Proc. R. Soc. Lond. A **391**, 305–350 (1984)
Printed in Great Britain

Critical properties of electron eigenstates in incommensurate systems

BY M. WILKINSON

H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

(*Communicated by M. V. Berry, F.R.S. – Received 7 July* 1983)

This paper describes some properties of the eigenvalue equation

$$\psi_{n+1}+\psi_{n-1}+2\alpha\cos(2\pi\beta n+\Delta)\psi_n = E\psi_n.$$

This is an example of the more general problem of a Hermitian eigenvalue equation in the form of a difference equation with periodic coefficients. These equations arise in solid state physics; they occur in connection with tight-binding models for electrons in one-dimensional solids with an incommensurate modulation of the structure, and in models for the energy bands of Bloch electrons moving in a plane with a perpendicular magnetic field.

The model studied has a critical point when $\alpha = 1$. Following some earlier work by Azbel (Azbel, M. Ya., *Phys. Rev. Lett.* **43**, 1954 (1979)), an approximate renormalization group transformation is derived. This predicts that the spectrum and eigenstates have a remarkable recursive structure at the critical point, which is dependent on the expansion of β as a continued fraction. Also, when β is an irrational number, there is a localization transition from extended states to localized states as α increases through the critical point. This localization transition, which was previously discovered by Aubry & André (Aubry, S. & André, G. *Ann. Israel phys. Soc.* **3**, 133 (1979)) using the Thouless formula for the localization length, is explained by the renormalization group transformation derived here.

1. INTRODUCTION

This paper describes some properties of the eigenvalue equation

$$\psi_{n+1}+\psi_{n-1}+2\alpha\cos(2\pi\beta n+\Delta)\psi_n = E\psi_n, \tag{1.1}$$

in which β is an irrational number. This equation, as well as being interesting in its own right, has applications in solid state physics; it has been used to model electron eigenstates in a one-dimensional solid with an incommensurate modulation of the structure (Aubry & André 1979), and in models of a Bloch electron in a magnetic field (see, for example, Harper 1955).

There is a considerable mathematical literature, reviewed by Simon (1982), about the problem of Schrödinger equations with a quasi-periodic potential, of which (1.1) is an example. It is known that, for a generic system of this type, the spectrum is a Cantor set of non-zero measure (see, for example, Simon 1982) and by using specialized perturbation theories it has been shown that localized and extended

eigenstates exist when the potential is very strong or very weak, respectively (i.e. $\alpha \gg 1$ or $\alpha \ll 1$ for (1.1)); see Craig (1983) and Dinaburg & Sinai (1976).

This paper is concerned with two remarkable effects that occur when $\alpha = 1$. First, Azbel (1964*a*, *b*, 1979) predicted, by using a W.K.B. method, that the spectrum of (1.1) is a Cantor set of zero measure when $\alpha = 1$, with a structure determined by the expansion of β as a continued fraction:

$$\beta = [n_1, n_2, n_3, \ldots] = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \ldots .}}} \tag{1.2}$$

Azbel predicted that there would be (approximately) n_1 bands, each splitting into n_2 sub-bands, each of which splits into n_3 sub-bands, etc.

Secondly, Aubry & André (1979) have shown, by using the Thouless formula for localization length (Thouless 1972), that for almost all irrational β, (1.1) has a localization transition at $\alpha = 1$ from having all eigenstates localized, for $\alpha < 1$, to all eigenstates extended, for $\alpha > 1$.

It is the aim in this paper to give a physical interpretation of Azbel's result, by using an approximate renormalization group (r.g.) method. This method also explains how the localization transition occurs, and a sufficient criterion for this type of critical behaviour is proposed.

The renormalization group method proposed here can be summarized as follows. The plan is to find approximate, localized solutions of (1.1), and to calculate the matrix elements of the Hamiltonian in a basis of these approximate solutions. These approximate solutions can be partitioned, in a natural way, into a set of energy bands, and matrix elements corresponding to inter-band transitions are ignored, as they are of little importance in localization problems. The remaining single band Hamiltonian equation has the same form as (1.1), and describes the system on a longer length scale and narrower energy scale, with new values of α and β. Under the action of this r.g. transformation, $\alpha = 1$ is an unstable fixed value, and β has a chaotic transformation related to its expansion as a continued fraction (1.2).

The use of a W.K.B. method to obtain approximate, localized solutions is described in § 2, and § 3 describes how the analysis of tunnelling effects leads to expressions for the intra-band matrix elements. Section 4 gives a description of the r.g. transformation, and uses the results of §§ 2, 3 to give an explicit solution for the r.g. recursion. Section 5 describes the properties of this r.g. transformation, including the prediction of a hierarchical splitting of the spectrum and the explanation of the localization transition. Section 6 describes the possible relevance of the results to problems in solid state physics, and § 7 gives a brief summary and describes further work in progress on this and related problems.

