Journal of Statistical Physics, Vol. 85, Nos. 3/4, 1996

Effect of Temperature on Polaronic and Bipolaronic Structures of the Adiabatic Holstein Model

C. Baesens[1, 2] and R. S. MacKay[1]

Received September 20, 1994; final March 15, 1996

It is proved that the polaronic and bipolaronic structures found in the adiabatic Holstein model at large electron–phonon coupling by Aubry, Abramovici, and Raimbault survive under connection of the electrons to a low-temperature heat bath, uniformly in the size of the system. Bounds are computed for one-dimensional nearest neighbor chains, and some sample solutions are continued numerically.

KEY WORDS: Electron–phonon interaction; temperature; polaron; Holstein model; bifurcation.

INTRODUCTION

Much of solid-state physics and chemistry has to do with the effects of interactions between electrons in a network of ions and distortions of the network. The latter are commonly called "phonons," whether or not they are wave-like in character. Because of the large mass ratio between ions and electrons, it is useful to study the *adiabatic approximation* where the phonon variables u_s describing the distortion of the network are regarded as classical and the electrons are assumed to relax instantaneously to equilibrium in the given phonon configuration. This approximation reduces the system to classical dynamics for the phonon variables $\mathbf{u} = (u_s)$ in an effective potential $W[\mathbf{u}]$.

The *Holstein model* is the simplest case of an electron–phonon system, where the electrons are assumed to be independent fermions on a network S with hopping amplitude $t\Delta_{rs}$ from site r to site s, the phonon variables

[1] DAMTP, Cambridge CB3 9EW, United Kingdom.
[2] On leave from Laboratoire de Physique URA CNRS 1796, Université de Bourgogne, Dijon, France.

0022-4715/96/1100-0471$09.50/0

u_s are independent one-degree-of-freedom harmonic oscillators, and the electron–phonon coupling energy is simply the product of u_s with the electronic density on site s. The shape of the network is not important for our purposes, the only requirements being that Δ be Hermitian and have finite l_2-norm (i.e., bounded spectrum). The assumption of independent harmonic oscillators can easily be relaxed to allow for some anharmonicity and coupling between the phonon variables, but we restrict to this case to keep the exposition simple. More general forms of electron–phonon coupling can be considered, but the above form can often be recovered by redefinition of u_s. Also, for simplicity of exposition we concentrate on the case of spinless fermions.

The most important aspects of the dynamics resulting from the adiabatic approximation are the set of local minima of W and the linearized motion around them. In ref. 2 a nearly complete understanding of these two aspects was found for the adiabatic Holstein model in the regime where electron–phonon interaction dominates electron hopping. The set of local minima of W was proved to be in one-to-one correspondence with the set of configurations of plus and minus signs on the network, and the spectrum for vibrations about each local minimum was proved to be bounded away from zero (the local minima have a "phonon gap"). The main open question is which of the local minima is the lowest, but this will depend crucially on the shape of the network S and we do not address it here.

In the present paper, these results are extended to the case of warm electrons, that is, instead of putting the electrons into their instantaneous ground state, we suppose they take up an instantaneous Fermi–Dirac distribution. We reach the same conclusions as ref. 2 provided the electronic temperature T and the chemical potential μ lie in a domain R to be given in (2.6).

One may criticize that it is inconsistent to do statistical mechanics for the electrons and deterministic mechanics for the phonons. It is true that when electron hopping is very weak, the time scale to achieve thermal equilibrium among the electrons might not be short, and it is true that maintaining the electrons at temperature T will lead to transfer of thermal energy to the phonons. However, our conclusions extend to significantly nonzero values of electron hopping (e.g., around 0.08 in dimensionless units for a 1D nearest neighbor chain), and because the mass ratio between ions and electrons is so large it is not implausible that this level of electron hopping is sufficient to lead to electron equilibration on a time scale shorter than that for motion of the ions. Second, it is plausible that at low enough temperatures transfer of energy from the electrons to the ions happens on a time scale longer than that for motion of the ions. We do not attempt to justify these assumptions quantitatively, but it seems that they

are often used in the theory of charge density waves. Ideally, we would like to study the statistical mechanics of the whole system. In the adiabatic approximation this reduces to studying the distribution $e^{-\beta W[u]}$, where $\beta = 1/T$. As this is dominated by the lowest local minima, and for weak electron hopping and chemical potential near $-\frac{1}{2}$ all the local minima turn out to have nearly the same height, our study of the local minima is an important first step.

Related problems are treated in refs. 7, 6, and 4. In particular, ref. 7 treats the adiabatic Holstein model on $\mathbb{Z}^d$ with translation symmetry, at the symmetry point where chemical potential $\mu = -\frac{1}{2}$; they find the ground states and prove existence of two low-temperature phases for $d \geqslant 2$. Our paper can be viewed as a first stage in analyzing the case of general μ and general networks.

The plan of the paper is as follows. The problem is formulated mathematically in Section 1. It is solved at the uncoupled limit in Section 2. In Section 3, it is proved that each of these solutions has a unique continuation for small electron hopping, uniformly in the size of the system. Our results are illustrated numerically on one-dimensional nearest neighbor chains in Section 4. First, a domain of electron hopping to which our continuation theorem applies is computed. Second, the ways certain solutions cease to be continuable if the electron hopping is too large are studied. The paper concludes with Section 5 on potential extensions of the work.

