mathcal{L}$ is obtained with

$$\xi^2 = \frac{\sqrt{3 c_2 c_6 (\Delta_{c,I} - \Delta) + c_4^2} - c_4}{3 c_6}. \quad (420)$$

Notice that ξ^2 continuously increases from 0 when Δ decreases from $\Delta_{c;I}$. Therefore, the critical value of Δ is given

by $\Delta_c = \Delta_{c;I}$. Moreover, using

$$\frac{d\tilde{O}(\Delta)}{d\Delta} = \left(\frac{\partial}{\partial\xi}\tilde{\mathcal{F}}^2(\Delta,\xi)\right)\mathcal{O}(\xi), \qquad (421)$$

we find that $\frac{d\tilde{O}(\Delta)}{d\Delta}$ is continuous and only has a kink at $\Delta = \Delta_c$. For $c_4 < 0$, when $\Delta > \Delta_{c;I}$ there is a local minimum at $\xi = 0$, while for $\Delta < \Delta_{c,I} + \frac{c_4^2}{3c_2c_6}$ there is a local minimum given by Eq. (420). The two saddle points need to be compared to obtain the global minimum, which yields $\Delta_c = \Delta_{c,I} + \frac{c_4^2}{4c_2c_6}$. Therefore, ξ jumps from zero to a finite value when Δ decreases from Δ_c, implying that $\frac{d\tilde{O}(\Delta)}{d\Delta}$ is discontinuous at $\Delta = \Delta_c$. It should be noted that for $c_4 > 0$, $\Delta_c = \Delta_{c;I}$ is exact, while for $c_4 < 0$, the expression $\Delta_c = \Delta_{c,I} + \frac{c_4^2}{4c_2c_6}$ is an approximation, and in this case one should use the exact form of $\mathcal{L}(\Delta, \xi)$ to determine Δ_c if precision is needed.

We now proceed to analytically compute the expansion for two cases: $J/U = 0$ with $N_{\mathrm{orb}} \geqslant 1$, and $J/U > 0$ with $N_{\mathrm{orb}} = 2$. We begin by expanding

$$\frac{\tilde{\mathcal{F}}^2(\Delta,\xi)}{\tilde{\mathcal{F}}^2(\Delta,0)} = 1 + a_2\xi^2 + a_4\xi^4 + a_6\xi^6 + \cdots, \qquad (422)$$

where

$$a_2 = 8 - \frac{2}{(1-2\Delta)^2}, \qquad (423)$$

$$a_4 = -\frac{16}{(1-2\Delta)^2} - \frac{1}{(1-2\Delta)^4} + 48, \qquad (424)$$

$$a_6 = -\frac{96}{(1-2\Delta)^2} - \frac{8}{(1-2\Delta)^4} - \frac{2}{(1-2\Delta)^6} + 256. \qquad (425)$$

It should be noted that a_2, a_4, and a_6 all monotonically decrease with increasing Δ for $\Delta \in [0, 1/4]$. It is straightforward to show that $a_2 = 0$ when $\Delta = 1/4$ and $a_4 = 0$ when $\Delta = \frac{1}{4}(2 - \sqrt{\frac{1}{3}(2 + \sqrt{7})}) \approx 0.188895$ and $a_6 = 0$ when $\Delta = 0.155281$. The remaining task is to compute $-\frac{\mathcal{O}(\xi)}{\mathcal{O}(0)}$, which we separately consider for the two aforementioned cases.

For $J/U = 0$ and $N_{\mathrm{orb}} \geqslant 1$, symmetry can be used to parametrize $\boldsymbol{\rho} = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = \sum_\Gamma \sqrt{p_{|N_\Gamma - N_{\mathrm{orb}}|}}|\Gamma\rangle$, where p_i is a variational parameter with $i = 0, \ldots, N_{\mathrm{orb}}$ and $N_\Gamma = \langle\Gamma|\sum_{\ell=0}^{2N_{\mathrm{orb}}}\hat{n}_\ell|\Gamma\rangle$ counts the number of electrons in state Γ. It is convenient to reparametrize p_i as

$$x_i = \sqrt{\binom{2N_{\mathrm{orb}}}{N_{\mathrm{orb}}}(2 - \delta_{i,0})\frac{(-i + N_{\mathrm{orb}} + 1)_i}{(N_{\mathrm{orb}} + 1)_i}}\sqrt{p_i}, \qquad (426)$$

such that $\mathrm{Tr}(\boldsymbol{\rho}) = \sum_{i=0}^{N_{\mathrm{orb}}} x_i^2$, where $(m)_n = \frac{\Gamma(m+n)}{\Gamma(m)} = m(m+1)\ldots(m+n-1)$ is the Pochhammer symbol. When taking an expectation value of an operator $\hat{A}$ in the qubit space, it is convenient to use a matrix representation $[\hat{A}]_{ij}$, where i and j take values from $0, \ldots, N_{\mathrm{orb}}$, such that

$$\mathrm{Tr}(\boldsymbol{\rho}\hat{A}) = \sum_{i,j=0}^{N_{\mathrm{orb}}} x_i[\hat{A}]_{ij}x_j. \qquad (427)$$

The nonzero entries of the effective matrices for $\hat{O}$ and $\hat{\sigma}_\ell^x$ are given as

$$[\hat{O}]_{ii} = \frac{1}{4}(2i^2 - N_{\mathrm{orb}}), \qquad (428)$$

$$\left[\hat{\sigma}_\ell^x\right]_{i,i+1} = \left[\hat{\sigma}_\ell^x\right]_{i+1,i} = \frac{\sqrt{(N_{\mathrm{orb}} - i)(i + N_{\mathrm{orb}} + 1)}}{\sqrt{2}N_{\mathrm{orb}}\sqrt{2 - \delta_{0,i}}}. \qquad (429)$$

Perturbation theory can then be used to obtain

$$-\frac{\mathcal{O}(\xi)}{\mathcal{O}(0)} = -1 + \frac{4\xi^2}{N_{\mathrm{orb}} + 1} + \frac{4b_4\xi^4}{N_{\mathrm{orb}}} + \frac{4b_6\xi^6}{N_{\mathrm{orb}}} + O(\xi^7), \qquad (430)$$

where

$$b_4 = \frac{N_{\mathrm{orb}}(7N_{\mathrm{orb}}(N_{\mathrm{orb}} + 1) + 2)}{4(N_{\mathrm{orb}} + 1)^3}, \qquad (431)$$

$$b_6 = \frac{N_{\mathrm{orb}}(N_{\mathrm{orb}}(N_{\mathrm{orb}} + 1)h + 24)}{36(N_{\mathrm{orb}} + 1)^5}, \qquad (432)$$

$$h = 209N_{\mathrm{orb}}(N_{\mathrm{orb}} + 1) + 146. \qquad (433)$$

The expansion coefficients of $\mathcal{L}$ are obtained as

$$\Delta_{c;I} = \frac{4N_{\mathrm{orb}} - \sqrt{4N_{\mathrm{orb}}^2 + 6N_{\mathrm{orb}} + 2} + 2}{8N_{\mathrm{orb}} + 4}, \qquad (434)$$

$$c_2 = \frac{16\sqrt{2}(2N_{\mathrm{orb}} + 1)^3}{((N_{\mathrm{orb}} + 1)(2N_{\mathrm{orb}} + 1))^{3/2}}, \qquad (435)$$

$$c_4 = \frac{N_{\mathrm{orb}}(N_{\mathrm{orb}}(32N_{\mathrm{orb}} + 55) + 27) + 6}{(N_{\mathrm{orb}} + 1)^3}, \qquad (436)$$

$$c_6 = \frac{N_{\mathrm{orb}}((h + 9211)N_{\mathrm{orb}} + 2774) + 384}{9(N_{\mathrm{orb}} + 1)^5}, \qquad (437)$$

$$h = N_{\mathrm{orb}}(N_{\mathrm{orb}}(3456N_{\mathrm{orb}} + 11729) + 15070). \qquad (438)$$

Given that $c_4 > 0$, we have $\Delta_c = \Delta_{c;I}$ and $\frac{d\tilde{O}(\Delta)}{d\Delta}$ has a kink at $\Delta = \Delta_c$.

We now discuss the case of $J/U > 0$ with $N_{\mathrm{orb}} = 2$. Using perturbation theory, we find

$$-\frac{\mathcal{O}(\xi)}{\mathcal{O}(0)} = -1 + 4\xi^2 + b_4\xi^4 + b_6\xi^6, \qquad (439)$$

where

$$b_4 = -\frac{4}{3}\left(\frac{8}{r} - \frac{6}{r-1} - 7\right), \qquad (440)$$

$$b_6 = \frac{32(r(r(7r(2r - 3) + 38) - 42) + 29)}{9(r-1)^2r^2}, \qquad (441)$$

where $r = J/U$. Therefore, we can compute the coefficients in the expansion of $\mathcal{L}$ as

$$\Delta_{c;I} = \frac{1}{4}(2 - \sqrt{2}) \approx 0.146447, \qquad (442)$$

$$c_2 = 16\sqrt{2}, \qquad (443)$$

$$c_4 = \frac{8}{3}\left(-\frac{4}{r} + \frac{3}{r-1} + 5\right), \qquad (444)$$

$$c_6 = \frac{16}{9}\left(\frac{8r + 58}{r^2} + \frac{36}{(r-1)^2} + 67\right). \qquad (445)$$

We can see that $c_4 < 0$, and therefore $\Delta_c > \Delta_{c,I}$, indicating that there is a discontinuity in $\frac{d\tilde{O}(\Delta)}{d\Delta}$ for $\Delta = \Delta_c$.

Finally, we numerically explore the cases of $N_{orb} > 2$, where $\Delta_{c;I}$ and c_2 are identical to the case of $N_{orb} = 2$. To be concrete, we consider $J/U = 0.25$, and we numerically compute the expansion coefficients by fitting $\mathcal{O}(\xi)$ to a sixth order polynomial. For $N_{orb} = 3$, we have $c_4 \approx -2.9$ and $c_6 \approx -147$. Notice that in this case $c_6 < 0$, and thus an expansion beyond sixth order is necessary, though plotting $\mathcal{L}(\Delta, \xi)$ indicates that $\Delta_c > \Delta_{c,I}$, yielding a discontinuity in $\frac{d\tilde{O}(\Delta)}{d\Delta}$ for $\Delta = \Delta_c$ (not shown). For $N_{orb} = 4$, $c_4 \approx -0.63$ and $c_6 \approx 39$. For $N_{orb} = 5$, $c_4 \approx 0.44$ and $c_6 \approx 35$. For $N_{orb} = 6$, $c_4 \approx 1.1$ and $c_6 \approx 33$. For $N_{orb} = 7$, $c_4 \approx 1.5$ and $c_6 \approx 34$. It should be noted that for $N_{orb} \geqslant 5$, we have $\Delta_c = \Delta_{c;I}$ and $\frac{d\tilde{O}(\Delta)}{d\Delta}$ has a kink at $\Delta = \Delta_c$.

2. Solution for the two-peak density of states

We now consider the two-peak density of states

$$D(\epsilon) = \frac{1}{2}\left(\delta\left(\epsilon - \frac{1}{N_{orb}}K_0\right) + \delta\left(\epsilon + \frac{1}{N_{orb}}K_0\right)\right), \quad (446)$$

where K_0 is the total noninteracting kinetic energy per site. For the two-peak density of states, $A = \sqrt{\Delta(2 - 4\Delta)}$, and the qubit trial energy can be written purely in terms of Δ as

$$E(\Delta) = (1 - 4\Delta)K_0 + U O_{tp}(\Delta), \quad (447)$$

$$O_{tp}(\Delta) = (2 - 4\Delta)^2 \Delta^2 \tilde{O}(\Delta). \quad (448)$$

Following the single orbital case, Δ can be used to determine U from $dE(\Delta)/d\Delta = 0$, which yields

$$U(\Delta) = -4|K_0|\left(\frac{dO_{tp}(\Delta)}{d\Delta}\right)^{-1}, \quad (449)$$

thus providing a succinct solution parametrized by Δ. The relation $U(\Delta)$ can be used to determine the nature of the Mott transition from $\tilde{O}(\Delta)$, given that this quantity will allow any observable to be expressed in terms of U. We plot $\frac{dO_{tp}}{d\Delta}$ versus Δ for various N_{orb} and J/U, which demonstrates three types of nonanalytical scenarios for $O_{tp}(\Delta)$ (see Fig. 7, panel a). First, for all $J/U = 0$, the $\frac{dO_{tp}}{d\Delta}$ is continuous with a positive slope and has a kink at $\Delta = \Delta_c$. Second, for $J/U > 0$ and small N_{orb}, the $\frac{dO_{tp}}{d\Delta}$ is discontinuous at $\Delta = \Delta_c$ with a negative slope for Δ_c^-. Third, for $J/U = 0.25$ and $N_{orb} = 7$, the $\frac{dO_{tp}}{d\Delta}$ is continuous and has a kink at $\Delta = \Delta_c$, and the slope is negative for Δ_c^-. The $\frac{dO_{tp}}{d\Delta}$ can now be used to determine $U(\Delta)$, which yields the order of the Mott transition [see Fig. 7(b)]. First, for all $J/U = 0$, U increases monotonically and continuously with Δ, with a kink at $\Delta = \Delta_c$, and therefore there are no metastable regions and the Mott transition is continuous. Second, for $J/U > 0$, there is an unstable region in the metal phase where $\frac{dU}{d\Delta} < 0$, and the total energy can be used to determine the transition between the metal and insulating phase, corresponding to a horizontal line, and the transition is first-order.

