PHYSICAL REVIEW B VOLUME 56, NUMBER 24 15 DECEMBER 1997-II

Small-bandwidth perturbation theory for highly covalent Mott insulators

Yen-Sheng Su, T. A. Kaplan, and S. D. Mahanti
Department of Physics and Astronomy and Center for Fundamental Materials Research, Michigan State University, East Lansing, Michigan 48824

(Received 13 June 1997)

Calculation of the elastic neutron scattering form factor by an essentially standard approach has given results that disagree seriously with experiment on La_2NiO_4. This has motivated us to look for a more fundamental approach to such a calculation (in Mott insulators). We have begun by considering perturbation approaches in the context of the three-band Hubbard model of cuprate CuO_2 planes due to Hybertsen *et al.* This was recently shown, for a small cluster, to have a straightforward small-bandwidth perturbation expansion of the Heisenberg exchange parameter J that is *nonconvergent*. We study the roles of one-body and two-body transformations on the basis set of states in converting nonconvergent many-body perturbation expansions into convergent ones. We choose the one-body transformations guided by the thermal single determinant approximation (TSDA), a variational generalization of the thermal Hartree-Fock approximation. All transformations preserve ''localization'' of copper d orbitals, and thus lead to low-lying states governed by a Heisenberg spin Hamiltonian, in leading order, provided the perturbation theory is convergent. We find the one-body transformations do make the perturbation expansion converge, although rather slowly; addition of two-body transformations gives significant improvement in the convergence rate. The reason for the limitation of the one-body transformation is given. [S0163-1829(97)02548-4]

I. INTRODUCTION

The question of how to set up a perturbation theory for Mott insulators like NiO, $KNiF_3$, Fe_2O_3, was considered in 1959 by Anderson.[1] The object was to take advantage of the obviously localized property of the magnetic electrons to construct the theory, where the interatomic overlap of localized orbitals is treated as the small parameter. Such a theory leads to low-lying states that are governed by the Heisenberg spin Hamiltonian, in leading order. At the time, it was known that most of the low-temperature properties of these materials could be understood on the basis of such a Hamiltonian.

Anderson, on considering how to determine the localized orbitals, stated that there exist ''exact'' localized orbitals (orthogonal, and thus called Wannier functions), which will make the perturbation theory converge well. These orbitals were required to be nonmagnetic (spin-up and down orbitals the same), as is natural, since the magnetic properties were to be derived from the spin Hamiltonian resulting from the perturbation theory (pt). Anderson gave an explicit prescription, namely Wannier functions derived from Hartree-Fock theory in which the spins of the magnetic ions were all parallel. It was noted some time later[2] that there was an inconsistency with this, since Hartree-Fock eigenstates with up and down spins differ in such a situation.[3] An alternative approach was studied[2] based on the so-called thermal single determinant approximation (TSDA),[4] a variational generalization of thermal Hartree-Fock theory. Solving this TSDA for nonmagnetic localized solutions at sufficiently low temperature amounted to minimizing the mean determinantal energy, averaged over all spin configurations of the magnetic ions, an approach adopted by Gondaira and Tanabe.[5] It was shown for a small cluster model[2,5] which we can call H-He-H (hydrogen-helium-hydrogen) and for a one-dimensional (1D) crystal model[2] (the corresponding H-He chain) that this TSDA choice indeed improved, quite dramatically, the convergence rate of perturbation theory for the Heisenberg exchange parameter J, through fourth order in the H-He overlap.

The overlap, or more precisely, the amplitude for cation-anion hopping, is small in the highly ionic materials considered at the time, e.g., NiO, $KNiF_3$, so the pt was expected to converge rapidly. For this reason higher order perturbations were not investigated.[2] However, materials that have recently become of interest in connection with high-temperature superconductivity, namely the cuprates and the related La_2NiO_4, are thought to be much more covalent. Doubts about the validity of perturbation theory in this connection, specifically for the three-band Hubbard model of a CuO_2 plane (Fig. 1), have been expressed.[6–8] And, in fact, it has recently been shown[9] that straightforward perturbation theory does *not* converge, specifically for the version of the three-band model due to Hybertsen *et al.*,[6] which is rather realistic for many physical observables,[10] and is quite similar to other versions of the three-band model.[11–14]

Clearly then it is of interest to try the TSDA approach to

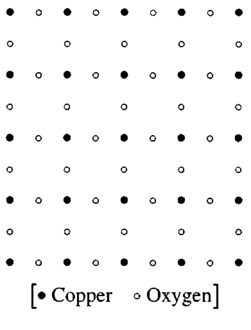

FIG. 1. CuO_2 plane.

find new localized orbitals, and to test the implied new perturbation expansion for convergence. In fact, the question of convergence of pt within the three-band model has been addressed,[7,8] and rapid convergence of a new pt was claimed. But the transformation in these works involved both one- and two-body transformations. (Finding new localized orbitals is a one-body transformation.) It is important to understand how far one can go with the one-body transformation alone, because of its relative conceptual simplicity and its historical interest. This is the purpose of the present paper.

We have been motivated to explore these questions by difficulties we have run into in connection with calculating the ground-state spin density in the cuprates and La_2NiO_4.[15,16] The type of perturbation theory discussed above, which in effect *separates the spin and space (or charge) degrees of freedom*, might be the only hope at present of making tractable *ab initio* calculations of this observable. This is due to the fact that the reduction of the ordered spin due to zero-point or quantum spin fluctuations (QSF's), which is large in these cases, constitutes a *macroscopic correlation effect*. That is, (i) in principle, it requires the thermodynamic limit for its very existence, and therefore one needs very large systems to estimate the QSF,[17] and (ii) the wave function representing such a state requires a linear combination of many Slater determinants. Thus, even with the remarkable advances in computational many-body physics (e.g., Refs. 18 and 19), the present problem is still beyond the reach of those approaches. Also standard band-theoretic approaches fail to capture essential physics [the local-spin-density approximation misses the antiferromagnetism and insulating property of La_2CuO_4,[20] and the Hartree-Fock method (unrestricted) misses the QSF].

Since the work reported here involves calculations only for a small cluster, the spin density of a crystal cannot be addressed (the spin density in a finite cluster with an even number of electrons is zero, since the ground state is a singlet). Instead we will be investigating the Heisenberg exchange parameter J as an indicator of convergence of the perturbation expansions. In Sec. II we define the Hamiltonian to be considered, review effective-Hamiltonian perturbation theory, and the TSDA. In Sec. III we describe the three types of single-particle transformations considered, which we call *site localization*, *cell localization*, and *no localization* (the names are after the transformed ligand p orbitals surrounding the magnetic cations), and find the best in each case according to the TSDA. Also in this section, J is calculated to high order in pt to examine its convergence properties. In Sec. IV we carry out the particular two-body transformation suggested by the work of Ref. 7, for the special model they considered, as a check; we also carry this out for the model of Hybertsen *et al.*[10] A summary and discussion are given in Sec. V. A brief overview of the results obtained can be seen by glancing at Figs. 3, 5, 7, 9, and 10, which show the nonconvergent series of ''straightforward pt,'' and the rather dramatic improvement obtained in each of the four modifications, respectively.

II. THE MODEL, EFFECTIVE-HAMILTONIAN PERTURBATION THEORY, AND TSDA

A. Model Hamiltonian

We consider the three-band Hubbard Hamiltonian (sometimes referred to as the Anderson lattice model) as parametrized by Hybertsen *et al.*:[6,10]

$$H=H_1+H_U+H_K, \tag{1a}$$

where

$$H_1=\varepsilon\sum_i n_i^p+t_{pd}\sum_{\langle il\rangle\sigma}(c_{i\sigma}^{d+}c_{l\sigma}^{p}+\text{H.c.})$$

$$+t_{pp}\sum_{\langle ll'\rangle\sigma}(c_{l\sigma}^{p+}c_{l'\sigma}^{p}+\text{H.c.}), \tag{1b}$$

$$H_U=U_d\sum_i n_{i\uparrow}^d n_{i\downarrow}^d+U_p\sum_l n_{l\uparrow}^p n_{l\downarrow}^p+U_{pd}\sum_{\langle il\rangle} n_i^d n_l^p, \tag{1c}$$

and

$$H_K=K_{pd}\sum_{\langle il\rangle\sigma\sigma'} c_{i\sigma}^{d+}c_{i\sigma'}^{d}c_{l\sigma'}^{p+}c_{l\sigma}^{p}$$

$$+K_{pp}\sum_{\langle ll'\rangle\sigma\sigma'} c_{l\sigma}^{p+}c_{l\sigma'}^{p}c_{l'\sigma'}^{p+}c_{l'\sigma}^{p}. \tag{1d}$$

$c_{i\sigma}^{\nu+}$ creates a hole in a Wannier function $w_{\nu i}$ of type ν at site i with spin σ, $n_i^\nu=\Sigma_\sigma n_{i\sigma}^\nu$, $n_{i\sigma}^\nu=c_{i\sigma}^{\nu+}c_{i\sigma}^{\nu}$. The orbital at a copper site is $d_{x^2-y^2}$; at each oxygen site there is one p orbital. The parameter values are $\varepsilon=3.6$, $t_{pd}=-1.3$, $t_{pp}=-0.65$, $U_d=10.5$, $U_p=4$, $U_{pd}=1.2$, $K_{pd}=-0.18$, $K_{pp}=-0.04$, all in eV.[21] [Note that for the convenience of having the d-p hopping parameters all the same, and similarly for the p-p hopping, as in Eq. (1b), one has to set up the underlying Wannier functions in the way that they may change signs under the crystal or cluster symmetry. The signs corresponding to orbital phases are chosen as follows: if $d_{x^2-y^2}$ exists at a particular Cu, then at the O's immediately to its right and left (along the x axis) the orbitals are p_x and $-p_x$, respectively; similarly, the orbitals at the nearest O's below and above are p_y and $-p_y$, respectively. The remaining phases are determined by having the nearest-neighbor d-p overlap always negative.] We limit ourselves to the case where the number of holes equals the number of Cu sites, i.e., the Mott insulator limit.

