Berry phases and pairing symmetry in Holstein-Hubbard polaron systems

K. Yonemitsu
Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan

J. Zhong and H.-B. Schüttler
Center for Simulational Physics, Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602

(Received 26 May 1998)

We study the tunneling dynamics of dopant-induced hole polarons that are self-localized by electron-phonon coupling in a two-dimensional antiferromagnet. Our treatment is based on a path-integral formulation of the adiabatic (Born-Oppenheimer) approximation, combined with many-body tight-binding, instanton, constrained lattice dynamics, and many-body exact diagonalization techniques. The applicability and limitations of the adiabatic approximation in polaron tunneling problems are discussed in detail and adiabatic results are compared to exact numerical results for a two-site polaron problem. Our results are mainly based on the Holstein-tJ and, for comparison, on the Holstein-Hubbard model. We also study the effects of second-neighbor hopping and long-range electron-electron Coulomb repulsion. The polaron tunneling dynamics is mapped onto an effective low-energy Hamiltonian that takes the form of a fermion tight-binding model with occupancy-dependent, predominantly second- and third-neighbor tunneling matrix elements, excluded double occupancy, and effective intersite charge interactions. Antiferromagnetic spin correlations in the original many-electron Hamiltonian are reflected by an attractive contribution to the first-neighbor charge interaction and by Berry phase factors that determine the signs of effective polaron tunneling matrix elements. In the two-polaron case, these phase factors lead to polaron-pair wave functions of either $d_{x^2-y^2}$-wave symmetry or p-wave symmetry with zero and nonzero total pair momentum, respectively. Implications for the doping-dependent isotope effect, pseudogap, and T_c of a superconducting polaron-pair condensate are discussed and compared to observed properties of the cuprate high-T_c materials. [S0163-1829(99)04602-0]

I. INTRODUCTION

The symmetry of the superconducting order parameter in the cuprate high-T_c superconductors had been controversial[1] before phase-sensitive experiments firmly established the $d_{x^2-y^2}$-wave pairing symmetry in $YBa_2Cu_3O_7$, using tricrystal ring magnetometry,[2] superconducting quantum interference device interferometry[3], and single-junction modulation.[4] Migdal-Eliashberg–type diagrammatic theories find d-wave pairing to be favored by antiferromagnetic (AF) spin-fluctuation exchange[5] and s-wave pairing by the conventional electron-phonon mechanism.[5,6] There is indeed strong experimental evidence for the importance of *both* AF spin correlations[7] *and* electron-phonon interactions[8] in the cuprates. However, when combined in the diagrammatic approach, the two mechanisms are mutually destructive, since d-wave pairing is strongly suppressed by phonons and s-wave pairing is suppressed by AF spin fluctuations, respectively. Also, the magnitude of the observed isotope effect in cuprate systems away from ''optimal'' doping[9] points towards an unusually strong electron-phonon effect that cannot be accounted for in the diagrammatic approaches.[6]

Strong-coupling studies,[10–13] going beyond the Migdal-Eliashberg regime, suggest that the AF spin correlations themselves can effectively enhance the electron-phonon effect by lowering the electron-phonon coupling threshold for polaron formation, that is, the threshold for electron-phonon–induced self-localization[14] of the dopant-induced carriers in the CuO_2 planes. In the present paper, we show how the tunneling dynamics of such self-localized holes in an AF-correlated spin background may lead to d-wave and other non-s-wave pairing states.

A Berry phase factor in finite systems with time-reversal symmetry has been relevant to the observation of half-odd-integer quantum numbers in the spectrum of the Na_3 molecule[15] to the cross section of the $H+H_2$ reaction and its isotope analogs[16] and to the problem of integer vs half-odd-integer spin tunneling in anisotropic potentials.[17] Contributions to the pair-binding energy in the C_{60} molecule have also been discussed in terms of Berry phase arguments.[18] In the present case, the non-s-wave symmetry is caused by a (-1) Berry phase factor, associated with predominantly second- and third-neighbor polaron tunneling processes. It also determines the total momentum: the one-polaron ground state has a momentum on the Fermi surface of the half-filled tight-binding model on the square lattice. The dynamics of few hole polarons reflects the local AF spin correlations of many electrons through the Berry phase factor.

This paper is organized as follows: In Sec. II, we introduce the basic Holstein-Hubbard and Holstein-tJ model Hamiltonians and their extensions to include second-neighbor hopping or long-range Coulomb repulsion. We then derive the effective action for the lattice degrees of freedom in the adiabatic approximation. In Sec. III, we illustrate the basic physical principles and formal concepts of our adiabatic treatment of the polaron tunneling in the context of a simple two-site model. In Sec. IV, we discuss the conditions under which the adiabatic approximation is valid, as well as its limitations when applied to polaronic systems on large or macroscopic lattice systems. In particular, we clear up some

recent misunderstandings concerning the applicability of the adiabatic approach to polaronic systems. In Sec. V, we use an instanton approach to elucidate the basic structure of the low-energy tunneling dynamics of hole polarons in Holstein-Hubbard or Holstein-tJ systems near half filling. We show that the dynamics of such hole polarons is governed by an effective tight-binding Hamiltonian which includes second- and third-neighbor hopping matrix elements and a first-neighbor attraction. In Sec. VI, we discuss the Berry phase factors and, with the help of lattice symmetry operations, we show how such phases can be properly assigned to each segment of a closed tunneling path. The Berry phase factors are then interpreted in terms of quasiparticle statistics and internal symmetries of the many-electron wave functions. In Sec. VII, we analytically solve the effective model to show how the Berry phase factors determine the total momenta and internal symmetries of the few-hole-polaron wave functions. In Sec. VIII, we report numerical results for the effective polaron hopping and effective pair-binding energy as functions of the phonon frequency and electron-phonon coupling strength. In Sec. IX, we discuss the implications of our numerical results for a possible superconducting pairing instability, the isotope effect, and the pseudogap in a hole polaron liquid at finite doping concentration in the nearly half-filled Holstein-tJ and Holstein-Hubbard systems and compare the results to experimental observations in the cuprates. In Sec. X, we summarize the present work. Part of the results presented in this paper were reported briefly in an unpublished paper and proceedings.[19]

II. MODEL AND EFFECTIVE ACTION

We use mainly the Holstein-tJ model[11,12] and occasionally the Holstein-Hubbard model for comparison. Later, we also include second-neighbor electron hopping and/or long-range electron-electron repulsion terms in the model. The total Hamiltonian is of the general form

$$H=H_{\mathrm{e}}+H_{\text{e-ph}}+H_{\mathrm{ph}}, \tag{1}$$

where H_{e} is the purely electronic tJ or Hubbard model part, defined on a two-dimensional (2D) square lattice with lattice sites $j=1,\ldots,N$ and on-site electron occupation numbers n_j, as specified below.

$$H_{\text{e-ph}}=C\sum_j u_j n_j \tag{2}$$

is the Holstein electron-phonon (EP) interaction, coupling the local oscillator displacement u_j to the electron on-site occupation n_j with an EP coupling constant C and

$$H_{\mathrm{ph}}=\frac{K}{2}\sum_j u_j^2+\frac{1}{2M}\sum_j p_j^2\equiv H_K+H_M \tag{3}$$

describes the noninteracting Einstein phonon system, consisting of the bare harmonic lattice potential H_K, with restoring force constant K, and of the lattice kinetic energy H_M with an atomic mass M and conjugate momenta $p_j\equiv -i\hbar\partial/\partial u_j$. If we rescale to dimensionless displacements and conjugate momenta

$$\bar{u}_j\equiv u_j/u_{\mathrm{P}},\quad \bar{p}_j\equiv -i\partial/\partial\bar{u}_j, \tag{4}$$

with the small polaron shift

$$u_{\mathrm{P}}\equiv\frac{C}{K}, \tag{5}$$

then $H_{\text{e-ph}}$ and H_{ph} can be completely parametrized in terms of only two characteristic energies, the bare Einstein phonon energy

$$\Omega\equiv\hbar\left(\frac{K}{M}\right)^{1/2}, \tag{6}$$

and the ionic-limit ($t\to 0$) small polaron binding energy

$$E_{\mathrm{P}}\equiv\frac{C^2}{K}. \tag{7}$$

All results in the following are therefore stated in terms of u_{P}, Ω, and E_{P} only.[11,14]

The tJ model is written as[20]

$$H_{\mathrm{e}}=-t\sum_{\langle i,j\rangle,\sigma}(c_{i\sigma}^{\dagger}c_{j\sigma}+\mathrm{H.c.})+J\sum_{\langle i,j\rangle}\left(\mathbf{S}_i\cdot\mathbf{S}_j-\frac{n_i n_j}{4}\right) \tag{8}$$

with first-neighbor electron hopping t and AF exchange coupling J. Here, $c_{i\sigma}$ annihilates an electron with spin σ at site i, $n_{i\sigma}=c_{i\sigma}^{\dagger}c_{i\sigma}$, $n_i=\Sigma_\sigma n_{i\sigma}$, $\mathbf{S}_i=\frac{1}{2}\Sigma_{\alpha,\beta}c_{i\alpha}^{\dagger}\sigma_{\alpha\beta}c_{i\beta}$ with $\sigma\equiv(\sigma_x,\sigma_y,\sigma_z)$ denoting the vector of Pauli spin matrices. The Hilbert space is restricted to states with no double occupancy at any site j, i.e., $n_j=0,1$ only.

The Hubbard model is written as

$$H_{\mathrm{e}}=-t\sum_{\langle i,j\rangle,\sigma}(c_{i\sigma}^{\dagger}c_{j\sigma}+\mathrm{H.c.})+U\sum_i n_{i\uparrow}n_{i\downarrow} \tag{9}$$

with on-site repulsion U and no restrictions on the on-site occupancy, i.e., $n_j=0,1,2$. In the following, we set $\hbar\equiv 1$, $t\equiv 1$ *and use* $J=0.5t$ *or* $U=8t$ *in the* tJ *or Hubbard model, respectively, unless stated otherwise*.

In addition to the standard tJ and Hubbard electronic model, we will also study the effects of additional, potentially important electronic terms, the second-neighbor hopping $H_{t'}$, and the long-range Coulomb repulsion H_{LC}. Namely,

$$H_{t'}=-t'\sum_{\{i,j\},\sigma}(c_{i\sigma}^{\dagger}c_{j\sigma}+\mathrm{H.c.}), \tag{10}$$

where $\{i,j\}$ denotes second-neighbor bonds and t' is the corresponding second-neighbor matrix element. The long-range $1/r$ Coulomb repulsion is

$$H_{\mathrm{LC}}=\frac{1}{2}V_{\mathrm{C}}\sum_{i\neq j}\frac{n_i n_j}{|r_{ij}|}, \tag{11}$$

where i and j are summed independently over all sites excluding $i=j$ and r_{ij} denotes the vector pointing from i to j, measured in units of the 2D lattice constant $a\equiv 1$. On a lattice with periodic boundary conditions we make the definition of $|r_{ij}|$ unique by requiring r_{ij} to be a vector of the shortest possible length connecting i to j, subject to all possible periodic boundary shifts. The matrix element V_{C} is thus the Coulomb repulsion energy between two electrons at first-neighbor distance.

To study the tunneling dynamics of self-localized holes, we consider the path integrals for transition amplitudes in imaginary time in the Born-Oppenheimer (adiabatic) approximation. Following the standard Feynman-Trotter approach,[21] we break up the Hamiltonian in the imaginary-time evolution operator

$$e^{-\beta H}=\lim_{L\to\infty}(e^{-\Delta\tau H_0}e^{-\Delta\tau H_M})^L, \tag{12}$$

where $\Delta\tau\equiv\beta/L$, H_M is the lattice kinetic energy defined in Eq. (3) and the zeroth-order part $H_0\equiv H-H_M$ commutes with all lattice displacement operators u_j. At each time slice $\tau_k\equiv k\Delta\tau$, with $k=1,\ldots,L$, we now insert a complete set of electron-phonon basis states $|\chi_u^{(\kappa)}\rangle$ that are chosen to be simultaneous eigenstates of H_0 and of all u_j. They can be written in the form

$$|\chi_u^{(\kappa)}\rangle=|\Psi^{(\kappa)}(u)\rangle\times|\Phi_u\rangle, \tag{13}$$

where $|\Phi_u\rangle$ is the lattice part and $|\Psi^{(\kappa)}(u)\rangle$ the electronic part of $|\chi_u^{(\kappa)}\rangle$. Written in first-quantized notation, the lattice part is simply

$$\Phi_u(x)=\delta(u-x)=\prod_j\ \delta(u_j-x_j) \tag{14}$$

with lattice coordinate vectors $x\equiv(x_1,\ldots,x_N)$ and $u\equiv(u_1,\ldots,u_N)$. The electronic part $|\Psi^{(\kappa)}(u)\rangle$ denotes the κth electronic eigenstate of the zeroth-order adiabatic Hamiltonian

$$H_0(u)=H_{\rm e}+H_{\rm e\text{-}ph}(u)+H_K(u), \tag{15}$$

at fixed u. That is, $H_0(u)$ is defined to act only on the electronic degrees of freedom at *fixed* (c-number) lattice displacement coordinates $u\equiv(u_1\ldots u_N)$ and

$$H_0(u)|\Psi^{(\kappa)}(u)\rangle=W_0^{(\kappa)}(u)|\Psi^{(\kappa)}(u)\rangle, \tag{16}$$

where $|\Psi^{(\kappa)}(u)\rangle$ and its eigenenergy $W_0^{(\kappa)}(u)$ depend parametrically on the lattice displacements u. The exact imaginary time evolution under H can thus be represented by a path integral with a Euclidean action, written at finite L as

$$S[u(\tau),\kappa(\tau)]=\sum_{k=1}^{L}\Bigg[(M/2)\sum_j\frac{[u_j(\tau_k)-u_j(\tau_{k-1})]^2}{\Delta\tau}+\Delta\tau W_0^{(\kappa_k)}[u(\tau_k)]-\ln\langle\Psi^{(\kappa_k)}[u(\tau_k)]|\Psi^{(\kappa_{k-1})}[u(\tau_{k-1})]\rangle\Bigg]. \tag{17}$$

The path integration is to be carried out both over the continuous lattice coordinates $u(\tau_k)\equiv[u_1(\tau_k)\ldots u_N(\tau_k)]$ and over the discrete electronic quantum numbers $\kappa_k\equiv\kappa(\tau_k)$.

In the zeroth-order adiabatic approximation, corresponding formally to the $M\to\infty$ limit, one neglects the imaginary time evolution of u altogether and replaces u_k by a τ-independent classical field. The first-order adiabatic approximation restores the τ dependence of the lattice coordinates u, under the simplifying assumption that the electrons follow the motion of the lattice adiabatically. That is, the path integration is restricted to configurations, where, during τ evolution, the electrons remain in the same eigenstate, i.e., $\kappa_k=\kappa_{k-1}\equiv\kappa=$const. Transitions between different electronic eigenstates $\kappa_k\neq\kappa_{k-1}$ are neglected. Formally, this approximation restores the leading-order $1/M$ corrections to the lattice dynamics. At sufficiently low temperatures, one restricts the path integral further to include only the electronic ground state $\kappa=0$. Suppressing the (κ) superscript altogether, one then arrives at the standard first-order adiabatic (Born-Oppenheimer) approximation, with an effective Euclidean action

$$S_{\rm ad}[u(\tau)]=\sum_{k=1}^{L}\Bigg[\frac{M}{2}\sum_j\frac{[u_j(\tau_k)-u_j(\tau_{k-1})]^2}{\Delta\tau}+\Delta\tau W_0[u(\tau_k)]-\ln\langle\Psi[u(\tau_k)]|\Psi[u(\tau_{k-1})]\rangle\Bigg]. \tag{18}$$

Note that $S_{\rm ad}$ depends explicitly only on the u coordinates of the lattice. The first $(M/2)$ term is the standard form of the lattice kinetic energy for discretized imaginary time (finite L). The electronic ground-state energy $W_0(u)$ plays the role of a zeroth-order (in $1/M$) effective lattice potential energy. The last term, containing the logarithms of the electronic ground-state wave function overlaps at adjacent time slices during τ evolution, contains the Berry phase and $1/M$ corrections to the potential energy, as we will now discuss.

In $\exp(-S_{\rm ad}[u(\tau)])$, the overlap product

$$Q[u(\tau)]\equiv\prod_{k=1}^{L}\langle\Psi[u(\tau_k)]|\Psi[u(\tau_{k-1})]\rangle \tag{19}$$

exists and which contains the Berry phase factor,

$$\exp(-i\theta[u(\tau)])\equiv Q[u(\tau)]/|Q[u(\tau)]|, \tag{20}$$

i.e., $\theta[u(\tau)]=-\,\mathrm{Im}\ln(Q[u(\tau)])$. Due to time-reversal symmetry, all $|\Psi[u(\tau_k)]\rangle$ have real amplitudes in an appropriately chosen electron basis and hence the phase factor is real: $\exp(-i\theta[u(\tau)])=\mathrm{sgn}(Q[u(\tau)])$. Taking $L\to\infty$, we can also rewrite $\mathrm{Re}(\ln Q[u(\tau)])\equiv\ln|Q[u(\tau)]|$ in $S_{\rm ad}[u(\tau)]$ as a $1/M$ correction to the effective lattice potential which thus becomes

$$W(u)\equiv W_0(u)+W_1(u), \tag{21}$$

with W_1 given by

$$W_1(u)=\frac{1}{2M}\sum_j\langle\partial_{u_j}\Psi(u)|\partial_{u_j}\Psi(u)\rangle. \tag{22}$$

Thus, the effective action for $L\to\infty$ becomes

$$S_{\rm ad}[u(\tau)]=\sum_{k=1}^{L}\Bigg[\frac{M}{2}\sum_j\frac{[u_j(\tau_k)-u_j(\tau_{k-1})]^2}{\Delta\tau}+\Delta\tau W[u(\tau_k)]\Bigg]+i\theta[u(\tau)]. \tag{23}$$

Equivalent results can be derived in the Hamiltonian approach to the adiabatic approximation. The basic idea here is to restrict the full electron-lattice Hilbert space to an ''adiabatic'' subspace that is spanned by the set of zeroth-order adiabatic electron-lattice eigenstates $|\chi_u^{(\kappa)}\rangle$ defined above in Eqs. (13)–(16) with κ restricted to the electronic groundstate $\kappa=0$. The adiabatic subspace thus consists of EP states of the general form

$$|\phi\rangle=\int d^N u\ \phi(u)|\Psi_u^{(0)}\rangle, \tag{24}$$

where $\phi(u)$ is an arbitrary (square-integrable) wave-function amplitude that depends only on the lattice coordinates u. The basic approximation step is then to project the full EP Hamiltonian H onto the adiabatic subspace. In this manner one arrives at a first-order effective Hamiltonian $H_{\rm ad}$ that is mathematically equivalent to the first-order adiabatic Euclidean action $S_{\rm ad}$ in Eq. (23), after $L\to\infty$. Since the adiabatic EP states $|\phi\rangle$ can be expressed entirely in terms of their wave-function amplitude $\phi(u)$, one can recast $H_{\rm ad}$ into the form of an effective Hamiltonian acting only on the lattice coordinates u in $\phi(u)$, without explicit reference to the underlying electronic ground-state wave function $|\Psi_u^{(0)}\rangle$ contained in $|\phi\rangle$. However, it is crucial to keep in mind the formal relationship (24) between the full adiabatic EP state $|\phi\rangle$ and its lattice wave-function amplitude $\phi(u)$ if one wants to properly compare first-order adiabatic results to exact results, obtained by, e.g., numerically diagonalizing the full EP Hamiltonian on small model clusters.

In systems obeying standard harmonic lattice dynamics, the zeroth-order Born-Oppenheimer ''energy surface'' $W_0(u)$ exhibits one unique global minimum configuration $u^{(\rm min)}$ which is, in terms of energy or in terms of configurational distance, well separated from other, if existent, local minima. In that case, the path integral is dominated by small-amplitude ''harmonic'' fluctuations around $u^{(\rm min)}$ and a description of the lattice dynamics in terms of renormalized harmonic oscillators, i.e., phonons, remains valid. Since the displacement excursions around $u^{(\rm min)}$ are small, so are the fluctuations in the electronic wave function $|\Psi(u)\rangle$; hence the small-amplitude (''phonon'') paths all have $\theta[u(\tau)]=0$ and Berry phase effects are negligible. Also, the u derivatives of $|\Psi(u)\rangle$ entering into W_1 are well behaved and the mth order u derivatives of the overlap matrix elements $\langle\partial_{u_j}\Psi(u)|\partial_{u_j}\Psi(u)\rangle$ are typical of the order of inverse lattice constants or inverse atomic distances raised to the $(m+2)$th power. The W_1 contribution to the harmonic restoring force constants, for example, are thus smaller than the zeroth-order W_0 contributions by factors of order of the fourth power of the lattice oscillator zero-point displacement amplitude over the lattice constant. Thus, the electronic overlap factor effects $W_1(u)$ and $\theta[u(\tau)]$ can be altogether neglected.