In summary, we have demonstrated that the nature of the Mott transition for the two-peak density of states is determined purely by $\tilde{O}(\Delta)$, and below we demonstrate that the Bethe lattice has the same behavior. Therefore, it appears that
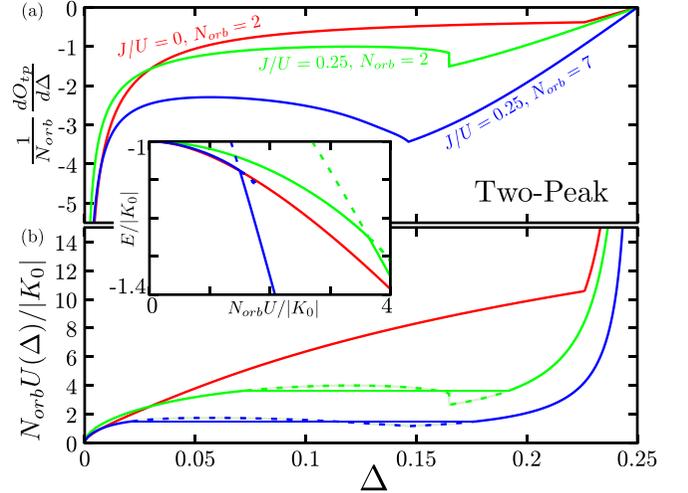

FIG. 7. Plots of $\frac{1}{N_{orb}}\frac{dO_{tp}}{d\Delta}$ (a) and $N_{orb}U(\Delta)/|K_0|$ (b) as functions of Δ for various N_{orb} and J/U for the two-peak model in $d = \infty$ solved using $\hat{\rho}_{G3}$. The inset plots $E/|K_0|$ vs $N_{orb}U/|K_0|$, where E is the total energy per site. The dotted lines represent unstable or metastable solutions.

$\tilde{O}(\Delta)$ is the essence of what determines the nature of the Mott transition in $d = \infty$.

3. Solution for a general density of states

We now execute a similar strategy for a general density of states. Using Eq. (416), the qubit trial energy can be written as

$$E(\Delta, A) = K(\Delta, A) + UA^4 \tilde{O}(\Delta). \quad (450)$$

The saddle-point equations are given as

$$\frac{\partial K}{\partial \Delta} = -4N_{orb}a = -UA^4 \frac{d\tilde{O}(\Delta)}{d\Delta}, \quad (451)$$

$$\frac{\partial K}{\partial A} = 2N_{orb}b = -4UA^3 \tilde{O}(\Delta). \quad (452)$$

Given that Δ and A are functions of a and b, one can solve a and b from Eqs. (451) and (452) for a given U, and then determine all physical quantities. We now demonstrate that $\frac{d\tilde{O}(\Delta)}{d\Delta} = 0$ indicates that the system is in the Mott phase. Given that the quasiparticle weight is given as $Z = a/\sqrt{a^2 + b^2}$ [see Eq. (373)], and that $A > 0$ for finite U [see Eq. (371)], the only scenario where $Z = 0$ is when $\frac{d\tilde{O}(\Delta)}{d\Delta} = 0$. Therefore, when $\Delta < \Delta_c$, the system is metallic, while $\Delta > \Delta_c$ the system is insulating. For a given U, one must minimize over Δ to determine nature of the ground state.

An alternate approach is to parametrize the solution in terms of Δ. Equations (451) and (452), in addition to the constraint on a and b for a given Δ, yield

$$-\frac{8a}{A(a, b)b} = \frac{d\ln|\tilde{O}(\Delta)|}{d\Delta}, \quad (453)$$

$$\Delta(a, b) = \Delta. \quad (454)$$

For a given Δ, a and b can be determined from Eqs. (453) and (454), A can be determined from $A(a, b)$, and U can be
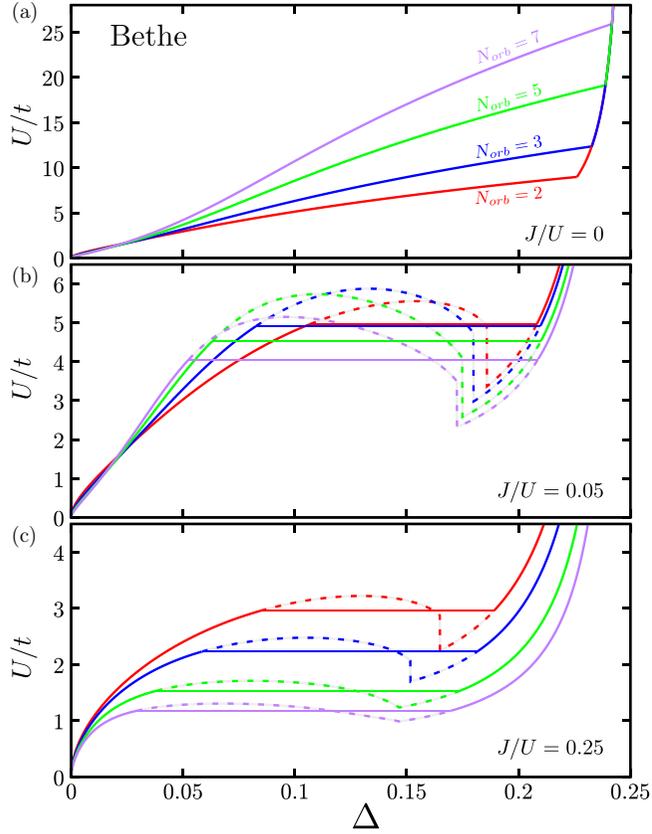
FIG. 8. Plots of U/t as a function of Δ for the multiorbital Hubbard model on the Bethe lattice in $d = \infty$ using $\hat{\rho}_{G3}$ for $N_{\text{orb}} = 2, 3, 5, 7$ with $J/U = 0$ (a), $J/U = 0.05$ (b), and $J/U = 0.25$ (c). The dotted lines indicate metastable or unstable solutions.

determined using Eq. (452) as

$$U(\Delta) = -\frac{N_{\text{orb}} b}{2A^3(a, b)\tilde{O}(\Delta)}, \qquad (455)$$

allowing for the evaluation of the total energy.

We now consider the Bethe lattice in $d = \infty$ for $N_{\text{orb}} = 2, 3, 5, 7$. Recall that for a given J/U, the $\tilde{O}(\Delta)$ yields a Δ_c which divides the metallic and insulating states, where $\Delta > \Delta_c$ indicates an insulating phase. For the case of a continuous transition, the U_c will be determined by Δ_c, while for a first-order transition, one must explicitly determine the U_c where the insulating and metallic states cross in energy. The algorithm is executed by evaluating A, U, the interaction energy, and the total energy as functions of Δ. We begin by plotting the U/t as a function of Δ [see Fig. 8(a)]. For $J/U = 0$, U is a monotonic and continuous function of Δ, implying a continuous phase transition at Δ_c, which can be identified as a kink. Alternatively, for $J/U = 0.05$, the U is not a monotonic nor a continuous function of Δ, implying that the there are regions of phase coexistence and unstable regions. The metallic curve exists for $\Delta < \Delta_c$, and the solution is only stable for $\Delta < \Delta_{c;1}$, where $\Delta_{c;1}$ is determined from $\frac{dU}{d\Delta} = 0$, and therefore the metallic solution is only stable for $U < U(\Delta_{c;1})$. The insulating curve exists for $\Delta > \Delta_c$ without any unstable regions, and therefore the insulating phase exists for $U > U(\Delta_c + 0^+)$. Given that $U(\Delta_c + 0^+) < U(\Delta_{c;1})$, there exists
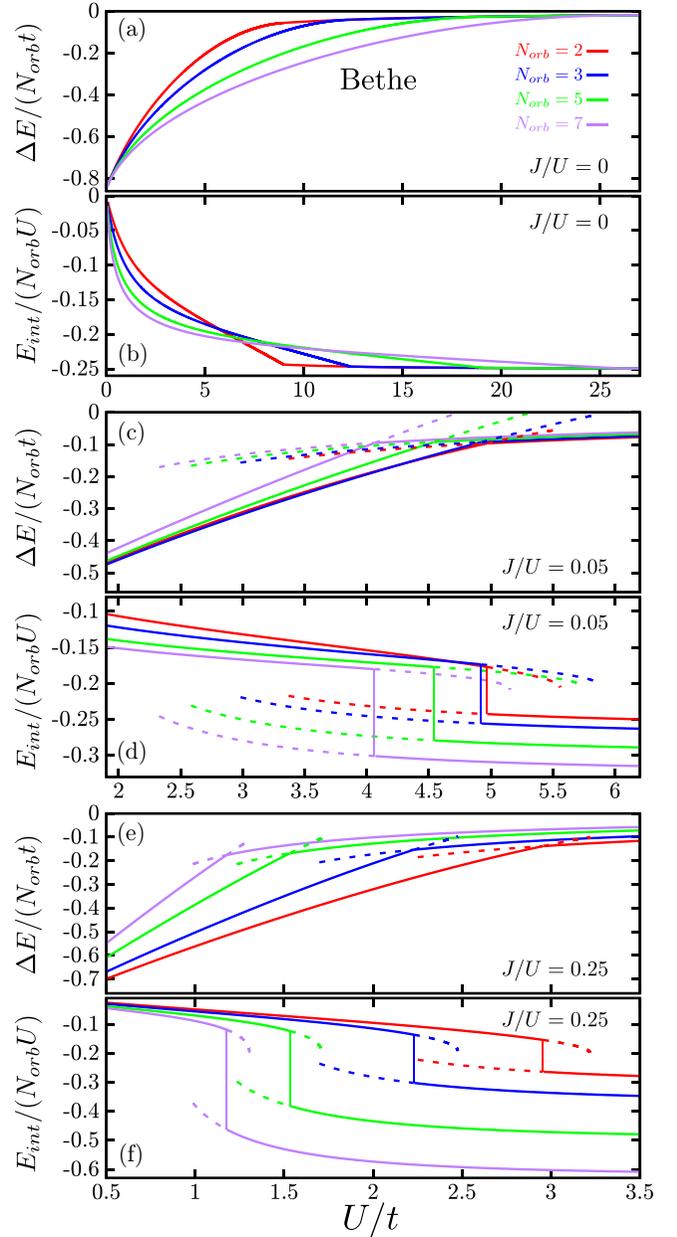

FIG. 9. Plots of the total energy and the interaction energy as functions of U/t for the multiorbital Hubbard model on the Bethe lattice in $d = \infty$ using $\hat{\rho}_{G3}$ for $N_{\text{orb}} = 2, 3, 5, 7$ with $J/U = 0$ (a), (b), $J/U = 0.05$ (c), (d), and $J/U = 0.25$ (e), (f). The dotted lines indicate metastable or unstable solutions.

a region of coexistence for the metallic and insulating solutions, and the total energy dictates the lowest energy solution. We now proceed to present the total energy and the interaction energy as functions of U/t for various J/U and N_{orb} (see Fig. 9), and the results are very similar to the two-peak case. Consistent with previous Gutzwiller [11], slave boson [82], and DMFT [83] studies, the Mott transition is continuous for $J/U = 0$ and first-order for $J/U > 0$.

Finally, we evaluate U_c for $J/U = 0$, which is explicitly given by $U_c = \frac{b_c}{8A_c^3}$, where b_c is determined from $\Delta(0, b_c) = \Delta_c$ and $A_c = A(0, b_c)$, where $\Delta(a, b)$ and $A(a, b)$ are defined

TABLE IV. Critical values for the SU($2N_{\text{orb}}$) Hubbard model ($J/U = 0$) on the Bethe lattice in $d = \infty$ solved using $\hat{\rho}_{G3}$.

N_{orb}	Δ_c	b_c/t	A_c	U_c/t
1	$\frac{1}{6}(3 - \sqrt{3}) \approx 0.211325$	5.34164	0.491668	5.61784
2	$\frac{1}{20}(10 - \sqrt{30}) \approx 0.226139$	8.80352	0.496835	8.97283
3	$\frac{1}{2} - \frac{1}{\sqrt{14}} \approx 0.232739$	12.2288	0.498345	12.3511
4	$\frac{1}{12}(6 - \sqrt{10}) \approx 0.236477$	15.6412	0.498984	15.7369
5	$\frac{1}{22}(11 - \sqrt{33}) \approx 0.238884$	19.0475	0.499314	19.1261
6	$\frac{1}{2} - \frac{\sqrt{\frac{7}{26}}}{2} \approx 0.240563$	22.4505	0.499505	22.5172
7	$\frac{1}{2} - \frac{1}{\sqrt{15}} \approx 0.241801$	25.8515	0.499627	25.9094

in Eqs. (372) and (371), respectively. The critical values for the Bethe lattice are listed in Table IV. For the case of large N_{orb}, we have

$$U_c/t \approx \frac{32 N_{\text{orb}}}{3\pi} + \frac{20}{3\pi} \tag{456}$$

$$\approx 3.39531 N_{\text{orb}} + 2.12207, \tag{457}$$

consistent with our numerical results in Ref. [17]. For a systematic exploration of how U_c depends on J/U and N_{orb}, see Fig. 5 of Ref. [18].

VIII. SUMMARY AND CONCLUSIONS

We begin by providing a high-level overview of the VDAT, such that the developments of the present work can be properly understood. VDAT is a variational approach to the many-body body problem that consists of two main components: a variational ansatz for the many-body wave function or density-matrix, known as the SPD, and a formalism for evaluating expectation values under the SPD, known as the discrete action theory [15,16]. The SPD has a natural mechanism to trade off between efficiency and accuracy, where the integer $\mathcal{N}$ monotonically increases the variational power of the SPD and guarantees the ability to recover the ground-state solution for $\mathcal{N} \to \infty$. Moreover, there are two distinct types of SPD which satisfy the properties of a many-body density matrix, denoted as G-type and B-type. The G-type $\mathcal{N} = 1$, $\mathcal{N} = 2$, and $\mathcal{N} = 3$ SPD encapsulate the Hartree-Fock wave function, the Gutzwiller wave function, and the Gutzwiller-Baeriswyl wave function, respectively. The key breakthrough using VDAT was the demonstration that the SPD can be *exactly* evaluated for multiorbital Hubbard models in $d = \infty$. We demonstrated that the G-type $\mathcal{N} = 3$ SPD accurately solves the Anderson impurity model on a ring [16], the single-band Hubbard model over all parameter space [16], the two orbital Hubbard model including a crystal field and the full rotationally invariant Hund's coupling [8], and the $SU(2N_{\text{orb}})$ Hubbard model for $N_{\text{orb}} \leqslant 8$ [17]. Moreover, we demonstrated that the computational cost of solving a G-type $\mathcal{N} = 3$ SPD is comparable to a G-type $\mathcal{N} = 2$ SPD, meaning that VDAT can provide a sufficiently accurate solution at a cost not far beyond the Gutzwiller approximation. The success of VDAT at $\mathcal{N} = 3$ motivated a search for the best possible algorithm for executing calculations using a G-type $\mathcal{N} = 3$ SPD [17], which is essential for detailed exploration of the

multiorbital Hubbard model and merging VDAT with realistic electronic structure methods.