2. The semiclassical approximation

This section describes how approximate solutions and eigenvalues of (1.1) can be obtained by a W.K.B. method. Approximate, localized, solutions obtained in this way are used, in § 4, to form a new basis for the Hamiltonian equation (1.1).

The equation (1.1) can be written in the form

$$\psi(x+\hbar)+\psi(x-\hbar)+2\alpha\cos x\psi(x) = E\psi(x), \tag{2.1}$$

where

$$x_n = n\hbar+\Delta, \quad \hbar = 2\pi\beta, \quad \psi(x_n) = \psi_n. \tag{2.2}$$

This way of writing difference equations will be used extensively in this paper. The symbol $\hbar$ is used because extensive use is made of semiclassical methods; it is not of course the physical Planck constant of quantum mechanics. The parameter Δ, which sets the origin for the sampling points x_n, will be referred to as the phase of the solution.

Equation (2.1) can also be written

$$\left.\begin{aligned} \hat{H}\psi(x) &= (2\cos\hat{p}+2\cos\hat{x})\,\psi(x) = E\psi(x), \\ \hat{p} &= -\mathrm{i}\hbar\,\mathrm{d}/\mathrm{d}x, \end{aligned}\right\} \tag{2.3}$$

so that (2.1) is obtained by quantizing a classical Hamiltonian given by

$$H(x,p) = 2\cos p+2\alpha\cos x. \tag{2.4}$$

In the limit in which $\hbar$ is small – the semiclassical limit – it is natural to obtain approximate solutions of (2.3) by the W.K.B. method: write

$$\left.\begin{aligned} &\psi(x) = \mathrm{e}^{\pm\mathrm{i}\Sigma(x)/\hbar}, \\ &\Sigma(x) = \int^x p(x')\,\mathrm{d}x'+\mathrm{i}\eta(x)\hbar+O(\hbar^2), \\ &2\cos p(x)+2\alpha\cos x = E; \end{aligned}\right\} \tag{2.5}$$

by requiring that (2.5) satisfy (2.1) to first order in $\hbar$, it is found that

$$\psi(x) = [\sin p(x)]^{-\frac{1}{2}}\exp\left[\pm\frac{\mathrm{i}}{\hbar}\int^x p(x')\,\mathrm{d}x'\right], \tag{2.6}$$

and better approximations are found by expanding $\Sigma(x)$ to higher order in $\hbar$.

The forms of the contours of $H(x,p)$ in the phase plane are very important in determining the nature of the solutions; these contours will be termed phase trajectories. The phase trajectories of (2.4) may be of three types; they can be closed curves, open in the p-direction, or open in the x-direction.

For $\alpha < 1$, all the phase trajectories are either open in x, or are closed curves. For $\alpha > 1$, all are open in p, or else are closed curves. When $\alpha = 1$, all the phase trajectories are closed curves, except for the separatrices at $E = 0$. These cases are illustrated in figure 1.

