Ground states of one-dimensional systems using effective potentials

Weiren Chou* and Robert B. Griffiths
Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213
(Received 12 May 1986)

A nonlinear eigenvalue equation whose solution is an "effective potential" is used to study the ground states of one-dimensional systems (such as the Frenkel-Kontorova model) whose Hamiltonian H is a sum of terms $V(u_n)+W(u_{n+1}-u_n)$, where the u_n are real and V is periodic. The procedure is not limited to convex W, and it yields the ground-state energy and orbit, in contrast to metastable or unstable states, and some information about "soliton" defects. It can be generalized to H a sum of $K(x_{n+1},x_n)$, where the arguments may be multidimensional. Numerical solutions of the eigenvalue problem are used to work out phase diagrams for W a parabola, and various choices of V. With V a cosine plus a small admixture of a second or third harmonic with the proper sign, we find first-order transitions between states of the same winding number ω but different symmetry. A piecewise parabolic V with continuous first derivative can yield sliding states (invariant circles) with rational ω.

I. INTRODUCTION

In recent years there has been an extensive study of phases and phase transitions in systems in which there are two natural length scales, such as the spacing between atomic planes and the pitch of the helix in a system with helical magnetization, whose ratio changes as a function of temperature or some other parameter and can "lock" onto a rational number corresponding to a "commensurate" state, or pass more or less smoothly through a series of values associated with an "incommensurate" state.[1–4]

As usual in theoretical physics, one way to study these systems is to construct mathematical models which are sufficiently complex to mimic the phenomena of interest while also sufficiently simple to allow an exact analysis or a well-defined series of approximations. One of the simplest models of this type is the ground state of an infinite one-dimensional system of atoms with energy

$$H=\sum_n [V(u_n)+W(u_n-u_{n-1})] , \tag{1.1}$$

where u_n is the position of the nth atom, V is a periodic potential of period 1,

$$V(1+u)=V(u) , \tag{1.2}$$

and W is a potential energy linking nearest neighbors; for example,

$$W(y)=\tfrac{1}{2}(y-\gamma)^2 \tag{1.3}$$

is often employed. The two length scales are the period of V and the distance γ between atoms in the ground state if V were equal to zero. These compete in determining the average separation between neighboring atoms, or "winding number"

$$\omega=\langle u_{n+1}-u_n \rangle , \tag{1.4}$$

which can be a complicated function of γ and the amplitude K of the potential V. For example, it is believed that ω as a function of γ will for a typical V exhibit a "devil's staircase" in which each rational value of ω is locked in over some finite interval of γ values.[5–7]

This model, along with various generalizations, has been discussed extensively in the literature. Some of the more recent work is described in Refs. 6 through 24. Finding its ground state is not easy, and the only case we are aware of where this has been done analytically is when V is a "scalloped" potential of repeating parabolas.[7,24–26] A common attack on the problem is to employ the force equilibrium equations

$$\partial H/\partial u_n=0 \tag{1.5}$$

to generate a nonlinear area-preserving map, and then to study this mapping numerically or to find general theorems applying to such maps. The map obtained using (1.3) and V a constant times $-\cos(2\pi u)$ is often called the "standard map" and has been studied extensively.[10,14–18,27–35] However, the fact that (1.5) also holds for metastable and unstable states means that the mapping problem is in some sense more complex than the ground-state problem, as Aubry[6,10] has emphasized. Whereas it is certainly possible to make a numerical comparison of the energies of different orbits for this map, there seems to be no other general procedure which will distinguish the ground state from other equilibrium states.

In order that the mapping corresponding to (1.5) be single valued, it is necessary that W be a strictly convex function. Aubry[6–10] has used the convexity of W and certain (not very restrictive) conditions on V to prove a number of properties about the ground state of (1.1) even when it cannot be obtained explicitly. [Mather[32,33] has obtained some of these results (and others) independently.] He has then used these properties in a numerical approach to finding the ground state.[9]

In this paper we present yet another approach to studying the ground state of (1.1) using an "effective potential" satisfying a nonlinear eigenvalue equation derived in Sec.

III below. (A preliminary account has appeared elsewhere.[36]) Its main advantage is that unlike approaches based on (1.5) it singles out the ground state, and unlike Aubry's procedures it works for both convex and nonconvex W. Its main disadvantage is that one must solve a functional equation, in general by numerical methods. In actual applications we have found it useful to combine the effective potential method with other numerical procedures, as noted in Sec. VI of this paper.

The effective potential arises rather naturally as the zero-temperature limit of an appropriate transfer operator ("matrix"), and in this form it has appeared earlier in the literature.[37] So far as we know, however, ours is the first attempt to exploit the properties of the eigenvalue equation in a systematic way to study the properties of the ground state of (1.1). This general type of eigenvalue problem arises in other fields of applied mathematics, and the book of Cuninghame-Green[38] is a useful reference. While his results do not cover our specific problem (which involves continuous functions), they are a useful guide and do apply directly to the discrete approximations which we use in numerical studies.

The effective potential approach can be applied to various generalizations of (1.1). From an abstract point of view it is convenient to write

$$H=\sum_n K(x_{n+1},x_n)\ , \tag{1.6}$$

where K is a continuous real-valued function whose arguments reside in a compact metric space. This form includes, for example, the mean-field approximation to the axial next-nearest-neighbor Ising (ANNNI) model,[17,39,40] with the x_j being points inside a two-dimensional square. While some of the general definitions in this paper employ (1.6), all of the applications presented in this paper are restricted to (1.1), i.e.,

$$K(u',u)=V(u')+W(u'-u)\ . \tag{1.7}$$

(Section II shows how to deal with the fact that the variables u and u' are not in a compact space.)

An outline of the rest of the paper is as follows. The appropriate notion of "ground state" for an infinite system is developed in Sec. II. In Sec. III we derive the minimization eigenvalue equation and discuss some of its properties. Associated with this equation is a (one-dimensional) map, some of whose properties are discussed in Sec. IV. This section also shows how to calculate the excitation energy of some simple defect structures. Our numerical procedure for solving the eigenvalue problem is the subject of Sec. V. Sections II and IV are somewhat technical and can be omitted in a first reading.

Several specific cases involving a harmonic W, (1.3), are considered in Sec. VI: V a cosine potential, a cosine potential perturbed by adding a second harmonic, and a piecewise parabolic potential with continuous first derivative. The last two cases yield phase diagrams with transitions between states with the same winding number but different symmetry, analogous to those found in Ref. 24 in a finite electric field.

II. MINIMUM-ENERGY (ENTHALPY) STATES AND GROUND STATES

We are interested in the states of lowest energy of (1.1) or (1.6) for an infinite chain, $-\infty<n<\infty$, and some care is needed in the definition since the total energy is in general infinite. Following Aubry,[10] we call $\{u_n\}$ a "minimum-energy configuration" if for any $p<q$,

$$\sum_{j=p}^{q-1}[V(u_{j+1})+W(u_{j+1}-u_j)] \tag{2.1}$$

will remain constant or increase if $u_{p+1},u_{p+2},\ldots,u_{q-1}$ are altered in any way, with u_p and u_q fixed. That is to say, the change in energy in altering a finite number of u's is always positive or zero.

Next define a "minimum-enthalpy configuration" as one in which for any choice of $p<q$, (2.1) can only increase or remain constant if $u_{p+1},\ldots,u_{q-1}$ are altered by any amount *and* u_q is changed by adding an integer (positive, negative, or zero). One can visualize this as a change in which the qth atom and all its neighbors to the right, i.e., all u_n for $n\geq q$, shift their positions by the same integer amount. Given that V is periodic and W depends only on differences, the change in (2.1) can plausibly be said to represent the change in energy in such a situation. Allowing changes of this type is roughly analogous to employing a constant pressure ensemble in statistical mechanics, which suggests the term enthalpy.

In the minimum-enthalpy condition, W in (1.1) can be replaced by

$$W^*(y)=\min_m W(m+y)\ , \tag{2.2}$$

assuming (as we shall) that the minimum over all integers m exists. The reason is that if for any nearest-neighbor separation u_p-u_{p-1}, W has a value greater than W^*, the energy (enthalpy) can be lowered by shifting all the u_n with $n\geq p$ by the same integer amount. Thus a minimum-enthalpy state relative to W is also one relative to W^*, and the converse is also true, since a minimum-enthalpy state for W^* can be changed into one for W by altering each nearest-neighbor distance by a suitable integer.

It follows from (2.2) that W^* is periodic,

$$W^*(1+y)=W^*(y) \tag{2.3}$$

and hence given any minimum-enthalpy configuration $\{u_n\}$ for W^*, another can be constructed by adding an arbitrary (n-dependent) integer to each u_n. Thus in searching for such configurations one may assume that every u_n belongs to the interval

$$0\leq u_n<1 \tag{2.4}$$

which should be thought of as a circle (1 the same point as 0). This circle is a compact metric space with the obvious choice of metric,

$$d(u',u)=\min\{|u'-u|,|u'-u-1|,|u'-u+1|\}\ . \tag{2.5}$$

For the more general case (1.6), a *minimum-energy config-*

uration is simply a configuration $\{x_n\}$, $-\infty < n < \infty$, such that for every $p < q$,

$$\sum_{j=p}^{q-1} K(x_{j+1},x_j) \tag{2.6}$$

remains constant or increases if $x_{p+1},x_{p+2},\ldots,x_{q-1}$ are altered in any way while x_p and x_q are held fixed. The previous discussion referring to (1.1) agrees with this if we set

$$K(y,x)=V(y)+W^*(y-x)\ , \tag{2.7}$$

except that the minimum-energy configuration is now what we previously called a minimum-enthalpy configuration.

As in the general case it is more natural to think of K as an energy rather than an enthalpy, and this is the common terminology in lattice models where n labels the lattice site, we shall hereafter employ the term *minimum-energy configuration* to refer to what we earlier termed a minimum-enthalpy configuration for (1.1).