By contrast, in polaronic systems the zeroth-order lattice potential W_0 exhibits a large number of nearly degenerate local minima. The low-energy lattice dynamics is dominated by tunneling processes between the local minima that requires anharmonic large-amplitude excursions of the local displacement coordinates u_j and large local rearrangements of the electronic wave function $|\Psi(u)\rangle$.[11] In that case electronic overlap effects arising from both $\theta[u(\tau)]$ and $W_1(u)$ can become quite important.

III. TWO-SITE PROBLEM

The two-site version of the Holstein-Hubbard model (1)–(3) (Refs. 22–24) is a simple toy problem that retains some essential physical features of the lattice polaron problem. We will use it here to elucidate the basic underlying physical ideas and formal concepts of our adiabatic treatment and, also, to test the validity and illustrate some important limitations of the adiabatic approximation. We restrict ourselves to the single-electron case on two sites, with an electronic intersite hybridization t. The adiabatic electronic wave function $|\Psi(u)\rangle$ can be solved exactly by diagonalizing $H_0(u)$ that reduces to a 2×2 matrix.

The two sites are labeled 1 and 2 with on-site oscillator coordinates u_1 and u_2 and on-site electron occupation numbers n_1 and n_2. With symmetrized coordinates

$$u_\pm=(u_1\pm u_2)/\sqrt{2}, \tag{25}$$

W_0 and W_1 can be written as

$$W_0(u)\equiv W_{0+}(u_+)+W_{0-}(u_-), \tag{26}$$

where

$$W_{0+}(u_+)=\frac{K}{2}u_+^2+\frac{C}{\sqrt{2}}u_+=\left[\frac{1}{2}\left(\frac{u_+}{u_{\rm P}}\right)^2+\frac{1}{\sqrt{2}}\left(\frac{u_+}{u_{\rm P}}\right)\right]E_{\rm P}, \tag{27}$$

$$W_{0-}(u_-)=\frac{K}{2}u_-^2-\sqrt{\frac{C^2u_-^2}{2}+t^2}$$
$$=\left[\frac{1}{2}\left(\frac{u_-}{u_{\rm P}}\right)^2-\sqrt{\frac{1}{2}\left(\frac{u_-}{u_{\rm P}}\right)^2+\left(\frac{t}{E_{\rm P}}\right)^2}\right]E_{\rm P}, \tag{28}$$

and

$$W_1(u)\equiv W_{1-}(u_-)=\frac{1}{4}\frac{\Omega^2}{E_{\rm P}}\frac{(t/E_{\rm P})^2}{[(u_-/u_{\rm P})^2+2(t/E_{\rm P})^2]^2}. \tag{29}$$

With the Berry phase $\theta[u(\tau)]\equiv0$, the problem is equivalent to solving the Hamiltonian of a quantum particle of mass M in a two-dimensional potential $W(u)\equiv W(u_+,u_-)=W_0(u)+W_1(u)$. Because of Eqs. (26) and (29) this dynamics is separable when written in terms of u_+- and u_--coordinates.

Since u_+ couples only to the total electron charge $n_+\equiv n_1+n_2$, W_{0+} is just a harmonic potential with shifted equilibrium position

$$u_+^{(0)}=-u_{\rm P}/\sqrt{2}, \tag{30}$$

and the electron ground state $|\Psi(u)\rangle\equiv|\Psi(u_-)\rangle$, and hence W_1, do not depend on u_+.

Since u_- couples to the charge imbalance $n_-\equiv n_1-n_2$ between the two sites, the shape of W_{0-} *is* renormalized by the EP coupling and W_1 depends on u_-. As shown in Fig.

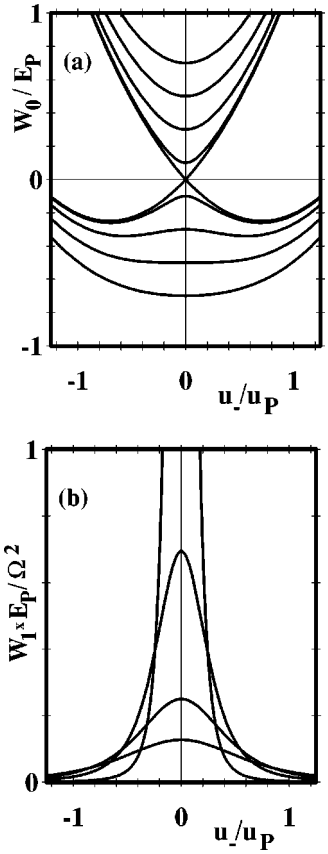

FIG. 1. Zeroth- and first-order contributions W_0 and W_1 to the effective adiabatic lattice potential in the 2-site Holstein model with 1 electron. In (a), the electronic ground-state energy W_0/E_P and the first-excited-state energy $W_0^{(1)}/E_P$ of the electronic Hamiltonian $H_0(u)$ are shown as functions of u_-/u_P, at $u_+=0$ for $t/E_P=0$, 0.1, 0.3, 0.5, and 0.7. In (b), $W_1\times E_P/\Omega^2$ vs u_-/u_P is shown at $u_+=0$ for $t/E_P=0.1$, 0.3, 0.5, and 0.7.

1(a), W_{0-} retains a single global minimum at $u_-=0$, for small E_P, with a harmonic restoring force constant

$$K_{0-}\equiv\frac{\partial^2}{\partial u_-^2}W_{0-}(u_-=0)=\left(1-\frac{E_P}{2t}\right)K \tag{31}$$

that softens with increasing E_P and changes sign when E_P reaches a critical value

$$E_P^{(\mathrm{crit})}=2t. \tag{32}$$

For $E_P>E_P^{(\mathrm{crit})}$, W_{0-} acquires two degenerate minima at $u_-=\pm u_-^{(0)}$, separated by a maximum at $u_-=0$, with

$$u_-^{(0)}=\sqrt{1-\left(\frac{2t}{E_P}\right)^2}\frac{u_P}{\sqrt{2}}, \tag{33}$$

where $u_P=C/K$ is the polaron shift (5). $u_-^{(0)}$ approaches $u_P/\sqrt{2}$ in the strong-coupling limit

$$E_P\gg t. \tag{34}$$

The height of the zeroth-order potential barrier separating the two minima,

$$\Delta_{B0}\equiv W_{0-}(0)-W_{0-}(\pm u_-^{(0)})=\left(\frac{1}{2}-\frac{t}{E_P}\right)^2E_P, \tag{35}$$

increases with E_P and approaches $E_P/4$ in the strong-coupling limit (34).

The physical origin of the double-well potential can be most easily understood starting from the ''ionic'' ($t=0$) limit of the model: For $t=0$, the two electronic eigenstates of $H_0(u)$,

$$|\Psi^{(l1)}\rangle\equiv|n_1=1,\ n_2=0\rangle \tag{36}$$

and

$$|\Psi^{(l2)}\rangle\equiv|n_1=0,\ n_2=1\rangle \tag{37}$$

have the electron completely localized on sites 1 and 2, respectively, with eigenenergies

$$W^{(l1,2)}(u_-)=\frac{K}{2}(u_-\pm u_P/\sqrt{2})^2-\frac{1}{4}E_P, \tag{38}$$

where the upper (lower) sign refers to $W^{(l1)}$ ($W^{(l2)}$), as shown by the two parabolic potential curves in Fig. 1(a). Assuming $C>0$, $|\Psi^{(l1)}\rangle$ is the ground state for $u_-<0$ and $|\Psi^{(l2)}\rangle$ for $u_->0$. At $u_-=0$, the two parabolic eigenenergy curves $W^{(l1)}(u_-)$ and $W^{(l2)}(u_-)$ intersect, both states are degenerate, and the ground-state wave function changes discontinuously as a function of u_-. When the hybridization t is turned on, the two fully localized wave functions $|\Psi^{(l1)}\rangle$ and $|\Psi^{(l2)}\rangle$ become mixed, the electronic degeneracy at $u_-=0$ is lifted, and a minimum excitation gap of $2t$ opens up between the electronic ground state and first excited state. The sharp cusp at $u_-=0$ in the $t=0$ double-parabolic potential function

$$\begin{aligned}W_{0-}(u_-)|_{t=0}&=\min[W^{(l1)}(u_-),W^{(l2)}(u_-)]\\&=\frac{K}{2}(|u_-|-u_P/\sqrt{2})^2-\frac{1}{4}E_P\end{aligned} \tag{39}$$

[see Fig. 1(a)] is rounded by the finite t; as a function of u_-, the ground-state wave function $|\Psi(u_-)\rangle$ now changes continuously at $u_-=0$. However, $|\Psi(u_-)\rangle$ still has predominantly $|\Psi^{(l1)}\rangle$ character near $u_-=-u_-^{(0)}$ and predominantly $|\Psi^{(l2)}\rangle$ character near $u_-=u_-^{(0)}$. With increasing t, the tunneling barrier height (35) decreases, initially by about t. The barrier vanishes when t reaches $t^{(\mathrm{crit})}=E_P/2$ that is equivalent to the above condition (32), for $E_P^{(\mathrm{crit})}$.

From the ground-state wave function $|\Psi(u_-)\rangle$ and its charge distribution $\langle\Psi(u_-)|n_j|\Psi(u_-)\rangle$ for $E_P>E_P^{(\mathrm{crit})}$, near the two potential minima $\pm u_-^{(0)}$, one thus finds the electron predominantly localized at site 1 when $u_-\cong-u_-^{(0)}$ and predominantly at site 2 when $u_-\cong+u_-^{(0)}$, assuming again $C>0$ here and in the following. By contrast, at the potential minimum $u_-=0$ in the regime $E_P<E_P^{(\mathrm{crit})}$, the electron charge is delocalized evenly between sites 1 and 2. Thus, at the level of the zeroth-order adiabatic approximation, the transition from the single-well potential case $E_P<E_P^{(\mathrm{crit})}$ to the double-well case $E_P>E_P^{(\mathrm{crit})}$ is essentially a transition from a delocalized nondegenerate ground state ($u_-=0$) to a localized degenerate ground state ($u_-=\pm u_-^{(0)}$). In the

former case, the electron's delocalization energy dominates and the electron wave function spreads out between the two sites. In the latter case, the EP coupling dominates; the lattice spontaneously distorts so as to set up an attractive EP ''potential well'' that binds and localizes the electron. The electronic binding energy thus gained in turn stabilizes the local lattice distortion. This self-localization mechanism is the essence of polaron formation.

Localizing the electron on either one of the two sites is energetically equivalent due to the reflection symmetry $[(1,2)\rightarrow(2,1)]$ of the underlying Hamiltonian. At the level of the zeroth-order adiabatic approximation, this symmetry is broken in the twofold degenerate zeroth-order ground states $u_- = \pm u_-^{(0)}$. The existence of two *degenerate* local minima in W_{0-} can thus be understood as a direct consequence of the symmetry breaking that accompanies the self-localization transition. In the first-order adiabatic approximation, the lattice kinetic energy restores this symmetry by inducing tunneling processes between the two potential minima, thus giving rise to a nondegenerate ground state in which the two degenerate zeroth-order states are admixed with equal probability weight.

The *lattice* tunneling processes, within the multiple-well Born-Oppenheimer potential, constitute the basic low-energy mechanism whereby self-localized electrons can move through the lattice. At higher temperatures, thermally activated hopping over the barrier may dominate the polaronic charge transfer;[14,22] this, again, can be described as a purely lattice dynamical phenomenon. Thus, within the framework of the first-order adiabatic approximation, polaron formation and polaron dynamics is fundamentally reduced to a problem of *nonlinear lattice* dynamics.

We now turn to the first-order potential correction $W_{1-}(u_-)$ (29) in the two-site problem, shown for several values of t/E_P in Fig. 1(b). Since $W_1(u)$, according to Eq. (22), is controlled by the u gradient of the electron wave function $|\Psi(u)\rangle$, we should expect it to exhibit peaks wherever $|\Psi(u)\rangle$ varies most rapidly with u. In the two-site problem, this occurs at $u_-=0$, where $|\Psi(u)\rangle$ changes its character from being predominantly localized on site 1 to being localized on site 2, as discussed above. For large $|u_-|$, $|\Psi(u_-)\rangle$ approaches a constant, either $|\Psi^{(l1)}\rangle$ or $|\Psi^{(l2)}\rangle$, hence $W_{1-}(u_-)\rightarrow 0$ for $|u_-|\rightarrow\infty$. In the polaron regime $E_P > E_P^{(\mathrm{crit})}$, the primary effect of W_1 is to enhance the tunneling barrier separating the two potential minima. In addition, $W_{1-}(u_-)$ will also tend to shift the two polaronic potential minima further apart, thus causing the tunneling barrier to become wider than in the zeroth-order potential W_0. Both of these W_1 effects tend to suppress the tunneling rate through the barrier. Even though $W_1(0)$ may be small compared to the zeroth-order barrier height Δ_{B0} (35), its effect on the polaron tunneling rates can be quantitatively of some importance, since tunneling rates are typically exponentially sensitive to changes in the tunneling barrier.

In the delocalized regime $E_P < E_P^{(\mathrm{crit})}$, the primary effect of W_1 is to soften the harmonic restoring force constant $K_- \equiv \partial_{u_-}^2 W_-(0)$ by an amount

$$K_{1-} \equiv \frac{\partial^2}{\partial u_-^2} W_{1-}(u_-=0) = -\frac{1}{8}\left(\frac{\Omega}{t}\right)^2\left(\frac{E_P}{t}\right)^2 K < 0. \tag{40}$$

Thus, W_1 also lowers the critical E_P for the onset of polaron formation. However, in the large-M limit where the adiabatic approximation is valid, that is, for $\Omega \ll t$ (see below), these corrections are smaller than the zeroth-order results (31) and (32) by factors of order $(\Omega/t)^2(E_P/t)^2$. Provided t and E_P are of comparable magnitude and $\Omega \lesssim t$ (see below), W_1 does not qualitatively alter the basic structure of the lattice potential W in either coupling-strength regime. However, W_1 can become qualitatively important in suppressing certain nonadiabatic processes in lattice systems, as will be discussed in the next section.

IV. VALIDITY AND LIMITATIONS OF THE ADIABATIC APPROXIMATION

The basic criterion for the validity of the adiabatic approximation is that the longest time scale of the electronic motion should be short compared to the shortest time scale of the lattice motion or, equivalently, the lowest electronic frequency scale should be large compared to the highest lattice frequency scale. In the two-site problem, the lowest electronic frequency scale is the excitation energy between the electronic ground state $|\Psi(u)\rangle$ and the first excited state that is at least $2t$ (at $u_-=0$) or larger. The highest lattice frequency scale is the phonon energy Ω and hence we expect the adiabatic approximation to work, provided that

$$\Omega \ll 2t. \tag{41}$$

In the polaron regime $E_P > E_P^{(\mathrm{crit})}$, the lattice (not the electron) motion acquires an additional, low-frequency scale, given by the polaronic tunneling splitting $2t_P$ between the ground state and first excited state in the double-well lattice potential $W(u)$. This tunneling energy scale is typically smaller than or, at most, comparable to the bare phonon energy scale Ω, given the conditions where a polaronic double-well forms in the first place. Hence, the basic criterion (41) applies in the polaronic regime just as well as in the delocalized regime, regardless of the electron-phonon coupling strength. Criterion (41) applies even in the strong-coupling regime (34) where $2t_P$ becomes orders of magnitude smaller than Ω.

While the foregoing considerations are well known,[14,22,24] we reemphasize them here because there has been some confusion about this in the more recent literature on the two-site problem. In more recent work, it is sometimes assumed that the polaron tunneling splitting $2t_P$, rather than $2t$, represents the lowest relevant electronic energy scale. Doing so, one then arrives at the much too restrictive validity criterion

$$\Omega \ll 2t_P. \tag{42}$$

If correct, this would imply that the polaron regime $E_P > E_P^{(\mathrm{crit})}$ cannot be treated in the adiabatic approximation, since typically $t_P \lesssim \Omega$ even under the most favorable conditions. In the strong-coupling regime (34), where $t_P \ll \Omega$, the adiabatic approximation should break down completely according to Eq. (42).

The flaw in the foregoing argument is that $2t_P$ is of course *not* the lowest electronic energy scale, but rather represents an energy scale of the lattice motion, as discussed above. The relevant lowest electronic energy splitting, be-

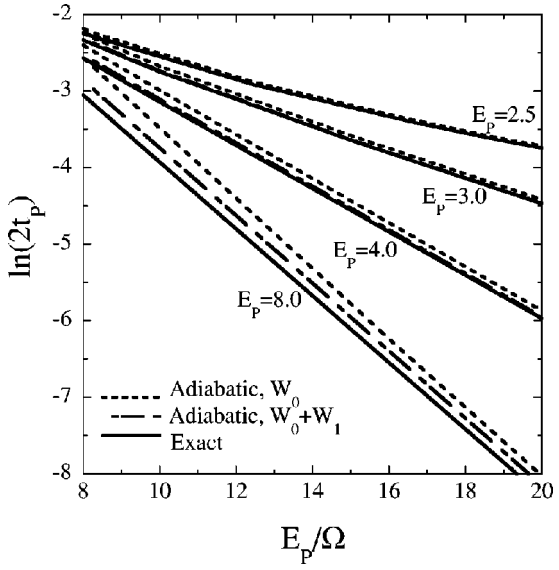

FIG. 2. Exact and adiabatic results for the tunneling energy splitting $2t_P$ between the ground-state and the first-excited-state electron-phonon eigenstate, plotted as a function of E_P/Ω, in the 2-site Holstein model with 1 electron, for $t=1$ and several values of E_P, as indicated.

tween the electronic ground state and first excited state *at fixed lattice coordinate u* is at least $2t$ in the two-site model, regardless of whether E_P is small or large.

To illustrate this point, we have generated exact numerical solutions of the two-site problem using the full Hamiltonian H without any approximation, and compared them to solutions of the first-order effective adiabatic effective Hamiltonian $H_{ad} \equiv H_M + W$, corresponding to the effective action S_{ad} from Eq. (23). For both the exact and the adiabatic problems, we have used a sufficiently fine discretization of the u_- coordinate and a sufficiently large cutoff at large u_- to ensure a numerical accuracy of better than 1% in the calculated energy splittings over the entire parameter range studied. In Fig. 2, we show the logarithm of the polaron tunneling splitting $2t_P$, that is, the excitation energy from the ground state to the first excited state of the full electron-phonon system, as a function of E_P/Ω for $t \equiv 1$ and four different EP couplings, $E_P = 2.5, 3, 4$, and 8 that are well inside the polaronic regime ($E_P > E_P^{(crit)}$).

In addition to the exact solution, we show two different adiabatic solutions in Fig. 2, one obtained with the full adiabatic lattice potential $W \equiv W_0 + W_1$, the other using only the zeroth-order potential, $W \cong W_0$. These are being referred to in the following as the ''full'' and as the ''simple'' adiabatic solutions, respectively. As expected from the Holstein-Lang-Firsov strong-coupling expansion[14,24,25] and from semiclassical (WKB) arguments, the tunneling splitting at fixed E_P and t decreases exponentially with $1/\Omega$, as indicated by a roughly linear dependence of $\ln(2t_P)$ on $1/\Omega$ in Fig. 2.