The VDAT algorithm in $d = \infty$ consists of two steps: the exact evaluation of the SPD via the SCDA and the optimization of the energy with respect to the variational parameters. The SCDA requires the numerical solution of a set of self-consistency conditions, and therefore can be inconvenient when minimizing over the variational parameters. For the case of a G-type $\mathcal{N} = 3$ SPD with certain restrictions (see Secs. I, VI A, and VI B for further details), the SCDA self-consistency condition can be automatically satisfied, which we refer to as the gauge constrained SCDA algorithm [17]. In the present work, we introduce the so-called qubit parametrization of the gauge constrained SCDA algorithm, which is mathematically equivalent to the original gauge constrained SCDA algorithm. The qubit parametrization offers several key improvements. The qubit parametrization analytically resolves some constraints over the variational parameters, thus reducing the number of variational parameters by one per spin orbital. Additionally, the variational parameters are physically intuitive and facilitate a deeper understanding of how the SPD captures Mott and Hund physics. Therefore, the qubit parametrization achieves the long sought goal of resolving the shortcomings of the Gutzwiller approximation while maintaining the computational simplicity and physical appeal.

The variational parameters of the qubit parametrization consist of the momentum density distribution, the noninteracting reference momentum density distribution, and the pure state of a qubit system with a dimension of the local Hilbert space. The qubit system naturally arises from reparametrizing the variational parameters of the interacting projector, and the renormalized correlations within the qubit space yield the physical local correlations. The variational parameters are restricted by two constraints per spin orbital, requiring that the local density computed from the momentum density distribution is the same as that computed from the noninteracting reference momentum density distribution, and the same as the density computed from the qubit system. The qubit trial energy has a very intuitive form: the kinetic energy is determined by the momentum density distributions, while the local interaction energy is the expectation value of an effective Hamiltonian within the qubit system. Interestingly, the effective Hamiltonian has the same form as the local interacting Hamiltonian where the local density operator is substituted by an effective density operator of the qubit system. The effective density operator for a given spin-orbital ℓ depends on five parameters: the density n_ℓ, the magnetization in the x direction for the ℓth qubit, denoted ξ_ℓ, and three quantities determined from the momentum density distribution, denoted as Δ_ℓ, $\mathcal{A}_{<\ell}$, and $\mathcal{A}_{>\ell}$. The quantity Δ_ℓ characterizes the number of electrons promoted across the reference Fermi surface, while $\mathcal{A}_{<\ell}$ and $\mathcal{A}_{>\ell}$ characterize the momentum density distribution below and above the reference Fermi surface, respectively. The quantity ξ_ℓ is an important variable which differentiates between a zero and nonzero quasiparticle weight when all variables are fully optimized, where $\xi_\ell = 0$ indicates zero quasiparticle weight for spin-orbital ℓ. The main computational cost for evaluating the ansatz is dictated by computing expectation values within the qubit system.

While evaluating the trial qubit energy ansatz is a straightforward task, optimizing over all the variational parameters remains nontrivial. In general, the qubit trial energy can be partially optimized over the momentum density distribution by introducing four Lagrange multipliers per spin-orbital, replacing the continuous momentum density distribution with four variables. For a system with $2N_{\text{orb}}$ spin-orbitals per site, there remain $2^{2N_{\text{orb}}} + 3 \times 2N_{\text{orb}} - 1$ variational parameters which must be optimized in general. However, this number may be greatly reduced by symmetry, and it is likely possible to compress these variables into a smaller number of parameters without a serious loss of fidelity. For the special case of half-filled orbitals with particle-hole symmetry, we demonstrate that one can efficiently minimize over all variational parameters, with a computational cost proportional to computing the ground state of a Hamiltonian defined within the qubit system.

To demonstrate the power of the qubit parametrization, we studied the ground-state properties of the multiorbital Hubbard model at half-filling with particle-hole symmetry for various J/U and $N_{\text{orb}} = 2 - 7$. For a given J/U, the majority of the energy minimization can be encapsulated into the computation of a single variable function $\tilde{O}(\Delta)$, which can then be used to obtain the solution at a negligible cost for an arbitrary U and density-of-states. The entire function $\tilde{O}(\Delta)$ is evaluated by solving a collection of qubit systems, which has a relatively small computational cost. For example, for $N_{\text{orb}} = 7$, the ground state for a given qubit system can be solved in several seconds on a typical single desktop computer core, and taking on the order of 100 samples, the entire function $\tilde{O}(\Delta)$ can be accurately obtained on the order of hundreds of seconds. The extreme computational efficiency of the qubit parametrization in this case allows one to easily map out all of parameter space, which is not possible with DMFT given the lack of efficient impurity solvers for the zero temperature multiorbital problem. We find that for $J/U = 0$, the Mott transition is continuous, while it is first-order for $J/U > 0$, consistent with previous Gutzwiller [11], slave boson [82], and DMFT [83] studies.

While the key result of this paper is formulating the qubit parametrization of the gauge constrained SCDA algorithm for a G-type $\mathcal{N} = 3$ SPD, we also demonstrate that the qubit parametrization can be applied to the G-type and B-type $\mathcal{N} = 2$ SPD. Moreover, we demonstrate that properly restricting the variational parameters of the qubit trial energy for the G-type $\mathcal{N} = 3$ SPD can recover the corresponding qubit trial energy for the G-type and B-type $\mathcal{N} = 2$ SPD. Interestingly, the qubit trial energy for the G-type $\mathcal{N} = 2$ SPD has an identical form to the slave spin mean-field theory (see Appendix B), and thus the $\mathcal{N} = 3$ qubit trial energy may provide insights for proceeding beyond mean-field theory in the slave spin formalism.

The qubit parametrization of the gauge constrained SCDA algorithm at $\mathcal{N} = 3$ is likely the optimal form when evaluating an SPD with a kinetic projector that is diagonal in both the momentum and spin-orbital indices and an interacting projector that consists of diagonal Hubbard operators. For the Hamiltonians treated in the present study, which have density-density interactions and hopping parameters that are diagonal in the spin-orbital index, the aforementioned restrictions on the SPD do not limit the variational power. When solving

a general Hamiltonian which includes the full rotationally invariant form of the Hund exchange or nondiagonal hopping terms, the qubit parametrization can still be applied and it will still yield an upper bound on the energy in $d = \infty$, but it will not contain the full variational power of $\mathcal{N} = 3$. Ongoing research is addressing how to generalize the qubit parametrization to handle an arbitrary $\mathcal{N} = 3$ G-type SPD, with aspirations of completely superseding our general decoupled minimization algorithm for $\mathcal{N} = 3$ [8].

DATA AVAILABILITY

The data that support the findings of this article are openly available [84].

APPENDIX A: ONE-BODY REDUCED DENSITY MATRIX FUNCTIONAL FOR THE HUBBARD MODEL

An important application of the qubit energy form for $\hat{\rho}_{G3}$ is the construction of a 1RDMF for the multiorbital Hubbard model, which is the focus of our companion manuscript [18]. Here we derive the corresponding results for $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$. Additionally, we evaluate existing 1RDMFs from the literature for the single-band Hubbard model at half-filling in $d = \infty$.

1. 1RDMFs for the multiorbital Hubbard model from $\hat{\rho}_{G2}$ and $\hat{\rho}_{B2}$

We begin by presenting the 1RDMF from $\hat{\rho}_{B2}$ using the qubit parametrization. The interaction energy is given as

$$E_{\text{int}}(\{n_{k\ell}\}) = \min_{\boldsymbol{\rho}} \langle H_{\text{loc}}(\{\hat{n}_{\text{eff},\ell}\}) \rangle_{\boldsymbol{\rho}} \tag{A1}$$

$$\text{subject to} \langle \hat{n}_\ell \rangle_{\boldsymbol{\rho}} = n_\ell, \tag{A2}$$

where $n_\ell = \int dk\, n_{k\ell}$, the effective density operator is $\hat{n}_{\text{eff},\ell} = n_\ell + \frac{A_\ell^2}{\xi_{\ell,0}^2}(\hat{n}_\ell - n_\ell)$ with $A_\ell = \int dk \sqrt{n_{k\ell}(1 - n_{k\ell})}$ and $\xi_{\ell,0} = \sqrt{n_\ell(1 - n_\ell)}$, and both $\boldsymbol{\rho}$ and $H_{\text{loc}}(\{\hat{n}_{\text{eff},\ell}\})$ are diagonal in the Pauli-Z basis of the qubit system. For the case of half-filling, the minimization yields

$$E_{\text{int}}(\{n_{k\ell}\}) = 16 U A^4 \mathcal{O}(0), \tag{A3}$$

where $\mathcal{O}(0) = -\frac{1}{4} N_{\text{orb}}(1 + (N_{\text{orb}} - 1)J/U)$.

We now present the 1RDMF for $\hat{\rho}_{G2}$. Unlike $\hat{\rho}_{G3}$ or $\hat{\rho}_{B2}$, the variational parameters do not explicitly contain $n_{k\ell}$, but instead $n_{k\ell} - n_\ell = \xi_\ell^2/\xi_{\ell,0}^2(n_{k\ell,0} - n_\ell)$, where $n_{k\ell,0} \in [0, 1]$. Therefore, for a given $\{n_{k\ell}\}$, we have $|\xi_\ell/\xi_{\ell,0}| \geqslant \sqrt{Z_{\ell,\min}}$, where $Z_{\ell,\min} = \max(\frac{n_{\ell,\max} - n_\ell}{1 - n_\ell}, \frac{n_\ell - n_{\ell,\min}}{n_\ell})$, where $n_{\ell,\max}$ and $n_{\ell,\min}$ are the maximum and minimum values of $\{n_{k\ell}\}$ for a given ℓ. It should be noted that the gauge symmetry can

be used to restrict $\xi_\ell \geqslant 0$. The interaction energy can then be written as

$$E_{\text{int}}(\{n_{k\ell}\}) = \min_\rho \langle H_{\text{loc}}(\{\hat{n}_\ell\}) \rangle_\rho \tag{A4}$$

$$\text{subject to} \quad \frac{1}{2}\langle \hat{\sigma}_\ell^x \rangle_\rho \geqslant \xi_{\ell;\min}, \quad \langle \hat{n}_\ell \rangle_\rho = n_\ell, \tag{A5}$$

where $\xi_{\ell;\min} = \xi_{\ell;0}\sqrt{Z_{\min}}$. Notice that the interaction energy will decrease with decreasing ξ_ℓ, and therefore the restriction in Eq. (A5) can be replaced with $\frac{1}{2}\langle \hat{\sigma}_\ell^x \rangle_\rho = \xi_{\ell;\min}$. For the case of half-filling, we have

$$E_{\text{int}}(\{n_{k\ell}\}) = U\mathcal{O}\left(\frac{1}{2}\sqrt{1 - 2n_{\ell;\min}}\right), \tag{A6}$$

where $\mathcal{O}(\xi)$ is defined in Eq. (415).

Using the above interaction energy functionals will yield results identical to the corresponding VDAT results, such as the ones provided in Fig. 3.

2. Results using published 1RDMFs for the single-orbital Hubbard model in $d = \infty$

We are not aware of any applications of existing 1RDMFs to the Hubbard model in $d = \infty$, though there are various studies of Hubbard clusters [85,86] and $d = 1$ [87–89] and $d = 2$ Hubbard models [90,91]. It is important to emphasize that $d = \infty$ is the most relevant test of local electronic correlations, and is emblematic of typical three-dimensional strongly correlated materials in that $d = \infty$ hosts a standard Fermi liquid in the metal phase and exhibits a Mott transition at a finite value of U. While $d = 1$ is exactly solvable via the Bethe ansatz, the Mott transition occurs at an infinitesimal U [92]. Alternatively, $d = 2$ is a testbed for the most advanced and expensive computational approaches, and the properties at half-filling and low temperatures are still actively studied [93–96]. While it is still very interesting to compare total energies from 1RDMFs in $d = 1$ and $d = 2$, if the goal is to determine whether or not a 1RDMF can describe the Mott transition, then it is most critical to first benchmark in $d = \infty$.

In this section, we discuss how to use various 1RDMFs to solve the one band Hubbard model at half-filling, including the MBB [97], CA [98], CGA [99], power [100], PNOF5 [101], PNOF7 [102], and dimer [91] functionals. The interaction energy is given by Ud, where d is the double occupancy and is a functional of $\{n_{k\sigma}\}$. If the interaction Hamiltonian is rewritten as $\frac{U}{2}\sum_{i\sigma\sigma'}\hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma'}^\dagger \hat{a}_{i\sigma'} \hat{a}_{i\sigma}$, the interaction energy for the MBB, CA, CGA, and power functionals can be viewed as a modification of the Hartree-Fock (HF) energy where the Fock term is altered [86]. We begin by writing the local interaction as $E_{\text{int}} = \frac{U}{2}\sum_{\sigma\sigma'}\langle \hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma'}^\dagger \hat{a}_{i\sigma'} \hat{a}_{i\sigma} \rangle$, which can be written in momentum space as

$$E_{\text{int}} = \frac{U}{2L^2} \sum_{k_1 k_2 k_1' k_2' \sigma\sigma'} \delta_{k_1+k_2, k_1'+k_2'} \langle \hat{a}_{k_1\sigma}^\dagger \hat{a}_{k_2\sigma'}^\dagger \hat{a}_{k_2'\sigma'} \hat{a}_{k_1'\sigma} \rangle, \tag{A7}$$

where L is the number of k points. The MBB, CA, CGA, and power functionals are all defined using the following approximation

$$\begin{aligned} \langle \hat{a}_{k_1\sigma}^\dagger \hat{a}_{k_2\sigma'}^\dagger \hat{a}_{k_2'\sigma'} \hat{a}_{k_1'\sigma} \rangle \\ \approx n_{k_1} n_{k_2} \delta_{k_1 k_1'} \delta_{k_2 k_2'} - \mathcal{F}(n_{k_1}, n_{k_2}) \delta_{k_1 k_2'} \delta_{k_2 k_1'} \delta_{\sigma\sigma'}, \end{aligned} \tag{A8}$$

TABLE V. $\mathcal{F}(n_i, n_j)$ and E_{int} for the HF, MBB, Power, CA, and CGA functionals for the single orbital Hubbard model at half-filling in the paramagnetic phase.