B. Formal perturbation theory

The form of the perturbation theory used involves an effective Hamiltonian. Given

$$H\Psi=E\Psi, \tag{2}$$

then the effective Hamiltonian H_{eff} satisfies

$$H_{\text{eff}}(E)P\Psi=EP\Psi, \tag{3a}$$

namely,

$$H_{\text{eff}}(E)=P\{H+H[Q(E-H)Q]^{-1}H\}P. \tag{3b}$$

Here, $P=1-Q$ is a projection operator which projects Ψ onto a given subspace. The ''inverse of the corner''[22] $G(E)\equiv[Q(E-H)Q]^{-1}$ has to be understood as the matrix with zeroes everywhere except in the Q subspace, where the matrix is the inverse of the Q projection of $E-H$. Similarly, $Q\equiv Q^{-1}$ is the matrix with zeroes everywhere except for the Q subspace, where it is the unit matrix. We choose the P subspace as that defined by the ground states of some chosen unperturbed Hamiltonian H_0, with eigenvalue E_0. Then $PHQ=P(H_0+V)Q=PVQ$, and Eq. (3a) can be rewritten as

$$P\{V+VG(E)V\}P\Psi=\delta EP\Psi, \tag{4}$$

where $\delta E=E-E_0$. Expanding $G(E)$ in powers of $\delta E-V$, and substituting the full left-hand side of Eq. (4) for $\delta EP\Psi$ everywhere it appears in the expansion, Eq. (4) becomes[23]

$$P\{H_0+V+VG_0V+(VG_0VG_0V-VG_0^2VPV)+\cdots\}P\Psi$$
$$=EP\Psi, \tag{5}$$

where

$$G_0=[Q(E_0-H_0)Q]^{-1}.$$

Straightforward perturbation theory takes the terms in H that involve hopping (t_{pd} and t_{pp}) plus the exchange terms (H_K) as the perturbation V (and $H_0=H-V$). We studied[9] this perturbation theory (pt) in the case of a crystal (infinite CuO$_2$ plane), where we obtained the first three terms (third, fourth, and fifth order) for the nearest Cu-Cu exchange parameter in the Heisenberg Hamiltonian.[24] We also studied this pt on the cluster Cu$_2$O$_7$, using the embedding scheme described in Ref. 10, and showed that the corrections to J increase in magnitude and oscillate in sign through fourteenth order.[9]

C. TSDA

To help in choosing transformations that are to define new pt's (i.e., new choices of V), we will appeal to the TSDA (Ref. 4) which, essentially, finds the ''best'' one-particle wave functions according to the free-energy variational principle (for the canonical ensemble),

$$F(\rho)\equiv\mathrm{tr}(\rho H)+\beta^{-1}\ \mathrm{tr}(\rho\ \ln\ \rho)\geqslant-\beta^{-1}\ \ln\ \mathrm{tr}\ \exp(-\beta H). \tag{6}$$

Here ρ is any density operator and β^{-1} is Boltzmann's constant times absolute temperature. Writing $\rho=Z_a^{-1}$ exp $(-\beta H_a)$ with $Z_a=\mathrm{tr}\exp(-\beta H_a)$, one can consider H_a as an approximate Hamiltonian. The TSDA is defined by taking H_a as a general real function of the number operators for some complete orthonormal set of one-particle wave functions ψ_ν and then varying the *functional form* of H_a and the ψ_ν's to minimize F. A complete set of eigenstates of H_a can be chosen as single Slater determinants with all the various subsets of the ψ_ν's occupied. It turns out[4] that the best function H_a for a given set of ψ_ν is simply the determinantal energy of the exact Hamiltonian H with respect to that set of ψ_ν. In other words, if H is expressed with respect to an arbitrary complete orthonormal set of one-particle wave functions ψ_ν as

$$H=\sum_{\nu\mu} h_{\nu\mu}c_\nu^+c_\mu+\tfrac{1}{2}\sum_{\nu\mu\lambda\kappa} v_{\nu\mu,\lambda\kappa}c_\nu^+c_\mu^+c_\kappa c_\lambda, \tag{7}$$

where the subscripts $\nu,\mu,\ldots$ each label both the spatial and spin quantum numbers and the quantities $h_{\nu\mu}$ and $v_{\nu\mu,\lambda\kappa}$ are matrix elements (which, by properly choosing the phase factors of the ψ_ν's, are taken to be real in this paper), respectively, of the one- and two-body terms in the exact Hamiltonian H, then H_a will be

$$H_a=\sum_\nu h_{\nu\nu}n_\nu+\tfrac{1}{2}\sum_{\nu\mu}(v_{\nu\mu,\nu\mu}-v_{\nu\mu,\mu\nu})n_\nu n_\mu, \tag{8}$$

which contains terms linear and quadratic in the number operators n_ν corresponding to the ψ_ν's. (The distinction between this and thermal Hartree-Fock theory is that in the latter, H_a is restricted to be linear in the occupation numbers, i.e., a one-electron operator.) H_a turns out to be equal to the diagonal part of H with respect to the single-determinantal states constructed from having the one-particle states $\mu,\nu,\ldots$ occupied in all possible ways. As an example, for the particular model (1) the determinants with the original Wannier functions occupied in all possible ways have energies given by

$$H_a=\varepsilon\sum_i n_i^p+H_U+K_{pd}\sum_{\langle il\rangle\sigma} n_{i\sigma}^d n_{l\sigma}^p+K_{pp}\sum_{\langle ll'\rangle\sigma} n_{l\sigma}^p n_{l'\sigma}^p, \tag{9}$$

the values of the occupation numbers being defined by the determinant being considered. Furthermore, stationarity of $F(\rho)$ with respect to the ψ_ν's implies[4]

$$\langle n_\nu-n_\mu\rangle h_{\nu\mu}+\sum_\lambda\langle(n_\nu-n_\mu)n_\lambda\rangle(v_{\nu\lambda,\mu\lambda}-v_{\nu\lambda,\lambda\mu})=0, \tag{10}$$

where the brackets $\langle\cdots\rangle$ mean average over the canonical ensemble with H_a as Hamiltonian, i.e., for any operator A, $\langle A\rangle$ means

$$\langle A\rangle\equiv\frac{\mathrm{tr}(e^{-\beta H_a}A)}{\mathrm{tr}(e^{-\beta H_a})}. \tag{11}$$

A physical interpretation of Eq. (10) is presented in Appendix A.

It should be noted that TSDA allows localized solutions in crystals, or analogously in molecules with high symmetry, in contrast to the thermal HF approximation.[4,25] This distinction is important in the Mott insulators we are considering. Further, the form of the localized orbitals in TSDA is dictated by the Hamiltonian H. As far as we are aware TSDA is the only variational theory in the literature with this formal property.

It would be natural to ask here why not simply use the TSDA as a variational approximation to the Hamiltonian H of Eq. (1)? The answer is that it is known to give a poor description of the low-energy physics when the hopping is very small (actually, specifically in the case of the single-band Hubbard model), so that the low-lying energies are accurately given by the Heisenberg model. E.g., it gives a Curie law susceptibility rather than the correct Curie-Weiss law at temperatures above the antiferromagnetic transition temperature, and it gives the latter about an order of magnitude too large.[26] Thus in the present work we are *not* considering TSDA in this way. Rather we are using it as a formal device to help discover a one-body transformation that (hopefully) will lead, via a rapidly convergent perturbation theory, to the proper effective Hamiltonian (it is the solution of the latter

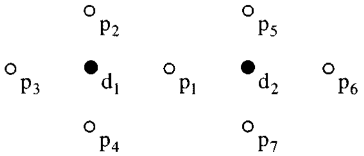

FIG. 2. The cluster Cu_2O_7. The symbols at each site serve both as labels and as orbitals.

that will determine accurately the low-energy physics of the model). The early success of this approach discussed in the Introduction has motivated our present investigation.

Thus in our TSDA considerations here, β will be considered as a parameter whose optimum value β_0 would ideally give the ''fastest'' convergence of the perturbation expansion (5). As is apparent from the above TSDA equations, the solution, i.e., the ψ_ν's, will depend on β_0. If we take a finite number of terms in the perturbation expansion (as one would do), then the resulting effective Hamiltonian will thus depend on β_0. For thermal properties, e.g., β_0 would then enter the partition function $\Sigma_n \exp[-E_n(\beta_0)/k_BT]$ where $E_n(\beta_0)$ are the eigenvalues of the spin Hamiltonian and T is the physical temperature. When working within the TSDA to find the ψ_ν's, we will think of β^{-1} intuitively as temperature to help find a ''good'' TSDA solution, i.e., a solution that gives rapid convergence of pt; in fact we will tend to consider large β, since the effective Hamiltonian is designed to give the low-energy states of H. (One could probably not explore the whole range of β in practice.)

One should realize that even for the small-basis-set model of the small cluster treated below, the one-body Hilbert space is too large to explore completely. So one needs *some* guide. There are alternate approaches to determining localized orbitals, widely used in quantum chemistry; see Ref. 19 and references contained therein and Ref. 27. These conceivably could be useful in the magnetism problem considered here, and should probably be studied in this connection.