In what follows, we shall always assume that D is a compact metric space and K is a continuous (real-valued) function on $D\times D$. [Note that (1.1) with W replaced by W^*, the u_n confined to the circle (2.4), and K given by (2.7) satisfies this continuity condition if both V and W^* are continuous, and that in most cases of interest—in particular for W given by (1.3)—the continuity of W^* is implied by the continuity of W.] One can then show that the energy per particle

$$\lambda=\lim_{p\to-\infty}\lim_{q\to\infty}(q-p)^{-1}\sum_{j=p}^{q-1}K(x_{j+1},x_j) \tag{2.8}$$

exists for any minimum-energy configuration $\{x_j\}$ and is independent of which minimum-energy configuration is considered, i.e., λ depends only on K. A *ground-state* configuration $\{x_n\}$ is defined in terms of

$$\bar{K}(y,x)=K(y,x)-\lambda \tag{2.9}$$

by the condition that for any $p<q$,

$$\sum_{j=p}^{q-1}\bar{K}(x_{j+1},x_j)=\Delta(x_q,x_p)\ , \tag{2.10}$$

where

$$\Delta(x',x)=\inf\left[\sum_{j=0}^{l}\bar{K}(y_{j+1},y_j):y_0=x,y_l=x'\right] \tag{2.11}$$

and l is allowed to be any positive integer, and the $y_1,y_2,\ldots,y_{l-1}$ can be chosen arbitrarily. That is to say, $\{x_n\}$ is a ground state provided the energy of any finite segment (from p to q) expressed in terms of $\bar{K}$ cannot be lowered by replacing this with another segment of (possibly) different length which has the same end points (x_p and x_q). As $\bar{K}$ differs from K by a constant, it is clear that a ground state is a minimum-energy configuration, but the converse is in general not true.

Given any configuration $\{y_n\}$, its *excitation energy* Λ relative to the ground state is defined by the formula

$$\Lambda(\{y_n\})=\lim_{p\to\infty}\left[\sum_{j=-p}^{p-1}\bar{K}(y_{j+1},y_j)-\Delta(y_p,y_{-p})\right]. \tag{2.12}$$

The quantity in large parentheses is obviously non-negative, and can be shown to be monotone nondecreasing in p. Hence the limit exists, though it may be $+\infty$. A configuration is a ground state if and only if $\Lambda=0$. States which are minimum-energy configurations but not ground states typically contain a defect[41–44] (sometimes called a "domain wall," "soliton," "kink," "discommensuration") which cannot be eliminated by displacing a finite number of atoms. For such configurations Λ may be thought of as the defect energy.

A configuration $\{u_n\}$ is *periodic* provided there are integers P and Q such that for n any integer,

$$u_{n+Q}=P+u_n\ . \tag{2.13}$$

The smallest positive Q for which this equation holds is the *period,* and the winding number is

$$\omega=P/Q\ . \tag{2.14}$$

In the more general case of a configuration $\{x_n\}$ of elements drawn from D, (2.13) is replaced by

$$x_{n+Q}=x_n\ , \tag{2.15}$$

and the winding number ω is in general undefined.

A periodic minimum-energy configuration $\{x_n\}$ is necessarily a ground state, and λ, (2.8), can be obtained by averaging over a cycle,

$$\lambda=Q^{-1}\sum_{j=0}^{Q-1}K(x_{j+1},x_j)\ . \tag{2.16}$$

Consequently

$$\sum_{j=0}^{Q-1}\bar{K}(x_{j+1},x_j)=0 \tag{2.17}$$

and from this one can infer the result that for any p and q,

$$\delta(x_p,x_q)=0\ , \tag{2.18}$$

where δ is defined by

$$\delta(x',x)=\Delta(x',x)+\Delta(x,x')\ , \tag{2.19}$$

and is always a non-negative quantity.

For later purposes it is useful to define a *pure ground state* $\{x_n\}$ as one in which (2.18) holds for any p and q. Such a state may or may not be periodic, and there can be ground states which are not pure ground states. What we term a pure ground state seems to be much the same thing as Aubry's ground state,[10] though we have not studied this in detail, and as his approach uses a somewhat different starting point and technical conditions, a direct comparison is not simple.

III. MINIMIZATION EIGENVALUE EQUATION

A. Derivation of the equation: Effective potentials R, S, and F

Consider the problem of finding the ground state for a system of atoms with interactions given by (1.1). As a first step in the derivation of the eigenvalue equation (3.3) below it is useful to consider a system of only two particles, the first moving in a periodic potential

$$U(1+u)=U(u) \tag{3.1}$$

and the second in the potential V, with the two connected by a spring with potential energy W, see Fig. 1. For a fixed position u' of the second atom, the minimum energy of the combined system is

$$\overline{U}(u')=\mathscr{K}U(u')=V(u')+\min_u[W(u'-u)+U(u)]\ , \tag{3.2}$$

where the right-hand side defines the action of the nonlinear functional transformation $\mathscr{K}$ acting on U.

In the particular case in which $U=V$, $\mathscr{K}V(u')$ is the minimum energy of two atoms in a common potential V, connected by a spring, provided the second atom is fixed at u'. Similarly, if the functional transformation is iterated n times, $\mathscr{K}^nV$ is the minimum energy of a chain of n atoms in the potential V provided the right-most atom is at a specified position. Note that $\mathscr{K}^nV$ has the same periodicity as V. Thus as n tends to infinity it is plausible that V and $\mathscr{K}^nV$ differ by $n\lambda$, where λ is the ground-state energy per particle, plus a term of order 1.

Rather than looking at iterates of $\mathscr{K}$, it is convenient to search for a solution to the nonlinear eigenvalue equation

$$\mathscr{K}R(u')=\lambda+R(u')=V(u')+\min_u[W(u'-u)+R(u)]\ , \tag{3.3}$$

where we impose the requirement that R have the same periodicity as V,

$$R(1+u)=R(u)\ . \tag{3.4}$$

We shall call R the "effective potential." It is at once clear from (3.3) that

$$\mathscr{K}^nR=n\lambda+R\ , \tag{3.5}$$

and thus R has the following property. If a finite chain of n atoms connected by springs is placed in a potential V, but the left-most atom is in a potential R rather than V, the minimum energy of the entire system as a function of the position of the right-most atom is given by $R+n\lambda$.

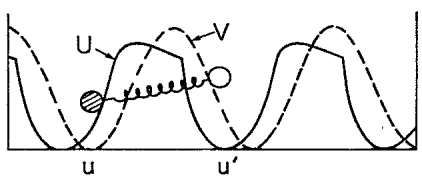

FIG. 1. Potentials $U(u)$ and $V(u')$.

Assuming that R is a bounded function, the energy per particle of such a system will tend to λ as n goes to infinity, and thus λ should be the average energy per particle in any ground state, since placing the first atom in a potential R rather than V only changes the total energy by a term of order 1.

One can think of $R(u)$ as the effective potential acting on or "seen by" the right-most atom in a semi-infinite chain, Fig. 2(a), when it is fixed at a position u and the other atoms are allowed to relax to a state of minimum energy (or enthalpy). Hence $R'(u)$ is the force which must be applied (externally) to the end atom to hold it at position u. This effective potential can only be defined up to an additive constant, consistent with the fact that adding a constant to any solution R of (3.3) yields another solution.

There is a corresponding effective potential $S(u)$ for the left-most atom in a semi-infinite chain extending to the right, Fig. 2(b), which satisfies the equation

$$\lambda+S(u)=V(u)+\min_{u'}[S(u')+W(u'-u)] \tag{3.6}$$

and the periodicity condition

$$S(1+u)=S(u)\ . \tag{3.7}$$

An atom in a doubly infinite chain, Fig. 2(c), will then experience an effective potential

$$F(u)=R(u)+S(u)-V(u)\ , \tag{3.8}$$

where V must be subtracted on the right-hand side to avoid counting it twice. In particular one expects F to take its minimum value whenever u is a point in some ground-state configuration. The value of F at other points can be useful in calculating defect energies, see Sec. IV D. [In cases in which (3.3) or (3.6) have multiple solutions for R and S, apart from the trivial addition of a constant, the effective potential F will depend in a nontrivial way on the choice of R and S on the right-hand side of (3.8).]

B. Minimization eigenvalue equation as the zero-temperature limit of a transfer operator or "matrix"

An alternative derivation for (3.3) begins with a finite temperature Gibbs distribution for a system with Hamiltonian (1.1), assuming that the u_n are bounded so that integrals exist. As is well known, the statistical properties of such a state can be obtained using a transfer operator (or matrix) which can be chosen in the form

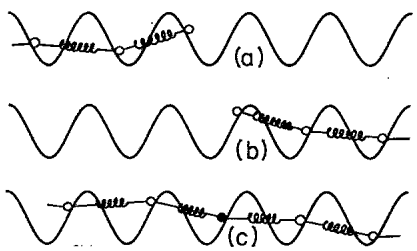

FIG. 2. Sketches showing the intuitive significance of the effective potentials: (a) $R(u)$; (b) $S(u)$; (c) $F(u)$.

$$\hat{T}(u',u)=\exp\{-\beta[V(u')+W(u'-u)]\}\ , \tag{3.9}$$

where β is the inverse temperature. In particular, all the thermodynamic properties are determined by the largest eigenvalue which, by the Perron-Frobenius theorem, is real and positive and corresponds to an eigenfunction which (apart from a constant factor) is also real and positive. Hence the linear eigenvalue problem for this particular eigenvalue can be written in the form

$$\int \hat{T}(u',u)e^{-\beta R(u)}du = e^{-\beta\lambda}e^{-\beta R(u')}\ . \tag{3.10}$$

Here λ and R depend on β, but it is plausible that as $\beta\rightarrow+\infty$ (zero temperature) they tend to finite limits. If we suppose that in this limit the integral in (3.10) can be approximated by the maximum value of the integrand, the result is (3.3). A similar line of reasoning employing the left rather than the right eigenfunction of $\hat{T}$ yields (3.6), and indicates that for finite but large β, $\exp[-\beta F(u)]$ is (approximately) proportional to the probability of finding a particle at position u.

While this approach to obtaining (3.3) is helpful in some ways, and has been discussed in the previous literature,[37] it is not easy to make it mathematically rigorous, and even for a simple intuitive understanding it is often simpler to deal with (3.3) and (3.6) directly.

C. Geometrical construction for the minimization transformation

The central nonlinear element in (3.3) is the minimization operation

$$\hat{R}(u')=\mathscr{W}R(u')=\min_{u}[W(u'-u)+R(u)]\ . \tag{3.11}$$

There are two geometrical perspectives which are helpful in thinking about this.[45] The first consists in noting that for a fixed u, the graph of $W(u'-u)+R(u)$ *as a function of* u' is simply the graph of the function $W(u')$ displaced horizontally (by an amount u) and vertically [by $R(u)$]. Then the right-hand side of (3.11) is the minimum or lower envelope of these displaced graphs as illustrated in Fig. 3, where the dashed curves are the original and displaced graphs of W, and the solid line is the function $\hat{R}$. From this construction it is evident that if W is a continuously differentiable function, $\hat{R}$ may possess an "upward kink" where its first derivative decreases discontinuously, as in Fig. 4(a), but it cannot have a "downward kink," Fig. 4(b), where its first derivative increases discontinuously.