Functional	$\mathcal{F}(n_i, n_j)$
HF	$n_i n_j$
MBB	$\sqrt{n_i n_j}$
power	$(n_i n_j)^\alpha$
CA	$n_i n_j + \sqrt{n_i(1-n_i)n_j(1-n_j)}$
CGA	$\frac{1}{2}(n_i n_j + \sqrt{n_i(2-n_i)n_j(2-n_j)})$

	E_{int}
HF	$\frac{U}{4}$
MBB	$\frac{U}{2}(1 - 2(\int dk\sqrt{n_k})^2)$
power	$\frac{U}{2}(1 - 2(\int dk n_k^\alpha)^2)$
CA	$U(\frac{1}{4} - (\int dk\sqrt{n_k(1-n_k)})^2)$
CGA	$U(\frac{3}{8} - \frac{1}{2}(\int dk\sqrt{n_k(2-n_k)})^2)$

where n_k is defined through $\langle \hat{a}_{k\sigma}^\dagger \hat{a}_{k'\sigma'} \rangle = n_k \delta_{kk'} \delta_{\sigma\sigma'}$. The $\mathcal{F}(n_i, n_j)$ and E_{int} are given for each functional in Table V, and our values of $\mathcal{F}(n_i, n_j)$ are identical to those provided in Ref. [86]. It is worth noting that the MBB, CA, and CGA recover the atomic limit where $E_{\text{int}} = 0$ when $n_k = 1/2$, while the power functional does not if $\alpha \neq 1/2$.

Now we discuss how to minimize the total energy over $\{n_k\}$. Given that MBB is a special case of the power functional, we first consider the power functional. Given that the interaction energy for the power functional is fixed for a given $\int dk n_k^\alpha$ and $\int dk n_k$, the kinetic energy can be minimized by introducing two Lagrange multipliers a and b, yielding a target functional $K = \int dk(\epsilon_k - a)n_k - b\int dk n_k^\alpha$ to be minimized, and the partially optimized n_k are obtained as

$$n_k = \begin{cases} 1 & \epsilon_k \leqslant a + \alpha b \\ \left(\frac{\epsilon_k - a}{\alpha b}\right)^{\frac{1}{\alpha-1}} & \epsilon_k > a + \alpha b \end{cases}. \tag{A9}$$

The ground-state energy for a given U is then obtained by optimizing the total energy over a and b.

We now consider the CA and CGA, which can be viewed in a unified way by writing the interaction energy functional as

$$E_{\text{int}} = U\left(\left(\frac{1}{2} - \frac{1}{4\theta}\right) - \frac{1}{\theta}\left(\int dk\sqrt{n_k(\theta - n_k)}\right)^2\right), \tag{A10}$$

where $\theta = 1$ corresponds to CA and $\theta = 2$ corresponds to CGA. We can similarly introduce two Lagrange multipliers, yielding a target functional

$$K = \int dk(\epsilon_k - a)n_k - b\int dk\sqrt{n_k(\theta - n_k)}, \tag{A11}$$

where $\theta \geqslant 1$ and the partially optimized value of n_k is given as

$$n_k = \begin{cases} 1 & \epsilon \leqslant a + \frac{b(\theta-2)}{2\sqrt{\theta-1}} \\ \frac{\theta}{2}\left(1 - \frac{(\epsilon_k - a)}{\sqrt{(\epsilon_k - a)^2 + b^2}}\right) & \epsilon > a + \frac{b(\theta-2)}{2\sqrt{\theta-1}} \end{cases}. \tag{A12}$$

For a given U, one can then optimize the Lagrange multipliers to obtain the ground-state energy and other observables. Plots of the double occupancy as a function of U/t are provided in Ref. [18].

An important drawback of the MBB, power, CA, and CGA functionals is that the antisymmetry of the two-body reduced density matrix (2RDM) is violated when $\mathcal{F}(n_i, n_j)$ deviates from the HF value [86]. The Piris natural orbital functionals (PNOF) [103] resolve this issue by approximating the form of the 2RDM $D_{ij,kl}^{\alpha\beta} = \frac{1}{2}\langle \hat{a}_{i\alpha}^\dagger \hat{a}_{j\beta}^\dagger \hat{a}_{l\beta} \hat{a}_{k\alpha}\rangle$ as [86]

$$D_{ij,kl}^{\sigma\sigma} = \frac{n_i n_j - \Delta_{ij}}{2}(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}), \quad \text{(A13)}$$

$$D_{ij,kl}^{\sigma\bar{\sigma}} = \frac{n_i n_j - \Delta_{ij}}{2}\delta_{ik}\delta_{jl} + \frac{\Pi_{ik}}{2}\delta_{ij}\delta_{kl}, \quad \text{(A14)}$$

where Δ_{ij} and Π_{ik} are explicitly defined in PNOF5 and PNOF7. In the previous calculations [86,104], it was found that the optimized natural orbital often breaks symmetry to further minimize the energy, and translational symmetry breaking in the Hubbard model allows PNOF5 and PNOF7 to recover the correct atomic limit [86]. In our context, we require translational symmetry to be respected, so we will examine PNOF5 and PNOF7 in this case. PNOF5 only includes intrapair correlation (i.e., $\Pi_{ik} = 0$ when i, k belong to different pairs), and therefore the deviation of the summation in Eq. (A7) from the corresponding Hartree-Fock value scales like L in the thermodynamic limit while there is a prefactor of $1/L^2$, and thus PNOF5 will yield the same energy form as HF in the thermodynamic limit. PNOF7 accounts for interpair correlation, which will provide a thermodynamic contribution, and for half-filling we find $E_{\text{int}} = \frac{1}{4} - \frac{1}{2}A^2$, which is similar to the CA but with different prefactor for A^2. However, PNOF7 was later renamed PNOF7(+), and another variant called PNOF7(−) was created [87], which will reproduce the CA result in the above case.

Finally, we discuss the interaction energy of the dimer functional at half-filling, given as

$$E_{\text{int}} = \frac{U}{4}(1 - \sqrt{1 - (K/K_0)^2}), \quad \text{(A15)}$$

where $K = 2\int dk\,\epsilon_k n_k$ and K_0 is the noninteracting kinetic energy. Minimizing the total energy over $\{n_k\}$ yields the double occupancy as a function of U as

$$d = \frac{1}{4}\left(1 - \sqrt{1 - \frac{1}{1 + (U/(4K_0))^2}}\right). \quad \text{(A16)}$$

APPENDIX B: EQUIVALENCE BETWEEN THE QUBIT ENERGY FORM FOR $\hat{\rho}_{G2}$ AND THE SLAVE SPIN MEAN FIELD THEORY

Here we prove that the slave spin mean-field theory (SSMF) is identical to the qubit energy form for ρ_{G2}. We begin by describing the SSMF using the conventions in our work, as there are trivial differences. We associate the spin state $|\uparrow\rangle$ with the fermionic state $|0\rangle$, while the standard SSMF associates $|\downarrow\rangle$ with $|0\rangle$. To restore the enlarged Hilbert space of the SSMF back to the physical space, we require that $\hat{n}_{i\ell} = \frac{1-\hat{\sigma}_{i\ell}^z}{2}$ holds for every local site i and spin-orbital

index ℓ. Then, the local Hamiltonian at site i can be written as $H_{\text{loc}}(\{\hat{n}_{i\ell}\}) = H_{\text{loc}}(\{\frac{1-\hat{\sigma}_{i\ell}^z}{2}\})$, while the hopping term can be written as $\hat{a}_{i\ell}^\dagger \hat{a}_{j\ell} = \hat{f}_{i\ell}^\dagger \hat{O}_{i\ell}^\dagger \hat{O}_{j\ell} \hat{f}_{j\ell}$, where $\hat{O}_{i\ell} = \hat{S}_{i\ell}^+ + c_{i\ell}\hat{S}_{i\ell}^-$, $\hat{S}_{i\ell}^\pm = \frac{\hat{\sigma}_{i\ell}^x \pm \hat{\sigma}_{i\ell}^y}{2}$, $c_{i\ell}$ is an arbitrary complex number, and $\hat{f}_{j\ell}$ is a fermionic annihilation operator. While the preceding transformations are exact for any $c_{i\ell}$, an appropriate choice of $c_{i\ell}$ is indeed critical when making mean-field approximations. In Ref. [49], which only addressed half-filling, the $c_{i\ell} = 1$ and $\hat{O}_{i\ell} = \hat{\sigma}_{i\ell}^x$, while later work generalized this result [53], choosing $c_{i\ell}$ such that the noninteracting limit can always be correctly recovered within the mean field approximation. We follow the latter choice. Finally, the Hamiltonian in the slave spin representation is given as

$$\hat{H} = \sum_{ij\ell} t_{i\ell,j\ell}\hat{f}_{i\ell}^\dagger \hat{O}_{i\ell}^\dagger \hat{O}_{j\ell}\hat{f}_{j\ell} + \sum_i H_{\text{loc}}\left(\left\{\frac{1 - \hat{\sigma}_{i\ell}^z}{2}\right\}\right), \quad \text{(B1)}$$

with the operator constraint $\hat{f}_{i\ell}^\dagger \hat{f}_{i\ell} = \frac{1-\hat{\sigma}_{i\ell}^z}{2}$ for every site and spin-orbital. There are two ways to derive the SSMF. The first approach introduces a mean-field decoupling for the hopping term and treats the operator constraint in the mean field level, yielding decoupled Hamiltonians for electrons and spins that must be solved self-consistently. The second approach [50,54,55,105] uses the variational principle, assuming a trial wave-function $|\Psi\rangle = |\Psi_f\rangle \otimes |\Psi_s\rangle$ which is a direct product of a fermionic part $|\Psi_f\rangle$ and a spin part $|\Psi_s\rangle$. Furthermore, the $|\Psi_f\rangle$ is assumed to be Slater determinant and the $|\Psi_s\rangle$ is assumed to be a direct product in real space, such that $|\Psi_s\rangle = \otimes_i|\Psi_{s;i}\rangle$. Therefore, the trial energy of Eq. (B1) is given as

$$E = \sum_{ij\ell} t_{i\ell,j\ell}\langle\Psi_f|\hat{f}_{i\ell}^\dagger \hat{f}_{j\ell}|\Psi_f\rangle\langle\Psi_{s;i}|\hat{O}_{i\ell}^\dagger|\Psi_{s;i}\rangle\langle\Psi_{s;j}|\hat{O}_{j\ell}|\Psi_{s;j}\rangle$$

$$+ \sum_i \langle\Psi_{s;i}|H_{\text{loc}}\left(\left\{\frac{1 - \hat{\sigma}_{i\ell}^z}{2}\right\}\right)|\Psi_{s;i}\rangle, \quad \text{(B2)}$$

subject to the constraint

$$\langle\Psi_f|\hat{f}_{i\ell}^\dagger \hat{f}_{i\ell}|\Psi_f\rangle = \langle\Psi_{s;i}|\frac{1 - \hat{\sigma}_{i\ell}^z}{2}|\Psi_{s;i}\rangle = n_\ell. \quad \text{(B3)}$$

Given translational symmetry and the correspondence $\boldsymbol{\rho} \leftrightarrow |\Psi_{s;j}\rangle\langle\Psi_{s;j}|$ and $n_{k\ell;0} \leftrightarrow \langle\Psi_f|f_{k\ell}^\dagger f_{k\ell}|\Psi_f\rangle$, we can prove that trial energy in Eq. (B2) is identical to the qubit trial energy for $\hat{\rho}_{G2}$ given in Eq. (50) and the constraint in Eq. (B3) is identical to Eq. (54). The key step is to evaluate

$$\langle\Psi_{s;j}|\hat{O}_{j\ell}|\Psi_{s;j}\rangle = (1 + c_\ell)\langle\tfrac{1}{2}\hat{\sigma}_\ell^x\rangle_{\boldsymbol{\rho}}, \quad \text{(B4)}$$

where gauge symmetry allows $\langle\hat{\sigma}_\ell^y\rangle_{\boldsymbol{\rho}} = 0$. To ensure $\langle\Psi_{s;j}|\hat{O}_{j\ell}|\Psi_{s;j}\rangle = 1$ in the noninteracting limit, we must choose c_ℓ such that $(1 + c_\ell)\xi_{\ell;0} = 1$, where $\xi_{\ell;0} = \sqrt{n_\ell(1 - n_\ell)}$ is the value for $\xi_\ell = \langle\tfrac{1}{2}\hat{\sigma}_\ell^x\rangle_{\boldsymbol{\rho}}$ in the noninteracting limit. Plugging c_ℓ into Eq. (B4), we have $\langle\Psi_{s;j}|\hat{O}_{j\ell}|\Psi_{s;j}\rangle = \xi_\ell/\xi_{\ell;0}$, completing the proof.

The preceding proof demonstrates the equivalence of the qubit energy form for $\hat{\rho}_{G2}$ and the trial energy for the SSMF. Now we demonstrate how to obtain a Hamiltonian form for the SSMF using the saddle-point equations for the trial energy. The total energy under a fixed $\{n_\ell\}$ can be minimized by introducing Lagrange multipliers for the electron and spin

systems, yielding

$$E = \frac{1}{L} \sum_{k\ell} \epsilon_{k\ell} \left(n_\ell + \frac{\xi_\ell^2 (n_{k\ell,0} - n_\ell)}{n_\ell (1 - n_\ell)} \right) + \langle H_{\text{loc}}(\{\hat{n}_\ell\}) \rangle_\rho$$

$$- \sum_\ell \left(\lambda_\ell \left(\frac{1}{L} \sum_k n_{k\ell;0} - n_\ell \right) + \lambda_\ell' (\langle \hat{n}_\ell \rangle_\rho - n_\ell) \right). \tag{B5}$$

Taking the derivative with respect to $\boldsymbol{\rho}$ yields the mean-field Hamiltonian for the spin system as

$$\hat{H}_{\text{eff}}^s = \frac{\partial E}{\partial \boldsymbol{\rho}} = H_{\text{loc}}(\{\hat{n}_\ell\}) + 2 \sum_\ell h_\ell \hat{\xi}_\ell - \lambda_\ell' \hat{n}_\ell + C, \tag{B6}$$

where $\hat{n}_\ell = \frac{1}{2}(1 - \hat{\sigma}_\ell^z)$, $\hat{\xi}_\ell = \frac{1}{2}\hat{\sigma}_\ell^x$, $h_\ell = \frac{1}{L} \frac{\xi_\ell}{n_\ell(1-n_\ell)} \sum_k \epsilon_{k\ell} n_{k\ell,0}$, C is a constant that does not influence the results, and we assume $\sum_k \epsilon_{k\ell} = 0$. Now consider the derivative of the electron components as

$$\frac{\partial E}{\partial n_{k\ell;0}} = Z_\ell \epsilon_{k\ell} - \lambda_\ell, \tag{B7}$$

where $Z_\ell = \frac{\xi_\ell^2}{n_\ell(1-n_\ell)}$, which can be connected to a noninteracting fermionic Hamiltonian as

$$\hat{H}_{\text{eff}}^f = \sum_{k\ell} (Z_\ell \epsilon_{k\ell} - \lambda_\ell) \hat{n}_{k\ell}. \tag{B8}$$

The ground states for Eqs. (B6) and (B8) will yield an updated $\boldsymbol{\rho}$ and $n_{k\ell}$, yielding new Hamiltonians for Eqs. (B6) and (B8), and this procedure is iterated until self-consistency is achieved.