III. ACHIEVEMENT OF CONVERGENCE VIA ONE-BODY TRANSFORMATIONS

As mentioned above, we applied[9] pt to the Hamiltonian in the original Wannier-function representation, i.e., Eq. (1), where the perturbation V incorporates all the hopping and exchange terms and H_0 the rest. For the crystal (an infinite CuO_2 plane) we obtained the first three contributions to the nearest-neighbor exchange parameter J (they occur in third, fourth, and fifth order), with no sign of convergence.[28] We also calculated J to very high order for the embedded cluster Cu_2O_7, shown in Fig. 2, with two holes. The results of the latter work are given in Fig. 3. Plotted there is

$$J^{(m)} = \sum_{n=1}^{m} J_n, \tag{12}$$

where J_n is the contribution to J (the splitting between the lowest singlet and triplet) from nth order pt. The nonconvergence is apparent, the result oscillating with increasing amplitude about the exact value (straight line) calculated by direct diagonalization.

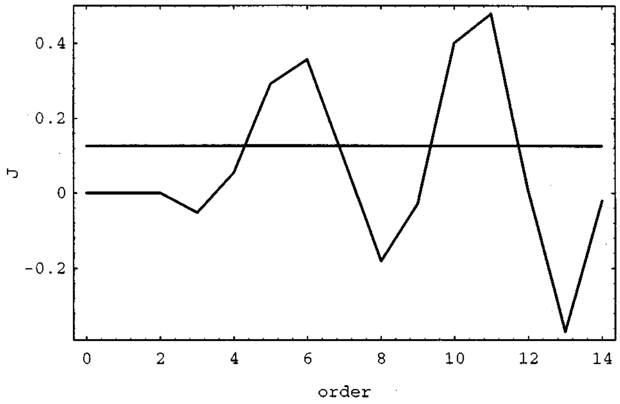

FIG. 3. The nth approximation to J vs n according to straightforward perturbation theory. The horizontal line shows the exact value of J. From Ref. 9.

To follow up on the idea[1] that an appropriate change of one-electron basis states, with its consequent change in the partition of H into unperturbed and perturbed pieces, might improve the convergence, we consider the unitary linear transformation,

$$c'_{i\sigma} = \sum_j A_{ij} c_{j\sigma}. \tag{13}$$

An essential assumption we make about this transformation is that it maintains localization and symmetry of the d states and that all states should be nonmagnetic. The reason for this is that we want to maintain the property of the straightforward pt that the low-lying energies are governed by a Heisenberg Hamiltonian with the symmetry of the nuclear structure. A_{ij} is a unitary matrix, which we take as real, since we assume the new Wannier functions are real. The spin independence of A_{ij} embodies the assumption that the new Wannier functions are nonmagnetic (they are each a product of a spatial orbital times a spin function, α or β, the spatial orbital being the same for either spin). These properties are assumed of course to hold for the original Wannier functions, i.e., those created by the $c^+_{i\sigma}$.

A. Site localization

We begin with our first scheme which we call site localization, defined by the requirement that the new Wannier functions have the same symmetry properties as the original ones. Referring to Fig. 2, let σ_v and σ_h be reflections in the vertical and horizontal symmetry axes, respectively. Then the d states are taken to satisfy

$$d'_1 = -\sigma_v d'_2 \tag{14}$$

(a similar relation being assumed, of course, for the original, unprimed, orbitals; the sign choice follows the discussion in Sec. II A.) Similarly, we require

$$p'_2 = -\sigma_v p'_5, \quad p'_3 = -\sigma_v p'_6, \quad p'_1 = -\sigma_v p'_1, \quad \text{etc.,} \tag{15}$$
$$p'_2 = \sigma_h p'_4, \quad p'_5 = \sigma_h p'_7, \quad p'_3 = \sigma_h p'_3, \quad \text{etc.}$$

Equations (14) and (15), satisfied also by the unprimed Wannier functions, define ''site localization.''

One can see that these requirements plus orthonormality are not sufficient to completely determine the transformation. A simple example is the three-site cluster $d_1 - p_1 - d_2$ or $\mathrm{Cu_2O}$: Then the two symmetry requirements, Eq. (14) and $p_1' = -\sigma_v p_1'$, give

$$d_1' = N_1(d_1 + A_1 d_2 + A_2 p_1),$$

$$d_2' = N_1(A_1 d_1 + d_2 + A_2 p_1),$$

$$p_1' = N_2[B(d_1 + d_2) + p_1].$$

The N_i are normalization constants, determined directly in terms of the three coefficients, A_1, A_2, and B. Orthogonality clearly gives two equations, leaving one independent coefficient. For the cluster $\mathrm{Cu_2O_7}$, it turns out that there are nine independent coefficients. It is useful to see this in detail. We can write the transformation as

$$\begin{aligned} d_1' &= N_1[d_1 + A_1 d_2 + A_2 p_1 + A_3(p_2 + p_4) + A_4 p_3 \\ &\quad + A_5(p_5 + p_7) + A_6 p_6], \\ p_1' &= N_2[B_1(d_1 + d_2) + p_1 + B_2(p_2 + p_4 + p_5 + p_7) \\ &\quad + B_3(p_3 + p_6)], \\ p_2' &= N_3(C_1 d_1 + C_2 d_2 + C_3 p_1 + p_2 + C_4 p_3 + C_5 p_4 + C_6 p_5 \\ &\quad + C_7 p_6 + C_8 p_7), \\ p_3' &= N_4[D_1 d_1 + D_2 d_2 + D_3 p_1 + D_4(p_2 + p_4) + p_3 \\ &\quad + D_5(p_5 + p_7) + D_6 p_6]. \end{aligned} \tag{16}$$

The first two and the last forms are dictated by the horizontal reflection symmetry; the remaining five primed orbitals are obtained from those given here by operating with σ_v and σ_h. Aside from the N_i there are 23 coefficients A_i, B_i, C_i, D_i. One can see that there are 14 orthogonality conditions, leaving nine free coefficients.

To determine these coefficients we turn to the TSDA. For every pair of states μ and ν, there is a corresponding TSDA equation (10) to be satisfied. Since both h and v are spin independent, the states labeled ν and μ in the TSDA equations (10) must correspond to the same spin, each such state also corresponding to an atomic site using the site-localized orbitals defined by Eq. (16). It is easy to see that if two sites are equivalent by symmetry, e.g., the two Cu sites or the two O sites labeled p_2 and p_5, then the corresponding equation vanishes identically, giving no information. Thus the only equations (10) with content are those where the sites in ν and μ are nonequivalent, e.g., a Cu and an oxygen, or p_2-p_3. The number of such nonequivalent pairs which are independent can be seen directly to be 9, the number of free variables to be determined, i.e., there is the correct number of equations (10) to determine these unknowns.

Because we are looking for basis functions on which a perturbation theory will be based with the purpose of deriving a spin Hamiltonian that will yield the low-lying energies and magnetic properties, we look for low-temperature nonmagnetic solutions such that for every spin-up state there is a spin-down state with the same spatial orbital. In the same spirit, we consider $K_{dd} \ll \beta^{-1} \ll \varepsilon$, where K_{dd} is the splitting of the lowest levels within the single-determinantal energies [it was 0 for the original Hamiltonian, Eq. (1), but as we will see later it has value of about 0.03 eV after the one-body transformations]. Unfortunately, the low temperatures of interest lead to a large disparity in the sizes of different contributions to the thermal averages $\langle n_\nu \rangle$, etc., this causing serious numerical difficulties. An approximation has been introduced to get rid of the exponential factors in, and thus simplify, the TSDA equations. Details of the calculation and discussion about its validity are given in Appendix B.

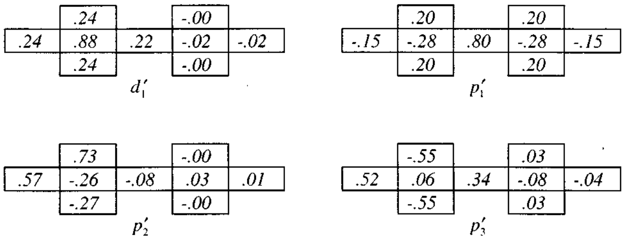

FIG. 4. TSDA orbitals within the *site localization* assumption—the numbers indicate the amplitudes of the transformed orbital at the original orbitals. (The positions of the numbers correspond to the $\mathrm{Cu_2O_7}$ cluster in Fig. 2.)

The results obtained after solving these approximate TSDA equations are shown in Fig. 4 which gives an idea of the ''shape'' of the new orbitals relative to the original ones. The Hubbard Hamiltonian (1) is then rewritten in terms of the new orbitals and thus the parameters in the model get renormalized. Note that the price to pay for the transformation is mainly that the Hamiltonian, which was restricted up to nearest-neighbor terms in Eq. (1), now resumes back to a general form like Eq. (7). We take together all those terms which are associated only with number operators as unperturbed Hamiltonian H_0, and the rest of the transformed Hamiltonian as perturbation.[29] The exchange parameter obtained from the new perturbation expansion is seen in Fig. 5 to *converge* (to the exact value). The errors in second and fourth order are 15.4 and 15.7%, respectively.

