UNUSUAL BAND STRUCTURE, WAVE FUNCTIONS AND ELECTRICAL CONDUCTANCE IN CRYSTALS WITH INCOMMENSURATE PERIODIC POTENTIALS

J.B. SOKOLOFF

Physics Department, Northeastern University, Boston, Massachusetts 02115, U.S.A.

NORTH-HOLLAND–AMSTERDAM

PHYSICS REPORTS (Review Section of Physics Letters) 126, No. 4 (1985) 189–244. North-Holland, Amsterdam

UNUSUAL BAND STRUCTURE, WAVE FUNCTIONS AND ELECTRICAL CONDUCTANCE IN CRYSTALS WITH INCOMMENSURATE PERIODIC POTENTIALS

J.B. SOKOLOFF

Physics Department, Northeastern University, Boston, Massachusetts 02115, U.S.A.

Received 26 February 1985

Contents:

Abstract:
The subject of electronic and phonon states in crystals with incommensurate periodic potentials is reviewed. The emphasis is on the physics behind the various theoretical methods of treating this problem and on showing the relationships between them. This review is meant to complement the reviews of mathematically rigorous results on the problem by Simon and by Bellissard. Although the review is primarily theoretical, there is a section which discusses possible experimental application of the theory, as well as experimental observations that have already been made. Some new theoretical results are presented and some old results are reinterpreted.

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 126, No. 4 (1985) 189–244.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 36.00, postage included.

0 370-1573/85/$19.60

1. Introduction

This article deals with the single particle states of almost periodic crystals. Almost periodic or incommensurate crystals are crystals which possess two or more periodic structures whose periods are incommensurate with each other. Examples of such crystals are crystals containing charge or spin density waves [1, 2], mercury chain compounds [3] and certain crystals which have distortion waves which are incommensurate with the underlying Bravais lattice [5]. Although these crystals are not random in the usual sense, they lack translational symmetry since there exist no translations which will leave the periods of all the periodic structures invariant. Nevertheless, there exist translations which almost leave them invariant. De Wolff and Janner and Janssen [6] were able to define a translation group in higher than d dimensions (where d is the actual physical dimension) which leaves the structure of an almost periodic crystal invariant.

In 1980, Aubry and André [7] studied the following one dimensional tight binding model for an almost periodic crystal:

$$\tfrac{1}{2}t(f_{n+1}+f_{n-1})+V_0\cos(Qna+ha)f_n=Ef_n\,, \tag{1}$$

where the wave function defined by the coefficient f_n is

$$\psi(x)=\sum_n f_n\,\phi(x-na)\,,$$

where $\phi(x-na)$ is a Wannier function centered on lattice site na, a is the lattice constant, t is the hopping matrix, V_0 is the potential strength, Q is the wave vector of the periodic structure in the tight binding lattice, ha the phase factor, and E is the energy eigenvalue. They were able to show that a metal–insulator transition occurred for $V_0=t$. This is to be contrasted to a random system in one dimension for which almost all states are localized. Equation (1) also represents phonons in an almost periodic crystal if we set $t\to -t$ and $E\to\omega^2-2t$ where ω is the phonon frequency. Many of the ideas discussed in the Aubry–André paper have been confirmed by succeeding workers in the field. In addition, quite a few rigorous results have been proven for this and related models in recent years [8]. Another interesting feature of almost periodic crystals can be anticipated by considering high order periodic systems (i.e. periodic systems with two or more commensurate periods in which the ratio of the periods is a ratio of large integers). For any periodic system, there should be a gap at half of any reciprocal lattice vector of the crystal, and all sums and differences of the reciprocal lattice vectors of the periodic structures are reciprocal lattice vectors. As the system becomes higher and higher order periodic (i.e. the integers become larger), the smallest reciprocal lattice vector of the crystal gets smaller and smaller. Thus, in the almost periodic limit, there should be a gap found in the vicinity of every wave vector, and hence every energy. This implies a highly fragmented band structure.

The question of localization in almost periodic crystals will be addressed in section 2, as well as the mathematical techniques that have been used to study localization. In section 3 the band structure of an almost periodic crystal will be discussed. Section 4 will discuss the type of state which occurs for a system which is at the borderline between having localized and extended states. Such states are neither extended in the usual sense nor exponentially localized. In section 5, generalization of the previously discussed one dimensional results to higher dimensions will be discussed. Section 6 will discuss the possible experimental consequences of almost periodicity. Although the fragmented nature of the band structure and the localization transition which occurs would be expected to lead to exotic transport properties, such effects have not been observed to date. The reasons for the difficulty of observing effects of almost periodicity will be discussed in this section. Section 7 will list results which have not been discussed earlier.

2. Aubry duality and the metal–insulator transition

2.1. Aubry self-duality

Aubry and André showed that eq. (1) possess self-duality [7]. Self-duality means that if we make the following transformation to reciprocal space:

$$f_n = \sum_{m=-\infty}^{\infty} g_m \exp\{\mathrm{i}m(nQa + ha)\} \exp\{\mathrm{i}kan\} \tag{2a}$$

$$g_m = \sum_{n=-\infty}^{\infty} f_n \exp\{-\mathrm{i}n(mQa + ka)\} \exp\{-\mathrm{i}kam\} \tag{2b}$$

where k is a wave vector, g_m satisfies

$$\tfrac{1}{2}V_0\,(g_{m+1} + g_{m-1})\, t\cos(mQa + ka)\, g_m = E\, g_m\,. \tag{3}$$

Thus, eq. (3) has the same form as eq. (1) but with the roles of V_0 and t interchanged. Clearly, if g_m is localized, in the sense that

$$\sum_m |g_m|^2$$

is finite (i.e. g_m is normalizable), the quantity

$$f(x) = \mathrm{e}^{\mathrm{i}kx} \sum_{m=-\infty}^{\infty} g_m\, \mathrm{e}^{\mathrm{i}mQx} \tag{4}$$

converges and is a Bloch function, and thus, an extended state. Since

$$f(x)|_{x=na+ha/Q} = \mathrm{e}^{\mathrm{i}kh/Q} f_n \tag{5}$$

f_n must be extended in the sense that

$$\sum_n |f_n|^2 = \infty\,. \tag{6}$$

By the symmetry between eqs. (1) and (3), we also conclude that if f_n is localized, g_m is extended.

Thouless has developed a formula which relates the exponent for the decay of a localized state to an integral over the density of states [9]. When applied to eq. (1), this formula is

$$\gamma_1(E) = \int \mathrm{d}E' \ln\left|\frac{2(E-E')}{t}\right| \rho(E') \tag{7}$$

where $\gamma_1(E)$ is the decay exponent and $\rho(E')$ is the density of states. The integral runs over the width of the band. Since the eigenvalue problems expressed by eqs. (1) and (3) must obviously have the same density of states, the exponent for eq. (3) $\gamma_2(E)$ is given by

$$\gamma_2(E) = \int \mathrm{d}E' \ln\left|\frac{2(E-E')}{V_0}\right| \rho(E') . \tag{8}$$

Then,

$$\gamma_1(E) = \gamma_2(E) + \ln(V_0/t) . \tag{9}$$

Since $\gamma_1(E)$ and $\gamma_2(E)$ are $\geqq 0$, we see from eq. (9) and the self-dual property discussed above eq. (7) that if the eigenstates of eq. (1) are exponentially localized, $\gamma_2(E) = 0$ and

$$\gamma_1(E) = \ln(V_0/t) > 0$$

which implies $V_0 > t$. Again by self-duality we conclude that if the eigenstates of eq. (1) are extended $\gamma_1(E) = 0$ and

$$\gamma_2(E) = \ln(t/V_0) > 0 ,$$

which implies that $t > V_0$. Thus, a metal–insulator transition occurs at $V_0 = t$. This result implies that in the localized regime (i.e. for $V_0 > t$) all states are exponentially localized for any irrational value of $Qa/2\pi$. This would imply a spectral density which consists of delta functions (a pure point spectrum). The previous arguments, which were based on the self-dual property possessed by eq. (1), can be applied to the following more general self-dual model:

$$t \sum_L Y_L f_{n+L} + V_0\, V(na+h)\, f_n = E f_n , \tag{10}$$

where Y_L is a decreasing function of L and where

$$V(x) = \sum_L Y_L\, \mathrm{e}^{\mathrm{i}LQx} . \tag{11}$$

Applying the previous argument based on Thouless's formula, we again find that a transition from all states extended to all states exponentially localized occurs at $V_0 = t$ [7]. The general self-dual model has the property that the Fourier transform of the hopping term has the same functional form as the potential. When they are equal in strength there is a metal–insulator transition. Another way of thinking of this is that $V_0 = t$ is a fixed point of the duality transformation represented by eq. (2), and therefore, we expect a metal–insulator transition, if it occurs, to occur at this point [7]. For non-self-dual models we do not seem to find a transition from all states extended to all states localized. For example, Aubry has shown by numerical calculation of the Thouless exponent that for a near neighbor hopping model with a potential of the form

$$V_0[\cos(Qan + h) + \cos 3(Qan + h)] \tag{12}$$

for certain values of V_0 there can exist both localized and extended states separated by a mobility edge. Soukoulis and Economou [10] have found a similar result for the potential

$$V[\cos(Qan + h) + \tfrac{1}{3}\cos 2(Qan + h)]\,. \tag{13}$$

The self-duality transformation in eq. (2) can be used to learn something about other interesting non-self-dual models. For example, a non-analytic potential (e.g. a square wave potential such as $V_0\,\mathrm{sign}[\cos(nQ + h)]$, where $\mathrm{sign}(x) = +1$ if x is positive and -1 if x is negative), has a Fourier series

$$V(x) = V_0 \sum_L Y_L\, e^{iLQx} \tag{14}$$

in which Y_L falls off very slowly with L. For example, for the square wave potential $Y_L \sim L^{-1}$ for large L. Then, the dual (i.e. Fourier transform) of this model has a long range hopping term. We would expect that long range hopping would lead to extended states in the dual model, which implies localized states in the original model. Actually, recent work on a particular square wave potential model shows that the eigenstates are more likely neither localized nor extended, but somewhere in between [11, 12]. This point will be discussed in section 4.

It should be mentioned here that Avron and Simon have shown that if $Qa/2\pi$ is an irrational number extremely well approximated by rationals (called a Liouville number), i.e. a number which satisfies the inequality

$$|Qa/2\pi - p_n/q_n| \le cn^{-q_n}\,,$$

where $\{p_n$ and $q_n\}$ is a sequence of integers such that p_n/q_n approximates $Qa/2\pi$, n is an integer and c is a constant, the spectral density in the $V_0 > t$ regime is singular continuous, which means that it is neither pure point nor absolutely continuous (i.e. a smooth function of energy) [13]. Rather it is a complete devil's staircase.

2.2. Numerical methods of studying the metal–insulator transition

The extended and localized nature of the eigenstates of an almost periodic crystal may be studied numerically [14] by using the method originally proposed by Anderson to study the localization problem in random systems [15, 16]. Consider the following time dependent generalization of eq. (1):

$$\frac{i}{h}\frac{\partial}{\partial T} f_n = \tfrac{1}{2}t\,(f_{n+1} + f_{n-1}) + V_0 \cos(Qna + ha)\, f_n\,. \tag{15}$$

If initially $f_n = \delta_{n,0}$, the solution to eq. (15) is

$$f_n(T) = G_{n,0}(T)$$

where G is the Green's function for eq. (15).

Then,

$$f_0(T) = G_{00}(T) = \frac{1}{2\pi}\int \mathrm{d}E \frac{e^{-iET}}{E - V_0\cos(ha) - M_{0,0}(E)} \tag{16}$$

where M is the self-energy. It has been shown that the condition for the occurrence of an extended state, $f_0(T) \to 0$ as $T \to \infty$, is equivalent to the condition that $M(E)$ have a branch cut on the real axis [14, 15]. The self-energy can be expressed as a continued fraction,

$$M = \tfrac{1}{2}t\,[M_+ + M_-] \tag{17a}$$

where

$$M_\pm = \cfrac{1}{\frac{2}{t}[E - V_0\cos(\pm Qa + ha)] - \cfrac{1}{\frac{2}{t}[E - V_0\cos(\pm 2Qa + ha)] - \cfrac{1}{\frac{2}{t}[E - V_0\cos(\pm 3Qa + ha)] + }}} \tag{17b}$$

If one of these continued fractions is cut off after n terms we may write

$$M_\pm = P_n/Q_n \tag{18}$$

where

$$x_n = \frac{2}{t}[E - V_0\cos(\pm Qna + ha)]\,x_{n-1} - x_{n-2} \tag{19}$$

where x_n represents P_n or Q_n, and $P_1(E) = 1$,

$$P_2(E) = \frac{2}{t}[E - V_0\cos(\pm 2Qa + ha)]\,, \qquad Q_1(E) = \frac{2}{t}[E - V_0\cos(\pm Qa + ha)]\,,$$

$$Q_2(E) = \frac{4}{t^2}[E - V_0\cos(\pm Qa + ha)]\,[E - V_0\cos(\pm 2Qa + ha)] - 1\,.$$

If the continued fractions in eq. (17) converge, M does not have a branch cut and the state at energy E is either localized or we are in a bandgap. If it does not converge, we have a branch cut, and hence, an extended state. Using eq. (19), we find

$$\frac{P_n}{Q_n}-\frac{P_{n-1}}{Q_{n-1}}=\frac{P_nQ_{n-1}-P_{n-1}Q_n}{Q_nQ_{n-1}}=\frac{1}{Q_nQ_{n-1}}. \tag{20}$$

Thus, we conclude that the continued fraction will converge and hence we will have a localized state or a gap if $|Q_nQ_{n-1}|\to\infty$ as $n\to\infty$. Therefore, to study localization, we need only solve the difference equation (19) for Q_n. To distinguish between a localized state or a gap in the regime in which the continued fraction for M converges, it is necessary to search the denominator of the Green function in eq. (16) for poles. This method was used in ref. [14] to study the localization of states for eq. (1). Dy and Ma [17] have proposed an alternative to the continued fraction method of ref. [14]. They have also done a more careful job than was done in ref. [14] of studying very narrow bands which can occur when $t>V_0$, and they have succeeded in demonstrating that these are indeed bands of extended states and not localized states. Soukoulis and Economou [10] have studied the localized or extended nature of the solutions of eq. (1) by direct diagonalization of eq. (1) for matrices as large as $10\,000\times10\,000$. To accomplish this they use special matrix diagonalization methods suitable for tridiagonal matrices.

Another method for studying localization is to study the behavior of the wave function as a function of distance using the method of transmission matrices:

$$\begin{pmatrix} f_n \\ f_{n-1}\end{pmatrix}=\begin{pmatrix} E-(2V/t)\cos(Qna+ha) & -1 \\ 1 & 0\end{pmatrix}\begin{pmatrix} f_{n-1} \\ f_{n-2}\end{pmatrix}=\begin{pmatrix} Q_n & P_n \\ Q_{n-1} & P_{n-1}\end{pmatrix}\begin{pmatrix} f_1 \\ f_0\end{pmatrix} \tag{21}$$

where P_n and Q_n are defined in the last paragraph. Thus, this method involves solving exactly the same difference equation as for the continued fraction method. A transmission coefficient may be defined by demanding that the potential be non-zero only for $1\le n\le N-1$ and zero outside this range. For $n\ge N$, we choose f_n to have the form

$$f_n=\mathrm{e}^{-\mathrm{i}kna}+r\,\mathrm{e}^{\mathrm{i}kna} \tag{22a}$$

and for $n\le 1$

$$f_n=t_0\,\mathrm{e}^{-\mathrm{i}kna} \tag{22b}$$

where t_0 and r are the transmission and reflection amplitudes, respectively. Then, eqs. (16) and (17) can be solved simultaneously for t_0 and r from which we find the transmission coefficient

$$T_0=|t_0^2|=\frac{4\sin^2 ka}{|P_{n-1}-P_n\,\mathrm{e}^{-\mathrm{i}ka}+Q_{n-1}\,\mathrm{e}^{\mathrm{i}ka}-Q_n|^2}. \tag{23}$$

As was shown by Landauer [18], a knowledge of the transmission coefficient allows us to calculate the conductivity. This will be elaborated on later in this article. This method of defining the transmission coefficient is valid only for $|E|\le 2t$, because the states in eq. (22) are not defined outside this range.

A third method of studying whether the states are extended or localized is to use a renormalization group method originally suggested by J. Jose [19, 20]. This method is carried out by writing down the equation corresponding to eq. (1) for f_{n+1} and f_{n-1} and using this to eliminate f_{n+1} and f_{n-1} from eq. (1). The resulting equation has a form similar to eq. (1) but with twice the lattice spacing. We then repeat this procedure with the resulting equation. If the states are localized, the hopping matrix will approach zero as we continue the procedure. Thus, this renormalization group method is suitable for studying the question of localization, despite the fact that it is not adequate for discussing the question of the scaling behavior of a commensurate band structure as the system approaches the incommensurate limit [11, 12] which will be discussed in section 3.

2.3. *Translation group symmetry in a higher dimensional representation of an almost periodic crystal and the breaking of this symmetry at the metal–insulator transition*

As we mentioned in the introduction, almost periodic potentials lack translational symmetry, yet they seem to possess more translational symmetry than disordered systems in some sense. One method of expressing this idea is by introducing a translation symmetry group in $p+d$ dimensions, where d is the physical dimension and p is the number of incommensurate periods. This was done by Romerio [21], de Wolff [5], and Janner and Janssen [6]. The extra dimensions represent translations of one periodic component with respect to the others. For example, if the potential can be expressed as

$$V = \sum_{j=1}^{p} v_j(\boldsymbol{r}) \tag{24}$$

where

$$v_j(\boldsymbol{r} + \boldsymbol{b}_j^{\alpha}) = v_j(\boldsymbol{r}) \tag{25}$$

($\alpha = 1$–3) where the sets $\{\boldsymbol{b}_j^{\alpha}\}$ for different j's are unrelated to each other, we may construct a $p+d$ dimensional space by introducing a set of p position vectors $\{\boldsymbol{r}_j\}$, and writing the position vector in this space as

$$\boldsymbol{R} = (\boldsymbol{r}_1 \boldsymbol{r}_2 \cdots \boldsymbol{r}_p)\,, \tag{26}$$

and the potential as $\sum_j v_j(\boldsymbol{r}_j)$. (In the physical problem, all $\boldsymbol{r}_j = \boldsymbol{r}$.) This potential is invariant under a translation in the $p+d$ dimensional space by a vector

$$\boldsymbol{b} = (n_1 \boldsymbol{b}_1^{\alpha_1}, n_2 \boldsymbol{b}_2^{\alpha_2}, \ldots, n_p \boldsymbol{b}_p^{\alpha_p}) \tag{27}$$

where $\{n_j\}$ are integers. It was suggested by Romerio [21] that this implies that the eigenfunction in the dimensional space satisfies Bloch's theorem

$$\psi(\boldsymbol{R} + \boldsymbol{b}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{b}}\,\psi(\boldsymbol{R}) \tag{28}$$

where $\boldsymbol{k}\cdot\boldsymbol{b} = \sum_{j,\alpha_j} \boldsymbol{k}_j \cdot \boldsymbol{b}_j^{\alpha_j}$, or

$$\psi(\boldsymbol{R}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}\,U(\boldsymbol{R})\,, \tag{29}$$

where $U(R+b)=U(R)$. The projection of this wave function onto the original d dimensional space, obtained by setting all $\boldsymbol{r}_j=\boldsymbol{r}$, is

$$\psi(\boldsymbol{r})=\mathrm{e}^{\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{r}}u(\boldsymbol{r})\,, \tag{30}$$

where $\boldsymbol{K}=\Sigma_j \boldsymbol{k}_j$ and $u(\boldsymbol{r})=U$ with all $\boldsymbol{r}_j$'s $=\boldsymbol{r}$. Then, $u(\boldsymbol{r})$ is clearly an almost periodic function, like the potential. Belokolos, Dinaburg and Sinai [22] and Bellissard, Lima and Testard [23] proved such a Bloch theorem for the energy of the particle sufficiently high. Although the above argument seems to imply that their proof may be extended to all energies, it will be argued here that eq. (29) does not actually imply that all states are extended, as was implied in ref. [21].

The arguments used in ref. [21] to argue that all states are extended in almost periodic systems depend on arguments concerning highly pathological differential operators which makes them questionable. Although there does not exist a proof that there must exist localized states in almost periodic crystals, some fairly convincing physical arguments will be presented for their existence. First of all deLange and Janssen [24] have calculated the wave functions for commensurate systems as a function of increasing order of commensurability. The model that they study is a Kroenig–Penny model in which the separation between delta function potentials varies periodically with position in the lattice. This is the model which was considered by Azbel [25]. They find, for sufficiently strong potentials, states which are highly localized about a point in the unit cell. The wave vector of the modulation of the delta function displacements is written as $Q=2\pi M/aN$ where a is the mean delta function separation and N and M are integers, which are chosen by cutting off the continued fraction representation for irrational number $Qa/2\pi$ at some point, so that M/N is a rational approximation to the irrational number. As we make better and better approximations to it (i.e., as we keep more and more terms in the continued fraction expansion), M and N, and hence, the unit cell size, obviously increase. These highly localized states then appear more and more like states localized about a specific point in the lattice in the sense that the wave function drops closer and closer to zero at the unit cell boundary. These states were also plotted as a function of the phase of the modulation relative to the lattice formed by the mean positions of the delta function potentials. As the phase is varied by a small amount the peak of the localized state jumps to a point which is far from where it was originally. It is argued in the next paragraph that this is the expected behavior for a localized state in an almost periodic potential problem.