1. MATHEMATICAL FORMULATION

The Hamiltonian for the adiabatic Holstein model with spinless fermions is

$$\tilde{H} = \sum_{s \in S} \tfrac{1}{2}(p_s^2 + u_s^2) + h \tag{1.1}$$

with

$$h = \sum_{s \in S} c n_s u_s - t\tilde{\Delta} \tag{1.2}$$

where

$$n_s = a_s^\dagger a_s \tag{1.3}$$

is the *number operator* on site s for fermions with creation and annihilation operators $a_s^\dagger$ and a_s, respectively, satisfying the anticommutation relations $\{a_r, a_s\} = \{a_r^\dagger, a_s^\dagger\} = 0$, $\{a_r^\dagger, a_s\} = \delta_{rs}$; and

$$\tilde{\Delta} = \sum_{r,s \in S} a_r^\dagger \Delta_{rs} a_s \tag{1.4}$$

is an off-diagonal Hermitian *hopping operator*, i.e., preserving the total number

$$N = \sum_{s \in S} n_s \tag{1.5}$$

but not the individual numbers n_s, with the only condition being that $\|\Delta\|_2 < \infty$, e.g.,

$$\tilde{\Delta} = \sum_{s \in Z} a_{s+1}^\dagger a_s + a_s^\dagger a_{s+1} \tag{1.6}$$

for a 1D nearest neighbor chain.

There are two parameters: c represents electron–phonon coupling and t represents the fermion hopping amplitude. It is easily seen that only the ratio c^2: t plays a role. We will be interested in the neighborhood of the *atomic limit* $t = 0$, which Aubry named the "anti-integrable limit" in analogy with a related problem in dynamical systems.

Aubry *et al.*[2] proved that all the local minima of the energy for the adiabatic Holstein model at the atomic limit (where they are easy to compute) have locally unique continuations for small t, and they keep phonon gap. The proof was subsequently improved in refs. 3 and 8.

The purpose of the present paper is to extend these results to warm fermions. The fermions are assumed to come from a bath with temperature T (whose inverse will be denoted β as usual) and chemical potential μ. Then instead of looking for local minima of the expectation of (1.1), we should look for local minima of the *free energy*

$$W[\mathbf{u}] = \sum_{s \in S} \tfrac{1}{2} u_s^2 + F_{\text{elec}}[\mathbf{u}] \tag{1.7}$$

where (e.g., ref. 5)

$$F_{\text{elec}}[\mathbf{u}] = -T \operatorname{tr} \log(1 + e^{-\beta(H-\mu)}) \tag{1.8}$$

and

$$H = \operatorname{diag}(u_s) - t\Delta \tag{1.9}$$

is the single-fermion Hamiltonian corresponding to $\tilde{H}$.

Differentiating (1.7), we find that a necessary condition for a local minimum is

$$\Phi[\mathbf{u}] := \mathbf{u} + \boldsymbol{\rho}[\mathbf{u}] = 0 \tag{1.10}$$

where $\boldsymbol{\rho} = (\rho_s)$ with

$$\rho_s = \partial F_{\text{elec}} / \partial u_s \tag{1.11}$$

which is easily verified to have the interpretation as the fermionic density on site s. Solutions of (1.10) are nondegenerate local minima if and only if

$$M := I + \partial \boldsymbol{\rho} / \partial \mathbf{u} \tag{1.12}$$

is positive definite there. Equations (1.10) and (1.12) will be used to continue local minima of (1.7) as t in (1.9) is increased from zero.

2. LOCAL MINIMA AT THE ATOMIC LIMIT

At the atomic limit ($t = 0$),

$$F_{\text{elec}}[\mathbf{u}] = -T \sum_{s \in S} \log(1 + e^{-\beta(u_s - \mu)}) \tag{2.1}$$

and so

$$W[\mathbf{u}] = \sum_{s \in S} V(u_s) \tag{2.2}$$

where

$$V(u) = \tfrac{1}{2}u^2 - T \log(1 + e^{-\beta(u-\mu)}) \tag{2.3}$$

The function V is a double-well potential for (μ, T) in the indicated region of Fig. 1, whose boundary is given by the condition for a double root of V', i.e., a simultaneous root of

$$V'(u) = u + (e^{\beta(u-\mu)} + 1)^{-1} = 0 \tag{2.4}$$

$$V''(u) = 1 - \frac{\beta}{4} \operatorname{sech}^2 \frac{1}{2} \beta(u - \mu) = 0 \tag{2.5}$$

The symmetry about $\mu = -\frac{1}{2}$ is due to the symmetry $a_s \mapsto a_s^\dagger$, $u_s \mapsto -1 - u_s$, $\mu \mapsto -1 - \mu$; which is a special but inessential feature of the model (1.1). The double-well region can be expressed as

$$R = \{(\mu, T) : 0 \leqslant T < 1/4, \mu_- < \mu < \mu_+\} \tag{2.6}$$

where

$$\mu_\pm = -T \log x_\pm - 1/(1 + x_\pm) \tag{2.7}$$

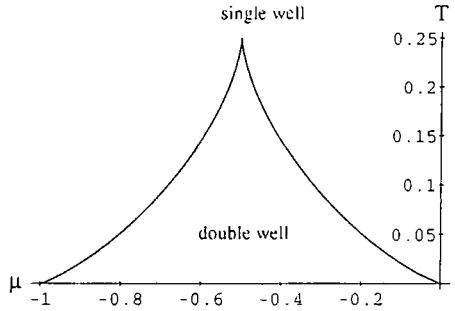

Fig. 1. The region of the plane of chemical potential μ and temperature T for which the function V of (2.3) is a double-well potential.

