A devil's staircase in the Zeeman response of the 1D half-filled adiabatic Holstein model

1994 J. Phys.: Condens. Matter 6 5891

(http://iopscience.iop.org/0953-8984/6/30/010)

J. Phys.: Condens. Matter 6 (1994) 5891–5918. Printed in the UK

A devil's staircase in the Zeeman response of the 1D half-filled adiabatic Holstein model

Christoph Kuhn and Serge Aubry

Laboratoire Léon Brillouin†, CE Saclay, 91191-Gif-sur-Yvette Cédex, France

Received 4 November 1993, in final form 17 March 1994

Abstract. The magnetization of the half-filled 1D adiabatic Holstein model in a magnetic field is investigated at 0 K. It has been proven elsewhere that at large enough electron–phonon coupling the ground state of this model with a large applied magnetic field is a mixed bipolaronic–polaronic configuration (at any band filling and at any dimension). In the half-filled case we propose as ground state for these polarons and bipolarons an ordered structure, which turns out to be equivalent to a periodic (or quasi-periodic) array of neutral solitons and which appears physically as spin density waves. We describe new improved numerical techniques for calculating these configurations and their energy.

It is predicted for large enough electron–phonon coupling k and confirmed numerically for $k > k_b \cong 1.4$ that the symmetry of the neutral soliton is broken and that there exist two kinds of neutral solitons (right and left).

The incommensurate arrays of neutral solitons exhibit a transition by breaking of analyticity at $k = k_c < 1.8$ (where k_c depends on the incommensurability ratio) with the same characteristic properties as for the incommensurate structures of several other models, especially of the prototype Frenkel–Kontorowa (FK) model. In the non-analytic regime we can determine the effective shape of the bipolaron and of the polaron, which takes into account their interactions.

The magnetization as a function of the magnetic field is found to vary as a devil's staircase, each plateau corresponding to a commensurate array of neutral solitons. The first plateau corresponds to a structure with no neutral soliton, while the magnetic threshold field required to cross its edge corresponds to the energy of this neutral soliton. In analogy with the FK model this devil's staircase is believed to be incomplete if there is an analytic incommensurate array of neutral solitons ($k < 1.8$), and complete otherwise.

The magnetic threshold field required to observe the beginning of these devil's staircases could be accessible in real charge density waves, if their electronic gap (which is found to be about four times the energy of the neutral soliton) does not range beyond 100 K. However, the value of this threshold should be reduced significantly by thermal fluctuations and by the existence of a Hubbard term, which both reduce the (free) energy for breaking the bipolarons into polarons and then favour the formation of arrays of neutral solitons.

1. Introduction

A polaron is a single electron localized in a potential well: induced by electron–phonon coupling and reinforced by the subsequent lattice deformation the polaron is formed self-consistently. A bipolaron is a pair of electrons with opposite spins localized self-consistently for the same reason. *A priori*, these concepts are only clear for a single polaron or a single bipolaron in the whole system. When there are many electrons in the system this definition becomes unclear, because the single-particle wavefunctions may not be localized in a single

† Laboratoire commun CEA-CNRS.

0953-8984/94/305891+28$19.50

potential well. Nevertheless, whether the electronic wavefunction is localized or not, a precise definition for the many bipolaron and polaron structures can be found.

The basic idea for defining these configurations is to follow them by uniform continuity from a limit of infinite electron–phonon coupling, where their existence is obvious at any density. At this limit, which is called 'anti-integrable' (Aubry 1994), the electrons are localized on arbitrary single sites and are thus strictly isolated from each other. There is a lattice distortion at the site, which is occupied by two electrons of opposite spin ('paired electrons'). By definition, this site is occupied by a bipolaron, which is not magnetic. When only one electron is present at the site ('unpaired electron'), the lattice distortion is weaker and, by definition, this site is occupied by a polaron, which is magnetic. The empty sites are holes.

Within the adiabatic Holstein model this anti-integrable limit is obtained by dropping the term corresponding to the electronic kinetic energy in the Hamiltonian. The persistence of arbitrary chaotic bipolaronic, polaronic or mixed polaronic–bipolaronic structures have been proven recently (Aubry *et al* 1992, Baesens and MacKay 1993, MacKay and Baesens 1993) for this model, when the electron kinetic energy is switched on up to a certain critical value or, equivalently, at large enough electron–phonon coupling. These results hold for the model on any lattice, at any finite dimension, at any electron density and with or without an applied magnetic field. These arbitrary distributions of bipolarons and (or) polarons (which can be periodic, chaotic or otherwise) depend continuously on the amplitude of this kinetic energy term and they remain distinct.

Because of this continuous dependence these states are still called bipolaronic, polaronic or mixed polaronic–bipolaronic configurations, even though the kinetic energy term no longer vanishes in the Hamiltonian. However, this result does not imply that each single-electron wavefunction remains localized, but concerns only the global electronic density function, where the electrons are indistinguishable. The electronic densities and the lattice distortions associated with the distribution of polarons and bipolarons do relax from their initial values, which means physically that the bipolarons and polarons do not remain strictly localized on single sites. The mathematical technique used for the proof of the existence of bipolaronic and polaronic structures can also be used numerically for finding these relaxed configurations explicitly from their anti-integrable limit (see appendix A).

In this regime of weak electronic kinetic energy or, equivalently, of strong electron–phonon coupling, the ground state of the adiabatic Holstein model has been proven to be one of the bipolaronic structures (i.e. a charge density wave (CDW)) obtained in the absence of magnetic field, and to become a mixed bipolaronic–polaronic structure (i.e. a charge density wave–spin density wave (CDW–SDW)) when the magnetic field becomes strong enough (Aubry *et al* 1992). However, no precise information concerning the ordering of these bipolarons and polarons and the nature of the phase transitions in a magnetic field is given by the theorems. It appears that the problem of finding the ordering of the bipolarons and polarons without and *a fortiori* with a magnetic field strongly depends on the details of the model and especially on its dimensionality. Finding the bipolaronic and polaronic ground states of coupled electron–phonon models remains an open problem, which until now has never been approached in the literature except in the simpler 1D Holstein model without a magnetic field (Le Daeron and Aubry 1983a, b, Aubry and Quemerais 1989).

We do not know yet any systematic numerical method for finding the ground states of such models, which we expect to become very complex in the general case. The aim of this paper is to study these mixed polaronic–bipolaronic ground state structures and their transitions on a specific example, which is the half-filled adiabatic Holstein model in one dimension and under a magnetic field. In this simpler case we found physical arguments for

conjecturing the precise configuration of the polarons and bipolarons in their ground state, thus allowing a numerical solution.

2. The adiabatic Holstein model

Let us recall the model (Holstein 1959) and give our notation. Its Hamiltonian is the sum of an elastic energy $\Phi_{\text{elast}}(\{u_i\})$, which depends on the lattice deformation $\{u_i\}$, and of an electronic Hamiltonian $H_{\text{el}}(\{u_i\})$ describing the motion of the electrons in the potential induced by this deformation. The quantum kinetic energy of the lattice is neglected†. The elastic energy of the dispersionless optical branch of N Einstein oscillators is

$$\Phi_{\text{elast}}(\{u_i\}) = \sum_{i=1}^{N} \tfrac{1}{2} m\omega_0^2 u_i^2 \tag{1a}$$

with the scalar variable u_i of site i, its frequency ω_0 and its mass m. In 1D models the magnetic field does not contribute directly to the Hamiltonian by orbital effects (except perhaps by slightly changing the model constants T and λ). It contributes mainly by the Zeeman term. Thus the tight-binding electronic Hamiltonian of the Holstein model in a magnetic field is

$$H_{\text{el}}(\{u_i\}) = -T\sum_{i,\sigma}(c^+_{i+1,\sigma}c_{i,\sigma} + c^+_{i,\sigma}c_{i+1,\sigma}) + \lambda\sum_i u_i(n_{i\uparrow}+n_{i\downarrow}) - \mu\sum_i(n_{i\uparrow}+n_{i\downarrow})$$
$$- \tfrac{1}{2}g\mu_B H\sum_i(n_{i\uparrow}-n_{i\downarrow}) \tag{1b}$$

with the standard notations: $c^+_{i,\sigma}$ and $c_{i,\sigma}$ are the creation and annihilation operators of an electron at site i with spin σ being up ($\sigma = \uparrow$) or down ($\sigma = \downarrow$). $2T$ is the half-bandwidth of the bare electrons, λ is the electron–phonon coupling and μ is the chemical potential, which fixes the total number of electrons in the band. H is the magnetic field, $g \cong 2$ is the Landé factor and the Bohr magneton is

$$\mu_B = e\hbar/2m_e \cong 9.274\ 10^{-24}\ \text{J T}^{-1} \cong 0.671\ \text{K T}^{-1}. \tag{1c}$$

To find the ground states of this model the minimization of the total energy can be performed in two steps. First, the atomic distortion $\{u_i\}$ and the number of electrons with spin up, $n_\uparrow$, and with spin down, $n_\downarrow$, are fixed. The total number of electrons, $n_\uparrow + n_\downarrow$, is thus given; in the half-filled case it is equal to the number of sites N. The electronic energy is minimal when the electrons occupy the lowest-energy states with eigenenergy $E_\nu(\{u_i\})$ defined by

$$-T\psi^\nu_{n+1} - T\psi^\nu_{n-1} + \lambda u_n\psi^\nu_n = E_\nu(\{u_i\})\psi^\nu_n. \tag{2a}$$

For convenience the indices $\nu = 1$ to N are chosen so the sequence $E_\nu(\{u_i\})$ is in increasing order. For a given configuration '$\{u_i\}$, $n_\uparrow$ and $n_\downarrow$' the energy of the ground state of the electronic Hamiltonian (1b) (with $H = \mu = 0$) is

$$\Phi_{\text{elect}}(\{u_i\}, n_\uparrow, n_\downarrow) = \sum_{\nu=1}^{n_\uparrow} E_\nu(\{u_i\}) + \sum_{\nu=1}^{n_\downarrow} E_\nu(\{u_i\}). \tag{2b}$$

† Note, however, that we have proven (Aubry 1991a) that in the situation of a Fröhlich conductivity with a strictly gapless phason mode, the Born–Oppenheimer approximation (and *a fortiori* the adiabatic approximation) for describing the CDWs is not valid. We conjecture that the effect of the quantum lattice fluctuations on the long-wavelength phasons could make the CDW become unstable. The adiabatic approximation, which is standard for the calculations of CDWs in quasi-1D systems, is valid only when the phason gap does not strictly vanish, that is, precisely in the bipolaronic and polaronic regime at large enough electron–phonon coupling.

The second step for finding the ground state is to search for the configuration '$\{u_i\}, n_\uparrow, n_\downarrow$' that yields the minimum value of the total energy

$$\Phi(\{u_i\}, n_\uparrow, n_\downarrow) = \Phi_{\text{elast}}(\{u_i\}) + \Phi_{\text{elect}}(\{u_i\}, n_\uparrow, n_\downarrow) - \mu(n_\uparrow + n_\downarrow) - \tfrac{1}{2}g\mu_B H(n_\uparrow - n_\downarrow). \quad (3)$$

It is convenient to define new parameters and dimensionless constants

$$u_i = u_i\sqrt{2T/m\omega_0^2} \qquad k = \lambda\sqrt{2/Tm\omega_0^2} \qquad h = \tfrac{1}{2}(g/2T)\mu_B H \qquad \mu = \mu/2T. \quad (4)$$

Then with $2T$ as the energy unit the initial Holstein Hamiltonian (1*b*) reduces to

$$H_{\text{red}} = \frac{1}{2}\sum_i u_i^2 - \frac{1}{2}\sum_{i,\sigma}(c_{i+1,\sigma}^+ c_{i,\sigma} + c_{i,\sigma}^+ c_{i+1,\sigma}) + \frac{k}{2}\sum_i u_i(n_{i\uparrow} + n_{i\downarrow}) - \sum_i(\mu(n_{i\uparrow} + n_{i\downarrow}) + h(n_{i\uparrow} - n_{i\downarrow})) \quad (5a)$$

and the total energy $F(\{u_n\}) = \Phi(\{u_i\})/2T$ becomes

$$F(\{u_i\}, n_\uparrow, n_\downarrow) = \frac{1}{2}\sum_i u_i^2 + \frac{1}{2}\sum_{\nu=1}^{n_\uparrow} E_\nu(\{u_i\}) + \frac{1}{2}\sum_{\nu=1}^{n_\downarrow} E_\nu(\{u_i\}) - \mu(n_\uparrow + n_\downarrow) - h(n_\uparrow - n_\downarrow) \quad (5b)$$

where

$$-\psi_{n+1}^\nu - \psi_{n-1}^\nu + ku_n\psi_n^\nu = E_\nu(\{u_i\})\psi_n^\nu. \quad (5c)$$

The dimensionless parameter k characterizes the strength of the electron–phonon coupling†. The extrema of this form (5*b*) fulfil

$$\partial F(\{u_i\}, n_\uparrow, n_\downarrow)/\partial u_n = u_n + (k/2)\rho_n = 0 \quad (6a)$$

where the local density of electrons $\rho_n = \rho_{n_\uparrow} + \rho_{n_\downarrow}$ is the sum of the density of the electrons

$$\rho_{n_\uparrow} = \sum_{\nu=1}^{n_\uparrow} |\psi_n^\nu|^2 \qquad \rho_{n_\downarrow} = \sum_{\nu=1}^{n_\downarrow} |\psi_n^\nu|^2 \quad (6b)$$

with spin $\uparrow$ and $\downarrow$ respectively.

At large enough electron–phonon coupling the theorems proven by Aubry *et al* (1992) and Baesens and MacKay (1993) assert that this form has infinitely many extrema, which correspond to polaronic and bipolaronic metastable configurations. Each of these configurations is obtained by relaxation of an initial configuration $\{\sigma_i\}$, where the electrons are localized on single sites. By definition, $\sigma_i = 0$ if there is no electron at site i, $\sigma_i = \frac{1}{2}$ if site i is occupied by a polaron and $\sigma_i = 1$ if it is occupied by a bipolaron.