Remarkably, the full adiabatic result agrees with the exact solution to better than 14% over a parameter region $0.15t < \Omega < 0.5t$ wherein $2t_P$ varies by more than nine orders of magnitude, including the regime where $2t_P$ is orders of magnitude smaller than Ω. The simple ($W \cong W_0$) adiabatic solution reproduces the qualitative features of the $1/\Omega$ and E_P dependence of $2t_P$ quite well, but the quantitative agreement is noticeably worse than for the full adiabatic solution. The agreement between the full adiabatic and the exact results is all the more convincing in light of the fact that the tunneling splitting is ''exponentially sensitive'' to small errors or changes in the wave function inside the tunneling barrier. Thus, our comparison of the tunneling splittings constitutes a much more stringent test of the underlying approximations than a comparison of, say, low-lying-state expectation values or wave-function amplitudes. Other exact numerical results for the two-site problem, are generally in equally good agreement with the corresponding adiabatic solution, provided, that is, one exercises enough care to use the proper adiabatic wave functions $|\phi\rangle$, Eq. (24), in carrying out the comparison.

As expected from Eq. (41), the agreement between adiabatic and exact results deteriorates at high phonon frequencies when Ω becomes comparable to t. As a practical matter, even for $\Omega \cong 2t$, the agreement is still quite acceptable. For applications to lattice systems, it is of interest to explore in some detail how the adiabatic approximation actually breaks down as one enters into the ''anti-adiabatic'' regime

$$t \ll \Omega. \tag{43}$$

As a limiting case, we consider the ionic limit $t \to 0$, already discussed above. Here, the Holstein-Hubbard problem can be trivially solved exactly.[26] Obviously there cannot be any electron tunneling between the two sites and the exact polaron tunneling splitting $2t_P$ vanishes.

By contrast, in the simple adiabatic approximation $W \cong W_0$, W_{0-} approaches the double-parabolic potential (39) for $t \to 0$, which has a tunneling barrier of finite height and width. The simple adiabatic approximation would thus predict a nonvanishing finite tunneling splitting $2t_P > 0$ even for $t = 0$, a clearly unphysical result.

If instead one uses the full adiabatic approximation, with $W = W_0 + W_1$, the correct qualitative physical behavior of $2t_P$ is restored by the W_1 term shown in Fig. 1(b): According to Eq. (29) the W_1 peak height (at fixed E_P and Ω) diverges as t^{-2}, while at the same time its peak width vanishes, but only linearly in t in the limit $t \to 0$. It is then easy to show that the transmission amplitude through the W_1 barrier vanishes, that is, the barrier becomes impenetrable in the limit $t \to 0$ that forces $2t_P \to 0$ for $t \to 0$. Thus, as far as the tunneling splitting $2t_P$ is concerned, the full adiabatic approximation reproduces qualitatively the correct physical behavior even in the extreme anti-adiabatic regime.

The actual failure of the full adiabatic approximation in the $t \to 0$ limit is a more subtle problem. It consists of the unphysical constraint being imposed on the dynamics of the u_- coordinate by the impenetrability of the W_1 barrier. For $t \to 0$, the W_1 barrier forces the lattice wave function $\phi(u)$ in Eq. (24) to vanish identically either to the right ($u_- > 0$) or to the left ($u_- < 0$) of the barrier. Thus, the amplitude for propagation from an initial $u_- < 0$ to a final $u_- > 0$ (or reverse) vanishes in the full adiabatic approximation at $t = 0$. In the exact solution of the $t = 0$ problem, this constraint does not exist; the lattice is free to propagate with some finite amplitude from $u_- < 0$ to $u_- > 0$. In the exact $t = 0$ solution, the lattice dynamics is governed either by the left or the right parabolic well, $W^{(l1)}$ or $W^{(l2)}$, corresponding respectively to

the left-localized or to the right-localized electron states, $|\Psi^{(l1)}\rangle$ or $|\Psi^{(l2)}\rangle$, discussed in Sec. III. The problem with the adiabatic approximation is that the $t=0$ electron ground state $|\Psi(u)\rangle$ exhibits a level crossing and thus changes discontinuously at $u_-=0$, as discussed above. The adiabatic approximation, by construction, excludes transitions between, say, the electronic ground state and first excited state. But this is just what happens at $u_-=0$ in the $t\to 0$-limit: If the lattice coordinate crosses $u_-=0$ from the left, say, under exact time evolution, the electron remains in its localized state $|\Psi^{(l1)}\rangle$, which is the ground state only for $u_-<0$, but becomes the first excited state when $u_->0$. The adiabatic approximation on the other hand forces the electron to remain in the ground state that changes discontinuously at $u_-=0$, from $|\Psi^{(l1)}\rangle$ to $|\Psi^{(l2)}\rangle$.

In the two-site problem, the foregoing impenetrability constraint causes only a small error, of order $\exp(-E_{\rm P}/\Omega)$, in the low-lying lattice eigenstates and energies if the lattice oscillator zero-point amplitude is small compared to the double-well separation $\sqrt{2}u_{\rm P}$, that is, if $\Omega \ll E_{\rm P}$. However, the impenetrability constraint may introduce a qualitative failure of the adiabatic approximation if applied to large systems $N\to\infty$ and tunneling processes that transfer a polaron in a single step over large distances, as we will now discuss.

Let us consider for simplicity the case of the Holstein model for just a single electron in a large lattice with sufficiently strong $E_{\rm P}$ to form a polaron. Suppose the polaron is localized at some site ξ, say, and we want to study the tunneling barrier for transferring the polaron in a single tunneling step to a distant site $\zeta=\xi+r$, i.e., with $|r|\gg a$, where a is the lattice constant.

Let $u^{(\xi)}\equiv(u_1^{(\xi)},\ldots,u_N^{(\xi)})$ denote that lattice configuration which minimizes $W_0(u)$ and localizes the polaron around the ''centroid site'' $\xi\in\{1,\ldots,N\}$. That is, $|u_l^{(\xi)}|$ and the corresponding electron charge density $\langle n_l\rangle^{(\xi)}$ are maximal at $l=\xi$ and die out exponentially at large distances $|l-\xi|$ from the centroid. Likewise, let $u^{(\zeta)}$ denote the lattice configuration that localizes the polaron around site ζ. By lattice translational invariance

$$u_l^{(\zeta)}=u_{l-r}^{(\xi)}, \tag{44}$$

if $\zeta=\xi+r$. Notice that polaron formation breaks the translational symmetry of the lattice in the zeroth-order adiabatic approximation. As a consequence, W_0 exhibits N degenerate local minima, corresponding to the N different, but translationally equivalent $u^{(\xi)}$ configurations on an N-site lattice with periodic boundary conditions. This is the lattice analog to the breaking of reflection symmetry in the two-site problem.

Let $u^{(\zeta\xi)}(s)$ denote the linear path segment in the N-dimensional u space connecting $u^{(\xi)}$ to $u^{(\zeta)}$, i.e.,

$$u^{(\zeta\xi)}(s)=\left(\frac{1}{2}-s\right)u^{(\xi)}+\left(\frac{1}{2}+s\right)u^{(\zeta)} \tag{45}$$

with $s\in[-\frac{1}{2},+\frac{1}{2}]$. In the following discussion, we consider Eq. (45) as a representative of low-action tunneling trajectories connecting $u^{(\xi)}$ to $u^{(\zeta)}$. The s coordinate can thus be regarded as the lattice analog to the u_- tunneling coordinate (25) in the two-site problem. Note in particular that $W_0[u^{(\zeta\xi)}(s)]$ has local minima at $s=-\frac{1}{2}$ and $s=+\frac{1}{2}$ that must, by continuity, be separated by (at least) one intervening maximum, i.e., by a tunneling barrier. The simplest scenario, normally borne out in the numerical calculations discussed below, is that there is only one barrier maximum, by symmetry located at $s=0$. Thus, along $u^{(\zeta\xi)}(s)$, W_0 has qualitatively the same structure as $W_{0-}(u_-)$ described above for the two-site problem.

The first crucial point to note here is that the width of this tunneling barrier, that is, the Euclidean distance from $u^{(\xi)}$ to $u^{(\zeta)}$ in their N-dimensional u space,

$$d(\zeta,\xi)\equiv|u^{(\zeta)}-u^{(\xi)}|\leqslant|u^{(\zeta)}|+|u^{(\xi)}|=2|u^{(\xi)}|\equiv d_\infty \tag{46}$$

is finite and bounded by an upper limit d_∞ that is independent of the spatial distance $|\zeta-\xi|=|r|$. Note that d_∞ is independent of ξ or ζ due to lattice translational invariance. Thus two polaron configurations $u^{(\xi)}$ and $u^{(\zeta)}$ are never further apart from each other than d_∞ in u space, regardless of how far apart their centroid sites ξ and ζ are in real space.

The second important point is that the height of the zeroth-order (W_0) tunneling barrier along $u^{(\zeta\xi)}(s)$ is also bounded independently of lattice distances $|\zeta-\xi|$. To see this, note that the EP potential Cu_l acting on the electron is attractive, i.e., $Cu_l<0$, for any u configuration along the path $u^{(\zeta\xi)}(s)$ between $s=0$ and $s=1$. Hence, the contribution to $W_0(u)$ from $H_{\rm e}+H_{\text{e-ph}}(u)$ is bounded from above by the electron ground-state energy of the undistorted lattice. Also, by an argument analogous to Eq. (46), the elastic energy contribution $H_K(u)$ is bounded from above by $\frac{3}{2}H_K(u^{(\xi)})$. Both of these upper bounds are independent of $|\zeta-\xi|$.

The foregoing considerations suggest that a manifold of tunneling trajectories exists, sufficiently close to $u^{(\zeta\xi)}(s)$, which will all connect $u^{(\xi)}$ to $u^{(\zeta)}$ through a W_0 barrier whose height and width is bounded by upper limits independent of $|\zeta-\xi|$. Within the simple adiabatic approximation, $W=W_0$, one thus arrives at the unphysical result that the polaron can tunnel in a single (''instanton'') tunneling step from any site ξ to any site ζ in the lattice with a tunneling matrix element $t_P(\zeta-\xi)$ that does *not* go to zero for $|\zeta-\xi|\to\infty$, but rather

$$\lim_{|\zeta-\xi|\to\infty}|t_P(\zeta-\xi)|\equiv t_{P\infty}>0. \tag{47}$$

The foregoing argument can be made formally more rigorous by employing instanton methods similar to those described in the next section for short-distance tunneling processes. We will not engage in that exercise here. Suffice it to say that the simple adiabatic result (47) for the lattice is analogous to the above-described two-site result in the $t=0$ limit: the simple adiabatic approximation allows tunneling solely on the basis of the W_0 electronic ground-state *energy* barrier, regardless of whether there is actually any electronic *wave-function overlap* between the initial and final u configurations of the tunneling process.

To account for wave-function overlap effects in long-distance tunneling processes, the W_1 term (22) has to be included in the total potential $W=W_0+W_1$. Let us consider the evolution of the electronic ground-state wave function $|\Psi(u)\rangle$ along the linear tunneling trajectory $u^{(\zeta\xi)}(s)$ (45) between two centroid sites ζ and ξ with $|\zeta-\xi|\gg l_P(u)$. Here

$l_P(u)$ denotes the exponential localization length of $|\Psi(u)\rangle$ for local lattice distortions comparable to $u^{(\xi)}$. As a simplest scenario, let us assume that the wave function $|\Psi(u)\rangle$ remains localized for all u along $u^{(\zeta\xi)}(s)$. This situation will be realized at EP coupling strengths $E_{\rm P}$ which are sufficiently large compared to $E_{\rm P}^{(\rm crit)}$. The electronic ground-state problem can then be qualitatively described as follows:

The EP potential $Cu_l^{(\zeta\xi)}(s)$, acting on the electron at sites l, consists of two localized wells, $C(\frac{1}{2}-s)u_l^{(\xi)}$ and $C(\frac{1}{2}+s)u_l^{(\zeta)}$, the former centered around site $l=\xi$, the latter around $l=\zeta$. As s is varied from $-\frac{1}{2}$ to $+\frac{1}{2}$, the EP well at ξ becomes shallower and the EP well at ζ deepens. At $s=0$, the two wells become degenerate. Assuming large real-space tunneling distances $|\zeta-\xi|$, the electron wave-function overlap between these two wells is exponentially small. Hence, the electron ground-state wave function $|\Psi[u^{(\zeta\xi)}(s)]\rangle$ will remain localized at site ξ for most $s<0$ until s gets very close to $s=0$. Within a very small interval around $s=0$, $|\Psi[u^{(\zeta\xi)}(s)]\rangle$ will then switch over from being localized around ξ to being localized around ζ. In that narrow s region around $s=0$, the electron wave function consists of the superposition of two almost nonoverlapping localized parts, one centered around ξ, the other around ζ. Since $|\Psi(u)\rangle$ changes very rapidly as a function of u near $u^{(\zeta\xi)}(0)$, $W_1(u)$ will exhibit a sharp peak along $u^{(\zeta\xi)}(s)$ that increases the tunneling barrier at $s=0$ and hence suppresses the tunneling amplitude.

Formally, this problem can be treated by a tight-binding ansatz for the electron ground-state wave function: $|\Psi[u^{(\zeta\xi)}(s)]\rangle$ near $s=0$ is approximated by a superposition of $|\Psi(u^{(\xi)}/2)\rangle$ and $|\Psi(u^{(\zeta)}/2)\rangle$, i.e., by the single-well ground states of the two EP wells $\frac{1}{2}Cu_l^{(\xi)}$ and $\frac{1}{2}Cu_l^{(\zeta)}$, discussed above. As s is varied near $s=0$, the response of $|\Psi[u^{(\zeta\xi)}(s)]\rangle$ to the changing EP well depths is then governed by the effective *electronic* hybridization overlap

$$t_{\rm eff}(\zeta-\xi)=\langle\Psi(u^{(\xi)}/2)|H_{\rm e}|\Psi(u^{(\zeta)}/2)\rangle\sim t\exp\left(-2\frac{|\zeta-\xi|}{l_{P,\,1/2}}\right), \tag{48}$$

where $l_{P,1/2}\equiv l_P(u^{(\xi)}/2)$ is the localization length of $|\Psi(u^{(\xi)}/2)\rangle$. Within the tight-binding ansatz, the problem then becomes analogous to the two-site problem in the $t\to 0$ limit, with the tight-binding basis states $|\Psi(u^{(\xi)}/2)\rangle$ and $|\Psi(u^{(\zeta)}/2)\rangle$ replacing the two-site basis states $|\Psi^{(l1)}\rangle$ and $|\Psi^{(l2)}\rangle$, respectively. $W_1[u^{(\zeta\xi)}(s)]$ exhibits a sharply peaked barrier at $s=0$, analogous to the $t\to 0$ limit of the two-site problem. The W_1 barrier will be roughly of the form given by Eq. (29), with u_- replaced by $u_-(s)\equiv d(\zeta,\xi)s$ and with t replaced by $t_{\rm eff}(\zeta-\xi)$. Thus, along with $t_{\rm eff}(\zeta-\xi)$, the transmission amplitude through the W_1 barrier and the effective polaron tunneling matrix element $t_P(\zeta-\xi)$ will decrease exponentially with the tunneling distance $|\zeta-\xi|$, analogous to the $t\to 0$ limit in the two-site problem.

The long-distance polaron tunneling processes are in the anti-adiabatic regime, since the relevant effective electronic hybridization overlap matrix elements $t_{\rm eff}(\zeta-\xi)$ become small compared to the phonon energy Ω at large tunneling distances $|\zeta-\xi|$ on large lattice sizes N. The W_1 potential ensures, at least qualitatively, that the effective polaron tunneling matrix elements $t_P(\zeta-\xi)$ are properly suppressed to zero at large tunneling distances. Hence, the full adiabatic approximation $W=W_0+W_1$ restores the correct long-distance behavior, as far as the polaron tunneling matrix element is concerned.

However, just as in the anti-adiabatic limit of the two-site problem, the W_1 term also imposes an unphysical constraint on the lattice coordinates. In the present case, involving long-distance tunneling on a lattice, this constraint acts to couple the lattice displacement coordinates at arbitrarily large distances $|\zeta-\xi|$, thereby introducing unphysical infinite-range interactions between the lattice coordinates.

Thus, in long-distance tunneling processes, the preconditions for the adiabatic approximation break down. However, from the foregoing discussion it is clear that the effective action for the corresponding paths increases exponentially and that the corresponding tunneling matrix element dies out exponentially with the tunneling distance. The simplest way of dealing with such long-distance tunneling processes is therefore to altogether neglect the corresponding tunneling paths in the path integral. This is what we will do in the following analysis. As far as the polaron tunneling dynamics is concerned, the short-distance processes will be dominant. The relevant effective electronic matrix elements $t_{\rm eff}$ for short-distance processes are of order of the first-neighbor t that is normally larger than or at least comparable to the phonon energy scale in typical solid-state situations. We *can* therefore use the adiabatic approximation to accurately estimate the effective action for short-distance tunneling paths. And it is only in this limited sense that the adiabatic approximation *will* be used in the following.

V. INSTANTONS AND EFFECTIVE HAMILTONIAN

The problem of polaron formation in the 2D Holstein-tJ and Holstein-Hubbard models has already been studied extensively.[10–12,14] In the nearly $\frac{1}{2}$-filled band regime, the dopant-induced hole carriers in the AF spin background can form polarons with much less EP coupling strength than is required for a single electron in an empty band. Thus $E_{\rm P}^{(\rm crit)}$ for forming a single-hole polaron in the $\frac{1}{2}$-filled system is reduced by a factor of about 4–5, compared to a single-electron polaron formation in the empty-band system. This reduction in $E_{\rm P}^{(\rm crit)}$ has been explained in terms of the hole mass enhancement and self-localization effect in the AF spin background of the nearly $\frac{1}{2}$-filled Hubbard system.[11] The basic idea here is that the coupling to the AF spin background already provides some form of self-localization of the hole relative to a self-induced local distortion of the AF spin correlations.[11,27] This spin polaron effect is manifested in the strongly reduced hole quasiparticle bandwidth, from $8t$ in the noninteracting system to $\sim 2J$ in Hubbard or tJ systems near half filling. In the presence of EP coupling, this electronic bandwidth reduction permits the hole quasiparticle to become self-trapped by a much weaker EP potential well; hence the reduction in $E_{\rm P}^{(\rm crit)}$. The fact that the polaron formation threshold $E_{\rm P}^{(\rm crit)}>0$ remains nonzero even in the strongly correlated systems is dictated by the so-called small-polaron dichotomy,[28] as discussed further in Sec. IX.

For a multihole system containing

$$P \equiv N - \sum_j n_j \tag{49}$$

doping-induced holes on an N-site lattice, there are $\binom{N}{P}$ possible configurations for accommodating the P localized holes on the N available sites. The lattice potential $W(u)$ is therefore expected to have up to $\binom{N}{P}$ nearly degenerate local minima, denoted by u^{ξ} in the following, corresponding to the $\binom{N}{P}$ different centroid configurations $\xi \equiv (\xi_1, \ldots, \xi_P)$.[11] Here, $\xi_i \equiv (\xi_{i,x}\xi_{i,y})$ denotes lattice (centroid) site occupied by the ith hole. As noted above, each of these local-minimum configurations breaks the translational symmetry of the lattice at the level of the zeroth-order adiabatic approximation. The symmetry is restored in the first-order adiabatic approximation by polaron tunneling processes between the different u^{ξ}.

At EP coupling strengths $E_{\rm P}$ larger than, but sufficiently close to $E_{\rm P}^{(\rm crit)}$, it is possible that some of the $\binom{N}{P}$ centroid configurations ξ do not have corresponding stable local-minimum configurations u^{ξ} in $W(u)$. This may happen, for example, in a two-hole system ($P=2$), if one tries to accommodate the two polarons at first-neighbor sites, ξ_1 and ξ_2, in the presence of a first-neighbor Coulomb repulsion $V_{\rm C}$. At sufficiently strong $V_{\rm C}$, the corresponding local minimum $u^{\xi} \equiv u^{(\xi_1,\xi_2)}$ becomes locally unstable, which is signaled by the smallest eigenvalue of the restoring force matrix $\partial^2 W/\partial u^2|_{u^{\xi}}$ becoming negative. In the following, we will not consider such situations, but rather restrict ourselves to parameter regions where all the local minimum configurations u^{ξ} are stable.