with

$$x_{\pm} = -1 + \beta/2 \pm \left(\frac{\beta^2}{4} - \beta\right)^{1/2} \tag{2.8}$$

We denote the two local minima of V by $U_-(\mu, T) < U_+(\mu, T)$, and their heights by $V_{\pm}(\mu, T)$ respectively. We deduce that for $(\mu, T) \in R$, there is a local minimum μ^{ξ} of the free energy corresponding to each choice of configuration $\xi = (\xi_s)_{s \in S}$ of plus and minus signs on the network S:

$$(\mathbf{u}^{\xi})_s = U_{\xi_s} \tag{2.9}$$

Its free energy is

$$W = N_+ V_+(\mu, T) + N_- V_-(\mu, T) \tag{2.10}$$

where $N_{\pm}$ are the numbers of $+$ and $-$ signs in the chosen configuration. Note that $V_- < V_+$ for $\mu > -\frac{1}{2}$ and $V_- > V_+$ for $\mu < -\frac{1}{2}$.

Note also that by (1.10), the electronic density on site s for an equilibrium state is simply $\rho_s = -u_s$. Hence for solutions at the atomic limit, ρ takes the values $\rho_{\pm} = -U_{\pm}$ and $\rho_- > \rho_+$. As $T \to 0$, we have $\rho_- \to 1$ and $\rho_+ \to 0$, so we can think of a minus sign as a polaron and a plus sign as an empty site.

3. CONTINUATION FOR SMALL HOPPING

We will prove that the local minima of Section 2 can be continued for small t (uniformly in the choice of configuration of plus and minus signs and in the size of the network S, but dependent on μ and T). The technique, as in refs. 3 and 8, is to use the implicit function theorem on Eq. (1.10),

using supremum norm ($\|\mathbf{u}\|_\infty = \sup\{|u_s| : s \in S\}$) on the space of phonon configurations $\mathbf{u}$. Thus we have to check that $\boldsymbol{\rho}[\mathbf{u}, t]$ is C^1 and that M of (1.12) is invertible (both statements being with respect to supremum norm). The conclusion is that solutions of (1.10) can be continued with respect to t as long as M remains invertible. Furthermore, they satisfy

$$d\mathbf{u}/dt = -M^{-1}\,\partial\boldsymbol{\rho}/\partial t \tag{3.0}$$

so we can obtain lower bounds on the distance in t for which they can be continued if we obtain upper bounds on the norms of M^{-1} and $\partial\boldsymbol{\rho}/\partial t$. Note that while M remains invertible, the solutions also remain local minima, because they cannot change index.

For H as in (1.9) and $E \in \mathbb{C}\backslash \operatorname{spec} H$, define the resolvent operator

$$G(E) = (E - H)^{-1} \tag{3.1}$$

Then using spectral projection (e.g., ref. 9), we can write (1.8) as

$$F_{\text{elec}} = -\frac{T}{2\pi i} \operatorname{tr} \int_\Gamma \log(1 + e^{-\beta(E-\mu)})\, G(E)\, dE \tag{3.2}$$

where Γ is any sum of closed contours whose winding number about each point of the spectrum of H is $+1$ and about each pole of the logarithm is 0. As will be seen shortly, in the regime of interest the spectrum of H is contained in two intervals $I^{\pm}$ on the real axis, and μ lies between them. An example of a contour Γ corresponding to this situation is sketched in Fig. 2. We denote the set of poles of the logarithm by

$$P = \{\mu + (2n+1)\, i\pi T : n \in \mathbb{Z}\} \tag{3.3}$$

From (3.1), for $E \notin \operatorname{spec} H$, G is differentiable with respect to any parameter for which H is, with derivative

$$G' = GH'G \tag{3.4}$$

Hence from (1.11), (3.2), and (1.9), ρ is given by

$$\rho_s = -\frac{T}{2\pi i} \operatorname{tr} \int_\Gamma \log(1 + e^{-\beta(E-\mu)})\, G\pi_s G\, dE \tag{3.5}$$

where π_s denotes the projection operator onto site s. Now

$$\operatorname{tr} G\pi_s G = (G^2)_{ss} \tag{3.6}$$

and then we can use the following lemma.