For an arbitrary distribution of polarons and bipolarons, $\{\sigma_i\}$, we fix for example as $\uparrow$ the up direction of the free spin of the polarons ($\sigma_i = \frac{1}{2}$). Then $\{\sigma_i\}$ also determine the number

† Aubry *et al* (1992) give the Holstein model in its full quantum form, but for mathematical convenience its Hamiltonian was transformed differently in order to have a small dimensionless coefficient $t = 1/k^2$ for the electronic kinetic energy instead of a large dimensionless electron–phonon coupling constant k. In that model, the phonons were represented by boson operators with the electron–phonon coupling constant $g = \lambda\sqrt{\hbar/2m\omega_0}$. The unit of energy was chosen to be $8g^2/\hbar\omega_0 = 4\lambda^2/m\omega_0^2$ and the deformation field $u_i = (4g/\hbar\omega_0)(a_i^+ + a_i)$ or equivalently $u_i = (2\lambda/m\omega_0^2)u_i$. The adiabatic Holstein model without magnetic field then reduces from (5*a*) to $\frac{1}{2}\sum_i[u_i^2 + u_i(n_{i\uparrow} + n_{i\downarrow}) - (t/2)\sum_{i,\sigma}(c_{i+1,\sigma}^+ c_{i,\sigma} + c_{i,\sigma}^+ c_{i+1,\sigma})$ showing that the regime of large electron–phonon coupling is also the regime of small electronic kinetic energy.

of electrons $n_\uparrow$ and $n_\downarrow$ with spin $\uparrow$ and $\downarrow$ respectively. The corresponding configuration of the adiabatic Holstein model is obtained at finite k by relaxation of the total energy (*5b*) from the initial configuration $\{u_i\} = -k\{\sigma_i\}$ (the bipolaronic or polaronic configuration at the limit $k = \infty$). This can be achieved in principle by following numerically the gradient flow of the potential energy given by (6) from this initial state. This method was used in early work (Le Daeron and Aubry 1983a, b), however, more efficient techniques can be implemented for faster numerical convergence (see appendix A). The improved theorems (Baesens and MacKay 1993) yield for the 1D Holstein model the following exact bounds, which determine minimum domains of existence:

for all bipolaronic configurations ($\sigma_i = 0$ or 1)	$k > 2.3463$
for all polaronic configurations ($\sigma_i = 0$ or $\frac{1}{2}$)	$k > 3.3181$
for all mixed polaronic–bipolaronic configurations ($\sigma_i = 0$, $\frac{1}{2}$ or 1)	$k > 3.682$.

When these conditions are fulfilled, two different initial states necessarily yield two different final metastable states $\{u_i\}$. When k becomes smaller, many local minima start to disappear by merging with one or more extrema of the variational form according to mathematical rules (Morse theory). In this many-variable variational form a complex and infinite cascade of bifurcations follows; these are still unexplored. These bifurcations start for small enough k and then the topological entropy of the whole set of metastable bipolaronic or polaronic states decreases, but does not vanish until the electron–phonon coupling constant k becomes strictly zero. This behaviour is similar to that observed in the Frenkel–Kontorowa (1938–1939) (FK) model (Aubry and Quemerais 1989). There are still many bipolaronic and polaronic configurations, which survive even for k very small! A partial extension of these theorems to the Holstein–Hubbard model has been recently obtained (Aubry 1993, 1994). It has been proven that the formation of bipolaronic and of mixed polaronic–bipolaronic structures at large electron–phonon coupling still holds when the electron–electron repulsion is not too large, but that there are only purely polaronic structures when this repulsion exceeds some critical value.

Furthermore, for the original adiabatic Holstein model theorems proven by Aubry *et al* (1992), at large k the ground state is purely bipolaronic (bipolaronic CDW) in the absence of the magnetic field and it becomes a mixed polaronic–bipolaronic structure (polaronic–bipolaronic CDW–SDW) for a strong enough magnetic field. Therefore the problem of finding the ground state of such models reduces to the problem of finding the corresponding configuration of pseudo-spins $\{\sigma_i\}$.

3. Bipolaronic ground states in the 1D Holstein model without a magnetic field

At small electron–phonon coupling k and in the absence of a magnetic field the ground state of this 1D model is expected to become a Peierls CDW, where $n_\uparrow = n_\downarrow$ (Peierls 1955). The wave vector of the periodic lattice distortion (PLD) associated with the CDW is twice the Fermi wave vector $2q_F = \zeta Q$ (where $\zeta = n_\uparrow/N = n_\downarrow/N$ is the band filling and Q the wave vector of the reciprocal lattice). The half-filled case, $\zeta = \frac{1}{2}$, has an exact solution corresponding to a dimerized chain (see Feinberg and Ranninger 1983), whilst the standard calculation of the CDW–PLD is only approximate for the incommensurate band filling (except in exceptional modified models).

In this incommensurate case, the transition to a bipolaronic ordering has been numerically observed by increasing the electron–phonon coupling k, as a second-order phase transition called transition by breaking of analyticity (TBA) (Le Daeron and Aubry

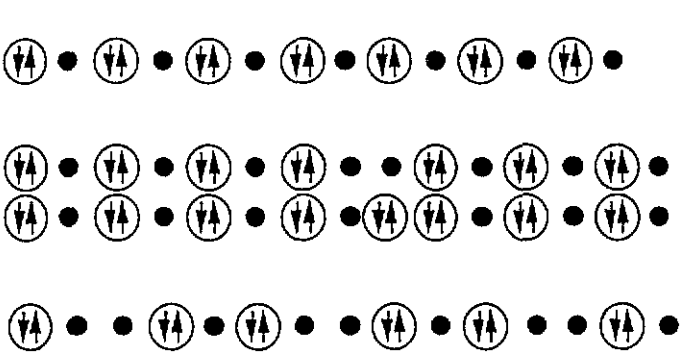

Figure 1. (*a*) A sequence of bipolarons and holes defined by (7) for the ground state at $\zeta = \frac{1}{2}$. Black dots are holes and circles with opposite arrows are bipolarons. (*b*) As (*a*) for the retarded discommensuration ($\zeta \to \frac{1}{2}$ with $\zeta < \frac{1}{2}$) and the advance discommensuration ($\zeta \to \frac{1}{2}$ with $\zeta > \frac{1}{2}$). (*c*) As (*a*) for $\zeta = \frac{2}{5}$. This structure can be also interpreted as a periodic sequence of retarded discommensurations as shown in (*b*).

1983a, b, Aubry and Quemerais 1989). For example the TBA was observed at $k \cong 1.58$ for $\zeta = (3 - \sqrt{5})/2$. This transition is physically characterized by the lattice pinning of the CDW and a series of associated physical features. On the basis of early numerical observations it is conjectured that the ground state associated with the bipolaronic CDW keeps the same Peierls wave vector $2q_F$ at all electron–phonon couplings. Thus finding the ground state of the adiabatic Holstein model in the absence of a magnetic field is equivalent to finding the pseudo-spin distribution $\{\sigma_i = 0 \text{ or } 1\}$ of the bipolaronic structure. This sequence is found to be identical to the sequence of pseudo-spins that describes the sequence of discommensurations in the ground state of the FK model (Vallet *et al* 1988). We interpret this feature by saying that (within this specific model) the bipolarons interact by repulsive forces, which are convex at all distances and which tend to give them an equidistant distribution†. This distribution of bipolarons is found by starting from an equidistant distribution $(n - \alpha)/\zeta$ (where $1/\zeta$ is the average distance and $0 \leqslant \alpha < 1$ some arbitrary phase) and by moving each bipolaron to the closest integer site. Then the nth bipolaron is located at site $i_n = \text{Int}((n - \alpha)/\zeta + \frac{1}{2})$ (Int(x) is the largest integer smaller than or equal to x), which yields equivalently

$$\sigma_i = \chi_\zeta(i\zeta + \alpha) \tag{7a}$$

where $\chi_\zeta(x)$ is a one-periodic function defined as

$$\chi_\zeta(x) = \text{Int}(x + \zeta/2) - \text{Int}(x - \zeta/2). \tag{7b}$$

This function $\chi_\zeta(x)$ is called the hull function of $\{\sigma_i\}$. Using the relaxation technique described in appendix A we claim to obtain the ground state of the adiabatic Holstein model without a magnetic field and with band filling ζ by starting from configuration $\{u_i = -k\sigma_i\}$, where $\{\sigma_i\}$ is given by (7*a*). Schemes of bipolaronic distributions given by (7) are represented in figure 1: the commensurate ground state for $\zeta = \frac{1}{2}$ (figure 1(*a*)), the retarded and the advanced discommensuration of this commensurate ground state (figure 1(*b*)) and the commensurate ground state for $\zeta = \frac{2}{5}$ (figure 1(*c*)).

† This result has not been proven rigorously and is restricted to the original Holstein model where the phonons are dispersionless. When a phonon dispersion (even small) is present, the elastic interaction between the bipolarons introduces a non-convex part which is responsible for phase separations between phases with different electronic densities when the electron–phonon coupling becomes large enough. This transition is indeed numerically observed. Note however that in physical situations, the long-range Coulomb interactions between the bipolarons prevent the formation of these phase separations but produce new incommensurate phases with wave vectors that are not $2q_F$ (Raimbault and Aubry 1994).

A retarded discommensuration is obtained by taking the limit of the sequence (7*a*) by lower values at $\zeta = \frac{1}{2} - 0$ with the phase $\alpha(\text{mod } 1) = \pm\frac{1}{4}$. The advanced discommensuration is obtained identically by taking the limit by upper values at $\zeta = \frac{1}{2} + 0$ with the phase $\alpha(\text{mod } 1) = \pm\frac{1}{4}$. All these sequences can be built by blocks following a sequence of inflation rules determined by the continued fraction expansion of ζ (Vallet *et al* 1988), which were initially involved in understanding the thermodynamical behaviour of the incommensurate structures in the FK model, but can be used here too.

For rational ζ the numerical studies for the half-filled 1D adiabatic Holstein model carried out elsewhere allow one to conjecture, in the case of this specific model only, that the ground state configuration is a continuous function of k and never bifurcates when k ranges from zero to infinity: the characteristic feature to be defectible is preserved. The discommensurations indeed correspond to defects in the bipolaronic ordering. The ground state can also accept chaotic distributions of discommensurations because of the intrinsic lattice pinning. Consequently, we can say by continuity that the ground state remains always a bipolaronic structure even for small electron–phonon coupling.

When ζ is irrational we observe the TBA at $k = k_c(\zeta)$. The value of $k_c(\zeta)$ is rather small and around 1.5 (1.58 at most). Since the amplitude of the CDW exponentially decreases and becomes negligible for $k < 1.4$, the regime of the standard Peierls–Fröhlich analytic CDW is restricted in rather a narrow domain. This TBA can be viewed in some extended sense as a bifurcation of infinite order. At this transition the ground state becomes undefectible (see e.g. Aubry and Quemerais 1989), which is characterized by the fact that the size of its defects diverges. These defects can be classified hierarchically as the discommensurations in the incommensurate pseudo-spin sequence (7) as described by Vallet *et al* (1988). Then we can say that all these defects of the ground state merge with the ground state itself, when the analyticity of the ground state is restored. Despite this TBA the sequence (7) still yields to the ground state by relaxation, even when the CDW is analytic: at this irrational electronic concentration the variational energy (5) has a unique minimum, which is the CDW (we call this state undefectible).

4. Mixed polaronic–bipolaronic ground states in the 1D Holstein model under a magnetic field: the half-filled case

The Zeeman response of a magnetic field applied to a metal makes the number $n_\uparrow = \zeta_\uparrow N$ of electrons with spin $\uparrow$ different from the number $n_\downarrow = \zeta_\downarrow N$ of electrons with spin $\downarrow$. This magnetization corresponds to the Pauli susceptibility. In the case of a 1D metal under a magnetic field coupled to a deformable lattice, the Peierls theory easily extends (figure 2). Within standard perturbation theory one should expect two superimposed periodic lattice distortions with wave vectors $2q_{F\uparrow} = \zeta_\uparrow Q$ and $2q_{F\downarrow} = \zeta_\downarrow Q$. The first one corresponds to a CDW for electrons with spin $\uparrow$ and the second one corresponds to a CDW for electrons with a spin $\downarrow$. Since the wave vectors of the modulations are different, the system develops in addition an SDW. This perturbative argument is approximately valid at small coupling, when these two waves are weakly interacting (see Brazovskii *et al* 1981, Brazovskii and Matveenko 1984) but becomes surely wrong at large coupling and when the wave vectors $Q\zeta_\uparrow$, $Q\zeta_\downarrow$ and Q have no commensurability relations (i.e. do not fulfil any relation $n_1\zeta_\uparrow + n_2\zeta_\downarrow + n_3 = 0$, where n_1, n_2 and n_3 are integers). Moreover, the interaction between the two modulations could even break the strict quasi-periodicity of the system. Indeed, when three incommensurate periods are competing, we have examples in simpler, but different models (Aubry *et al* 1987, Aubry 1989, 1991b), where a 'weakly periodic'

structure becomes the ground state. The Fourier spectrum of such a structure is no longer a sum of Dirac peaks as for a quasi-periodic structure but a 'singular continuous' measure with scaling properties. Unfortunately we have not yet found any tool applicable to the Holstein model for finding the ground state and its properties in the general case.

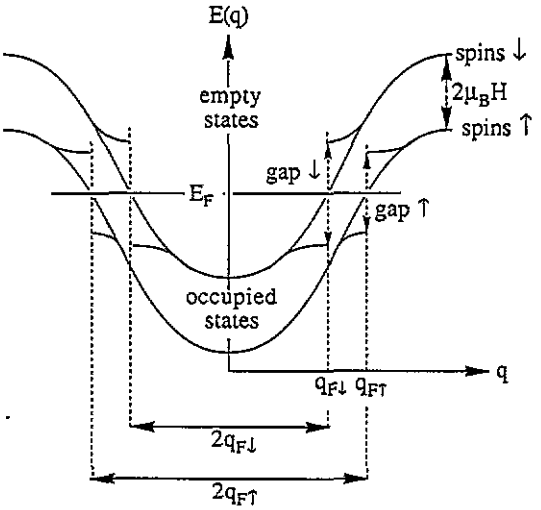

Figure 2. When a magnetic field is applied (in the direction corresponding to spin ↑) on a 1D conductor, the Peierls instability occurs simultaneously and opens two electronic gaps. The first one is at wave vectors $\pm q_{F\uparrow}$ for electrons with spin ↑ and the second one at $\pm q_{F\downarrow}$ for electrons with spin ↓. When the electron–phonon coupling is small, the resulting CDW is expected to be well approximated by a linear superposition of these two periodic CDWs with wave vectors $2q_{F\uparrow}$ and $2q_{F\downarrow}$.