To establish the basic structure of the effective polaron tunneling dynamics, we treat the path integral for the effective action $S_{\rm ad}$ (23) or its equivalent Hamiltonian $H_{\rm ad}$ by a lattice dynamical many-body tight-binding approach. The basic idea behind this approach is that an effective polaron tunneling Hamiltonian $H_{\rm P}$ can be defined that operates in a ''low-energy'' subspace of nearly orthogonal tight-binding basis states $|\phi_{\xi}\rangle$, labeled by the localized polaron centroid configurations ξ. Each such basis state represents a lattice wave function $\phi_{\xi}(u)$ that is assumed to be localized in u space around the corresponding local potential minimum configuration u^{ξ}. For example, $\phi_{\xi}(u)$ could be chosen as the vibrational (''phonon'') ground state obtained in a harmonic approximation by expanding $W(u)$ to quadratic order around u^{ξ}. By restricting the lattice Hilbert space to such a set of basis states ϕ_{ξ}, all vibrational excited states around the polaronic local minima are neglected. Thus, formally, our approach can be regarded as a tight-binding approximation, formulated for the quantum dynamics of the multiple-well lattice potential $W(u)$ in the N-dimensional lattice configuration (u-) space.

In the simplest tight-binding approach one would then simply estimate the matrix elements of $H_{\rm P}$ by projecting the adiabatic lattice Hamiltonian $H_{\rm ad}$ onto the corresponding tight-binding low-energy subspace spanned by all ϕ_{ξ}. In such a first-order projection approach, one neglects all effects arising from virtual excitations out of the low-energy subspace.

Since tunneling matrix elements are exponentially sensitive to small corrections in, for example, the tunneling barriers, such a tight-binding projection could cause severe quantitative errors in the estimation of the magnitude of tunneling matrix elements. Also, as a practical matter, the accurate evaluation of Hamiltonian matrix elements with basis functions defined on the N-dimensional u space can become quite difficult. Lastly, in the first-order projection approach, it is difficult to include the tunneling Berry phases into $H_{\rm P}$.

To avoid the foregoing difficulties, we have adopted an approach that is based on a direct mapping of imaginary-time tunneling paths, rather than a mapping of Hamiltonian matrix elements. Formally, this is accomplished by the path-integral instanton method.[29–31] In this method, an ''instantaneous'' polaron hopping process $\xi \rightarrow \zeta$ induced by $H_{\rm P}$ between two polaron centroid configurations $\xi \equiv (\xi_1, \ldots, \xi_P)$ and $\zeta \equiv (\zeta_1, \ldots, \zeta_P)$ is identified with the (restricted) path sum of instanton tunneling paths connecting u^{ξ} to u^{ζ} in u space. The effective action $S_{\rm P}$ of the instantaneous hopping paths, so obtained, can then be immediately translated into matrix elements of the effective tunneling Hamiltonian $H_{\rm P}$. Since only tunneling paths, but no basis states, enter into the mapping, the results do not depend on any particular choice of tight-binding basis states ϕ_{ξ}.

As a specific starting point, we consider the trace of the resolvent operator at complex energies E,

$$\mathrm{Tr}(E-H)^{-1} = -\int_0^{\infty} d\beta\, e^{\beta E} \mathrm{Tr}\, e^{-\beta H}, \tag{50}$$

written in the imaginary-time domain in path-integral form

$$\mathrm{Tr}\, e^{-\beta H} = \int_{u(\beta)=u(0)} \mathcal{D}u(\tau) e^{-S_{\rm ad}[u(\tau)]}. \tag{51}$$

The trace operation in Eq. (50) leads to periodic boundary conditions on the imaginary time interval $[0,\beta]$ in Eq. (51). These periodic boundary conditions in Eq. (51) impose not only the closed-path constraint $u(\tau)=u(0)$, but also the condition that the initial and final electron wave functions must be the same, including their phase factors. That is, for the electron wave functions $|\Psi[u(\tau)]\rangle$ entering into $S_{\rm ad}$ via Eq. (19), the constraint $\langle \Psi[u(\beta)]|\Psi[u(0)]\rangle = +1$ must be imposed for all paths $u(\tau)$ integrated over in Eq. (51). The latter requirement ensures that the Berry phase contribution to $S_{\rm ad}$ in Eq. (51) is uniquely defined for every closed path $u(\tau)$, independent of the choice of phase for each individual electronic wave function $|\Psi[u(\tau)]\rangle$ along such a path. Quantized eigenenergies can be found from Eq. (50) by searching for the poles of the trace of the resolvent operator on the real E axis.

The main contributions to the low-energy part of Eq. (50) arise from instanton path configurations, i.e., u paths that are almost always close to one of the centroid configurations, occasionally transfer from one to another centroid configuration by an almost instantaneous polaron hopping process, and finally return to the initial u configuration at imaginary time β, in order to satisfy the closed-path constraint. Important closed-path tunneling processes for polaron states with $P=1$ and 2 dopant-induced holes are shown in Fig. 3. Each black circle represents an occupied polaron centroid site in the initial configuration ξ of the hopping process. Arrows indicate the hopping processes transferring the initial configuration ξ into the final configuration ζ. Thus, in u space

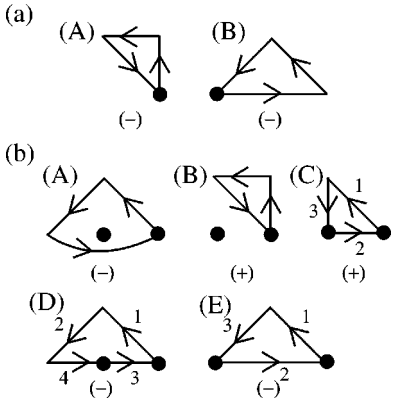

FIG. 3. (a) One-polaron and (b) two-polaron closed tunneling paths and their Berry phase factors. Black circles indicate the polaron locations for the initial u configuration of the path. The numbers on the two-polaron exchange paths in (b) indicate the order of the single-polaron tunneling steps.

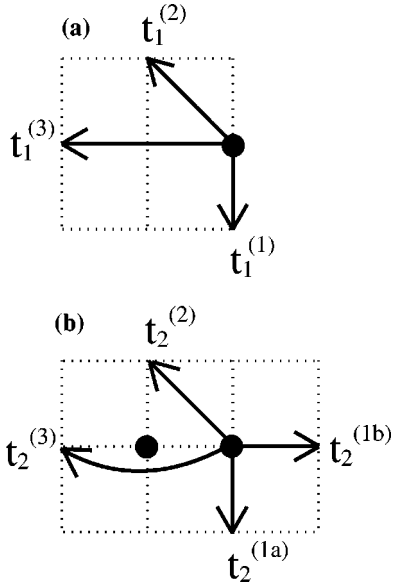

FIG. 4. Important single-polaron tunneling processes with matrix elements $t_P^{(\nu)}$ to the νth neighbor sites for (a) $P=1$ and (b) $P=2$ polaron states on the 2D square lattice.

each arrow corresponds to a set of instanton-type tunneling paths that connects the two respective minimum-W end-point configurations u^{ξ} and u^{ζ} and traverse the W barrier separating the two minima.[11] Note that, as discussed above, via such tunneling paths, a hole *polaron* can tunnel in a single step between second-, third-, etc. neighbor sites even if the original *electron* Hamiltonian (1) contains only a first neighbor t.[11]

First, we consider the case of $P=1$. For the time being, we take into account only the second- and third-neighbor processes denoted by amplitudes $t_1^{(2)}$ and $t_1^{(3)}$ in Fig. 4(a). Single-polaron *intersublattice* processes are strongly suppressed by the AF spin correlations.[32] Hence, the first-neighbor amplitude $t_1^{(1)}$ can be much smaller than or, at most, comparable to $t_1^{(2)}$ and $t_1^{(3)}$ (to within 20–30 %) in the case of $P=1$. Then, instanton path configurations are classified according to the numbers of intrasublattice processes: n_x counts the number of $t_1^{(3)}$ processes to the right, m_x the $t_1^{(3)}$ processes to the left, n_y the $t_1^{(3)}$ processes to the upper, m_y the $t_1^{(3)}$ processes to the lower, n_u the $t_1^{(2)}$ processes to the upper-right, m_u the $t_1^{(2)}$ processes to the lower-left, n_v the $t_1^{(2)}$ processes to the lower-right, and m_v the $t_1^{(2)}$ processes to the upper-left neighbors. Path integration over the corresponding instanton paths gives[29–31]

$$\begin{aligned}\mathrm{Tr}\, e^{-\beta H} &= e^{-\beta W(u^{(\min,1)})} \sum_{n_x,\ldots,m_v} \frac{1}{n_x!m_x!n_y!m_y!n_u!m_u!n_v!m_v!} \int \frac{dp_x}{2\pi} e^{ip_x(2n_x-2m_x+n_u-m_u+n_v-m_v)} \\ &\times \int \frac{dp_y}{2\pi} e^{ip_y(2n_y-2m_y+n_u-m_u-n_v+m_v)} (e^{-\delta R_1^{(2)}-i\theta_1^{(2)}} J_1^{(2)} K_1^{(2)} \beta)^{n_u+m_u+n_v+m_v} (e^{-\delta R_1^{(3)}-i\theta_1^{(3)}} J_1^{(3)} K_1^{(3)} \beta)^{n_x+m_x+n_y+m_y} \\ &= \int \frac{dp_x}{2\pi} \frac{dp_y}{2\pi} \exp[-\beta\{W(u^{(\min,1)}) + 2t_1^{(2)}[\cos(p_x+p_y)+\cos(p_x-p_y)] + 2t_1^{(3)}[\cos(2p_x)+\cos(2p_y)]\}]. \end{aligned} \tag{52}$$

The effective hopping matrix elements $t_P^{(\nu)}$ are obtained as

$$t_P^{(\nu)} = -J_P^{(\nu)} K_P^{(\nu)} e^{-\delta R_P^{(\nu)} - i\theta_P^{(\nu)}}. \tag{53}$$

$W(u^{(\min,1)})$ is the absolute minimum lattice potential energy obtained at a minimum-W configuration $u^{(\min,1)} \equiv u^{(\xi_1)}$ for $P=1$. Factorial factors such as $n_x!$, etc. are introduced to account for identical species of instantons. The p_x and p_y integrals are introduced to enforce the imaginary-time periodic boundary condition. The quantity $\delta R_P^{(\nu)}$ is the single-instanton contribution to the real part of the action for the path segment of the corresponding tunneling process $t_P^{(\nu)}$ and $\theta_P^{(\nu)}$ is the corresponding Berry phase contribution. The assignment of a unique Berry phase factor $e^{-i\theta_P^{(\nu)}}$ to each such open path segment requires more detailed symmetry considerations and will be postponed until Sec. VI. The quantity $K_P^{(\nu)}$ in Eq. (53) represents the $-\frac{1}{2}$ power of the fluctuation determinant for the instanton solution with the zero mode excluded divided by that for the static solution at $u^{(\min,1)}$, and $J_P^{(\nu)}$ is the Jacobian needed for a special treatment of the corresponding zero mode. They are defined as in Eqs. (10.13) and (10.14) of Ref. 29 for the periodic potential problem. Substituting the result of the path integral (52) into the formula (50), we obtain the dispersion relation shown in

the parentheses of $\exp[-\beta\{\ldots\}]$ in Eq. (52). Note that the effective hopping matrix element is defined such that it is positive if the corresponding path segment carries a nontrivial (-1) Berry phase factor: the sign convention of our polaron tunneling matrix elements $t_P^{(\nu)}$ is opposite to that used in the original electron Hamiltonian (1).

Next, we consider the case of $P=2$. Since self-localization reduces substantially the polaron kinetic energy scale, it is favorable for two polarons in an AF spin background to be bound in a pair: the binding energy can easily be of the order of the effective polaron nearest-neighbor attraction, i.e., comparable to the AF spin exchange coupling J (Ref. 33) in the Holstein-tJ model. As a first approximation, we therefore restrict the path integration to include only first-neighbor pair configurations $u^{(\xi_1,\xi_2)}$ and the instanton tunneling paths connecting them. Our numerical studies described below suggest that these first-neighbor configurations represent the *absolute* minimum of $W(u)$ for $P=2$. Other, more distant pair configurations with $|\xi_1-\xi_2|>1$ are either represented by local W minima $u^{(\xi_1\xi_2)}$ of higher energy or they do not form local minima in $W(u)$ at all. We are thus limiting ourselves, for now, to the tunneling processes $t_2^{(2)}$ and $t_2^{(3)}$ between the degenerate, absolute-minimum u configurations as shown in Fig. 4(b).

The technique used above for $P=1$ can be generalized in a straightforward manner to the present case $P=2$. Here, in addition to the lattice translational degeneracy of the minimum-W u configurations, the $P=2$ system exhibits twofold internal degeneracy, corresponding to the two possible orientations of the first-neighbor polaron pair, along either the x or along the y axis. Because of this twofold internal degree of freedom, the instanton exponential function in the path integral takes the form of the trace over a 2×2 matrix exponential, namely,

$$\mathrm{Tr}\, e^{-\beta H}=\int \frac{dp_x}{2\pi}\frac{dp_y}{2\pi}\mathrm{Tr}\exp\left[-\beta W(u^{(\min,2)})\begin{pmatrix}1&0\\0&1\end{pmatrix}-\beta\begin{pmatrix}2t_2^{(3)}\cos p_x & 4t_2^{(2)}\cos(p_x/2)\cos(p_y/2)\\ 4t_2^{(2)}\cos(p_x/2)\cos(p_y/2) & 2t_2^{(3)}\cos p_y\end{pmatrix}\right], \tag{54}$$

where $W(u^{(\min,2)})$ denotes the absolute minimum lattice potential energy for $P=2$, obtained at $u^{(\min,2)}\equiv u^{(\xi_1\xi_2)}$ with ξ_1 and ξ_2 denoting first-neighbor centroid sites. The tunneling matrix elements $t_P^{(\nu)}$ are expressed analogous to Eq. (53) in terms of the action contributions, fluctuation determinants, and Jacobians of the respective instanton path segments. The two low-lying eigenenergies of the polaron pair at total momentum $p\equiv(p_x,p_y)$ are obtained by diagonalizing the 2×2 matrix in $\exp[-\beta\{\ldots\}]$ of Eq. (54).

The generalization of the foregoing path-integral approach to $P>2$ hole polaron states is in principle straightforward, but becomes practically difficult to implement with increasing polaron number P. Analogous to Eq. (54), the approach leads to a momentum integral over the trace of a matrix exponential where the matrix dimension reflects the number of (nearly) degenerate, translationally inequivalent polaron centroid configurations $(\xi_1,\ldots,\xi_P)$ included in the tunneling analysis. In the following, we restrict ourselves to the cases $P=0$, 1, and 2 that will allow us to extract the effective one-polaron tunneling and two-polaron interaction matrix elements.

The low-lying tunneling eigenenergies identified by the foregoing instanton path-integral method (and their corresponding eigenstates) can be equivalently represented in terms of an effective polaron tunneling Hamiltonian H_{P} where each polaron is represented as a spin-$\frac{1}{2}$ fermion. H_{P} is thus defined to operate in an effective spin-$\frac{1}{2}$ fermion Hilbert space with the effective fermions occupying sites ξ_i on the 2D square lattice of the original EP Hamiltonian. A P-polaron centroid configuration $(\xi_1,\ldots,\xi_P)$ is mapped onto the corresponding state of P site-localized fermions with minimum possible total spin, i.e., with $S_{tot}=\frac{1}{2}$ (0) for odd (even) P. The latter mapping condition reflects the fact that the absolute electron ground states $|\Psi(u)\rangle$, numerically calculated on finite clusters, exhibit minimum total spin quantum number. Notice however that by representing the polaron as an effective spin-$\frac{1}{2}$ fermion, we are actually including low-energy spin excitations into the effective Hamiltonian description. In order to derive the effective polaron spin-spin interactions, our adiabatic path-integral treatment can be straightforwardly generalized to include restricted electron ground states in Hilbert space sectors of higher total spin quantum numbers $S_{tot}\geqslant1$. In this manner, the spin-$\frac{1}{2}$ fermion representation can be extended well beyond the scope of our original adiabatic approximation that retains only the (minimum-spin) absolute electron ground state $|\Psi(u)\rangle$. In the following analysis, we limit ourselves to the absolute ground state only. Hence we are only studying the total spin-singlet pair state in the $P=2$ case. Using our numerical Berry phase results, we will show in Sec. VI that each single polaron in such a singlet pair behaves indeed as a spin-$\frac{1}{2}$ fermion.

In generalizing the above first-neighbor approach, it is also straightforward to include intersublattice hopping processes: the dimension of the matrix increases, the **k**-independent term is no longer proportional to the unit matrix, and $t_P^{(1)}$ (more precisely, $t_1^{(1)}$, $t_2^{(1a)}$, and $t_2^{(1b)}$) are defined as above. Then, the effective Hamiltonian describing the polaron tunneling dynamics and effective polaron-polaron interactions can be written in the form

$$H_{\mathrm{P}}=\sum_{i\neq j,\sigma}\left(t_{ij}+\sum_k \Delta t_{ijk}n_{dk}\right)\times(1-n_{dj,-\sigma})d_{j\sigma}^{\dagger}d_{i\sigma}(1-n_{di,-\sigma})-\sum_{\langle i,j\rangle}V_{\mathrm{P}}n_{di}n_{dj}\,. \tag{55}$$

Thus, $d_{j\sigma}^{\dagger}$ creates a spin-$\frac{1}{2}$ fermion polaron with spin σ at site j, $n_{dj}=\Sigma_\sigma d_{j\sigma}^{\dagger}d_{j\sigma}=0,1$ and $P=\Sigma_j n_{dj}=1,2$. The hopping term is to include, appropriately, the amplitudes $(t_{ij}+\Sigma_k\Delta t_{ijk}n_{dk})\equiv t_1^{(1)}$, $t_1^{(2)}$, $t_1^{(3)}$, $t_2^{(1a)}$, $t_2^{(1b)}$, $t_2^{(2)}$, or $t_2^{(3)}$ (with appropriate sign according to the corresponding Berry phase factor) for $i\rightarrow j$ tunneling processes shown in Fig. 4. Note here that the sign convention for the polaron tunneling amplitudes $t_P^{(\nu)}$ in Eq. (55) is opposite to that used in the underlying electron Hamiltonians (8) and (9).

The first-neighbor attraction V_P in Eq. (55) is estimated as

$$V_P = 2W(u^{(\min,1)}) - W(u^{(\min,2)}) - W(u^{(\min,0)}), \quad (56)$$

where $W(u^{(\min,P)})$ is the (absolute) minimum potential energy $W(u)$ for the u configurations $u^{(\min,P)} \equiv u^{(\xi_1,\ldots,\xi_P)}$ that minimize $W(u)$ for P holes. For $P=2$, our numerical calculations suggest that the absolute minimum-W u configuration does indeed correspond to the first-neighbor pair. For purposes of estimating V_P numerically on small model clusters, we have minimized W_0 instead of $W=W_0+W_1$, thus neglecting the effect of W_1 on $u^{(\min,P)}$. In the physical parameter regime of interest, $\Omega \ll t, E_P$, these W_1 corrections to $u^{(\min,P)}$ are indeed small, of order $(\Omega/t)^2$. The full potential $W=W_0+W_1$ was used to calculate $W(u^{(\min,P)})$.

To obtain order of magnitude estimates for $t_P^{(\nu)}$, we have used both the dilute instanton-gas approach,[29–31] as explained above, and a constrained lattice dynamics approach[11] that is more straightforward and adopted in Sec. VIII. The two approaches have given similar results. In the latter approach, the lattice Schrödinger equation corresponding to $S_{\rm ad}$ is solved exactly for u constrained to the linear tunneling path $u^{(\zeta\xi)}(s)$ that is defined analogous to Eq. (45) and connects the two energetically degenerate, minimum-W polaron end-point u configurations u^{ξ} and u^{ζ} of the respective hop. The hopping matrix element $|t_P^{(\nu)}|$ is then estimated as one half of the ground-state—to—first-excited-state energy splitting.