In order to understand the dependence of localized state on the phase, consider two periodic potentials of infinite spatial extent with periods which are incommensurate with each other. Let the relative phase be chosen so that two minima coincide at one point. Clearly they will coincide at only one point because the periods are incommensurate. If we now change the phase by an infinitesimal amount, these minima will no longer coincide, but two different minima may now coincide. These minima will in general be very far away from the original minima. If we call the point where two minima coincide the origin, it is clear that if after our infinitesimal change of the phase we move the origin to the new point of coincidence of minima, the system will appear exactly the same as it did before changing the phase. From this argument we may draw two conclusions. First, the energy spectrum of an almost periodic crystal with two periods is invariant under an infinitesimal change in the relative phase of the two potentials. Second, a localized state whose maximum lies a given distance from the original origin will be translated to a point the same distance from the new origin, which will in general be a large distance from where it was located. This is a reason for expecting non-analytic behavior of the localized states as a function of the phase. This may be what Aubry and André [7] mean by "localization due to breaking of analyticity". It should be noted that what we are speaking of here is not a lack of analyticity of the wave function as a function of position; rather we are speaking of a lack of analyticity as a function of the phase variable. The phase is only a position variable in the higher dimensional space introduced in order to define a translation symmetry group for the almost periodic system and not a physical position variable. Furthermore, this lack of analyticity does not imply that if the phase variable of an almost periodic potential were changed suddenly that an electron occupying a state localized at a particular point in the crystal would discontinuously jump to a point which is far away from where it is presently located. It would simply no longer occupy the same localized state.

The lack of analyticity of localized states in the phase variable is actually a very well known phenomenon, which can be understood quite simply using Aubry duality. Consider eq. (3) for small V_0 (i.e. the "nearly free electron limit"). Then clearly we are in the regime where each solution g_m is localized about a particular m value. If we were speaking of the periodic instead of the almost periodic case, we would expect that when k passed through half of any reciprocal we would pass through a gap into the next band. As this happens g_m would switch from being localized about one wave vector (i.e. $k + mQ$) to another. Thus g_m will switch over from being localized around $m = m_1$ to $m = m_2$, when the unperturbed energies (i.e. the eigenvalues for $V_0 = 0$) for $m = m_1$ and m_2 are equal, i.e.

$$2t\cos(k + m_1 Q)a = 2t\cos(k + m_2 Q)a\,,$$

whose solution is

$$k = \tfrac{1}{2}[(m_1 + m_2)Q + (2\pi/a)L]\,. \tag{31}$$

The term in brackets is recognized as a reciprocal lattice vector of the system. As we go to the limit in which Q and $2\pi/a$ are incommensurate, the values of k given by eq. (31), which are the locations of the gaps, become closer and closer together. Most of the closely spaced values of k given by (31) correspond to values of m_1 and m_2 which can be very far apart. By Aubry's duality transformation, what we have just said could equally well describe the configuration space problem if we make the following substitutions $k \to h$, $V_0 \to t$. Then, we see that it is obvious that an infinitesimal change in h can make a localized state jump from being localized around one site to being localized around a far away site.

2.4. Equivalence of difference and differential almost periodic Schroedinger equations

Since much of the work on almost periodic Schroedinger equations centers on studies of tight binding models (or difference equation models), one might question the generality of such results. There have been several works, however, which show that many difference equation models are equivalent to differential Schroedinger equation models. For example Azbel [25] studied a delta function Kroenig–Penny model in which the spacing of the delta functions is a periodic function of distance with period incommensurate with the mean delta function spacing. He showed that this model could be described by an almost periodic difference equation. DeLange and Janssen [24] have used the same difference equation. This equation is, however, more complicated than the Aubry model [eq. (1)]. An interesting connection between a differential equation and difference equation almost periodic problem was made by Bellissard et al. [26]. They showed that the Schroedinger equation for a periodic array of delta functions of potentials (which can be incommensurate with the lattice of delta functions) is equivalent to the Aubry model. The potential strength and energy parameters in this Aubry model are functions of the particle's energy, however.

Let us consider the general case of an array of delta function potentials whose positions and strengths are periodically modulated. We wish to derive a relationship relating ψ_n, ψ'_n to ψ_{n-1}, ψ'_{n-1} where ψ_n and ψ'_n are the values of the wave function and its spatial derivative on the right side of the nth delta function potential site. We can find ψ_n and ψ'_n in terms of ψ_{n-1} and ψ'_{n-1} by solving the free electron Schroedinger equation (which is valid between delta functions), using the values of $\bar{\psi}_{n-1}$ and $\bar{\psi}'_{n-1}$ as the boundary conditions on the wave function, where $\bar{\psi}_{n-1}$ and $\bar{\psi}'_{n-1}$ are the wave function and derivative on the left side of the $n-1$st delta function. Using the continuity of $\psi(x)$ and the known discontinuity of $\psi'(x)$ across a delta function, we obtain

$$\psi_n = \cos k(x_n - x_{n-1})\,\psi_{n-1} + \frac{\sin k(x_n - x_{n-1})}{ka}\,\psi'_{n-1} \tag{31a}$$

$$\psi'_n = -ka \sin k(x_n - x_{n-1})\,\psi_{n-1} + \cos k(x_n - x_{n-1})\,\psi'_{n-1} - \beta_n\psi_n\,, \tag{31b}$$

where β_n is the strength of the delta function potential at the point $x = x_n$, a is the mean delta function spacing, and $k = (2mE/h^2)^{1/2}$. Eliminating ψ'_n and ψ'_{n-1} from eqs. (31a, b) we obtain

$$t_{n+1,n}\psi_{n+1} + t_{n,n-1}\psi_{n-1} + (\beta_n/ka)\psi_n = \varepsilon_n\psi_n\,, \tag{32}$$

where

$$t_{n,n+1} = [\sin k(x_{n+1} - x_n)]^{-1}$$

and

$$\varepsilon_n = t_{n+1}t_{n-1} \sin k(x_{n+1} - x_{n-1})\,.$$

The localized or extended nature of the solutions to eq. (32) can now be studied by the continued fraction or renormalization group methods discussed earlier in this section. For the case of evenly spaced delta functions (i.e. $x_n = na$), eq. (32) reduces to

$$\psi_{n+1} + \psi_{n-1} + \frac{\beta_n \sin ka}{ka}\,\psi_n = 2\cos ka\,\psi_n\,, \tag{33}$$

which is Bellissard et al.'s result.

Although differential Schroedinger equations with delta function potentials are more general than single band tight binding-like models (i.e. difference equations), since the delta functions are singular potentials, the model is somewhat unphysical. As pointed out in ref. [20], however, it can be shown that an almost periodic array of finite potential barriers, with regions of zero potential between the barriers is equivalent to an almost periodic array of delta function potentials (i.e. an array of delta function potentials whose strength and position vary almost periodically). This result may be proven using the method of transmission matrices. Let the zero potential region to the left of a barrier be denoted by region 1 and the region to the right of the barrier, region 2. Then, we may write

$$\psi_j = A_j\,e^{ikx} + B_j\,e^{-ikx}$$

and with $j = 1$ or 2 in regions 1 and 2 respectively and where $k = (2E/\hbar)^{1/2}$. The transmission matrix is defined by

$$\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} = R \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}, \tag{34}$$

Erdos and Herndon [27] show that for any general potential barrier

$$R = \begin{pmatrix} (1 - \mathrm{i}t_0)\, \mathrm{e}^{-\mathrm{i}\beta}/2k & (-\mathrm{i}t_0/2k)\exp(-2\mathrm{i}kx_0) \\ (\mathrm{i}t_0/2k)\exp(2\mathrm{i}kx_0) & (1 + \mathrm{i}t_0)\,\mathrm{e}^{\mathrm{i}\beta}/2k \end{pmatrix} \tag{35}$$

where t_0 is the transmission amplitude and t_0, x_0 and β are energy dependent. If we could perform a transformation which eliminates the parameter β, eq. (35) would have the same form as the transmission matrix of a delta function potential located at $x = x_0$. The required transformation was shown by Erdos and Herndon [27] to be

$$M = U(v_1)\, R\, U^*(v_2), \tag{36}$$

where

$$U(v_j) = \begin{pmatrix} \exp(\mathrm{i}kv_j) & 0 \\ 0 & \exp(-\mathrm{i}kv_j) \end{pmatrix}$$

where

$$kv_1 = \tfrac{1}{2}\beta + k\,\Delta x \tag{37a}$$

$$kv_2 = \tfrac{1}{2}\beta - k\,\Delta x \tag{37b}$$

where Δx can be set equal to zero if we want at this stage, but it will be needed in the next step. We find that

$$M = \begin{pmatrix} (1 - \mathrm{i}t_0) & -\mathrm{i}t_0 \exp\{-2\mathrm{i}k(x_0 - \Delta x)\} \\ \mathrm{i}t_0 \exp\{2\mathrm{i}k(x_0 - \Delta x)\} & (1 + \mathrm{i}t_0) \end{pmatrix}. \tag{38}$$

Equation (38) is recognized as the transmission matrix for a δ function located at $x = x_0 - \Delta x$ [27]. If we have an array of N potential barriers

$$R = R_1 R_2 \cdots R_N \tag{39}$$

where each R has the form of eq. (35). Equation (39) may be written as

$$\begin{aligned} R = {}& U^*(v_{1,1})\, M_1\, U(v_{2,1})\, U^*(v_{1,2})\, M_2\, U(v_{2,2})\, U^*(v_{1,3})\, M_3\, U(v_{2,3}) \\ & \cdots U(v_{2,N-1})\, U^*(v_{1,N})\, M_N\, U(v_{2,N}). \end{aligned} \tag{40}$$

Then, to make the potential barrier array problem equivalent to the delta function array problem, we must choose the Δx's such that

$$0 = k(v_{2,n-1} - v_{1,n}) = \tfrac{1}{2}(\beta_{n-1} - \beta_n) + k(\Delta x_{n-1} - \Delta x_n).$$

Hence, we must choose the Δx's to satisfy

$$\Delta x_n = \Delta x_{n-1} + \frac{1}{2k}(\beta_{n-1} - \beta_n). \tag{41}$$

With this choice, eq. (40) becomes

$$R = U^*(v_{1,1}) M_1 M_2 M_3 \cdots M_N U(v_{2,N}) \tag{42}$$

where

$$M_n = \begin{pmatrix} (1 - \mathrm{i}t_0)/2k & -(\mathrm{i}t_0/2k)\exp\{-2\mathrm{i}k(x_n - \Delta x_n)\} \\ (\mathrm{i}t_0/2k)\exp\{2\mathrm{i}k(x_n - \Delta x_n)\} & (1 + \mathrm{i}t_0)/2k \end{pmatrix}.$$

Thus, if the barriers are far enough apart so that $x_n - \Delta x_n > x_{n-1} - \Delta x_{n-1}$ for all n, the almost periodic potential barrier problem is equivalent to a problem with an array of delta functions whose positions and strengths are almost periodically modulated. By the methods of the last paragraph, this problem is equivalent to the generalized tight binding problem represented by eq. (32). Thus, this tight binding problem, which may be studied by the renormalization group and continued fraction methods discussed earlier in this section, is more general than one might have thought at first. Of course, since the parameters in the equivalent delta function problem are all energy dependent, a different problem must be solved for each energy in the spectrum. Despite this fact, the discussion in this subsection shows that the question of localization in any Schroedinger equation whose potential consists of an almost periodic array of finite and smooth potential barriers, separated by regions of zero potential energy, can be answered by solving a tight binding-like Schroedinger equation of the form of eq. (32). Thus, the information gained by solving such models applies directly to fairly general realistic differential Schroedinger equations.

2.5. The tangent potential model

To complement physical arguments for the existence of localized states in almost periodic Schroedinger equations presented earlier in this section, it would be useful to have an exactly soluble almost periodic problem which exhibits localization. Such a model was provided by Prange, Grempel and Fishman [28, 29]. They consider the following generalization of the tight binding model of eq. (1)

$$\sum_L h_L f_{n+L} + V(n) f_n = E f_n. \tag{43}$$

Here, L is an integer, h_L is a general range hopping matrix element and $V(n)$ is an almost periodic potential. For the special case where $V(n) = \tan[(\omega - QK(n))/2]$, this model can be shown to be equivalent to the following Schroedinger equation for a quantum rotor which is periodically given an angular impulse:

$$\mathrm{i}\,\partial\psi(\theta,t)/\partial t = [QK(-\mathrm{i}\partial/\partial\theta) + A(\theta)\Delta(t)]\,\psi(\theta,t)\,, \tag{44}$$

where $A(\theta+2\pi)=A(\theta)$, where $\Delta(t)=\Sigma_n\,\delta(t-nt_0)$, and t_0 is the time between impulses. For the case in which $K(x)=x$, and Q, an irrational multiple of 2π, eq. (43) is a tight binding almost periodic Schroedinger equation, and eq. (44), to which it is equivalent, is exactly soluble. The equivalence of eqs. (43) and (44) is shown using the following arguments: Equation (44) can be written as

$$\mathrm{i}\,\partial\ln\psi/\partial t + \mathrm{i}Q\,\partial\ln\psi/\partial\theta = A(\theta)\sum_n \delta(t-nt_0)\,. \tag{45}$$

We now integrate this equation over time from $t=nt_0^-$ to nt_0^+. Assuming that ψ and hence the second term on the right hand side of eq. (45) has a discontinuity at $t=nt_0$ but remains finite, we get

$$\psi^+(\theta,t) = \psi^-(\theta,t)\,\mathrm{e}^{-\mathrm{i}A(\theta)}\,. \tag{46}$$

Between $t=nt_0^+$ and $t=(n+1)t_0^-$, ψ satisfies eq. (44) with the second term on the right hand side absent (i.e. a free particle Schroedinger equation), which can be rewritten as

$$\psi^-(\theta,t+t_0) = \exp\{-\mathrm{i}t_0QK(-\mathrm{i}\partial/\partial\theta)\}\,\psi^+(\theta,t) \tag{47a}$$

or

$$\psi^-(\theta,t+t_0) = \exp\{-t_0Q\partial/\partial\theta\}\,\psi^+(\theta,t) = \psi^+(\theta-Q,t)\,. \tag{47b}$$

Since the potential on the right hand side of eq. (44) is periodic in time we may apply a Bloch theorem to the time variable to show that $\psi(\theta,t)$ must have the form

$$\psi(\theta,t) = \mathrm{e}^{-\mathrm{i}\omega t}u(\theta,t)\,, \tag{48}$$

where $u(\theta_0,t+t_0)=u(\theta,t)$. Here ω is called the "quasi-energy" in analogy to the quasi-momentum in a spatially periodic problem: The equation (47a) can be written as

$$u^-(\theta,t) = \exp\{\mathrm{i}\omega t_0\}\exp\{-\mathrm{i}t_0QK(-\mathrm{i}\partial/\partial\theta)\}\,u^+(\theta,t)\,. \tag{49}$$

If we write

$$\mathrm{e}^{-\mathrm{i}A(\theta)} = \frac{1+\mathrm{i}h(\theta)}{1-\mathrm{i}h(\theta)}, \tag{50}$$

and substitute into eq. (46), we obtain

$$[1-\mathrm{i}\,h(\theta)]\,u^+(\theta,t) = [1+\mathrm{i}\,h(\theta)]\,u^-(\theta,t)\,. \tag{51}$$

Since in order for the solution to eq. (44) to be single valued $u(\theta, t)$ must $= u(\theta+2\pi, t)$ and since $A(\theta)$ and hence $h(\theta)$ are periodic, we may write these three functions in a Fourier series,

$$u(\theta, t) = \sum_m u_m(t)\,\mathrm{e}^{im\theta} \tag{52a}$$

etc. where

$$u_m = \frac{1}{2\pi}\int_0^{2\pi} u(\theta, t)\,\mathrm{e}^{-im\theta}\,\mathrm{d}\theta\,. \tag{52b}$$

Then, Fourier transforming (49) over θ, we have

$$u_m^-(t) = \mathrm{e}^{\mathrm{i}(\omega - Qm)}\,u_m^+(t)\,. \tag{53a}$$

Fourier transforming eq. (51) and substituting u_m^- from (53a), we have

$$u_m^+ - \mathrm{i}\sum_L h_L u_{m+L}^+ = \mathrm{e}^{\mathrm{i}(\omega-Qm)}\Big[u_m^+ + \mathrm{i}\sum_L h_L u_{m+L}^+\Big] \tag{53b}$$

where h_L is the Fourier transform of $h(\theta)$. This can be rewritten as

$$-\tan(\omega - Qm)\,\bar{u}_m + \sum_{L\neq 0} h_L \bar{u}_{m+L} = -h_0\bar{u}_m\,, \tag{54}$$

which is precisely eq. (43) if we identify $f_m \to \bar{u}_m$, $-h_0 \to E$ and $-\tan(\omega Qm) \to V(m)$.

The reason that the equivalence that we have just proven is useful is because, as will be shown below, eq. (44) is exactly soluble.

We now will study the solutions to eq. (44) in order to learn about the solutions to eq. (43). Combining eqs. (46), (47b) and (48), we obtain

$$u^+(\theta, t) = \exp\{\mathrm{i}(\omega t_0 - A(\theta))\}\,u^+(\theta - Q, t)\,. \tag{55a}$$

This implies that

$$|u^+(\theta, t)| = |u^+(\theta - Q, t)|\,. \tag{55b}$$

Since we are interested in the case in which the potential in eq. (43) is almost periodic we choose Q to be an irrational multiple of 2π. This combined with the fact that $u^+(\theta+2\pi, t) = u^+(\theta, t)$ and eq. (55b) implies that the magnitude of $u^+(\theta, t)$ is independent of θ. Then, we have that

$$u^+(\theta, t) \sim \mathrm{e}^{\mathrm{i}\phi(\theta, t)}\,. \tag{56}$$

By the single valuedness of u^+, we obtain

$$\phi(\theta+2\pi, t) = \phi(\theta, t) + 2\pi\nu\,, \tag{57}$$

where ν is an integer. Then, from eq. (57) we see that $\phi(\theta, t)$ may be written as

$$\phi_\nu(\theta, t) = \nu\theta + F(\theta, t) \tag{58}$$

where $F(\theta+2\pi, t) = F(\theta)$. Then, from eq. (55a)

$$\exp\{i(\nu\theta + F(\theta, t))\} = \exp\{i[\omega t_0 - A(\theta)]\}\exp\{i[\nu(\theta-Q) + F(\theta-Q, t)]\}\,.$$

Writing F and A in terms of their Fourier series we obtain

$$\nu\theta + \sum_m F_m(t)\,e^{im\theta} = \omega t_0 + \nu\theta - \nu Q - \sum_m A_m\,e^{im\theta} + \sum_m F_m(t)\,e^{im(\theta-Q)}\,. \tag{59}$$

We may find a solution if we set

$$\omega t_0 = A_0 + \nu Q \tag{60a}$$

and

$$F_m = -A_m/(1-e^{-imQ})\,. \tag{60b}$$

From eq. (55b), we have

$$\sum_n |u_n^+| = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta\,|u^+(\theta, t)| = \text{a finite number}\,. \tag{61}$$

Then, all the eigenfunctions of eq. (43) are normalizable, and hence localized for any strength potential. In order to study the way in which they decay to zero for large values of $|n|$ we must first find the form of A_n. Using eq. (60b) if we arbitrarily choose $A(\theta) \propto \cos\theta$, the Fourier series for A, and hence $F(\theta)$, has only two terms which give $F(\theta) = B(t)\sin(\theta + Q/2)$ where $B(t)$ is some function of t. Substituting this expression into eq. (58) and eq. (59) into (56), and using

$$e^{B\sin\theta} = \sum_m J_m(B)\,e^{im\theta}$$

we find that $u_m \propto J_m(B)$. Using the asymptotic behavior of the Bessel function $J_m(B)$, we find $u_m \sim B^m/2^m m!$ for $m \gg B$, which decays for large m at a rate faster than exponential. For near neighbor hopping $h_m = h_1\,(\delta_{m,1} + \delta_{m-1})$ and hence $h(\theta) = 2h\cos\theta$. Substituting for A from eq. (50), finding its

Fourier transform and substituting in eqs. (60), (58) and (56), we find that the wave function of the tight binding problem [i.e. eq. (43)] decays as

$$u_n^+ \propto C_{|n-\nu|}\,\mathrm{e}^{-\gamma|n-\nu|} \tag{62a}$$

where

$$2h_1 \cosh\gamma = \sqrt{1+(E+h_1)^2} + \sqrt{1+(E-h_1)^2} \tag{62b}$$

where $C_n < n\exp(2\sqrt{n/p_n})$ with $p_n \approx \min[L\sin(L\tau/2)]$. The dependence of C on n turns out to be much weaker than the exponential factor. The Fourier series for $F(\theta, t)$ converges when $Q = 2\pi$ times an irrational number not too well approximated by rationals. When $Q/2\pi$ is a rational number or an irrational extremely well approximated by rationals (i.e. a Liouville number) what we have just said about exponential decay is no longer correct. The discussion of the wave function for the model in this case will be presented in section 4. The fact that for $Q/2\pi$ an irrational not well approximated by rationals there exist only localized states, in contrast to the Aubry model for which it is believed that there can be both localized and extended states, is probably due to the divergent nature of the tangent potential in eq. (43). Although this is an unphysical feature, many of the qualitative phenomena found for the model are characteristic of continuous potential models. Still one must keep in mind the fact that although this is an exactly soluable almost periodic problem which exhibits localized states, the singular nature of the potential obviously aids localization.