In the half-filled case corresponding to the condition $\zeta_\uparrow + \zeta_\downarrow = 1$ we have a commensurability relation and thus only two independent wave vectors. On the basis of known results obtained mostly with the FK model this situation is expected to be simpler and to yield usually incommensurate structures. The wave vectors $Q\zeta_\uparrow$ and $Q\zeta_\downarrow = Q - Q\zeta_\uparrow$ in fact correspond to the same period for both modulations. Then we expect that the phases of the CDWs for spins ↑ and ↓ will lock on to each other, as this usually happens when incommensurate structures become commensurate. According to the theorems valid at large enough coupling the ground state has to become a mixed polaronic–bipolaronic structure described by a sequence of pseudo-spins $\{\sigma_i = 0, \frac{1}{2} \text{ or } 1\}$. We can write

$$\sigma_i = \tfrac{1}{2}(\sigma_{i\uparrow} + \sigma_{i\downarrow}) \tag{8}$$

where for $k = \infty$ $\sigma_{i\uparrow}$ and $\sigma_{i\downarrow}$ take the two values, 0 or 1. By definition, $\sigma_{i\uparrow} = 1$ if there is an electron with spin ↑ at site i for $k = \infty$ and $\sigma_{i\uparrow} = 0$ in the opposite case. $\sigma_{i\downarrow}$ is defined identically for the electrons with spin ↓.

To find the ground state of the half-filled 1D adiabatic Holstein model under a magnetic field we first study the ground state $\{u_i\}$ for a given $\zeta_\uparrow$, which has the energy per site

$$\Psi_0(\zeta_\uparrow) = \lim_{N\to\infty} \frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{2}u_i^2 + \frac{1}{2}\sum_{\nu=1}^{n_\uparrow} E_\nu(\{u_n\}) + \frac{1}{2}\sum_{\nu=1}^{n_\downarrow} E_\nu(\{u_n\}) \right) \tag{9a}$$

and next we determine $\zeta_\uparrow$ by minimization of the energy

$$\Psi(\zeta_\uparrow) = \Psi_0(\zeta_\uparrow) - h(2\zeta_\uparrow - 1) - \mu \tag{9b}$$

as a function of the magnetic field. To fix the ideas we assume that the magnetic field is in direction $\uparrow$, which implies $n_\uparrow > n_\downarrow$ or equivalently $\zeta_\uparrow \geqslant 1/2$. We propose as conjectures two simple statements, which are compatible with the above Peierls prediction valid at small k and extrapolate the numerical results at zero magnetic field:

(i) for given electronic densities $\zeta_\uparrow$ and $\zeta_\downarrow$ the ground state configuration $\{u_i\}$ depends continuously on k and

(ii) the sequences $\{\sigma_{i\uparrow}\}$ and $\{\sigma_{i\downarrow}\}$, which determine the mixed polaronic–bipolaronic structure at large k by equation (8), tend to yield independently two equidistant distributions, which are given by the same law as in (7*a*)

$$\sigma_{i\uparrow} = \chi_{\zeta_\uparrow}(i\zeta_\uparrow + \alpha_\uparrow) \tag{10a}$$

$$\sigma_{i\downarrow} = \chi_{\zeta_\downarrow}(i\zeta_\downarrow + \alpha_\downarrow). \tag{10b}$$

Since $\zeta_\uparrow + \zeta_\downarrow = 1$, these two functions $\chi_{\zeta_\uparrow}(x)$ and $\chi_{\zeta_\downarrow}(x)$ have in fact the same period and moreover we have

$$\chi_{\zeta_\downarrow}(x) = 1 - \chi_{\zeta_\uparrow}(x + \tfrac{1}{2}). \tag{10c}$$

The two phases $\alpha_\uparrow$ and $\alpha_\downarrow$ are not arbitrary, but lock on to each other since $\zeta_\uparrow + \zeta_\downarrow = 1$. We found by numerical tests that $\alpha_\uparrow = -\alpha_\downarrow$ yields the configuration with the smallest energy and that corresponds to a symmetrical hull function $\chi^*(x)$ for $\sigma_i = 0$, $\frac{1}{2}$ or 1 (see figure 3). This condition tends first to maximize the number of bipolarons (which occupy the sites i where $\sigma_{i\uparrow} = \sigma_{i\downarrow} = 1$) and second to form these bipolarons at the sites with phase $i\zeta_\uparrow + \alpha_\uparrow$ (mod 1) as close as possible to zero. In summary we assume that the pseudo-spin sequence $\{\sigma_i\}$

$$\sigma_i = \tfrac{1}{2}(1 + \chi_{\zeta_\uparrow}(i\zeta_\uparrow + \alpha_\uparrow) - \chi_{\zeta_\uparrow}(i\zeta_\uparrow + \alpha_\uparrow + \tfrac{1}{2})) = \chi^*_\zeta(i\zeta_\uparrow + \alpha_\uparrow) = 0, \tfrac{1}{2} \text{ or } 1 \tag{11}$$

yields the distribution of polarons and bipolarons corresponding to the ground state.

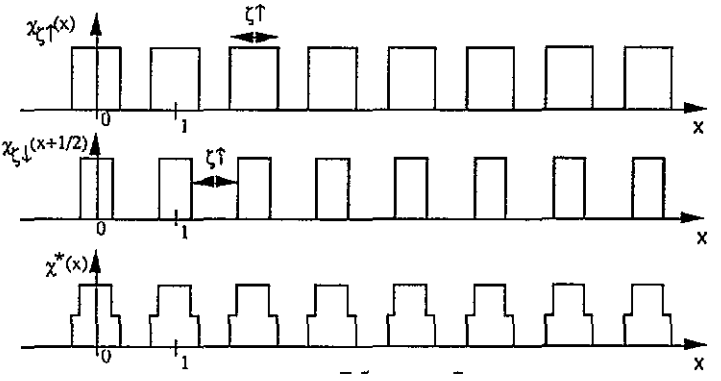

Figure 3. Hull functions $\chi_{\zeta_\uparrow}(x)$, $\chi_{\zeta_\downarrow}(x)$ and $\chi^*(x)$ of the pseudo-spin sequences $\sigma_{i\uparrow}$, $\sigma_{i\downarrow}$ and σ_i.

We have not proved rigorously that the ground state is obtained by relaxation of this mixed polaronic–bipolaronic structure. However, no contradiction will appear in our results. This can be considered as a good argument for having the true ground state of our system for half band filling.

Note, by contrast, that in the general case when $\zeta_\uparrow$, $\zeta_\downarrow$ and 1 do not fulfil any commensurability relation, the sequence of pseudo-spins $\{\sigma_{i\uparrow}\}$ and $\{\sigma_{i\downarrow}\}$ cannot be described by the form (10*a*) and (10*b*) for any square function. If such a form were valid, the ground state would contain polarons with opposite spins, but correctly the spins of all the polarons (if any) have to be aligned in the direction of the magnetic field. This remark invalidates, in the general case, the assumption suggested for continuum models (Brazovskii *et al* 1981, Brazovskii and Matveenko 1984) that the ground state can be well described by the superposition of two modulations with different wave vector for the electrons with spin $\uparrow$ and $\downarrow$ respectively.

Now we analyse first the limit configuration obtained from (11) when $\zeta_\uparrow \to \frac{1}{2}$, which is a bipolaronic configuration with only one polaron. This defect is the so-called neutral soliton.

5. Ground state with a single polaron: the neutral soliton

Neutral solitons carrying a spin were first discovered within the SSH model (Su *et al* 1979, 1980), a 1D model with a half-filled band (different from the Holstein model) initially proposed for describing the quantum chemical properties of polyacthylene (a long hydrocarbon molecule with bond length alternation and an electronic gap). The neutral soliton is, by definition, a defect in the bond length alternation, which appears topologically in a large ring with $2N+1$ sites and $2N+1$ electrons ($N \to \infty$). It has no charge because the density of electrons is strictly preserved: there is exactly one electron per site, but it must exhibit a free spin because the number of electrons is odd. A neutral soliton is a topological defect, which means it is the ground state of the model under appropriate boundary conditions. The word soliton implies *a priori* the dynamical properties of a travelling solitary wave, which this defect does not have necessarily, since it might be pinned to the lattice. We shall keep however this standard terminology.

A neutral soliton also exists in the half-filled 1D adiabatic Holstein model for the same reasons. As a consequence of the theorem on the ground states of the adiabatic Holstein model (Aubry *et al* 1992) it must become a mixed polaronic–bipolaronic structure at large enough k. Since it is a ground state of the system (without magnetic field) with an equal and odd number of sites and electrons, this mixed structure just exhibits a single polaron. It clearly appears that there is no symmetric sequence $\{\sigma_i\}$ on the closed ring of length $2N+1$, which could determine the mixed polaronic–bipolaronic distribution and, consequently, because of this symmetry breaking there are at least two neutral solitons in our model when k is large enough. This symmetry breaking is a consequence of the discreteness of the lattice. In the original model, the continuum approximation was used and therefore no breaking of symmetry was expected (Su *et al* 1979, 1980, 1981).

The sequence $\{\sigma_i\}$, which determines the sequence of bipolarons and polarons of the neutral soliton, can be obtained from (11), when $\zeta_\uparrow \to \frac{1}{2}$ (by upper values). The two determinations are obtained by choosing the phase either $\alpha_\uparrow \bmod 1 = +\frac{1}{4}$ or $\alpha_\uparrow \bmod 1 = -\frac{1}{4}$ (see figure 4(*a*) and (*b*)). Both can be viewed as the superposition of an advanced discommensuration for the electrons with spin $\uparrow$ and a retarded discommensuration for the electrons with spin $\downarrow$ (compare with figure 1(*b*)), which determines a sequence of bipolarons and holes with just a single polaron. The two possible determinations are obtained by changing the relative phase of the two discommensurations. (If $\zeta_\uparrow \to \frac{1}{2}$ by lower value, one obtains the same neutral solitons but with an opposite spin.)

We checked numerically that the symmetry breaking of the neutral soliton does exist at large k according to our prediction. At small k the neutral soliton is found to remain

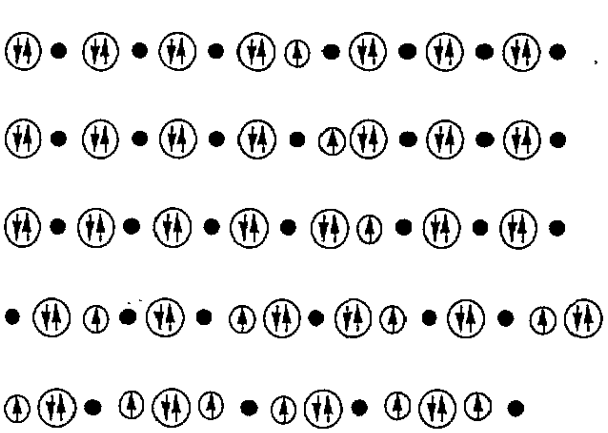

Figure 4. (*a*) A sequence (defined by equation (11)) of bipolarons, polarons and holes for the ground state in the limit $\zeta_\uparrow \to \frac{1}{2}+$, which characterizes the left neutral soliton. (*b*) As (*a*) for the right neutral soliton. (*c*) As (*a*) for the left neutral soliton shifted by one unit cell. (*d*) A sequence (defined by equation (11)) of bipolarons, and polarons and holes for the ground state at $\zeta_\uparrow = \frac{5}{8}$ and $\zeta_\downarrow = \frac{3}{8}$ (two unit cells). This can be viewed as an alternate sequence of right and left neutral solitons. (*e*) A sequence (defined by equation (11)) of bipolarons, polarons and holes for the ground state at $\zeta_\uparrow = \frac{5}{7}$ and $\zeta_\downarrow = \frac{2}{7}$ (two unit cells).

asymmetric although it tends to become almost symmetric because its width diverges. We take care to confirm this asymmetry by detailed numerical analysis of the shape of the Peierls–Nabarro energy barrier† of the neutral soliton.

For this purpose we have to determine a continuous path in the space of configurations that translates the neutral soliton by one unit cell and that overcomes the smallest energy barrier. For k large enough the translation of a neutral soliton by one unit cell has to be performed in two steps, since there are two equivalent neutral solitons. We have to find first a path of configurations connecting the left neutral soliton (figure 4(*a*)) with the right neutral soliton (figure 4(*b*)), which yields a minimax configuration associated with the maximum of the energy along this path. We have to find next a second minimax path connecting this right neutral soliton (figure 4(*b*)) to the left neutral soliton shifted by one unit cell (figure 4(*c*)). Since our half-filled model possesses electron–hole symmetry, which consists in replacing the bipolarons by holes and the holes by bipolarons while the polarons are kept identical at the same sites, we note that by this transformation figures 4(*b*) and (*c*) becomes equivalent to figures 4(*a*) and (*b*) respectively. Consequently these two minimaxes correspond by the electron–hole symmetry and their heights in energy are equal.