VI. SYMMETRY OPERATIONS AND BERRY PHASES

Before going into numerical estimations of effective model parameters, we need to settle the quasiparticle statistics and the signs of effective polaron hopping processes by calculating Berry phase factors. To calculate $\exp(-i\theta[u(\tau)])$ for tunneling paths $u(\tau)$ shown in Fig. 3, we discretize τ with at least 5 τ points per linear path segment and obtain $|\Psi[u(\tau)]\rangle$ of the Holstein-tJ model by the Lanczos exact diagonalization method on an $N=4\times4$ lattice with periodic boundary conditions. The electron Hilbert space is restricted to the sector of minimum total spin ($S=0,1/2,0$ for $P=0,1,2$, respectively), which comprises the absolute ground state $|\Psi(u)\rangle$ for u configurations near the local W minima. The results for all paths in Fig. 3 are summarized by

$$\theta[u(\tau)] = \pi(m^{(2)} + m^{(3)} + m_2^{(1)}), \quad (57)$$

where $m^{(\nu)}$ is the number of νth neighbor hops with $\nu=2$, 3, and $m_2^{(1)}$ for $P=2$ denotes the number of first-neighbor hops indicated by the dashed bonds shown in Fig. 5(a) by the first polaron in close proximity to the second, static polaron, indicated as a black circle. The effect of the $m_2^{(1)}$ term can be illustrated, for example, by comparing the Berry phase factors $e^{-i\theta}$ of the triangular paths (a) (A) and (b) (B) shown in Fig. 3. In both paths, a single polaron is taken around the triangle in three steps, consisting of two first-neighbor and one second-neighbor transfer. For the one-polaron case, (a) (A), the phase factor is (-1), for the two-polaron case (b) (B) it is $(+1)$. Thus, the close proximity of the second, static polaron in (b) (B) has altered the Berry phase of the first polaron tunneling around a closed path.

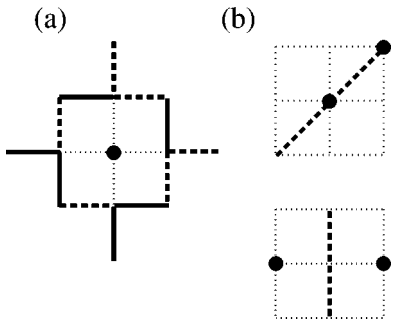

FIG. 5. (a) Berry phase contributions from single-polaron first-neighbor processes in the vicinity of a second, static polaron (black circle). Full and dashed bonds indicate $(+1)$ and (-1) Berry phase contributions, respectively. (b) Internal parity of second- and third-neighbor polaron pairs is odd under reflection along the dashed line.

The origin of the $m_2^{(1)}$ term can be traced back to the internal symmetry of $|\Psi(u)\rangle$: For the local minimum-W u configurations of second- and third-neighbor polaron pairs, $|\Psi(u)\rangle$ is odd under reflection along the dashed lines shown in Fig. 5(b), i.e., along the pair axis for the second-neighbor pair and perpendicular to the pair axis for the third-neighbor pair. Suppose, for example, that the first polaron hops from (1,0) to (1,1) in a first step and from (1,1) to (0,1) in a second step with the second polaron staying fixed at (0,0). These are the first two steps of path (b) (B) in Fig. 3. Note that the two steps generate the same final centroid configuration as a reflection along the dashed (0,0)-(1,1) line, shown in Fig. 5(b). Because of this odd ''internal'' parity of $|\Psi(u)\rangle$ for the intermediate (second-neighbor polaron pair) configuration, one of the two first-neighbor hops must contribute an additional factor (-1). Assigning this (-1) phase factor to one of the two first-neighbor steps in path (b) (B) of Fig. 3 is to some extent arbitrary. The pattern of dashed-line and full-line bonds surrounding the static polaron in Fig. 5(a) represents one possible assignment that is consistent with all the closed-path Berry phase results in Fig. 3(b). As a consequence of its odd internal parity, the second-neighbor polaron pair configuration is actually allowed to contribute with finite amplitude to polaron pair wave functions of $d_{x^2-y^2}$ symmetry, in spite of the fact that the second-neighbor pair axis points along the nodal axis of the $d_{x^2-y^2}$ pair wave function.

The $m^{(2)}$ and $m^{(3)}$ terms can be regarded as due to strong antiferromagnetic correlation. Suppose a polaron is initially located at (0,0) and hops to (2,0), (1,1), and then back to (0,0) along the path (a) (B) in Fig. 3. The electron initially located at (2,0) hops to (0,0) and then to (1,1), while the electron initially located at (1,1) hops to (2,0). Thus, if one approximates the AF spin background by a Néel state, two electrons of like spin are exchanged. This produces a fermionic (-1) factor. More generally, when a closed path consists of an odd number of second- or third-neighbor hopping processes, an even number of electrons within a sublattice are cyclically permuted, producing the (-1) factor within the Néel approximation to the AF spin background. In order for this to occur, the AF spin correlation has to be strong, but it need not be long ranged. For $P=1$, the Berry phase rule can be completely explained in this way.

For both $P=1$ and 2, $\theta[u(\tau)]$ is given by a sum of independent single-polaron hopping contributions and

$\exp(-i\theta[u(\tau)])$ does *not* depend on whether or not the two polarons are being adiabatically exchanged in a given path [Fig. 3(b)].[34] Thus, for example, paths (b) (B) and (b) (C) in Fig. 3 contain the same first- and second-neighbor hops and they have the same Berry phase. The two paths differ only in that (b) (C) exchanges the two polarons, whereas (b) (B) does not. Since the pair is a total spin singlet, this implies that each single polaron in the pair behaves effectively as a spin-1/2 fermion or as a spin-0 boson. Only the spin-1/2 fermion representation is consistent with the half-odd-integer total spin in odd-P systems and, as discussed in the previous sections, it is the one we have adopted. Equation (57) rules out the possibility of representing dopant-induced hole polarons as spin-0 fermions or as spin-1/2 bosons.

To settle the signs of effective polaron hopping processes, we need to define Berry phase factors for the corresponding single-hop open-path segments. Let the initial u configuration of such a single-hop path segment be denoted by $u^{(\xi)}$ and the final u configuration by $u^{(\zeta)}$. The assignment of a Berry phase to such a path segment can be made unique by fixing the phase of the corresponding wave function $|\Psi(u^{(\zeta)})\rangle$ relative to that of $|\Psi(u^{(\xi)})\rangle$ in some unique manner. Given $u^{(\xi)}$ and $|\Psi(u^{(\xi)})\rangle$, let $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ denote such a final-state reference wave function. Also, let $|\Psi^{(\mathrm{ad})}(u^{(\zeta)})\rangle$ denote that ground-state wave function $|\Psi(u^{(\zeta)})\rangle$ that one obtains by adiabatically evolving $|\Psi(u)\rangle$ along the tunneling path segment, without discontinuity in phase, beginning with $|\Psi(u^{(\xi)})\rangle$. The Berry phase of the path segment is then defined as the phase difference between $|\Psi^{(\mathrm{ad})}(u^{(\zeta)})\rangle$ and $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$, that is, as the phase of the wave-function overlap $\langle\Psi^{(\mathrm{ref})}(u^{(\zeta)})|\Psi^{(\mathrm{ad})}(u^{(\zeta)})\rangle$. If, for example, $u^{(\xi)}$ and $u^{(\zeta)}$ are related by a lattice symmetry operation, we can choose $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ as the ground-state wave function $|\Psi(u^{(\zeta)})\rangle$ generated by applying to $|\Psi(u^{(\xi)})\rangle$ the symmetry operation which transforms $u^{(\xi)}$ into $u^{(\zeta)}$. If $u^{(\xi)}$ and $u^{(\zeta)}$ are related by several different symmetry operations giving different $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$, we need to specify the reference $|\Psi(u^{(\zeta)})\rangle$, i.e., which symmetry operation is chosen to generate the reference $|\Psi(u^{(\zeta)})\rangle$ from $|\Psi(u^{(\xi)})\rangle$. There does not always exist such a symmetry operation to relate $|\Psi(u^{(\zeta)})\rangle$ and $|\Psi(u^{(\xi)})\rangle$, e.g., the second- or third-neighbor pair and the first-neighbor one. Then, we can arbitrarily choose the phase of $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$. The Berry phase factor for the corresponding path is also arbitrary. Figure 5(a) is an example. If we chose the different phase (i.e., the negative) of $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ for the second (third)-neighbor pair, the signs of all the $t_2^{(1a)}$ ($t_2^{(1b)}$) processes would be reversed.

For $P=1$, we first fix, arbitrarily, the phase of $|\Psi(u^{(\xi)})\rangle$ for the centroid configuration $\xi=[(0,0)]$. All the $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ are then uniquely defined by either translation or rotation operations. The Berry phase factors for $P=1$ are summarized in Fig. 6(a). Here, the translation $(x,y)\rightarrow(x+a,y+b)$ is denoted by $T(a,b)$, and the rotation $(x,y)\rightarrow(x\cos\phi-y\sin\phi, x\sin\phi+y\cos\phi)$ is denoted by $R(\phi)$. The left-hand side shows $|\Psi^{(\mathrm{ad})}(u^{(\zeta)})\rangle$. The right-hand side shows the possible choices of $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$, generated from the same $|\Psi(u^{(\xi)})\rangle$ by the appropriate lattice symmetry operations. Note that, for second- and third-neighbor hops, both translation and rotation operations generate the same $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$. The first three lines of Fig. 6(a) give $\exp[-i\theta_1^{(1)}]=+1, \exp[-i\theta_1^{(2)}]=-1, \exp[-i\theta_1^{(3)}]=-1$, and thus

$$t_1^{(1)}<0, \quad t_1^{(2)}>0, \quad t_1^{(3)}>0. \tag{58}$$

Obviously, these results are consistent with relation (57).

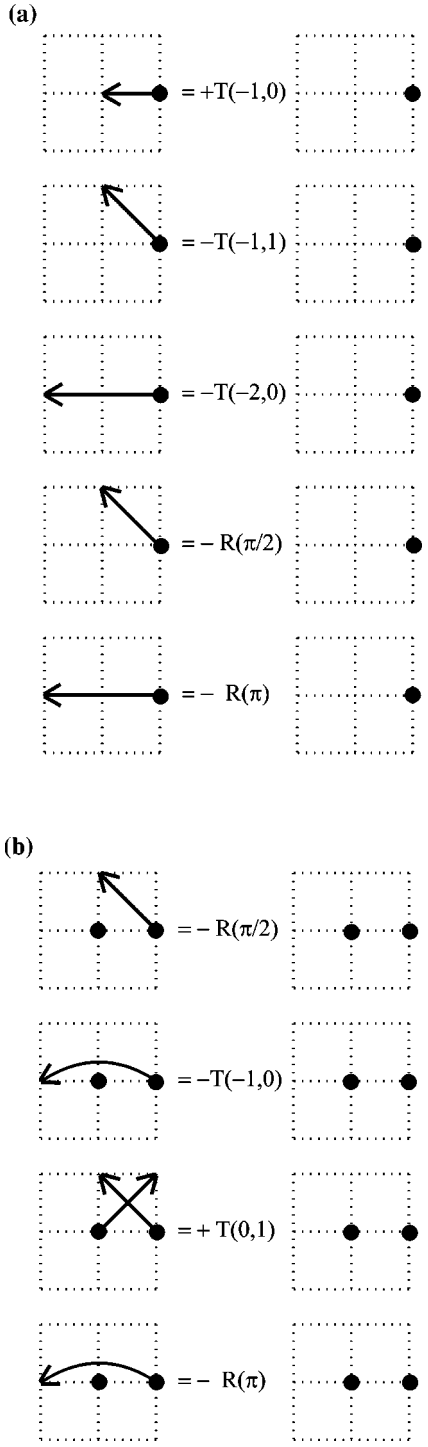

FIG. 6. (a) One-polaron and (b) two-polaron open tunneling paths and their Berry phase factors. Black circles indicate the polaron locations for the initial u configuration of the path. $T(a,b)$ denotes the translation by vector (a,b), $R(\phi)$ the rotation by angle ϕ.

For $P=2$, we first fix the phase of $|\Psi(u^{(\xi)})\rangle$ for the centroid configuration $\xi\equiv(\xi_1,\xi_2)=[(0,0),(1,0)]$. Rotating this $|\Psi(u)\rangle$ by angle $\pi/2$, we define the $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ for $\zeta\equiv(\zeta_1,\zeta_2)=[(0,0),(0,1)]$. The $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ for the other first-neighbor pair configurations ζ are defined by applying translation operations to either $|\Psi(u^{(\xi)})\rangle$ or to its rotated version $R(\pi/2)|\Psi(u^{(\xi)})\rangle$. The resulting Berry phase factors for $P=2$ are summarized in Fig. 6(b). The first two lines imply that $\exp[-i\theta_2^{(2)}]=-1$,$\exp[-i\theta_2^{(3)}]=-1$, and thus

$$t_2^{(2)}>0, \quad t_2^{(3)}>0. \tag{59}$$

The last two lines in Fig. 6(b) imply, by comparison to the first two lines, that available rotations would generate the same $|\Psi^{(\mathrm{ref})}(u^{(\zeta)})\rangle$ as the translations. The first-neighbor hops $t_2^{(1a)}$ and $t_2^{(1b)}$ are positive or negative for the processes indicated by the dashed and full bonds, respectively, of Fig. 5(a), as discussed above.

VII. BERRY PHASES AND QUANTUM NUMBERS

Using the effective Hamiltonian (55) with parameters $t_1^{(1)}, t_1^{(2)}, t_1^{(3)}, t_2^{(1a)}, t_2^{(1b)}, t_2^{(2)}, t_2^{(3)}$ (with signs determined above), and V_{P}, we can now calculate the low-energy eigenstates for the $P=1$ and $P=2$ polaron systems.

In the case $P=1$, the dispersion relation from H_{P} is given by

$$\epsilon_1(\mathbf{p})=2t_1^{(1)}[\cos p_x+\cos p_y]+2t_1^{(2)}[\cos(p_x+p_y)+\cos(p_x-p_y)]+2t_1^{(3)}[\cos(2p_x)+\cos(2p_y)]. \tag{60}$$

As mentioned in Sec. V, on finite lattices, $t_1^{(1)}$ is smaller than the second- and third-neighbor t's for $P=1$. As the cluster size increases, the overlap between the two minimum-W wave functions connected by the $t_1^{(1)}$ process, $|\Psi(u^{(\xi)})\rangle$ and $|\Psi(u^{(\zeta)})\rangle$ for a first-neighbor bond (ξ,ζ), becomes small. Then, the potential energy $W(u)$ would develop a higher barrier for the first-neighbor hop, due to $W_1(u)$, so that $t_1^{(1)}$ would continue to become smaller. Allowing for arbitrary values of $t_1^{(2)}$ and $t_1^{(3)}$ but $t_1^{(1)}=0$, the one-polaron band minimum is located at momentum $\mathbf{p}=(\pi/2,\pi/2)$ for $|t_1^{(2)}|<2t_1^{(3)}$, at $(\pi,0)$ for $|t_1^{(2)}|>2t_1^{(3)}$ and $t_1^{(2)}>0$, and at $(0,0)$ and (π,π) for $|t_1^{(2)}|>2t_1^{(3)}$ and $t_1^{(2)}<0$, as shown in Fig. 7(a). For the physically relevant signs implied by the Berry phase factors, $t_1^{(2)},t_1^{(3)}>0$, the momentum of the one-polaron band minimum is thus at $(\pi/2,\pi/2)$ or $(\pi,0)$ which lies on the Fermi surface of the noninteracting nearest-neighbor tight-binding band model at half filling. The one-polaron bandwidth is given by

$$B_1=\begin{cases}4t_1^{(2)}+8t_1^{(3)} & \text{for} \quad 0<t_1^{(2)}<2t_1^{(3)} \\ 8t_1^{(2)} & \text{for} \quad 0<2t_1^{(3)}<t_1^{(2)}.\end{cases} \tag{61}$$

For the cluster geometries studied here ($N=\sqrt{8}\times\sqrt{8}$, $N=\sqrt{10}\times\sqrt{10}$, and $N=4\times4$) and with only nearest-neighbor terms (t,J) included in the original Hamiltonian (1), certain ''accidental'' symmetries exist that cause $t_P^{(2)}=t_P^{(3)}$. As a consequence, the band minimum is at $(\pi/2,\pi/2)$, and the eigenvalues of the inverse effective-mass tensor at this point are

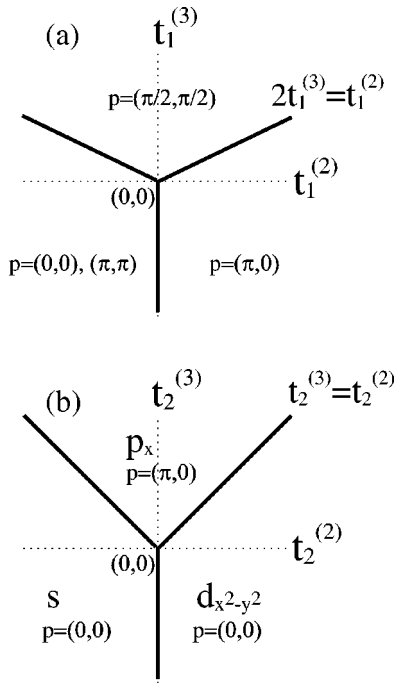

FIG. 7. Total momentum (p_x,p_y) of the (a) $P=1$ and (b) $P=2$ polaron ground state of H_{P} in a $t_P^{(3)}$-vs-$t_P^{(2)}$ phase diagram. Also shown in (b) are the internal symmetries of the respective two-polaron ground states.

$$(m_1^{-1})_r=8t_1^{(3)}+4t_1^{(2)}=12t_1^{(2,3)}, \tag{62}$$

$$(m_1^{-1})_\phi=8t_1^{(3)}-4t_1^{(2)}=4t_1^{(2,3)}, \tag{63}$$

where the subscript r is for the $(1,1)$ direction and ϕ is for the $(1,-1)$ direction. If we include finite and negative $t_1^{(1)}$ (representing, e.g., the $N\to\infty$ limit at finite, fixed hole density, rather than at fixed hole number $P=1$), then the band minimum is shifted by $t_1^{(1)}$ from $(\pi/2,\pi/2)$ to some point (p,p) with $p<\pi/2$ that would fall on the Fermi surface of the noninteracting band model at corresponding filling.

For $P=2$, we first consider the tightly bound pair limit, $V_{\mathrm{P}}\gg|t_2^{(\nu)}|$, where we can approximate the polaron-pair ground state by including only nearest-neighbor pair configurations, thus retaining only the $t_2^{(2)}$ and $t_2^{(3)}$ matrix elements of H_{P}. The pair dispersion relations are then given by

$$\epsilon_2^{\pm}(\mathbf{p})=-V_{\mathrm{P}}+t_2^{(3)}(\cos p_x+\cos p_y)\pm\left[t_2^{(3)2}(\cos p_x-\cos p_y)^2+\left(4t_2^{(2)}\cos\frac{p_x}{2}\cos\frac{p_y}{2}\right)^2\right]^{1/2}. \tag{64}$$

Allowing for arbitrary values of $t_2^{(2)}$ and $t_2^{(3)}$, the pair wave function in the nearest-neighbor pair approximation for $|t_2^{(2)}|>t_2^{(3)}$ has $d_{x^2-y^2}$-wave symmetry if $t_2^{(2)}>0$, and s-wave symmetry if $t_2^{(2)}<0$, and, in either case, total momentum $\mathbf{p}=(0,0)$, as shown in Fig. 7(b), at the band minimum. For $|t_2^{(2)}|<t_2^{(3)}$, the pair ground states are multiply degenerate:

the horizontal pair (with pair axis parallel to the x axis) with total momentum $\mathbf{p}=(\pi,p_y)$ for arbitrary $|p_y|\leqslant\pi$ and the vertical pair (with pair axis parallel to the y axis) with total momentum $\mathbf{p}=(p_x,\pi)$ for arbitrary $|p_x|\leqslant\pi$ all have the same energy. The two-polaron bandwidth is given by

$$B_2\equiv\max_{\mathbf{p}}\epsilon_2^-(\mathbf{p})-\min_{\mathbf{p}}\epsilon_2^-(\mathbf{p})=4||t_2^{(2)}|-t_2^{(3)}|. \tag{65}$$

If we take account of first-, second-, and third-neighbor pair configurations, including also the first-neighbor hopping terms, $t_2^{(1a)}$ and $t_2^{(1b)}$, in second-order perturbation theory, the initially degenerate energy along $\mathbf{p}=(\pi,p_y)$,

$$\epsilon_2^-(\pi,p_y)=-V_{\mathrm{P}}-2t_2^{(3)}, \tag{66}$$

is lowered by

$$\delta\epsilon_2^-(\pi,p_y)=-\frac{4\left(2t_2^{(1a)2}\cos^2\frac{p_y}{2}+t_2^{(1b)2}\right)}{V_{\mathrm{P}}+2t_2^{(3)}}. \tag{67}$$

It is reasonable for the $t_2^{(1a)}$ term to favor $p_y=0$ because the process of $(\xi_1,\xi_2)=[(0,0),(1,0)]\rightarrow[(0,0),(1,1)]$ and the process of $[(0,0),(1,1)]\rightarrow[(0,1),(1,1)]$, for example, are in phase, as shown in Fig. 5(a). The ground states are still doubly degenerate: the horizontal pair with $\mathbf{p}=(\pi,0)$ and the vertical pair with $\mathbf{p}=(0,\pi)$, which would correspond to p_x and p_y wave though they are total-spin singlets.[35] For $t_2^{(2)},t_2^{(3)}>0$ implied by the Berry phase factors, we thus get either $d_{x^2-y^2}$- or $p_{x(y)}$-pairing symmetry with total momentum $\mathbf{p}=(0,0)$ or $\mathbf{p}=(\pi,0)\,[(0,\pi)]$, respectively.