B. Cell localization

To see if we could do better, we considered another type of symmetry assumption, named *cell localization* with re-

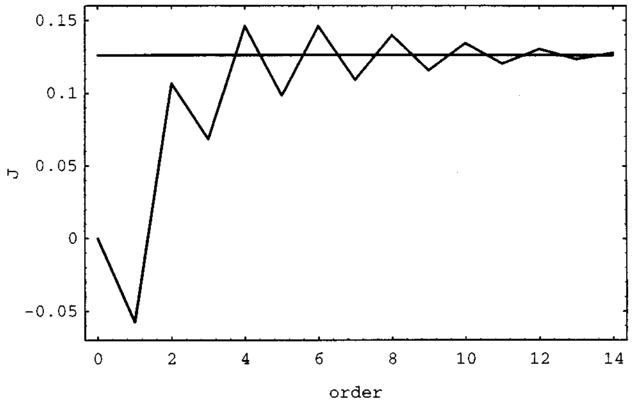

FIG. 5. Perturbation result for *site localization.*

spect to the transformed p orbitals. This was inspired by the work of Zhang and Rice,[30] which also played an important role in the work of Refs. 7 and 8. Instead of demanding that each new p orbital is localized at a site, we require each of them, with the exception of the central p, to be localized either within the cell on the left (the d plus the four p's) or the one on the right, maintaining a certain symmetry, analogous to Eqs. (14) and (15). There is no distinction between *site* and *cell localization* for the d orbitals, so they obey Eq. (14):

$$d_1' = -\sigma_v d_2'. \tag{17}$$

The distinction between *site* and *cell localization* is seen in the following symmetry requirements for the (*cell localization*) p orbitals:

$$\begin{gathered} p_i^l = -\sigma_v p_i^r, \quad p_i^{l(r)} = \sigma_h p_i^{l(r)} \quad \text{with } i = 1,2, \\ p_3^l = -\sigma_v p_3^r, \quad p_3^{l(r)} = -\sigma_h p_3^{l(r)}, \\ p^c = -\sigma_v p^c, \quad p^c = \sigma_h p^c. \end{gathered} \tag{18}$$

The superscripts l and r stand for left and right, and c for central. Thus, for example, rather than σ_h taking orbital p_2' into an orthogonal orbital p_4', as in the *site localization*, it takes p_1^l into itself. The explicit transformation in this case is

$$\begin{aligned} d_1' &= N_1[d_1 + A_1 d_2 + A_2 p_1 + A_3(p_2 + p_4) + A_4 p_3 \\ &\quad + A_5(p_5 + p_7) + A_6 p_6], \\ p^c &= B_1(d_1 + d_2) + B_2 p_1 + B_3(p_2 + p_4 + p_5 + p_7) \\ &\quad + B_4(p_3 + p_6), \\ p_1^l &= C_1 d_1 + C_2 d_2 + C_3 p_1 + C_4(p_2 + p_4) + C_5 p_3 \\ &\quad + C_6(p_5 + p_7) + C_7 p_6, \\ p_2^l &= p_1^l \quad \text{with } C_i \text{ replaced by } D_i, \\ p_3^l &= \frac{1}{\sqrt{2}}(p_2 - p_4). \end{aligned} \tag{19}$$

Following a counting similar to that done in the *site localization* case, one can show that, for the *cell localization*, the number of free coefficients after orthonormalization and the number of independent nontrivial TSDA equations are equal, as in the previous case.

The analogous figures to the *site localization* case for the *cell localization* are shown in Figs. 6 and 7. The errors in the exchange parameter in second and fourth order are 15.4 and 13.6%, respectively. It shows that the *cell localization* is a slightly better choice for the perturbation calculation than the *site localization*.

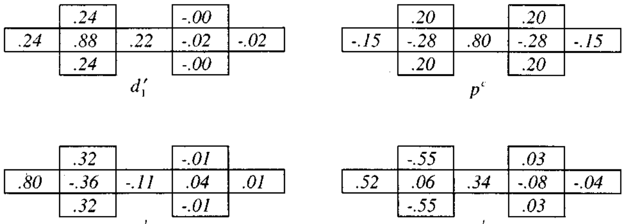

FIG. 6. TSDA orbitals for *cell localization*.

C. No localization

The third choice of the symmetry of the new orbitals is named *no localization*. The term means releasing any localization requirement on the new p orbitals (but still keeping the new d orbitals localized). This is analogous to the case in a crystal where the p orbitals may form a p band while the d orbitals remain localized. The new orbitals are required to satisfy the following symmetry properties:

$$d_1' = -\sigma_v d_2', \tag{20}$$

which is the same as in the previous two cases, and

$$\begin{gathered} p_i^{-+} = -\sigma_v p_i^{-+}, \quad p_i^{-+} = \sigma_h p_i^{-+} \quad \text{with } i = 1,2,3, \\ p_i^{++} = \sigma_v p_i^{++}, \quad p_i^{++} = \sigma_h p_i^{++} \quad \text{with } i = 1,2, \\ p^{--} = -\sigma_v p^{--}, \quad p^{--} = -\sigma_h p^{--}, \\ p^{+-} = \sigma_v p^{+-}, \quad p^{+-} = -\sigma_h p^{+-}, \end{gathered} \tag{21}$$

where the first sign in the superscripts denotes the symmetry under σ_v, and the second sign, the symmetry under σ_h. The explicit transformations for some of the orbitals are listed below:

$$\begin{aligned} d_1' &= N_1[d_1 + A_1 d_2 + A_2 p_1 + A_3(p_2 + p_4) + A_4 p_3 \\ &\quad + A_5(p_5 + p_7) + A_6 p_6], \\ p_1^{-+} &= B_1(d_1 + d_2) + B_2 p_1 + B_3(p_2 + p_4 + p_5 + p_7) \\ &\quad + B_4(p_3 + p_6), \\ p_2^{-+}(p_3^{-+}) &= p_1^{-+} \quad \text{with } B_i \text{ replaced by } C_i(D_i), \\ p_1^{++} &= E_1(d_1 - d_2) + E_2 p_1 + E_3(p_2 + p_4 - p_5 - p_7) \\ &\quad + E_4(p_3 - p_6), \end{aligned}$$

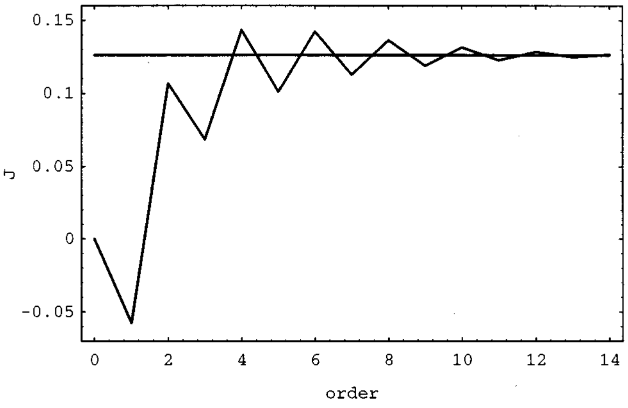

FIG. 7. Perturbation result for *cell localization*.

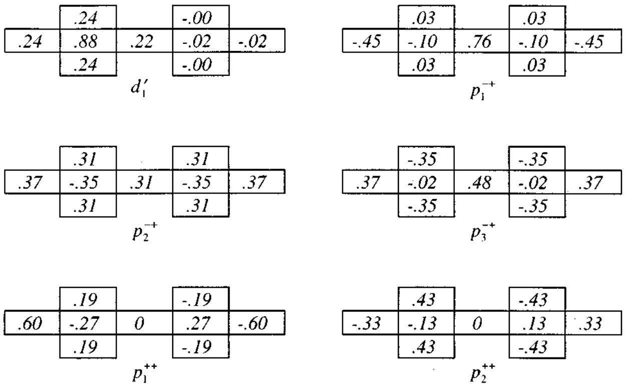

FIG. 8. TSDA orbitals for *no localization*.

$$p_2^{++}=p_1^{++} \quad \text{with } E_i \text{ replaced by } F_i, \qquad (22)$$

$$p^{--}=\tfrac{1}{2}\,(p_2-p_4+p_5-p_7),$$

$$p^{+-}=\tfrac{1}{2}\,(p_2-p_4-p_5+p_7).$$

The transformed orbitals and the perturbation results for the *no localization* case is shown in Figs. 8 and 9. The errors at second and fourth order in the exchange parameter J are 15.5 and 14.8%, respectively, not as good as the *cell localization.*

D. Comments on the TSDA solutions for the three localization choices

Because the symmetry requirements in these three choices are different for p orbitals but the same for d orbitals, and since the lowest energy states in which only the d orbitals are occupied dominate at low T, it turns out that the transformed d orbitals for the three localization choices are the same. After TSDA transformation, the energies of the lowest two single-determinantal states of the half-filled Cu_2O_7 cluster are -2.7718 and -2.7430 (all in eV) for the spins of the two holes, one at each transformed d orbital, being parallel and antiparallel, respectively. This can be compared with the energies using the original orbitals, 0 for both parallel and antiparallel spin configurations, and with the exact eigenenergies from direct diagonalization of the Hamiltonian matrix, -2.7738 and -2.9000 for the triplet and the singlet, respectively.

FIG. 9. Perturbation result for *no localization*.

After transformations, all parameters $h_{\nu\mu}$ and $v_{\nu\mu,\lambda\kappa}$ in the Hamiltonian get renormalized. For example, some new parameters resulting from the TSDA solutions are

$$t_{dd}'=-0.403, \quad U_d'=6.910, \quad U_{dd}'=0.113, \quad K_{dd}'=-0.029, \qquad (23)$$

where the first, third, and fourth ones were absent in the original form of the Hamiltonian, Eq. (1).

IV. FURTHER IMPROVEMENT VIA TWO-BODY TRANSFORMATION

Although the TSDA guided one-body transformations did lead to convergent perturbation expansions as we had hoped, we were somewhat disappointed about the relatively slow convergence rate compared to the achievement of Refs. 7 and 8, in which transformation is not restricted to be one-body. An interesting question was then asked: What is the main factor that limits the convergence rate of one-body-transformation-induced pt? The answer is given below and a minimal two-body transformation is added to improve the convergence rate.