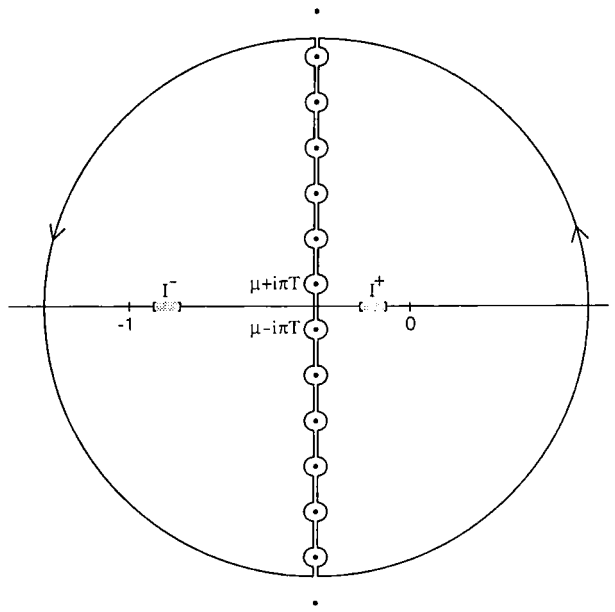

Fig. 2. A contour Γ for Eq. (3.2).

Lemma. For any analytic function f and closed contour Γ,

$$\int_\Gamma f(E)\, G^2\, dE = \int_\Gamma f'(E)\, G\, dE \tag{3.7}$$

Proof. Integration by parts, using $dG/dE = -G^2$, from (3.1). ∎

Hence

$$\rho_s = \frac{1}{2\pi i} \int_\Gamma \sigma(E)\, G_{ss}\, dE \tag{3.8}$$

where

$$\sigma(E) = (e^{\beta(E-\mu)} + 1)^{-1} \tag{3.9}$$

will be recognized as the Fermi–Dirac filling factor, which confirms the interpretation of ρ_s as fermionic density on site s.

Checking that $\boldsymbol{\rho}$ is C^1 with respect to $\mathbf{u}$ and t is virtually the same as the zero-temperature case treated in ref. 3, so we skip that step here. Now

we are ready to estimate the derivatives of $\boldsymbol{\rho}$ with respect to $\mathbf{u}$ and t. Using (3.8), (3.4), and (1.9), we have

$$\frac{\partial \rho_s}{\partial u_r} = \frac{1}{2\pi i} \int_\Gamma dE\, \sigma(E)\, G_{sr} G_{rs} \tag{3.10}$$

$$\frac{\partial \rho_s}{\partial t} = -\frac{1}{2\pi i} \int_\Gamma dE\, \sigma(E) (G\, \Delta G)_{ss} \tag{3.11}$$

The simplest choice for the contour Γ is a circle around μ of radius $2N\pi T$, N large, plus a small backward circle around each point of the set P of (3.3) inside the large circle. It is easily seen that the contribution of the large circle goes to zero as its radius goes to infinity, hence it does not contribute. Each small circle contributes $-2\pi i$ times the residue of the pole of σ it surrounds. Hence for any function X (analytic at P)

$$\frac{1}{2\pi i} \int_\Gamma dE\, \sigma(E)\, X(E) = T \sum_{E \in P} X(E) \tag{3.12}$$

This can be seen as a discretization of the integral along a vertical line that we used in the zero-temperature case in refs. 3 and 8.

We estimate the sizes of $G_{sr} G_{rs}$ and $(G\, \Delta G)_{ss}$ as in ref. 3. The first step is to show that the spectrum of H does not approach too closely to the set P. As in refs. 2, 3, and 8, if the u_s are within τ of their initial positions ($U_\pm$), then the spectrum of H is contained in the union of two intervals $I^\pm$ of length at most $2(\tau + t \|\Delta\|_2)$ centered on $U_\pm$. Hence, if

$$U_- + \tau + t \|\Delta\|_2 < \mu < U_+ - \tau - t \|\Delta\|_2 \tag{3.13}$$

(as is the case initially in the double-well regime), then the line through P avoids spec H by at least

$$\tfrac{1}{2}\varepsilon := \min(U_+ - \tau - t \|\Delta\|_2 - \mu, \mu - U_- \tau - t \|\Delta\|_2) \tag{3.14}$$

There are only two significant changes from ref. 3 in the remainder of the estimations. The first is the change of the integrals of ref. 3 to sums as in (3.12), and the second is that the "zeroth-order" part of the diagonal terms $\partial \rho_s / \partial u_s$ no longer vanishes, but is instead given by

$$T \sum_{E \in P} (E - u_s)^{-2} \tag{3.15}$$

Reversing the transition (3.12) between integral and sum and using residue calculus, this is easily evaluated to be

$$-\frac{\beta}{4}\operatorname{sech}^2\frac{1}{2}\beta(u_s-\mu) \tag{3.16}$$

which can be recognized as the electronic contribution to $V''(u_s)$, visible in (2.5). So we write

$$M=\operatorname{diag}(V''(u_s))+j \tag{3.17}$$

Completing the estimation analogously to ref. 3, we obtain the bounds

$$\begin{aligned}\|j\|_\infty \leqslant J := 6t^2 S_\Delta \varepsilon^{-3}&\left(\tanh\frac{\varepsilon\beta}{4}-\frac{\varepsilon\beta}{4}\operatorname{sech}^2\frac{\varepsilon\beta}{4}\right)\\ &+6t^4 S_\Delta^2\varepsilon^{-5}\left(\tanh\frac{\varepsilon\beta}{4}-\frac{\varepsilon\beta}{4}\operatorname{sech}^2\frac{\varepsilon\beta}{4}\right.\\ &\left.-\frac{\varepsilon^2\beta^2}{24}\tanh\frac{\varepsilon\beta}{4}\operatorname{sech}^2\frac{\varepsilon\beta}{4}\right)\end{aligned} \tag{3.18}$$

$$\begin{aligned}\left\|\frac{\partial\boldsymbol{\rho}}{\partial t}\right\|_\infty \leqslant L := 8\,|t|\,S_\Delta\pi^{-1}\varepsilon^{-2}&\left(1+\frac{4\pi}{\varepsilon\beta}\right)+2\,\|\Delta\|_2\,t^2 S_\Delta\varepsilon^{-3}\\ &\times\left(\tanh\frac{\varepsilon\beta}{4}-\frac{\varepsilon\beta}{4}\operatorname{sech}^2\frac{\varepsilon\beta}{4}\right)\end{aligned} \tag{3.19}$$