The accidental symmetries for our finite-cluster geometries [in the absence of longer-range terms in the original Hamiltonian (1) studied here] lead to $t_2^{(2)}=t_2^{(3)}$, which is exactly on the d-p phase boundary where B_2 vanishes due to a frustration effect.[32] So, the energy $\epsilon_2^-(\mathbf{p})$ is independent of $\mathbf{p}$. If we take account of first-, second-, and third-neighbor pair configurations again, in second-order perturbation theory, the initially degenerate energy on the d-p phase boundary,

$$\epsilon_2^-(\mathbf{p})|_{t_2^{(2)}=t_2^{(3)}}=-V_{\mathrm{P}}-2t_2^{(2,3)}, \tag{68}$$

is lowered by

$$\delta\epsilon_2^-(\mathbf{p})|_{t_2^{(2)}=t_2^{(3)}}=-\frac{f(\mathbf{p})}{V_{\mathrm{P}}+2t_2^{(2,3)}} \tag{69}$$

with

$$f(\mathbf{p})=4t_2^{(1a)2}(2+\cos p_x+\cos p_y)+4t_2^{(1b)2}\frac{1-\cos p_x\cos p_y}{2+\cos p_x+\cos p_y}. \tag{70}$$

Then the ground state has $d_{x^2-y^2}$ symmetry with $\mathbf{p}=(0,0)$ for $\sqrt{2}|t_2^{(1a)}|>|t_2^{(1b)}|$ and $p_{x(y)}$ wave with $\mathbf{p}=(\pi,0)\,[(0,\pi)]$ otherwise. For the $N=4\times4$ Holstein-tJ cluster with periodic boundary conditions, an accidental symmetry leads to $|t_2^{(1a)}|=|t_2^{(1b)}|$ and thus $d_{x^2-y^2}$-pairing symmetry. It is reasonable for the $t_2^{(1a)}$ term to favor the $d_{x^2-y^2}$-wave state because the process of $(\xi_1,\xi_2)=[(0,0),(1,0)]\rightarrow[(0,0),(1,1)]$ and the process of $[(0,0),(1,1)]\rightarrow[(0,0),(0,1)]$, for example, have opposite signs. Also, it is reasonable for the $t_2^{(1b)}$ term to favor the p_x-wave state because the process of $(\xi_1,\xi_2)=[(0,0),(1,0)]\rightarrow[(0,0),(2,0)]$ and the process of $[(0,0),(2,0)]\rightarrow[(1,0),(2,0)]$, for example, have opposite signs, as shown in Fig. 5(a). Once again we note that the second-neighbor polaron pair configuration contributes to the polaron-pair wave function of $d_{x^2-y^2}$ symmetry.

VIII. EFFECTIVE HOPPING AND ATTRACTION

We have seen how total momenta and internal symmetries of few-hole polaron states are determined by the signs and relative magnitudes of the effective polaron tunneling matrix elements. In this section, we show numerical estimates of them with effective polaron nearest-neighbor attraction and effective pair binding energy to see the energy scales of polaron dynamics. The relative energy scale of kinetic energy to interaction strength is controlled by the phonon frequency in the original Hamiltonian (1). It is noted again that we use a constrained lattice-dynamics approach and exactly solve the lattice Schrödinger equation corresponding to the effective action (23) for the lattice-displacement configurations constrained to the linear tunneling path of the respective hop, as described at the end of Sec. V. The effective lattice potentials are based on Lanczos calculations on finite clusters with periodic boundary conditions. The numerical results should be regarded as very rough order-of-magnitude estimates only. The nearest-neighbor attraction V_{P} is calculated according to formula (56). The pair binding energy Δ is estimated in the nearest-neighbor pair approximation, according to

$$\Delta=2\epsilon_1(\mathbf{p}_1^{(\min)})-\epsilon_2^-(\mathbf{p}_2^{(\min)}), \tag{71}$$

where $\epsilon_1(\mathbf{p})$ and $\epsilon_2^-(\mathbf{p})$ are defined in Eq. (60) (with $t_1^{(1)}=0$) and Eq. (64), respectively, measured relative to the $P=0$ ground-state energy, and $\mathbf{p}_P^{(\min)}$ are the respective (thus different) momenta at the band minima discussed above. Note that the sign of Δ is so defined that $\Delta>0$ signifies a net attraction, $\Delta<0$ repulsion.

Figure 8(a) shows the logarithm of the dominant second- and third-neighbor hopping amplitudes $t_P^{(2)}$ and $t_P^{(3)}$ for $P=1,2$ in the Holstein-tJ model on an $N=\sqrt{8}\times\sqrt{8}$ cluster. As expected in a polaronic system,[14] all $t_P^{(\nu)}$ are suppressed, roughly exponentially, with increasing E_{P}/Ω and strongly reduced compared to the bare electronic t. However, for E_{P} near $E_{\mathrm{P}}^{(\mathrm{crit})}$, the $t_P^{(\nu)}$ can become comparable to the phonon energy scale Ω. For $P=2$, the proximity of the second, static polaron strongly enhances the amplitudes $t_2^{(2)}$ and $t_2^{(3)}$ relative to $t_1^{(2)}$ and $t_1^{(3)}$. It is worth noting that this effect occurs only in the presence of strong electron correlations where bipolaron formation is prevented by the strong on-site Coulomb repulsion. By contrast, this effect never occurs in ordinary polaronic systems with the electron-phonon interaction E_{P} larger than the local Coulomb repulsion. In the latter case small bipolarons will form[36] that are much heavier than polarons. To generate the above-described delocalization (and

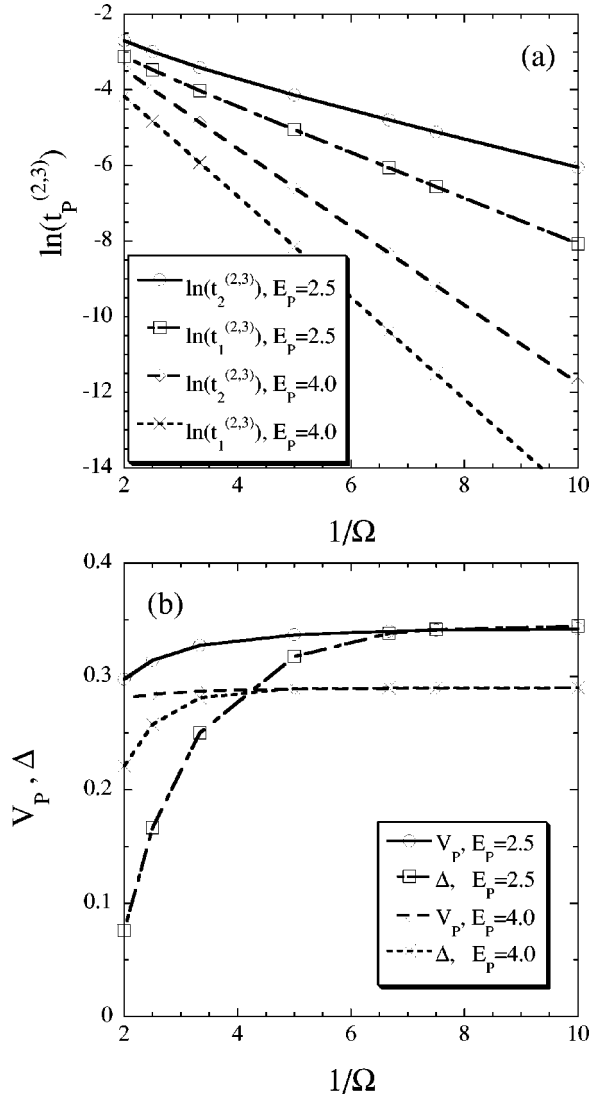

FIG. 8. (a) Logarithm of the effective polaron hopping amplitudes $t_P^{(\nu)}$ for $P=1$ and 2 holes and ν as second- and third-neighbor processes vs inverse phonon energy $1/\Omega$ (with $t_P^{(2)}=t_P^{(3)}$ due to accidental cluster symmetries). (b) Effective polaron nearest-neighbor attraction V_P and two-polaron binding energy Δ vs $1/\Omega$. All results are for $t\equiv 1, J=0.5t$, with $E_P\equiv C^2/K=2.5t$ and $4.0t$, on an $N=8$ lattice with periodic boundary conditions.

hence mobility) enhancement effect, it is essential to keep the two polarons spatially separated by strong enough on-site Coulomb effects.

The accidental symmetries leading to $t_P^{(2)}=t_P^{(3)}$ will be lifted on larger lattices and, more importantly, by inclusion of longer-range couplings, such as second-neighbor hopping t', Eq. (10), and extended Coulomb repulsion V_C, Eq. (11), in the original EP Hamiltonian (1), as will be shown below. Due to the exponential dependence of the delocalization matrix elements $t_P^{(\nu)}$ on the lattice potential parameters, such additional couplings can substantially affect the magnitudes of the $t_P^{(\nu)}$ parameters, without necessarily altering the Berry phase factors or the predominance of the second- and third-neighbor hopping terms ($t_P^{(2,3)}>t_P^{(1)}$) and their two-polaron enhancement ($t_2^{(\nu)}>t_1^{(\nu)}$). The Berry phase factors should be a robust feature of our model, since they reflect the topological properties of the relevant tunneling paths relative to certain singular manifolds of the lattice action in u space.

Figure 8(b) shows the nearest-neighbor attraction V_P and the pair binding energy Δ, where the latter quantity is given by

$$\Delta = V_P + 2t_2^{(2,3)} - 8t_1^{(2,3)} \tag{72}$$

for $t_P^{(2)}=t_P^{(3)}$, using Eq. (71). Since two self-localized nearest-neighbor holes mutually inhibit their delocalization, the t term in the original Hamiltonian (1) gives a repulsive contribution to V_P: in the parameter range shown in the figure, $V_P<0.342J$ $(0.316J)$ is substantially reduced compared to $V_P(t=0)=1.00J$ $(0.926J)$ on $N=\sqrt{8}\times\sqrt{8}$ $(\sqrt{10}\times\sqrt{10})$ sites in the $t\to 0$ limit. Compared to the tJ model, V_P can be larger or smaller: self-localization reduces the effective polaron hopping processes, giving an attractive contribution, and it is more effective in the one-hole state than in the two-hole state, giving a repulsive contribution. The binding energy Δ is enhanced by the two-polaron hopping amplitudes $t_2^{(2,3)}$, but it is smaller, in most of the parameter range shown in the figure, than V_P due to the restricted hopping processes for the polaron pair and due to the non-negligible $t_1^{(2,3)}$ term for large Ω. In a more realistic theory, the possible competition between polaron pairing and phase separation[33] would need to be considered for finite density of holes.

In order to see a finite-size effect, we have calculated the effective model parameters on $N=\sqrt{10}\times\sqrt{10}$ sites (not shown) to compare with those on $N=\sqrt{8}\times\sqrt{8}$ sites above. We find no qualitative difference between them. In the parameter range shown in the figure, the values of $t_P^{(\nu)}$ are different by a factor of 2 at most, but these values are rough order-of-magnitude estimates in any case. The values of V_P for $N=\sqrt{10}\times\sqrt{10}$ are smaller by a factor of 0.8–0.9.

For the Holstein-Hubbard model, we find results (Fig. 9) quite similar to those shown above. However, the values of V_P are only 30% of those in the Holstein-tJ model, which are reminiscent of the fact that the hole binding energy is larger for the tJ model than for the Hubbard model. Furthermore, the values of $t_2^{(2,3)}$ are smaller than those of the Holstein-tJ model for $\Omega<0.2t$, and the values of $t_1^{(2,3)}$ are larger by a factor of 1.6–3.7 in the parameter range shown in the figure. All these results make the pair-binding energy smaller in the Holstein-Hubbard model. For large Ω, the polaron pair becomes unbound, though our results are based on the adiabatic approximation and the nearest-neighbor pair approximation so that they are less reliable for large Ω.

We now turn to the effects of second-neighbor electron hybridization and long-range Coulomb couplings that lift the accidental finite-cluster degeneracy, $t_P^{(2)}=t_P^{(3)}$, and thus shift the system off the d-p phase boundary for $P=2$, already in the absence of $t_2^{(1a,1b)}$ processes. The second-neighbor electron-hopping term in the original Hamiltonian (1) enhances the second-neighbor hopping $t_2^{(2)}$, lowers the third-neighbor one $t_2^{(3)}$, and thus favors the $d_{x^2-y^2}$-wave symmetry if t' is positive by the definition in Sec. II (Fig. 10), and the effects are opposite if t' is negative (Fig. 11). Note that, in the noninteracting tight-binding model, the positive t' raises the energy of $p=(\pi,0)$ state [thus the energy of $p=(\pi/2,\pi/2)$ state is relatively lowered] and makes the Fermi

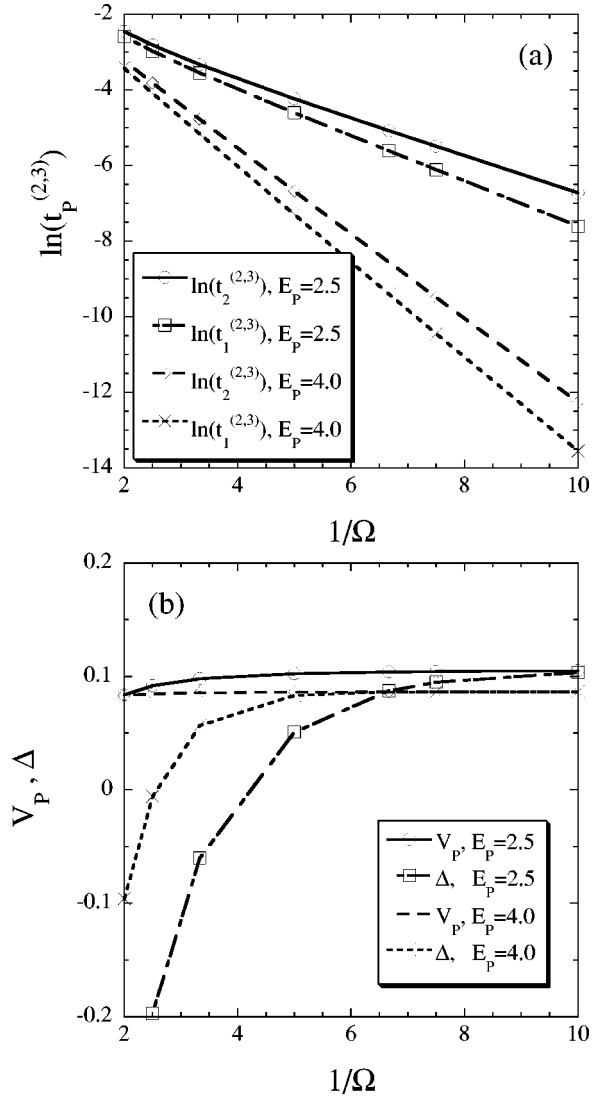

FIG. 9. (a) Logarithm of the effective polaron hopping amplitudes $t_P^{(\nu)}$ for $P=1$ and 2 holes and ν as second- and third-neighbor processes vs inverse phonon energy $1/\Omega$ (with $t_P^{(2)}=t_P^{(3)}$ due to accidental cluster symmetries). (b) Effective polaron nearest-neighbor attraction V_P and two-polaron binding energy Δ vs $1/\Omega$. All results are for $t\equiv 1, U=8t$, with $E_P=2.5t$ and $4.0t$, on an $N=8$ lattice with periodic boundary conditions.

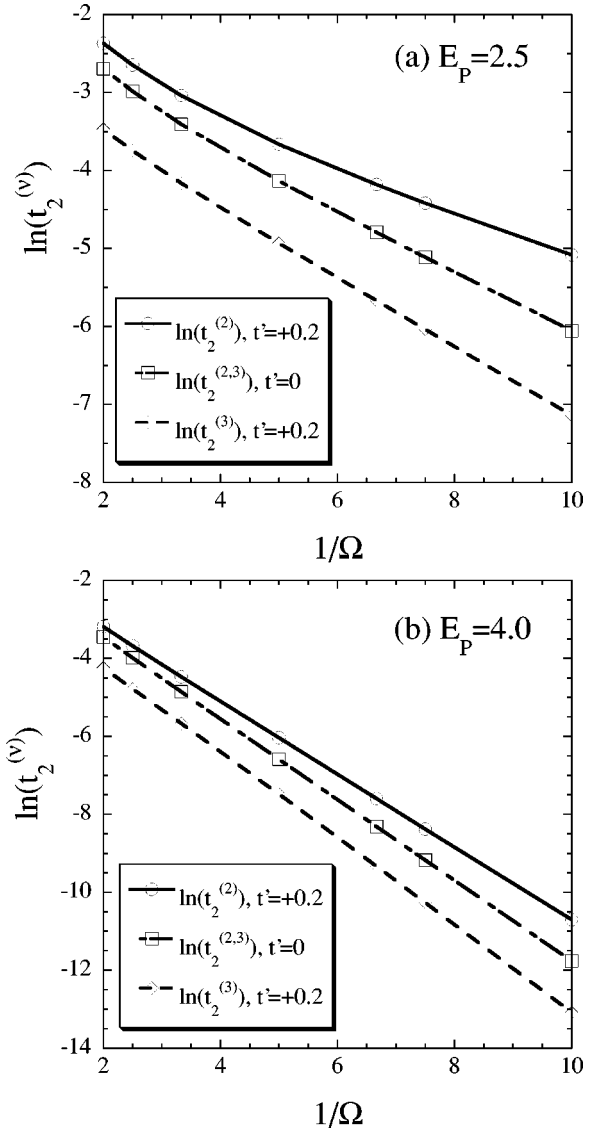

FIG. 10. Logarithm of the effective polaron hopping amplitudes $t_2^{(\nu)}$ for ν as second-and third-neighbor processes vs inverse phonon energy $1/\Omega$, with and without inclusion of next-nearest-neighbor hopping $t'=+0.2t$, for $t\equiv 1, J=0.5t$, (a) $E_P=2.5t$, and (b) $E_P=4.0t$, on an $N=8$ lattice with periodic boundary conditions.

surface convex. A t' term that helps the second-neighbor electron hopping also helps the second-neighbor polaron hopping.

The long-range repulsion term enhances $t_2^{(3)}$ more than $t_2^{(2)}$, so that it favors the $p_{x(y)}$-wave symmetry (Fig. 12). This can be understood if we recall the second-order perturbation theory with respect to $t_2^{(1a,1b)}/V_P$. The V_C term raises the energy of the intermediate second-neighbor pair favoring the $d_{x^2-y^2}$-wave symmetry, compared to that of the intermediate third-neighbor pair favoring the $p_{x(y)}$-wave symmetry. Note that V_C enhances both of the $t_2^{(2)}$ and $t_2^{(3)}$ processes. This happens because the lattice distortion and thus the localization potential is weakened by V_C. If V_C is too strong, however, it may overcome the nearest-neighbor attraction and the polaron pairing will then be suppressed altogether. This will be discussed further in the next section.

Other modifications of our model, going beyond the basic Einstein phonon model with Holstein EP coupling, can be considered. Such modifications include, for example, spatially extended EP coupling terms in $H_{\text{e-ph}}$, Eq. (2), and dispersion in the bare phonon spectrum, which introduces spatially extended elastic couplings into H_K, Eq. (3). There is no reason to believe that either modification will fundamentally alter our conclusions concerning (i) the basic conditions of polaron formation in the near-$\frac{1}{2}$-filled Hubbard or tJ electron system, (ii) Berry phases associated with the tunneling of such polarons, and (iii) the qualitative parameter dependences of their effective polaron tunneling amplitudes discussed in the present section.