Thus it is sufficient to study only a minimax path between the left and right neutral solitons. For this purpose we first calculate the configuration $\{u_i^\downarrow\}$ (with $-N \leqslant i \leqslant N$) of the left neutral soliton and its energy for k not too small ($k > 1.5$) and for a large system of $2N+1$ sites and $2N+1$ electrons ($N \simeq 50$) with cyclic boundary conditions ($u_{i+2N+1} = u_i$). We use the technique described in appendix A taking as initial state the

† Let us consider the family of continuous paths $\{x_n(\xi)\}$ with $0 \leqslant \xi \leqslant 1$, which connects two equal local minima of the variational energy (5) at $\{x_n(0)\} = \{u_n\}$ and $\{x_n(1)\} = \{v_n\}$. For each path $\{x_n(\xi)\}$ we consider the maximum energy variation reached along this path. The minimax, $\min_{\{u_i(\xi)\}} \max_{0\leqslant\xi\leqslant 1} F(\{x_n(\xi)\}, n_\uparrow, n_\downarrow) - F(\{x_n(0)\}, n_\uparrow, n_\downarrow)$, is defined as the minimum over the whole family of continuous paths $\{u_i(\xi)\}$ of this maximum. The minimax configuration is the configuration $\{y_i\}$ that yields this minimax value. The Peierls–Nabarro energy barrier is defined as the maximum of the energy variation along a path that realizes the minimax. This path is not unique, but its maximum remains the same.

mixed bipolaronic–polaronic sequence shown in figure 4(a). To fix these ideas, we set the polaron location at site 0 ($\sigma_0 = \frac{1}{2}$). The right neutral soliton $\{u_i^r\}$ with the sequence of figure 4(b) is obtained simply by the symmetry relation $u_i^r = u_{1-i}^l$. It is convenient to define the coordinate $u_0(\xi) = u_0^l + \xi(u_0^r - u_0^l)$ of atom 0 and to parametrize the continuous path of configurations $\{u_i(\xi)\}$ connecting $\{u_i^l\}$ and $\{u_i^r\}$. For each $0 < \xi < 1$ we calculate a neutral soliton $\{u_i(\xi)\}$, which minimizes the variational energy (5) with the same initial state as the left neutral soliton but with the constraint of atom 0 fixed at $u_0(\xi)$. When ξ varies from zero to unity, we find numerically that there is a point ξ_s where the configuration of the neutral soliton under constraint becomes symmetric, $u_i(\xi_s) = u_{1-i}(\xi_s)$, for each i and that $\{u_i(\xi)\}$ appears as a continuous function of ξ for $0 < \xi < \xi_s$. However, it becomes discontinuous for $\xi_s \leqslant \xi \leqslant 1$. Then it is more convenient to define the coordinates $u_1(\xi) = u_1^r + (2\xi_s - \xi)/\xi_s(u_1(\xi_s) - u_1^r)$ of atom 1 for $\xi_s < \xi < 2\xi_s$ in order to complete continuously and symmetrically this path of configurations: for each $\xi_s < \xi < 2\xi_s$ we calculate a neutral soliton $\{u_i(\xi)\}$, which minimizes the variational energy (5) with the constraint of atom 0 fixed at $u_1(\xi)$. Since $\{u_i(\xi)\}$ minimizes the energy of the system with only one constraint, this continuous path of configurations must reach the minimax configuration at its maximum of energy. The curve corresponding to the energy variation of the configuration $\{u_i(\xi)\}$ as a function of ξ is the shape of the Peierls–Nabarro barrier (minimax) between the left and right neutral solitons (figure 5(a)). This continuous path, which reaches the minimax at its maximum, is not unique but this choice is convenient. It has a single maximum, which corresponds to the symmetric neutral soliton and which is the minimax configuration. It appears that, for all values of k that can be investigated, the shape of this Peierls–Nabarro barrier does not change significantly when k varies apart from the change of energy scale, and the minimum of the energy always corresponds to the asymmetric neutral soliton while the minimax corresponds to the symmetric neutral soliton. The height of the Peierls–Nabarro barrier is found to vanish very fast for small k (figure 5(b)).

It has been noted that the SSH model also yields in the discrete case two kinds of neutral solitons, which were called strong and weak (Bryant and Glick 1982). Although the SSH model does not possess a well defined anti-integrable limit, because the model becomes unstable at large electron–phonon coupling, in a modified (and more physical) SSH model that accepts large electron–phonon coupling, the existence of an anti-integrable limit is suggested by the existence of a TBA (Le Daëron and Aubry 1983b) observed numerically. This is backed up by the standard chemical representation of configurations of such models as a sequence of double bonds and single bonds (Kuhn 1989). Such a representation appears to be equivalent to the pseudo-spin representation $\{\sigma_i = 0 \text{ or } 1\}$, which is proven to be valid in our present model in the absence of a magnetic field. However, we suggest that the observation of Bryant and Glick (1982) is similar to our observation except that the minima and the minimax appear to be exchanged in the SSH model. The symmetrical neutral solitons (called strong and weak) correspond to the minima of energy while the asymmetric neutral solitons (not considered by Bryant and Glick) should become the minimax. We also know similar situations in some extended FK models where the minimax and minima of discommensuration can be exchanged.

The energies of an asymmetric neutral soliton, of an advanced (or retarded) discommensuration and of the electronic gap for the dimerized ground state without defects are compared in figure 6. (To calculate these energies, we choose $\mu = -k/2\langle u_i \rangle = -k^2/4$ in (5a) to have electron–hole symmetry.) According to the predictions of continuous models, it is found that for small k ($k < 1.4$) the energies of the neutral soliton and of the discommensuration are both equal to a quarter of the electronic gap. For larger k the

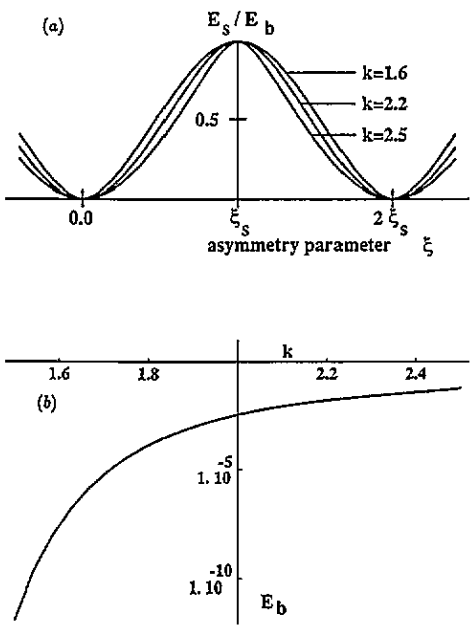

Figure 5. (*a*) The shape of the Peierls–Nabarro energy barrier along the path $u_i(\xi)$ of the neutral soliton. (*b*) The Peierls–Nabarro barrier of the neutral soliton versus k (log plot, energy units $2T$).

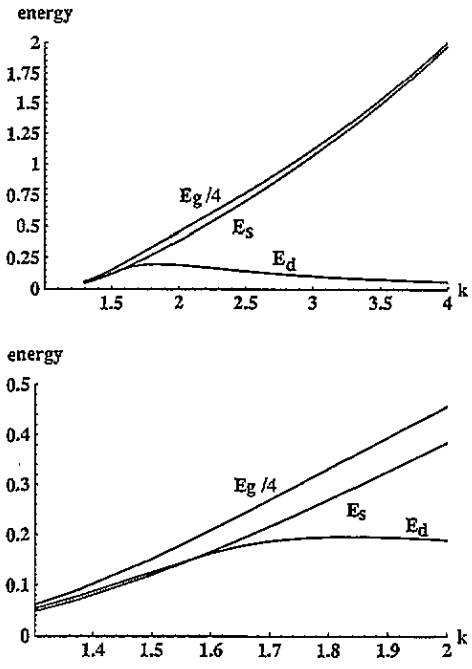

Figure 6. The energy versus k of the asymmetric neutral soliton E_s, the discommensuration E_d and of the electronic gap E_g for the half-filled 1D adiabatic Holstein model (energy unit $2T$).

effect of the lattice discreteness becomes obvious. The bipolarons tend to become localized on single sites. Their interaction decays to zero and thus the discommensuration energy, that is, the energy difference of the configurations described by figure 1(*a*) and (*b*), decays to zero. For larger k the energy of a neutral soliton is approximately the energy required to break a bipolaron into two polarons, which behaves as $k^2/8$. It is a quarter of the electronic gap, which behaves as $k^2/2$.

6. Quasi-periodic arrays of neutral solitons: the transition by breaking of analyticity

The mixed polaronic–bipolaronic ground state configurations with pseudo-spin sequences $\{\sigma_i\}$ generated by equation (11) can be viewed as periodic or quasi-periodic arrays of neutral solitons with density per site $\zeta_\uparrow - \zeta_\downarrow$. Figures 4(*d*) and (*e*) shows as examples the sequences of polarons and bipolarons obtained for $\zeta_\uparrow = \frac{5}{8}$ ($\zeta_\downarrow = \frac{3}{8}$) and for $\zeta_\uparrow = \frac{5}{7}$ ($\zeta_\downarrow = \frac{2}{7}$) respectively. When $\zeta_\uparrow$ becomes an irrational number, we can reasonably expect a TBA similar to the TBA proven to exist for an array of discommensurations within the FK model. For $\zeta_\uparrow$ irrational the lattice distortion $\{u_i\} = -k/2\{\rho_i\}$ or, equivalently, the electronic density can be described by a hull function

$$\rho_i = \rho_{i\uparrow} + \rho_{i\downarrow} = g(i\zeta_\uparrow + \alpha_\uparrow) \tag{12a}$$

which has the same period as the hull function of the pseudo-spin sequence (11). It is convenient to define separately two hull functions $g_\uparrow(x)$ and $g_\downarrow(x)$ for the electronic

densities (6*b*) of electrons with spins ↑ and ↓ respectively

$$\rho_{i\uparrow} = g_\uparrow(i\zeta_\uparrow + \alpha_\uparrow) \tag{12b}$$

$$\rho_{i\downarrow} = g_\downarrow(i\zeta_\downarrow + \alpha_\downarrow) \tag{12c}$$

where $g_\uparrow(x)$ and $g_\downarrow(x)$ have the same period and the same phase shift $\alpha_\uparrow - \alpha_\downarrow = \frac{1}{2}$ as $\chi_{\zeta_\uparrow}(x)$ and $\chi_{\zeta_\downarrow}(x)$ in (10*a*) and (10*b*). Of course we have $g(x) = g_\uparrow(x) + g_\downarrow(x)$.

The numerical calculation of the incommensurate mixed polaronic–bipolaronic structures are performed with the improved contracting operator technique described in appendix A. We study in practice finite systems, which are determined by the sequence of the 'best' rational approximates r_n/s_n of $\frac{1}{2} < \zeta_\uparrow < 1$ (see e.g. Hardy and Wright 1960) obtained by successive truncation of the continued fraction expansion of $\zeta_\uparrow$ at order n.

$$\zeta_\uparrow = 1/[\![a_1 + 1/\{a_2 + 1/[\ldots + 1/(a_n + \ldots)]\}]\!] = \{a_1, a_2, \ldots, a_n, \ldots\}. \tag{13}$$

This expansion also reveals the block structure of the pseudo-spin sequence $\{\sigma_i\}$ (Vallet *et al* 1988). The system studied to approach the incommensurate structure is a ring of $N = s_n$ sites with $n_\uparrow = r_n$ electrons with spin ↑ and $n_\downarrow = s_n - r_n$ electrons with spin ↓, a system with $s_n - r_n$ bipolarons and $2r_n - s_n$ polarons. We can also say that the dimerized chain contains $2r_n - s_n$ neutral solitons. The plot of $\rho_{i\uparrow}$, $\rho_{i\downarrow}$ and ρ_i versus $i(r_n/s_n)\text{mod } 1$ yields a fit of the periodic hull functions $g_\uparrow(x)$, $g_\downarrow(x)$ and $g(x)$ in (12). For $s_n \simeq 100$ these curves become independent of the order of the approximation (within the graph accuracy). Several hull functions $g_\uparrow(x)$ and $g_\downarrow(x)$ are shown in figure 7(*a*). It can be checked that

$$g_\uparrow(x) = g_\uparrow(-x) \tag{14a}$$

$$g_\uparrow(x) = 1 - g_\downarrow(x + \tfrac{1}{2}) \tag{14b}$$

which, as expected, have the same symmetry properties as the characteristic function (10*c*). In addition we also have

$$g(x) = \tfrac{1}{2}(g_\uparrow(x) + g_\downarrow(x)) = \tfrac{1}{2}(1 + g_\uparrow(x) - g_\downarrow(x + \tfrac{1}{2})). \tag{14c}$$

These curves appear as smooth curves for small values of k but, when k goes beyond a critical value $k_c(\zeta_\uparrow)$, which depends on $\zeta_\uparrow$, they become sharply discontinuous. There are few large discontinuities, which remain practically unchanged, when the size of the system (the order of the rational approximate) increases, and many other smaller discontinuities, which require an accurate analysis on larger systems.

A TBA is easily recognized. The same kind of transition with the same phenomenology was observed first in the FK model, but also in several models of Peierls chains and now in many other models for incommensurate structures (Aubry and Quemerais 1989). We also expect here that in the limit of infinite size each of these hull functions $g(x)$ has infinitely many discontinuities, although it remains a bounded variation function. The scaling properties of many critical quantities at the TBA have been studied in details in the FK model by a renormalization approach (MacKay 1992).