2.6. Broadening of states in almost periodic potentials

We close section 2 with another application of localization in almost periodic potentials. Consider eq. (1) for $V_0 < t$. All states will be extended in configuration space in this case but localized in k-space. According to the arguments given earlier in this section, this means that the self-energy for the k-space representation of the Green function will not have a branch cut. This implies that the electronic states will not be broadened by the almost periodic potential. Another way of saying this is that wave vector is a good quantum number. Since eq. (1) also describes phonons in a chain of atoms situated in a periodic potential incommensurate with the lattice (with suitable identifications of t and V_0 with the force constants and E with $-\omega^2 + 2t$), phonons will not be broadened by the almost periodic potential despite the fact that such a crystal lacks translation group symmetry. This was pointed out by Sokoloff et al. [30]. Of course when the modulation potential reaches the critical strength for the metal–insulator transition, all phonon modes will suddenly broaden out as they become localized in configuration space.

3. Band structure

If in eq. (1) $Q = 2\pi p/qa$ where p and q are integers, the crystal is periodic and has a unit cell of length qa, and hence, a Brillouin zone size $2\pi/qa$. Such a periodic system has gaps at the Brillouin zone boundaries $\pi L/qa$ where L is an integer and bands in between the gaps. Clearly in the incommensurate limit (i.e. the limit as $q \to \infty$), the bands become more and more fragmented with gaps and the bands between the gaps become infinitesimally narrow. Yet we have seen in the last section that for weak potentials, the system has extended states and hence metallic behavior. The questions that we wish to address in this section on the band structure of almost periodic crystals are

1. How can the existence of extended states for weak potentials be reconciled with the existence of a completely fragmented band structure?

2. What are the sizes and locations of the bands and gaps?

3.1. Perturbation theoretic methods

For weak potentials eq. (1) can easily be solved by applying perturbation theory in the potential term in eq. (1) (i.e., the nearly free electron model [31]). To do this, let us apply Wigner perturbation theory to the dual model, given in eq. (3). Consider $g_{m=m_0} \approx 1$ and all other $|g_m| \ll 1$. Then, from eq. (3), we find

$$E \approx 2t\cos(m_0 Qa + ka) + V_0(g_{m_0+1} + g_{m_0-1}), \tag{63}$$

where g_m is found by a perturbation theory developed by iterating eq. (3) for $m > m_0$ (upper sign) and $m < m_0$ (lower sign),

$$g_{m_0 \pm m} = \frac{V_0^m}{\varepsilon_1 \varepsilon_2 \cdots \varepsilon_m} \times \frac{1}{\left[1 - \dfrac{V_0^2}{\varepsilon_1\varepsilon_2}\right]\left[\dfrac{1 - V_0^2/(\varepsilon_2\varepsilon_3)}{1 - V_0^2/(\varepsilon_1\varepsilon_2)}\right] \cdots \dfrac{[1 - V_0^2/(\varepsilon_m\varepsilon_{m-1})]}{1 - \dfrac{V_0^2/(\varepsilon_{m-1}\varepsilon_{m-2})}{1 - \dots}}} g_{m_0} + \frac{\dfrac{V_0}{\varepsilon_m} g_{m_0 \pm (m+1)}}{1 - \dfrac{V_0^2/(\varepsilon_m\varepsilon_{m-1})}{1 - \dfrac{V_0^2/(\varepsilon_{m-1}\varepsilon_{m-2})}{1 - \dots}}} \tag{64}$$

where $\varepsilon_m = E - 2t\cos[Q(m_0 \pm m)a + h]$. Setting $g_{m_0} = 1$ in eq. (64) and substituting in the right hand side of eq. (63) for $g_{m_0 \pm 1}$, we obtain a perturbation theory expression for the energy of the form

$$E - 2t\cos(m_0 Qa + ka) = \sum_m \sum_{\pm} \frac{V_0^{m+2}}{\varepsilon_1 \varepsilon_2 \cdots \varepsilon_{m-2} \varepsilon_{m-1}^2} \times \frac{1}{\left[1 - \dfrac{V_0^2}{\varepsilon_1\varepsilon_2}\right]\left[\dfrac{1 - V_0^2/(\varepsilon_2\varepsilon_3)}{1 - V_0^2/(\varepsilon_1\varepsilon_2)}\right] \cdots \left[\dfrac{1 - V_0^2/(\varepsilon_{m-1}\varepsilon_{m-2})}{1 - \dfrac{V_0^2/(\varepsilon_{m-2}\varepsilon_{m-3})}{1 - \cdots}}\right]\left[\dfrac{1 - V_0^2/(\varepsilon_m\varepsilon_{m-1})}{1 - \dfrac{V_0^2/(\varepsilon_{m-1}\varepsilon_{m-2})}{1 - \cdots}}\right]}. \tag{65}$$

A solution to eq. (65) for small values of V_0 will occur in the vicinity of an energy for which the right hand side of eq. (65) has a pole. This is illustrated in fig. 1. This will occur when one of the last two continued fractions in the denominator vanishes. The residue of this pole is of the order of $V_0^{m+2}/(\varepsilon_1\varepsilon_2 \cdot$

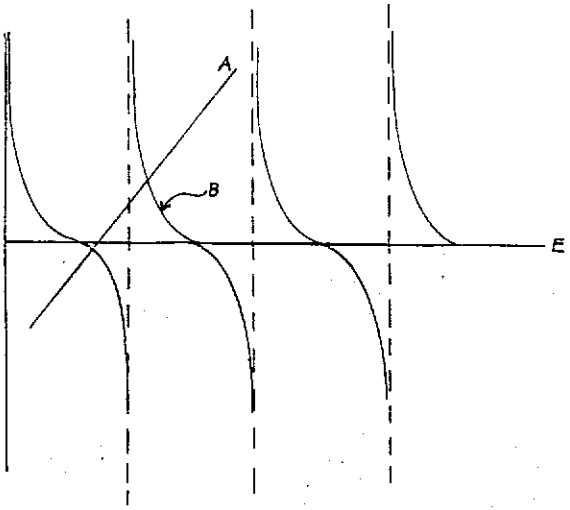

Fig. 1. Figure on solution of Wigner perturbation theory in eq. (65) for the energy. The horizontal axis gives the energy E and the vertical axis gives the value of the function. The line labeled A is the left hand side of eq. (65) and the line labeled B is the right hand side. The dotted line denotes the location of a pole in eq. (64).

$\cdots \varepsilon^2_{n-1}$) [the other continued fractions do not change this result much when they do not vanish]. When the denominator of this expression for the residue does not vanish, it may be approximated by its geometric mean, which gives a residue of order V_0^{m+2}/t^m. There is a solution to eq. (65) in the vicinity of each such pole, corresponding to a pair of bands. The two bands for the nth order pole are separated by a gap of order $V_0(V_0/t)^{m/2}$. Every order in the perturbation expression in eq. (65) gives another pair of bands and hence another gap in the spectrum, but for $V_0 < t$, most of these gaps will be negligibly small.

The same perturbation theory method can be applied for $V_0 > t$ to the dual model [i.e. eq. (1)]. Now we replace $g \to f$, $V_0 \to t$, $t \to V_0$. The wave function is now found from eq. (64) to be localized with mean localization length $a \ln(V_0/t)$, as found by Aubry.

3.2. Quasiclassical methods

Another way to study the fragmented nature of the band structure in the metallic regime is to use the quasi-classical approximation, valid when the potential is very slowly varying as a function of position. When Q in eq. (1) is very small (or very close to $2\pi/a$), we may clearly use the quasi-classical approximation to study the band structure. This is the method used by Azbel [25]. As will be seen in section 6, this is the limit which is relevant for the problem of an electron in a magnetic field. Here this method will be generalized to the case in which Qa is very close to being equal to a rational multiple of 2π [32]. Since any irrational number, except perhaps those which are extremely poorly approximated by rationals, may be approximated to any desired degree of accuracy by a rational number of sufficiently high order, this is a useful limit. Since we may always write $Q = Q_0 + \Delta Q$ where Q_0 is a rational approximation to Q and ΔQ is the correction, eq. (1) may be written as

$$t(f_{n+1} + f_{n-1}) + 2V_0 \cos(Q_0 na + ha + \Delta Q\, na) f_n = E f_n\,. \tag{66}$$

This is not the usual quasi-classical limit which we are considering because although $\Delta Q\, na$ varies slowly with n for small ΔQ, the potential generally varies rapidly with n. Nevertheless, if ΔQ is sufficiently small we may diagonalize eq. (66), treating $h+\Delta Q\,na$ as a constant phase variable, giving a set of band energies which are functions of $\Delta Q\,na$ as well as of the wave vector. These can be thought of as classical trajectories which form the starting point for a WKB approximation calculation. This approximation may be developed by looking for solutions of eq. (66) of the form

$$f_n = f(x) = \sum_m A_m(x)\,e^{i(S(x)+mQx)} \tag{67}$$

where $x = na$ and $dS(x)/dx$ and $A_m(x)$ are assumed to be slowly varying functions of x, which must be the case because for $\Delta Q \equiv 0$, eq. (67) with A independent of x and $S \propto x$ is clearly the solution to the resulting commensurate problem. Since for the case of $t > V_0$, we showed in section 2 that $f(x)$ is an analytic function of x, we may approximate the difference operator in eq. (66) by a differential operator [i.e. $f_{n+1}+f_{n-1} = 2\cos((1/i)a\,d/dx)f(x)$]. Then, substituting $A_m \exp\{i[S(x)+mQ]\}$ into eq. (66) and keeping only first derivatives of S (i.e. assuming that $a \ll$ the distance over which $dS(x)/dx$ varies), we have

$$\begin{aligned} &2t\cos(aS'(x)+mQa)\,A_m(x)\exp\{i(S(x)+mQx)\} \\ &\quad + 2V_0\,[\exp\{i(x\,\Delta Q+ha)\}\,A_m(x)\exp\{i[S(x)+(m+1)Qx]\} \\ &\quad + \exp\{-i(x\,\Delta Q+ha)\}\,A_m(x)\exp\{i[S(x)+(m-1)Qx]\}] = E\,A_m(x)\exp\{i[S(x)+mQx]\}\,. \end{aligned} \tag{68}$$

Equation (68) may be formally diagonalized if we pretend that $S'(x) = dS\,dx$ and $x\,\Delta Q$ are independent of x. We obtain

$$E_\alpha(S'(x), x\,\Delta Q+ha)\,f(x) = Ef(x)\,, \tag{69}$$

where $E_\alpha(k,h)$ is the αth band of the commensurate problem obtained by setting $\Delta Q = 0$ in eq. (69). Equation (69) is the starting point for the WKB approximation if we solve for $S'(x)$ in eq. (69). At "classical turning points" we simply replace $S'(x)$ by $(a/i)\,d/dx$ and solve the resulting Schroedinger equation to get connecting formulas or use the Landau and Lifshitz analytic continuation method [33]. If we apply this procedure, we find that the states in the classically localized regime, i.e. the energy region in which the classical trajectories given by equation are localized, the classically localized trajectories are broadened into narrow bands. Well into the band (i.e. in the classically allowed regime) we clearly have a continuous region of extended band-like states, since the extended "classical" trajectories are clearly a good approximation to the wave functions. In the classically allowed region, but close to the beginning of the classically localized regime (i.e. the "classical" mobility edge), we may solve the WKB equations in a way similar to that done by Zilberman [34]. We find that many small gaps are introduced whose size decays as $\exp\{-c|E-E_B|/\Delta Q\}$ where E_B is the classical mobility edge defined above and c is a constant. Thus, for small ΔQ we see that these gaps become negligibly small as we move well into the bands of the commensurate problem (i.e. the problem for $Q = Q_0$). Hence, we expect the band structure of the incommensurate problem to look much like that of the commensurate problem, but with a series of negligibly small band gaps introduced near the band edges. Clearly, as V_0 approaches t, these gaps must become more important and we must use a higher order rational Q_0 to approximate Q in order to use this method.

Unfortunately, a systematic formulation of the quasi-classical approximation to the almost periodic potential problem, based on equations [66–69] has as yet not materialized. Wilkinson has, however, recently developed a systematic way of formulating the quasi-classical approximation, by using the transmission matrix method discussed in section 2.2 of this article [34a]. For Q_0, the rational approximation to Q, given by

$$Q_0 = \frac{2\pi}{a}\frac{p}{q},$$

where p and q are integers, it is convenient to define the matrix

$$M(ha) = \begin{pmatrix} Q_{q-1} & P_{q-1} \\ Q_{q-2} & P_{q-2} \end{pmatrix}, \tag{70}$$

in the notation of eq. (21). If Q were rational (i.e. $\Delta Q = 0$), the allowed energy eigenvalues would be determined by the requirement that the eigenvalues of M^n do not diverge as the integer n approaches infinity. This means that the eigenvalues of M lie on the unit circle. This commensurate case will be discussed in the next section.

For the incommensurate case,

$$\Psi(x') = G(x', x)\,\Psi(x), \tag{71}$$

where

$$\Psi(x) = (f_{n-1}, f_{n-2}) \qquad \text{and} \qquad \Psi(x') = (f'_{n-1}, f'_{n-2}), \tag{72}$$

where

$$G(x', x) = M(x' - \beta)\cdots M(x + 2\beta)\,M(x + \beta)\,M(x)\,\Psi(x), \tag{73}$$

where $\beta = q\,\Delta Q\,a$.

Assuming that M can be diagonalized, we may write

$$M(x) = U^{-1}(x)\,D(x)\,U(x), \tag{74}$$

where $D(x)$ is diagonal. Then,

$$G(x', x) = U^{-1}(x')\,g(x', x)\,U(x), \tag{75}$$

where

$$g(x', x) = D(x')\,[1 + \beta\,V(x' - \beta)]\,D(x' - \beta)\,[1 + \beta\,V(x' - 2\beta)]\cdots[1 + \beta\,V(x)]\,D(x), \tag{76}$$

where

$$1+\beta\, V(x)=U(x+\beta)\, U^{-1}(x)\,. \tag{77}$$

To a good approximation

$$V(x)=\lim_{\beta\to 0}\,(\mathrm{d}U(x)/\mathrm{d}x)\, U^{-1}(x)\,. \tag{78}$$

Here β plays the role of h in the present quasi-classical method. Collecting powers of β in eq. (76), we obtain

$$g(x',x)=g_0(x',x)+\beta\sum_{n=0}^{N} g_0(x',x''+\beta)\, V(x'')\, g_0(x'',x)+\cdots \tag{79}$$

where

$$x''=x+n\beta\,,$$

which when summed gives

$$g(x',x)=g_0(x',x)+\beta\sum_{n=0}^{N} g_0(x',x''+\beta)\, V(x'')\, g(x'',x)\,, \tag{80}$$

$$g_0(x',x)=D(x')\,D(x'-\beta)\cdots D(x+\beta)\,D(x)\,. \tag{81}$$

Equation (80) can be solved within the adiabatic approximation to yield

$$G(x',x)=U^{-1}(x')\, g_0(x',x)\exp\left[\int_x^{x'}\mathrm{d}x''\, V(x'')\right] U(x)+\mathrm{O}(\beta) \tag{82}$$

where

$$g_0(x',x)=\exp\left[\frac{\mathrm{i}}{2}(S'(x)+S'(x'))\right]\exp\left[\frac{\mathrm{i}}{\beta}S(x',x)\right]+\mathrm{O}(\beta)\,, \tag{83}$$

where $S(x)$ and $S(x',x)$ are diagonal matrices defined by

$$D(x)=\exp\{\mathrm{i}S'(x)\} \tag{84}$$

and

$$S(x',x)=\int_x^{x'}\mathrm{d}x''\, S'(x'') \tag{85}$$

where $S'(x)=\mathrm{d}S(x)/\mathrm{d}x$.

By making use of a current operator (which is conserved under the transformations), Wilkinson was able to derive the Bohr–Sommerfeld quantization conditions for the incommensurate potential problem with wave vector close to an arbitrary commensurate value, and in this way obtain the quantized energy levels for closed orbits. One surprising result is that the wave function possesses a non-holonomic phase (i.e. when the wave function is transported clockwise around a closed path in phase space, it picks up a phase factor $\exp(i\gamma_c)$). This result leads to a modification of the usual Bohr–Sommerfeld quantization condition. For Q close to the fundamental registry, i.e. $Q \approx 2\pi n/a$ where $n = 0$ or any integer, of course, this correction to the quantization condition vanishes. Further splitting of the energy levels caused by tunneling between them was not considered.

3.3. Band structure at the metal–insulator transition critical point

Much attention has been given recently to the critical point $V_0 = t$ for the Aubry model, where the metal–insulator transition takes place. As we have seen, for either $V_0 < t$ or $V_0 > t$, all states are localized in position space and extended in wave vector space or all states are extended in position space but localized in wave vector space. Since the density of states for the position space problem and its dual, the wave vector space problem, must obviously be the same, we conclude from the contents of the last two subsections that it must be continuous like that of an ordinary metallic system. For $V_0 = t$, however, since the band structure becomes completely fragmented in both wave vector and position space the density of states must also be fragmented. Furthermore, all the classical orbits except one in the center of the bands are localized for $V_0 = t$ in both the quasi-classical approximation for small Q and for $Q \approx$ a rational multiple of $2\pi/a$ [35]. These are all indications that something unusual happens at this point. In this section we will discuss the behavior of the spectrum at the critical point. In section 4, the wave functions will be considered.

The unusual spectral structure which occurs for $V_0 = t$ was first considered by Azbel [36]. He argued, using the WKB approximation, that if we chose Q in eq. (1) to have the form $Q = (2\pi/a)\beta_0$ where

$$\beta_j^{-1} = N_j + \beta_{j+1}, \tag{86}$$

with each member of the set of integers $\{N_j\}$ large we may proceed as follows: If we approximate β_0 by N_0^{-1} we may simply diagonalize eq. (1). There will clearly be N_0 bands of extended states. Almost all of these bands, however, will be quite narrow. If the N_j's are all large, stopping at $\beta_0 = N_0^{-1}$ should be a good approximation, but if we wish to obtain the next approximation we keep $\beta_1 = N_1/(N_0N + 1)$. Then, in this approximation we have $N_0N_1 + 1$ bands, and hence we expect that each of the N_0 bands in the previous approximation will be split into approximately N_1 bands. Again, we expect each of these bands to be quite narrow. In the next approximation to β_0, we have $N_0N_1N_2 + N_0 + N_2$ bands, and hence, for large N's, each band from the previous approximation is split approximately N_2 times, etc. For small β_0, the quasi-classical approximation should be good. Since most of the classical trajectories are far from each other, we do not expect much splitting of these levels. Instead it should be a good approximation to neglect the tunneling between them (which leads to the splitting), and to determine the allowed orbits using the Bohr–Summerfeld quantization condition. For orbits close to touching, however, the splittings described above become important and Azbel used the WKB approximation to calculate the spectrum [36]. A similar approach has recently been introduced by Wilkinson [35]. Since almost all classical trajectories are localized in both momentum and position, he determines a set of approximate levels by using Bohr–Sommerfeld quantization, and then uses these localized levels as a basis for constructing the eigenstates, including effects of tunneling between orbits perturbationally. (There may be a problem, however, with applying this method for $V_0 = t$. Since the metal–insulator transition occurs at this point and f_n is not analytic for localized states, it is not obvious that the WKB approximation, which depends on replacing difference by differential operators may be used.) The energy levels are found to have the form

$$E_m = E_m^0 + C_m \cos(2\pi h/a\beta_0) \tag{87}$$

where E_m^0 is the energy predicted by the Bohr–Sommerfeld method and C_m is the coupling between orbits, h is the phase variable in eq. (1) and m labels the Bohr–Sommerfeld levels. Since h gives the locations of the localized orbits $h = na$, and hence

$$E_m = E_m^0 + C_m \cos(2\pi n/\beta_0) = E_m^0 + C_m \cos(2\pi n\beta_1) \tag{88}$$

using eq. (86). Then, each level is split into N_1 new levels. Equation (88) implies an equation for the amplitudes of the localized orbits, with the same form as eq. (1) but with β_1 replacing β_0. Repeating the procedure, we find that each level is split into N_2 new levels. This procedure may be continued, showing that each level is split into an infinite manifold of levels. This behavior of the spectrum has recently been verified numerically by Rubinstein and Azbel [37]. Azbel studied this problem in connection with the problem of an electron in both a magnetic field and a periodic lattice [36]. It will be shown in section 6 that this problem is identical with the one-dimensional almost periodic problem expressed by eq. (1) for both the tight binding and extremely weak lattice potential limits. In the magnetic field problem, β_0 is proportional to the magnetic field, and hence this analysis shows that as the magnetic field (actually the number of flux quanta through a unit cell) is varied through rational and irrational values, the level structure will change according to how the continued fraction expansion of β_0 given in eq. (86) changes. The degree to which this might be experimentally observable will be discussed in section 6.