As far as polaron formation in dimensions $d\geqslant 2$ is concerned, the polaron formation threshold $E_P^{(\text{crit})}$ will of course depend on the extended EP and elastic coupling parameters, but it will remain finite for short-range extended couplings by general scaling arguments.[28] For Fröhlich-type long-range EP couplings,[37] $E_P^{(\text{crit})}$ may be reduced, compared to the

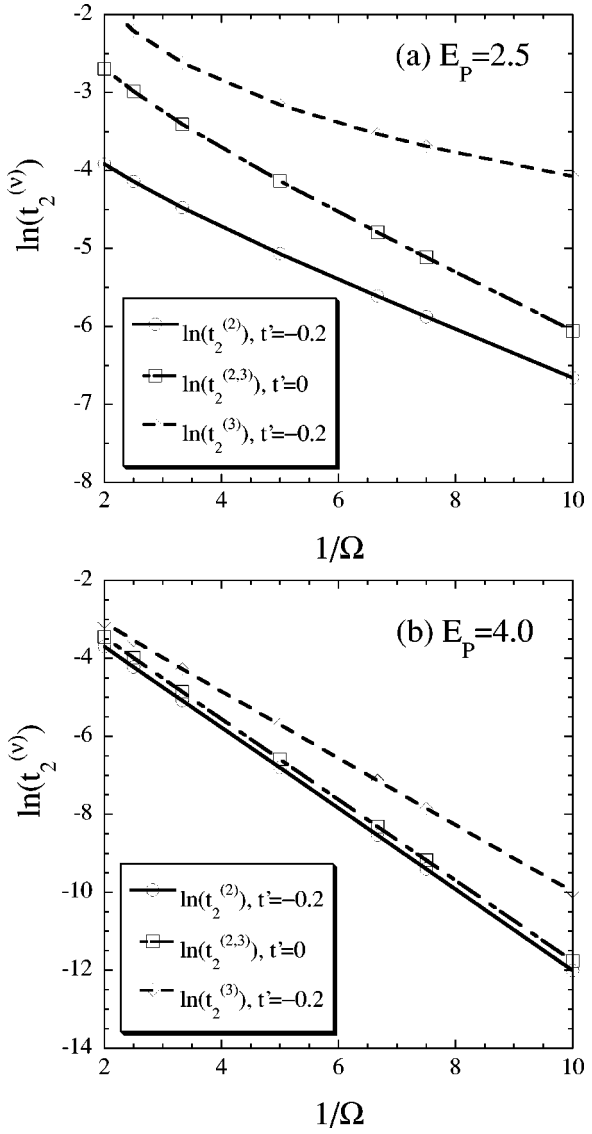

FIG. 11. Logarithm of the effective polaron hopping amplitudes $t_2^{(\nu)}$ for ν as second- and third-neighbor processes vs inverse phonon energy $1/\Omega$, with and without inclusion of next-nearest-neighbor hopping $t'=-0.2t$, for $t\equiv 1, J=0.5t$, (a) $E_P=2.5t$, and (b) $E_P=4.0t$, on an $N=8$ lattice with periodic boundary conditions.

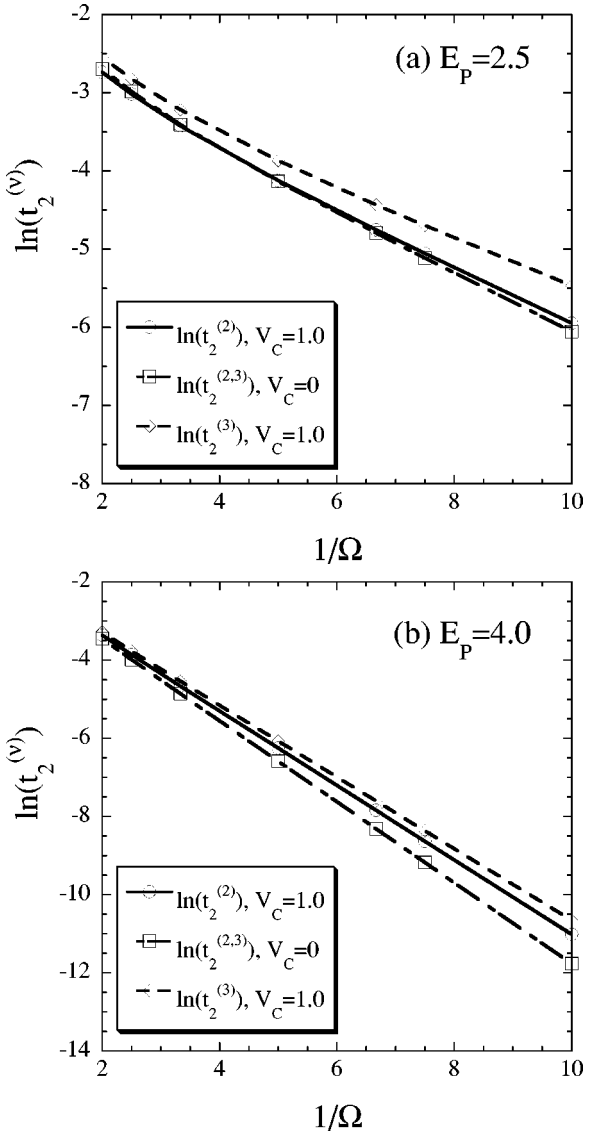

FIG. 12. Logarithm of the effective polaron hopping amplitudes $t_2^{(\nu)}$ for ν as second- and third-neighbor processes vs inverse phonon energy $1/\Omega$, with and without inclusion of long-range repulsion $V_C=1.0t$, for $t\equiv 1, J=0.5t$, (a) $E_P=2.5t$, and (b) $E_P=4.0t$, on an $N=8$ lattice with periodic boundary conditions.

short-range case. The latter type of coupling could favor a spatially more extended ''large'' polaron formation, as discussed further below. Most importantly, however, the basic physical arguments for an AF spin-correlation-induced reduction of $E_P^{(\mathrm{crit})}$, are quite independent of the structure of the phonon spectrum or the EP coupling.[11] The existence of self-localized states has been discussed for other types of phonon branches[38] and, in near-$\frac{1}{2}$-filled Hubbard electron systems, for other types of EP coupling.[11]

Due to their topological and *electronic* nature, we also believe that our Berry phase results are likely to be very robust against model extensions. Recall here that the Berry phase, as given in Eq. (20), is a property of the electronic wave function only. Hence, the foregoing should be taken with the proviso that the extensions of the model do not change the character or symmetry of the self-localized electronic wave function. That is to say, our Berry phase results will remain valid for any polaronic state in the near-$\frac{1}{2}$-filled Hubbard system in which the localizing EP potential wells Cu_j are centered around the (fourfold symmetric) Cu sites. Our Berry phase results may *not* hold in polaronic systems where the localizing EP potential well and the self-localized state forms, e.g., along a first neighbor Cu-Cu bond. Such ''bond-localized'' polaronic states are conceivable in EP models, such as the Su-Schrieffer-Heeger model,[24,39] where the lattice displacement modulates, for example, the first-neighbor electron intersite hybridization, rather than the electron on-site energy.

Both spatially extended EP and spatially extended elastic couplings *can* have important effects on effective polaron-polaron interactions, that is, on the V_P term in Eq. (55). For example, introducing either a first-neighbor elastic coupling K' or a first-neighbor EP coupling C' into H_{ph}, Eq. (3), or into $H_{\text{e-ph}}$, Eq. (2), respectively, will either increase or decrease the first-neighbor attraction V_P, depending on the sign of K'/K or C'/C, respectively.

Similar effects are found in extended EP coupling models involving either breathing or buckling modes of the oxygen atoms in the CuO_2 planes.[26,40–43] The former involve the displacements of the planar O atoms parallel to their in-plane Cu–O bonds, the latter involves the planar O displacements perpendicular to the CuO_2 plane. In the former, the simplest Cu–O first-neighbor EP coupling causes an effective electronic Cu–Cu first-neighbor repulsion, in the latter it causes a first-neighbor attraction.

The foregoing examples illustrate the general point that EP coupling to several different phonon branches can lead to subtle, competing effects when spatially extended pairing interactions are being considered.[6] This is very much in contrast to the conventional phonon-mediated on-site *s*-wave attraction where all phonon branches contribute attractively, regardless of the details of the EP couplings or phonon dispersion.[44] We caution, therefore, that one may not realistically try to construct a *phonon-mediated* $d_{x^2-y^2}$-pairing theory based on a phonon model that selectively includes only the EP coupling to a single branch. Whether phonons ultimately contribute to—or subtract from—the *d*-wave (or other non-*s*-wave) pairing potentials in the cuprates[45] is presently an open question that will require further study. The renormalization of both extended phonon-mediated and extended Coulomb interactions by the strongly correlated Hubbard electron system will be a central issue in such future studies.[46–48]

With regards to the present paper, we emphasize that we are *not* proposing a particular microscopic pairing mechanism. Rather, our primary purpose here is to explore the consequences of the polaron tunneling matrix elements and Berry phases on the pairing state, *provided that* a first-neighbor attractive pairing mechanism, of whatever microscopic origin, exists, i.e., provided that V_P in Eq. (55) is attractive.

IX. POLARON LIQUIDS AND THE CUPRATE SUPERCONDUCTORS

To the extent that the qualitative features of the above-discussed effective Hamiltonian (55) and the resulting tunneling and pairing dynamics remain intact at finite hole doping concentrations, the foregoing results have some potentially interesting consequences for the physical properties of the polaron liquid formed at finite polaron densities. In the present section, we will speculate on some of these properties and compare them to experimental observations in the cuprate high-T_c superconductors.[19]

If the above-discussed polaron-pair state remains stable and delocalized at finite hole doping, then formation of a superconducting polaron-pair condensate can occur at low enough temperatures. The foregoing discussion has focused primarily on the tightly bound pair limit where such a condensate would be formed via Bose condensation of the *preexisting* polaron pairs. However, the qualitative Ω dependences of the delocalization energies $t_P^{(\nu)}$ and of the pairing potential V_P, shown in Fig. 8, suggest that with increasing Ω (and fixed electronic parameters *t, J,* and E_P), such a condensate may exhibit a crossover from tightly bound pair to a BCS-like, extended-pair behavior: For small Ω, the delocalization matrix elements $t_2^{(\nu)}$ and resulting polaron pair bandwidth become small compared to the pairing potential V_P. Hence tightly bound pairs will form, as described above, with a pair wave function extending only over 1-2 lattice constants. For large Ω on the other hand, the polaron bandwidths (B_1 and B_2) can become comparable or larger than the pairing potential, thus leading to a BCS-like extended pair state, with a pair wave function extending over several/many lattice constants.

In the tightly bound pair regime, the Bose condensation T_c is controlled by the pair density $x_{\rm pair}$ and the pair bandwidth B_2, that is, roughly

$$T_c \sim x_{\rm pair} B_2 , \tag{73}$$

where B_2 is the pair bandwidth (65) and

$$x_{\rm pair} = \frac{1}{2}(1-\langle n\rangle) = \frac{1}{2}x \tag{74}$$

is the pair concentration, i.e., half of the hole concentration *x*. $B_2(\Omega)$ and hence T_c decreases with decreasing Ω.

In the BCS-like extended-pair regime, T_c is controlled by the pair binding energy Δ that decreases with increasing delocalization energy and hence with increasing Ω. As a consequence, there must exist, somewhere in the crossover regime between the tightly bound pair and the BCS (extended-pair) limits, an optimal phonon frequency Ω_0 where the transition temperature $T_c(\Omega)$ is maximized. Ω_0 is roughly determined by the condition

$$B_2(\Omega_0) \sim V_P , \tag{75}$$

and the maximum possible T_c (as a function of Ω), estimated by extrapolation of Eq. (73) from the tightly bound pair side, is of order

$$T_{c0} \equiv T_c(\Omega_0) \sim x_{\rm pair} B_2(\Omega_0) \sim x_{\rm pair} V_P , \tag{76}$$

where $B_2(\Omega)$ is the polaron pair bandwidth, Eq. (65), as a function of phonon frequency Ω.

One crucial, experimentally observable difference between the tightly bound and the extended pair condensate is the relation between pair formation and superconducting transition: In the tightly bound pair regime, the pairs, and hence the pairing gap Δ in the excitation spectrum, can be *preformed*. That is, the polaron pairs and the energy gap for pair breaking exist already at temperatures $T \sim \Delta$ that could be well above T_c, provided that $\Delta \gg T_c$. By contrast, in the extended-pair BCS-like regime we expect the pair formation to coincide with the superconducting transition, that is, the pairing gap should be observable only at temperatures T below T_c and should vanish at T_c.

The existence of such an optimum phonon frequency implies that T_c exhibits a vanishing isotope exponent α when $\Omega = \Omega_0$. To show this, we note that the isotopic mass dependence enters into the theory only via the phonon frequency Ω, if the electron-phonon Hamiltonian is parametrized, as in Eqs. (7) and (6), in terms of E_P and Ω, since electron-phonon potential constants (C) and harmonic restoring force constants (K), and hence E_P, are of purely electronic origin, i.e., do *not* depend on atomic/isotope masses. Using $\Omega \propto M^{-1/2}$, from Eq. (6), we obtain

$$\alpha \equiv -\left.\frac{\partial \ln T_c}{\partial \ln M}\right|_{\rm el} = \frac{1}{2}\left.\frac{\partial \ln T_c}{\partial \ln \Omega}\right|_{\rm el}, \tag{77}$$

which vanishes at the T_c maximum $\Omega=\Omega_0$. The notation $\ldots|_{\rm el}$ here means that the derivatives are to be taken with all purely electronic model parameters ($t, U, J, E_{\rm P}, u_{\rm P}$, etc.) held constant. The T_c maximum at Ω_0 also implies that α is positive in the tightly bound pair regime $\Omega_0>\Omega$, but negative in the extended-pair regime $\Omega_0<\Omega$. The vanishing of α at $\Omega=\Omega_0$ does, however, *not* imply that α is generically a small number. Quite to the contrary, because of the strong Ω dependence of the polaron bandwidth parameters, we should expect α to attain quite substantial magnitudes, with $|\alpha| \sim \mathcal{O}(1)$, as the system is tuned away from the optimal phonon frequency, i.e., when $\Omega \neq \Omega_0$.

It is tempting to compare the foregoing features of a finite-density polaron liquid to the observed properties of the cuprates. The doping dependence of the superconducting and normal-state properties of the cuprates is, in some respects, very much reminiscent of a crossover from tightly bound pair to BCS/extended-pair behavior: In the underdoped cuprates, there is now a substantial body of evidence suggesting that the superconducting gap is pre-existing, in the form of a ''pseudogap,'' at temperatures well above T_c.[49] With increasing hole doping concentration x, T_c approaches a maximum, while the pseudogap above T_c is gradually suppressed, and, in close proximity to the optimal doping concentration x_0, the pseudogap above T_c vanishes. Well inside the overdoped regime $x>x_0$, there is no detectable pseudogap and T_c rapidly decreases with increasing x.

The isotope exponents α in the underdoped cuprates are typically quite large in magnitude, of order of the classical BCS value $\alpha_{\rm BCS}=\frac{1}{2}$ or larger. However, in contrast to conventional BCS-type phonon-mediated superconductors, α can be very sensitive to changes in doping and other system properties such as impurity concentration and crystal structure. With increasing hole-doping concentration, the observed, usually positive oxygen isotope exponent α decreases and becomes very small, typically <0.05, at the optimal doping concentration x_0.[9] It is presently not clear whether α changes its sign for $x>x_0$. Negative α values have been observed in copper isotope substitutions on less than optimally doped cuprate materials.[50]

In comparing these experimental results to the foregoing theoretical picture of a polaron liquid, it is important to note that, experimentally, the T_c maximum and the surmised crossover from tightly bound pair to extended-pair BCS-like behavior is observed as a function of doping concentration x, whereas, in our above theoretical considerations, we have discussed the crossover as a function of phonon frequency Ω. To see how such a crossover could arise in our model as a function of doping, we need to consider the doping dependence of the polaron delocalization matrix elements $t_P^{(\nu)}$.

As indicated in Figs. 8–12, the polaron delocalization matrix elements, and hence the polaron pair bandwidth B_2 are rapidly increasing functions of Ω. At finite doping, these delocalization matrix elements will also become dependent on the hole doping concentration $x=1-\langle n\rangle$ by the following mechanism: As the polaron density increases, the localized wave functions of nearby holes will begin to overlap and the holes will begin to mutually screen out each other's tunneling barriers. This effect can be clearly seen in comparing the 1- and 2-hole results in Fig. 8. For $P=2$, the mere proximity of the second, static polaron strongly enhances the tunneling matrix element of the first, moving polaron, hence $t_2^{(\nu)}>t_1^{(\nu)}$ for $\nu=2,3$. Treated at the mean-field level, at finite polaron density, this tunneling enhancement effect will cause the (mean-field average) tunneling matrix elements to increase with the hole doping concentration. Thus the effective polaron-pair bandwidth $B_2=B_2(\Omega,x)$ becomes a strongly increasing function of the hole doping concentration x.

According to the crossover criterion (75) it may then be possible to drive the polaron liquid from the tightly bound pair into the extended-pair regime by changing either the phonon frequency Ω or the doping concentration x, if $V_{\rm P}$ is only weakly dependent on Ω and x. Another way of stating the same result is to say that the optimal phonon frequency $\Omega_0=\Omega_0(x)$, from Eq. (75), is a decreasing function of the hole doping concentration x. The underdoped region corresponds to the tightly bound pre-existing-pair regime in this picture; the overdoped region is identified with the extended-pair BCS-like regime. The superconducting transition temperature T_c as a function of x reaches a maximum at an optimal doping concentration x_0 not too far from the concentration x_Ω, where $\Omega_0(x_\Omega)=\Omega$ and the isotope exponent vanishes. Notice here that the point $x_0=x_0(\Omega)$ [where $T_c(\Omega,x)$ reaches its maximum as a function of x at fixed Ω] need not exactly coincide with the point x_Ω [where the optimal phonon frequency $\Omega_0(x)$ equals the actual phonon frequency Ω].

At sufficiently large hole doping the polaron-polaron wave-function overlap and the mutual screening of the hole-localizing potential wells Cu_j may become so strong that the holes become unbound, that is, the polarons become unstable towards forming free carriers. This finite-density polaron unbinding can be regarded as analogous to the Mott delocalization transition in moderately doped semiconductors. The primary difference is that the Mott transition in semiconductors involves the screening of localizing potential wells due to static impurities whereas, in the present case, the localizing potential wells are due to local lattice distortions that are induced, via the EP coupling, by the polaronic holes themselves. In the adiabatic potential $W(u)$, this unbinding will manifest itself in the (gradual or abrupt) disappearance of local minimum configurations $u^{(\xi)}$. Whether, in the thermodynamic limit, this occurs as a sharp transition or as a continuous crossover is presently unclear and needs further study.[12] The nature of the polaron unbinding and the characteristic concentration x_u where the unbinding occurs will also be influenced by the long-range Coulomb interaction $V_{\rm C}$ and, in more general EP models, by the spatial range of the EP interaction.[37]

If the optimal polaronic doping concentrations x_0 and x_Ω are close to the polaron unbinding concentration x_u, the polaron unbinding will likely dominate the crossover into the extended pair regime: In this scenario ($x_u \cong x_0, x_\Omega$), the crossover from underdoping to overdoping takes the system directly from the tightly bound polaron-pair liquid into a BCS-like superconductor of extended pairs of nonpolaronic carriers. The effective mass of the nonpolaronic carriers in the overdoped regime is much less enhanced by the electron-

phonon coupling and, more importantly, the mass enhancement is independent of the isotopic mass of the ions. The latter is suggested by the conventional weak-coupling electron-phonon theory where the mass enhancement factor is given by $(1+\lambda_z)$ and the Eliashberg parameter λ_z is independent of the isotope mass.[6,44] If the pairing attraction is of predominantly electronic (i.e., nonphonon) origin, one will then obtain a very small isotope exponent,[6] $|\alpha| \ll 1$, throughout the overdoped regime $x > x_u$.