Here,

$$S_\Delta=\sup_{r\in S}\sum_{s\in S}|\Delta_{rs}|^2 \tag{3.20}$$

Let us define

$$f(\tau)=\min(V''(U_-+\tau),\,V''(U_+-\tau)) \tag{3.21}$$

which is a lower bound for $V''(u_s)$, provided that the u_s do not pass the inflection points of V at $\mu\pm x$, where x satisfies

$$\operatorname{sech}^2\tfrac{1}{2}\beta x=4T \tag{3.22}$$

This is true provided τ does not pass a zero of f. Then

$$\|M^{-1}\|_\infty\leqslant(f(\tau)-J)^{-1} \tag{3.23}$$

From Eq. (3.0) we deduce that the solutions can be continued with respect to t, with the bound

$$|d\tau/dt| \leqslant L/(f(\tau) - J) \tag{3.24}$$

as long as the denominator remains positive. Note that the right-hand side of (3.24) starts positive, because the local minima at which we start are nondegenerate, and $J = 0$ at $t = 0$. This differential inequality can be studied by integrating the pair

$$\begin{aligned} d\tau/dz &= L(t, \tau) \\ dt/dz &= f(\tau) - J(t, \tau) \end{aligned} \tag{3.25}$$

with respect to a pseudo-time z, from $\tau = t = 0$, until the first time that $f = J$ (note that this happens before f reaches 0, so we do not need to check that condition separately). Denoting by t_0 the value of t reached then, we deduce that the solutions can be continued for $|t| < t_0$. By the construction, t_0 depends only on $(\mu, T) \in R$, and $\|\Delta\|_2$ and S_Δ.

Furthermore, the solutions retain a phonon gap as long as they can be continued by our method. This is because the phonon spectrum for an equilibrium of the adiabatic Holstein model is precisely that of M, and thus there is phonon gap iff M is invertible in l_2 (and then its value is $\|M^{-1}\|_2^{-1}$). But M is invertible in l_2 as long as it is invertible with respect to supremum norm (in fact $\|M^{-1}\|_2 \leqslant \|M^{-1}\|_\infty$), because it is symmetric. Incidentally, for more general form of phonon kinetic energy, the phonon spectrum is not directly related to the spectrum of M, but it is still true that there is a phonon gap if and only if M is invertible in supremum norm.

4. NUMERICAL ILLUSTRATIONS

As an illustration of the continuation proof given in Section 3, Fig. 3 indicates the bound t_0 obtained by numerical solution of the differential equations (3.25) for a one-dimensional nearest neighbor chain ($\Delta_{rs} = 1$ for $|r - s| = 1$, and 0 otherwise), as a function of μ and T. For this calculation, we used the bounds $\|\Delta\|_2 \leqslant 2$, $S_\Delta \leqslant 2$. Note that these are optimal bounds for the infinite chain, and for all finite chains of length at least 3 with periodic boundary conditions, and the bound $S_\Delta \leqslant 2$ is also optimal for chains of length at least 3 with free ends.

Note that at $T = 0$, μ plays no role as long as it is in the electronic gap between occupied and unoccupied energy levels. However, our procedure for obtaining bounds for general temperature is not able to take this into

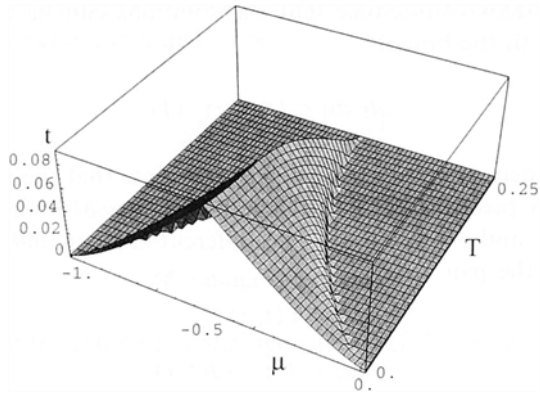

Fig. 3. A numerically computed region of (μ, T, t) space in which continuation is guaranteed, for a 1D nearest neighbor chain.

account, and so the bound t_0 obtained in Fig. 3 depends on μ even at $T=0$. In a sense, in refs. 3 and 8 we chose $\mu=-\frac{1}{2}$, which optimized the results.