Because of the interest in the variation of the energy level spectrum with β_0 in connection with the magnetic field problem, Hofstadter has studied the energy level structure as a function of β_0 numerically and has drawn some interesting conclusions [38]. Hofstadter studied the spectrum of eq. (1) for rational values of Q only. For rational values of $Q = 2\pi p/aq$ with p and q integers, eq. (21) may be written as

$$\begin{pmatrix} f_{Nq} \\ f_{Nq-1} \end{pmatrix} = \begin{pmatrix} Q_q & P_q \\ Q_{q-1} & P_{q-1} \end{pmatrix}^N \begin{pmatrix} f_1 \\ f_0 \end{pmatrix} = (M)^N \begin{pmatrix} f_1 \\ f_0 \end{pmatrix}. \tag{89}$$

From eq. (21), we conclude that the determinant of M is 1. Then, if λ_1 and λ_2 are the eigenvalues of M, $\lambda_1 = 1/\lambda_2$. Clearly, if λ_1 and λ_2 are real $|\mathrm{Trace}\, M| = |\lambda_1 + \lambda_2| > 2$. For this case, eq. (89) shows that the wave function grows exponentially. Then, since eq. (89) satisfies Bloch's theorem only when λ_1 and λ_2 are both complex numbers of unit magnitude, we conclude that only when $|\mathrm{Trace}\, M| < 2$ we are in a band. This criterion was used to locate the allowed energy bands as a function of β_0. Obviously, for β_0 rational, we will have a finite number of bands. Hofstadter uses these values of β_0 to form boundaries of subcells in the allowed energy versus β_0 plot. He then shows that each subcell contains a distorted replica of the band structure of the system. In fact, each subcell contains another set of subcells, each of which contains a replica of the entire spectrum. To understand this procedure, let us first consider what Hofstadter calls the "pure cases", $\beta_0 = N_0^{-1}$ and $\beta_0 = 1 - N_0^{-1}$. For these values of β_0 the spectrum contains N_0 bands. We define a subcell bounded by $\beta_0 = (1 + N_0)^{-1}$ and N_0^{-1} and by the minimum energy of the band (i.e. $E=-2t$) and the lowest gap and another type of subcell bounded by the maximum energy $2t$ and the highest gap, and also, we may define a subcell bounded by $\beta_0 = 1-(N_0+1)^{-1}$ and $1-N_0^{-1}$ and by the lowest (highest) gap and the minimum (maximum) energy. Then, for example, we may write a value of β_0 in the first type of subcell as $\beta_0=(N_0+\beta_1)^{-1}$ where β_1 is a number between 0 and 1. Clearly, for most values of β_1, β_0 is not a rational number, and hence we might expect there to be an infinite number of bands at general points in the cell. In fact, what Hofstadter finds is that if we plot out the band structure inside this cell as a function of β_1, we obtain a spectrum that is a distorted version of the original energy spectrum as a function of β_0. The procedure may now be repeated with β_1, etc., ad-infinitum. For a given value of β_0, the spectrum follows the continued fraction expansion of β_0, as proposed by Azbel [36].

In order to consider more general irrational values of β_0, Hofstadter defines what he calls the "special cases", $\beta_0 = N_0/(2N_0+1)$ and $(N_0+1)/(2N_0+1)$ for $N_0 \geq 2$. He then defines a set of subcell boundaries by two successive values of N_0, and hence, inside a cell, we may write

$$\beta_0[2+(N_0+\beta_1)^{-1}]^{-1}$$

(and also there is a cell with $\beta_0 = 1-$ the above value of β_0) where β_1 goes from 0 to 1, to interpolate between $N_0/(2N_0+1)$ and $(N_0+1)/[2(N_0+1)+1]$. The other two boundaries are the lowest and highest gaps at the two values of β_0 which bound the cell. Then, there are two cells which Hofstadter calls C_0 and C_{-1} which are bounded by the upper and lower gaps at $\beta_0=\frac{1}{3}$ and $\frac{2}{3}$ and by $\beta_0=0$ and $\beta_0=1$ respectively. All of these cells are replicas of the whole spectrum as a function of β_0 with β_1 replacing β_0, and again the procedure may be continued ad infinitum. For a given irrational value of β_0, the splitting of the spectrum according to the continued function of β_0 predicted by Azbel was not found by Hofstadter in the latter subcells.

3.4. *Total bandwidth as a function of potential strength*

Hofstadter's and Azbel's ideas, expressed in the last paragraph do not really say precisely what the widths of the various subbands are compared to the size of the gaps separating them. Aubry and André, however, have calculated the sum of the subbandwidths of eq. (1) as a function of the potential strength in the metallic regime for Q equal to a rational multiple of $2\pi/a$ [7]. They consider this functional form for the rational approximation of $Qa/2\pi$. For the incommensurate limit, these curves approach the line $4(t-V_0)$. In the insulating regime, the total bandwidth for a fixed value of the phase is obviously zero, but total width of the energy bands which result from allowing the phase to vary is proportional to V_0-t. For any commensurate approximation, however, the bandwidth versus V_0 curves seem to lie above this curve in Aubry and André's calculation. In fact, Thouless has proven that the quantity $4|t-V_0|$ is a lower bound to the total bandwidth for any commensurate system [39]. The proof uses matrix evaluation methods valid for tridiagonal matrices and puts limits on the matrix elements. Thus, the spectrum considered by Azbel and Hofstadter for the case $t=V_0$ in eq. (1) has zero total bandwidth. Therefore, the energy level structure consists of an infinite number of flat levels for incommensurate values of $Qa/2\pi$.

3.5. *Counting of the number of states below a gap*

Wannier and Claro and Wannier have proven using the recursive level structure found by Hofstadter [40, 41] that any gap in the spectrum can be labeled by a pair of integers n and m (positive or negative) and that the fraction of the total number of states below that gap is equal to

$$\frac{a}{2\pi}|nQ + m(2\pi/a)|\,. \tag{90}$$

This theorem has been proven rigorously by Johnson and Moser [42] and by Delyon and Souillard [43], and has been called by Simon the "gap labeling theorem" [8] because it gives a way of labeling the infinite hierarchy of gaps in the spectrum of the almost periodic problem. The physical basis for this theorem can be understood using the following simple arguments: In one dimension, the fraction of states below the Fermi level is given by $(2a/2\pi)k_F$ where k_F is the Fermi wave vector. For rational values of $Qa/2\pi$, it is clear that a gap will occur at half of any reciprocal lattice vector or $\frac{1}{2}[nQ + m(2\pi/a)]$. Then, if the Fermi energy lies in this gap (which is specified by the integers n and m), k_F must be equal to this wave vector, and hence, we have eq. (90) for the fraction of states below this gap. Since this result holds for *any* rational value of $Qa/2\pi$ (including very high order rationals) it is likely to hold for irrational values of $Qa/2\pi$ too. It will be shown in section 6 to be quite important for understanding the quantized Hall effect in periodic lattices, and it was shown by Avron, Seiler and Simon [44] to be important for characterizing the bands of a periodic lattice.

3.6. The approach to incommensurability and scaling behavior of the spectrum

There have been several recent studies of the band structure of high order commensurate systems (which clearly have well defined energy bands) in the limit as $Q/2\pi$ approaches an irrational number in order to study self-similarity of the band structure for an almost periodic system. These studies focus on the scaling behavior of the band structure which occurs as the incommensurate limit is approached. For example, Kohmoto [46] has studied the Aubry model [i.e. eq. (1)] with $V_0 = t$ and with wave vector $Q_L/2\pi = F_{L-1}/F_L$ where F_L is the Fibonacci number, which is defined by the difference equation

$$F_L = F_{L-1} + F_{L-2} \tag{91}$$

with $F_0 = F_1 = 1$ (probably using one of the numerical methods discussed in section 2). For $L = 3$, there are 3 bands. For higher L these bands split further. As $L \to \infty$, $Qa/2\pi$ approaches $w = \frac{1}{2}(\sqrt{5} - 1)$. For $L = 6$, for example, the middle band for $L = 3$ has now split into three bands, which if magnified, look exactly like the bands for the $L = 3$ case. Also, the middle cluster of bands for $L = 6$ consists of three bands. The middle cluster for $L = 9$ consists of a group of three bands which are the middle three band clusters of the $L = 6$ case rescaled by a factor α. Kohmoto has carried this study up to $L = 21$ ($F_{21} = 17.711$). In all cases, if the middle cluster of bands for the $L + 3$ case is rescaled it reproduces the L case for all L. The reason that this works for $L + 3$ and L is that in the Fibonacci sequence, every third Fibonacci number is even. He finds that as L increases, α_0 approaches 14.01. He also finds some sort of scaling for $V_0 < t$, but the results are not conclusive.

Kohmoto also finds that the total bandwidth scales as a constant times $F_L^{-\delta}$ for a given Q_L, and he finds an upper bound for the power δ of $\frac{1}{3}\log \alpha/(\log w - 1)$. Thouless finds by numerical means that δ is 1 [39]. In fact, Thouless argues that the total bandwidth B given by the finite size scaling form

$$B = (t - V_0)\, f(\xi/F_L) \tag{92}$$

where ξ is the localization length (for $t > V_0$ the states are extended in configuration space but are localized in wave vector space with localization length ξ). For large ξ but finite F_L, the result for B should become independent of ξ, whereas for large F_L but finite ξ, the result becomes independent of F_L. Hence, for large F_L but not infinite and finite ξ, $f(\xi/F_L)$ should be linear in ξ/F_L, and here we get the previous result of $b \sim F_L^{-1}$ whereas for large ξ/F_L, f should go to a constant limit.

Many of the results in the incommensurate potential problem, such as the Cantor set band structure, the metal–insulator transition accompanied by a break in the analyticity of the tight binding wave functions as a function of phase and the chaotic (i.e. divergent) behavior of the map described by eqs. (18) and (19) are reminiscent of classical dynamical mapping theory. In order to attempt to understand the connection between dynamical mapping theory and the present problem, Kohmoto, Kadanoff and Tang [11], as well as Ostlund, Pandit, Rand, Schellnhuber and Siggia [12] have studied the following tight binding model:

$$t(f_{n+1}+f_{n-1})+V(nQa)f_n=Ef_n \tag{93}$$

where $Q=(2\pi/a)F_{L-1}/F_L$ and $V(x+2\pi)=V(x)$. The potential $V(x)$ is chosen to be a step potential of the following form:

$$\begin{aligned} V(x) = {} & V_0 \qquad \text{for } -2\pi w < x \le -2\pi w^3 \\ & V_1 \qquad \text{for } -2\pi w^3 < x \le 2\pi w^2 , \end{aligned} \tag{94}$$

where w is the inverse Golden Mean $[1/2(\sqrt{5}-1)]$. The periodicity of $V(x)$ is easily shown using the relation $w+w^2=1$. References [11] and [12] show that the study of the spectrum of this model reduces to the study of a two dimensional dynamical mapping problem. Their method is based on an equation similar to eq. (21), which may be written as

$$\Psi_n = M^{(n)}(Qa)\,\Psi_1 \tag{95a}$$

where

$$M^{(1)}(x)=\begin{pmatrix} E-V(x) & -1 \\ 1 & 0 \end{pmatrix}, \tag{95b}$$

$$\Psi_n=\begin{pmatrix} f_n \\ f_{n-1} \end{pmatrix}, \tag{95c}$$

and

$$M^{(L)}(nQa)\Psi_n = \Psi_{n+L} \tag{95d}$$

$$M^{(m_1+m_2)}(x) = M^{(m_1)}(x+m_2Qa)\,M^{(m_2)}(x)\,, \tag{96}$$

we may write

$$M^{(F_{L+1})}(x) = M^{(F_{L-1})}(x+F_LQa)\,M^{(F_L)}(x)\,, \tag{97}$$

using eq. (91). If we set $x = 0$ in eq. (97), we may also replace $F_L Q$ by zero if $Q = 2\pi w/a$. The reason for this is that our choice of potential in eq. (94), the definition of $M^{(1)}$ in eq. (95b), and the periodicity of the potential (i.e. $V(x+2\pi) = V(x)$) guarantees that $F_L Q$, modulo 2π falls between $-2\pi w^3$ and $2\pi w^2$. This is because $F_1 w = F_{L-1} - (-w)^{L+1}$ and w^{L+1} approaches zero as L approaches infinity. Therefore, $M^{(F_{L-1})}(F_L Q) = M^{(F_{L-1})}(0)$ for all L. Then we may write eq. (97) with $x = 0$ as

$$M_{L+1} = M_{L-1} M_L \,, \tag{98}$$

where $M_L = M^{(F_{L-1})}(0)$. Equation (98) is also valid for $Q = (2\pi/a)F_{m-1}/F_m$ if m is sufficiently large. Equation (98) may also be written as

$$M_{L-2}^{-1} = M_{L-1} M_L^{-1} \,. \tag{99}$$

Adding eqs. (98) and (99) gives

$$M_{L+1} + M_{L-2}^{-1} = M_{L-1} M_L + M_{L-1} M_L^{-1} \,. \tag{100}$$

Since each of the M matrices has a determinant of 1, if λ_l is one eigenvalue of M_l, λ_l^{-1} is the other, and therefore

$$\mathrm{Tr}\, M_l = \lambda_l^{-1} + \lambda_l = \mathrm{Tr}\, M_L^{-1} \,,$$

and

$$\mathrm{Tr}\, M_{L-1}(M_L + M_L^{-1}) = (\mathrm{Tr}\, M_{L-1})\,(\mathrm{Tr}\, M_L) \,,$$

where we have used the fact that since the determinant of $M_L = 1$, $M_L + M_L^{-1} = \overleftrightarrow{1}\, \mathrm{Tr}\, M_L$ where $\overleftrightarrow{1}$ is the unit matrix. Then, taking the trace of eq. (100) using these results we find

$$x_{L+1} = 2x_{L-1}x_L - x_{L-2} \tag{101}$$

where $x_L = \frac{1}{2}$ trace M_L. Clearly if $|\mathrm{Tr}\, M_L|$ is greater than 2, the eigenvalues are both real (one greater than and one less than 1 in magnitude). Then, if $Q = (2\pi/a)F_{L-1}/F_L$,

$$M^{NF_L}(0) = M^{(F_L)}((N-1)Qa) \cdots M^{(F_L)}(2Qa)\, M^{(F_L)}(Qa)\, M^{(F_L)}(0) = (M_L)^N \tag{102}$$

using $M^{(F_L)}(x + 2\pi) = M^{(F_L)}(x)$. Then, clearly if $|\mathrm{Tr}\, M_L| > 2$, the product of matrices in eq. (86) will diverge as N increases, meaning that the wave function will diverge, which means that we are in a gap. If for a particular energy $|\mathrm{Tr}\, M_L| < 2$ (i.e. $x_L < 1$), both eigenvalues are complex numbers of unit magnitude and the wave function will therefore remain finite meaning that we are in a band. Thus, by studying the difference equation in eq. (101), we can determine that location of the bands. The initial conditions are $x_1 = \frac{1}{2}(E - V_1)/t$, $x_0 = \frac{1}{2}(E - V_0)/t$ and $x_{-1} = 1$. Thus, eq. (101) can be used in place of solving eq. (12) for P_m and Q_m and then taking the trace of M using eq. (21). Equation (101) has been used by Kohmoto et al. to study the band structure of the model of eq. (94) as the incommensurate limit is approached. This model seems to lie at the critical point of the metal-insulator transition (which corresponds to $V_0 = t$ in the Aubry model) because the total bandwidth becomes zero in the incommensurate limit, and as we will see later, the wave function is neither exponentially localized nor extended, but lies somewhere in between.

So far in our discussion, the methods used by Kohmoto et al., do not accomplish anything which could not have been accomplished by the methods described in section 2 of this article. The main contribution of their work, however, is the connection with dynamical mapping theory which will now be described. From the discussion under eq. (97), we conclude that the mapping expressed by eq. (101) is valid for any value of Q of the form $(2\pi/a)F_{L-1}/F_L$. Since $F_L w = F_{L-1} - (-w)^{L+1}$, if we choose $Q = L(\pi/a)w$ and iterate eq. (101) up to a given large value of L it gives the value of x_L for the commensurate problem corresponding to that value of L (i.e. corresponding to $Q = (2\pi/a)F_{L-1}/F_L$). Thus, if $|x_L| < 1$ for a given value of L, we are in one of the bands corresponding to that commensurate problem. If as we increase L, $|x_L|$ becomes greater than 1, we are no longer in a band for the problem represented by this larger value of L. Equation (101) can be written as a three dimensional map

$$r_{L+1} = Tr_L = (2x_L y_L - z_L, x_L, y_L) \tag{103}$$

where $r_L = (x_L, x_{L-1}, x_{L-2}) = (x_L, y_L, z_L)$. The quantity

$$\lambda^2 = -1 + x_{L+1}^2 + x_L^2 + x_{L-1}^2 - 2x_{L+1}x_L x_{L-1} \tag{104}$$

is shown to be independent of this mapping (i.e. independent of L) by substituting x_{L-1} using eq. (101). This means that the mapping defined by eq. (103) maps the point r_L on a two dimensional surface defined by eq. (104), which is defined uniquely by the value of λ. By substituting the values of x_{-1}, x_0 and x_1 (since after eq. (102)) into eq. (104) with $L = 0$, we find that $\lambda = \frac{1}{2}|V_1 - V_0|/t$. For a given value of λ, the point r_L is confined to this two dimensional surface determined by λ and cannot reach a surface determined by a different value of λ. For a given value of λ, a given initial value of the energy determines the initial point $r_1 = (x_1, x_0, x_{-1})$ and hence uniquely determines the orbit. Now the methods of classical mapping theory may be applied to the present problem. Here, it should be remembered that x_L represents half of the trace of the matrix $M^{(F_L)}(0)$, and hence, describes the spectrum of a commensurate problem with a unit cell of length $F_L a$. (Note that although $M^{(F_L)}(0)$ is evaluated for $Q = (2\pi/a)w$ rather than $(2\pi/a)F_{L-1}/F_L$, because of the step function nature of the potential chosen $V(QF_L)$ is the same for either value of Q. Therefore, x_l does describe the commensurate problem.) If $|x_L| < 1$ for a given energy, that energy lies in a band, whereas if $|x_L| > 1$ it is in a gap. As $L \to \infty$ (i.e. we apply the mapping operator T an infinite number of times), for most energies x_L will escape to infinity (and hence certainly be greater than 1 in magnitude). Thus, most energies will lie in gaps. If N_0 initial points are determined by N_0 uniformly spaced energies, the number of points which have escaped to infinity after L iterations is given by $N_L \sim N_0 F_L^{-\delta'(\lambda)}$, where $\delta' = 0.2202$ at $\lambda = 0.4$ and $\delta' = 0.423$ at $\lambda = \sqrt{2}/2$, for example. The total bandwidth is found to decay in the same way as $L \to \infty$, which is not unexpected. By studying the behavior near the fixed points of the map T^6, Kohmoto and Oono [45] were also able to determine the scaling index of the spectrum (the parameter α discussed earlier in this section) by looking at the linearized version of T^6 (α is argued to be equal to the square root of the exponent of the linearized map). They study the heteroclinic fixed points for the map. This leads to a Smale horseshoe structure which explains the infinite hierarchical structure of the Cantor set spectrum. The method of Ostlund et al. is basically the same, except they discuss it in the language of a decimation technique on a string of matrices. Their decimation technique consists of grouping a product of many matrices in alternating groups of F_L and F_{L-1} matrices (where F_L is a Fibanaci number) and using eq. (98) to convert it to a string of alternating groups of F_{L+1} and F_L matrices.