The second perspective employs

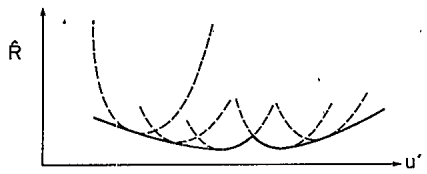

FIG. 3. Function $\hat{R}$ is a lower envelope of translations of W (dashed curves).

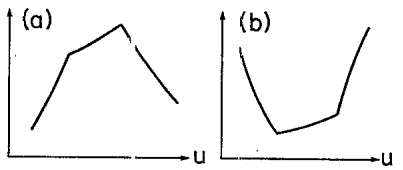

FIG. 4. Functions with (a) two upward kinks and (b) two downward kinks.

$$w(u)=-W(-u)\ , \tag{3.12}$$

a function whose graph is obtained from that of W by a 180° rotation about the origin (Fig. 5). The function $\hat{R}(u')$ can then be constructed by displacing the graph of w horizontally by an amount u' and then vertically until it touches but does not cross the graph of R, as in Fig. 6. This vertical displacement is the value of $\hat{R}(u')$. To see that this construction is correct, note that (3.11) is equivalent to

$$\hat{R}(u')+w(u-u')\leq R(u)\ , \tag{3.11a}$$

with equality for at least one value of u. The left-hand side of this inequality as a function of u is a formula for the displaced graph of w. Also note that if the point where w crosses the ordinate is regarded as a fiducial point rigidly attached to the graph of w (the solid circle in Figs. 5 and 6), its locus for all the conditions in which the displaced graph of w touches that of R in the manner described above is the graph of $\hat{R}(u')$, apart from an additive constant.

D. Some properties of the effective potentials

Let us assume that V and W are continuous functions of their arguments, V is periodic, (1.2), and W^* in (2.2) is defined (ie., the minimum exists for all y) and continuous. Then it can be shown that there is always some continuous and periodic, in the sense of (3.4), function R which satisfies (3.3), and the corresponding λ is unique (for a given V and W). Similarly there is at least one continuous function S satisfying (3.6) and (3.7), and the (unique) λ is the same as in (3.3). Furthermore, as long as one is only considering periodic solutions of (3.3) and (3.6), W may be replaced by W^*.

Even if V and W have continuous first derivatives, the first derivatives of R and S need not be continuous. (For an example, see Sec. VI A.) However, under this condi-

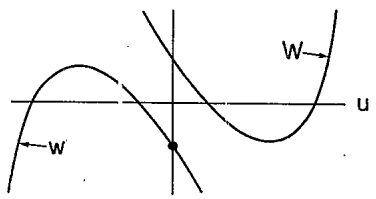

FIG. 5. Graph of w is obtained from that of W by a 180° rotation in agreement with Eq. (3.12).

tion R and S *cannot* have downward kinks, as in Fig. 4(b), as explained above in Sec. III C.

It is obvious that adding a constant to some R satisfying (3,3) yields an equally good solution. However, there may exist multiple solutions which are not related to each other in this trivial manner, and which may be thought of as corresponding to different possible "coexisting phases." The situation has been analyzed in some detail by Cuninghame-Green[38] for the finite matrix analog of (3.3) and (3.6), or their generalizations (3.16) and (3.17) in Sec. III E. A complete discussion is outside the scope of the present paper, but the following remarks will be needed for some of the applications in Sec. VI.

If there can be multiple solutions to (3.3), then the general solution is of the form

$$R(u)=\min_{\alpha}[R_{\alpha}(u)+C_{\alpha}]\ , \tag{3.13}$$

where the $R_{\alpha}(u)$, $\alpha=1,2,\ldots$ are "basis functions" corresponding to different "pure phases," and the C_{α} are real constants. [It is conceivable that an infinite number of pure phases will be present, in which case the minimum in (3.13) should be replaced by an infimum over a suitable index set.] The same remarks about multiple solutions apply to S, and whenever there are multiple solutions to (3.3) there will be multiple solutions to (3.6), and vice versa.

In the case of a symmetrical V,

$$V(u)=V(-u)\ , \tag{3.14}$$

a solution to (3.6) can be constructed from a solution to (3.3), or vice versa, by setting

$$S(u)=R(-u)\ . \tag{3.15}$$

[When there are multiple solutions to either equation, (3.15) need not necessarily yield the S (or the R) which is relevant to a particular ground-state configuration under study.]

E. Minimization equation for general K

The generalizations of (3.3) and (3.6) to the case of H of the form (1.6) are

$$\lambda+R(y)=\min_{x}[K(y,x)+R(x)]\ , \tag{3.16}$$

$$\lambda+L(x)=\min_{y}[L(y)+K(y,x)]\ , \tag{3.17}$$

where R and L are right and left eigenvectors of K, and x and y are drawn from a common domain D. The counterpart of (3.8) is

$$F(x)=R(x)+L(x)\ . \tag{3.18}$$

The case where x and y take on only a finite number of values, so that K is a square matrix, has been discussed at length by Cuninghame-Green.[38] In particular, he has demonstrated the existence of left and right eigenvectors and the uniqueness of the eigenvalue, and has provided a constructive method for obtaining the eigenvalue and all of the eigenvectors. The generalization of his work which is most useful for our purposes is that in which D is a compact metric space and K is a continuous real-valued function on $D\times D$. One can then show that (3.16) and (3.17) possess continuous solutions R and L, and that λ is the same in both equations and unique. By using (2.7) and by defining

$$S(u)=L(u)+V(u) \tag{3.19}$$

one can show that (3.3) and (3.6) are indeed of the form (3.16) and (3.17).

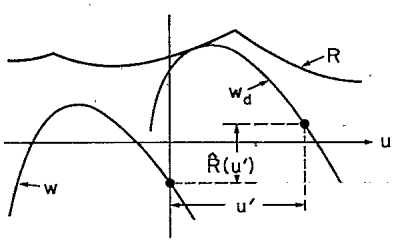

FIG. 6. Geometrical construction for $\hat{R}(u')$ using the displaced graph w_d of w.

IV. MAPS AND ORBITS ASSOCIATED WITH EFFECTIVE POTENTIALS

A. Maps τ and σ

Given a solution R to (3.3), the map $\tau(u')$ associated with R is the set of points u where the minimum on the right-hand side is actually achieved. There is at least one such point but there may be more than one, so it is convenient to think of τ as mapping points of the real line onto nonempty subsets. Similarly, $\sigma(u)$ denotes the set of points u' where the minimum on the right-hand side of (3.6) is achieved, given some solution S. In terms of Fig. 2(a), τ gives the position of the atom just to the left of any given atom in terms of the latter's location; and similarly in the situation in Fig. 2(b), σ gives the position of the atom to the right of any atom as a function of the location of the latter. When there is more than one solution to (3.3) (in a nontrivial sense) τ depends on which solution is employed, and of course the same comment applies to σ and the solutions of (3.6). Even when these solutions are unique, τ is not in general the inverse of σ.

The periodicity of R, (3.4), and of S, (3.7), implies that

$$\tau(1+u)=1+\tau(u)\ , \tag{4.1}$$

$$\sigma(1+u)=1+\sigma(u)\ . \tag{4.2}$$

If in addition $V(u)$ is $V(-u)$ and (3.15) applies, then

$$\tau(u)=-\sigma(-u)\ . \tag{4.3}$$

(Note that the left and right sides of these equations are sets.)

In the case of a general K, τ and σ are defined in the same manner: $\tau(y)$ as the set of points where the minimum in (3.16) is achieved, and $\sigma(x)$ the corresponding set in (3.17).

B. Half orbits and orbits

The infinite sequence of points $\{u_n\}$ with $n=p,p-1,p-2,\ldots$ (the numbering is chosen to increase

from left to right in Fig. 2) is an *R half orbit* provided

$$u_{n-1}\in\tau(u_n) \tag{4.4}$$

or, equivalently,

$$\lambda+R(u_n)=V(u_n)+W(u_n-u_{n-1})+R(u_{n-1}) \tag{4.5}$$

for all $n\leq p$. Similarly $\{u_n\}$, $n\geq q$, is an S half orbit if

$$u_{n+1}\in\sigma(u_n) \tag{4.6}$$

or, equivalently,

$$\lambda+S(u_n)=V(u_n)+S(u_{n+1})+W(u_{n+1}-u_n)\ . \tag{4.7}$$

Given a doubly infinite sequence $\{u_n\}$, we shall call it an *R orbit* if (4.4) or (4.5) holds for all n, $-\infty<n<\infty$, and an *S orbit* if (4.6) or (4.7) holds for every integer. The same terminology for orbits and half orbits applies for a general K, except that S is replaced by L, see (3.17), and (4.5) and (4.7) are appropriately modified. Note that if there is more than one solution to (3.3), in a nontrivial sense, the R orbit(s) will in general depend on which R is chosen.

Given our basic assumption that D is compact and K is continuous (the latter is implied by the continuity of V and W^*), one can show that an R orbit is always a ground state, and that every *pure* ground state is an R orbit, for every R satisfying the minimization eigenvalue equation. The same statement holds for L or S orbits. Furthermore there is always at least one pure ground state. Given any ground state, there is some solution R to (3.3) or (3.16) such that this ground state is an R orbit.

By combining (4.5) and (4.7) one sees that F is constant at all points on a configuration which is both an R and an S orbit, and consequently on the points of any pure ground state. When R and S are unique up to additive constants ("single phase"), this constant is the minimum value of F, and F is strictly larger than its minimum at any point which is not part of a ground state. Thus, in particular, if there is only one phase and if F is a constant, every point has the property that there is at least one ground-state configuration which includes it. This can arise if there is a sliding state (Sec. VI), but in some other cases as well.

Half orbits (R, S, or L), in contrast to orbits, can be begun at any initial point u_0. One's intuition, see Figs. 2(a) or 2(b), suggests that as $|n|\rightarrow\infty$, the half orbit will tend to look more and more like an orbit, and indeed this can be proved, provided "tend" is understood in a suitable weak sense. In our numerical studies (Sec. V), we use iterations of τ to find the ground-state configuration.