Our attempt to incorporate both one- and two-body transformations is inspired by the work of Zhang and Rice,[30] as well as Jefferson, Eskes, and Feiner.[7] On the CuO_2 planes, a cell is defined as any one copper and its four surrounding oxygens atoms. Turning on the perturbation (p-d hopping) will induce hybridization among the copper d orbital and the oxygen p orbitals. If the perturbation is small, the new orbitals would remain localized within a cell. The copper d orbital is expected, by symmetry, to *equally* hybridize with its four surrounding oxygen p orbitals. Let us define the symmetric p orbital of each cell as

$$p_i^s=\frac{1}{2}\sum_{l\in\{i\}} p_l ,$$

where the summation index indicates that l is taken over the (four) p sites in cell i. It is intuitive to expect that the largest hybridization will occur between the d_i and p_i^s. The symmetric p_i^s orbital has its analogue in our *cell localization* choice of the one-body transformation for the Cu_2O_7 cluster, namely, p_1^l and p_1^r, for the left cell and the right cell, respectively. Therefore, our following investigation is based on the *cell localization* choice.

Because the Hamiltonian (1) conserves the total spin S and its projection on the z axis S_z, the Hamiltonian matrix elements form isolated blocks for different values of S_z. From here to the end of this section we only consider the $S_z=0$ case. Through the process of the one-body transformations, the hybridization between the d and p orbitals have actually been taken into account. Because of the expected large hybridization between the d and the symmetric p^s orbitals, we are interested in those states which consist of only d_1' (d_2') and p_1^l (p_1^r). Among the total 81 states for the half-filled Cu_2O_7 cluster (i.e., only two holes on it) with $S_z=0$, this means the states of interest are

$$|d'_{1\uparrow}p^l_{1\downarrow}\rangle,\quad |d'_{1\downarrow}p^l_{1\uparrow}\rangle,\quad |d'_{1\uparrow}d'_{1\downarrow}\rangle,\quad |p^l_{1\uparrow}p^l_{1\downarrow}\rangle, \tag{24}$$

and four similar states for the right cell. Hereafter we consider only the left cell because the expressions for the right cell are exactly the same. After the one-body transformation, the Hamiltonian submatrix in the basis set (24) is

$$\begin{bmatrix} 5.483 & -1.149 & -2.239 & -0.148 \\ -1.149 & 5.483 & 2.239 & 0.148 \\ -2.239 & 2.239 & 4.055 & 0.984 \\ -0.148 & 0.148 & 0.984 & 11.922 \end{bmatrix} \tag{25}$$

[row and column indices correspond to the states in Eq. (24) in order]. Two things are worth noting. First, the off-diagonal matrix elements are large. This is due to d'_1 and p^l_1 being mainly strong mixtures of d_1 and p^s_1. This hybridization leads to large renormalized model parameters for hole-transfer terms (between d'_1 and p^l_1) in the transformed Hamiltonian, and thus contributes to the off-diagonal matrix elements. It indicates that these four states are far from the eigenstates and may lead to slow convergence if adopted as unperturbed states in the perturbation calculation. Second, the energy of the state of double occupancy of the d orbital, 4.055, the third diagonal element of Eq. (25), is dramatically reduced from its original value 10.5, also a result of strong hybridization. Each new d orbital contains part of the original d orbital and part of the original p orbitals. Therefore double occupancy in the new d orbital only gains energy from a fraction of the large (original) U_d, and that causes the large change of this energy level. Direct diagonalization of the Hamiltonian matrix shows that the energy levels associated with double occupancy at the d orbitals remain high (~ 12.7). Thus, in spite of giving good approximate low-lying energy levels, the one-body transformation also leads to some high-lying levels that are not at all good approximations of corresponding eigenstates, and these high-energy eigenstates do play an important role in the perturbation theory. In conclusion, it is the strong hybridization (covalence effect) and the large on-site Coulomb energy U_d (correlation effect) together that prevents a one-body transformation from generating a good approximation for the whole energy spectrum.

In order to take care of the above limitation, we diagonalize the matrix (25) and use the four resulting eigenstates (the two-body transformation comes in here), plus the four counterpart states for the right cell and the remaining 73 (unchanged) states, as the basis set to perform the perturbation calculation. In some sense, by this we allow relaxation in the intermediate states of the pt, where one hole hops into the neighbor cell that has already been occupied by another hole. We call this scheme the *minimal two-body* transformation where ''minimal'' means that, among the various p orbitals, only p^l_1 (p^r_1) are involved in the two-body transformation. The result is shown in Fig. 10. We have also tried including p^l_2 (p^r_2) in the two-body transformation [p^l_3 (p^r_3) are out of our consideration because they possess different symmetry from d'_1 (d'_2)]. However, it only made a tiny change from that of the minimal two-body transformation and so the result is omitted here.

In Ref. 7, Jefferson, Eskes, and Feiner described another transformation scheme that is in spirit, similar to our one-body+minimal two-body transformation, and demonstrated rapid convergence to second-order pt on a model with a restricted parameter set: $\varepsilon=3$, $t_{pd}=-1$, $t_{pp}=-.5$, $U_d=\infty$, and all other parameters equal to zero. (The straightforward pt for this parameter set is divergent, too.) For comparison, we applied our method to the same parameter set. The result showed that the one-body transformation alone fails to lead to a convergent perturbation expansion. This is not surprising since the renormalization in U_d, the double-occupation energy at d orbitals, caused by the one-body transformation is fatal to the pt in the case of $U_d=\infty$. Nevertheless, we achieved precision similar to theirs after we added the minimal two-body transformation to the pt.

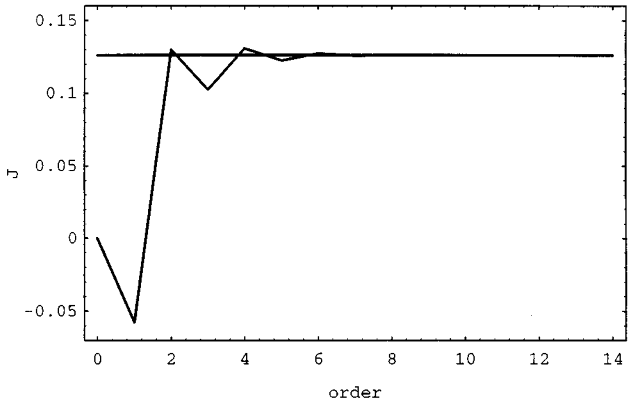

FIG. 10. Perturbation result for *cell localization* one-body+minimal two-body transformation.

V. SUMMARY AND DISCUSSION

For simplicity, we sought one-body transformations to convert the pt from divergent to convergent in the context of strongly covalent CuO_2 planes. Taking TSDA as a guide, several one-body transformations were found, in the Cu_2O_7 case, to achieve this goal, but the convergence rate in each case was rather slow. Besides those suggested by TSDA, we have also tried some other one-body transformations, of which the best one gives errors of about 13 and 6% at second- and fourth-order pt, respectively, better than the best TSDA result (*cell localization*, where the errors at second- and fourth-order pt are 15.4 and 13.6%). However, the convergence is still slow, particularly in higher order where the results oscillate with slowly decreasing amplitude, similarly to TSDA (see Figs. 5, 7, and 9). The details of these other transformations are omitted here for the above reason and the fact that they were obtained in a rather *ad hoc* way, not as systematically as in TSDA. A striking difference between the best *ad hoc* trial and the *cell localization* TSDA is the following. Two of the transformed p orbitals of the *ad hoc* trial (the one with errors 13 and 6%) were deliberately guided to be close to the Zhang-Rice-type orbital[30] (where a transformed orbital localized in a cell consists of four oxygen p orbitals in the cell with *equal* weight). On the other hand, the p^l_1 and p^r_1 in the *cell localization* choice of TSDA are further from the Zhang-Rice-type orbital, which is a result of the low symmetry of the Cu_2O_7 cluster plus the free-energy minimization. This suggests that being closer to the Zhang-

Rice-type orbital helps the convergence. However, one can see that this issue becomes moot when the *cell localization* TSDA is applied to the *crystal*, where the fourfold symmetry at a Cu site holds and therefore the *cell localization* TSDA will give the Zhang-Rice-type orbital.

After all our considerations of one-body transformations, we guess that we are close to their limit in connection with improving convergence. If one wants to do better, two- or more-body transformation is probably necessary. It is the large covalence and strong correlation in the problem that limits the effect of one-body transformations: a one-body transformation designed to give some zero-order energy levels accurately will likely spoil other levels, all of which enter pt in high order.

We then found that a minimal two-body transformation dramatically improved the convergence rate. Only transformed d_i-like and p_i^s-like orbitals are essential to be taken into account in extracting the Hamiltonian submatrix to solve for ‘‘good’’ two-particle basis states. In comparison, our cell-localization one-body+minimal two-body transformation scheme achieves similar (high) precision by second-order pt as obtained in the work of Ref. 7.

ACKNOWLEDGMENT

This work was partially supported by NSF Grant No. DMR 92-13824.

APPENDIX A: A PHYSICAL INTERPRETATION OF THE TSDA EQUATIONS

In the one-body terms of the Hamiltonian, $h_{\nu\mu} c_\nu^+ c_\mu (\nu \neq \mu)$ is the so-called hopping term that causes a particle initially in the one-particle state μ to hop to the one-particle state ν, provided the one-particle state ν is empty initially. Therefore, for an arbitrary single-determinantal state $|\Psi\rangle$, the expectation value of the operator $h_{\nu\mu} n_\mu (1-n_\nu)$ gives the amplitude of the hopping $\mu \rightarrow \nu$ when $|\Psi\rangle$ is operated on by the hopping terms of the Hamiltonian; i.e.,

$$\langle \Psi | h_{\nu\mu} n_\mu (1-n_\nu) | \Psi \rangle = \langle \Psi' | H'_{\text{1-body}} | \Psi \rangle, \tag{A1}$$

where

$$H'_{\text{1-body}} \equiv (H-H_a)_{\text{1-body}} = \sum_{\alpha \neq \beta} h_{\alpha\beta} C_\alpha^+ C_\beta$$

and $|\Psi'\rangle$ is the single-determinantal state that differs from $|\Psi\rangle$ by replacing, if there is any, the one-particle state μ in $|\Psi\rangle$ with the one-particle state ν. For the same reason, the operator $h_{\nu\mu} n_\nu (1-n_\mu)$ (note that $h_{\mu\nu} = h_{\nu\mu}$ are assumed to be real here) corresponds to the reverse hopping $\nu \rightarrow \mu$. The difference between these two operators,

$$h_{\nu\mu} n_\nu (1-n_\mu) - h_{\nu\mu} n_\mu (1-n_\nu) = h_{\nu\mu} (n_\nu - n_\mu), \tag{A2}$$

reflects, in some sense, the ‘‘net’’ hopping $\nu \rightarrow \mu$. Therefore, we define it as the net hopping operator due to the one-body terms of the Hamiltonian.