APPENDIX C: CENTRAL POINT EXPANSION

The CPE can be viewed as an approach to evaluate both a G-type and B-type SPD at $\mathcal{N} = 2$ by expanding about a reference SPD referred to as the central point. The CPE can be applied in arbitrary dimensions, and it is formally exact if all orders are summed, though it has only ever been applied at first-order. Interestingly, for the G-type $\mathcal{N} = 2$ SPD, the first-order CPE yields the same result as the GA and the SCDA in any dimension, and the derivation offers an alternative perspective from the GA and the SCDA. For the B-type $\mathcal{N} = 2$ SPD, the first-order CPE yields the same result as the SCDA in $d = \infty$, while providing a different approximation in finite dimensions. The CPE was originally developed in the context of the OET [56], where the CPE was renormalized using both weak coupling and strong coupling perturbation theory to ensure the correct limiting behavior, yielding excellent results for the Hubbard model in $d = 1, 2, \infty$.

1. CPE for $\hat{\rho}_{G2}$

In this section, we use the first-order CPE to evaluate $\hat{\rho}_{G2}$, which will be shown to be equivalent to the GA and the SCDA. The first-order CPE can be motivated by the fact that the GA relation between $n_{k\ell} - n_\ell$ and $n_{k\ell;0} - n_\ell$ is a linear form [see Eq. (111)] given by

$$n_{k\ell} - n_\ell = \mathcal{Z}_\ell (n_{k\ell;0} - n_\ell), \tag{C1}$$

which is valid for an arbitrary $n_{k\ell;0} \in [0, 1]$ with a constraint $(1/L) \sum_k n_{k\ell;0} = n_\ell$, where L is the number of sites in the lattice. This remarkably simple relation motivates the use of $n_{k\ell;0} - n_\ell$ as the expansion parameters, where $\mathcal{Z}_\ell$ can be determined using a $\hat{\rho}_{G2}$ with $n_{k\ell;0}$ slightly deviating from the uniform distribution n_ℓ at first-order. Alternatively, the relation between $n_{k\ell;0}$ and $\gamma_{k\ell}$ is highly nonlinear, suggesting that a first-order approximation in $\gamma_{k\ell}$ cannot be applied to the Gutzwiller wavefunction.

The CPE begins by choosing an expansion point for $\hat{\rho}_{G2} = \hat{P}_1 \hat{K}_2 \hat{P}_1$ as $\hat{\rho}_{G2}^\star = \hat{P}_1 \hat{K}_2^\star \hat{P}_1$, where $\hat{K}_2^\star = \exp(\sum_{k\ell} \gamma_\ell^\star \hat{n}_{k\ell})$ and $\gamma_\ell^\star$ is chosen to reproduce the noninteracting local density $n_{\ell;0} \equiv \langle \hat{n}_{i\ell} \rangle_{\hat{K}_2} = \langle \hat{n}_{i\ell} \rangle_{\hat{K}_2^\star}$ through $n_{\ell;0} = 1/(1 + \exp(-\gamma_\ell^\star))$. Considering a kinetic projector that deviates slightly from $\hat{K}_2^\star$ as $\hat{K}_2 = \exp(\sum_{k\ell} \gamma_{k\ell} \hat{n}_{k\ell})$, where $\gamma_{k\ell} = \gamma_\ell^\star + \delta\gamma_{k\ell}$, we compute the response of the expectation value $\langle \hat{O} \rangle_{\hat{\rho}_{G2}}$ to $\gamma_{k\ell}$ to the first-order about the central point, given as

$$\left. \frac{\partial \langle \hat{O} \rangle_{\hat{\rho}_{G2}}}{\partial \gamma_{k\ell}} \right|_{\hat{\rho}_{G2} = \hat{\rho}_{G2}^\star} = \left. \frac{\partial}{\partial \gamma_{k\ell}} \frac{\text{Tr}(\hat{P}_1 \hat{K}_2 \hat{P}_1 \hat{O})}{\text{Tr}(\hat{P}_1 \hat{K}_2 \hat{P}_1)} \right|_{\hat{\rho}_{G2} = \hat{\rho}_{G2}^\star} \tag{C2}$$

$$= \frac{\text{Tr}(\hat{P}_1 \sqrt{\hat{K}_2^\star} \hat{n}_{k\ell} \sqrt{\hat{K}_2^\star} \hat{P}_1 \hat{O})}{\text{Tr}(\hat{\rho}_{G2}^\star)}$$

$$- \frac{\text{Tr}(\hat{P}_1 \sqrt{\hat{K}_2^\star} \hat{n}_{k\ell} \sqrt{\hat{K}_2^\star} \hat{P}_1) \text{Tr}(\hat{\rho}_{G2}^\star \hat{O})}{\text{Tr}(\hat{\rho}_{G2}^\star)^2} \tag{C3}$$

$$= \langle \hat{n}_{k\ell}; \hat{O} \rangle_{\hat{\rho}_{G2}^\star} - \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}^\star} \langle \hat{O} \rangle_{\hat{\rho}_{G2}^\star}, \tag{C4}$$

where we have utilized that fact that $\hat{K}_2^\star$ commutes with $\hat{P}_1$ [56], $\sqrt{\hat{\rho}_{G2}^\star} = \hat{P}_1 \sqrt{\hat{K}_2^\star} = \sqrt{\hat{K}_2^\star} \hat{P}_1$, and the notation $\langle \hat{A}; \hat{B} \rangle_{\hat{\rho}}$ is defined as

$$\langle \hat{A}; \hat{B} \rangle_{\hat{\rho}} = \langle \hat{B}; \hat{A} \rangle_{\hat{\rho}} = \text{Tr}(\sqrt{\hat{\rho}} \hat{A} \sqrt{\hat{\rho}} \hat{B}) / \text{Tr}(\hat{\rho}). \tag{C5}$$

The response coefficient of $\langle \hat{O} \rangle_{\hat{\rho}_{G2}}$ to $\gamma_{k\ell}$ at a general $\hat{\rho}_{G2}$ can be conveniently expressed via the integer time correlation function in the compound space as

$$\frac{\partial \langle \hat{O} \rangle_{\hat{\rho}_{G2}}}{\partial \gamma_{k\ell}} = \langle \underline{\hat{n}}_{k\ell}^{(1)} \underline{\hat{O}}^{(2)} \rangle_{\hat{\rho}_{G2}} - \langle \underline{\hat{n}}_{k\ell}^{(1)} \rangle_{\hat{\rho}_{G2}} \langle \underline{\hat{O}}^{(2)} \rangle_{\hat{\rho}_{G2}}. \tag{C6}$$

While Eq. (C6) is difficult to evaluate in general, the CPE circumvents this problem by observing that Eq. (C6) becomes trivial to evaluate at the central point given that $\hat{\rho}_{G2}^\star$ and $\sqrt{\hat{\rho}_{G2}^\star}$ are direct product states in real space. Though it is not yet clear, it will prove critical to exploit the gauge symmetry of the SPD to restrict $\hat{P}_1$ such that $\langle \hat{n}_{i\ell} \rangle_{\hat{\rho}_{G2}^\star} = \langle \hat{n}_{i\ell} \rangle_{\hat{K}_2^\star} = n_{\ell;0}$. The response for the momentum density distribution $n_{k\ell}$ to $\gamma_{k'\ell'}$ can be obtained as

$$\left. \frac{\partial \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}}}{\partial \gamma_{k'\ell'}} \right|_{\hat{\rho}_{G2} = \hat{\rho}_{G2}^\star} = \langle \hat{n}_{k\ell}; \hat{n}_{k'\ell'} \rangle_{\hat{\rho}_{G2}^\star} - n_{\ell;0} n_{\ell';0}, \tag{C7}$$

where the correlation function $\langle \hat{n}_{k\ell}; \hat{n}_{k'\ell'} \rangle_{\hat{\rho}_{G2}^\star}$ can be computed by transforming to real space as

$$\langle \hat{n}_{k\ell}; \hat{n}_{k'\ell'} \rangle_{\hat{\rho}_{G2}^\star} = \frac{1}{L^2} \sum_{j_1 j_2 j_3 j_4} e^{ik(j_1 - j_2) + ik'(j_3 - j_4)}$$

$$\times \langle \hat{a}_{j_1\ell}^\dagger \hat{a}_{j_2\ell}; \hat{a}_{j_3\ell'}^\dagger \hat{a}_{j_4\ell'} \rangle_{\hat{\rho}_{G2}^\star}. \tag{C8}$$

Utilizing the fact that $\hat{\rho}_{G2}^\star = \otimes_i \hat{\rho}_{G2;i}^\star$, where $\hat{\rho}_{G2;i}^\star$ is the local reduced density matrix for $\hat{\rho}_{G2}^\star$, then $\langle \hat{a}_{j_1\ell}^\dagger \hat{a}_{j_2\ell}; \hat{a}_{j_3\ell'}^\dagger \hat{a}_{j_4\ell'} \rangle_{\hat{\rho}_{G2}^\star}$ is only nonzero in the following cases: (1) when $j_1 = j_2$, $j_3 = j_4$, and $j_1 \neq j_4$, the value is $n_{\ell;0}n_{\ell';0}$; (2) when $j_1 = j_4$, $j_2 = j_3$, $j_1 \neq j_2$, and $\ell = \ell'$, the value is A_ℓ^2, where $A_\ell \equiv \langle \hat{a}_{j\ell}^\dagger; \hat{a}_{j\ell'} \rangle_{\hat{\rho}_{G2}^\star} = \langle \hat{a}_{j\ell}; \hat{a}_{j\ell'}^\dagger \rangle_{\hat{\rho}_{G2}^\star}$; (3) when $j_1 = j_2 = j_3 = j_4$, and $\ell = \ell'$, the value is $n_{\ell;0} = A_{\ell;0}^2 - n_{\ell;0}^2$, where $A_{\ell;0} \equiv \langle \hat{a}_{j\ell}^\dagger; \hat{a}_{j\ell} \rangle_{\hat{K}_2^\star} = \sqrt{(1 - n_{\ell;0})n_{\ell;0}}$; (4) when $j_1 = j_2 = j_3 = j_4$, and $\ell \neq \ell'$, the value is $\langle \hat{n}_{i\ell}\hat{n}_{i\ell'} \rangle_{\hat{\rho}_{G2}^\star}$. Combining these four cases, we have

$$\langle \hat{a}_{j_1\ell}^\dagger \hat{a}_{j_2\ell}; \hat{a}_{j_3\ell'}^\dagger \hat{a}_{j_4\ell'} \rangle_{\hat{\rho}_{G2}^\star} = \delta_{j_1 j_2}\delta_{j_3 j_4} n_{\ell;0}n_{\ell';0}$$
$$+ \delta_{j_1 j_4}\delta_{j_2 j_3}\delta_{\ell\ell'}A_\ell^2 + \delta_{j_1 j_4}\delta_{j_2 j_3}\delta_{j_1 j_3}C_{\ell\ell'}, \tag{C9}$$

where $C_{\ell\ell'}$ is defined as

$$C_{\ell\ell'} = \begin{cases} A_{\ell;0}^2 - A_\ell^2 & \ell = \ell' \\ \langle \hat{n}_{i\ell}\hat{n}_{i\ell'} \rangle_{\hat{\rho}_{G2}^\star} - n_{\ell;0}n_{\ell';0} & \ell \neq \ell' \end{cases}. \tag{C10}$$

Therefore, the response coefficient of $n_{k\ell}$ to $\gamma_{k'\ell'}$ is given as

$$\frac{\partial \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}}}{\partial \gamma_{k'\ell'}} \bigg|_{\hat{\rho}_{G2}=\hat{\rho}_{G2}^\star} = \delta_{\ell\ell'}\delta_{kk'}A_\ell^2 + C_{\ell\ell'}/L, \tag{C11}$$

which consists of two contributions: (1) the coherent contribution $\delta_{\ell\ell'}\delta_{kk'}A_\ell^2$, which reflects the hopping renormalization captured in the GA, and (2) the incoherent contribution $C_{\ell\ell'}/L$, which has no contribution to the momentum density distribution given that the constraint $\langle \hat{n}_{i\ell} \rangle_{\hat{K}_2} = \langle \hat{n}_{i\ell} \rangle_{\hat{K}_2^\star}$ requires

$$\sum_k \delta\gamma_{k\ell} = 0. \tag{C12}$$

When applying Eq. (C11) with $\hat{P}_1 \rightarrow 1$ such that $\hat{\rho}_{G2}^\star \rightarrow \hat{K}_2^\star$, the response for the bare momentum density distribution $n_{k\ell;0}$ to $\gamma_{k'\ell'}$ is given by

$$\frac{\partial \langle \hat{n}_{k\ell} \rangle_{\hat{K}_2}}{\partial \gamma_{k'\ell'}} \bigg|_{\hat{K}_2=\hat{K}_2^\star} = \delta_{\ell\ell'}\delta_{kk'}A_{\ell;0}^2. \tag{C13}$$

The next stage is to find the relation between $\delta n_{k\ell} = \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}} - \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}^\star}$ and $\delta n_{k\ell;0} = \langle \hat{n}_{k\ell} \rangle_{\hat{K}_2} - \langle \hat{n}_{k\ell} \rangle_{\hat{K}_2^\star}$. We first compute the response of $\delta n_{k\ell}$ and $\delta n_{k\ell;0}$ for a given $\delta\gamma_{k\ell}$ with the constraint given by Eq. (C12), and then solve $\delta\gamma_{k\ell}$ from $\delta n_{k\ell;0}$ and express $\delta n_{k\ell}$ in terms of $\delta n_{k\ell;0}$, given as

$$n_{k\ell} - n_{\ell;0} = \frac{A_\ell^2}{A_{\ell;0}^2}(n_{k\ell;0} - n_{\ell;0}). \tag{C14}$$

The constraint given by Eq. (C12) for $\delta\gamma_{k\ell}$ naturally yields the following relation:

$$n_\ell \equiv \frac{1}{L}\sum_k n_{k\ell} = n_{\ell;0} = \frac{1}{L}\sum_k n_{k\ell;0}, \tag{C15}$$

which is the constraint imposed in the GA.