In the present model the physical characteristics of the TBA are those of a metal–insulator transition. For $k < k_c(\zeta_\uparrow)$ the phonon spectrum associated with the phase fluctuations is gapless. In the presence of an additional electric field the system should carry an electric current (and a spin current) by the uniform phase rotation of $\alpha_\uparrow$ according to the Fröhlich mechanism (Fröhlich 1954). (We pointed out in the earlier footnote that the adiabatic approximation used in this paper is not valid in the case of a Fröhlich conductivity, thus

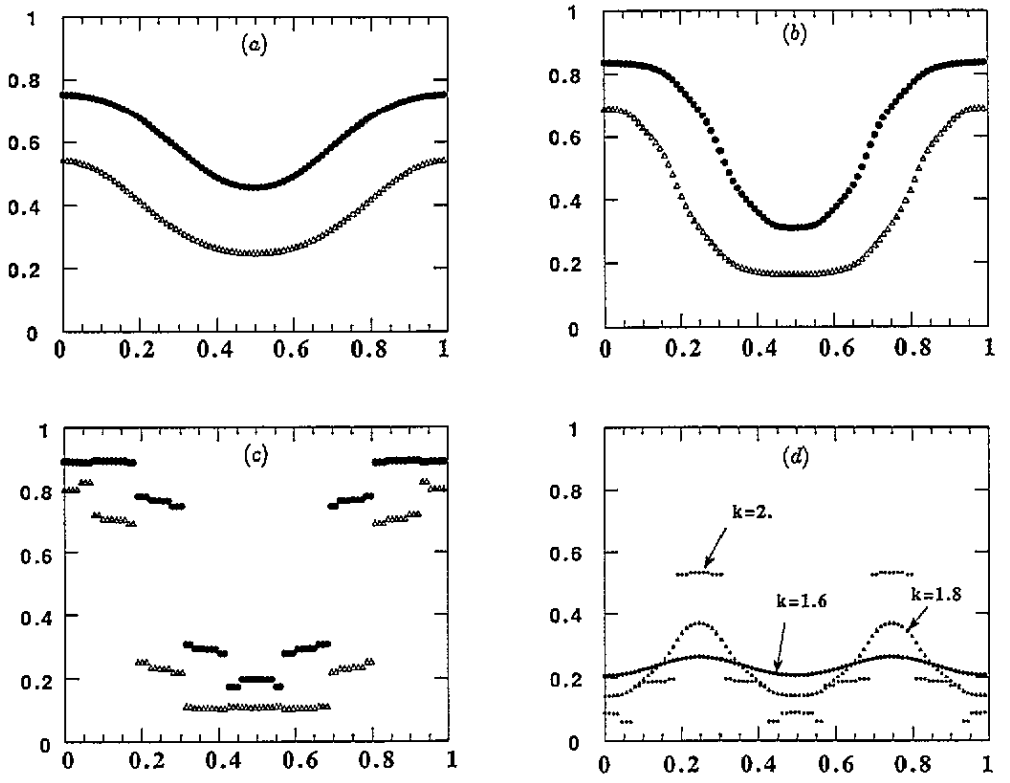

Figure 7. (*a*) Hull functions $g_\uparrow(x)$ (filled circles) and $g_\downarrow(x)$ (open triangles) for $k = 1.6$ and $\zeta_\uparrow = \frac{55}{89} \cong (\sqrt{5}-1)/2$. (*b*) As (*a*) for $k = 1.8$. (*c*) As (*a*) for $k = 2$. (*d*) Hull functions of the SDWS $\rho_{i\uparrow} - \rho_{i\downarrow}$ for several values of k.

the quantum lattice fluctuations have to play an essential role.) For $k > k_c(\zeta_\uparrow)$ a non-vanishing phason gap opens and the CDW–SDW pins to the lattice; it becomes an ordered quasi-periodic array of polarons and bipolarons. The propagation of electrons becomes a diffusive propagation of polarons and bipolarons within the mixed polaronic–bipolaronic structure, which is thermally activated and insulating at 0 K.

It has been rigorously proven for the FK model (Aubry *et al* 1991) and shown numerically for Peierls chains (Aubry *et al* 1990) that the non-analytic ground states can be decomposed into quasi-periodic arrays of well defined objects called effective discommensuration (in fact its bond modulation) in the FK model, but which could be named differently (e.g. effective bipolaron etc) according to the model and its physical context. More precisely the lattice distortion is obtained as a *linear* sum of *localized distortions* associated with each object: their positions are determined by a generalized pseudo-spin sequence equal to the coding sequence, which characterizes their distribution at the anti-integrable limit. At this limit the localized distortion extends only at a single site (or bond). An incommensurate bond modulation thus appears to be the convolution of its coding sequence by shape factors, which can be physically interpreted as effective shapes of different kinds of objects after their interactions with the other objects have been taken into account. In the simplest examples (incommensurate structures in the FK model, 1D Holstein model and SSH model without a magnetic field) there is only one kind of such objects, which is the effective discommensuration, but it has been also proven that there may exist more than one kind of such objects (but necessarily a finite number of kinds!) associated with a coding sequence

of several components. Numerical examples of extended FK models with several kinds of effective discommensurations were shown by Aubry *et al* (1991), where the coding sequence has several components. Even if the coding sequence is unknown, these objects can be associated with classes of discontinuities of the non-analytic hull function and thus can be identified numerically.

These effective discommensurations do not only describe the incommensurate ground states, but they also describe accurately the configurations of the system at low temperature, when the disorder of the coding sequence becomes weak. This weak disorder physically corresponds to phase fluctuations with a small phase gradient, which appear because the slightly random coding sequences can be constructed hierarchically by blocks up to a relatively high order (Vallet *et al* 1988). We have some mathematical results, which confirm this intuitive representation†.

We apply these general ideas to the incommensurate mixed polaronic–bipolaronic structures of our half-filled 1D Holstein model in a magnetic field. These structures are characterized by two pseudo-spin sequences $\{\sigma_{i\uparrow}\}$ and $\{\sigma_{i\downarrow}\}$ given by (10*a*) and (10*b*). If the hull functions $g(x)$, $g_\uparrow(x)$ and $g_\downarrow(x)$ (defined by (12)) have a discontinuity at x_0, they also have a discontinuity at $x_n = x_0 + n\zeta_\uparrow (\mathrm{mod}\ 1)$ for n integer positive or negative (Aubry *et al* 1991). This set of discontinuities $\{x_n\}$ is called a class of discontinuities. Since at the limit $k \to \infty$ these hull functions become $2\chi^*_\zeta(x)$ defined by (11) and $\chi_{\zeta_\uparrow}(x)$ and $\chi_{\zeta_\downarrow}(x)$ defined by (7*b*) we find that we have two discontinuities at $x_\uparrow = \zeta_\uparrow/2$ and $x_\downarrow = \zeta_\downarrow/2 = \frac{1}{2} - \zeta_\uparrow/2$ (see figure 3). For most $\zeta_\uparrow$ the classes of discontinuities are distinct, because there is no integer n such that $x_\uparrow - x_\downarrow = \frac{1}{2} = n\zeta_\uparrow (\mathrm{mod}\ 1)$. We conclude that we must have at least two classes of discontinuities for $g(x)$, $g_\uparrow(x)$ and $g_\downarrow(x)$. Our numerical analysis (see appendix B) shows the existence of two well localized shape factors $\{b_i^+\}$ and $\{b_i^-\}$, which allow one to represent the spin $\uparrow$ electronic density as a convolution of the pseudospin sequence by this shape factor

$$\rho_{i\uparrow} = \sum_n [\sigma_{n_\uparrow} b^+_{i-n} + (1 - \sigma_{n_\downarrow}) b^-_{i-n}] \tag{15a}$$

and are independent of the system size (see figure 8(*a*)). We find also the symmetry property $b_i^+ = b_{-i}^+$ and $b_i^- = b_{-i}^-$. (15*a*) is equivalent to writing the hull function (12*b*) as

$$\begin{aligned} g_\uparrow(x) &= \sum_n (b_n^+ \chi_{\zeta_\uparrow}(x - n\zeta_\uparrow) + b_n^- (1 - \chi_{\zeta_\downarrow}(x - n\zeta_\downarrow))) \\ &= \sum_n (b_n^+ \chi_{\zeta_\uparrow}(x - n\zeta_\uparrow) + b_n^- \chi_{\zeta_\uparrow}(x - n\zeta_\uparrow + \tfrac{1}{2})). \end{aligned} \tag{15b}$$

Since we have $\sum_i \rho_{i\uparrow} = n_\uparrow$, equation (15*a*) implies

$$\sum_i (b_i^+ + b_i^-) = 1. \tag{16}$$

† Let us consider an incommensurate configuration $\{u_i\} = \{\delta_i\}^*\{\sigma_i\}$, which can be represented as the convolution of an effective object $\{\delta_i\}$ with the coding sequence $\{\sigma_i\}$. A coding sequence $\{\tau_i\}$ is said to be *weakly disordered at order n with respect to the referring sequence* $\{\sigma_i\}$ when any subblock of symbols with length $n\{\tau_i\}_{i=p+1,p+n}$ of the infinite sequence $\{\tau_i\}$ also belongs to the sequence $\{\sigma_i\}$. (i.e. there exists q such that $\tau_i = \sigma_{i+q}$ for $p + 1 \leqslant i \leqslant p + n$). Then, if $\{v_i\}$ is the exact configuration associated with $\{\tau_i\}$ and $\{v_i'\} = \{\delta_i\}^*\{\tau_i\}$ is the configuration obtained by convolution of the sequence $\{\tau_i\}$ by the effective object $\{\delta_i\}$, we have $\sup_i |v_i - v_i'| < K \exp(-\gamma n/2)$ where γ is the Lyapounov coefficient or inverse coherence length of the incommensurate configuration $\{u_i\}$ and K is a constant.

Because of (14a), a similar relation holds for $\{\rho_{i\downarrow}\}$. We have

$$g_\downarrow(x) = 1 - g_\uparrow(x+\tfrac{1}{2}) = 1 - \sum_n [b_n^+ \chi_{\zeta_\uparrow}(x - n\zeta_\uparrow + \tfrac{1}{2}) + b_n^- \chi_{\zeta_\uparrow}(x - n\zeta_\uparrow)]$$

$$= \sum_n [b_n^+ \chi_{\zeta_\downarrow}(x + n\zeta_\downarrow) + b_n^- (1 - \chi_{\zeta_\uparrow}(x + n\zeta_\downarrow))] \tag{17a}$$

and

$$\rho_{i\downarrow} = g_\downarrow(i\zeta_\downarrow) = \sum_n [b_n^+ \chi_{\zeta_\downarrow}((n+i)\zeta_\downarrow) + b_n^- (1 - \chi_{\zeta_\uparrow}((n+i)\zeta_\uparrow))]$$

$$= \sum_n [\sigma_{n_\downarrow} b_{n-i}^+ + (1 - \sigma_{n_\uparrow}) b_{n-i}^-] \tag{17b}$$

which yields the electronic density

$$\rho_i = \rho_{i\uparrow} + \rho_{i\downarrow} = \sum_n (\sigma_{n_\uparrow} + \sigma_{n_\downarrow}) b_{n-i}^+ + (2 - \sigma_{n_\uparrow} - \sigma_{n_\downarrow}) b_{n-i}^-$$

$$= 1 + \sum_n (\sigma_{n_\uparrow} + \sigma_{n_\downarrow} - 1)(b_{n-i}^+ - b_{n-i}^-) \tag{18a}$$

and the spin density

$$s_i = \rho_{i\uparrow} - \rho_{i\downarrow} = \sum_n (\sigma_{n_\uparrow} - \sigma_{n_\downarrow})(b_{n-i}^+ + b_{n-i}^-). \tag{18b}$$

As expected we find in the non-analytic phase for $k > k_c(\zeta_\uparrow)$ that the electronic densities $\{b_i^+ - b_i^-\}$ and $\{b_i^+ + b_i^-\}$ are well defined and localized. They are independent of the size of the system, when it is large compared to their width. Formula (18a) has an explicit physical interpretation. If only polarons are present in the system ($\sigma_{n_\uparrow} + \sigma_{n_\downarrow} - 1 = 0$ for all n), the electronic density is uniform, $\rho_i = 1$. If a bipolaron is created at site 0 ($\sigma_{0\uparrow} + \sigma_{0\downarrow} - 1 = +1$), its contribution to the electronic density of the whole system is $b_i^+ + b_i^-$ at site i, while for a hole ($\sigma_{0\uparrow} + \sigma_{0\downarrow} - 1 = -1$) it is the same with just the opposite sign.

As we noted above this formula should also apply in case of weak disorder for describing the electronic density at low temperature providing the number of electrons remains strictly one electron per site, i.e. the number of bipolarons and holes is the same. We have not yet performed the block construction that generates the double sequence $\{\sigma_{i\uparrow}, \sigma_{i\downarrow}\}$ defined by (10a) and (10b) from the continued fraction expansion of $\zeta_\uparrow$ and which allows one to generate also the weakly disordered sequences by truncation. However, this construction is not essential in the framework of this paper, but would be useful for understanding the low-temperature behaviour of this model according to Vallet *et al* (1988).

As expected the formula (18b) shows that bipolarons and holes ($\sigma_{n_\uparrow} = \sigma_{n_\downarrow}$) do not contribute to the spin density. For a polaron at site 0 ($\sigma_{0\uparrow} = 1$ and $\sigma_{0\downarrow} = 0$), the contribution to the spin density at site i is $b_i^+ + b_i^-$ and that to the electronic density is $b_i^+ - b_i^-$. Figure 8(a) shows some examples of the effective polaron for several values of k and the same value of $\zeta_\uparrow$. Its width clearly diverges at $k_c(\zeta_\uparrow)$. Figure 8(b) shows that its shape also depends on the electronic density. For a given k, it is more extended for the best irrational numbers $\zeta_\uparrow$ (noble numbers†).

This phenomenon is a characteristic of the TBA between the bipolaronic–polaronic ground state and a Peierls–Fröhlich CDW–SDW (with gapless phason), where the polarons

† Noble numbers are characterized by a continued fraction expansion (13) where for some finite n_0, the integer sequence $\{a_n\}$ becomes $a_n = 1$ for $n > n_0$.