C. Special properties of convex W

Let us assume that V and W in (3.3) are continuously differentiable functions, that W is strictly convex (i.e., its derivative is a monotone strictly increasing function), and that $W(y)$ tends to $+\infty$ as $|y|\rightarrow\infty$ Then one can show that $u<u'$ implies that

$$\tau(u)<\tau(u')\ , \tag{4.8}$$

i.e., all elements in the left set are strictly smaller than those in the right. This condition together with (4.1) is all that is needed to demonstrate the existence of a unique winding number ω, (1.4), by an easy extension of the standard argument.[46]

The relationship of R orbits and ground states discussed previously (Sec. IV B) then implies that every ground state has a well-defined winding number (since it is an R orbit for some R), and if there are several distinct ground states for a given V and W, they all have the same winding number (equal to that of some pure ground state which is an R orbit for every R). It is also clear that half orbits have the same winding number as the ground state. [Note that σ satisfies (4.8), and pure ground states are also S or L orbits.] We suspect that minimum-energy configurations have the same winding number as the ground state(s) for the same V and W, but have not established this. Note that Aubry[10] has established the existence of a winding number under somewhat different technical conditions.

In the special case where R has a continuous first derivative (under the conditions given above for V and W), $\tau(u)$ is an ordinary single-valued, continuous, and monotone strictly increasing function. Similarly, if S has a continuous first derivative, $\sigma(u')$ has these same properties.

D. Excitation energies for certain defects

A knowledge of the effective potentials R, S, and F makes it possible to compute the excitation energy Λ, see (2.12), of certain types of defects. We shall define a type-I defect as a configuration $\{\bar{u}_n\}$ in which there is some l such that the semi-infinite configuration with $n\leq l$ is an R half orbit, and that with $n\geq l$ is an S half orbit. An example is shown in Fig. 7(b). Note that $\bar{u}_l$ is the end point of both half orbits. In a type-II defect, as in Fig. 8(b), the atoms with $n\leq l$ constitute an R half orbit and those with $n\geq l+1$ an S half orbit, so that the end points $\bar{u}_l,\bar{u}_{l+1}$ of the two half orbits are nearest neighbors.

The defect energy can be computed if we assume, as suggested in Figs. 7 and 8, that as $|n|\rightarrow\infty$ both half orbits approach the positions of the atoms belonging to a single ground state $\{u_r\}$. That is to say, for p sufficiently large, and with an appropriate renumbering of the atoms if necessary,

$$\bar{u}_{-p}=u_{-p};\quad \bar{u}_{p+m}=u_p\ , \tag{4.9}$$

with a negligible error. Here m is the number of atoms added to form the defect (m can be negative). Then using (2.10) to (2.12), we see that Λ will be given by

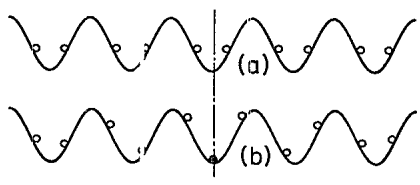

FIG. 7. Type-I defect (b) obtained by adding an atom to the ground state in (a). The solid circle is the atom $\bar{u}_l$.

$$\Lambda=\sum_{j=-p}^{p+m-1}\bar{K}(\bar{u}_{j+1},\bar{u}_j)-\sum_{j=-p}^{p-1}\bar{K}(\bar{u}_{j+1},\bar{u}_j)\ . \tag{4.10}$$

Suppose the defect is type I with the end atom common to both half orbits at $l=0$. Then for $j\le 0$, we have

$$R(\bar{u}_j)=\bar{K}(\bar{u}_j,\bar{u}_{j-1})+R(\bar{u}_{j-1})\ , \tag{4.11}$$

and for $j\ge 0$,

$$L(\bar{u}_j)=L(\bar{u}_{j+1})+\bar{K}(\bar{u}_{j+1},\bar{u}_j)\ . \tag{4.12}$$

Note in addition that the ground state $\{u_n\}$ is an R orbit, so that

$$\begin{aligned}\Lambda&=R(\bar{u}_0)-R(\bar{u}_{-p})+L(\bar{u}_0)-L(\bar{u}_{p+m})\\&\quad-[R(u_p)-R(u_{-p})]=F(\bar{u}_0)-F_m\ ,\end{aligned} \tag{4.13}$$

where we have used (3.18), and F_m denotes the value of F everywhere on the ground state, in particular at $u_p=\bar{u}_{p+m}$. An analogous calculation yields the expression

$$\begin{aligned}\Lambda&=R(\bar{u}_0)+L(\bar{u}_1)+\bar{K}(\bar{u}_1,\bar{u}_0)-\lambda-F_m\\&=R(\bar{u}_0)+S(\bar{u}_1)+W^*(\bar{u}_1-\bar{u}_0)-\lambda-F_m\end{aligned} \tag{4.14}$$

for a type-II defect, where $\bar{u}_0$ and $\bar{u}_1$ are the end points of the R and S half orbits, respectively. [One can also derive (4.13) and (4.14) by imagining adding or deleting a particle, reconnecting the springs, and letting the other particles relax. However, the relaxation may produce unexpected results, which is why we prefer the derivation given above.]

Note that (4.13) and (4.14) do not by themselves specify the values of $\bar{u}_0$, or $\bar{u}_0$ and $\bar{u}_1$. One can sometimes use symmetry in order to make a plausible guess, but in general the minimal-energy excitation for a given m must be found by minimizing Λ with respect to $\bar{u}_0$ (and $\bar{u}_1$ for type II) subject to the constraint that the configuration of atoms is of the appropriate sort—in particular, that they form suitable half orbits.

An important special case is that in which W is a harmonic potential (1.1), or

$$W(u'-u)=\tfrac{1}{2}(u'-u)^2-\gamma(u'-u)\ , \tag{4.15}$$

ignoring the $\gamma^2/2$, which does not affect Λ. One frequently finds (Sec. VI) that for a given V, ω is constant over a range of γ values and within this range the ground-state orbit $\{u_n\}$ is independent of γ. [This is plausible because for this W, γ does not appear in the area-preserving map based on the equilibrium equations (1.5).] If we assume the same is true for one of the defect states discussed above, then over this range of γ values

$$d\Lambda/d\gamma=m\omega\ , \tag{4.16}$$

where m is the number of additional atoms [see (4.9)]. The derivation is as follows: The quantity $\bar{K}$ depends on γ both through (4.15) and because λ depends on γ. Indeed, if the ground-state configuration $\{u_n\}$ does not depend on γ, (4.15), which omits a $\gamma^2/2$ term, implies that $d\lambda/d\gamma$ is $-\omega$. The dependence of Λ on γ can then be found from (4.10). The γ-dependent contributions to the two sums from the W part of $\bar{K}$ exactly cancel because of (4.9), but the dependence of $\bar{K}$ on λ does not cancel for $m\neq 0$, and this yields (4.16).

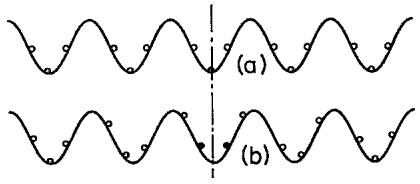

FIG. 8. Type-II defect (b) obtained by adding an atom to the ground state in (a). The two solid circles denote the atom $\bar{u}_l$ and $\bar{u}_{l+1}$.

V. NUMERICAL METHOD FOR SOLVING THE MINIMIZATION EIGENVALUE EQUATION

The minimization eigenvalue equation (3.3) can sometimes be solved analytically (see examples in Sec. VI D), but in general must be attacked numerically. We do this by first restricting u and u' to a set of N (≈ 100) discrete values uniformly distributed on the interval [0,1): $u=i/N$, with $i=0,1,\ldots,N-1$. The resulting finite matrix problem can be solved (in principle) by the methods of linear programming, see the discussion in Chap. 25 of Ref. 38, but we have instead employed a minor modification of the following iteration procedure.

Let $R^{(0)}(u)$ be any function (e.g., $R^{(0)}=V$), and define $R^{(j)}$ iteratively by means of the equation

$$R^{(j+1)}(u)=\tfrac{1}{2}[\mathscr{K}R^{(j)}(u)+R^{(j)}(u)]-C_j\ , \tag{5.1}$$

where C_j is a constant chosen so that the minimum value of $R^{(j+1)}$ is zero. The iteration is continued until

$$\max_u\,|\,\overline{\mathscr{K}}R^{(j)}(u)-R^{(j)}(u)\,| \tag{5.2}$$

is less than some specified value, typically 10^{-4}, where $\overline{\mathscr{K}}$ is defined in terms of $\mathscr{K}$ by

$$\overline{\mathscr{K}}R(u)=\mathscr{K}R(u)-\min_u\mathscr{K}R(u)\ . \tag{5.3}$$

This final $R^{(j)}$ (typically 20 to 30 iterates are required) is the numerical approximation for R, and the minimum value of $\mathscr{K}R^{(j)}$ is the numerical approximation for λ. As $V(u)=V(-u)$ for the cases we have considered, we set $S(u)$ equal to $R(-u)$ [see (3.15)]. The value of u for which the minimum is (3.3) actually occurs gives the discrete version of the map $\tau(u')$ (Sec. IV A).

The ground-state orbit is obtained by iterating τ starting at some arbitrary value of u' and continuing until an earlier value is repeated. This cycle yields the period and the winding number. The starting value of u' does not seem to matter except near a phase transition where R can be a "mixture," see (3.13), of functions R_α corresponding to the "pure phases." In this case the starting value of u' determines which of the ground states is reached by iterating τ. An example will be given in Sec. VI C.

We found that a simple iteration scheme, in which the right-hand side of (5.1) is replaced by

$$\mathscr{K}R^{(j)}(u)-C_j\ , \tag{5.4}$$

does not work in general, as it fails to converge. On the other hand, (5.1) always converges, but the rate of convergence varies somewhat. It is particularly slow near a first-order phase transition (the horizontal bars of Sec. VI C and the pinching points of Sec. VI D), or when the final R function has lots of upward kinks, or when the strength K characterizing the potential V, see (6.1), (6.8), and (6.9), is small.

VI. SPECIFIC APPLICATIONS

A. Sinusoidal V and harmonic W (Frenkel-Kontorova model)

1. Solutions to the minimization eigenvalue equation, and the phase diagram

The system (1.1) with

$$V(u)=[K/(2\pi)^2][1-\cos(2\pi u)]\ , \tag{6.1}$$

$$W(y)=\tfrac{1}{2}(y-\gamma)^2\ , \tag{6.2}$$

is often called the Frenkel-Kontorova model.[47] There have been a number of previous studies of the ground state of this model using a variety of methods.[10,14–18,34,35,48] However, only Ying[48] has constructed a phase diagram, which in some respects is almost surely in error.