Recently Ostlund and Pandit [46] have extended the methods of ref. [12] and have been able to carry out a detailed renormalization group study of the scaling behavior of the spectrum of the Aubry model [eq. (1)] for $Q = 2\pi w/a$, where w is the inverse Golden Mean. In their analysis the matrices M_{L-1} and M_L, where M_L is defined above eq. (98), are expanded in a Taylor series in their argument (i.e. the position variable) with 60 terms. This is possible because in eq. (99), which is the basis of their renormalization group transformation, the argument of M_{L-1} is very close to the argument of M_L modulo an integer, as discussed above eq. (99) in this article. For the step potential previously considered, the small deviation of the argument of M_{L-1} from that of M_L could be neglected because the potential did not depend on position over a finite region. For the Aubry model potential, however, this is not true, but the position dependence of the potential in the M-matrices may be expanded in a Taylor series expansion since the position variable modulo 1 is always very small. The calculations were all carried out for $E = 0$. The renormalization group transformation defined by eq. (98) can then be considered as a transformation acting on the parameters in the Taylor series expansion of the elements of these matrices. Let us denote this transformation by T and the initial values of the collection of parameters on which it acts by f_0. The system actually moves cyclically among 6 fixed points after many applications of T, but we can find unique fixed points of T^6 by

$$\lim_{L\to\infty} T^{6L} f_0 = f^* . \tag{105}$$

If we denote by $u\delta$, where δ is a small number, a small variation of f_0 away from the fixed point, we may write

$$T^{6L}(f_0 + u\delta) \approx T^{6L} f_0 + \delta J_{T^{6L}} J_{T^{6L-1}} \cdots J_{T^6} u , \tag{106}$$

where $J_{T^{6L}}$ is the Jacobian matrix of T^{6L}.

As $L \to \infty$, this reduces to

$$T^{6L}(f_0 + u\delta) = f^* + \delta \lambda_1^L e_1 + \cdots , \tag{107}$$

where λ_1 is the largest eigenvalue for $J_{T^{6L}}$ and e_1 the corresponding right eigenvector. This relationship may be used to find the scaling behavior of the energy with K as follows: If we vary the matrices by changing the energy, δ denotes this deviation of the energy from $E = 0$. From eq. (107) we see that δ/λ_1, for $L+1$ gives the same result as δ for L. Then,

$$\Delta K'(\delta) = \Delta K'(\delta/\lambda_1) \tag{108}$$

where $\Delta K'(\delta)$ is the change in the wave vector in the rescaled variables corresponding to the change δ in the energy. The wave vector $\Delta K'$ depends only on the wave functions (i.e. the M matrices), which are unchanged when we go from L to $L+1$ and change δ to δ/λ_1 by the above argument. To get ΔK in the non-rescaled variables, we write $\Delta K'(\delta) = F_L\, \Delta K(\delta)$, where ΔK is the change in wave vector in unscaled variables. Then, we have

$$F_L\,\Delta K(\delta) = F_{L+6}\,\Delta K(\delta/\lambda_1)\,. \tag{109}$$

Hence

$$w^{-6}\,\Delta K(\delta) = \Delta K(\delta/\lambda_1)\,, \tag{110}$$

as $L\to\infty$, using $w^{-6} = F_L/F_{L+6}$ as $L\to\infty$. Since $\delta = E(K+\Delta K) - E(K)$, we may solve this equation for δ/λ_1 by inverting $\Delta K(\delta/\lambda_1)$ to give

$$\delta/\lambda_1 = \frac{E(K+\Delta K) - E(K)}{\lambda_1} = E(K + w^{-6}\,\Delta K) - E(K)\,. \tag{111}$$

As seen from eq. (107) this relationship is valid asymptotically in the limit as δ, and hence ΔK, approaches zero. Therefore, if we write ΔK as $xw^{-6(n-1)}$ where x is some number, eq. (111) can be written in the form

$$\lambda_1 = w^{-6\tilde{\gamma}} = \lim_{n\to\infty} \frac{E(K + w^{-6(n-1)}x) - E(K)}{E(K + w^{-6(n)}x) - E(K)} \tag{112}$$

where $\tilde{\gamma}$ is not equal to 1 for the almost periodic system. This is the scaling relation written down by Ostlund and Pandit. It shows how the size of an energy region corresponding to an asymptotically small region in K space of width ΔK changes when ΔK is rescaled by w^{-6}. The reason for defining the scaling index $\tilde{\gamma}$ in the way that we have is that $\tilde{\gamma}$ actually represents the degree of non-analyticity of the energy bands as a function of K due to the Cantor set band structure in this problem. If $E(K)$ were analytic, as they are for the periodic problem, the limit in eq. (112) would clearly give w^{-6} on the left hand side. Equation (109) is the main result of Ostlund and Pandit's work. The amazing thing about their work is that they have been able to obtain information about scaling behavior of the energy bands for the almost periodic problem by studying periodic problem approximations to the problem via the renormalization group. References [11], [12] and [46] obtained scaling relations which show how the band energies for successive commensurate approximations to the almost periodic problem scale for fixed K as the order of commensurability is increased. Specifically, their relationship is

$$w^{-6\tilde{\gamma}} = \lim_{n\to\infty} \frac{E_{L+6(n-1)}(K) - E_L(K)}{E_{L+6n}(K) - E_L(K)}\,. \tag{113}$$

To show that this is equivalent to eq. (112), we consider eq. (107) with the right hand side set equal to a constant, to keep K fixed. By the previous arguments we obtain eq. (113). Ostlund and Pandit's renormalization group calculation could only be carried out for a single value of the phase ha in eq. (1), namely

$$ha = \lim_{L\to\infty} wn_L \tag{114}$$

where n_L denotes the location of successive maxima of the quantity $\Sigma_{ij}\,(M^{n_L})ij^2$. The latter quantity is vaguely related to the square of the wave function. They showed that this limit indeed existed. For other choices of phase, $\Sigma_{ij} (M^{n_L})_{ij}$ diverges and the calculation cannot be carried out.

Ostlund and Pandit show that there are two directions in which one can flow away from the fixed points of T^6 when one changes the form of the M matrices by a small amount, one corresponding to varying the amplitude of the potential and one corresponding to varying the energy, if we demand that the constraints on the matrices be satisfied (e.g. that the determinant be unity and others). In this sense they have proven universality for their results. It is possible, however, that large deviations from the Aubry cosine potential could modify their scaling results.

Ostlund and Pandit's method depends heavily on the properties of the inverse Golden Mean. In contrast, the renormalization group methods of Wilkinson and Suslov [35], discussed in section 3.3 are valid for period ratios other than the inverse Golden Mean. These methods, however, are most likely valid only for irrational period ratios which are very close to a low order rational number (i.e. ratio of small integers).

3.7. Miscellaneous results on band structure

Peres [48] has also studied a delta function Kronig–Penny model with sinusoidally varying potential strength incommensurate with the delta function spacing. He uses a transfer matrix method which is based on replacing the Schroedinger equation by a rotation in a Minkowski space whose coordinates are linear combinations of products of the wave function and its time derivatives. The transfer matrix represents a rotation in this space. He finds a fragmented band structure like that discussed in this section, with band splittings on all energy scales. He does not, however, attempt to distinguish localized from gap states.

Grempel et al. have shown that the almost periodic Schroedinger equation with a tangent potential that was discussed in section 2 has a density of states which is identified with that of Lloyd's model of a disordered lattice. Simon [49] has shown that their model is one member of a whole class of Schroedinger equations which have identical densities of states. These models, however, have widely differing spectral densities, and hence, the equality of their densities of states is not a good indication of the physics of the problem.

4. Unusual wave function behavior, the singular continuous spectrum and unusual electrical transport

Avron and Simon [13] and Simon [8] have shown that the Aubry model [i.e. eq. (1)] for $Qa/2\pi$ a Liouville number (i.e., irrational extremely well approximated by rationals) exhibits a singular continuous spectrum. That is, the spectral density is neither a continuous function nor a series of delta functions (i.e. a pure point spectrum). They suggested that the wave functions corresponding to a singular continuous spectrum would have unusual behavior, decaying practically to zero and then recovering to large values. Over very long distances the envelope of the wave function does gradually decay to zero, but it does not decay to zero exponentially. Avron and Simon [13] also suggested that such behavior would occur for the case $V_0 = t$ for $Qa/2\pi$ equal to an ordinary irrational number. Since Aubry and André [7] have shown that the localization length for $V_0 = t$ is infinite in both the configuration space model and its self-dual k-space model, there does not exist a length scale in the problem over which the wave function can decay. Therefore, such unusual wave function behavior is expected in this case.

4.1. Unusual wave functions in the tangent potential model

Prange et al. [28, 29] have found such unusual wave function behavior in their tangent potential almost periodic model for $Qa/2\pi$ a Liouville number. Their model allows one to obtain a physical picture of how such behavior comes about. To this end, consider $u^{+}(\theta, t) \sim \exp\{i\phi(\theta, t)\}$ in eq. (53) whose Fourier transform on θ gives the solution to eq. (43), the Schroedinger equation for the almost periodic problem considered by Prange et al. [28, 29]. We showed in section 2 that the Fourier series for ϕ could be written as

$$\phi_\nu(\theta, t) = \nu\theta + \sum_m F_m(t)\, e^{im\theta},$$

where

$$F_m = -\frac{A_m}{1 - e^{-imQa}},$$

[eq. (57b)], where A_m is defined in section 2. In that section, we discussed the behavior of the Fourier transform of $u^{+}(\theta, t)$ when mQa never becomes too close to a multiple of 2π. This is the case of usual irrational values of $Qa/2\pi$. In such a case the Fourier series for $\phi(\theta, t)$ converges, and we showed that the resulting solutions of eq. (43) decayed rapidly at large distances. When $Qa/2\pi$ is extremely well approximated by rationals, however, mQa can become arbitrarily close to a multiple of 2π, and for values of m where this occurs, some denominators in eq. (60b) can become very small. Such terms in the Fourier series for $\phi(\theta, t)$ must be treated separately. For example, consider the case where there is only one such small denominator when $m = m_1$. Then

$$u(\theta) = \exp\{i\phi'(\theta)\} \exp\{iF_{m_1} \sin(m_1\theta + h_1)\}, \tag{115}$$

where $\phi'(\theta)$ represents the sum of all terms in the Fourier series for $\phi(\theta)$ except the $m = \pm m_1$ terms. If we let $\exp\{i\phi'(\theta)\} = u'(\theta)$ and use a well known Bessel function identity to express the second exponential factor in eq. (115) as a Fourier series, we find that the Fourier transform of $u(\theta)$ has the form

$$u_m = \sum_L u'_{m-Lm_1} J_L(F_{m_1}) \exp\{-iLm_1h_1\}. \tag{116}$$

Since the quantity F_{m_1} in eq. (112) is assumed large, we may use the asymptotic behavior of the Bessel function for large argument to conclude that u_m represents a series of rescaled versions of the exponentially decaying wave function u'_m over larger and larger distances, separated by m. (The reasons that u'_m is exponentially decaying are given in section 2.) If there are j small denominators in the Fourier series for $\phi(\theta, t)$, the arguments leading to eq. (106) may be generalized to yield

$$u_m^{(j)} = \sum_{L_1, L_2, \ldots, L_j} u'_{m-L_1m_1-L_2m_2\ldots L_jm_j} J_{L_1}(F_{m_1}) J_{L_2}(F_{m_2}) \cdots J_{L_j}(F_{m_j}).$$

In this wave function, the localized wave function u_m^j is repeated on j length scales. Although the limit $j \to \infty$, which must be relevant for the Liouville number case, might not exist, the present discussion gives a physical picture of approximately how such a wave function could come about.

The spectral density due to this wave function can be understood as follows: If we wish to calculate the spectral density at a lattice site m, we must consider the fact that the wave functions are neither localized nor extended. Clearly, the energy eigenvalue of a given state will depend on the site around which it is quasi-localized. Therefore, if $\omega_{m'}$ is the energy of the state localized on site m', our spectral density will have an energy level at energy $\omega_{m'}$ for all m' in addition to that at ω_m. The weight of that energy level is proportional to $J_{m-m'}(F_{m-m'})$. Since the nature of this almost periodic system is that there must exist states extremely close in energy quasi-localized about sites which almost repeat in space (they would have precisely equal energy for a true periodic system), there is an almost continuum of discrete levels in the spectral density. The weights of some of these levels, however, can be very small, being weighted by Bessel functions of very large argument. This discussion gives an approximate physical picture of how a singular continuous spectrum could occur. It is hoped that some of these features will carry over to other models with singular continuous spectra.

Prange, Grempel and Fishman [49a] have recently calculated the low frequency AC conductivity for the tangent potential model. The calculations are straightforward, following directly from the calculated wave functions. For most incommensurate wave vectors, the AC conductivity is given by

$$\sigma \sim \exp\{-(\omega_0/\omega)^{\sigma}\}$$

where ω_0 is a constant and σ is a number > the dimensionality d. For the case where the wave vector is a Liouville number times $2\pi/a$, however,

$$\sigma \sim \omega^x (\ln \omega)^y$$

where $y = d+1$ and $-1 < x < 1$. This is because of the existence of extremely distant localized states of *nearly* equal energy.

4.2. Approximate scaling behavior of the wave functions

Thouless and Niu [50] have proposed that the absolute value of a wave function at the critical point $V_0 = t$ should be a product of periodic functions, the period of each function being one of the approximate periods of the almost periodic problem obtained by cutting off the continued fraction representation of $Qa/2\pi$ at a particular stage. This argument makes use of a relationship similar to eq. (64) to eliminate intermediate sites, resulting in a difference equation of form similar to that of eq. (1), valid for $t < V_0$. The number of sites eliminated is chosen to correspond to one of the approximate periods for the almost periodic problem. Successive on-site potential differences in the new Schroedinger equation will thus be made as small as possible by this choice of number of sites to be eliminated. Provided that successive on-site energy differences are not too small, the effective hopping matrix elements obtained in successive Schroedinger difference equations obtained by successive applications of the above procedure get smaller and smaller compared to the on-site potential strength. Since after an infinite number of applications of this procedure the hopping terms become zero, we conclude that the wave functions must be localized. When $Qa/2\pi$ is a Liouville number successive on-site potential differences can be so small that the hopping matrix elements do not become small under successive applications of the above renormalization group procedure. Thus, the states are no longer localized. For non-Liouville numbers, as V_0 approaches t, eq. (64) breaks down and successive hopping terms become comparable to the on-site potential. Thus, sites are coupled on all distance scales. If after a given number of applications of the above renormalization group procedure we approximate the on-site potential in the resulting Schroedinger equation by the periodic potential which most closely approximates it, the eigenfunctions must be Bloch functions. Hence, the magnitude of the wave function is periodic with the period of the potential in this Schroedinger equation. Of course, the lattice sites in this renormalized problem are separated by many atomic sites. To learn about the variation of the wave function over smaller distances, we stop the renormalization group procedure one stage earlier and approximate the potential by the periodic approximation for this stage. Continuing this procedure we find that the wave function must have a series of envelope factors periodic with period corresponding to each periodic approximation to the almost periodic problem. Because at some point several of these periodic envelope functions can come close to being in phase (but never can they actually be in phase except for the periodic case), a peak at the origin will reappear at distant points but with successively reduced amplitudes. This is similar to what the model of Prange (discussed in the last paragraph) gives for the case of $Qa/2\pi$ a Liouville number.

4.3. Transmission coefficients and conductivity for periodic and almost periodic crystals

The transmission coefficient has been calculated numerically as a function of system length using eqs. (19) and (23) for $V_0 = t$ for $E = 0$ for the Aubry model [51]. As V_0 approaches t from below, the widths of the subbands shrink to zero, and therefore, it is difficult to find the discrete energies of the eigenstates for $V_0 = t$. For $E = 0$, it was possible to determine, using symmetry and other arguments, that there was definitely a state at that energy. These calculations were done for $Qa/2\pi = 0.5(\sqrt{5} - 1)$ and $1/(2\sqrt{10})$. The transmission coefficient was found at times to drop as low as 10^{-7} only to recover to about 10^{-2} when the length of the system was further increased. Systems as long as 50 000 sites were studied and this behavior was found to persist. Such behavior is consistent with what was found by Prange et al. for the singular tangent potential model [29]. Such studies of the transmission coefficient are important because it was shown by Landauer in 1970 [18] that the electrical conductivity of a one dimensional system is proportional to $T/(1 - T)$. Thus, such studies give us information about an experimentally observable quantity.

Although it is easily shown that the trace of the transmission matrices is independent of the phase of the potential, the transmission coefficient given by eq. (23) will depend on the phase. As discussed in section 3 Ostlund and Pandit [45] showed that the sequence of maxima of the matrix elements diverged for all except for one special choice of phase. In the light of the fact that they speculated that the physics of the problem should not depend on the choice of phase, it is interesting that the transmission coefficient is a physical quantity which does depend on choice of phase. The reason for this is probably its dependence on the choice of location of the edges of the finite sample being studied.

In order to attempt to gain some insight into the behavior of the transmission coefficient of an almost periodic crystal, let us consider the transmission coefficient for a periodic system and how it behaves in the incommensurate limit. To proceed further we recognize that for a periodic system for energy inside a band, the eigenvalues of the transmission matrix defined in eq. (21), which relates the column matrix (f_1, f_0) to (f_n, f_{n-1}), have unit magnitude (since the determinant of the matrix is 1) and are complex conjugates of each other. If the unit cell length of a particular periodic system is qa (where q is an integer), the eigenvalues of the matrix

$$M^{(q)} = \begin{pmatrix} Q_q & P_q \\ Q_{q-1} & P_{q-1} \end{pmatrix} \tag{117}$$

can be written as

$$\exp\{\pm \mathrm{i} Kqa\} = x \pm \mathrm{i}(1 - x^2)^{1/2} \tag{118}$$

where x is half the trace of $M^{(q)}$. Here K is actually the wave vector of the state corresponding to the energy that we are considering. The matrix which diagonalizes $M^{(q)}$ (i.e. the matrix S such that $M_\mathrm{D} = S^{-1}M^{(q)}S$ is diagonal), is given by

$$S = \begin{pmatrix} -P_q & -P_q \\ Q_q - \mathrm{e}^{\mathrm{i}Kqa} & Q_q - \mathrm{e}^{-\mathrm{i}Kqa} \end{pmatrix}. \tag{119}$$

At the band edges $K = 0$ or π/qa and S becomes singular. From eq. (23), we can easily show using $M^{(q)} = SM_\mathrm{D}S^{-1}$ and eqs. (118) and (119) that for small K (i.e. near the band edges) but KNa = an odd multiple $\pi/2$ the transmission coefficient is given by

$$T_0 = \frac{4C_2^2 \sin^2 ka}{[D - F\cos ka]^2 + F^2 \sin^2 ka}(Kqa)^2, \tag{120}$$

where $E = t\cos ka$ defines k and $D = C_1^2 + C_2^2$, $F = C_1(1 + C_2)$, $C_1 = Q_q - 1$, and $C_2 = -P_q$. Thus T_0 can be very small if (Kqa) is. It is easily shown using eqs. (118), (119) and (23) that whenever KNa is equal to a multiple of π, $T_0 = 1$. Thus, near the band edges, T_0 can become very small and then grow to much larger values. The reason for this behavior is most likely that the band edge is a point close to where T_0 switches from decaying exponentially (in the gap) to remaining large (in the band). It is tempting to speculate for the almost periodic case that when V_0 becomes equal to t, the two band edges move together and only transmission coefficient behavior typical of the band edge occurs. Unfortunately, there is no evidence either for or against this idea.

In fact, eq. (118) points out that one must be careful in generalizing from periodic to almost periodic systems. For any commensurate system, no matter how high order commensurate, Kqa will vary from zero to π as the energy moves across the band (no matter how narrow the band). Thus, if we take the thermodynamic limit before we take the incommensurate limit, so that the system size is always larger than the unit cell size, Kqa will always take all values from 0 to π as the energy moves across the band even though the bands become infinitesimally wide. This implies that the transmission coefficient calculated in this way is not defined since it takes on all values in an infinitesimal energy interval.

Studies of the commensurate system do, however, give us information about another puzzling feature of almost periodic systems. An almost periodic system should have a gap in the neighborhood of every energy. Nevertheless, we saw in section 3 that for V_0 less than t, the system behaves just like a periodic system as far as the energy spectrum is concerned. In fact, numerical calculations of T_0 for $V_0 < t$ like those discussed earlier in this section for $V_0 = t$, show that there exist bands of energy for which T_0 never gets small, no matter how long we make the system. To attempt to understand this apparent paradox, let us consider the behavior of T_0 inside one of the small gaps. In a gap, the wave vector defined by eq. (118) becomes imaginary. Its magnitude gives the rate at which the wave functions, and hence T_0 decays. The quantity x becomes greater than 1 in magnitude inside a gap. Then if x varies smoothly as a function of energy, we expect that if the gaps are much narrower than the bands separating them, $|x|$ will never get a chance to become too much greater than 1. Thus, we expect that inside the extremely small gaps the decay rate of T_0 will be so slow that it most likely would not be found in numerical studies and in fact would most likely never be observed in an experimental system. Therefore, although there is a gap in the neighborhood of every energy in an almost periodic system, and thus, any random choice of energy has a good chance of falling in a gap, the rate of decay of the transmission coefficient in a very small gap as function of system length might be so slow as to make the gap physically irrelevant.