Finally, to connect Eq. (C14) with Eq. (107), we need to show $\mathcal{R}_\ell = A_\ell/A_{\ell;0}$, which can be accomplished by proving that $\hat{\rho}_{G2}$ and $\hat{\rho}_{G2}^\star$ have the same local reduced density matrix. To prove this, we need to consider the expectation of a given Hubbard operator under $\hat{\rho}_{G2}$, which can be accomplished by

considering the linear response of a general diagonal Hubbard operator $\hat{X}_{i\Gamma}$ at site i to $\gamma_{k\ell}$ as

$$\frac{\partial \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}}}{\partial \gamma_{k\ell}} \bigg|_{\hat{\rho}_{G2}=\hat{\rho}_{G2}^\star} = \langle \hat{X}_{i\Gamma}; \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}^\star} - \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}^\star}n_\ell, \tag{C16}$$

$$\langle \hat{X}_{i\Gamma}; \hat{n}_{k\ell} \rangle_{\hat{\rho}_{G2}^\star} = \frac{1}{L}\sum_{jj'} e^{ik(j-j')}\langle \hat{X}_{i\Gamma}; \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{\rho}_{G2}^\star}, \tag{C17}$$

$$\langle \hat{X}_{i\Gamma}; \hat{a}_{j\ell}^\dagger \hat{a}_{j'\ell} \rangle_{\hat{\rho}_{G2}^\star} = \delta_{jj'}\langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}^\star}n_\ell + \delta_{jj'}\delta_{ij}C_{\Gamma\ell}, \tag{C18}$$

$$C_{\Gamma\ell} = \langle \hat{X}_{i\Gamma}\hat{n}_{i\ell} \rangle_{\hat{\rho}_{G2}^\star} - \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}^\star}n_\ell, \tag{C19}$$

$$\frac{\partial \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}}}{\partial \gamma_{k\ell}} \bigg|_{\hat{\rho}_{G2}=\hat{\rho}_{G2}^\star} = \frac{1}{L}C_{\Gamma\ell}. \tag{C20}$$

Given the constraint on $\gamma_{k\ell}$ [see Eq. (C12)], the first-order change in $\langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}}$ is zero and therefore we have

$$\langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}} = \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{G2}^\star}. \tag{C21}$$

Similarly, with $\hat{P}_1 \rightarrow 1$ and $\hat{\rho}_{G2} \rightarrow \hat{K}_2$, we have

$$\langle \hat{X}_{i\Gamma} \rangle_{\hat{K}_2} = \langle \hat{X}_{i\Gamma} \rangle_{\hat{K}_2^\star}. \tag{C22}$$

Therefore, we have recovered Eq. (102). Moreover, we have

$$\frac{A_\ell}{A_{\ell;0}} = \frac{\text{Tr}(\sqrt{\hat{\rho}_{G2}^\star}\hat{a}_{i\ell}^\dagger\sqrt{\hat{\rho}_{G2}^\star}\hat{a}_{i\ell})/\text{Tr}(\hat{\rho}_{G2}^\star)}{\text{Tr}(\sqrt{\hat{K}_2^\star}\hat{a}_{i\ell}^\dagger\sqrt{\hat{K}_2^\star}\hat{a}_{i\ell})/\text{Tr}(\hat{K}_2^\star)} = \mathcal{R}_\ell, \tag{C23}$$

where we have used the fact that the local reduced density matrix of $\hat{\rho}_{G2}^\star$ is same as the local reduced density matrix of $\hat{\rho}_{G2}$, as indicated in Eq. (C21), and represented as ρ_{loc} in Eq. (105). Similarly, the local reduced density matrix of $\hat{K}_2^\star$ is the same as the local reduced density matrix of $\hat{K}_2$, as indicated in Eq. (C22), and represented as $\rho_{\text{loc};0}$ in Eq. (106).

In summary, we have demonstrated that the first-order CPE is equivalent to the GA and the SCDA. The key steps in the proof include: (1) the local reduced density matrix for $\hat{\rho}_{G2}$ is the same as for $\hat{\rho}_{G2}^\star$, with no dependency on the details of $n_{k\ell;0}$ except its average value and (2) the momentum density distribution $n_{k\ell}$ is uniformly shrunk toward n_ℓ through $\mathcal{Z}_\ell$, which is uniquely determined by $\hat{\rho}_{G2}^\star$. Finally, it would be interesting to explore the behavior of the CPE beyond first-order in finite dimensions, which provides insight beyond the GA for evaluating $\hat{\rho}_{G2}$.

2. CPE for $\hat{\rho}_{B2}$

In this section, we use the first-order CPE to evaluate $\hat{\rho}_{B2} = \hat{K}_1\hat{P}_1\hat{K}_1$, where $\hat{P}_1 = \exp(\sum_{i\Gamma} \upsilon_\Gamma \hat{X}_{i\Gamma})$ and $\hat{K}_1 = \exp(\sum_{k\ell} \gamma_{k\ell}\hat{n}_{k\ell})$, demonstrating the equivalence to the SCDA in $d = \infty$. The CPE for $\hat{\rho}_{B2}$ can be viewed as a dual version of the CPE for the $\hat{\rho}_{G2}$, which has various correspondences [56]. First, in the CPE for the $\hat{\rho}_{G2}$, the central projector $\hat{K}_2$ determines the local density, and is invariant after applying the interacting projector $\hat{P}_1$. Correspondingly, in the CPE for $\hat{\rho}_{B2}$, the central projector $\hat{P}_1$ determines the local density, and is invariant after applying $\hat{K}_1$. Second, in the CPE for the $\hat{\rho}_{G2}$, the physical momentum density distribution is linearly related to the bare momentum density distribution, and the local reduced density matrix is independent of the details of the momentum density distribution. Correspondingly, in the

CPE for the $\hat{\rho}_{B2}$, the local reduced density matrix is linearly related to the reference local reduced density matrix, and the momentum density distribution is independent of the details of the local reduced density matrix.

We proceed in evaluating $\hat{\rho}_{B2}$ via the CPE by choosing the central point for $\hat{\rho}_{B2}$ as $\hat{\rho}_{B2}^{\star} = \hat{K}_1 \hat{P}_1^{\star} \hat{K}_1$, where $\hat{P}_1^{\star} = \exp(\sum_{i\ell} \gamma_\ell^{\star} \hat{n}_{i\ell})$ and $\gamma_\ell^{\star}$ is chosen such that $n_\ell^{\star} \equiv \langle \hat{n}_{i\ell} \rangle_{\hat{P}_1} = \langle \hat{n}_{i\ell} \rangle_{\hat{P}_1^{\star}}$. We can also rewrite $\hat{P}_1^{\star} = \exp(\sum_{i\Gamma} \upsilon_\Gamma^{\star} \hat{X}_{i\Gamma})$ by expressing $\hat{n}_{i\ell}$ as a linear combination of Hubbard operators, where $\upsilon_\Gamma^{\star} = \sum_\ell \gamma_\ell^{\star} \Gamma(\ell)$. Considering the linear response of υ_Γ about $\upsilon_\Gamma^{\star}$, similar to Eqs. (C2)–(C4), we have

$$\left. \frac{\partial \langle \hat{O} \rangle_{\hat{\rho}_{B2}}}{\partial \upsilon_{i\Gamma}} \right|_{\hat{\rho}_{B2}\hat{\rho}_{B2}^{\star}} = \frac{\partial}{\partial \upsilon_{i\Gamma}} \frac{\text{Tr}(\hat{K}_1 \hat{P}_1 \hat{K}_1 \hat{O})}{\text{Tr}(\hat{K}_1 \hat{P}_1 \hat{K}_1)} \tag{C24}$$

$$= \frac{\text{Tr}(\hat{K}_1 \sqrt{\hat{P}_1^{\star}} \hat{X}_{i\Gamma} \sqrt{\hat{P}_1^{\star}} \hat{K}_1 \hat{O})}{\text{Tr}(\hat{\rho}_{B2}^{\star})}$$

$$- \frac{\text{Tr}(\hat{K}_1 \sqrt{\hat{P}_1^{\star}} \hat{X}_{i\Gamma} \sqrt{\hat{P}_1^{\star}} \hat{K}_1) \text{Tr}(\hat{\rho}_{B2}^{\star} \hat{O})}{\text{Tr}(\hat{\rho}_{B2}^{\star})^2} \tag{C25}$$

$$= \langle \hat{X}_{i\Gamma}; \hat{O} \rangle_{\hat{\rho}_{B2}^{\star}} - \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{B2}^{\star}} \langle \hat{O} \rangle_{\hat{\rho}_{B2}^{\star}}, \tag{C26}$$

where we similarly utilize the fact that $\hat{P}_1^{\star}$ commutes with $\hat{K}_1$ and $\sqrt{\hat{\rho}_{B2}^{\star}} = \hat{K}_1 \sqrt{\hat{P}_1^{\star}} = \sqrt{\hat{P}_1^{\star}} \hat{K}_1$. We proceed by computing the response

$$\left. \frac{\partial \langle \hat{X}_{i'\Gamma'} \rangle_{\hat{\rho}_{B2}}}{\partial \upsilon_{i\Gamma}} \right|_{\hat{\rho}_{B2}=\hat{\rho}_{B2}^{\star}} = \langle \hat{X}_{i\Gamma}; \hat{X}_{i'\Gamma'} \rangle_{\hat{\rho}_{B2}^{\star}} - \langle \hat{X}_{i\Gamma} \rangle_{\hat{\rho}_{B2}^{\star}} \langle \hat{X}_{i'\Gamma'} \rangle_{\hat{\rho}_{B2}^{\star}}. \tag{C27}$$

Directly computing Eq. (C27) is cumbersome, and this can be avoided by decomposing $\hat{X}_{i\Gamma}$ into an alternate form using Eq. (3), resulting in

$$\hat{X}_{i\Gamma} = \prod_\ell \hat{X}_{i\Gamma;\ell} = \prod_\ell (X_{i\Gamma;\ell}^{\star} + \delta \hat{X}_{i\Gamma;\ell}), \tag{C28}$$

where

$$X_{i\Gamma;\ell}^{\star} \equiv \langle \hat{X}_{i\Gamma;\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = (1 - n_\ell^{\star}) \delta_{0,\Gamma(\ell)} + n_\ell^{\star} \delta_{1,\Gamma(\ell)}, \tag{C29}$$

$$\delta \hat{X}_{i\Gamma;\ell} \equiv \hat{X}_{i\Gamma;\ell} - X_{i\Gamma;\ell}^{\star} = (\hat{n}_{i\ell} - n_\ell^{\star})(\delta_{1,\Gamma(\ell)} - \delta_{0,\Gamma(\ell)}). \tag{C30}$$

We introduce the density fluctuation operator $\delta \hat{D}_{iI} = \prod_{\ell \in I} \delta \hat{n}_{i\ell}$, where $\delta \hat{n}_{i\ell} = \hat{n}_{i\ell} - n_\ell^{\star}$, and $\delta \hat{D}_I = \hat{1}$ when $I = \{\}$, and we use $\delta \hat{D}_{iI}$ as a new basis for the CPE expansion. The diagonal Hubbard operator $\hat{X}_{i\Gamma}$ can be written as

$$\hat{X}_{i\Gamma} = \sum_I \left(\prod_{\ell \in \bar{I}} X_{i\Gamma;\ell}^{\star} \right) \left(\prod_{\ell \in I} (\delta_{1,\Gamma(\ell)} - \delta_{0,\Gamma(\ell)}) \right) \delta \hat{D}_{iI}, \tag{C31}$$

where I enumerates over all subsets of $\{1, 2, \ldots, 2N_{\text{orb}}\}$ and $\bar{I} = \{1, 2, \ldots, N_{\text{orb}}\} - I$. The interacting projector can be written as $\hat{P}_1 = \exp(\sum_{iI} \eta_{iI} \delta \hat{D}_{iI})$, and at the central point $\eta_{i\{\}}^{\star} = \gamma_\ell^{\star}$, with $\eta_{i\{\}}^{\star}$ absorbing the constant contribution, while $\eta_{iI}^{\star} = 0$ if $|I| > 1$. Similar to Eq. (C26), for $|I| > 0$ we have

$$\left. \frac{\partial \langle \hat{O} \rangle_{\hat{\rho}_{B2}}}{\partial \eta_{iI}} \right|_{\hat{\rho}_{B2}=\hat{\rho}_{B2}^{\star}} = \langle \delta \hat{D}_{iI}; \hat{O} \rangle_{\hat{\rho}_{B2}^{\star}}, \tag{C32}$$

given that $\langle \delta \hat{D}_{iI} \rangle_{\hat{\rho}_{B2}} = 0$ in this case. It should be noted that $\eta_{i\{\}}$ has no effect on the expectation values so we implicitly only consider cases with $|I| > 0$. The response of $\langle \delta \hat{D}_{i'I'} \rangle_{\hat{\rho}_{B2}}$ to η_{iI} is given as

$$\left. \frac{\partial \langle \delta \hat{D}_{i'I'} \rangle_{\hat{\rho}_{B2}}}{\partial \eta_{iI}} \right|_{\hat{\rho}_{B2}=\hat{\rho}_{B2}^{\star}} = \langle \delta \hat{D}_{iI}; \delta \hat{D}_{i'I'} \rangle_{\hat{\rho}_{B2}^{\star}}. \tag{C33}$$