Our approach was to solve the minimization eigenvalue equation numerically, as explained in Sec. V, for a large number of γ and K values. Typically we used a grid of $N=100$ points. As an example, Fig. 9 shows the effective potential R for the case $\gamma=0.36$, $K=3$, and Fig. 10 shows the corresponding τ. The winding number ω is $\frac{1}{3}$. Note that the function R is continuous, but has upward kinks at the same points where τ is discontinuous. There are three of these discontinuities in every period, but only two of the corresponding kinks in R are readily visible in the figure. The points where τ is discontinuous are also points where it takes multiple values, at the top and bottom of the discontinuity. The appearance of kinks in R and discontinuities in τ is a sign that the ground state is "pinned,"[9,15] i.e., the atoms cannot slide along the potential while remaining in the ground state.

By evaluating ω for different γ and K (see Sec. V), we constructed the phase diagram in Fig. 11. Only a few of the regions corresponding to rational values of ω with small denominators are shown. Presumably there is a separate "tongue"—most of them extremely narrow—for

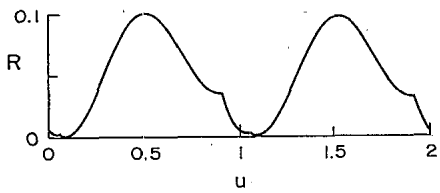

FIG. 9. Effective potential $R(u)$ for $K=3$ and $\gamma=0.36$ in Eqs. (6.1) and (6.2).

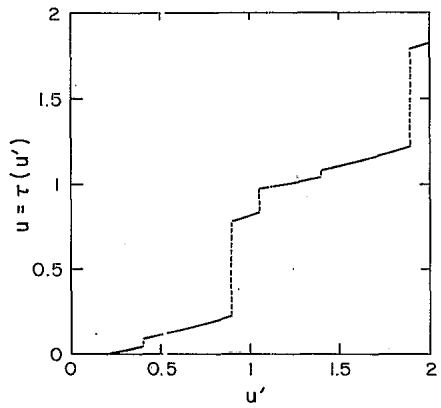

FIG. 10. Mapping $u=\tau(u')$ corresponding to Fig. 9.

every rational value of ω, and these fill up, to some extent, the regions between those whose winding numbers are shown explicitly. The result appears to have the same topological structure which Aubry[25] found in the case of a scalloped potential V (consisting of repeated parabolas), though the dependence of the widths of the tongues on K at small K is clearly different. We did not find any triple points, in contrast to Ying.[48] As a function of γ at fixed K, one expects ω to show a "devil's staircase" behavior, and this is supported by the numerical calculations at $K=1$, shown in Fig. 12.

Our numerical approach was less successful for small values of K: The iteration procedure for R took longer to converge, and the results were less reliable. However, ω was never radically different from what one would suppose to be a reasonable value. (Better precision could of course be obtained with larger values of N.)

2. Defect excitation energies

It is plausible that the edges of a tongue associated with a particular rational winding number ω, Fig. 11, come about when the energy of one of the defect structures of the sort shown in Figs. 7 and 8 goes to zero. Since the corresponding excitation energy Λ varies linearly with γ, (4.16), in the case of a harmonic W, (6.2), it should suffice to evaluate R at a single value of γ for a particular K, within the region where this ω is stable, in order to find both the left and right edges of the tongue at this K.

We employed this procedure for $\omega=0$, $\frac{1}{2}$, and $\frac{1}{3}$, and

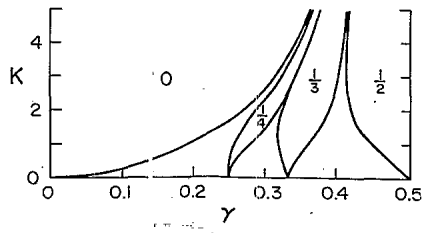

FIG. 11. Phase diagram corresponding to Eqs. (6.1) and (6.2). The numbers are values of the winding number ω. The unlabeled regions contain additional structure as in Fig. 25 of Ref. 1(b).

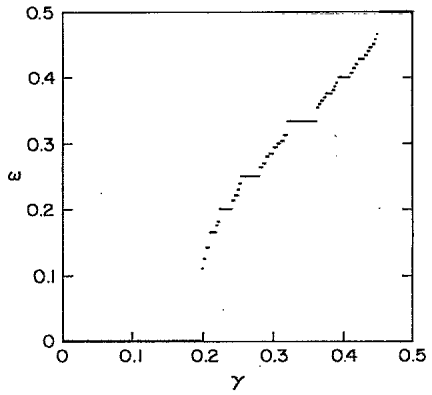

FIG. 12. Winding number ω as a function of γ at $K=1$ (Fig. 11) showing a devil's staircase structure.

found that it worked very well. For $\omega=P/Q$ with Q even, it is plausible that the defect of interest is of type I, see Fig. 7, with the central atom at a potential minimum, so that Λ is easily evaluated using (4.13). When Q is odd, one expects a situation like that in Fig. 8, so it is plausible that $-\bar{u}_0$ is $\bar{u}_1$ in (4.14). One must, however, still minimize Λ with respect to this single parameter, which can be a bit tricky for reasons discussed above in Sec. IV D. The case $\omega=0$, see Fig. 13, turns out to be equivalent to $\omega=1$, i.e., the defect of interest is of type II, and the boundary of the $\omega=0$ region falls at $\gamma=\Lambda$, where Λ is the defect energy evaluated at $\gamma=0$.

The phase boundaries determined in this manner were indistinguishable from those computed previously by the straightforward method of Sec. VI A 1 to within the accuracy of the latter, typically 0.002 in γ. Since the method based on excitation energies requires only one solution of the minimization eigenvalue equation for each K value, for a given ω, it requires less computation than a straightforward search. On the other hand, the latter is not based on any assumption about the nature of the transition at which a particular phase terminates. The fact that the two coincide in the present case provides numerical support for the idea that the limits of stability of the phases with rational ω are determined by "solitons" of the type under discussion.

B. Comments on a special case of nonconvex W

An interesting example of a nonconvex W is provided by the "chiral XY model" in which the angular variable θ_n at site n, $0\le\theta_n<2\pi$, is subject to an anisotropic potential

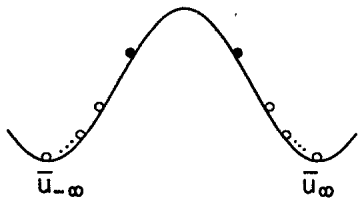

FIG. 13. Type-II soliton defect for $\omega=0$ (schematic).

$$\bar{V}(\theta_n)=-K\cos(p\theta_n)\ , \tag{6.3}$$

with $p>0$ an integer, and the nearest-neighbor interaction is

$$\bar{W}(\theta_{n+1}-\theta_n)=1-\cos(\theta_{n+1}-\theta_n-2\pi\gamma)\ . \tag{6.4}$$

So that the notation conforms to that used elsewhere in this paper we let

$$x_n=p\theta_n/2\pi\ , \tag{6.5}$$

and define

$$V(x_n)=\bar{V}(\theta_n)=V(1+x_n)\ , \tag{6.6}$$

$$W(x_{n+1}-x_n)=\bar{W}(\theta_{n+1}-\theta_n)=W(p+x_{n+1}-x_n)\ . \tag{6.7}$$

In the case $p=1$, K can be thought of as a "magnetic field," and the phase diagram shows important effects due to the nonconvexity of W.[49] From their numerical studies of the case $p=2$, Banerjea and Taylor[21] concluded that only the convex part of W plays a role in determining the ground state (at least for those values of K and γ they considered). We did some numerical studies for $p=3$ and came to the same conclusion.

The fact that only the convex part of W need be considered in calculating the ground state for $p\ge2$ follows from the fact that W, which by (6.7) has period p, can be replaced by W^*, (2.2). The relationship is indicated in Fig. 14 for $p=2$: The smooth curves are the graphs of $W(y)$ and $W(1+y)$, while their minimum $W^*(y)$ is indicated by the heavy line. (These are drawn for $\gamma=0$, but a nonzero γ merely shifts the origin of y.) The latter is convex between the upward kink points, and thus it could have been obtained using (2.2) starting from a convex function W_c equal to W for $-\frac{1}{2}\le y\le\frac{1}{2}$, and defined outside this interval by the dashed lines in Fig. 14. That is, both W and W_c inserted on the right-hand side of (2.2) yield precisely the same W^*, and hence give rise to equivalent ground state problems. The same construction works for any $p\ge2$.

There is, to be sure, one weak point in the above analysis for $p=2$ though not for $p\ge3$: The second derivative of W, and hence also of W_c, actually vanishes at $y=\pm\frac{1}{2}$, so that theorems (such as those of Aubry) which demand that the second derivative of W be bounded below by a positive constant might not apply. Hence it would be nice if one could demonstrate that in a ground

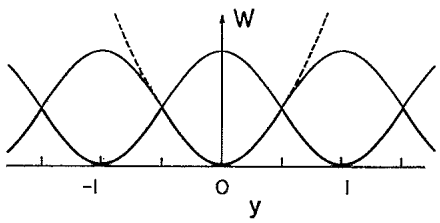

FIG. 14. Heavy curve $W^*(y)$ is the minimum of the smooth curves $W(y)$ and $W(1+y)$. The convex function $W_c(y)$ is obtained by extending $W^*(y)$ for $|y|\ge\frac{1}{2}$ as indicated by the dashed lines.

state $x_{n+1}-x_n$ will always stay a finite distance away from "dangerous" values where W^* has upward kinks.

This can, indeed, be shown to be the case, with an appropriate restriction on V, using the geometrical constructions introduced in Sec. III C. In particular, let w be the function [see (3.12) and Fig. 5] obtained by rotating the graph of W^* in Fig. 14 by 180° about the origin. Then in the construction in Fig. 6, the graph of w cannot make contact with that of R at one of the downward kinks of the former (corresponding to an upward kink of W^*) if the latter has no downward kinks. But if V has no downward kinks, R cannot have a downward kink, since it is, by (3.3), the sum of V and a function $\hat{R}$, (3.11), which lacks downward kinks because of the geometrical construction indicated in Fig. 3. It is not difficult to turn these geometrical arguments into a formal proof as long as K in (6.3) is finite, and thus the second derivative of V is bounded above.

C. Perturbed sinusoidal V and harmonic W

We now consider the effect of adding a small amount of a second harmonic to the potential V of the Frenkel-Kontorova model (6.1)

$$V(u)=[K/(2\pi)^2][1-\cos(2\pi u)+\epsilon(1-\cos(4\pi u))] . \tag{6.8}$$

With $\epsilon \leq 0$, the perturbation is "innocuous" in that the topological properties of the phase diagram remain the same as in Fig. 11. However, as soon as ϵ is positive there are important changes.