Next we give some numerical illustrations of the continuation of specific equilibrium states for one-dimensional chains. These are computed by using Newton's method on Eq. (1.10), where $\boldsymbol{\rho}$ and $\partial\boldsymbol{\rho}/\partial\mathbf{u}$ are evaluated from the eigenvalues and eigenvectors of H, Eq. (1.9),

$$H\Psi^\nu = E^\nu \Psi^\nu \tag{4.1}$$

as follows:

$$\rho_s = \sum_\nu \sigma(E^\nu)\,|\Psi_s^\nu|^2 \tag{4.2}$$

$$\partial\rho_s/\partial u_r = \sum_{\nu,\alpha} \Psi_s^\nu \Psi_r^{\nu *} \Psi_r^\alpha \Psi_s^{\alpha *} g(E^\alpha, E^\nu) \tag{4.3}$$

where

$$g(E, E') = \begin{cases} (\sigma(E)-\sigma(E'))/(E-E') & \text{if } E \neq E' \\ \sigma'(E) & \text{otherwise} \end{cases} \tag{4.4}$$

We take chains with free ends so that H is tridiagonal, which simplifies the computation of its spectrum.

Figure 4 shows the continuation of the state $+\,+\,+\,+\,+\,+\,-\,+\,+\,+\,+\,+\,+$ from $t=0$ for a chain of length $n=13$ at $\mu=-\frac{1}{2}$, $T=0.05$, until M

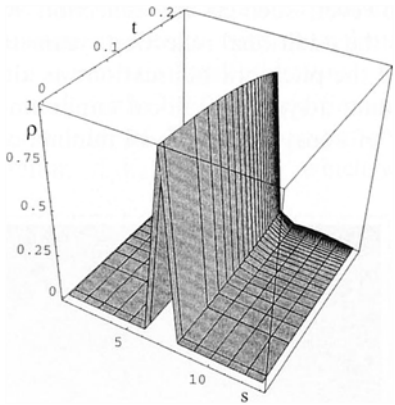

Fig. 4. Continuation of a polaron on the central site of a chain of length 13 with $\mu = -\frac{1}{2}$, $T = 0.05$, from $t = 0$ until it is annihilated at about $t = 0.21363$.

ceases to be invertible, which happens at about $t = 0.21363$. This is above our bound of Fig. 3, as indeed it must be, but by a factor of less than 3, indicating that our bound is not grossly conservative.

Note that the total electronic charge

$$Q = \sum_s \rho_s \tag{4.5}$$

associated with an equilibrium state typically varies with t. If it is desired to maintain constant electronic charge, then μ should be simultaneously varied to achieve it, by adding (4.5) (or more easily, $\sum u_s = -Q$) to the system of equations (1.10) to be solved, and μ to the set of variables $\mathbf{u}$ to be solved for.

An interesting question is to find with which solution(s) the above solution bifurcates when M loses invertibility. Usually, for a critical point of a family of smooth functions on a finite-dimensional manifold, degeneracy is lost by collision with one or more other critical points, which is called *bifurcation*. There are some constraints on the ways this can happen, which are provided by Morse theory. The simplest constraint is that the sum of the parities of the indices of the critical points involved in a bifurcation is conserved. Recall that the *index* of a nondegenerate critical point is the number of negative squares in the Lagrange normal form for the second variation, or equivalently the number of negative eigenvalues. The *parity* is $+1$ for even index, -1 for odd index. In generic one-parameter families, the only way that a local minimum (index 0) can become degenerate is by annihilation with a saddle (index 1) in a fold bifurcation. Our system has

some symmetries, however, such as the reflection $R_1: u_s \mapsto u_{n+1-s}$, and when $\mu = -\frac{1}{2}$ there is the additional reflection symmetry $R_2: u_s \mapsto -1-u_s$. These symmetries add the pitchfork bifurcations as alternative robust ways to lose a local minimum: a symmetric local minimum changes to a saddle, giving birth to a pair of nonsymmetric local minima or absorbing a pair of

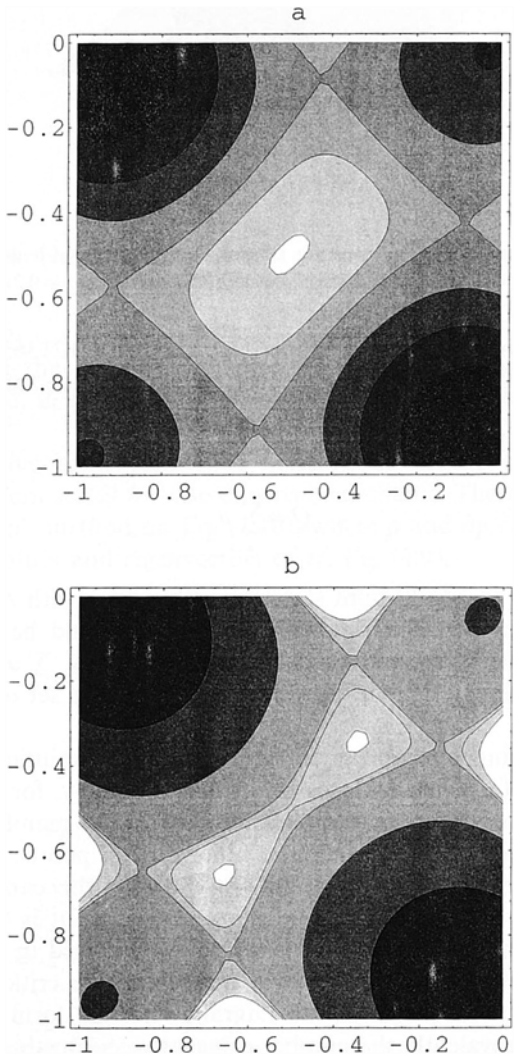

Fig. 5a–e. Contour plots of the free energy for the two-site Holstein model with $\mu = -\frac{1}{2}$, $T = 0.1$ at five values of $t = 0.18$, 0.225, 0.255, 0.35 and 0.6 (figures 5a–e, respectively); darker regions have lower free energy.

nonsymmetric saddles. They also provide an alternative way to lose a nonsymmetric local minimum: it merges with its reflection and a symmetric saddle, giving rise to a symmetric local minimum.