Now, let us turn to the two-body terms of the Hamiltonian. The analogous hopping terms which cause a particle in the one-particle state μ to hop to another one-particle state ν while other particles in the system do not move are $v_{\nu\lambda,\mu\lambda} c_\nu^+ c_\lambda^+ c_\lambda c_\mu$ and $v_{\nu\lambda,\lambda\mu} c_\nu^+ c_\lambda^+ c_\mu c_\lambda$. What is different from that of the one-body terms is that this hopping must occur in the presence of another particle in the orbital λ. Also keep in mind that the second term, $v_{\nu\lambda,\lambda\mu} c_\nu^+ c_\lambda^+ c_\mu c_\lambda$, introduces an extra minus sign in the final state from that of the first term, $v_{\nu\lambda,\mu\lambda} c_\nu^+ c_\lambda^+ c_\lambda c_\mu$. Analogous to the case of the one-body terms of the Hamiltonian, for an arbitrary single-determinantal state $|\Psi\rangle$ the expectation value of the operator $\Sigma_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda n_\mu (1-n_\nu)$ gives the amplitude of the hopping $\mu \rightarrow \nu$ when $|\Psi\rangle$ is operated on by the two-body terms of the Hamiltonian,

$$\left\langle \Psi \middle| \sum_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda n_\mu (1-n_\nu) \middle| \Psi \right\rangle = \langle \Psi' | H'_{\text{two-body}} | \Psi \rangle, \tag{A3}$$

where $H'_{\text{two-body}}$ denotes the two-body terms of $H-H_a$ and $|\Psi'\rangle$ has the same meaning as in Eq. (A1); and correspondingly, $\Sigma_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda n_\nu (1-n_\mu)$ accounts for the reverse hopping $\nu \rightarrow \mu$. The difference,

$$\sum_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda n_\nu (1-n_\mu) - \sum_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda n_\mu (1-n_\nu) = \sum_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda (n_\nu - n_\mu), \tag{A4}$$

is defined as the net hopping $\nu \rightarrow \mu$ due to the two-body terms of the Hamiltonian. Then it is natural that the total net hopping operator is

$$h_{\nu\mu} (n_\nu - n_\mu) + \sum_\lambda (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) n_\lambda (n_\nu - n_\mu), \tag{A5}$$

which is just the left-hand side of the TSDA equation (10) without the thermal average acting on it. Thus we get another view of TSDA; namely, the solutions of TSDA give zero thermal average of the net hopping operator (again, note that here the thermal averages are not with respect to the exact Hamiltonian H but to the approximation H_a).

APPENDIX B: SOLVING THE TSDA EQUATIONS

Since the use of the TSDA in the present context is not commonly seen, we present the details to aid in further pursuit of the matter. The numerical difficulty mentioned in the text and its significance are discussed in this section. In the following, we consider *site localization* as an example to explain the difficulty and show how we circumvent it.

The TSDA equations are [Eq. (10) of the text]

$$\langle n_\nu - n_\mu \rangle h_{\nu\mu} + \sum_\lambda \langle (n_\nu - n_\mu) n_\lambda \rangle (v_{\nu\lambda,\mu\lambda} - v_{\nu\lambda,\lambda\mu}) = 0 \tag{B1}$$

for any pair of orbitals labeled ν,μ. The thermal average bracket, $\langle\cdots\rangle$, is defined as, for any operator A,

$$\langle A\rangle \equiv \frac{\mathrm{tr}(e^{-\beta H_a}A)}{\mathrm{tr}(e^{-\beta H_a})} = \frac{\sum \langle\cdots,N_\nu,\cdots|e^{-\beta H_a}A|\cdots,N_\nu,\cdots\rangle}{\sum \langle\cdots,N_\nu,\cdots|e^{-\beta H_a}|\cdots,N_\nu,\cdots\rangle}, \tag{B2}$$

where the summation is over all possible occupations. The subscripts in Eq. (B1) label both the spatial and spin quantum numbers of the one-particle states. In the Hubbard model, the one-body and two-body matrix elements, $h_{\nu\mu}$ and $v_{\nu\mu,\lambda\kappa}$, preserve spin (i.e., $h_{\nu\mu}\equiv h_{i\sigma,j\sigma'}=0$ if $\sigma\neq\sigma'$; similarly $v_{\nu\mu,\lambda\kappa}\equiv v_{i\sigma js,k\sigma' ls'}=0$ if $\sigma\neq\sigma'$ or $s\neq s'$) and are spin independent (i.e., $h_{i\sigma,j\sigma}\equiv h_{ij}$; similarly, $v_{i\sigma js,k\sigma ls}\equiv v_{ij,kl}$). So the TSDA equations in this case can be written in the following form:

$$\langle n_{i\sigma}-n_{j\sigma}\rangle h_{ij}+\sum_{k,\sigma'}\langle (n_{i\sigma}-n_{j\sigma})n_{k\sigma'}\rangle (v_{ik,jk}-\delta_{\sigma\sigma'}v_{ik,kj}) =0 \tag{B3}$$

for any pair of spin orbitals $\psi_{i\sigma}$ and $\psi_{j\sigma}$. In the large β regime in which we are interested, the exponential factors in the thermal averages of the number operators in Eq. (B3) cause numerical difficulty and need special treatment.

In viewing the diagram of the $\mathrm{Cu_2O_7}$ cluster (Fig. 2), it is apparent that all the orbitals can be divided into four groups:

$$(d_1',d_2'),\quad (p_1'),\quad (p_2',p_4',p_5',p_7'),\quad (p_3',p_6'). \tag{B4}$$

Each orbital is said to be equivalent by symmetry to any other orbital in the same group. Because no original orbital (unprimed) will occur in this discussion, the prime on the new orbitals will be suppressed hereafter for brevity. By symmetry requirements, it is easy to see that Eq. (B3) for any pair of orbitals is automatically satisfied if the two orbitals of this pair are in the same group. These trivial equations give us no information in determining the new orbitals. Further, careful counting shows that, among those remaining nontrivial TSDA equations, there are precisely nine independent equations; they are the TSDA equations for the pairs

$$\begin{gathered}(d_1\uparrow,p_1\uparrow),\quad (d_1\uparrow,p_2\uparrow),\quad (d_1\uparrow,p_3\uparrow),\\ (d_1\uparrow,p_5\uparrow),\quad (d_1\uparrow,p_6\uparrow),\quad (p_1\uparrow,p_2\uparrow),\\ (p_1\uparrow,p_3\uparrow),\quad (p_2\uparrow,p_3\uparrow),\quad (p_2\uparrow,p_6\uparrow).\end{gathered} \tag{B5}$$

(The corresponding equations for the down-spin pairs are identical to these.) If β^{-1} is far below the transfer energy $\varepsilon_p-\varepsilon_d$ but much larger than the parallel-antiparallel splitting of the underlying states, i.e., $K_{dd}\ll\beta^{-1}\ll\varepsilon_p-\varepsilon_d$, then the states that have one hole in each of the Cu d orbitals will dominate the thermal averages and the weights they contribute to the thermal average are essentially equal no matter what their spin configurations are. For the $\mathrm{Cu_2O_7}$ cluster this means

$$\langle n_{d_1\uparrow}\rangle=\langle n_{d_1\downarrow}\rangle=\langle n_{d_2\uparrow}\rangle=\langle n_{d_2\downarrow}\rangle\cong\frac{1}{2}$$

$$\langle n_{p_i\sigma}\rangle\cong 0 \quad \text{for all } i=1\ldots 7 \text{ and } \sigma=\uparrow,\downarrow. \tag{B6}$$

Because the TSDA equations for the last four pairs of Eq. (B5) only involve the thermal averages $\langle n_{p_i\sigma}\rangle$ and $\langle n_{p_i\sigma}n_\nu\rangle$, their numerical values for the left-hand side of Eq. (B3) are very small. This causes difficulty in solving numerically the nine simultaneous TSDA equations resulting from Eq. (B5) since the computer program may only ''see'' the first five equations of Eq. (B5) and regard the last four as having been solved. To circumvent this problem, we keep only the leading terms in the thermal averages, substitute them into the TSDA equations, and then extract the limiting forms of the equations in the large β regime, as shown below.