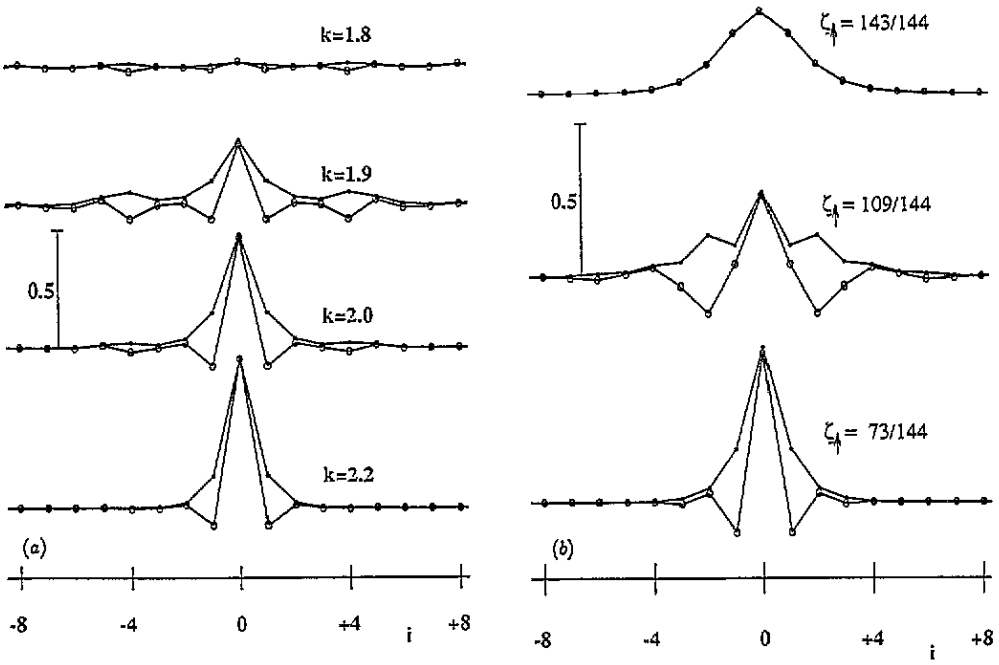

Figure 8. (*a*) Effective polarons (electron density $\{b_i^+ - b_i^-\}$ (open circles) and spin density $\{b_i^+ + b_i^-\}$ (filled circles)) for $\zeta_\uparrow = (\sqrt{5}-1)/2$ and several values of $k > k_c(\zeta_\uparrow)$. (*b*) Effective polarons (electron density $b_i^+ - b_i^-\}$ (open circles) and spin density $\{b_i^+ + b_i^-\}$ (filled circles)) for $k = 2.0$ and several values of $\zeta_\uparrow$.

and bipolarons disappear. The phenomenology of this transition, which has been studied in detail in other models (Aubry and Quemerais 1989), is the same in this model. The bipolaronic–polaronic ground states are insulating, while the Peierls–Fröhlich CDW–SDWs are conducting.

This TBA is not necessarily visible for all physical quantities. For example the electronic density of states shows basically three subbands (as predicted by Aubry *et al* (1992)), which does not change qualitatively above and below the TBA. The lowest-energy band corresponds to the doubly occupied electronic states. The midgap subband corresponds to the singly occupied states or neutral solitons and the upper subband to the empty states.

7. The devil's staircase as a Zeeman response

When a strong enough magnetic field is applied to the system, a non-linear Zeeman response is expected, because of the creation of an array of neutral solitons, which induce a magnetization. For a given concentration $2\zeta_\uparrow - 1$ of polarons or neutral solitons the bipolaronic–polaronic distribution of the ground state has been assumed to be given by the distribution (10). In the presence of a magnetic field this concentration of polarons is obtained by minimizing the energy per site $\Psi(\zeta_\uparrow)$ defined by (9*b*). For that purpose one calculates numerically the configurations corresponding to the series of values of $\zeta_\uparrow$ and their energy per site $\Psi_0(\zeta_\uparrow)$ defined by (9*a*) (see appendix C for technical details). Then the minimization of (9*b*) yields $\zeta_\uparrow$ as an implicit function of h by the equation

$$d\Psi_0(\zeta_\uparrow)/d\zeta_\uparrow = 2h. \tag{19}$$

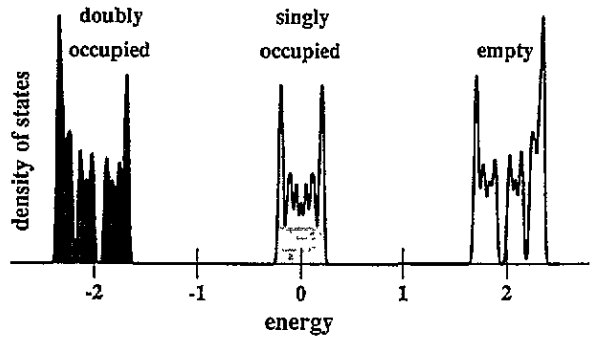

Figure 9. The electronic density of states (smoothed by Gaussians) for the mixed polaronic–bipolaronic state of figure 7(*c*) ($k = 2.0$, $\zeta_\uparrow = \frac{89}{144}$).

Function $\Psi_0(\zeta_\uparrow)$ is convex, because its derivative has been found to be monotonically increasing. This feature implies that the curve $\zeta_\uparrow = D(h)$ defined implicitly by (19) is well defined and monotonically increasing. This observation confirms that our choice for the bipolaronic–polaronic distribution (10) of the ground state is consistent (although it does not prove rigorously that this is the right choice). If indeed $\Psi_0(\zeta_\uparrow)$ were found not to be convex, this would imply that for certain values of $\zeta_\uparrow$ a lower energy for the ground state would be obtained by a phase separation into two phases with different $\zeta_\uparrow$, but if this situation occurred, the distribution (10) would not correspond to those of the ground state. This is indeed what we found sometimes for large k in the non-half-filled case ($\zeta_\uparrow + \zeta_\downarrow \neq 1$), if we made a careless extension of formulae (10) for finding the ground state. As we suggested in section 1, this feature also confirms that the bipolaronic–polaronic distribution for the ground state is more complex in the non-half-filled case than in the half-filled case and requires a specific study.

As for the FK model we expect that at all rationals $\zeta_\uparrow = r/s$ (r and s are irreducible) the left derivative of $\Psi_0(\zeta_\uparrow)$ is unequal and strictly smaller than the right derivative, but their difference goes to zero exponentially with the order s of the rational, thus we can confirm this feature only for the rational with an order s not too large. This curve $D(h)$, which is continuous and monotonically increasing, has constant plateaux at all rational $\zeta_\uparrow$, corresponding to the situations where the array of neutral solitons becomes commensurate with the period of the underlying lattice. Such a curve is called a devil's staircase (DS).

According to the Lebesgue decomposition theorem any monotonic function $D(h)$ is the sum of two monotonic functions $D_{ac}(h)+D_{sc}(h)$, where $D_{ac}(h)$ is an absolutely continuous function (by definition such a function is equal to the integral of its derivative) and $D_{sc}(h)$ a singular continuous curve, which is continuous but has a zero derivative almost everywhere. Apart from a constant term this decomposition is unique. The DS $D(h)$ is called complete if $D_{ac}(h)$ is constant and incomplete in the opposite case.

The existence of TBA for the array of neutral solitons implies consequences for the Zeeman response of this 1D adiabatic Holstein model, which has the same phenomenology as those observed and proved in the FK model. When for irrational $\zeta_\uparrow$ the corresponding incommensurate hull functions $g_\uparrow(x)$ and $g_\downarrow(x)$ are analytic functions, the derivative $\mathrm{d}\zeta_\uparrow(h)/\mathrm{d}h$ is defined and not zero. Then $D(h)$ has an absolutely continuous part $D_{ac}(h)$ and the DS $D(h)$ is incomplete. Truly incommensurate structures of neutral solitons can be obtained for a non-vanishing measure set of magnetic field and these structures, which are

unpinned from the lattice, are conducting (the phonon spectrum† is gapless). Our numerical study for $k = 1.8$ (see figure 10(*a*)) shows that for low-order rationals of $\zeta_\uparrow$ such as $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, etc the system is insulating (non-zero phason gap) and the DS is complete in this region, while for high-order rationals of $\zeta_\uparrow$ the system is conducting (vanishing phason gap) and the DS is incomplete there.

The set of values of irrational $\zeta_\uparrow$, where $k < k_c(\zeta_\uparrow)$ has a finite measure, exhibits the typical structure predicted by the KAM theory with a gap at all rationals. In principle, the DS becomes complete for $\zeta_\uparrow$ in each of these gaps (in some interval beyond the edges of each plateau of the DS). In practice, except for the low-order rationals, the gaps associated with these rationals become negligible. When the magnetic field is applied the system undergoes some re-entrant phase transitions between insulating and conducting states (see figure 10(*a*)). As conjectured by Aubry (1991a) we expect that the analytic mixed CDW–SDWs are unstable with respect to the quantum lattice fluctuations (which were neglected in this paper). Thus the conducting states should be simply metallic without charge modulation and with a uniform spin distribution.

In addition, we have already pointed out for the FK model, the incomplete DS transformation is reversible with almost no hysteresis since the intermediate incommensurate configurations are unpinned from the lattice, while most commensurate configurations are weakly pinned. When $k > \sup\{\zeta_\uparrow\} k_c(\zeta_\uparrow)$ the hull function $g_\uparrow(x)$ is no longer analytic for any irrational $\zeta_\uparrow$. As a result the variation $D(h)$ of $\zeta_\uparrow$ versus the magnetic field h is a complete DS. The structure always remains an insulating mixed bipolaronic–polaronic structure. The set of values of the magnetic field, which yields an incommensurate structure, is infinite and uncountable, although it has a zero (Lebesgue) measure. In principle, if the system is supposed to be in equilibrium at any magnetic field, it is commensurate with probability one although the commensurability order can be very large. Our numerical study for $k = 2.0$ (figure 10(*b*)) shows that the system is insulating except at very high magnetic fields, where it is conducting ($\sup\{\zeta_\uparrow\} k_c(\zeta_\uparrow) \cong 2.0$). The global phase diagram is shown in figure 11

We pointed out many years ago for the FK model that a complete DS cannot be observed strictly in real experiments. Unlike the case where the DS is incomplete, the Peierls–Nabarro barrier (which characterizes the lattice pinning) does not vanish even for irrational $\zeta_\uparrow$ and always remains larger than a strictly non-zero bound. Only the main plateaux of the DS could be experimentally observed, because the defect energies of the low-order commensurate phases become relatively large. For the commensurate phase with a high order or for the incommensurate phases many metastable states exist, which cost very little energy (in our model they are characterized by weakly disordered pseudo-spin sequences as explained in an earlier footnote). Thus the complete DS should appear in real systems as an irreversible transformation with a global hysteresis and memory effects. Except for the main commensurate phases the system is generally trapped in metastable states, which are out of equilibrium and exhibit a more or less pronounced glassy behaviour. The larger the

† The phonon spectrum is obtained by applying perturbation theory for the second derivative of the total energy $F(\{u_i\}, n_\uparrow, n_\downarrow)$ from (5*b*) and then diagonalizing the matrix

$$M_{n,m} = \frac{d^2 F\{u_i\}, n_\uparrow, n_\downarrow)}{du_n du_m} = \delta_{n,m} - \frac{k^2}{2} \sum_{\nu,\nu' \neq \nu}^{N} \sigma_{\nu\uparrow} \left(\frac{\psi_n^{\nu*} \psi_n^{\nu'} \psi_m^{\nu'*} \psi_m^{\nu}}{E_{\nu'} - E_\nu} + \text{cc} \right)$$

$$= \delta_{n,m} - k^2 \sum_{\nu,\nu'}^{N} \{\sigma_{\nu\uparrow}(1 - \sigma_{\nu'\uparrow}) + \sigma_{\nu\downarrow}(1 - \sigma_{\nu'\downarrow})\} \frac{\psi_n^{\nu*} \psi_n^{\nu'} \psi_m^{\nu'*} \psi_m^{\nu}}{E_{\nu'} - E_\nu}.$$

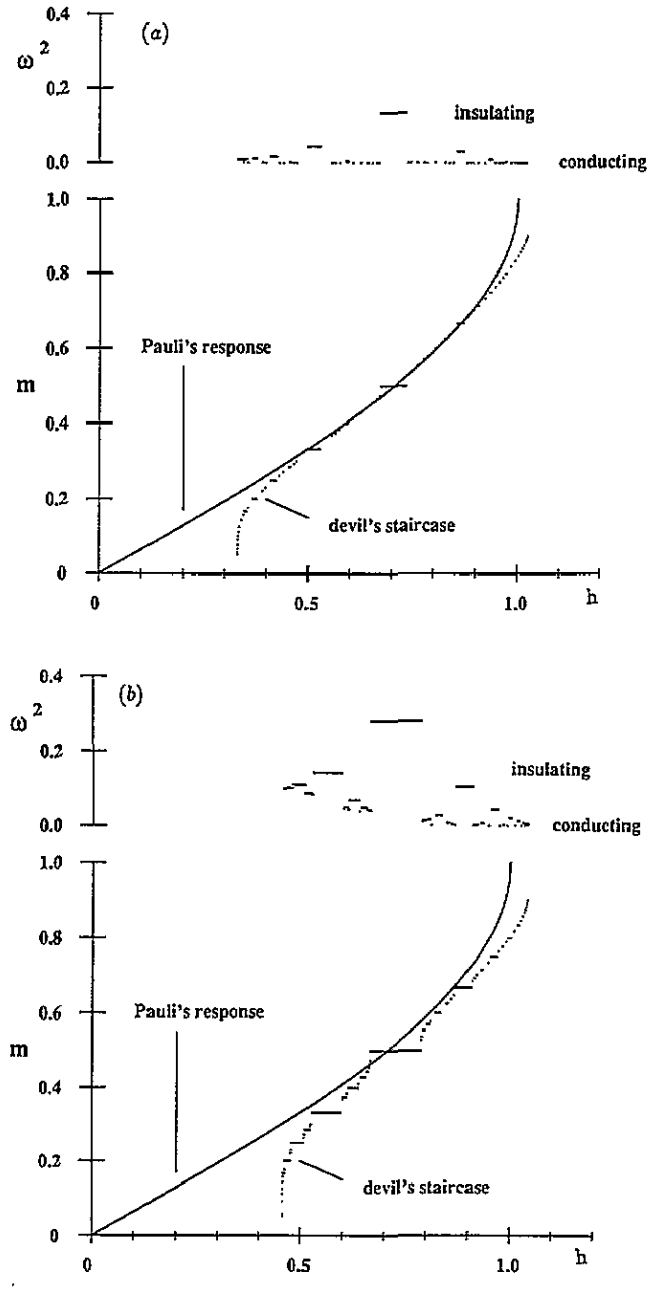

Figure 10. (*a*) The magnetization $m = 2\zeta_\uparrow - 1$ against h, $k = 1.8$, with all rationals $\frac{1}{2} < r/s < 1$ of the commensurability ratio $\zeta_\uparrow = r/s$ obtained by the Farey construction up to the order 20: insulating (non-zero phason gap ω^2) for low-order rationals of $\zeta_\uparrow$; conducting (zero phason gap ω^2) for high-order rationals of $\zeta_\uparrow$. (*b*) The magnetization $m = 2\zeta_\uparrow - 1$ against h, $k = 2.0$, with all rationals $\frac{1}{2} < r/s < 1$ of the commensurability ratio $\zeta_\uparrow = r/s$ obtained by the Farey construction up to order 20: conducting (zero phason gap ω^2) only at high magnetic field h.