Figure 15 shows a portion of the phase diagram for $\epsilon=0.1$. It was constructed in the same manner as Fig. 11, by evaluating the minimization eigenvalue equation at a large number of points in the γ,K plane. The simple tongues corresponding to rational values of ω are now split by a series of horizontal bars at which there is a phase transition between phases of the same winding number but different symmetry. In particular, the phases marked A are similar to those obtained with $\epsilon=0$, in which the ground state may have an atom at the potential minimum [as in Fig. 8(a)], but never an atom at the potential maximum. By contrast, in phase B there is always an atom at the potential maximum. This means that when K is sufficiently large the B-type phase cannot be stable (assuming the orbit is periodic), and this is indeed what we find.

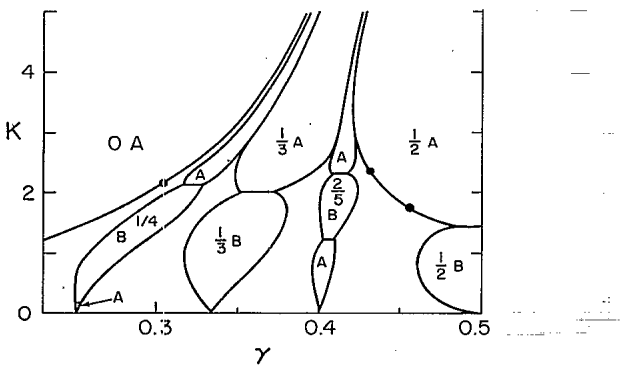

FIG. 15. Phase diagram corresponding to (6.2), and (6.8) with $\epsilon=0.1$. The solid circles are some of the points on the boundaries of $\omega=0$ and $\omega=\frac{1}{2}$ where horizontal bars accumulate.

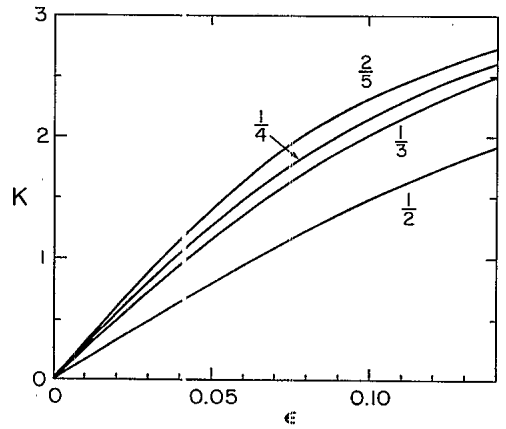

FIG. 16. Maximum value of K for which a horizontal bar occurs in the phase diagram, as a function of ϵ for $\omega=\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$, and $\frac{2}{5}$.

This new feature was checked in the following way. The force equilibrium equations (1.5) were solved for atomic positions for various choices of ω, for phases having symmetries corresponding to A and B. The energies were then compared, and the values of K at which they cross were obtained with much higher precision than is possible using (3.3) alone. We found that the corresponding K values go to zero with ϵ, Fig. 16, in a continuous manner, and that the number of bars increases with the denominator Q of $\omega=P/Q$. Since solutions to the force equilibrium equations need not yield the actual ground state (even if the symmetry is appropriate), these results were occasionally checked by solving (3.3) and using the corresponding R and τ to generate the ground state.

The numerical studies just mentioned strongly suggest that the horizontal bars possess points of accumulation in the γ,K plane. Three of these are indicated by solid circles in Fig. 15, one on the boundary of the $\omega=0$ phase and two on the boundary of $\omega=\frac{1}{2}$. We find numerically that these accumulation points occur at the same values of K at which the minimum-energy soliton in the corresponding $\omega=0$ and $\omega=\frac{1}{2}$ phases changes its character, as indicated in the next paragraph.

We assume that the defect of interest always possesses a point of reflection symmetry. This point can occur either at the maximum or minimum of V, and either there is or

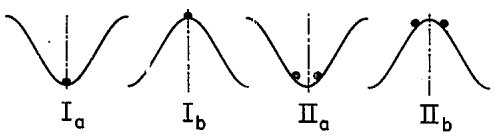

FIG. 17. Atoms at the center of a minimum-energy soliton defect indicating the different symmetry types I_a, I_b, II_a, and II_b. [Compare with Figs. 7(b) and 8(b).]

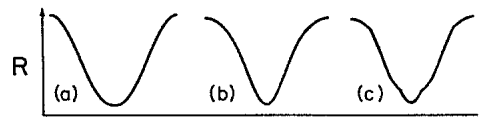

FIG. 18. Effective potential R for $\gamma=\frac{1}{2}$ and K at the horizontal bar of $\omega=\frac{1}{2}$ (Fig. 15). The R corresponding to the pure phases A and B is shown in (a) and (b), while (c) is the mixture produced by our numerical procedure.

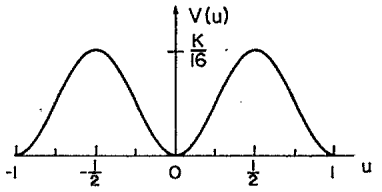

FIG. 19. Piecewise parabolic potential $V(u)$ of (6.9).

is not a particle at the symmetry point. These cases are shown in Fig. 17, which indicates a notation for the different symmetry types. In the $\omega=0$ phase the minimum-energy soliton changes from II_b to I_b as K decreases below the value indicated by the accumulation point. In the $\omega=\frac{1}{2}$ phase the change is from I_a to I_b at the upper point and then back to I_a as K decreases through the value of the lower point. Since one expects that the minimum-energy soliton will determine the location of the phase boundary, it is plausible that the latter will have a discontinuous slope at the accumulation point. Our numerical studies indicate such discontinuities which are, however, quite small. There is evidence to suggest that in the $\omega=0$ phase there are an infinite number of such accumulation points as K tends to zero.[50]

The horizontal bars separating the A and B types of phase behave in many respects like first-order phase transitions. Thus the configuration changes discontinuously, and so does $(\partial\lambda/\partial K)_\gamma$, upon crossing the bar. We have studied the solutions to the minimization eigenvalue equation for the $\omega=\frac{1}{2}$ case and find the following behavior. At $\gamma=\frac{1}{2}$, there are distinct solutions R_A and R_B as one approaches the transition from larger- and smaller-K values, respectively. For K very near the transition value the solution we find, Fig. 18, is a mixture of these two [see (3.13)] which is undoubtedly a reflection of the numerical approximation procedure. Both of the pure ground states are R orbits, and which of these is reached by iterating the corresponding τ map depends on the choice of the initial u'. However, as γ decreases towards the value at the left edge of the bar, R_A and R_B approach each other and probably tend towards a common unique solution. Note that the discontinuity in $(\partial\lambda/\partial K)_\gamma$ remains constant, as do the configurations for K just above and just below the transition, over the entire range of γ values included in the bar. (Note that the phase diagram has a mirror symmetry at $\gamma=\frac{1}{2}$, so the behavior for $\gamma>\frac{1}{2}$ mimics that for $\gamma<\frac{1}{2}$.)

We also studied the case where (6.8) includes a third harmonic rather than a second harmonic, i.e., 4π is replaced by 6π. In this case $\epsilon\geq 0$ is innocuous (yields a phase diagram like Fig. 11), whereas $\epsilon<0$ produces a diagram which is similar to Fig. 15. However, it differs in some details; in particular for $\omega=\frac{1}{2}$ there is only one phase.

Phase diagrams which are similar to Fig. 15 in that they contain horizontal bars representing first-order transitions between phases of different symmetry have been noted by Aubry *et al.*[24] for a mode in which V is a scalloped potential (of repeating parabolas), but with two atoms per "unit cell" and, in addition, a staggered electric field. (See Fig. 11 of Ref. 24.) Our work shows that such features can arise in even simpler models. Indeed the original Frenkel-Kontorova model with V a simple cosine seems in some sense on the borderline between two very different types of behavior, one being the simple tongues found previously by Aubry,[25] and the other in which the tongues are crossed by horizontal bars.

D. Piecewise parabolic V and harmonic W; Numerical errors

1. Hamiltonian

In this model W is given by (6.2) and V by

$$V(u)=\tfrac{1}{2}Ku^2,\quad -\tfrac{1}{4}\leq u\leq\tfrac{1}{4}$$
$$=K/16-\tfrac{1}{2}K(u-\tfrac{1}{2})^2,\quad \tfrac{1}{4}\leq u\leq\tfrac{3}{4} \qquad (6.9)$$

and for other values of u by periodicity, $V(1+u)=V(u)$. Thus V is piecewise parabolic with a continuous first derivative and a second derivative which is discontinuous at $u=-\frac{1}{4},\frac{1}{4},\frac{3}{4}$, etc. See Fig. 19.

TABLE I. Five closed-form solutions to the minimization eigenvalue equation for V given by (6.9).

No.	γ	K	ω	Vertices: $(u,r(u))$
1	0	$\frac{9}{10}$	0	$(-\frac{1}{2},-\frac{3}{10}),(-\frac{1}{4},-\frac{3}{8}),(\frac{1}{4},\frac{3}{8}),(\frac{1}{2},\frac{3}{10}),(\frac{1}{2},-\frac{3}{10})$
2	$\frac{1}{20}$	$\frac{9}{10}$	0	$(-\frac{5}{12},-\frac{3}{8}),(-\frac{1}{4},-\frac{17}{40}),(\frac{1}{4},\frac{13}{40}),(\frac{7}{12},\frac{9}{40}),(\frac{7}{12},-\frac{3}{8})$
3	$\frac{1}{4}$	$\frac{4}{3}$	0	$(0,-\frac{1}{4}),(\frac{1}{4},\frac{1}{4}),(1,-\frac{1}{4})$
4	$\frac{13}{32}$	$\frac{4}{3}$	$\frac{1}{2}$	$(-\frac{1}{4},-\frac{5}{32}),(-\frac{3}{16},-\frac{1}{32}),(0,-\frac{5}{32}),(\frac{3}{16},\frac{7}{32}),(\frac{3}{4},-\frac{5}{32})$
5	$\frac{11}{32}$	1	$\frac{1}{3}$	$(-\frac{1}{4},-\frac{3}{32}),(0,-\frac{3}{32}),(\frac{1}{4},\frac{5}{32}),(\frac{3}{4},-\frac{3}{32})$

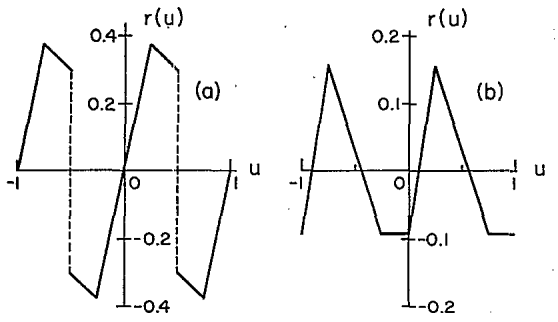

FIG. 20. Derivative r of the effective potential R for (a) example 1 and (b) example 5 in Table I.