Since $\hat{\rho}_{B2}^{\star}$ is diagonal in ℓ, the correlation function on the right side of Eq. (C33) is only nonzero when $I = I'$, and it can be written as the product of contributions from each relevant spin-orbital as

$$\langle \delta \hat{D}_{iI}; \delta \hat{D}_{i'I'} \rangle_{\hat{\rho}_{B2}^{\star}} = \delta_{II'} \prod_{\ell \in I} \langle \delta \hat{n}_{i\ell}; \delta \hat{n}_{i'\ell} \rangle_{\hat{\rho}_{B2}^{\star}}, \tag{C34}$$

requiring the evaluation of

$$\langle \delta \hat{n}_{i\ell}; \delta \hat{n}_{i'\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = \langle \hat{n}_{i\ell}; \hat{n}_{i'\ell} \rangle_{\hat{\rho}_{B2}^{\star}} - (n_\ell^{\star})^2, \tag{C35}$$

and $\langle \hat{n}_{i\ell}; \hat{n}_{i'\ell} \rangle_{\hat{\rho}_{B2}^{\star}}$ can be evaluated similarly to Eq. (C8) as

$$\langle \hat{n}_{j\ell}; \hat{n}_{j'\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = \frac{1}{L^2} \sum_{k_1 k_2 k_3 k_4} e^{i(k_2-k_1)j + i(k_4-k_3)j'}$$

$$\times \langle \hat{a}_{k_1\ell}^{\dagger} \hat{a}_{k_2\ell}; \hat{a}_{k_3\ell}^{\dagger} \hat{a}_{k_4\ell} \rangle_{\hat{\rho}_{B2}^{\star}}. \tag{C36}$$

Given that $\hat{\rho}_{B2}^{\star} = \otimes_{k\ell} \hat{\rho}_{B2;k\ell}^{\star}$ is a direct product state in momentum space, $\langle \hat{a}_{k_1\ell}^{\dagger} \hat{a}_{k_2\ell}; \hat{a}_{k_3\ell}^{\dagger} \hat{a}_{k_4\ell} \rangle_{\hat{\rho}_{B2}^{\star}}$ is only nonzero in the following cases. (1) $k_1 = k_2$, $k_3 = k_4$, and $k_1 \neq k_4$ results in $n_{k_1\ell}^{\star} n_{k_3\ell}^{\star}$, where $n_{k\ell}^{\star} = \langle \hat{n}_{k\ell} \rangle_{\hat{\rho}_{B2}^{\star}}$, (2) $k_1 = k_4$, $k_2 = k_3$, and $k_1 \neq k_2$ results in $A_{k_1\ell} A_{k_2\ell}$, where $A_{k\ell} \equiv \langle \hat{a}_{k\ell}^{\dagger}; \hat{a}_{k\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = \langle \hat{a}_{k\ell}; \hat{a}_{k\ell}^{\dagger} \rangle_{\hat{\rho}_{B2}^{\star}}$ and $A_{k\ell} = \sqrt{n_{k\ell}^{\star}(1 - n_{k\ell}^{\star})}$, (3) $k_1 = k_2 = k_3 = k_4$ results in $n_{k\ell}^{\star} = (n_{k\ell}^{\star})^2 + A_{k\ell}^2$. In summary, we have

$$\langle \hat{a}_{k_1\ell}^{\dagger} \hat{a}_{k_2\ell}; \hat{a}_{k_3\ell}^{\dagger} \hat{a}_{k_4\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = \delta_{k_1 k_2} \delta_{k_3 k_4} n_{k_1\ell}^{\star} n_{k_3\ell}^{\star} + \delta_{k_1 k_4} \delta_{k_2 k_3} A_{k_1\ell} A_{k_2\ell}, \tag{C37}$$

which can be used to obtain

$$\langle \hat{n}_{j\ell}; \hat{n}_{j'\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = \left(\frac{1}{L} \sum_k n_{k\ell}^{\star} \right)^2 + \left| \frac{1}{L} \sum_k e^{ik(j-j')} A_{k\ell} \right|^2,$$

where L is the number of sites in the lattice. A constraint is imposed on $\hat{K}_1$ such that

$$\langle \hat{n}_{i\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = n_\ell^{\star}, \tag{C38}$$

resulting in $(1/L) \sum_k n_{k\ell}^{\star} = n_\ell^{\star}$ and

$$\langle \delta \hat{n}_{j\ell}; \delta \hat{n}_{j'\ell} \rangle_{\hat{\rho}_{B2}^{\star}} = \left| \frac{1}{L} \sum_k e^{ik(j-j')} A_{k\ell} \right|^2. \tag{C39}$$

In infinite dimensions, only local contributions need to be accounted for, which is used in the following steps. Considering the response $\delta D_{iI} = \langle \delta \hat{D}_{iI} \rangle_{\hat{\rho}_{B2}}$ and $\delta D_{iI;0} = \langle \delta \hat{D}_{iI} \rangle_{\hat{P}_1}$ to first-order in η_{iI} and solving for η_{iI} as a function of $\delta D_{iI;0}$, the response in infinite dimensions is

$$\delta D_{iI} = \left(\prod_{\ell \in I} \mathcal{F}_\ell \right) \delta D_{iI;0}, \tag{C40}$$

where the renormalization factor is given by

$$\mathcal{F}_\ell \equiv \left(\sum_k A_{k\ell}\right)^2 \bigg/ \left(\sum_k A_{k\ell;0}\right)^2, \qquad \text{(C41)}$$

where $A_{k\ell} = \sqrt{n_{k\ell}^\star(1-n_{k\ell}^\star)}$ and $A_{k\ell;0} = \sqrt{n_\ell^\star(1-n_\ell^\star)}$.

Finally, we need to consider how the momentum density distribution is influenced by η_{iI} using the response

$$\frac{\partial \langle \hat{n}_{k\ell}\rangle_{\hat{\rho}_{B2}}}{\partial \eta_{iI}}\bigg|_{\hat{\rho}_{B2}=\hat{\rho}_{B2}^*} = \langle \delta\hat{D}_{iI}; \hat{n}_{k\ell}\rangle_{\hat{\rho}_{B2}^\star} \qquad \text{(C42)}$$

$$= \prod_{\ell' \in I} \langle \delta\hat{n}_{i\ell'}; \hat{n}_{k\ell}\rangle_{\hat{\rho}_{B2}^\star}, \qquad \text{(C43)}$$

which is only nonzero when $I = \{\ell\}$. To first-order, $\eta_{i\{\ell\}} = \eta_{i\{\ell\}}^\star$ and the contribution from η_{iI} where $|I| > 0$ yields

$$n_{k\ell} \equiv \langle \hat{n}_{k\ell}\rangle_{\hat{\rho}_{B2}} = \langle \hat{n}_{k\ell}\rangle_{\hat{\rho}_{B2}^\star} \equiv n_{k\ell}^\star, \qquad \text{(C44)}$$

and

$$n_\ell \equiv \langle \hat{n}_{i\ell}\rangle_{\hat{\rho}_{B2}} = \langle \hat{n}_{i\ell}\rangle_{\hat{\rho}_{B2}^\star} = n_\ell^\star. \qquad \text{(C45)}$$

In conclusion, if we write the local Hamiltonian as $\hat{H}_{\text{loc}} = \sum_{iI} E_{\text{loc};I} \delta\hat{D}_{iI}$, then the total energy per site for $\hat{\rho}_{B2}$ is given as

$$E = \sum_\ell \int dk\, \epsilon_{k\ell} n_{k\ell} + \sum_I E_{\text{loc};I}\left(\prod_{\ell\in I} \mathcal{F}_\ell\right)\delta D_{iI;0},$$

where $\delta D_{iI;0} = \langle \delta\hat{D}_{iI}\rangle_{\hat{P}_1}$ with $\delta D_{iI;0} = 0$ for $|I| = 1$, and $n_{k\ell} \equiv \langle \hat{n}_{k\ell}\rangle_{\hat{\rho}_{B2}} \in [0, 1]$ is the physical momentum density distribution and is constrained by $(1/L)\sum_k n_{k\ell} = n_\ell$. The $n_{k\ell}$ can be viewed as variational parameters determined from the $\gamma_{k\ell}$ within $\hat{K}_1$. Combining Eq. (C41) and Eq. (C44), we can explicitly express $\mathcal{F}_\ell$ as

$$\mathcal{F}_\ell = \left|\frac{\int dk \sqrt{n_{k\ell}(1-n_{k\ell})}}{\sqrt{n_\ell(1-n_\ell)}}\right|^2. \qquad \text{(C46)}$$

To evaluate the Hamiltonian in the usual Hubbard operator representation for a given $X_{i\Gamma;0} = \langle \hat{X}_{i\Gamma}\rangle_{\hat{P}_1}$ while respecting the density constraint, we can compute $\delta D_{iI;0}$ from $X_{i\Gamma;0}$, and then use Eq. (C40) to evaluate δD_{iI} and then use Eq. (C31) to compute $X_{i\Gamma} = \langle \hat{X}_{i\Gamma}\rangle_{\hat{\rho}_{B2}}$. Finally, we remark that Eq. (C40) is derived by ignoring nonlocal contributions, which is equivalent to taking the limit of infinite dimensions. One could straightforwardly account for nonlocal contributions using Eq. (C39), which will be explored in the future work.

[1] R. M. Fernandes and A. V. Chubukov, Low-energy microscopic models for iron-based superconductors: A review, Rep. Prog. Phys. **80**, 014503 (2017).

[2] M. Imada, A. Fujimori, and Y. Tokura, Metal-insulator transitions, Rev. Mod. Phys. **70**, 1039 (1998).

[3] 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).

[4] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, Rev. Mod. Phys. **78**, 865 (2006).

[5] G. Kotliar and D. Vollhardt, Strongly correlated materials: Insights from dynamical mean-field theory, Phys. Today **57**(3), 53 (2004).

[6] D. Vollhardt, Dynamical mean-field theory for correlated electrons, Ann. Phys. **524**, 1 (2012).

[7] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte-Carlo methods for quantum impurity models, Rev. Mod. Phys. **83**, 349 (2011).

[8] Z. Cheng and C. A. Marianetti, Precise ground state of multiorbital Mott systems via the variational discrete action theory, Phys. Rev. B **106**, 205129 (2022).

[9] J. Bunemann, The Gutzwiller approximation for degenerate bands: A formal derivation, Eur. Phys. J. B **4**, 29 (1998).

[10] J. Bunemann and F. Gebhard, Equivalence of Gutzwiller and slave-boson mean-field theories for multiband Hubbard models, Phys. Rev. B **76**, 193104 (2007).

[11] J. Bunemann and W. Weber, Generalized Gutzwiller method for n⩾2 correlated bands: First-order metal-insulator transitions, Phys. Rev. B **55**, 4011 (1997).

[12] N. Lanata, P. Barone, and M. Fabrizio, Fermi-surface evolution across the magnetic phase transition in the Kondo lattice model, Phys. Rev. B **78**, 155127 (2008).

[13] J. P. Lu, Metal-insulator transitions in degenerate Hubbard models and $A_x C_{60}$, Phys. Rev. B **49**, 5687 (1994).

[14] T. Okabe, Generalization of Gutzwiller approximation, J. Phys. Soc. Jpn. **66**, 2129 (1997).

[15] Z. Cheng and C. A. Marianetti, Foundations of variational discrete action theory, Phys. Rev. B **103**, 195138 (2021).

[16] Z. Cheng and C. A. Marianetti, Variational discrete action theory, Phys. Rev. Lett. **126**, 206402 (2021).

[17] Z. Cheng and C. A. Marianetti, Gauge constrained algorithm of variational discrete action theory at $n = 3$ for the multiorbital Hubbard model, Phys. Rev. B **108**, 035127 (2023).

[18] Z. Cheng and C. A. Marianetti, companion paper, Mott transition from the nonanalyticity of the one-body reduced density-matrix functional, Phys. Rev. B **113**, L081108 (2026).

[19] N. Lanata, T. H. Lee, Y. X. Yao, and V. Dobrosavljevic, Emergent Bloch excitations in Mott matter, Phys. Rev. B **96**, 195126 (2017).

[20] T. H. Lee, N. Lanatà, and G. Kotliar, Accuracy of ghost rotationally invariant slave-boson and dynamical mean field theory as a function of the impurity-model bath size, Phys. Rev. B **107**, L121104 (2023).

[21] T. H. Lee, C. Melnick, R. Adler, N. Lanatà, and G. Kotliar, Accuracy of ghost-rotationally-invariant slave-boson theory for multiorbital Hubbard models and realistic materials, Phys. Rev. B **108**, 245147 (2023).

[22] S. Giuli, C. Mejuto-Zaera, and M. Capone, Altermagnetism from interaction-driven itinerant magnetism, Phys. Rev. B **111**, L020401 (2025).

[23] C. Mejuto-Zaera and M. Fabrizio, Efficient computational screening of strongly correlated materials: Multiorbital phenomenology within the ghost Gutzwiller approximation, Phys. Rev. B **107**, 235150 (2023).

[24] T. H. Lee, C. Melnick, R. Adler, X. Sun, Y. Yao, N. Lanata, and G. Kotliar, Charge self-consistent density functional theory plus ghost rotationally invariant slave-boson theory for correlated materials, Phys. Rev. B **110**, 115126 (2024).

[25] G. Kotliar and A. E. Ruckenstein, New functional integral approach to strongly correlated Fermi systems: The Gutzwiller approximation as a saddle-point, Phys. Rev. Lett. **57**, 1362 (1986).

[26] J. Bunemann, A slave-boson mean-field theory for general multi-band Hubbard models, Phys. Status Solidi B **248**, 203 (2011).

[27] M. Chatzieleftheriou, M. Berovic, P. V. Arribi, M. Capone, and L. de' Medici, Enhancement of charge instabilities in Hund's metals by breaking of rotational symmetry, Phys. Rev. B **102**, 205127 (2020).

[28] L. de' Medici, Hund's induced Fermi-liquid instabilities and enhanced quasiparticle interactions, Phys. Rev. Lett. **118**, 167003 (2017).

[29] A. Isidori, M. Berovic, L. Fanfarillo, L. de' Medici, M. Fabrizio, and M. Capone, Charge disproportionation, mixed valence, and Janus effect in multiorbital systems: A tale of two insulators, Phys. Rev. Lett. **122**, 186401 (2019).

[30] A. Isidori and M. Capone, Rotationally invariant slave bosons for strongly correlated superconductors, Phys. Rev. B **80**, 115120 (2009).

[31] F. Lechermann, A. Georges, G. Kotliar, and O. Parcollet, Rotationally invariant slave-boson formalism and momentum dependence of the quasiparticle weight, Phys. Rev. B **76**, 155102 (2007).