Peierls–Nabarro barrier, the larger should be the entropy of the residual disorder allowed for these metastable states. An investigation of these metastable states, their entropy and their physical role has only been made in the 1D FK model, which has no long-range order

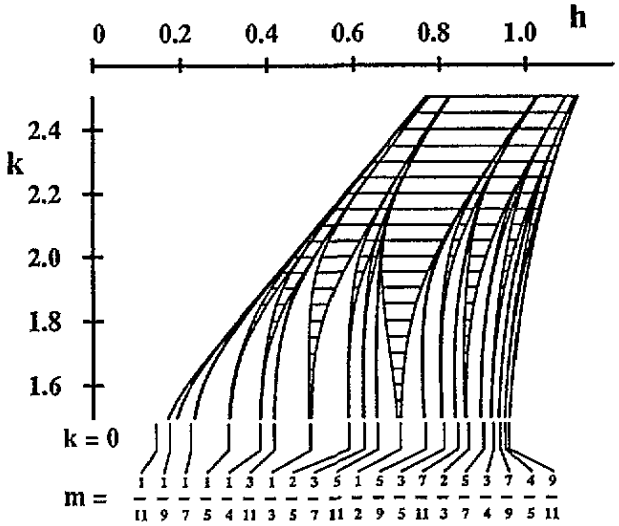

Figure 11. Plateaux of the devil's staircase: magnetization $m = 2\zeta_\downarrow - 1$ against k for all rationals $\frac{1}{2} < r/s < 1$ with $s \leqslant 11$ of the commensurability ratio $\zeta_\uparrow = r/s$. Values of magnetization m against h for $k = 0$ (Pauli susceptibility) are given.

at finite temperature (Vallet *et al* 1988) and does not exhibit really frozen states.

8. Application to physical systems: discussion and concluding remarks

The characteristic energy required for a magnetic field to observe the edge of the first plateau of the DS at $\zeta_\uparrow = \frac{1}{2}$ is the energy of a neutral soliton E_n. To observe at least the beginning part of this DS in quasi-1D conductors the magnetic field h has to be larger than the edge h_c of the first plateau of $D(h)$ at $\zeta_\uparrow = \frac{1}{2}$. The potential energy for creating a neutral soliton is $2Te_s$, while the magnetic energy gain is $\frac{1}{2}g\mu_B H = 2Th$. Thus the critical field h_c is equal to the energy of the neutral soliton e_s, which is approximately a quarter of the electronic gap, which itself is approximately the energy $k_B T_{MF}$ corresponding to the mean field critical temperature T_{MF} of the dimerized CDW (in the absence of a magnetic field). Since the largest uniform magnetic field realizable does not exceed 40 T, this critical magnetic field will be observable for CDWs with a mean field critical temperature not larger than 25 K (the energy of the Bohr magneton being 0.671 K T^{-1}). We expect that this argument on the critical magnetic field also applies for quasi-1D conductors with a non-half-filled band, although the DS found in this particular model is expected to become a more complex curve and requires a specific study. Real CDWs with a half-filled band and such a low critical temperature seem to be exceptional.

Let us note that this critical value of the magnetic field is expected to be lowered if the electron–electron repulsion is taken into account, because the energy of breaking a bipolaron into two polarons (which is also the energy of two neutral solitons) will be reduced. However, if this electronic repulsion is sufficiently strong, we will have a polaronic SDW instead of a bipolaronic CDW (Aubry 1993). This is what might happen in some compounds of the $(TMTSF)_2X$ series (Beechgaard salts), which exhibit SDWs in the absence of a magnetic field with a low critical temperature (Jérome 1990). Many phase transitions under a magnetic field were observed in these materials, but explained on the basis of magnetic orbital effects. Some similar organic compounds such as $TTF(M(dmit)_2)_2$ with M = Ni (Brossard *et al* 1990)

exhibit a CDW instead of an SDW with a critical temperature around 40 K, but this can be reduced under pressure (before the system reaches a superconducting state). We suggest that it should be easier to find quasi-1D CDWs in the family of the organic compounds with low critical temperature; these could be good candidates for exhibiting a DS-like magnetic response related to the formation of arrays of neutral solitons.

However, this work has been focused on a relatively simple situation that is understandable. It suggests that there are more complex situations for the Zeeman response not only for bipolaronic CDWs but also for polaronic SDWs. Although the formation of mixed bipolaronic–polaronic structures under the Zeeman response for non-half-filled models is also highly non-linear, we have not yet found the bipolaronic–polaronic distribution of their ground state, and although our general theory predicts the persistence of bipolaronic–polaronic ground states in two and three dimensions, finding their distribution becomes a tougher problem (even in the absence of a magnetic field if the band filling is not $\frac{1}{2}$).

The role of the temperature has not been analysed. We know examples where the DS persists in mean field models at non-zero temperature (Floria *et al* 1992). The width of the plateaux of the DS is expected to be reduced and the critical value of the magnetic field required for generating neutral solitons could then be lowered (entropy reduction of the resulting mixed bipolaronic–polaronic structure), but a specific study should be devoted to this effect.

The contribution to the magnetic response of the orbital response of the bipolaronic and polaronic structures in two and three dimensions might also be important, if the system is not strictly 1D. We have recently proved very generally that in 2D and 3D models (Aubry and Kuhn 1995 in preparation) a magnetic field localizes a single electron forming a polaron as soon as the electron–phonon coupling is not strictly zero. This effect appears to be essential in quasi-1D systems, but still has to play also an important role for 2D systems, i.e. in the physical situation encountered in the quantum Hall effect. This localization effect becomes very weak in isotropic 3D systems.

A correct treatment for the magnetic response of real systems should involve both the orbital and the Zeeman response. The bipolaronic–polaronic models for CDW–SDWs will give a wide variety of physical behaviours, which are still hardly explored.

Acknowledgments

We thank Drs J L Raimbault and J P Gosso for many valuable discussions. One of us (CK) thanks the late Dr J Rossat Mignod for hospitality at the LLB and the Swiss National Foundation for financial support.

Appendix A. Numerical technique for finding bipolaron, polaron and mixed structures

Three methods were used. The last one is presently the most efficient. We also briefly describe the first methods, which were used in early works.

A1. Gradient method (Le Daëron and Aubry 1983a)

This is the most intuitive method. Since the bipolaronic and polaronic structures are local minima of the variational energy (*5b*), it is natural to follow the gradient flow of this energy to reach one of these local minima. To find the configuration associated with a pseudo-spin

sequence $\{\sigma_i\}$ we start from an initial configuration $u_i(0) = -k\sigma_i$ and we integrate the differential equation

$$du_n/dt = -\partial F(\{u_i\}, n_\uparrow, n_\downarrow)/\partial u_n = -(u_n(t) + k/2(\rho_{n_\uparrow} + \rho_{n_\downarrow})) \tag{A1}$$

where $\rho_{n_\uparrow}$ and $\rho_{n_\downarrow}$ are functions of $u_n(t)$ obtained with (6*b*) from the eigenvectors of (5*c*) with their eigenvalues in increasing order. The program is stopped at $u_n(\infty)$ for large enough t, when the gradient becomes smaller than a given accuracy ε. We know from the theorems given by Aubry *et al* (1992) that for large enough k the initial configuration $\{-k\sigma_i\}$ belongs to the basin of attraction of the final metastable state $\{u_n(\infty)\}$, which is the required mixed polaronic–bipolaronic configuration. For smaller k we follow continuously the metastable configuration through its cascades of bifurcations. The flaw of this method is that it requires a continuous integration on t, thus many steps for convergence and a matrix diagonalization at each step.

A2. Contracting operator method (Quemerais 1987, Raimbault 1990)

The contracting operator $\{u'_n\} = T(\{u_n\})$, which has been used for proving the theorems of Aubry *et al* (1992) and Baesens and MacKay (1993), is iterated from the initial configuration $\{-k\sigma_i\}$. For large k these theorems assert that this configuration belongs to the basin of attraction of the corresponding mixed polaronic–bipolaronic structure. This operator is defined as

$$u'_n = -k/2(\rho_{n_\uparrow} + \rho_{n_\downarrow}). \tag{A2a}$$

It can be obtained as an approximation of (A1) with a coarse discretization of the continuous variable t

$$u_n(t+1) - u_n(t) = -(u_n(t) + k/2(\rho_{n_\uparrow} + \rho_{n_\downarrow})). \tag{A2b}$$

This algorithm converges much faster (by a factor of more than 10) to the same limit as the gradient method, which requires a much thinner discretization. This method still has the flaw that it requires a matrix diagonalization at each step.

A3. Improved contracting operator method

In this paper we use an improved contracting operator method, which does not require this matrix diagonalization at each step. This new method is obtained by a combination of the above method and diagonalizing a matrix by power iteration.

Finding the minima of the variational form (5*b*) is equivalent to finding the absolute minima over both $\{u_i\}$ and a set of n orthonormal vectors $\{\psi_n^\nu\}$ of variational energy

$$F(\{u_i\}, \{\psi_n^\nu\}) = \sum_i \frac{1}{2}u_i^2 + \frac{1}{2}\sum_{\nu=1}^{n_\uparrow}\langle\nu|H(\{u_i\})|\nu\rangle + \frac{1}{2}\sum_{\nu=1}^{n_\downarrow}\langle\nu|H(\{u_i\})|\nu\rangle \tag{A3a}$$

where $|\nu\rangle = \{\psi_n^\nu\}$ is the νth vector of a set of n orthonormal vectors and

$$H(\{u_i\}) = \sum_n(-|n\rangle\langle n+1| - |n\rangle\langle n-1| + ku_n|n\rangle\langle n|) \tag{A3b}$$

is the Hamiltonian (5*c*). Then (A3*a*) can be written more explicitly as

$$F(\{u_i\}, \{\psi_n^\nu\}) = \sum_n \frac{1}{2}u_n^2 + \frac{k}{2}u_n(\rho_{n_\uparrow} + \rho_{n_\downarrow}) - \frac{1}{2}\sum_{\nu=1}^{n_\uparrow}\sum_n(\psi_{n+1}^{\nu^*}\psi_n^\nu + \psi_n^{\nu^*}\psi_{n+1}^\nu) - \frac{1}{2}\sum_{\nu=1}^{n_\downarrow}\sum_n(\psi_{n+1}^{\nu^*}\psi_n^\nu + \psi_n^{\nu^*}\psi_{n+1}^\nu) \tag{A4}$$

with $\rho_{n_\uparrow}$ and $\rho_{n_\downarrow}$ defined by (6*b*). The minimum of (A4) with respect to u_n is obtained for $u_n = -k/2(\rho_{n_\uparrow} + \rho_{n_\downarrow})$ and yields

$$F_0(\{\psi_n^\nu\}) = -\frac{k^2}{8}\sum_n\left(\sum_{\nu=1}^{n_\uparrow}|\psi_n^\nu|^2 + \sum_{\nu=1}^{n_\downarrow}|\psi_n^\nu|^2\right)^2 - \frac{1}{2}\sum_{\nu=1}^{n_\uparrow}\sum_n(\psi_{n+1}^{\nu^*}\psi_n^\nu + \psi_n^{\nu^*}\psi_{n+1}^\nu) - \frac{1}{2}\sum_{\nu=1}^{n_\downarrow}\sum_n(\psi_{n+1}^{\nu^*}\psi_n^\nu + \psi_n^{\nu^*}\psi_{n+1}^\nu) \quad \text{(A5}a\text{)}$$

or equivalently

$$F_0(\{\psi_n^\nu\}) = -\frac{k^2}{8}\left(\sum_{\nu=1}^{n_\uparrow}\langle\nu|\sum_n|n\rangle\langle n||\nu\rangle + \sum_{\nu=1}^{n_\downarrow}\langle\nu|\sum_n|n\rangle\langle n||\nu\rangle\right)^2 - \frac{1}{2}\sum_{\nu=1}^{n_\uparrow}\langle\nu|\left(\sum_n|n+1\rangle\langle n| + |n\rangle\langle n+1|\right)|\nu\rangle - \frac{1}{2}\sum_{\nu=1}^{n_\downarrow}\langle\nu|\left(\sum_n|n+1\rangle\langle n| + |n\rangle\langle n+1|\right)|\nu\rangle. \quad \text{(A5}b\text{)}$$

The set of vectors $\{\psi_n^\nu\}$ spans the two subspaces called $\mathcal{E}_{n_\uparrow}$ for $\nu = 1$ to $n_\uparrow$ and $\mathcal{E}_{n_\downarrow}$ for $\nu = 1$ to $n_\downarrow$. This variational form appears to be only a function of the trace of such operators, which are truncated to $\mathcal{E}_{n_\uparrow}$ or $\mathcal{E}_{n_\downarrow}$, and thus is invariant under any vector rotation that leaves invariant these subspaces $\mathcal{E}_{n_\uparrow}$ and $\mathcal{E}_{n_\downarrow}$. The problem of minimization consists of finding these global subspaces, but not a specific set of vectors.