4.4. *The renormalization group model*

Bellissard, Bessis and Moussa [52] have put forward a model Hamiltonian with a spectrum which is a Cantor set of zero measure, which can be solved exactly. The Hamiltonian is defined to be a fixed point of the following renormalization group transformation:

$$D\,F(H)\,D^{-1} \tag{121}$$

where H is the Hamiltonian. This means $D\,F(H)\,D^{-1} = H$. Here

$$F(x) = x^2 - \lambda \tag{122}$$

where λ is a real number which must be greater than or equal to 2 for H to be Hermetian, and $Df_n = f_{2n}$ where f_n is the tight binding wave function. The solution is

$$Hf_n = t_{n+1}f_{n+1} + t_n f_{n-1} \tag{123}$$

where

$$t_0 = 0 \tag{124a}$$

$$t_{2n}^2 + t_{2n+1}^2 = \lambda \tag{124b}$$

$$t_{2n}^2 t_{2n-1}^2 = t_n^2 \, . \tag{124c}$$

For $\lambda = 2$, $t_0 = 0$, $t_1 = \sqrt{2}$ and $t_n = 1$ for $n \geq 2$, which obviously has a single band extending from -2 to $+2$ because it is the ordinary periodic tight binding model for $n \geq 2$. Because the Hamiltonian is a fixed point of the transformation (i.e. the Hamiltonian is invariant under it), the spectrum is invariant under it. Since the spectral density is simply the imaginary part of the Green function, which is the inverse of E minus the Hamiltonian, it too is invariant. If we apply $F^{-1}(x) = \pm\sqrt{\lambda + x}$ an infinite number of times to any energy in the spectrum, we get a Cantor set of zero total bandwidth. By studying the behavior of t_n under a large number of doublings of n, they show that t_n can come arbitrarily close to its initial value. Thus, it is almost periodic. Iterating eq. (124) and using the transmission matrix method discussed in section 2 [i.e. iterating eqs. (19) and (21)] gives the wave functions. By applying the Thouless exponent formula to the spectrum they show that the decay exponent is zero implying that there are no exponentially localized states. A better understanding of how the spectrum of zero measure (i.e. zero total bandwidth) comes about may be obtained by studying t_n. We find that t_n is usually of order unity but occasionally it drops to very small values at certain points. Thus, there will be regions in the crystal in which the lattice sites are strongly coupled, but these regions are only weakly coupled to each other. Thus, from each such cluster, we get a set of discrete levels. These are split further by the weak coupling between the clusters. The couplings are never strong enough, however, to broaden the levels into bands. If we include at each level weaker and weaker intercluster couplings, the levels are split again and again but never made into bands.

4.5. Numerical and renormalization group wave function calculations

Ostlund and Pandit [45] have recently studied the behavior of the wave functions of the Aubry model using the transmission matrix method expressed by eq. (21). For $V_0 < t$, they find that the envelope of the wave function is smooth and the wave function is extended. For $V_0 > t$, the envelope of the wave function is found to decay exponentially. For $V_0 = t$, for $E = 0$ (i.e., the band center) the envelope of the wave function decays to very small values as a function of distance only to grow again to a value which is a fraction ξ of its original value. The process then repeats itself, the successive large values of the wave function decaying to zero at best as a power law. There is no exponential decay. The structure of the wave function in each region in which it becomes large again is approximately a rescaled version of what it was in the previous such region. In order to obtain their wave function they choose a value of the phase in eq. (1), ha, given by

$$ha = \lim_{L\to\infty} wn_L \bmod 1 = \tfrac{1}{4} - w\,,$$

where $\{n_L\}$ are the locations of maxima of the sum of the squares of the transmission matrix elements, for zero energy states (the center of the band). For this choice of phase the wave function has an absolute maximum at some point and then its envelope decays to either side of this maximum algebraically. [This behavior only occurs for one choice of the matrix column matrix (f_1, f_0) in eq. (23).] Although the wave functions do not possess translational symmetry for the almost periodic problem at $V_0 = t$, Ostlund and Pandit have found for $E = 0$ that the wave functions f_n have the following symmetry: f_n takes on values which when n is translated by F_1 and then by $F_2, F_3, \ldots F_\infty$, where F_L is the Lth member of the Fibonacci sequence, the f's begin to almost repeat. Thus, there seems to be a type of translation symmetry on a "skewed lattice". Ostlund and Pandit have also studied the $V_0 > t$ regime and find results similar to the Aubry–André results for the localization length as a function of potential strength. For $V_0 < t$, they find that the gaps which are seen when we look at smaller and smaller energy scales approach zero in size. This agrees with the results discussed in subsections 3.1 and 3.2, found by perturbation theoretic and quasi-classical methods.

5. Higher dimensional almost periodic lattices

5.1. One dimensional modulations in higher dimensional lattices

Some of the results discussed in previous sections for one dimensional lattices may be generalized to higher dimensions. A trivial higher dimensional case which was discussed by the present author a few years ago is the case of a one dimensional modulation potential in a higher dimensional periodic lattice [14]. An example of a one dimensional modulation potential is a potential of the form $V_0 \cos \boldsymbol{Q} \cdot \boldsymbol{R}$ where $\boldsymbol{R}$ is the location of a lattice site, $\boldsymbol{Q}$ is the wave vector of the potential and V_0 its amplitude. The case in which $\boldsymbol{Q}$ is along a crystallographic axis for a simple cubic lattice was discussed in ref. [14]. Let us now generalize this result.

Consider the following tight binding Schroedinger equation:

$$t \sum_{\boldsymbol{a}} f(\boldsymbol{R}+\boldsymbol{a}) + V_0 \cos \boldsymbol{Q} \cdot \boldsymbol{R} f(\boldsymbol{R}) = E f(\boldsymbol{R}), \tag{125}$$

where $f(\boldsymbol{R})$ is the tight binding wave function on site $\boldsymbol{R}$, and $\boldsymbol{a}$ runs over nearest neighbor sites. Consider the Fourier transform of eq. (125):

$$\varepsilon(\boldsymbol{k})\, g(\boldsymbol{k}) + \tfrac{1}{2} V_0 [g(\boldsymbol{k}+\boldsymbol{Q}) + g(\boldsymbol{k}-\boldsymbol{Q})] = E\, g(\boldsymbol{k}), \tag{126}$$

where

$$\varepsilon(\boldsymbol{k}) = t \sum_{\boldsymbol{a}} \exp\{\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{a}\}, \tag{127a}$$

and

$$f(\boldsymbol{R}) = \frac{1}{N} \sum_{\boldsymbol{k}} g(\boldsymbol{k}) \exp\{\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{R}\}, \tag{127b}$$

where N is the number of lattice sites. Consider first the case in which $\boldsymbol{Q}$ is along one of the primitive reciprocal lattice vectors G_1. That is, we consider

$$\boldsymbol{Q} = \alpha \boldsymbol{G}_1 .$$

Then, eq. (126) may be rewritten as

$$2t \cos 2\pi n\alpha\, g_n + \tfrac{1}{2} V_0 (g_{n+1} + g_{n-1}) = [E - \varepsilon'(k)]\, g_n \tag{128a}$$

where

$$g_n = g(\boldsymbol{k} + n\boldsymbol{Q}), \tag{128b}$$

$$\varepsilon'(k) = t \sum_{\boldsymbol{a} \neq \pm \boldsymbol{a}_1} \exp\{\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{a}\} \tag{128c}$$

where $\boldsymbol{G}_1 \cdot \boldsymbol{a}_1 = 2\pi$ and $\boldsymbol{G}_1 \cdot \boldsymbol{a} = 0$ for all $\boldsymbol{a} \neq \boldsymbol{a}_1$. This clearly has the same form as eq. (1). Hence, the solutions will be localized in $\boldsymbol{k}$ space along the direction of $\boldsymbol{Q}$ for $4t > V_0$ and extended if $4t < V_0$. To understand what this tells us about the behavior of $f(\boldsymbol{R})$, consider the case in which the cosine potential localizes $f(\boldsymbol{R})$ along the $\boldsymbol{a}_1$ direction. Then,

$$g_n = \sum_R f(\boldsymbol{R}) \exp\{i(\boldsymbol{k} + n\alpha \boldsymbol{G}_1) \cdot \boldsymbol{R}\}. \tag{129}$$

Since $\boldsymbol{R} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$ where $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ are the primitive lattice vectors and n_1, n_2 and n_3 are integers, and since we may write $\boldsymbol{k} = k_1\boldsymbol{G}_1 + k_2\boldsymbol{G}_2 + k_3\boldsymbol{G}_3$, we may rewrite eq. (129) as

$$g_n = \sum_{n_1} h_{n_1}(k_2, k_3) \exp\{i2\pi(k_1 + n\alpha)n_1\} \tag{130a}$$

where

$$h_{n_1}(k_2, k_3) = \sum_{n_2 n_3} f(n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3) \exp\{i2\pi(k_2n_2 + k_3n_3)\}. \tag{130b}$$

Since $f(\boldsymbol{R})$ is localized in the $\boldsymbol{a}_1$ direction, the sum over n_1 in eq. (130a) converges, and hence, g_n is extended in n. If g_n is localized in n, then from eq. (127b), we conclude that $f(\boldsymbol{R})$ is extended in the $\boldsymbol{a}_1$ direction. Thus, we again have a form of "Aubry duality", and hence, by studying the localized or extended nature of solutions of an equation of exactly the same form as eq. (1) [namely eq. (126)], we determine the nature of the solutions of eq. (125).

Let us now consider a more general one dimensional modulation in which $\boldsymbol{Q}$ is not along a primitive reciprocal lattice vector. Then, we may write

$$\boldsymbol{Q} = \alpha_1\boldsymbol{G}_1 + \alpha_2\boldsymbol{G}_2 + \alpha_3\boldsymbol{G}_3. \tag{131}$$

In this case, the equation satisfied by g_n defined in eq. (128b) is more complicated than eq. (128a). It has the form

$$\varepsilon(\boldsymbol{k} + n\boldsymbol{Q})\, g_n + \tfrac{1}{2}V_0[g_{n+1} + g_{n-1}] = Eg_n. \tag{132}$$

Since $\varepsilon(\boldsymbol{k} + n\boldsymbol{Q})$ is also an almost periodic function of n, the methods discussed in section 2 should be applicable to eq. (132) to determine whether or not the solutions are localized. From eqs. (127) and (125b) we have

$$g_n = \sum_R f(\boldsymbol{R}) \exp\{i\boldsymbol{k} \cdot \boldsymbol{R}\} \exp\{in\boldsymbol{Q} \cdot \boldsymbol{R}\}. \tag{133}$$

If $|f(\boldsymbol{R})| \to 0$ as $|R_\parallel| = |\boldsymbol{R} \cdot \boldsymbol{Q}|/Q \to \infty$, it is clear that g_n given by eq. (130) will not approach zero as $|n| \to \infty$. Thus, g_n is an extended solution of eq. (132). If g_n is a localized solution, then, from eq. (127b), $f(\boldsymbol{R})$ will be extended in the $\boldsymbol{Q}$ direction. The nature of the solutions of (132) can be studied using the numerical methods discussed in section 2.

Bellissard, Lima and Scoppola [53] have studied a multi dimensional model similar to eq. (125). They consider the Schroedinger equation

$$t \sum_a f(\boldsymbol{R} + \boldsymbol{a}) + V(\boldsymbol{Q} \cdot \boldsymbol{R}) f(\boldsymbol{R}) = E f(\boldsymbol{R}) \tag{134}$$

where $\boldsymbol{R}$ is restricted to lie on a simple cubic lattice ($\boldsymbol{a}$ is a near neighbor lattice vector for a simple cubic lattice), but V is a member of a class of functions of which the cosine potential is a special case. Their model is more general than eq. (125) in the sense that V is more general but less general in that they only consider a simple cubic lattice. They show that for a class of potentials V and $\boldsymbol{Q}$ satisfying a diophantine condition which essentially requires that $\boldsymbol{Q}\cdot\boldsymbol{a}/2\pi$ not be extremely well approximated by rationals (i.e. not a Liouville number) and t sufficiently small, that the spectrum is pure point with only exponentially localized states (i.e. the usual case of Anderson localization).

5.2. *A two dimensional modulation in a two dimensional lattice*

So far we have only spoken of the case of a one dimensional potential in a higher dimensional lattice. In ref. [14] a two dimensional potential in a two dimensional lattice was considered. Equation (134) was studied in two dimensions (i.e. for a square lattice) with the potential replaced by the potential

$$V_0(\cos Q_x X + \cos Q_y Y)\,, \tag{135}$$

which is a two dimensional potential, with $Q_x a/2\pi$ and $Q_y a/2\pi$, where a is the lattice constant of the lattice, irrational numbers. The approximate Anderson localization function method [15, 16] was applied to this model. Although this method is not rigorous, it should be correct for many purposes. It was found that for $V_0 > 8t$, all states are localized, for $V_0 < 2t$, all states are extended and for $2t < V_0 < 8t$, there exist both localized and extended states separated by a mobility edge. The method could easily be generalized to higher dimensions. The disadvantage of this method is that the Anderson localization function method involves an uncontrolled approximation, namely cutting off the continued fractions in the expansion of the self-energy at the first term.

The particular model represented by eq. (135) in fact points out deficiencies of the localization function method. For example, when the crystallographic axes line up with the symmetry axes of the potential (i.e. the X and Y axes), the model is separable, and we get the model of eq. (1) for motion along both the X and Y axes. Then, clearly, all states will be localized for $V_0 > 2t$ and extended for $V_0 < 2t$. Perhaps the approximate localization function method gives an average over all relative orientations of the axes of the potential and the lattice. It should also be noted that the model possesses a type of self-duality for all orientations. This is easily seen by Fourier transforming the Schroedinger equation. The Fourier transformed model will have the axes of the potential and lattice interchanged from what they were in the configuration space Schroedinger equation. The transformed model will, therefore, be identical to the original model for $V_0 = 2t$ only for certain orientations of the axes.

6. Possible experimental application

6.1. *Difficulties in observing experimental consequences of the exotic band structure of almost periodic crystals*

The presence of minibands and minigaps on all energy scales in almost periodic crystals affords an opportunity to measure a number of interesting effects of an applied electric field. For example, a strong electric field can induce negative differential conductivity as it pushes electrons toward the region of the energy versus wave vector dispersion curve that begins to bend as it approaches the region of a gap.

Furthermore, for higher electric fields one should begin to see the oscillations of the electric current due to the Stark ladder and interband or Zener tunneling. These effects have been considered for a long time in connection with the study of ordinary periodic crystals. They are very difficult to observe experimentally, however, as they require extremely strong electric fields and very long relaxation times. Almost periodic crystals could potentially bring all of these effects within the possibility of experimental test, because the widths of the minibands and the size of the minigaps are considerably smaller than those of ordinary periodic crystals. Although theoretically an almost periodic crystal has gaps on all energy scales, we saw in the previous sections of this article that in order to have the very small gaps present, the crystal must be coherent over distances much larger than the size of most experimental samples. Thus, the hierarchy of gap structure discussed earlier can only be approximated in real crystals. Furthermore, except for the case of a modulation potential comparable in strength to the lattice potential, we saw that most of the gaps are negligibly small, and the decay constant for an electron inside one of these extremely small gaps is probably longer than the crystal. Thus, it is only for a modulation potential of sufficient strength to put us at the phase boundary for the metal–insulator transition that we can hope to see experimental effects of incommensurability.

There exist many crystals in nature which contain a modulation incommensurate with the underlying lattice structure. Examples of such crystals are crystals containing charge or spin density waves [13], mercury chain compounds [3], and some ferroelectrics such as thiourea [4]. In most of these systems, however, the potential produced by the modulation, which acts on either the electronic or phonon states, is weak compared to the bandwidths. We saw in section 3 that the band structure for a weak incommensurate potential is almost identical to that of the closest commensurate approximation to that potential. It is only extremely close to the band edges of the commensurate problem used to approximate the almost periodic problem that the fragmented band structure characteristic of the almost periodic potential problem is found if the closest commensurate approximation is higher than second order registry (i.e. the ratio of the periods is greater than 2) [54]. For first and second order registry cases it has been shown that structure will occur deep inside the commensurate band gaps [54, 55]. The first order registry case corresponds to thiourea and $(NH_4)_2BeF_4$ [4] near the commensurate–incommensurate transition. The second order registry case corresponds to polyacetylene [55]. These two cases would be good candidates for studying the unusual band structure due to incommensurability. Optical methods in which one looks for transitions between sub-bands would be good ways of making such studies.

The sinusoidal potential of the Aubry model [eq. (1)] is not expected to be a good description of the potential seen by the electrons or phonons due to the modulation when the crystal is almost commensurate (e.g. in the vicinity of the commensurate–incommensurate transition). Rather, an almost commensurate crystal should have large commensurate regions in the crystal separated by relatively short domain walls called discommensurations. Since the potential due to such a configuration should have many Fourier components, we might expect extensive gap structure even for fairly weak potentials. It was shown in ref. [54], however, that for parameters appropriate to most almost periodic crystals, it is still difficult to observe effects of incommensurability on the band structure. The reason for this is that the largest wave vector contained in Fourier transform of the domain wall lattice potential is comparable to 2π divided by the domain width. Since the modulation is weak in most crystals, the domain wall size, which is proportional to the square root of the ratio of the force constant of the lattice to the force constant contribution due to the modulation, will generally be quite large. As mentioned earlier, the exception is the first order and second order registries.

One promising possibility for studying effects of incommensurability is to study them for almost periodic superlattices in semiconductors and metals grown by molecular beam epitaxy. One difficulty with such a program is that the degree of perfection of such superlattices is still not at a level sufficient for studying effects of incommensurability on the band structure. Furthermore, these crystals are not coherent over sufficiently long distances to distinguish the band structure of an almost periodic from a periodic crystal. In fact, it should be remembered that much of the fine structure in the bands of an almost periodic crystal results from the beating of the two incommensurate periods, and this only happens over long distances. In fact in order to study the hierarchical nature of the band structure discussed in section 3 in detail, it would probably be necessary to have crystals which are coherent over distances comparable or larger than the size of most experimental samples. Nevertheless, it would still be possible to study gross features of the band structure if we had fairly good crystals.

An almost periodic layered structure, grown by molecular beam epitaxy, in which two different substances alternate in a square wave manner, could approximate an almost crystal at the metal–insulator transition point, even for relatively weak modulation potentials, since in subsection 3.6 it was found that square wave potentials always lead to critical behavior. Of course, since the steps in the potential will be rounded in real crystals, this can only be approximately correct. The system will certainly be critical for a sufficiently strong modulation potential due to having layered structure with two different substances, however.

Studies of localization effects, however, do not require such a degree of crystal perfection. In fact, such localization effects may already have been observed in a disordered substance based on $NbSe_3$ namely $Fe_{1+x}Nb_{3-x}Se_{10}$ [56, 57]. This compound possesses a charge density wave which appears at sufficiently low temperatures. When the charge density wave appears, a metal–insulator transition occurs, with the conductivity falling over 9 orders of magnitude. The temperature dependence of the conductivity in the insulating state is described by Mott's variable range hopping theory, which signifies that some sort of Anderson localization has occurred. Since the charge density wave essentially represents a modulation whose wave vector is along the c-axis of the crystal, we would expect it to cause localization of the electronic states primarily along the c-axis. This would make the system into a two dimensional system and the impurities would localize the states in the other directions since impurities cause localization in two dimensions. It is expected, however, that since the impurity potential is comparable to or greater than the potential due to the charge density wave, this simple picture will not be an accurate description of this compound. Approximate Anderson localization function calculations for a system containing both a modulation potential and impurities, however, do show that the charge density wave potential will certainly aid the localization due to the impurities.

6.2. The optical spectrum of hollandite

As stated earlier, one of the main difficulties in observing effects of incommensurability is that most almost periodic crystals occurring in nature consist of a crystal potential and a relatively weak modulation potential incommensurate with the crystal potential. As a result, the effects of the fact that the modulation potential is incommensurate with the lattice are only a small perturbation on the band structure for closest commensurate approximation to the modulation. Hollandite, however, is a crystal with a relatively strong modulation potential. Hollandite is a one dimensional ionic conductor which has channels containing mobile potassium ions which order in a structure incommensurate with the lattice potential inside the channels [58]. In the case of hollandite, the channel potential is strong enough compared to the potassium–potassium interaction to cause large displacements of the potassium atoms from equal spacing, but it is not large enough to force all potassium ions to reside at its minima.