Now for (μ, T) in R with $T \neq 0$, in addition to 2^n local minima, a chain of length n has a total of 3^n critical points at the atomic limit, by using the local maximum of V [Eq. (2.3)], which we denote by U_m, for

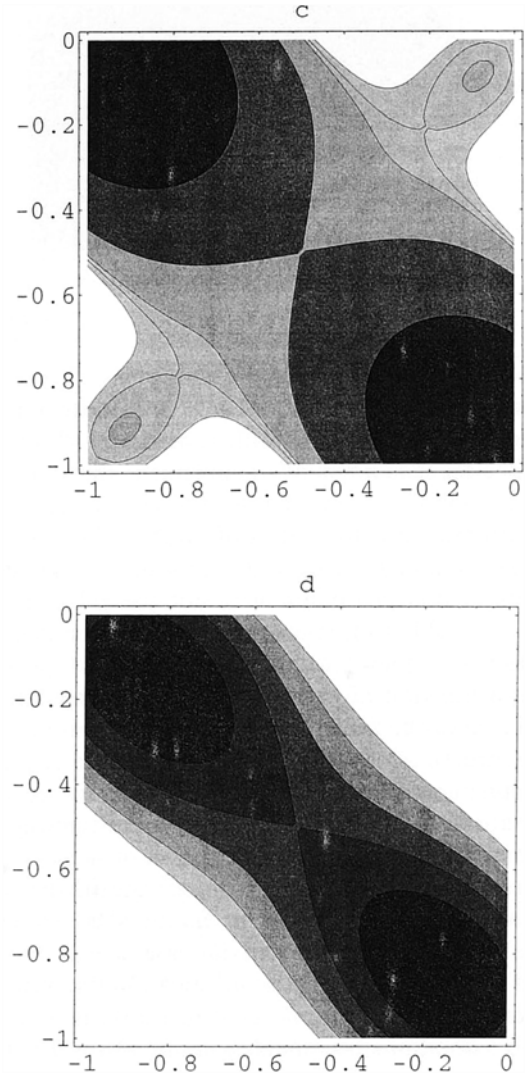

Fig. 5. (*Continued*)

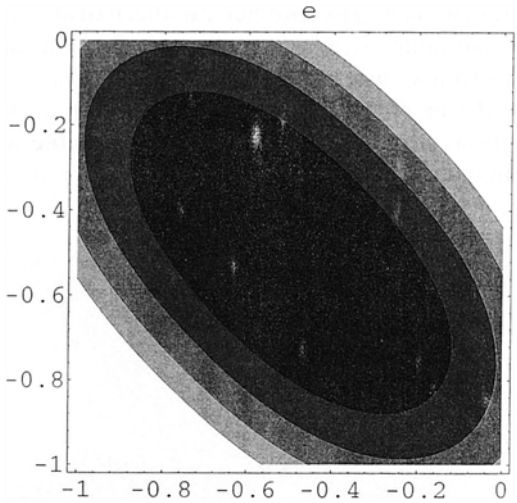

Fig. 5. (*Continued*)

some sites instead of the minima $U_{\pm}$. These 3^n critical points are all nondegenerate and can all be continued with respect to t, though the bound t_0 applies only to the local minima. We supplement the symbols $\{+, -\}$ of Section 2 by the symbol m, representing the choice U_m, so we label the solutions at the atomic limit by strings of symbols from $\{-, m, +\}$. Note that the local maximum of V ceases to be smooth at $T=0$. For this reason, the critical points of nonzero index were not treated in refs. 2, 3, and 8. Nevertheless, it is possible to first eliminate the phonon degrees of freedom instead of the electronic ones, and hence view the $T=0$ problem as finding critical points of a function of electronic degrees of freedom; this function is smooth and has many critical points other than local minima. The catch is that they are nearly all degenerate at the atomic limit, forming subvarieties, so a refinement of our methods is needed to see what continues to $t>0$, but we are in the process of treating this situation.

We tried continuing some saddles, e.g., $++++++m++++++$, but none of the ones we tried became degenerate in the same place [in $(\mathbf{u}, t)$] as $++++++-++++++$. So to shed some light on the problem, we studied the simplest nontrivial case: $n=2$.