Thus, the overall magnitude of α can serve as a distinguishing feature between extended pairs of polaronic and nonpolaronic carriers in the overdoped regime. In the former scenario, already described above, $\alpha(x)$ changes sign near optimal doping, but $|\alpha|$ well inside the overdoped regime can become as large as in the underdoped regime, reflecting the fact that the underlying pair constituents are still single-hole polarons. By contrast, in the latter (unbound carrier) scenario, $|\alpha|$ becomes small in the overdoped regime, without necessarily incurring a sign change in α, reflecting the nonpolaronic nature of the pair constituents. Further experimental studies of the isotope exponent in the overdoped cuprate systems would be desirable.

The foregoing features of the underdoped polaron liquid model and its crossover into the overdoped regime exhibit strong similarities with the observed pairing symmetry and doping dependences of T_c, isotope exponent, and pseudogap in the cuprates. However, in its present form, the model also suffers from several potential drawbacks that arise from the *small*-polaron character of the self-localized hole. Small-polaron formation necessarily implies bandwidths B_1 and B_2 that cannot be substantially larger than the phonon energy scale Ω, as shown in Figs. 8–12. As a consequence, small-polaron carriers may be easily localized by disorder and/or long-range Coulomb interaction effects. Also, by Eq. (76), the overall magnitude of the optimal $T_{c0} \sim \frac{1}{2} x B_2 \lesssim \frac{1}{2} x \Omega$ cannot exceed some fraction of Ω. With $x \sim 0.10-0.20$ and $\Omega \lesssim 1000$ K,[8] this upper limit on T_c is of order 50–100 K in the cuprates and it is reached if $E_{\rm P}$ just barely exceeds $E_{\rm P}^{(\rm crit)}$. For substantially larger $E_{\rm P}$, B_2 and T_c are rapidly (exponentially) suppressed with $E_{\rm P}$. It is not clear from the experimental data whether observed carrier mobilities, effective masses, and T_c's in the underdoped cuprates actually exhibit such a strong sensitivity to changes in $E_{\rm P}$ and/or to disorder or long-range Coulomb interactions.

The foregoing limitations of the small-polaron system can ultimately be traced back to the short-range nature of the assumed Holstein EP coupling in our model. Scaling arguments show that, at the level of the zeroth-order adiabatic approximation in spatial dimensions $D \geqslant 2$, short-range EP models are subject to a dichotomy whereby single carriers either form small polarons, if $E_{\rm P}$ exceeds a certain threshold $E_{\rm P}^{(\rm crit)} > 0$, or they do not form polarons at all, if $E_{\rm P} < E_{\rm P}^{(\rm crit)}$.[28] By contrast, in systems with additional longer-range EP couplings, such as the Fröhlich model,[28,37] as well as in 1D short-range EP models,[38,51] it is possible to form large polarons at arbitrarily weak $E_{\rm P}$, i.e., with $E_{\rm P}^{(\rm crit)} = 0$. It has been argued[37] that large-polaron and large-bipolaron models can remedy some of the above-described deficiencies of the small-polaron picture, while retaining most of the desirable physical features described above. Thus, in a large-polaron model, there is still preformed pair formation above the superconducting T_c in the underdoped regime; and the carrier bandwidth is still strongly reduced and dependent on isotope mass. Also, the possibility of a crossover to a BCS-type free-carrier superconductor, as a function of doping, via polaron unbinding, is retained in a large-polaron theory. However the dependence of the bandwidth and T_c on EP coupling strengths $E_{\rm P}$ and on phonon frequencies Ω is only algebraic, rather than exponential, and the overall magnitude of the large-polaron bandwidths can become substantially larger than in a small-polaron model, thus allowing for larger T_c's. The proposed large-polaron theories studied so far[37] have been based on phenomenological continuum models that of course cannot reproduce lattice-related features, such as the location of band minima and pairing symmetries discussed above for our 2D lattice model. It will therefore be of interest to extend our present work to lattice models with longer-range EP couplings. Such future studies should explore the possibility of large-polaron formation and the band structure and pair wave-function symmetry of large-polaron pairs.

Another critical problem in the above described polaron models is the inclusion of long-range Coulomb effects. Rough estimates based on a point charge model and measured long-wavelength dielectric constants[47] suggest that $V_{\rm C}/t$ in the cuprates could be as large as $1-2$, if only the electronic contribution to the dielectric screening, that is, only ϵ_∞, is taken into account. If additional screening from phonons, i.e., ϵ_0, is included, the estimated $V_{\rm C}/t$ is reduced to $0.15-0.3$. The former, $V_{\rm C}/t \sim 1-2$, would be sufficient to completely suppress the polaron pairing attraction in a system containing only two isolated holes, that is, in the limit of vanishing polaron density. The latter, $V_{\rm C}/t \sim 0.15-0.3$, may be overcome by the AF-mediated first-neighbor attraction, but the net attraction strength would still be substantially reduced by $V_{\rm C}$.[33,52] The suppression of extended pairing states, such as $d_{x^2-y^2}$ pairing, by the long-range part of the Coulomb interaction is a common problem in all extended pairing models that are currently under investigation. Recent studies of the metallic (in addition to insulating dielectric) screening of the extended Coulomb potential at finite doping density[47,48] have suggested that the screened Coulomb potential becomes substantially reduced, or even attractive, at doping concentrations $x \sim 0.1-0.2$. However, the foregoing studies are based on weak-coupling or diagrammatic approaches that do not include polaronic strong-EP effects. It therefore remains to be seen whether metallic screening in a finite-density polaron liquid will be sufficient to ''rescue'' the AF-driven pairing attraction from the repulsive long-range Coulomb interactions.

It is also worth re-emphasizing[37] the strong phonon contribution to the dielectric screening in the cuprates, as evidenced by the large measured dielectric constant ratio $\epsilon_0/\epsilon_\infty \gtrsim 6$.[37,47] This phonon contribution, which acts to reduce long-range Coulomb potentials, can be equivalently regarded as a long-range attraction, mediated by long-range (dipolar) EP interactions. This long-range EP interaction is in fact a primary agent causing (bi)polaron formation in the above-cited[37] phenomenological large-polaron models. It is therefore quite conceivable that, in a realistic model of the

cuprates, both AF- *and* EP-mediated attractions contribute to the overall pairing potential and that the EP contribution may even be the predominant one.

X. SUMMARY

In conclusion, we have developed a treatment of polaron tunneling dynamics on the basis of a path-integral formulation of the adiabatic approximation. The adiabatic treatment of polaron tunneling has been tested by comparison to exact numerical results for a two-site Holstein system. The breakdown of the adiabatic approach in the anti-adiabatic regime has been discussed and the resulting limitations of applicability for long-range polaron tunneling processes in lattice models have been identified. Using a combination of path-integral, many-body tight-binding, and exact diagonalization techniques, we have then explored the Berry phases and effective matrix elements for single- and two-polaron tunneling, the two-polaron quasiparticle statistics, effective two-polaron interactions, and polaron pairing states in the 2D Holstein-tJ and Holstein-Hubbard models near half filling. The effect of second-neighbor electron hybridization and long-range Coulomb repulsion has also been studied. Due to the AF spin correlations, single-polaron hopping is dominated by *intrasublattice* second- and third-neighbor processes. These processes are strongly enhanced by close proximity of a second polaron. The Berry phases imply either $d_{x^2-y^2}$- or $p_{x(y)}$-wave pair symmetries and effective spin-1/2-fermion quasiparticle statistics of dopant-induced polaron carriers. For the Holstein-tJ and Holstein-Hubbard models on the 8-, 10-, and 16-site clusters, the $d_{x^2-y^2}$-wave state is stable for two polarons. The second-neighbor hopping $H_{t'}$ favors the $d_{x^2-y^2}$-wave pair for $t'>0$, while the long-range Coulomb repulsion H_{LC} favors the $p_{x(y)}$-wave pair.

The strong on-site Hubbard-U Coulomb repulsion plays a crucial role in the formation of these pairing states. By keeping the electrons spatially separated and preventing on-site bipolaron formation, the Hubbard-U interaction acts, effectively, to greatly enhance the polaron tunneling bandwidths and, hence, their mobility, in the nearly $\frac{1}{2}$-filled regime.

For a hypothetical superconducting polaron pair condensate, our results imply qualitative doping dependences of the isotope effect, T_c, and pseudogap that are similar to those observed in the cuprates. Potential limitations of the present polaron model, arising from the short-range nature of the assumed EP coupling, have been pointed out. Further studies to include longer-range EP couplings, in combination with extended Coulomb interactions, have been outlined.

ACKNOWLEDGMENTS

One of us (H.B.S.) would like to thank D. Emin, J.P. Franck, K. Levin, and M. Norman for helpful discussions. This work was supported by Grant No. DMR-9215123 from the National Science Foundation and a Grant-in-Aid for Scientific Research on Priority Area ''Nanoscale Magnetism and Transport'' from the Ministry of Education, Science, Sports and Culture, Japan. Computing support from UCNS at the University of Georgia is gratefully acknowledged.

[1] See, for example, B.G. Levi, Phys. Today **46** (No. 5), 17 (1993); and references therein.

[2] C.C. Tsuei, J.R. Kirtley, C.C. Chi, Lock See Yu-Jahnes, A. Gupta, T. Shaw, J.Z. Sun, and M.B. Ketchen, Phys. Rev. Lett. **73**, 593 (1994); J.R. Kirtley, C.C. Tsuei, J.Z. Sun, C.C. Chi, L.S. Yu-Jahnes, A. Gupta, M. Rupp, and M.B. Ketchen, Nature (London) **373**, 225 (1995).

[3] D.A. Brawner and H.R. Ott, Phys. Rev. B **50**, 6530 (1994); A. Mathai, Y. Gim, R.C. Black, S. Amar, and F.C. Wellstood, Phys. Rev. Lett. **74**, 4523 (1995).

[4] D.A. Wollman, D.J. Van Harlingen, J. Giapintzakis, and D.M. Ginsberg, Phys. Rev. Lett. **74**, 797 (1995); J.H. Miller, Jr., Q.Y. Ying, Z.G. Zou, N.Q. Fan, J.H. Xu, M.F. Davis, and J.C. Wolfe, *ibid.* **74**, 2347 (1995).

[5] For a review, see D.J. Scalapino, Phys. Rep. **250**, 329 (1995), and references therein.

[6] H.-B. Schüttler and C.-H. Pao, Phys. Rev. Lett. **75**, 4504 (1995); C.-H. Pao and H.-B. Schüttler, J. Supercond. **8**, 633 (1995); J. Phys. Chem. Solids **56**, 1745 (1995); Phys. Rev. B **57**, 5051 (1998); cond-mat/9809321 (unpublished); J. Zhong and H.-B. Schüttler, cond-mat/9809383 (unpublished).

[7] Y. Endoh, K. Yamada, R.J. Birgeneau, D.R. Gabbe, H.P. Jenssen, M.A. Kastner, C.J. Peters, P.J. Picone, T.R. Thurston, J.M. Tranquada, G. Shirane, Y. Hidaka, M. Oda, Y. Enomoto, M. Suzuki, and T. Murakami, Phys. Rev. B **37**, 7443 (1988); J.M. Tranquada, D.E. Cox, W. Kunnmann, H. Moudden, G. Shirane, M. Suenaga, P. Zolliker, D. Vaknin, S.K. Sinka, M.S. Alvarez, A.J. Jacobson, and D.C. Johnston, Phys. Rev. Lett. **60**, 156 (1988); S.-W. Cheong, G. Aeppli, T.E. Mason, H. Mook, S.M. Hayden, P.C. Canfield, Z. Fisk, K.N. Clausen, and J. L. Martinez, *ibid.* **67**, 1791 (1991).

[8] *Lattice Effects in High-T_c Superconductors*, edited by Y. Bar-Yam, T. Egami, J. Mustre-de Leon, and A.R. Bishop (World Scientific, Singapore, 1992); A.R. Bishop, in *Strongly Correlated Electronic Materials: The Los Alamos Symposium 1993*, edited by K. Bedell, Z.Q. Wang, D. E. Meltzer, A.V. Balatsky, and E. Abrahams (Addison-Wesley, Reading, MA, 1994); T. Egami and S.J.L. Billinge, Prog. Mater. Sci. **38**, 359 (1994); in *Physical Properties of High Temperature Superconductors V*, edited by D.M. Ginsberg (World Scientific, Singapore, 1996); and references therein.

[9] M.K. Crawford, M.N. Kunchur, W.E. Farneth, E.M. McCarron III, and S.J. Poon, Phys. Rev. B **41**, 282 (1990); M.K. Crawford, W. E. Farneth, E.M. McCarron III, P.L. Harlow, and A.H. Modden, Science **250**, 1390 (1990); J.P. Franck, S. Gygax, G. Soerensen, E. Alshuler, A. Hnatiw, J. Jung, M.A.-K. Mohamed, M.K. Yu, G.I. Sproule, J. Chrzanowski, and J.C. Irwin, Physica C **185-189**, 1379 (1991); J.P. Franck, S. Harker, and J.H. Brewer, Phys. Rev. Lett. **71**, 283 (1993); J.P. Franck, in *Physical Properties of High T_c Superconductors IV*, edited by D.M. Ginsberg (World Scientific, Singapore, 1994), p. 189, and references therein.

[10] K. Yonemitsu, A.R. Bishop, and J. Lorenzana, Phys. Rev. Lett. **69**, 965 (1992); Phys. Rev. B **47**, 12 059 (1993).

[11] J. Zhong and H.-B. Schüttler, Phys. Rev. Lett. **69**, 1600 (1992).
[12] H. Röder, H. Fehske, and H. Bütter, Phys. Rev. B **47**, 12 420 (1993); H. Fehske, H. Röder, A. Mistriotis, and H. Bütter, J. Phys.: Condens. Matter **5**, 3565 (1993); H. Fehske, H. Röder, G. Wellein, and A. Mistriotis, Phys. Rev. B **51**, 16 582 (1995); G. Wellein, H. Röder, and H. Fehske, *ibid.* **53**, 9666 (1996); B. Bäuml, G. Wellein, and H. Fehske, *ibid.* **58**, 3663 (1998).
[13] A. Greco and A. Dobry, Solid State Commun. **99**, 473 (1996); D. Poilblanc, T. Sakai, D.J. Scalapino, and W. Hanke, Europhys. Lett. **34**, 367 (1996); T. Sakai, D. Poilblanc, and D. J. Scalapino, Phys. Rev. B **55**, 8445 (1997); S. Ishihara, T. Egami, and M. Tachiki, *ibid.* **55**, 3163 (1997).
[14] T. Holstein, Ann. Phys. (N.Y.) **8**, 325 (1959); **8**, 343 (1959).
[15] G. Delacrétaz, E.R. Grant, R.L. Whetten, L. Wöste, and J.W. Zwanziger, Phys. Rev. Lett. **56**, 2598 (1986).
[16] Y.-S.M. Wu and A. Kuppermann, Chem. Phys. Lett. **201**, 178 (1993); D.E. Adelman, H. Xu, and R.N. Zare, *ibid.* **203**, 573 (1993).
[17] D. Loss, D.P. DiVincenzo, and G. Grinstein, Phys. Rev. Lett. **69**, 3232 (1992); J. von Delft and C.L. Henley, *ibid.* **69**, 3236 (1992).
[18] A. Auerbach, Phys. Rev. Lett. **72**, 2931 (1994).
[19] K. Yonemitsu, J. Zhong, and H.-B. Schüttler (unpublished); H.-B. Schüttler, K. Yonemitsu, and J. Zhong, J. Supercond. **8**, 555 (1995).
[20] P.W. Anderson, Science **235**, 1196 (1987); F.C. Zhang and T.M. Rice, Phys. Rev. B **37**, 3759 (1988); H.-B. Schüttler and A.J. Fedro, *ibid.* **45**, 7588 (1992).
[21] R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[22] D. Emin and T. Holstein, Ann. Phys. (N.Y.) **53**, 439 (1969).
[23] J. Ranninger and U. Thibblin, Phys. Rev. B **45**, 7730 (1992).
[24] For the most recent numerical and analytical work on the two-site problem and on EP strong-coupling expansions, see M. Capone, W. Stephan, and M. Grilli, Phys. Rev. B **56**, 4484 (1997); W. Stephan, M. Capone, M. Grilli, and C. Castellani, Phys. Lett. A **227**, 120 (1997).
[25] I.G. Lang and Yu.A. Firsov, Zh. Éksp. Teor. Fiz. **43**, 1843 (1962) [Sov. Phys. JETP **16**, 1301 (1963)].
[26] H.-B. Schüttler, J. Zhong, and A.J. Fedro, in *Electronic Properties and Mechanisms of High T_c Superconductors*, edited by T. Oguchi, K. Kadowaki, and T. Sasaki (Elsevier Science Publishers, 1992), p. 295.
[27] C.-X. Chen and H.-B. Schüttler, Phys. Rev. B **41**, 8702 (1990).
[28] D. Emin and T.D. Holstein, Phys. Rev. Lett. **36**, 323 (1976).
[29] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982), Chap. 10.
[30] S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1985), Chap. 7.
[31] K. Yonemitsu, Phys. Rev. B **50**, 2899 (1994).
[32] S.A. Trugman, Phys. Rev. B **37**, 1597 (1988).
[33] V.J. Emery, S.A. Kivelson, and H.Q. Lin, Phys. Rev. Lett. **64**, 475 (1990); T. Barnes and M.D. Kovarik, Phys. Rev. B **42**, 6159 (1990).
[34] D. Arovas, J.R. Schrieffer, and F. Wilczek, Phys. Rev. Lett. **53**, 722 (1984).
[35] In Ref. 19, the momenta of the *p*-wave ground states were given incorrectly.
[36] A. S. Alexandrov and J. Ranninger, Phys. Rev. B **23**, 1796 (1981); **24**, 1164 (1981); **45**, 13 109 (1992).
[37] D. Emin, Phys. Rev. Lett. **62**, 1544 (1989); D. Emin and M.S. Hillery, Phys. Rev. B **39**, 6575 (1989); D. Emin, *ibid.* **48**, 13 691 (1993); **49**, 9157 (1994); Phys. Rev. Lett. **72**, 1052 (1994).
[38] H.-B. Schüttler and T. Holstein, Phys. Rev. Lett. **51**, 2337 (1983); Ann. Phys. (N.Y.) **166**, 93 (1986).
[39] W.P. Su, J.R. Schrieffer, and A.J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).
[40] A.J. Fedro and H.-B. Schüttler, Physica C **185-189**, 1673 (1991).
[41] J. Song and J.F. Annett, Phys. Rev. B **51**, 3840 (1995); **52**, 6930 (1995).
[42] N. Bulut and D.J. Scalapino, Phys. Rev. B **54**, 14 971 (1996).
[43] A. Nazarenko and E. Dagotto, Phys. Rev. B **53**, 2987 (1996).
[44] P.B. Allen and B. Mitrovic, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1982), Vol. 37, pp. 1–91.
[45] A.A. Abrikosov, Phys. Rev. B **52**, 15 738 (1995); **51**, 11 955 (1995); Physica C **222**, 191 (1994); **233**, 102 (1994); **244**, 243 (1995); A. A. Abrikosov, J.C. Campuzano, and K. Gofron, *ibid.* **214**, 73 (1993).
[46] R. Zeyher and M.L. Kulic, Phys. Rev. B **53**, 2850 (1996).
[47] H.-B. Schüttler, C. Gröber, H.G. Evertz, and W. Hanke, cond-mat/9805133 (unpublished).
[48] G. Esirgen, H.-B. Schüttler, and N.E. Bickers, cond-mat/9806264 (unpublished).
[49] H. Ding, T. Yokoya, J.C. Campuzano, T. Takahashi, M. Randeria, M.R. Norman, T. Mochiku, K. Kadowaki, and J. Giapinzakis, Nature (London) **382**, 51 (1996); A. G. Loeser, Z.-X. Shen, D.S. Dessau, D.S. Marshall, C.H. Park, P. Fournier, and A. Kapitulnik, Science **273**, 325 (1996).
[50] J.P. Franck and D.D. Lawrie, Physica C **235-240**, 1503 (1994); J. Supercond. **8**, 591 (1995); J.P. Franck, Phys. Scr. **T66**, 220 (1996).
[51] L.A. Turkevich and T.D. Holstein, Phys. Rev. B **35**, 7474 (1987); T. D. Holstein and L. A. Turkevich, *ibid.* **38**, 1901 (1988); **38**, 1923 (1988).
[52] See the work by Barnes and Kovarik in Ref. 33 and by C. Gazza, G.B. Martins, J. Riera, and E. Dagotto, , cond-mat/9803314 (unpublished).