The lowest energy group of the half-filled $\mathrm{Cu_2O_7}$ cluster is the set of four single determinants,

$$\Psi(d_1\sigma,d_2\sigma'), \tag{B7}$$

where σ and σ' are either ↑ or ↓. The next three higher-energy groups are

$$\begin{aligned}&1.\ \Psi(d_1\sigma,p_1\sigma'),\quad \Psi(d_2\sigma,p_1\sigma'),\\ &2.\ \Psi(d_2\sigma,p_2\sigma'),\quad \Psi(d_2\sigma,p_4\sigma'),\\ &\quad\ \Psi(d_1\sigma,p_5\sigma'),\quad \Psi(d_1\sigma,p_7\sigma'),\\ &3.\ \Psi(d_2\sigma,p_3\sigma'),\quad \Psi(d_1\sigma,p_6\sigma').\end{aligned} \tag{B8}$$

[We do not know the energy order of the three groups in Eq. (B8) before we solve the TSDA equations.] To circumvent the numerical difficulty mentioned above, we make two assumptions about β: (i) β^{-1} is sufficiently low such that in the thermal average the exponential weight due to a lower-energy group numerically overwhelms the weights due to other higher-energy groups. (ii) But at the same time β^{-1} is sufficiently high such that the parallel-antiparallel splittings are to be neglected and average energies (over four possible spin configurations) are to be used in substitution of real individual energies. For example, define $E(d_1\sigma,p_1\sigma')$ as the energy of state $\Psi(d_1\sigma,p_1\sigma')$ and $E(d_1p_1)$ as the average of $E(d_1\sigma,p_1\sigma')$ over the four possible spin configurations,

$$\begin{aligned}E(d_1p_1)&=\tfrac{1}{4}\sum_{\sigma\sigma'}E(d_1\sigma,p_1\sigma')\\ &=\tfrac{1}{4}\sum_{\sigma\sigma'}(h_{d_1d_1}+h_{p_1p_1}+v_{d_1p_1,d_1p_1}-\delta_{\sigma\sigma'}v_{d_1p_1,p_1d_1})\\ &=h_{d_1d_1}+h_{p_1p_1}+v_{d_1p_1,d_1p_1}-\tfrac{1}{2}v_{d_1p_1,p_1d_1}.\end{aligned} \tag{B9}$$

Because there is essentially no energy resolution among the different spin configurations by our assumption, we are to replace $E(d_1\sigma,p_1\sigma')$ with $E(d_1p_1)$ in our calculation. The energies of other states are similarly treated. We will come back to our assumption about β later. Some thermal averages of the number operators are listed below as examples:

$$\begin{aligned}\langle n_{d_1\sigma}\rangle&\cong\frac{1}{Z}\,(e^{-\beta E(d_1\sigma,d_2\uparrow)}+e^{-\beta E(d_1\sigma,d_2\downarrow)})\\ &\cong 2\langle n_{d_1\sigma}n_{d_2\sigma'}\rangle\cong\tfrac{1}{2},\end{aligned}$$

$$\langle n_{p_1\sigma}\rangle \cong \frac{1}{Z}\,(e^{-\beta E(p_1\sigma,d_1\uparrow)}+e^{-\beta E(p_1\sigma,d_1\downarrow)}+e^{-\beta E(p_1\sigma,d_2\uparrow)}$$
$$+e^{-\beta E(p_1\sigma,d_2\downarrow)})\cong 4\langle n_{p_1\sigma}n_{d_1\sigma'}\rangle\cong 0,$$

$$\langle n_{p_2\sigma}\rangle \cong \frac{1}{Z}\,(e^{-\beta E(p_2\sigma,d_2\uparrow)}+e^{-\beta E(p_2\sigma,d_2\downarrow)})\cong 2\langle n_{p_2\sigma}n_{d_2\sigma'}\rangle$$
$$\cong 0,$$

$$\langle n_{p_3\sigma}\rangle \cong \frac{1}{Z}\,(e^{-\beta E(p_3\sigma,d_2\uparrow)}+e^{-\beta E(p_3\sigma,d_2\downarrow)})\cong 2\langle n_{p_3\sigma}n_{d_2\sigma'}\rangle$$
$$\cong 0, \tag{B10}$$

where Z denotes the partition function. The TSDA equation for the pair $(d_1\uparrow,p_1\uparrow)$, therefore, reduces approximately to a simple form in our assumed β regime:

$$(\langle n_{d_1\uparrow}\rangle-\langle n_{p_1\uparrow}\rangle)h_{d_1p_1}+\sum_{k,\sigma}\,(\langle n_{d_1\uparrow}n_{k\sigma}\rangle-\langle n_{p_1\uparrow}n_{k\sigma}\rangle)$$
$$\times(v_{d_1k,p_1k}-\delta_{\uparrow\sigma}v_{d_1k,kp_1})=0$$
$$\Rightarrow(\tfrac{1}{2}-0)h_{d_1p_1}+(\langle n_{d_1\uparrow}n_{d_2\uparrow}\rangle-0)$$
$$\times(v_{d_1d_2,p_1d_2}-v_{d_1d_2,d_2p_1})+(\langle n_{d_1\uparrow}n_{d_2\downarrow}\rangle-0)$$
$$\times(v_{d_1d_2,p_1d_2})=0$$
$$\Rightarrow h_{d_1p_1}+(v_{d_1d_2,p_1d_2}-\tfrac{1}{2}v_{d_1d_2,d_2p_1})=0.$$

The TSDA equations for the second to fifth pairs in Eq. (B5) are similar in form to that for $(d_1\uparrow,p_1\uparrow)$. Now we examine the TSDA equation for the sixth pair $(p_1\uparrow,p_2\uparrow)$:

$$(\langle n_{p_1\uparrow}\rangle-\langle n_{p_2\uparrow}\rangle)h_{p_1p_2}+\sum_{k,\sigma}\,(\langle n_{p_1\uparrow}n_{k\sigma}\rangle-\langle n_{p_2\uparrow}n_{k\sigma}\rangle)(v_{p_1k,p_2k}-\delta_{\uparrow\sigma}v_{p_1k,kp_2})=0$$
$$\Rightarrow(\langle n_{p_1\uparrow}\rangle-\langle n_{p_2\uparrow}\rangle)h_{p_1p_2}+(\langle n_{p_1\uparrow}n_{d_1\uparrow}\rangle-0)(v_{p_1d_1,p_2d_1}-v_{p_1d_1,d_1p_2})+(\langle n_{p_1\uparrow}n_{d_1\downarrow}\rangle-0)(v_{p_1d_1,p_2d_1})+(\langle n_{p_1\uparrow}n_{d_2\uparrow}\rangle$$
$$-\langle n_{p_2\uparrow}n_{d_2\uparrow}\rangle)(v_{p_1d_2,p_2d_2}-v_{p_1d_2,d_2p_2})+(\langle n_{p_1\uparrow}n_{d_2\downarrow}\rangle-\langle n_{p_2\uparrow}n_{d_2\downarrow}\rangle)(v_{p_1d_2,p_2d_2})=0$$
$$\Rightarrow(2\langle n_{p_1\uparrow}n_{d_i\sigma}\rangle-\langle n_{p_2\uparrow}n_{d_2\sigma}\rangle)h_{p_1p_2}+\langle n_{p_1\uparrow}n_{d_i\sigma}\rangle(v_{p_1d_1,p_2d_1}-\tfrac{1}{2}v_{p_1d_1,d_1p_2})$$
$$+(\langle n_{p_1\uparrow}n_{d_i\sigma}\rangle-\langle n_{p_2\uparrow}n_{d_2\sigma}\rangle)(v_{p_1d_2,p_2d_2}-\tfrac{1}{2}\,v_{p_1d_2,d_2p_2})=0,$$

where in the last expression $d_i\sigma$ can be either $d_1\uparrow$, $d_1\downarrow$, $d_2\uparrow$, or $d_2\downarrow$. If $E(p_1d_1)<E(p_2d_2)$, then $\langle n_{p_1\uparrow}n_{d_1\sigma}\rangle \gg \langle n_{p_2\uparrow}n_{d_2\sigma}\rangle$ by the former assumption about β and the previous equation reduces to

$$2h_{p_1p_2}+(v_{p_1d_1,p_2d_1}-\tfrac{1}{2}v_{p_1d_1,d_1p_2})+(v_{p_1d_2,p_2d_2}-\tfrac{1}{2}v_{p_1d_2,d_2p_2})=0;$$

on the other hand, if $E(p_1d_1)>E(p_2d_2)$, then $\langle n_{p_1\uparrow}n_{d_1\sigma}\rangle \ll \langle n_{p_2\uparrow}n_{d_2\sigma}\rangle$ and the equation reduces to

$$h_{p_1p_2}+(v_{p_1d_2,p_2d_2}-\tfrac{1}{2}v_{p_1d_2,d_2p_2})=0.$$

The limiting forms of the TSDA equations for the seventh to ninth pairs in Eq. (B5) can be obtained in a manner similar to that for $(p_1\uparrow,p_2\uparrow)$. All nine equations are summarized below:

$$(d_1\uparrow,p_i\uparrow):h_{d_1p_1}+(v_{d_1d_2,p_id_2}-\tfrac{1}{2}v_{d_1d_2,d_2p_i})=0\quad\text{for } i=1,2,3,5,6,$$
$$(p_1\uparrow,p_2\uparrow):2h_{p_1p_2}+(v_{p_1d_1,p_2d_1}-\tfrac{1}{2}v_{p_1d_1,d_1p_2})+(v_{p_1d_2,p_2d_2}-\tfrac{1}{2}v_{p_1d_2,d_2p_2})=0\quad\text{if } E(p_1d_1)<E(p_2d_2),$$
$$h_{p_1p_2}+(v_{p_1d_2,p_2d_2}-\tfrac{1}{2}v_{p_1d_2,d_2p_2})=0\quad\text{if } E(p_1d_1)>E(p_2d_2),$$
$$(p_1\uparrow,p_3\uparrow):2h_{p_1p_3}+(v_{p_1d_1,p_3d_1}-\tfrac{1}{2}v_{p_1d_1,d_1p_3})+(v_{p_1d_2,p_3d_2}-\tfrac{1}{2}v_{p_1d_2,d_2p_3})=0\quad\text{if } E(p_1d_1)<E(p_3d_2),$$
$$h_{p_1p_3}+(v_{p_1d_2,p_3d_2}-\tfrac{1}{2}v_{p_1d_2,d_2p_3})=0\quad\text{if } E(p_1d_1)>E(p_3d_2),$$
$$(p_2\uparrow,p_3\uparrow):h_{p_2p_3}+(v_{p_2d_2,p_3d_2}-\tfrac{1}{2}v_{p_2d_2,d_2p_3})=0,$$
$$(p_2\uparrow,p_6\uparrow):h_{p_2p_6}+(v_{p_2d_2,p_6d_2}-\tfrac{1}{2}v_{p_2d_2,d_2p_6})=0\quad\text{if } E(p_2d_2)<E(p_3d_2),$$
$$h_{p_2p_6}+(v_{p_2d_1,p_6d_1}-\tfrac{1}{2}v_{p_2d_1,d_1p_6})=0\quad\text{if } E(p_2d_2)>E(p_3d_2). \tag{B11}$$