2. *Some exact solutions for the effective potential R*

Because V has such a simple form it is possible in some cases to find explicit solutions to the minimization eigenvalue equation (3.3). Some examples are given in Table I. As the graph of

$$r(u)=dR/du \tag{6.10}$$

is piecewise linear (in these examples) and periodic with period 1, it can be reconstructed from the end points, $(u,r(u))$, of the straight-line segments listed in the column labeled "vertices." Examples 1 and 5 are sketched in Fig. 20. Note that in examples 1 and 2, r is discontinuous at the point where R has an upward kink. In examples 3, 4, and 5, r is continuous, so R has no kinks and F is constant. (These represent somewhat special points in the phase diagram; see below.)

3. *Phase diagram and pinching points*

Figure 21 shows the phase diagram in the γ,K plane constructed using numerical solutions to the minimization eigenvalue equation along with closed-form solutions, some of which are listed in Table I. As in Fig. 15, one finds transitions between phases of different symmetry for a given rational ω; A and B have the same significance as in Sec. VI C above.

However, these transitions take place at single "pinching points" rather than along horizontal bars of finite width in the γ,K plane. At the pinching point there is a sliding ground state in the sense that every atom can be simultaneously and continuously displaced to the right, in such a manner that the whole configuration remains a ground state, until each atom is in the position formerly occupied by its right-hand neighbor. Thus the ground state is infinitely degenerate. The fact that this can occur for rational ω seems to reflect the piecewise parabolic property of V, which means that the equilibrium equations are actually (inhomogeneous) linear equations in the atomic positions as long as these do not move past points where $V''(u)$ is discontinuous. The situation is somewhat reminiscent of the Kolmogorov-Arnold-Moser (KAM) trajectories,[28,31,33,51] the proof of whose existence demands, however, a higher degree of differentiability than is present in a piecewise parabolic V.

Given such a sliding mode it is impossible to produce a soliton by adding or deleting a particle since the other particles will simply relax into a new periodic orbit. Equivalently, the defect energy for adding or deleting a particle is zero, and hence by the discussion in Sec. VI A, the interval in γ over which this ω is the ground state should be zero.

The area-preserving map associated with the equilibrium equations (1.5) depends on K but not on γ, and the transition from A to B at a pinching point occurs in the following way. For K near the transition value, there is a hyperbolic periodic orbit intertwined with an elliptic orbit of the same period. These become part of an invariant circle as K passes through the transition, and emerge on the other side with the previous hyperbolic orbit now elliptical and vice versa. The ground state corresponds to the hyperbolic orbit, consistent with the expectation that an elliptic orbit cannot be a ground state when V is smooth and W is convex and twice differentiable.[25,52] (In the present case V is not smooth, but the corresponding orbits only pass through the points where V'' is discontinuous when K has its transition value.)

As is the case of the bars in Fig. 15, it is more efficient to locate the pinching points in Fig. 21 by solving the force equilibrium equations than by finding a solution to the minimization eigenvalue equation. They can best be found by looking for a ground state such that for the K and ω of interest, a particle falls precisely on each of the "division points" where $V''(u)$ is discontinuous, as in Fig. 22. By assigning the particles at the division points to the parabolic arc lying to its right, it is evident that a displacement of the whole configuration in which each atom moves to the position previously occupied by its neighbor to its right carries the system into a ground state of the

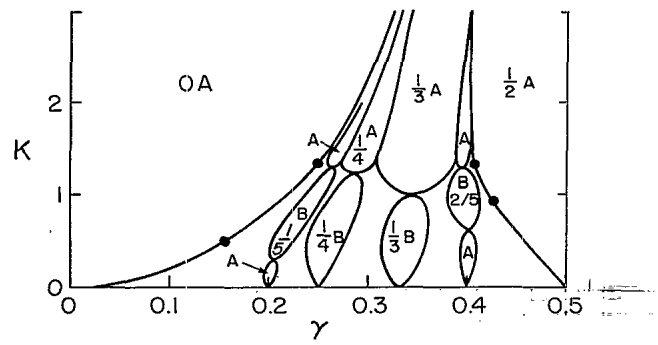

FIG. 21. Phase diagram corresponding to (6.2) and (6.9). The solid circles are some of the accumulation points of the pinching points separating phases A and B.

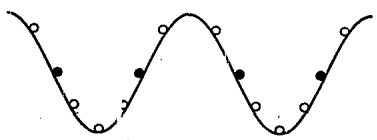

FIG. 22. Example of a ground state giving rise to a pinching point in the phase diagram of Fig. 21. The solid circles are atoms lying precisely at the points where $V''(u)$ is discontinuous.

same energy. The linearity of the equilibrium equations (see above) then guarantees that the intermediate configurations have the same energy.

The pinching points of phases with $\omega=P/Q$, $Q\leq 13$, are plotted in the ω,K plane in Fig. 23. The points appear to lie on various curves which terminate on accumulation points. For example, the sequence $(1/n, K_n)$, with K_n the largest K for a pinching point for $\omega=1/n$, converges to $\omega=0$, $K=\frac{4}{3}$, with the asymptotic behavior

$$\tfrac{4}{3}-K_n\sim 3^{-n}\,. \tag{6.11}$$

We suspect that at $K=\frac{4}{3}$, every ω is an accumulation point of pinching points, though we expect there to be lots of accumulation points at other values of K as well. Figure 23 also suggests the presence of some self-similar structures, but this needs to be confirmed by further study.

Corresponding to these accumulation points in the ω,K plane there should be accumulation points in the γ,K plane on the boundaries of phases with rational values of ω. Examples 3 and 4 in Table I, on the right and left edges of the $\omega=0$ and $\omega=\frac{1}{2}$ regions, respectively, in Fig. 21 are of this type. These also appear to correspond to sliding ground states, but in a sense which is somewhat more complex than that described earlier. Thus in example 3 in Table I, there is a ground-state configuration in which $u_n=0$ for all n, and another ground state in which the atoms are disposed in the fashion shown in Fig. 13. That is to say, there is a soliton defect, but its excitation energy Λ, (2.12), is zero. Furthermore this soliton is a sliding ground state in the sense that the atoms can simultaneously and continuously be displaced till each reaches the position formerly occupied by its right-hand neighbor while the configuration remains a ground state. Thus the soliton itself is not pinned but can move with zero energy change.

As K passes through the value corresponding to one of these accumulation points, one expects a change in the symmetry of the minimum-energy soliton, just as noted previously in Sec. VI C. Indeed, in the $\omega=0$ phase the minimum-energy soliton is of type II_b (Fig. 17) for $K>\frac{4}{3}$ and of type I_b for $\frac{1}{2}<K<\frac{4}{3}$, while for $\omega=\frac{1}{2}$ the type is I_a for $K>\frac{4}{3}$, and then I_b down to $K=0.92$. Using a closed-form solution for R at $\gamma=0$ we have (as explained in Sec. VI A) calculated the phase boundary of the $\omega=0$ region for $K\geq\frac{1}{2}$ assuming that this is determined by the γ value where the energy of the corresponding soliton goes to zero, with results in agreement with the numerical calculations. (At smaller K values it is more difficult to construct the closed-form solution for R.)

The existence of sliding ground states for rational values of ω implies that the width in γ of the region corresponding to this ω in the phase diagram must go to zero at the corresponding K, as noted earlier. This is consistent with our numerical solutions of the minimization eigenvalue equation, but the latter are not precise enough to tell us how the width shrinks to zero as K approaches the pinching point, so that the precise form shown in Fig. 21 should not be taken too seriously. The numerical studies do indicate that upon approaching the pinching point from either phase A or B, the R function approaches a common limit, and this limit has, as expected, a continuous first derivative (no kinks), and the corresponding F is constant. Note that example 5 in Table I is the closed-form solution at the pinching point for $\omega=\frac{1}{3}$. Despite the uniqueness of R, the phase transition at a pinching point is "first order" in the sense that $(\partial\lambda/\partial K)_\gamma$ is discontinuous (the same as at one of the bars in Fig. 15) due to the discontinuous change in the configuration going from phase A to phase B.

4. *Dependence of numerical errors on grid size*

We checked the accuracy of our numerical procedures (Sec. V), and in particular their dependence on the size $h=1/N$ of the grid imposed on the unit interval, by comparing the resulting R and λ with the exact quantities R_e and λ_e for the cases listed in Table I. Two measures of error,

$$\delta R=\max_u |R(u)-R_e(u)|\,,$$
$$\Delta R=N^{-1}\sum_u [R(u)-R_e(u)]^2\,, \tag{6.12}$$

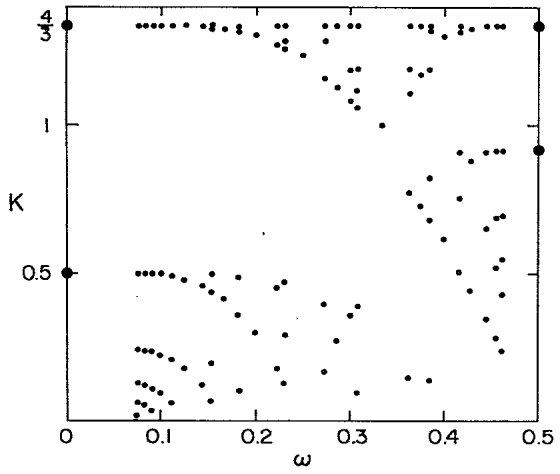

FIG. 23. Pinching points in the $\omega=P/Q,K$ plane for $Q\leq 13$. The large solid circles correspond to the accumulation points shown in Fig. 21. There are many other accumulation points.