For a chain of length 2, with $\mu=-\frac{1}{2}$ and $T=0.1$, we continued each of the critical points from the atomic limit in t until it became degenerate. The $++$ and $--$ solutions become degenerate at $t\simeq 0.264$, and the $+-$ and $-+$ solutions at $t\simeq 0.49$. But the saddles $+m$, $-m$, $m+$, $m-$ all become degenerate before these, at $t\simeq 0.246$, and the maximum mm even

earlier, at $t \simeq 0.206$. To see how the branches connect, we plotted the free energy as a function of $\mathbf{u} = (u_1, u_2)$ for intermediate values of t in Fig. 5. We see that two additional branches of critical points are created by a pitchfork from the central maximum mm at $t \simeq 0.206$, they absorb the saddles by a pitchfork at $t \simeq 0.246$, and finally annihilate with the minima $++$ and $--$, respectively at $t \simeq 0.264$.

We expect it to be a general feature that there will be bifurcations creating further critical points of nonzero index before any of the local minima bifurcate, especially as $T \to 0$, because the zone of negative curvature around U_m is small and so the critical points of positive index are less "stable" under continuation. Thus we expect that it will often happen that the local minima from the atomic limit are destroyed by bifurcation with secondary critical points which are not direct continuations from the atomic limit.

5. EXTENSIONS

Extension to spin-$\frac{1}{2}$ fermions is straightforward, as in ref. 3. One introduces chemical potentials $\mu_\uparrow, \mu_\downarrow$, and has to sum along two lines of poles, but the ideas remain the same. This leads to local minima consisting of arbitrary configurations of polarons, bipolarons, and holes, for weak enough electron hopping.

Finite coherence length, i.e., exponential decay of the response to a localized external force, can also be proved for networks with suitable form of coupling, along the lines of refs. 2 and 8.

The phonon energy can easily be modified from $\sum_{s \in S} \frac{1}{2} u_s^2$ to other functionals thus allowing nonlinearity and dispersion (as in refs. 2 and 3), provided that at the atomic limit the configuration has a phonon gap.

A natural question is whether we have found all the local minima for $|t| < t_0$. By constructing domains of contraction (as in ref. 2) for the Newton step used in the proof of the implicit function theorem about each atomic solution, and showing that outside the union of these neighborhoods either $\Phi[\mathbf{u}]$ is too far from zero or $D\Phi$ is negative in some directions, we expect that one can prove there exists a $t_1 > 0$, depending only on $(\mu, T) \in R$, and $\|\Delta\|_2$ and S_Δ, such that there are no other local minima for $|t| < t_1$, but we have not yet done so.

The same strategy of proof can also be used for (μ, T) outside R, but the results are less interesting. If $(\mu, T) \notin R$, then V has a single well, so there is only one local minimum at the atomic limit. We can prove that it continues for at least some interval of t. Furthermore, it should be easy to prove that it remains the only critical point for some interval of t.

For $(\mu, T) \in R$, it would be very interesting to find out which of the local minima is the lowest, and to study properties of the distribution $e^{-\beta W[\mathbf{u}]}$.

Finally, it would be most interesting to extend the results to the cases of quantum phonons or the Holstein–Hubbard model, which includes electron–electron interactions.[1]

ACKNOWLEDGMENTS

We thank Serge Aubry for suggesting this problem and making helpful comments. Part of the work was done during a visit of C.B. to the Institut des Hautes Etudes Scientifiques at Bures-sur-Yvette, January–April 1994, and we wish to express our thanks to them for their hospitality. The work was also supported by British Council Alliances between Warwick and the Laboratoire Léon Brillouin at Saclay and between Warwick and the Centre de Dynamique des Systèmes Complexes in Dijon, the U.K. Science and Engineering Research Council, and two E.C. network grants on Nonlinear Approach to Coherent and Fluctuating Phenomena in Condensed Matter and Optical Physics, and Stability and Universality in Classical Mechanics.

REFERENCES

1. S. Aubry, Bipolaronic charge density waves and polaronic spin density waves in the Holstein–Hubbard model, in Proceedings ECRYS-93 (Carry-le-Rouet, France, June 1993), *J. Phys. IV Coll. C2* (Paris) **3**:349–355 (1993).
2. S. Aubry, G. Abramovici, and J. L. Raimbault, Chaotic polaronic and bipolaronic states in the adiabatic Holstein model, *J. Stat. Phys.* **67**:675–780 (1992).
3. C. Baesens and R. S. MacKay, Improved proof of existence of chaotic polaronic and bipolaronic states for the adiabatic Holstein model and generalisations, *Nonlinearity* **7**:59–84 (1994).
4. V. Bach, and E. H. Lieb, and J. P. Solovej, Generalized Hartree–Fock theory and the Hubbard model, *J. Stat. Phys.* **76**:3–89 (1994).
5. R. Balescu, *Equilibrium and Non-equilibrium Statistical Mechanics* (Wiley, New York 1975).
6. T. Kennedy and E. H. Lieb, An itinerant electron model with crystalline or magnetic long range order, *Physica A* **138**:320–358 (1986).
7. J. L. Lebowitz and N. Macris, Low temperature phases of itinerant fermions interacting with classical phonons: the Static Holstein model, *J. Stat. Phys.* **76**:91–123 (1994).
8. R. S. MacKay and C. Baesens, A new paradigm in quantum chaos: Aubry's theory of equilibrium states for the adiabatic Holstein model, in *Quantum Chaos*, G. Casati, I. Guarneri, and U. Smilansky, eds. (North-Holland, Amsterdam, 1993), pp. 51–75.
9. M. Reed and B. Simon, *Functional Analysis* (Academic Press, New York, 1980).