There are $3!=6$ possible orderings of the energies $E(p_1d_1),E(p_2d_2),E(p_3d_2)$. For each possible order, we solve Eqs. (B11) and then check the consistency of the solutions with their respective presumed energy order. Those that are not self-consistent are, of course, dropped. Although the forms of the Eqs. (B11) are pretty simple, the transformed parameters $h_{\nu\mu}$ and $v_{\nu\mu,\lambda\kappa}$ in Eq. (B11) are each a polynomial function of the transformation coefficients defined in Eq. (16) of order 2 and 4, respectively. Therefore, there indeed exist many solutions. We solved Eq. (B11) numerically. By varying the starting point and length of searching steps, we got several solutions of Eq. (B11). These solutions were examined by comparing their resulting TSDA free energies and also by cross checking with the solutions obtained in the *cell localization* and *no localization* cases to confirm that the real TSDA solution (which means the one which, within the single-determinantal and the particular localization-choice framework, gives the lowest free energy) has been found and identified. It turns out $E(p_2d_2)<E(p_1d_1)<E(p_3d_2)$ for the real TSDA solution.

Now we examine the consistency of the solution with our assumption about β, which was made to simplify the TSDA equations and therefore circumvent the numerical difficulty it leads to. Based on the solution, we calculate the energies of various states. The average energies (over four possible spin configurations) of the first four lowest-energy groups, i.e., those of Eqs. (B7) and (B8), are

$E(d_1d_2)$	$E(p_1d_1)$	$E(p_2d_2)$	$E(p_3d_2)$
-2.75742	3.80976	3.40425	4.6475

(B12)

The smallest energy gap among the above four groups is between the second and the third ones, which is about 0.41. For our assumption to be right, the β^{-1} must be well below 0.41 to make a lower-energy group numerically overwhelm other higher-energy groups in contributing the weight in thermal average. However, the parallel-antiparallel splittings, which we have neglected in order to replace the energy of every state of a group with the average energy of that group, for the four groups in Eq. (B12) are

$K(d_1d_2)$	$K(p_1d_1)$	$K(p_2d_2)$	$K(p_3d_2)$
0.0288	0.7605	0.0097	0.0823

(B13)

(Note that the second one is big because the p_1 and d_1 are nearest neighbors to each other so that they are from strong hybridization of the original p_1 and d_1, and that strong hybridization also leads to big exchange energy between the new p_1 and d_1 orbitals.) So β^{-1} must be much larger than 0.76 for the neglect of parallel-antiparallel splittings to be legitimate. An inconsistency thus occurs.

We have two reasons to believe that this inconsistency does no harm to our work here. First, in addition to what we presented so far (which we call the average case), we have done the same calculation on two artificial cases: one is to assume the spins of the two particles in the Cu_2O_7 cluster can only be parallel and the other, antiparallel. As a matter of fact, the parallel case is not really artificial because it is exactly the limiting situation when $\beta^{-1}\to 0$. (Note that a state with a pair of spin-parallel particles is always lower in energy than its antiparallel counterpart by an amount of the exchange energy K between these two particles. When β^{-1} is sufficiently low, the parallel states will dominate the thermal average.) At first glance, the inconsistency suggests that we shouldn't assume β^{-1} is sufficiently high and neglect the parallel-antiparallel splittings. It turns out that the TSDA solution of the average case is somewhere between the TSDA solutions of the parallel and antiparallel cases, and, more importantly, *they only slightly differ from each other*. This suggests that the solution we obtained is close to the solution one might get by solving the exact (unsimplified) TSDA equations (for $\beta^{-1}\ll 0.41$). Second, the average case itself can also be viewed as the exact calculation for another artificial case, namely, a system that has no spin degrees of freedom and has the average energies as the energies of its levels. Our purpose is to seek a definite guideline that can lead us to a good transformation on the basis set such that the perturbation converges rapidly. So the model that we use to derive the transformation may be closely related to the system of interest, but it does not have to be the same. Also, recall that β^{-1} here is rather a parameter than a physical temperature. Therefore, the inconsistency is not really our concern as long as we can obtain a suitable transformation.

---

[1] P. W. Anderson, Phys. Rev. **115**, 2 (1959); Solid State Phys. **14**, 99 (1963).

[2] R. S. Tu and T. A. Kaplan, in Magnetism and Magnetic Materials, edited by C. D. Graham, D. H. Lander, and J. J. Rhyne, AIP Conf. Proc. No. 24 (AIP, New York, 1974), p. 218; R. S. Tu, Ph.D. thesis, Michigan State University, 1975.

[3] See B. Brandow, Adv. Phys. **26**, 651 (1977) for other criticisms and a general history of the early work.

[4] T. A. Kaplan and P. N. Argyres, Ann. Phys. **92**, 1 (1975).

[5] K.-I. Gondaira and Y. Tanabe, J. Phys. Soc. Jpn. **21**, 1527 (1966).

[6] M. S. Hybertsen, M. Schluter, and N. E. Christensen, Phys. Rev. **39**, 9028 (1989).

[7] J. H. Jefferson, H. Eskes, and L. F. Feiner, Phys. Rev. B **45**, 7959 (1992).

[8] V. I. Belinicher and A. L. Chernyshev, Phys. Rev. B **49**, 9746 (1994).

[9] T. A. Kaplan, S. D. Mahanti, Y-S Su, and K. Kubo, J. Appl. Phys. **79**, 6433 (1996).

[10] M. S. Hybertsen, E. B. Stechel, M. Schluter, and D. R. Jennison, Phys. Rev. B **41**, 11 068 (1990).

[11] A. K. McMahan, R. M. Martin, and S. Satpathy, Phys. Rev. B **38**, 6650 (1988).

[12] H. Eskes, G. A. Sawatzky, and L. F. Feiner, Physica C **160**, 424 (1989).

[13] V. Emery, Phys. Rev. Lett. **58**, 2794 (1987).

[14] C. M. Varma, S Schmitt-Rink, and E. Abrahams, Solid State Commun. **62**, 681 (1987).

[15] Hyunju Chang, J. F. Harrison, T. A. Kaplan, and S. D. Mahanti,

Phys. Rev. B **49**, 15 753 (1994).

[16] T. A. Kaplan, Hyunju Chang, S. D. Mahanti, and J. F. Harrison, in *Electronic Properties of Solids Using Cluster Methods*, edited by T. A. Kaplan and S. D. Mahanti (Plenum, New York, 1995), p. 73.

[17] J. D. Reger and A. P. Young, Phys. Rev. B **37**, 5978 (1989); M. Gross, E. Sanchez-Velasco, and E. Siggia, *ibid.* **39**, 2484 (1989).

[18] L. Mitas, in *Electronic Properties of Solids Using Cluster Methods*, edited by T. A. Kaplan and S. D. Mahanti (Plenum, New York, 1995), p. 131.

[19] R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Chem. Phys. **103**, 1481 (1995).

[20] W. Pickett, Rev. Mod. Phys. **61**, 433 (1989).

[21] These are the same parameter values as those given in Refs. 6 and 10 except that the hopping parameters t are the negatives of those given there. We thank Dr. Ellen Stechel (private communication) for verifying that these negative values are the values actually used in Ref. 10.

[22] P. O. Löwdin, J. Math. Phys. **3**, 969 (1962).

[23] I. Lindgren and J. Morrison, *Atomic Many-Body Theory*, edited by J. Peter Toennies, *Springer Series in Chemical Physics* Vol. 13 (Springer-Verlag, New York, 1982), p. 207 ff.

[24] A similar calculation was carried out for a Hamiltonian less general than Eq. (1), by H. Eskes and J. H. Jefferson, Phys. Rev. B **48**, 9788 (1993).

[25] T. A. Kaplan and P. N. Argyres, Phys. Rev. B **11**, 1775 (1975).

[26] T. A. Kaplan and R. A. Bari, J. Appl. Phys. **41**, 845 (1970); T. A. Kaplan, in *Magnetism and Magnetic Materials*, edited by J. J. Rhyne, AIP Conf. Proc. No. 5 (AIP, New York, 1972), p. 1305.

[27] W. England, L. S. Salmon, and K. Ruedenberg, *Molecular Orbitals* (Springer-Verlag, New York, 1971); H. Weinstein, R. Pauncz, and M. Cohen, *Advance in Atomic & Molecular Physics*, Vol. 7 (Academic, New York, 1971).

[28] The same contributions were found in Ref. 7 for less general parameters in H.

[29] There is an ambiguity for the exchange terms, $K_{ij}c^{+}_{i\sigma}c_{i\sigma'}c^{+}_{j\sigma'}c_{j\sigma}$, because they can also be expressed in terms of number operators when $\sigma=\sigma'$. For convenience, we take all the exchange terms as a part of the perturbation. In fact, due to the small values of the exchange parameters K_{ij}, this convention hardly affects the convergence rate.

[30] F. C. Zhang and T. M. Rice, Phys. Rev. B **37**, 3757 (1988). These authors worked within straightforward pt.