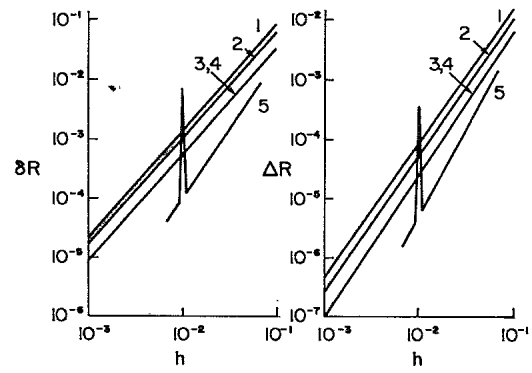

FIG. 24. Errors in R produced by the finite grid for examples 1 to 5 of Table I. Example 5 produces irregular results for values of h smaller than those shown here.

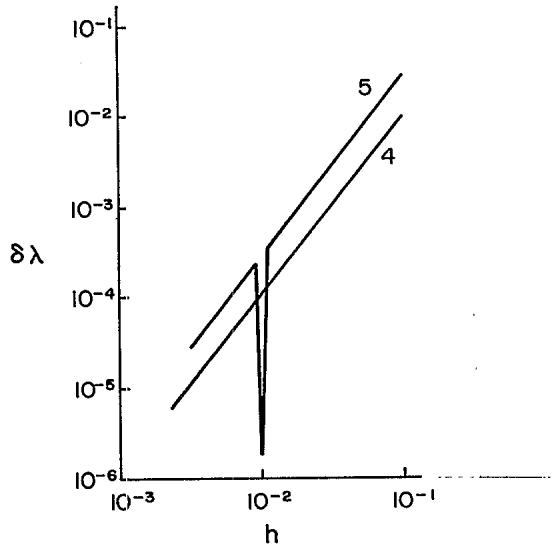

FIG. 25. Errors in λ produced by the finite grid for examples 4 and 5 of Table I.

where both R and R_e have the same minimum value of zero, are plotted against h in Fig. 24. Both vary as h^2, apart from example 5, which also shows a resonance phenomenon which we do not understand. The error in λ has the same behavior, Fig. 25, though it is zero in the (exceptional) case in which the ground-state orbit lies on points of the grid.

VII. CONCLUSION

The method of effective potentials as developed in this paper provides a useful and apparently new approach to the problem of calculating the ground states of one-dimensional systems with short-range interactions. In principle it yields both the ground-state energy and the corresponding particle configuration without the ambiguities associated with metastable (or unstable) states which arise in approaches based on solving the force equilibrium equations. It works both for a convex nearest-neighbor interaction W and in the nonconvex case. And, in addition, it can be used to calculate the energies of certain types of defect configurations.

While the effective potential approach can be used to establish certain abstract mathematical results (such as the existence of a ground state), it also turns out to be a useful tool for numerical analysis of the ground-state phase diagrams of these systems. In particular, the diagrams in Figs. 11, 15, and 21 were constructed primarily by numerical solutions of the functional equation (3.3). All of these diagrams represent new results, and while the general topological features in Fig. 11 are not unexpected on the basis of previous calculations for other choices of V, the bars and pinching points in Figs. 15 and 21 were not anticipated in advance of the numerical calculations.

Once the general features of the phase diagram have been established using the effective potential method, further refinements can be made by a numerical solution of the force equilibrium equations for a periodic orbit. The latter provides more accurate answers for the positions of the atoms in a periodic ground state (and thus for the energy) than can be obtained using a relative coarse grid for solving (3.3), and with much less computation. We have, in particular, used the equilibrium equations to find the precise K values for the bars and pinching points in Figs. 15 and 21.

To be sure, the numerical procedures we have used to solve the minimization eigenvalue equation are relatively primitive and leave much room for improvement. The use of a simple grid of points on the unit interval is roughly comparable to employing a trapezoid rule for numerical integration. Attempts to improve on this need to take account of the fact that while the eigenfunction R is continuous, it can have upward kinks, as in Fig. 9. While the minimum in (3.3) cannot occur at such kinks, their existence complicates the problem of representing the function by its values at a finite number of points. The direct approach of simply using many more points in the grid demands a more efficient way of solving (3.3) than the simple iteration method of (5.1).

The minimization eigenvalue equation can also be employed, in principle, to study incommensurate phases: There is no requirement that the ground state be periodic in order to use (3.3). However, there are obvious difficulties which face an attempt to study such phases by numerical methods, and thus far we have not tried to deal with these.

There are a number of other one-dimensional systems of short-range character to which the methods of this paper should be applicable. The case of a chiral XY model in a magnetic field, the $p=1$ case in Sec. VI B, is an interesting situation in which the nonconvex nature of W creates interesting features in the phase diagram. The mean-field approximation to the ANNNI model presents a greater challenge, as in this case each x_n in (1.6) lies in a two-dimensional square, making numerical computations based on discretization more difficult.

There are some problems of a purely mathematical character connected with the minimization eigenvalue problem, (3.16) or (3.17), in the case where the variables x and y are continuous. For example, the structure of the space of solutions has been worked out by Cuninghame-Green in the case in which K is a finite matrix, and this needs to be appropriately generalized. In addition, the effective potential approach should prove useful for proving certain mathematical properties about the ground states of systems of the type (1.1) and (1.6) even when these cannot be obtained explicitly.

ACKNOWLEDGMENTS

We wish to thank R. Duffin, J. Mather, D. Ruelle, L. Tang, M. Widom, and C. Yokoi for helpful discussions on the topic of this paper and closely related matters. Financial support for this research has been provided by the National Science Foundation through Grant No. DMR-8108310.

*Present address: Light source project, Argonne National Lab, Argonne, Illinois 60439.
[1]The following review articles contain numerous references: (a) R. Pynn, Nature **281**, 433 (1979); (b) P. Bak, Rep. Prog. Phys. **45**, 587 (1982); (c) Y. A. Izyumov, Sov. Phys.—Usp. **27**, 845 (1984) [Usp. Fiz. Nauk **144**, 439 (1984)].
[2]J. Rossat-Mignod *et al.*, Phys. Rev. B **16**, 440 (1977).
[3]P. Fischer *et al.*, J. Phys. C **11**, 345 (1978).
[4]D. Gibbs *et al.*, Phys. Rev. Lett. **55**, 234 (1985).
[5]B. Mandelbrot, *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
[6]S. Aubry, Physica D **7**, 240 (1983).
[7]S. Aubry, J. Phys. C **16**, 2497 (1983).
[8]L. de Seze and S. Aubry, J. Phys. C **17**, 389 (1984).
[9]S. Aubry, J. Phys. (Paris) **44**, 147 (1983).
[10]S. Aubry and P. Y. Le Daeron, Physica D **8**, 381 (1983).
[11]M. Peyrard and S. Aubry, J. Phys. C **16**, 1593 (1983).
[12]Y. G. Sinai, J. Stat. Phys. **29**, 401 (1982).
[13]V. F. Lazutkin and D. Y. Terman, Commun. Math. Phys. **94**, 511 (1984).
[14]V. L. Pokrovsky, J. Phys. (Paris) **42**, 761 (1981).
[15]S. N. Coppersmith and D. S. Fisher, Phys. Rev. B **28**, 2566 (1983).
[16]S. R. Sharma, B. Bergersen, and B. Joos, Phys. Rev. B **29**, 6335 (1984).
[17]P. Bak, Phys. Rev. Lett. **46**, 791 (1981).
[18]B. Joos, Solid State Commun. **42**, 709 (1982).
[19]B. Joos, B. Bergersen, R. J. Gooding, and M. Plischke, Phys. Rev. B **27**, 467 (1983).
[20]S. R. Sharma and B. Bergersen, Phys. Rev. B **30**, 6586 (1984).
[21]A. Banerjea and P. L. Taylor, Phys. Rev. B **30**, 6489 (1984).
[22]A. Milchev, Phys. Rev. B **33**, 2062 (1986).
[23]M. Kardar, Phys. Rev. B **30**, 6368 (1984).
[24]S. Aubry, F. Axel, and F. Vallet, J. Phys. C **18**, 753 (1985).
[25]S. Aubry, in *Solitons and Condensed Matter Physics,* edited by A. R. Bishop and T. Schneider (Springer, Berlin, 1978), p. 264.
[26]S. Aubry, in *The Riemann Problem, Complete Integrability and Arithmetic Applications,* edited by D. and G. Chudnovski (Springer, Berlin, 1980), p. 221.
[27]B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[28]J. M. Greene, J. Math. Phys. **20**, 1183 (1979).
[29]J. M. Greene, R. S. MacKay, F. Vivaldi, and M. J. Feigenbaum, Physica D **3**, 468 (1981).
[30]R. S. MacKay and I. C. Percival, Commun. Math. Phys. **98**, 469 (1985).
[31]S. J. Shenker and L. P. Kadanoff, J. Stat. Phys. **27**, 631 (1982).
[32]J. N. Mather, Topology **21**, 457 (1982).
[33]J. N. Mather, Ergodic Theor. Dyn. Sys. **4**, 301 (1984).
[34]J. E. Sacco and J. B. Sokoloff, Phys. Rev. B **18**, 6549 (1978).
[35]S. Aubry, Phys. Rep. **103**, 127 (1984).
[36]R. B. Griffiths and W. Chou, Phys. Rev. Lett. **56**, 1929 (1986).
[37]J. E. Sacco, A. Widom, and J. B. Sokoloff, J. Stat. Phys. **21**, 497 (1979).
[38]R. A. Cuninghame-Green, *Minimax Algebra* (Springer, Berlin, 1979).
[39]F. Axel and S. Aubry, J. Phys. C **14**, 5433 (1981).
[40]W. Selke and P. M. Duxbury, Z. Phys. B **57**, 49 (1984).
[41]F. C. Frank and J. H. van der Merwe, Proc. R. Soc. London, Ser. A **198**, 205 (1949); **198**, 216 (1949).
[42]W. L. McMillan, Phys. Rev. B **14**, 1496 (1976).
[43]P. Bak and J. von Boehm, Phys. Rev. B **21**, 5297 (1980).
[44]J. Villain and M. B. Gordon, J. Phys. C **13**, 3117 (1980).
[45]This can be compared with the geometrical construction for the Legendre transformation; see, for instance, H. B. Callen, *Thermodynamics: An Introduction* (Wiley, New York, 1960).
[46]Z. Nitecki, *Differentiable Dynamics* (MIT, Cambridge, Mass., 1971).
[47]J. Frenkel and T. Kontorova, Phys. Z. Sowjetunion **13**, 1 (1938).
[48]S. C. Ying, Phys. Rev. B **3**, 4160 (1971).
[49]C. Yokoi, L. Tang, and W. Chou (unpublished).
[50]C. Yokoi (private communication).
[51]W. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968), Sec. 21.
[52]E. Allroth, J. Phys. A **16**, L497 (1983).