The absolute minimum of $F_0(\{\psi_n^\nu\})$ is obtained by choosing for example for $\{\psi_n^\nu\}$ the $n_\uparrow$ first eigenvectors of the (non-linear) eigenequation

$$-\psi_{n+1}^\nu - \psi_{n-1}^\nu - (k^2/2)(\rho_{n_\uparrow} + \rho_{n_\downarrow})\psi_n^\nu - 2\psi_n^\nu = (E_\nu - 2)\psi_n^\nu \quad \text{(A6)}$$

where $\rho_{n_\uparrow} = \sum_{\nu=1}^{n_\uparrow}|\psi_n^\nu|^2$ and $\rho_{n_\downarrow} = \sum_{\nu=1}^{n_\downarrow}|\psi_n^\nu|^2$ are determined self-consistently. (We introduced $-2\psi_n^\nu$ in both members of (A6) so that all the eigenvalues $(E_\nu - 2)$ are negative, the ones with the largest modulus being the smallest.)

To find $\mathcal{E}_{n_\uparrow}$ and $\mathcal{E}_{n_\downarrow}$ we take the successive power of the member of (A6). For that we define the operator $\tilde{T}$, which transforms a given set of $n_\uparrow$ orthonormal vectors $\{\varphi_n^\nu\}$ into a new set of $n_\uparrow$ orthonormal vectors $\{\tilde{\varphi}_n^\nu\} = \tilde{T}(\{\varphi_n^\nu\})$. We first define a set of $n_\uparrow$ vectors $\{\varphi_n'^\nu\}$ as

$$\varphi_n'^\nu = -\varphi_{n+1}^\nu - \varphi_{n-1}^\nu - \frac{k^2}{2}\left(\sum_{\nu=1}^{n_\uparrow}|\varphi_n^\nu|^2 + \sum_{\nu=1}^{n_\downarrow}|\varphi_n^\nu|^2\right)\varphi_n^\nu - 2\varphi_n^\nu \quad \text{(A7)}$$

which next are orthogonalized and normalized by using the standard Graham–Schmidt method and yield $\{\tilde{\varphi}_n^\nu\}$. The sequence of subspaces spanned by $\{\varphi_n^\nu\}$ converges after a few iterations to a limit space, which yields a local minima for $F_0(\{\psi_n^\nu\})$ although each $\{\varphi_n^\nu\}$ has not yet converged to a good eigenvector.

For finding the mixed polaronic–bipolaronic state defined by a given sequence $\{\sigma_i = 0, \frac{1}{2}$ or $1\}$ we take as initial vectors the vectors φ_n^ν, which are localized at single sites, where $\sigma_i \neq 0$. We choose first but in an arbitrary relative order a block of vectors ($\nu = 1$ to $n_\downarrow$) localized at the bipolaronic sites where $\sigma_i = 1$ and next also in an arbitrary relative order a second block of vectors ($\nu = n_\downarrow + 1$ to $n_\uparrow$) localized at the polaronic sites, where $\sigma_i = \frac{1}{2}$. To preserve the initial polaronic and bipolaronic ordering the vectors of the first and second

blocks must not be exchanged at each iteration, but the choice of the relative order in each block remains arbitrary. Then we apply recursively the transformation $\tilde{T}$ on this initial set of vectors and stop the transformation when both electronic densities $\rho_{n_\uparrow}$ and $\rho_{n_\downarrow}$ converge to a certain limit for all n. This limit state essentially depends on $\{\sigma_i\}$, but not on the details of the procedure.

The rate of convergence for finding this pair of invariant subspaces $\mathcal{E}_{n_\uparrow}$ and $\mathcal{E}_{n_\downarrow}$ is determined by the two main electronic gaps of $\tilde{T}$, which separate the two invariant subspaces $\mathcal{E}_{n_\downarrow}$ from $\mathcal{E}_{n_\uparrow} - \mathcal{E}_{n_\downarrow}$ and $\mathcal{E}_{n_\uparrow}$ from its complementary space. (Our theorem predicts that these electronic gaps persist for large systems between the $n_\uparrow$th and the $(n_\uparrow + 1)$th eigenvalue and between the $n_\downarrow$th and the $(n_\downarrow + 1)$th eigenvalue.) The rate of convergence does not depend on the relative distance of the closest eigenvalues as in a standard diagonalization (which has the flaw that this distance becomes very small when the size of the system becomes large), but essentially depends on the main electronic gaps, which are independent of the size of the system.

We checked that this third technique yields the same results as the contracting operator technique, but within a CPU time about 10 times less. Its advantage is that it works with only a coarse diagonalization of the operators. (However, it is convenient to perform one complete diagonalization of (A6) just after the last iteration to obtain the detailed electronic spectrum.)

Appendix B. Numerical calculation of the effective bipolarons, holons and spinons

The incommensurate bipolaronic–polaronic configuration $\{u_i\}$ with incommensurability ration $\zeta_\uparrow$ is approximated by a finite system with $n_\uparrow = r_n$ electrons with spin $\uparrow$, $n_\downarrow = s_n - r_n$ electrons with spin $\downarrow$ and s_n sites. The integers r_n and s_n are obtained from the series of best approximates r_n/s_n of $\zeta_\uparrow$, which are given by the truncation of the continued fraction expansion (13) at order n. Plotting the electronic densities $\rho_{i\uparrow}$, $\rho_{i\downarrow}$ and ρ_i as a function of the variable $x_i = ir_n/s_n \bmod 1 = p_i/s_n$ where the integers p_i take all values from zero to $s_n - 1$ yields discrete approximations of the hull functions $g_\uparrow(x)$, $g_\downarrow(x)$ and $g(x)$ respectively. The smallest phase shift of this commensurate structure is $1/s_n$.

We search first a periodic sequence of numbers $a_i^\uparrow$ with a period s_n such that

$$\rho_{i\uparrow} = \sum_m \sigma_{m\uparrow} a_{i-m}^\uparrow \tag{B1a}$$

where

$$\sigma_{m\uparrow} = \chi_{\zeta_\uparrow}(mr_n/s_n) \tag{B1b}$$

is the pseudo-spin sequence for the electrons with spin $\uparrow$ of the initial configuration calculated with the technique of appendix A.

This sequence $a_{i-m}^\uparrow$ is simply obtained by solving a set of s_n linear equations with s_n variables. The sequence $a_i^\uparrow$ shown in figure B1 appears as two peaks, which can be split into the sum of two peaks at separation q_n

$$a_i^\uparrow = b_i^+ + b_{i+q_n}^- \tag{B2}$$

where q_n is an integer that depends on the size s_n of the system. When the order n of the approximate increases, we find that the shape of each of these two peaks does not change,

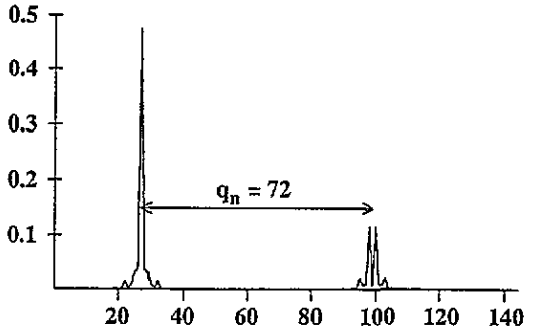

Figure B1. $a_i^{\uparrow}$ can be split into a sum of two peaks at separation q_n ($k = 2.0$ and $\zeta_\uparrow = \frac{55}{89}$).

but q_n increases with n. The essential point is that we find $q_n r_n/s_n \bmod 1 = \frac{1}{2}$ for s_n even and $q_n r_n/s_n \bmod 1 = \frac{1}{2} \pm 1/2s_n \cong \frac{1}{2}$ for s_n odd.

$$\rho_{i\uparrow} = \sum_m \sigma_{m\uparrow}(b^+_{i-m} + b^-_{i+q_n-m}) = \sum_m \sigma_{m\uparrow} b^+_{i-m} + \sum_m \sigma_{m+q_n\uparrow} b^-_{i-m}. \tag{B3}$$

We can now write (B3) and this yields (15a) with

$$\sigma_{m+q_n\uparrow} = \chi_{\zeta_\uparrow}((m+q_n)r_n/s_n) = \chi_{\zeta_\uparrow}(mr_n/s_n + \tfrac{1}{2}) = 1 - \chi_{\zeta_\downarrow}(mr_n/s_n)$$
$$= 1 - \chi_{\zeta_\downarrow}(m(1 - r_n/s_n)) = 1 - \sigma_{m\downarrow}. \tag{B4}$$

Appendix C. Numerical calculation of the devil's staircase

We calculate the energy $\Psi_0(\zeta_\uparrow)$ for a sequence of irreducible rationals r/s between $\frac{1}{2}$ and 1 (obtained by the well known Farey construction). We study systems around 100 sites. For this purpose we use a multiplicative factor q such that $qs \simeq 100$ and study the system with qr electrons with spin $\uparrow$, $(s-r)q$ electrons with spin $\downarrow$ and sq sites. We also calculate the energy of the same system but with a single retarded or advanced discommensuration. We consider the finite continuous fraction expansion

$$r/s = 1/\{a_1 + 1/[a_2 + 1/(\ldots + 1/a_n)]\} = \{a_1, a_2, \ldots, a_n = 1\}. \tag{C1}$$

With the remainder $\pm 1/q$ we obtain rationals

$$r^+/s^+ = \{a_1, a_2, \ldots, a_n, +q\} = (qr + r')/(qs + s') \tag{C2a}$$

and

$$r^-/s^- = \{a_1, a_2, \ldots, a_n, -q\} = (qr - r')/(qs - s') \tag{C2b}$$

where

$$r'/s' = \{a_1, a_2, \ldots, a_{n-1}\} \tag{C3}$$

is before the last truncation of (C1). The systems with $qr \pm r'$ electrons with spin $\uparrow$, $(s-r)q \pm (s'-r')$ electrons with spin $\downarrow$ and $sq \pm s'$ sites contain a single advanced discommensuration or a single retarded discommensuration in their unit cell respectively. We have $r^+/s^+ - r/s = (-1)^n/ss^+$ and $r^-/s^- - r/s = -(-1)^n/ss^-$.

The right and left derivatives $\Psi_0'^{\pm}(r/s)$ of $\Psi_0(\zeta_\uparrow)$ for $\zeta_\uparrow = r/s$ are obtained by calculating the energy differences $[\Psi_0(r^\pm/s^\pm) - \Psi_0(r/s)]/(r^\pm/s^\pm - r/s)$. They are found to be different with $\Psi_0'^{-}(r/s) < \Psi_0'^{+}(r/s)$, which yields the edge of the plateau of $D(h)$ corresponding to $\zeta_\uparrow = r/s$.

References

Aubry S 1989 *J. Physique Coll.* **50** C3 97–106

—— 1991a *Microscopic Aspects of Non-Linearity in Condensed Matter Physics* (NATO ASI Series B) vol 264, ed A R Bishop, V L Pokrovsky and V Tognetti (New York: Plenum) pp 105–14

—— 1991b *Geometry and Thermodynamics* (NATO ASI Series B: Physics 229) ed J C Toledano (New York: Plenum) pp 281–300

—— 1993 *J. Physique IV Coll.* **3** C2 349–55

—— 1994 *Physica* D **71** 196–221

Aubry S, Abramovici G and Raimbault J-L 1992 *J. Stat. Phys.* **67** 675–780

Aubry S, Godreche C and Luck J M 1987 *Europhys. Lett.* **4** 639–43

—— 1988 *J. Stat. Phys.* **51** 1033–75

Aubry S, Gosso J P, Abramovici G, Raimbault J L and Quemerais P 1991 *Physica* D **47** 461–97

Aubry S and Kuhn C 1995 in preparation

Aubry S and Quemerais P 1989 *Low-Dimensional Electronic Properties of Molybdenum Bronzes and Oxides* ed C Schlenker (Deventer: Kluwer)

Aubry S, Quemerais P and Raimbault J L 1990 *3rd Eur. Conf. on Low Dimensional Conductors and Superconductors (Dubrovnik)* ed S Barisic *Fisica* **21** (supplement 3) 98–101

Baesens C and MacKay R S 1994 *Nonlinearity* **7** 59–84

Brazovskii S A and Matveenko S I 1984 *Sov. Phys.–JETP* **60** 804

Brazovskii S A, Matveenko S I and Kirova N 1981 *Sov. Phys.–JETP* **54** 1209

Brossard L, Canadell E, Ravy S, Pouget J P, Legros J P and Valade L 1990 *Physica* **21** (supplement 3) 15

Bryant G W and Glick A J 1982 *Phys. Rev.* B **26** 5855–66

Feinberg D and Ranninger J 1983 *J. Phys. C: Solid State Phys.* **16** 1875–85

Floria L M, Quemerais P and Aubry S 1992 *J. Phys.: Condens. Matter* **4** 5921–46

Fröhlich H 1954 *Proc. R. Soc.* A **223** 296

Hardy G H and Wright E M 1960 *An Introduction to the Theory of Numbers* (Oxford: Clarendon)

Holstein T 1959 *Ann. Phys.* **8** 325–43

Jérome D 1990 *3rd Eur. Conf. on Low Dimensional Conductors and Superconductors (Dubrovnik)* ed S Barisic *Physica* **21** (supplement 3) 1–6

Kontorowa T and Frenkel I 1938 *Phys. Z. Sov. Un.* **13** 1

—— 1939 *Fis. Z.* **1** 137

Kuhn C 1989 *Phys. Rev.* B **40** 7776–87

Le Daëron P Y and Aubry S 1983a *J. Phys. C: Solid State Phys.* **16** 4827–38

—— 1983b *J. Physique Coll.* **C3** 1573–7

MacKay R 1992 *Physica* D **50** 71–9

MacKay R and Baesens C 1993 *Quantum Chaos* ed G Casati, I Guarneri and U Smilansky (Amsterdam: North-Holland) pp 51–75

Peierls R E 1955 *Quantum Theory of Solids* (Oxford: Oxford University Press) p 108

Quemerais P 1987 *PhD Dissertation* University of Nantes

Raimbault J L 1990 *PhD Dissertation* University of Nantes

Raimbault J L and Aubry S 1994 in preparation

Su W P and Schrieffer J R 1981 *Phys. Rev. Lett.* **46** 738

Su W P, Schrieffer J R and Heeger A J 1979 *Phys. Rev. Lett.* **42** 1968

—— 1980 *Phys. Rev.* B **22** 2099

Vallet F, Schilling R and Aubry S 1988 *J. Phys. C: Solid State Phys.* **21** 67–105