Therefore, the phonon equations of motion will contain a term in the dynamical matrix due to the channel potential which is comparable in magnitude to the part due to potassium–potassium interactions. This term is periodic with a period incommensurate with the potassium lattice. Calculations of the absorption coefficient and Raman scattering cross section for a model for the potassium ions in hollandite show many strong peaks due to the breaking of translational symmetry by the incommensurate channel potential. That is, the channel potential makes phonons with wave vector equal to a multiple of the channel potential wave vector observable by absorption and scattering of light. The experimental situation is unclear at this point, however, because it is difficult to grow crystals with a range of potassium concentrations [59].

The possibility also exists that the phonon peaks will be broadened by the almost periodic potential, as was discussed in subsection 2.6. If crystals in the hollandite family could be grown with variable mobile ion concentrations, one would expect a phase transition in the phonon linewidth at a critical value of the ratio of the channel potential to the interionic potential strength, as one crosses the metal–insulator transition critical point.

6.3. The electron in a magnetic field problem

An unexpected application of the theory of electrons in almost periodic potentials occurs in the electron in a magnetic field problem [34, 36]. This application was originally pointed out by Harper [60]. The energy spectrum of a two dimensional free electron gas in a magnetic field is well known to consist of a series of discrete Landau levels. In the presence of a periodic potential the spectrum becomes more interesting.

Let us consider first the case of a weak periodic potential. We will treat the periodic potential as a perturbation on a free electron gas in a magnetic field. That is we will assume that we are in the strong magnetic field limit in which the spacing between Landau levels is large compared to the matrix elements of the periodic potential. Then our unperturbed Hamiltonian for a uniform magnetic field is

$$H = \frac{1}{2m}(\pi_x^2 + \pi_y^2) \tag{136}$$

where

$$\pi_x = p_x + \frac{e}{2c} By \tag{137a}$$

$$\pi_y = p_y - \frac{e}{2c} Bx \tag{137b}$$

and p_x, p_y are components of the momentum operator and B is the magnetic field. If we choose our Landau level wave functions to be of the form

$$\exp\{iky\} \exp\left\{-i \frac{e}{2c} Bxy\right\} \phi\left(x - \frac{c\hbar k}{eB}\right), \tag{138}$$

ϕ will satisfy a Schroedinger equation for a harmonic oscillator centered about the coordinate

$x = c\hbar k/eB$, with energy given by

$$(n+\tfrac{1}{2})\hbar\omega_c \tag{139}$$

where $\omega_c = eB/mc$, independent of k. Thus, the Landau level wave functions are independent of k and hence highly degenerate. Applying generalized Born–Von Karmon boundary conditions in the y direction by requiring $e^{ikL} = e^{i\eta}$ where η is some number and L is the length of the box and requiring the orbit center to be inside the sample, we obtain the following degeneracy factor:

$$BL^2/\phi_0 = m + \eta/2\pi \tag{140}$$

where $\phi_0 = hc/e$ is the flux quantum and m is an integer. Thus, the degeneracy of each Landau level is equal to the number of flux quanta through the sample.

The presence of a weak periodic potential will break the degeneracy of the Landau levels in general by mixing different values of k. Because of its simplicity, let us first consider the effect of a weak rectangular lattice potential of the form

$$V(x,y) = V_0(\cos G_1x + \cos G_2y)\,, \tag{141}$$

where $G_1 = 2\pi/a$, $G_2 = 2\pi/b$ where a and b are the lattice constants. Expanding our wave function in the unperturbed wave functions of the lowest Landau level [61] as follows:

$$\psi(x,y) = \sum_n f_n \exp\{i(k+nG_2)(y-y_0)\}\exp\left\{-i\frac{eB}{2c}xy\right\}\phi\left[x-x_0-\frac{c\hbar}{eB}(k+nG_2)\right] \tag{142}$$

where (x_0, y_0) is the origin of the coordinate system, we find that in lowest order degenerate perturbation theory f_n satisfies Harper's equation [60]

$$\tfrac{1}{2}V_1(f_{n+1}\exp(iG_2y_0) + f_{n-1}\exp(-iG_2y_0)) + V_2\cos(Qn+\delta)f_n = Ef_n\,, \tag{143}$$

where

$$Q = 2\pi\phi_0/Bab \tag{144a}$$

$$\delta = k\phi_0/Ba + 2\pi x_0/a \tag{144b}$$

$$V_1 = V_0\exp\{-\tfrac{1}{4}(\phi_0/BbL)^2\} = V_0\exp\{-\tfrac{1}{4}(2\pi)\phi_0/Bb^2\} \tag{144c}$$

$$V_2 = V_0\exp\{-\tfrac{1}{4}G_1^2L^2\} = V_0\exp\{-\tfrac{1}{4}(2\pi)\phi_0/Ba^2\} \tag{144d}$$

where L is the radius of a Landau orbit radius ($L = (\phi_0/2\pi B)^{1/2}$).

Equation (143) is equivalent to eq. (1). Then, from eq. (144) we see that if $b < a$, $V_2 > V_1$ and all solutions to eq. (143) will be localized in n if $Q/2\pi$ (the flux through a unit cell of the rectangular lattice) is an irrational number. According to eq. (142), this means that $\psi(x,y)$ is localized in the x-direction and extended in the y-direction. When $a < b$, it is easily shown using the dual model to eq. (143) that the states are extended in the x-direction and localized in the y-direction. For $a = b$, we get the unusual wave functions described in section 4 as solutions to eq. (143), and the spectrum is believed to be singular continuous. By studying an experimental isotropic square lattice system one would be able to study such unusual behavior, and by applying uniaxial stress one could study the spectrum of an almost periodic system extremely near (but not at) the metal–insulator transition. The one problem with this method is that, as we see from eqs. (142) and (144a), the quantity Q gives the ratio of the spacing between the centers of the harmonic oscillator functions which form the basis of the wave function $\psi(x, y)$ and the lattice constant in the x direction a. For a field of 10 000 gauss and lattice constants a and b of the order of 3 Å, $Q \approx 10^{-5}$ Å^{-1}. Thus, to see these effects the crystal would have to be coherent and the magnetic field would have to be uniform over a distance of 10 microns since the Landau orbit centers will be $\approx 10^5$ Å apart. Therefore, one cannot even hope to see these effects for ordinary crystal lattices. For the cases in which the periodic lattice is determined by a charge density wave with a unit cell which is twenty or thirty lattice constants long or longer however, the required crystal size becomes smaller. Also, it might be possible to see such unusual magnetic field effects with artificially grown superlattices which could have such large unit cells. For an anisotropic triangular lattice the equation analogous to eq. (143), obtained by the same methods as used to find eq. (143) has the form

$$2V_1 \cos \tfrac{1}{2}[(n-\tfrac{1}{2})Q+\delta]f_{n-1} + 2V_1 \cos \tfrac{1}{2}[(n+\tfrac{1}{2})Q+\delta]f_{n+1} + 2V_2 \cos(nQ+\delta)f_n = Ef_n . \tag{145}$$

Thouless's work on this model suggests that the total bandwidth of the spectrum of eq. (145) becomes zero when the crystal is isotropic (i.e. $V_1 = V_2$) [39], which implies that there will be unusual wave function behavior and a singular continuous spectrum in this case. This is an important result because there exist many real two dimensional crystals with a triangular symmetry superlattice structure [62]. Such systems would be good candidates for studying the band structure of an almost periodic crystal.

So far we have considered the weak periodic potential case. In the tight binding limit, we also obtain an equation with the same form as eq. (143) by the following argument: If $\varepsilon(\boldsymbol{k})$ is the band energy for a two dimensional system in the tight binding approximation, by the usual methods, the Schroedinger equation when a magnetic field is applied is obtained by replacing $\varepsilon(\boldsymbol{k})$ by $\varepsilon(\boldsymbol{k} - (e/ch)\boldsymbol{A})$. In the Landau gauge $\boldsymbol{A} = (0, Bx)$. Then, the effective Schroedinger equation becomes for a rectangular lattice for which $\varepsilon(\boldsymbol{k}) = \varepsilon_1 \cos k_x a + \varepsilon_2 \cos k_y$,

$$\left[\varepsilon_1 \cos k_x a + \varepsilon_2 \cos\left(k_y - \frac{e}{c\hbar} Bx\right) b\right]\psi = E\psi . \tag{146}$$

Replacing k_x by $(1/\mathrm{i})\partial/\partial x$, eq. (146) becomes

$$\frac{\varepsilon_1}{2}[\psi(x+a) + \psi(x-a)] + \varepsilon_2 \cos\left(\frac{eb}{ch} Bx - k_y\right)\psi(x) = E\psi(x) ,$$

using the fact that $\cos((a/\mathrm{i})\partial/\partial x)$ is a translation operator, or

$$\tfrac{1}{2}\varepsilon_1(f_{n+1} + f_{n-1}) + \varepsilon_2 \cos(nQ - \delta)f_n = Ef_n , \tag{147}$$

where

$$f_n = \psi(x + na) , \tag{148a}$$

$$Q = 2\pi Bab/\phi_0\,, \tag{148b}$$

and δ is the same as in (144b). The new $Q/2\pi$ is the inverse of the $Q/2\pi$ defined in eq. (144a). An equation of the same form as (145) can also be easily derived in this way for the triangular lattice. All of the qualitative conclusions discussed for the weak coupling case will also hold for the tight binding case.

The relationship between the electron in a magnetic field problem and the almost periodic potential problem also allows us to draw conclusions about the quantized Hall effect in the presence of a periodic lattice [63]. The classical high field Hall conductivity is given by

$$\sigma_{\mathrm{H}} = nec/B \tag{149}$$

where n is the number density per unit area. If the Fermi level lies in the gap between the m and $m+1$ Landau levels (i.e., we have m Landau levels completely filled), the density is given by

$$n = mB/\phi_0 = meB/ch\,, \tag{150}$$

using eq. (140). Substituting into eq. (149) we get the quantized Hall effect with quantum number m. The reason that the Hall steps have non-zero width is either that the two dimensional system is always in equilibrium with a bulk solid which fixes the position of the Fermi level [64], or there are localized states in the gap which do not contribute to the transport but which can pin the position of the Fermi level [65]. Since we have just seen that in the presence of a periodic potential the problem reduces to that of an electron in an almost periodic potential, the lowest Landau level will have a hierarchy of gaps introduced into it. If a fraction ν of the states in the Landau level lie below a particular gap, we might naively expect from eq. (149)

$$\sigma_{\mathrm{H}} = \nu e^2/h\,. \tag{151}$$

In fact Streda [66] and Thouless et al. [61] showed that this conclusion is not correct. Streda uses a formula for the Hall conductivity which both he [67] and Widom [68, 69] derived by different methods. We will give Widom's derivation which is simpler; it is based on Faraday's law [68]. Consider a planar region with a time varying flux passing through it. Then, the voltage or EMF around the border of the region is given by

$$V = \frac{1}{c}\frac{\mathrm{d}\phi}{\mathrm{d}t}\,. \tag{152}$$

There will be a Hall current I_{H} out of the border of the region given by

$$I_{\mathrm{H}} = eA\,\mathrm{d}n/\mathrm{d}t = \sigma_{\mathrm{H}} V \tag{153}$$

where n is the number density of electrons, A is the area of the region and e is the electronic charge. Combining eqs. (152) and (153) we find the relationship

$$\sigma_{\mathrm{H}} = ec\,\partial n/\partial B\,. \tag{154}$$

This is the Widom–Streda formula. Since the Fermi level lies in a gap, we may use eqs. (90), (144) (148b) and (140) to find n as a function of B. In the tight binding limit, eq. (90) divided by the unit cell area gives n. Then, from eq. (154) we find that σ_{H} must be an integral multiple of e^2/h. This is in conflict with the experimental observation of a fractional Hall conductivity. For the weak periodic potential case n is given by eq. (90) divided by the unit cell area and multiplied by the degeneracy factor for the Landau level given by eq. (140). On applying eq. (154), we obtain the same qualitative result as for the tight binding case. This result is consistent with Laughlin's arguments [70], and holds whenever the unit cell area is independent of the density. For a periodic lattice due to a Wigner lattice or charge density wave, however, this is not the case [71]. If $n\Omega$ = a constant, where Ω is the unit cell area, as is the case for a charge density wave, it is easily shown using eqs. (90) and (154) that

$$\sigma_{\mathrm{H}} = \nu e^2/h \tag{155}$$

where ν is the filling factor (i.e., the fraction of the Landau level which is filled). The quantity ν is clearly not an integer. Although this result appears to be in contradiction with refs. [61, 66, 70], it actually is not because the Wigner lattice and charge density wave cases are many body problems and these references treat non-interacting systems.

Thouless et al. [61] have given a different proof of the theorem which states that for a non-interacting system, the Hall conductivity must be an integral number of conductivity quanta. For the case in which $Q/2\pi$ is a rational number, the solutions to the Schroedinger equation for the magnetic field problem are Bloch functions because in that case the system has translational symmetry. The periodic part of the Bloch function will then satisfy a Schroedinger equation with a Hamiltonian in the Landau gauge given by

$$\hat{H}(k_x, k_y) = \frac{1}{2m}\left(\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x} + \hbar k_x\right)^2 + \frac{1}{2m}\left(\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial y} + \hbar k_y - \frac{e}{c}Hx\right)^2 + V(x, y)\,, \tag{156}$$

where V is the periodic potential. The velocity operator is easily shown to be given by $(1/\hbar)(\partial\hat{H}/\partial k_x, \partial\hat{H}/\partial k_y)$. Then, the Kubo formula for the Hall conductivity is

$$\sigma_{\mathrm{H}} = \frac{\mathrm{i}e^2}{A\hbar}\sum_{\varepsilon_\alpha < E_{\mathrm{F}}}\sum_{\varepsilon_\beta > E_{\mathrm{F}}} \frac{\langle\alpha|\partial\hat{H}/\partial k_x|\beta\rangle\langle\beta|\partial\hat{H}/\partial k_y|\alpha\rangle - \langle\alpha|\partial\hat{H}/\partial k_y|\beta\rangle\langle\beta|\partial\hat{H}/\partial k_x|\alpha\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2}\,, \tag{157}$$

where A is the area of the system and α and β label the bands. It can easily be shown that

$$\langle\alpha|\nabla_k\hat{H}|\beta\rangle = \nabla_k\varepsilon_\alpha(k)\delta_{\alpha\beta} + (\varepsilon_\beta(k) - \varepsilon_\alpha(k))\int \mathrm{d}^3r\,\mu^*_{k\alpha}\nabla_k(r)\mu_{k\beta} \tag{158}$$

where $\nabla_k = (\partial/\partial k_x,\ \partial/\partial k_y)$ and $\mu_{k\alpha}(r)$ is the periodic part of the Bloch function for the αth band. Substituting eq. (158) into eq. (157), and using completeness of the μ's, we find

$$\sigma_{\mathrm{H}} = \frac{\mathrm{i}e^2}{2\pi h}\sum_\alpha\int \mathrm{d}^2k\int \mathrm{d}^2r\left[\frac{\partial\mu^*_{k\alpha}(r)}{\partial k_x}\frac{\partial\mu_{k\alpha}(r)}{\partial k_y} - \frac{\partial\mu^*_{k\alpha}(r)}{\partial k_y}\frac{\partial\mu_{k\alpha}(r)}{\partial k_x}\right], \tag{159}$$

where the sum over α is over occupied bands only. By Stokes theorem

$$\sigma_{\mathrm{H}} = \frac{\mathrm{i}e^2}{4\pi h} \sum_{\alpha} \int \mathrm{d}^2 r \oint \sum_{j} \mathrm{d}k_j \left(\mu^*_{k\alpha} \frac{\partial \mu_{k\alpha}}{\partial k_j} - \frac{\partial \mu^*_{k\alpha}}{\partial k_j} \mu_{k\alpha} \right), \tag{160}$$

where the line integral on $\boldsymbol{k}$ is around the border of the Brillouin zone. For high order rational values of $Qa/2\pi$, all the discussions in earlier sections of this article show that the bands do not overlap, in which case $\kappa_{k\alpha}$ is a single valued function of $\boldsymbol{k}$. This integrand in eq. (160) reduces to $\partial\theta/\partial k_j$ where θ is the phase of $\mu_{k\alpha}(\boldsymbol{r})$. Since the function $\mu_{k\alpha}$ is single valued, the integral of $\partial\theta/\partial k_j$, gives 2i times 2π for the change of phase on going around the Brillouin zone. Thus, from eq. (160) σ_{H} equals an integer times e^2/h. By using methods similar to the perturbation theory in eqs. (64) and (65), Thouless et al. [61] were able to calculate the value of the integer. This integer jumps quite wildly for small changes in magnetic field. Avron, Seiler and Simon have shown this theorem to be important for characterizing the band structure [44].

One experiment which has actually shown effects of incommensurability on the band structure is the measurement of the de Haas–van Alphen effect in the mercury chain compounds [3] by Razavi et al. [72]. In this experiment, new de Haas–van Alphen frequencies due to incommensurability were found. It should be noted that Ehrenfeund, Ron and Weger have used this picture to explain the nuclear spin relaxation in these compounds using such a picture [73]. The idea is that because of incommensurability, new segments of Fermi surface occur which are obtained from the commensurate crystal Fermi surface by translation by one of the reciprocal lattice vectors of the mercury chain. In a weak magnetic field the electron orbits on the Fermi surface should flow into these segments. Since for an incommensurate system there should be a hierarchy of such Fermi surface segments, there should be extremely small Fermi surface orbits, which will contribute some extremely small de Haas–van Alphen frequencies. In principle this should be the ideal way to observe effects of incommensurability on band structure, but in practice, because of the presence of impurities in the sample, it is necessary to apply relatively strong magnetic fields in order to see the de Haas–van Alphen effect and this causes magnetic break-down, which results in the electron orbits not following most of these small Fermi surface sections. Thus, the presence of this predicted hierarchy of de Haas–van Alphen frequencies due to incommensurability has not been observed. Recent band structure calculations on this compound have been performed to explain this effect [74].

6.4. Superconducting networks

A very exciting experimental manifestation of incommensurability has been found recently by Pannetier et al. [75]. They have discovered structures in the phase boundary (i.e. critical field versus critical temperature curve) for periodic superconducting networks. On the phase boundary it is sufficient to solve the linearized Landau–Ginzberg equation (since the amplitude is small there) [76],

$$[-1/\xi_s^2 + (\mathrm{i}\,\partial/\partial s - K)^2]\Delta = 0\,, \tag{161}$$

where Δ is superconducting amplitude, ξ_s is the coherence length ($\xi_s^{-2} \propto (T_c - T)/T_c$), $s = \hat{\mu}\cdot\boldsymbol{r}$, $K = (2e/\hbar c)\hat{\mu}\cdot\boldsymbol{A}$, where $\hat{\mu}$ is a unit vector along the superconducting strand in the network. At the two ends of the strand ij

$$\Delta_{ij}(s_i) = \Delta_i \qquad \text{and} \qquad \Delta_{ij}(s_j) = \Delta_j \tag{162}$$

where Δ_i and Δ_j are respectively the values of Δ at the i and j ends of the strand. Then, solving eq. (161) subject to these conditions gives

$$\Delta_{ij}(s) = \left[\Delta_i \frac{\sin((l_j - s)/\xi_s)}{\sin(l_{ij}/\xi_s)} + \Delta_j \exp(\mathrm{i}\gamma_{ij}) \frac{\sin(s/\xi_s)}{\sin(l_{ij}/\xi_s)}\right] \exp(-\mathrm{i}k_j s)\,, \tag{163}$$

where $\gamma_{ij} \approx K_{ij} l_{ij}$. Applying the following matching condition on the derivatives of Δ at the boundaries:

$$\sum_j \left[\left(\mathrm{i}\frac{\partial}{\partial s_{ij}} - K_{ij}\right)\Delta_{ij}(s)\right]\bigg|_{s=s_i} = 0\,, \tag{164}$$

we obtain

$$-\Delta_i \sum_j \cot(l_{ij}/\xi_s)/\xi_s + \sum_j \Delta_j \exp\{\mathrm{i}\gamma_{ij}/\xi_s\} \sin(l_{ij}/\xi_s) = 0\,. \tag{165}$$

For a square lattice and in the Landau gauge, $A_x = A_z = 0$, $A_y = Bx$, this equation reduces to

$$0 = -4\Delta_{nm} \cos(a/\xi_s) + \Delta_{n+1,m} + \Delta_{n-1,m} + \mathrm{e}^{\mathrm{i}\gamma n}\Delta_{n,m+1} + \mathrm{e}^{-\mathrm{i}\gamma n}\Delta_{n,m-1}\,, \tag{166}$$

where

$$\gamma = 2\pi B a^2/\phi_0$$

where a is the strand length. If we write

$$\Delta_{nm} = f_n(q)\,\mathrm{e}^{\mathrm{i}qm}\,,$$

we have

$$2\cos(\gamma n + q) f_n + f_{n-1} + f_{n+1} = \varepsilon f_n \tag{167}$$

where

$$\varepsilon = 4\cos(a/\xi_s)\,.$$

This is just the Aubry model with ε given by the above value. Clearly the solution to eq. (167) for a given B corresponding to the highest temperature, and hence the highest value of ε, represents the temperatures on the phase boundary corresponding to that value of B. The structure in the phase boundary observed by Pannetier corresponds to the structure that one would expect on the basis of Hofstadter's work when Ba^2/ϕ_0 takes on rational values. Since a is of the order of microns, it is not difficult to make this quantity take on values of order unity, unlike the case of a crystal lattice. Pannetier et al. have also studied triangular and honeycomb lattice networks. This work represents the first definitive test of the theories discussed in this review.