[32] C. Piefke and F. Lechermann, Rigorous symmetry adaptation of multiorbital rotationally invariant slave-boson theory with application to Hund's rules physics, Phys. Rev. B **97**, 125154 (2018).

[33] E. Arrigoni and G. C. Strinati, Beyond the Gutzwiller approximation in the slave-boson approach: Inclusion of fluctuations with the correct continuum-limit of the functional integral, Phys. Rev. Lett. **71**, 3178 (1993).

[34] T. Jolicoeur and J. C. LeGuillou, Fluctuations beyond the Gutzwiller approximation in the slave-boson approach, Phys. Rev. B **44**, 2403 (1991).

[35] M. Lavagna, Functional-integral approach to strongly correlated Fermi systems: Quantum fluctuations beyond the Gutzwiller approximation, Phys. Rev. B **41**, 142 (1990).

[36] T. Li, Y. S. Sun, and P. Wolfle, Dynamic-response functions of Hubbard-model in Gutzwiller approximation, Z. Phys. B **82**, 369 (1991).

[37] T. Li and P. Benard, Slave-boson calculation of the Landau parameters of the one-band Hubbard model, Phys. Rev. B **50**, 17837 (1994).

[38] R. Raimondi and C. Castellani, Lower and upper Hubbard bands: A slave-boson treatment, Phys. Rev. B **48**, 11453 (1993).

[39] W. Zimmermann, R. Fresard, and P. Wolfle, Spin and charge structure factor of the two-dimensional Hubbard model, Phys. Rev. B **56**, 10097 (1997).

[40] J. Bünemann, S. Wasner, E. Von oelsen, and G. Seibold, Exact response functions within the time-dependent Gutzwiller approach, Philos. Mag. **95**, 550 (2015).

[41] M. Fabrizio, Quantum fluctuations beyond the Gutzwiller approximation, Phys. Rev. B **95**, 075156 (2017).

[42] J. Lorenzana and G. Seibold, Dynamics of metallic stripes in cuprates, Phys. Rev. Lett. **90**, 066404 (2003).

[43] E. V. Oelsen, G. Seibold, and J. Bünemann, Time-dependent Gutzwiller theory for multiband Hubbard models, Phys. Rev. Lett. **107**, 076402 (2011).

[44] E. Von oelsen, G. Seibold, and J. Bünemann, Time-dependent Gutzwiller theory for multi-band Hubbard models, New J. Phys. **13**, 113031 (2011).

[45] M. Schiro and M. Fabrizio, Time-dependent mean field theory for quench dynamics in correlated electron systems, Phys. Rev. Lett. **105**, 076401 (2010).

[46] M. Schiro and M. Fabrizio, Quantum quenches in the Hubbard model: Time-dependent mean-field theory and the role of quantum fluctuations, Phys. Rev. B **83**, 165105 (2011).

[47] G. Seibold and J. Lorenzana, Time-dependent Gutzwiller approximation for the Hubbard model, Phys. Rev. Lett. **86**, 2605 (2001).

[48] N. Lanata, Operatorial formulation of the ghost rotationally invariant slave-boson theory, Phys. Rev. B **105**, 045111 (2022).

[49] L. de' Medici, A. Georges, and S. Biermann, Orbital-selective Mott transition in multiband systems: Slave-spin representation and dynamical mean-field theory, Phys. Rev. B **72**, 205124 (2005).

[50] M. Crispino, M. Chatzieleftheriou, T. Gorni, and L. de' Medici, Slave-spin mean field for broken-symmetry states: Neel antiferromagnetism and its phase separation in multiorbital Hubbard models, Phys. Rev. B **107**, 155149 (2023).

[51] A. B. Georgescu and S. Ismail-beigi, Generalized slave-particle method for extended Hubbard models, Phys. Rev. B **92**, 235117 (2015).

[52] T. Gorni, D. Florez-ablan, and L. de' Medici, Rationalizing doping and electronic correlations in LaFe$_2$As$_2$, Phys. Rev. B **107**, 125166 (2023).

[53] S. R. Hassan and L. de' Medici, Slave spins away from half filling: Cluster mean-field theory of the Hubbard and extended Hubbard models, Phys. Rev. B **81**, 035106 (2010).

[54] A. K. Maurya, M. Sarder, and A. Medhi, Ground state of a three-band Hubbard model with Hund's coupling: Janus-faced behavior in presence of magnetic order, J. Phys.: Condens. Matter **33**, 425603 (2021).

[55] A. K. Maurya, M. 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 (2022).

[56] Z. Cheng and C. A. Marianetti, Off-shell effective energy theory: A unified treatment of the Hubbard model from $d = 1$ to $d = \infty$, Phys. Rev. B **101**, 081105(R) (2020).

[57] J. Liu, Y. Yao, J. H. Zhang, V. Antropov, K. M. Ho, and C. Z. Wang, *Ab initio* study on fcc Pr with correlation matrix renormalization theory, Phys. Rev. B **106**, 205124 (2022).

[58] J. Liu, Y. Yao, J. H. Zhang, K.-M. Ho, and C.-Z. Wang, Role of Coulomb interaction in the phase formation of fcc Ce: Correlation matrix renormalization theory, Phys. Rev. B **104**, L081113 (2021).

[59] J. Liu, X. Zhao, Y. X. Yao, C. Z. Wang, and K. M. Ho, Correlation matrix renormalization theory in multi-band lattice systems, J. Phys.: Condens. Matter **33**, 095902 (2021).

[60] 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).

[61] M. Dzierzawa, D. Baeriswyl, and M. Distasio, Variational wave functions for the Mott transition: The 1/r Hubbard chain, Phys. Rev. B **51**, 1993(R) (1995).

[62] D. Baeriswyl, Variational schemes for many-electron systems, in *Nonlinearity in Condensed Matter*, 1st ed., edited by A. R. Bishop, D. K. Campbell, and S. E. Trullinger (Springer-Verlag, Berlin, 1987), pp. 183–193.

[63] H. Otsuka, Variational Monte-Carlo studies of the Hubbard-model in one-dimension and 2-dimension—Off-diagonal intersite correlation-effects, J. Phys. Soc. Jpn. **61**, 1645 (1992).

[64] M. Suzuki, Generalized Trotters formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, Commun. Math. Phys. **51**, 183 (1976).

[65] H. Yokoyama and H. Shiba, Variational Monte-Carlo studies of Hubbard-model 1, J. Phys. Soc. Jpn. **56**, 1490 (1987).

[66] H. Yokoyama and H. Shiba, Variational Monte-Carlo studies of Hubbard-model 2, J. Phys. Soc. Jpn. **56**, 3582 (1987).

[67] H. Yokoyama and H. Shiba, Variational Monte-Carlo studies of superconductivity in strongly correlated electron-systems, J. Phys. Soc. Jpn. **57**, 2482 (1988).

[68] S. Koike, T. Yanagisawa, and K. Yamaji, Variational Monte Carlo study of the superconducting condensation energy on the dependence of $\epsilon_p \epsilon_d$ in the two-dimensional $d-p$ model, Physica C **388-389**, 65 (2003).

[69] R. Levy, M. A. Morales, and S. W. Zhang, Automatic order detection and restoration through systematically improvable variational wave functions, Phys. Rev. Res. **6**, 013237 (2024).

[70] S. Sorella, Systematically improvable mean-field variational ansatz for strongly correlated systems: Application to the Hubbard model, Phys. Rev. B **107**, 115133 (2023).

[71] K. Yamaji, T. Yanagisawa, T. Nakanishi, and S. Koike, Variational Monte Carlo study on the superconductivity in the two-dimensional Hubbard model, Physica C **304**, 225 (1998).

[72] T. Yanagisawa, Crossover from weakly to strongly correlated regions in the two-dimensional Hubbard model off-diagonal wave function Monte Carlo studies of Hubbard model II, J. Phys. Soc. Jpn. **85**, 114707 (2016).

[73] T. Yanagisawa, Antiferromagnetism, superconductivity and phase diagram in the two-dimensional Hubbard model—off-diagonal wave function Monte Carlo studies of Hubbard model III, J. Phys. Soc. Jpn. **88**, 054702 (2019).

[74] T. Yanagisawa, Enhancement of superconductivity due to kinetic-energy effect in the strongly correlated phase of the two-dimensional Hubbard model, Phys. Lett. A **403**, 127382 (2021).

[75] T. Yanagisawa, S. Koike, and K. Yamaji, Off-diagonal wave function Monte Carlo studies of Hubbard model I, J. Phys. Soc. Jpn. **67**, 3867 (1998).

[76] T. Yanagisawa, S. Koike, and K. Yamaji, d-wave state with multiplicative correlation factors for the Hubbard model, J. Phys. Soc. Jpn. **68**, 3608 (1999).

[77] T. Yanagisawa, M. Miyazaki, and K. Yamaji, Phase diagram of cuprate high-temperature superconductors based on the optimization Monte Carlo method, Mod. Phys. Lett. B **34**, 2040046 (2020).

[78] T. Yanagisawa, K. Yamaji, and M. Miyazaki, On the kinetic energy driven superconductivity in the two-dimensional Hubbard model, Condens. Matter **6**, 12 (2021).

[79] T. Yanagisawa, S. Koike, and K. Yamaji, Ground state of the three-band Hubbard model, Phys. Rev. B **64**, 184509 (2001).

[80] T. Yanagisawa, M. Miyazaki, and K. Yamaji, Ground-state phase diagram of the three-band d-p model, Europhys. Lett. **134**, 27004 (2021).

[81] Y. X. Chen, L. F. Zhang, E. Weinan, and R. Car, Hybrid auxiliary field quantum Monte Carlo for molecular systems, J. Chem. Theory Comput. **19**, 4484 (2023).

[82] H. Hasegawa, Slave-boson functional-integral approach to the Hubbard model with orbital degeneracy, J. Phys. Soc. Jpn. **66**, 1391 (1997).

[83] Y. Ono, M. Potthoff, and R. Bulla, Mott transitions in correlated electron systems with orbital degrees of freedom, Phys. Rev. B **67**, 035119 (2003).

[84] Z. Cheng, https://github.com/chengzhengqian/vdatqubithalf (2025).

[85] E. Kamil, R. Schade, T. Pruschke, and P. E. Blochl, Reduced density-matrix functionals applied to the Hubbard dimer, Phys. Rev. B **93**, 085141 (2016).

[86] I. Mitxelena, M. Piris, and M. Rodriguez-Mayorga, On the performance of natural orbital functional approximations in the Hubbard model, J. Phys.: Condens. Matter **29**, 425602 (2017).

[87] I. Mitxelena, M. Rodriguez-mayorga, and M. Piris, Phase dilemma in natural orbital functional theory from the n-representability perspective, Eur. Phys. J. B **91**, 109 (2018).

[88] R. Lopez-Sandoval and G. M. Pastor, Density-matrix functional theory of the Hubbard model: An exact numerical study, Phys. Rev. B **61**, 1764 (2000).

[89] I. Mitxelena and M. Piris, An efficient method for strongly correlated electrons in one dimension, J. Phys.: Condens. Matter **32**, 17LT01 (2020).

[90] I. Mitxelena and M. Piris, An efficient method for strongly correlated electrons in two-dimensions, J. Chem. Phys. **152**, 064108 (2020).

[91] M. Saubanere, M. B. Lepetit, and G. M. Pastor, Interaction-energy functional of the Hubbard model: Local formulation and application to low-dimensional lattices, Phys. Rev. B **94**, 045102 (2016).

[92] E. H. Lieb and F. Y. Wu, Absence of Mott transition in an exact solution of short-range 1-band model in 1 dimension, Phys. Rev. Lett. **20**, 1445 (1968).

[93] M. Chatzieleftheriou, S. Biermann, and E. A. Stepanov, Local and nonlocal electronic correlations at the metal-insulator transition in the two-dimensional Hubbard model, Phys. Rev. Lett. **132**, 236504 (2024).

[94] L. Geng, J. Yan, and P. Werner, Two-particle self-consistent approach combined with dynamical mean field theory: A real-frequency study of the square-lattice Hubbard model, Phys. Rev. B **111**, 115143 (2025).

[95] T. Schafer, N. Wentzell, F. Simkovic, Y. Y. He, C. Hille, M. Klett, C. J. Eckhardt, B. Arzhang, V. Harkov, F. M. Le regent,

A. Kirsch, Y. Wang, A. J. Kim, E. Kozik, E. A. Stepanov, A. Kauch, S. Andergassen, P. Hansmann, D. Rohe, Y. M. Vilk, *et al.*, Tracking the footprints of spin fluctuations: A multimethod, multimessenger study of the two-dimensional Hubbard model, Phys. Rev. X **11**, 011058 (2021).

[96] A. Tanaka, Metal-insulator transition in the two-dimensional Hubbard model: Dual fermion approach with Lanczos exact diagonalization, Phys. Rev. B **99**, 205133 (2019).

[97] A. Muller, Explicit approximate relation between reduced 2-particle and one-particle density-matrices, Phys. Lett. A **105**, 446 (1984).

[98] G. Csanyi and T. A. Arias, Tensor product expansions for correlation in quantum many-body systems, Phys. Rev. B **61**, 7348 (2000).

[99] G. Csanyi, S. Goedecker, and T. A. Arias, Improved tensor-product expansions for the two-particle density matrix, Phys. Rev. A **65**, 032510 (2002).

[100] S. Sharma, J. K. Dewhurst, N. N. Lathiotakis, and E. Gross, Reduced density matrix functional for many-electron systems, Phys. Rev. B **78**, 201103(R) (2008).

[101] M. Piris, X. Lopez, F. Ruiperez, J. M. Matxain, and J. M. Ugalde, A natural orbital functional for multiconfigurational states, J. Chem. Phys. **134**, 164102 (2011).

[102] M. Piris, Global method for electron correlation, Phys. Rev. Lett. **119**, 063002 (2017).

[103] M. Piris, A natural orbital functional based on an explicit approach of the two-electron cumulant, Int. J. Quantum Chem. **113**, 620 (2013).

[104] M. Piris, J. M. Matxain, X. Lopez, and J. M. Ugalde, The one-electron picture in the Piris natural orbital functional 5 (PNOF5), Theor. Chem. Acc. **132**, 1298 (2013).

[105] A. B. Georgescu and S. Ismail-Beigi, Symmetry breaking in occupation number based slave-particle methods, Phys. Rev. B **96**, 165135 (2017).