6.5. Spin waves in incommensurate antiferromagnets

Several theoreticians have suggested that an antiferromagnetic with a linear spin density incommensurate with the crystal lattice should have its spin waves highly broadened by the non-periodic nature of the resulting incommensurate spin structure [77]. This will not happen for a spiral spin structure because such a structure does possess translation group symmetry (except that each of the "translations" in this group consists of a translation plus a rotation of the spin structure). The presence of strong damping seems to be consistent with experiment [78].

Starting with the Hamiltonian

$$\mathcal{H} = -\sum_{ij} J_{ij} \boldsymbol{S}_i \boldsymbol{S}_j - D \sum_i (S_i^z)^2 , \tag{168}$$

the equation of motion for S_i^+ in the random phase approximation is

$$\frac{\hbar}{\mathrm{i}} \dot{S}_i^+ = -2 \sum_j J_{ij} [\langle S_i^z \rangle S_j^+ - \langle S_j^z \rangle S_i^+] + 2D \langle S_i^z \rangle S_i^+ . \tag{169}$$

If we take

$$\langle S_i^z \rangle = S \cos(\boldsymbol{Q} \cdot \boldsymbol{R}_i) , \tag{170}$$

we have

$$\frac{\mathrm{i}}{\hbar} \dot{S}_i^- = [2D + J(\boldsymbol{Q})] S \cos(\boldsymbol{Q} \cdot \boldsymbol{R}_i) S_i^+ - S \cos(\boldsymbol{Q} \cdot \boldsymbol{R}_i) \sum_j J_{ij} S_j^+ . \tag{171}$$

For near neighbor exchange, this equation is similar in form to the Schroedinger equation for the Aubry model, except that in this case the hopping term (the second term on the right hand side) has a spatial dependence. In this system the potential term (the first term on the right hand side), is often larger than the hopping term because of the anisotropy, and preliminary calculations on the above model by the present author similar to those described in section 1 seem to show that the spin wave states are localized. Thus, the observed damping of the spin wave modes is not surprising. Since the modes are localized, the wave vector is not a good quantum number. These magnetic systems clearly provide an experimental example of the incommensurate potential problem in a regime in which the potential term is larger than the hopping term.

6.6. The rings of Saturn

A completely different application of the theory of almost periodic Schroedinger equations was suggested by Avron and Simon [79]. They suggested that the complex structure observed by Voyager I for rings of Saturn could be understood by studying almost periodic Schroedinger equations. Their method is based on Hill's method for studying linear stability of planetary orbits. Consider a planetary orbit defined by the position vector for the planet $\boldsymbol{r}(t)$, which must be a solution to Newton's second law

$$\ddot{\boldsymbol{r}} = \frac{1}{m} \boldsymbol{f}(\boldsymbol{r}) . \tag{172}$$

If we perturb the orbit by replacing $\boldsymbol{r} \to \boldsymbol{r} + \hat{z} w(t)$, for example, where w is assumed small and $\hat{z}$ is a unit vector in the z-direction, w satisfies the eigenvalue equation

$$Hw = Ew \tag{173a}$$

with $E = 0$, where

$$H = -\mathrm{d}^2/\mathrm{d}t^2 + V(t) , \tag{173b}$$

where $V(t) = -\partial F_z(\boldsymbol{r}(t))/\partial z$. Equation (173a) is a periodic "Schroedinger" equation. The condition for stability of the orbit is that the "Hamiltonian" have a zero eigenvalue in its spectrum. If such a solution does not exist, our assumption that w remains small is clearly invalid and w must grow with time, and hence, the orbit is unstable. Let us now consider the stability of an orbit of a particle in Saturn's rings under perturbations due to its moons and the Sun. To do this, we add to $\boldsymbol{F}$ a force $\boldsymbol{F}_1(\boldsymbol{r}(t))$ due to these perturbations. For weak F_1, $\boldsymbol{r} \to \boldsymbol{r} + \boldsymbol{u}$ where $\boldsymbol{u}$ is a small perturbation. Then $\boldsymbol{u}$ satisfies the following approximate form of Newton's law

$$\ddot{\boldsymbol{u}} = (\boldsymbol{u} \cdot \nabla)\boldsymbol{F}(\boldsymbol{r}(t)) + \boldsymbol{F}_1(\boldsymbol{r}(t)) . \tag{174}$$

Applying Hill's criterion for stability of the orbit described by this equation (i.e. adding a perturbation $\hat{z}w(t)$ to $\boldsymbol{r}(t) + \boldsymbol{u}(t)$ and linearizing) we obtain eq. (173a) with

$$H = \frac{-\partial^2}{\partial t^2} + \frac{\partial F(\boldsymbol{r})}{\partial z} + \frac{\partial F_1(\boldsymbol{r})}{\partial z} + \frac{\partial}{\partial z} \boldsymbol{u} \cdot \nabla F(\boldsymbol{r}) . \tag{175}$$

This results in an almost periodic Schroedinger equation in general because the period of the orbit on one hand and the moons of Saturn and the Sun are generally incommensurate. The structure in the eigenvalue spectrum (i.e. its Cantor set nature) is reflected in the structure of the orbits. Using this picture Avron and Simon were able to explain some of the gaps in the rings of Saturn by using perturbation theory. The smaller gaps, inside the larger rings, however, are not explained correctly by their perturbation theory arguments. Most likely the problem is that since the perturbing potential is very weak, this problem is probably equivalent to the case of a very weak modulation potential discussed in section 3. For such a case, the band structure does not differ too much from that of the nearest commensurate case, and this is probably reflected in the orbit structure.

7. Miscellaneous results not previously discussed

This article concludes with a discussion of some rigorous results which did not fit into any previous discussions, and some new results which the author became aware of after the article was completed.

Bellissard and Simon have succeeded in proving in some sense that the support of the spectral density of the Aubry model is a Cantor set [80]. This article proves for the Aubry model what they had previously proven for the differential Schroedinger equation with a type of almost periodic potential called a limit periodic potential. [A limit periodic potential has the form $\Sigma_L V_L \cos(\pi/2^L)$] [81]. Physically, the occurrence of a Cantor set spectrum means that there is a gap in the neighborhood of every energy in the spectrum.

More recently, Avron and Simon [82] have shown completeness of the eigenfunctions of Grempel et al. [28].

Deift and Simon [83] have derived many results about the absolutely continuous spectrum. In addition to Bellissard, Lima and Scopola's [53] work, which was mentioned earlier, Craig [84] and Posuhrl [85] have constructed examples of almost periodic Schroedinger equations with a pure point spectrum.

Recently Llois et al. [86] have studied localization in almost periodic tight binding models for various types of potentials. They basically use the transmission matrix method discussed in section 2 to calculate the wave function, from which they find a localization exponent. They find a mobility edge in all but the cosine potential, in agreement with refs. [7] and [10].

Chao, Riklund and Wahlstrom [87] have recently studied the metal–insulator transition in the Aubry model using the renormalization group method proposed by Jose [19, 20] (discussed in section 2.2 of this article). They find variation with energy of the band structure and the number of iterations necessary to reach the fixed point as one approaches the values of t/V_0 at which the metal–insulator transition takes place. Thus, it appears that even if the transition occurs at $t/V_0 = 1$ for all energies, certain properties of the system, characteristic of the metal–insulator transition, are energy dependent.

Fishman et al. and Pastur and Figotin [88] have recently shown that a three dimensional generalization of the tangent potential model exhibits all localized states. Their model is a near neighbor tight binding model on a simple cubic lattice with a potential proportional to $\tan \boldsymbol{Q}\cdot\boldsymbol{R}$ where $\boldsymbol{R}$ is the positron vector of a lattice site and $\boldsymbol{Q}$ is chosen so the ratios of its components are irrational numbers and $\boldsymbol{Q}$ is incommensurate with the lattice. This guarantees that the planes of constant potential will not coincide with crystal planes, which is probably why there is localization within a plane of constant potential (since lattice sites within a plane are only connected by paths which takes the particle onto neighboring planes).

Machida [89] has considered the self-consistent solutions of the Hartree–Fock approximation Schroedinger equation for linear spin density waves in chromium in a continuum approximation. Although all harmonics of the spin density are included in the calculation, the resulting electronic spectrum possesses only two gaps, because in the continuum approximations, the Schroedinger equation reduces to Lamay's equation, which is known to produce only few gaps.

Acknowledgements

I wish to thank the U.S. Department of Energy (contract number DE-AC02-81ER10866) and the National Science Foundation (grant number DMR-8205480) for their financial support. I also wish to thank R.E. Prange for discussing his work and for giving some insights that he had concerning the model of ref. [52]. I would also like to thank S. Ostlund for carefully explaining his work with Pandit. I would also like to thank B. Simon, J. Bellissard, M. Azbel, M. Kohmoto and M. Wilkinson for their critical reading of the manuscript prior to publication and for their extremely valuable comments.

References

[1] P. Bruesch, S. Strassler and H.R. Zeller, Phys. Rev. B12 (1975) 210.
[2] L.B. Coleman, M.J. Cohen, D.J. Sandman, F.G. Yagamishi, A.F. Garito and A.J. Heeger, Solid State Commun. 12 (1973) 1125.
[3] C.K. Chiang, R. Spal, A. Denenstein, A.J. Heeger, N.D. Miro and A.G. MacDiarmid, Solid State Commun. 22 (1977) 293.
[4] J.D. Axe, in: Proc. Intern. Symp. on Neutron Inelastic Scattering, Vienna, 1977 (Int. Atomic Energy Agency, Vienna, 1978); Y. Shiozaki, Ferroelectrics 2 (1971) 245; P.M. deWolff, Acta Cryst. A30 (1974) 777.
[5] P.M. deWolff, Acta Cryst. A30 (1974) 777; P.M. deWolff, T. Janssen and A. Janner, Acta Cryst. A37 (1981) 625.
[6] A. Janner and T. Janssen, Phys. Rev. B15 (1977) 643.
[7] S. Aubry and C. André, Proc. Israel Physical Society, ed. C.G. Kuper 3 (Adam Hilger, Bristol, 1979) p. 133.
[8] See B. Simon, Adv. Appl. Math. 3 (1982) 463. More recent rigorous results will be discussed later on.
[9] D.J. Thouless, J. Phys. C5 (1972) 77.
[10] C.M. Soukoulis and E.N. Economou, Phys. Rev. Lett. 48 (1982) 1043.
[11] M. Kohmoto, L.P. Kadanoff and C. Tang, Phys. Rev. Lett. 50 (1983) 1870.
[12] S. Ostlund, R. Pandit, H.J. Schellnhuber and E. Siggia, Phys. Rev. Lett. 50 (1983) 1873.
[13] J. Avron and B. Simon, Bull. American Math. Soc. 6 (1982) 81.
[14] J.B. Sokoloff, Phys. Rev. B22 (1980) 5823; Solid State Comm. 40 (1981) 633.
[15] P.W. Anderson, Phys. Rev. 109 (1958) 1492.
[16] E.N. Economou and M.H. Cohen, Phys. Rev. B5 (1972) 2931; B4 (1972) 396.
[17] K.S. Dy and T.C. Ma, J. Phys. C: Solid State Phys. 15 (1982) 6971.
[18] R. Landauer, Phil. Mag. 21 (1970) 863.
[19] J.V. Jose, J. Phys.: A. Math. Gen. 16 (1983) L205; Proc. Latin American Summer School, Cali, Columbia (1982), eds. B. Gomez et al. (World Scientific, 1983) p. 209.
[20] J.B. Sokoloff and J.V. Jose, Phys. Rev. Lett. 49 (1982) 334, 700 (err.).
[21] M.V. Romerio, J. Math. Phys. 12 (1971) 552.
[22] E.J. Dinaberg and Ya.G. Sinai, Func. Annal. Appl. 9 (1976) 279.
[23] J. Bellissard, R. Lima and D. Testard, unpublished.
[24] C. deLange and T. Janssen, Phys. Rev. B28 (1983) 195.
[25] M.Ya Azbel, Phys. Rev. Lett. 43 (1979) 1954.
[26] J. Bellissard, A. Formoso, R. Lima and D. Testard, Phys. Rev. B26 (1982) 3020; B. Simon, private communication.
[27] P. Erdos and R.C. Herndon, Adv. Phys. 31 (1982) 65; Helv. Phys. Acta 50 (1977) 513.
[28] D.R. Grempel, S. Fishman and R.E. Prange, Phys. Rev. Lett. 49 (1982) 833.
[29] D.R. Grempel, R.E. Prange and S. Fishman, Phys. Rev. A29 (1984) 1639.
[30] J.B. Sokoloff, J.E. Sacco and J.F. Weisz, Phys. Rev. Lett. 41 (1978) 1561.
[31] J.B. Sokoloff, Phys. Rev. B23 (1981) 6422.
[32] J.B. Sokoloff, Phys. Rev. B23 (1981) 2039.
[33] L.D. Landau and E.M. Lifshitz, Quantum Mechanics – Non-Relativistic Theory (Pergamon, New York, 1977) pp. 80–81.
[34] G.E. Zil'berman, Zh. Eksp. Teor. Fiz. 32 (1957) 296 [Sov. Phys. JETP 5 (1957) 208]; 33 (1957) 387 [6 (1958) 299].
[34a] M. Wilkinson, Proc. Roy. Soc. (London), in press; Ph.D. Thesis (unpublished).
[35] M. Wilkinson, Proc. Royal Society (London) A39 (1984) 305. A similar calculation has been given by I.M. Suslov, Zh. Eksp. Teor. Fiz. 83 (1982) 1079 [Sov. Phys. JETP 56 (1982) 612].
[36] M.Ya. Azbel, Zh. Eksp. Teor. Fiz. 46 (1964) 929 [Sov. Phys. JETP 19 (1964) 634]; Dokl. Akad. Nauk. 159 (1964) 703 [Sov. Math. Dokl. 5 (1964) 1549].
[37] M.Ya. Azbel and M. Robinstein, Phys. Rev. B27 (1983) 6530.
[38] D.R. Hofstadter, Phys. Rev. B14 (1976) 2239; Ph.D. Thesis, University of Oregon, 1976, unpublished.
[39] D.J. Thouless, Phys. Rev. B28 (1983) 4272.
[40] G.H. Wannier, Phys. Status Solidi B88 (1978) 757.
[41] F.H. Claro and G.H. Wannier, Phys. Rev. B19 (1979) 6068.
[42] R. Johnson and J. Moser, Commun. Math. Phys. 84 (1982) 403; 90 (1983) 317 (err.).
[43] F. Delyon and B. Souillard, Commun. Math. Phys. 89 (1983) 415.
[44] J.E. Avron, R. Seiler and B. Simon, Phys. Rev. Lett. 51 (1983) 51.
[45] S. Ostlund and R. Pandit, Phys. Rev. B29 (1984) 1394.
[46] M. Kohmoto, Phys. Rev. Lett. 51 (1983).
[47] M. Kohmoto and Y. Oono, 1983, unpublished.
[48] A. Peres, Phys. Rev. B27 (1983) 6493.

[49] B. Simon, Phys. Rev. B27 (1983) 3859.
[49a] R.E. Prange, D.R. Grempel and Samuel Fishman, Phys. Rev. Lett. 53 (1984) 1582.
[50] D.J. Thouless and Q. Niu, J. Phys. A.: Math. Gen. 16 (1983) 1911.
[51] J.B. Sokoloff, J. Phys. C: Solid State Phys. 17 (1983) 1703.
[52] J. Bellissard, D. Bessis and P. Moussa, Phys. Rev. Lett. 49 (1982) 701.
[52a] I.M. Suslov, Zh. Eksp. Teor. Fiz. 83 (1982) 1079 [Sov. Phys. JETP 56 (1982) 612].
[53] J. Bellissard, R. Lima and E. Scoppola, Comm. Math. Phys. 88 (1983) 465.
[54] J.B. Sokoloff, Phys. Rev. B25 (1982) 5901.
[55] E.J. Melee and M.J. Rice, Phys. Rev. B23 (1981) 5397.
[56] R.J. Cava, V.L. Himes, A.D. Mighell and R.S. Roth, Phys. Rev. B24 (1981) 3634.
[57] S.J. Hillenus, R.V. Coleman, R.M. Fleming and R.J. Cava, Phys. Rev. B23 (1981) 1567.
[58] H.U. Beyeler, Phys. Rev. Lett. 37 (1976) 1557;
H.U. Beyeler, L. Pietronero, S. Strässler and H.J. Wiessman, Phys. Rev. Lett. 38 (1977) 1532.
[59] C.H. Perry, B. Alzyab and D.-W. Liu, Bull. Am. Phys. Soc. 28 (1983) 284; Solid State Ionics, in press.
[60] P.G. Harper, Proc. Phys. Soc. Lond. A68 (1955) 874.
[61] D.J. Thouless, M. Kohmoto, P. Nightingale and M. den Nijs, Phys. Rev. Lett. 49 (1982) 405.
[62] J.A. Wilson, F.J. DiSalvo and S. Mahajan, Adv. Phys. 24 (1975) 117.
[63] See review by B.I. Halperin, Helv. Phys. Acta 56 (1983) 75.
[64] G.A. Baraff and D.C. Tsui, Phys. Rev. B24 (1981) 2274.
[65] R.E. Prange and R. Joynt, Phys. Rev. B25 (1982) 2943;
R.E. Prange, Phys. Rev. B23 (1981) 4802.
[66] P. Streda, J. Phys. C: Solid State Phys. 15 (1982) L1299.
[67] P. Streda, J. Phys. C: Solid State Phys. 15 (1982) L717.
[68] A. Widom and T.D. Clark, J. Phys. D: Appl. Phys. 15 (1982) L181.
[69] A. Widom, Phys. Lett. 90A (1982) 474.
[70] R.B. Laughlin, Phys. Rev. B23 (1981) 5632.
[71] R.S. Markiewicz, A. Widom and J.B. Sokoloff, Phys. Rev. B28 (1983) 3654.
[72] F.S. Razavi, W.R. Datars, D. Chartier and R.J. Gillespie, Phys. Rev. Lett. 42 (1979) 1182.
[73] E. Ehrenfreund, A. Ron and M. Weger, Phys. Rev. Lett. 47 (1981) 942.
[74] J.J.M. Buiting, M. Weger and F.M. Mueller, Solid State Comm. 46 (1983) 857;
J.J.M. Buiting, M. Weger and F.M. Mueller, J. Phys. F14 (1984) 2343.
[75] B. Pannetier, J. Chaussy and R. Rammal, J. Phys. Lett. 44 (1983) 853;
B. Pannetier, Y.Y. Wang and J. Chaussy, LT-17, DH2 (1983);
B. Pannetier, J. Chaussy, R. Rammal and J.C. Villegier, unpublished.
[76] S. Alexander, Phys. Rev. B27 (1983) 1541.
[77] S.H. Liu, J. Mag. Mag. Mat. 22 (1980) 93;
B.R. Cooper, Solid State Physics Vol. 21, eds. F. Seitz, D. Turnbull and H. Ehrenreich (Academic Press, N.Y., 1968) p. 393;
P.A. Lingard, J. Mag. Mag. Mat. 31–34 (1983) 603.
[78] W.G. Stirling, K.A. McEwen and H. Stanley, 1984 Conf. on Mag. and Mag. Mat.
[79] J.E. Avron and B. Simon, Phys. Rev. Lett. 46 (1981) 1166.
[80] J. Bellissard and B. Simon, J. Functional Analysis 48 (1982) 408.
[81] J.E. Avron and B. Simon, Comm. Math. Phys. 82 (1981) 101.
[82] J.E. Avron and B. Simon, unpublished.
[83] Deift and B. Simon, Comm. Math. Phys. 90 (1983); 89 (1983) 227;
Kotani, Proc. 1982 Kyoto Conf. Stochastic Processes (to be published).
[84] W. Craig, Comm. Math. Phys. 88 (1983) 113.
[85] Posuhrl, Comm. Math. Phys. 88 (1983) 447.
[86] A.M. Llois, M. Weissmann and N.V. Cohan, Phys. Rev. B27 (1983) 7379.
[87] K.A. Chao, R. Riklund and G. Wahlstrom, J. Phys. C. Lett.
[88] L.A. Pastur and A.L. Figotin, Pis'ma Zh. Eksp. Teor. Fiz. 37 (1983) 575;
S. Fishman, D.R. Grempel and R.E. Prange, Phys. Rev. B29 (1984) 4272.
[89] K. Machida, Phys. Rev. B30 (1984) 418;
K. Machida and M. Fujita (unpublished).