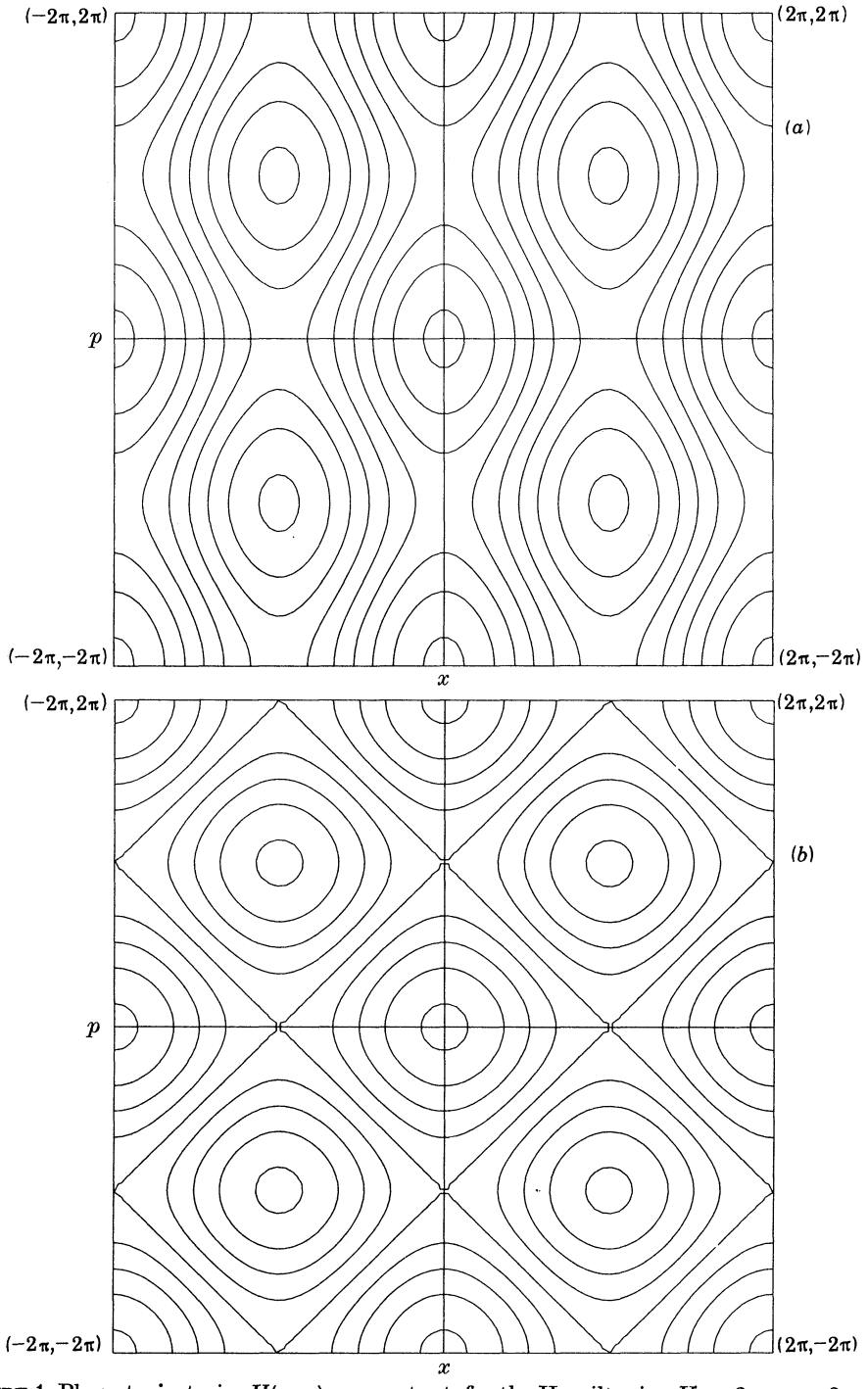

FIGURE 1. Phase trajectories $H(x, p) = \text{constant}$, for the Hamiltonian $H = 2\cos p + 2\alpha \cos x$. (*a*) $\alpha = 2$. The phase trajectories for $1/\alpha = \frac{1}{2}$ are of the same form, but with the x- and p-axes reversed. (*b*) $\alpha = 1$.

In the W.K.B. approximation, ignoring the effects of tunnelling (which will be described in § 3), the phase trajectories open in x support extended states, and the other two types support localized eigenstates.

The energies for which the phase trajectories are open have a band spectrum, in this approximation. For energy ranges over which the phase trajectories are closed orbits, a discrete spectrum is predicted, in the absence of tunnelling effects; in this case the spectrum is obtained using the Bohr–Sommerfeld quantization rule. It is shown in Appendix A that the classical turning points of (2.4) are of the usual first order type except at the separatrix $E = E_s$, so that in the limit $\hbar \to 0$ the eigenvalues, E_m, are given by the usual formula (Landau & Lifshitz 1958)

$$\left.\begin{aligned} S(E_m) &= 2\pi(m+\tfrac{1}{2})\,\hbar, \\ S(E) &= \oint_{H=E} p(x)\,\mathrm{d}x. \end{aligned}\right\} \tag{2.7}$$

Finally, as has been pointed out by Sokoloff (1981), the semiclassical method can be extended to cases in which $\beta = 2\pi\hbar$ is not small, but merely sufficiently close to any rational number; in fact in this method the semiclassicality condition can usually be satisfied to any desired accuracy. The basis of this method is described briefly in Appendix B.

3. Tunnelling between classical trajectories

In the semiclassical case, the localization properties of the eigenstates, and the fine structure of the spectrum of the Hamiltonian (2.3), are determined by tunnelling between classical trajectories of the Hamiltonian. Tunnelling may occur in both the x-direction and the p-direction.

For clarity of presentation, the cases of tunnelling in the x- and p-directions will be illustrated by two systems in which they occur separately.

First it is shown that the effect of tunnelling in the x-direction, between two closed orbits, is equivalent to that of a matrix element, related to the x-tunnelling coefficient, coupling states defined by adjacent Bohr-quantized orbits.

Secondly, it is shown that, for an eigenvalue equation in the form of a difference equation, the dependence of the eigenvalue E for a Bohr-quantized orbit, on the phase Δ of the sampling points, is related to the p-tunnelling coefficient.

Both of these results will be very important in calculating an explicit asymptotic solution for the renormalization group equations in § 4.

(i) *x-Tunnelling*

To illustrate the notation used, consider the well known problem of tunnelling between the classically bounded orbits in a periodic potential.

The Hamiltonian equation is:

$$\left.\begin{aligned} \hat{H}\psi(x) &= (\hat{p}^2/2m + V(x))\,\psi(x) = E\psi(x), \\ V(x+2\pi) &= V(x). \end{aligned}\right\} \tag{3.1}$$

The classical phase trajectories are sketched in figure 2 for the Mathieu equation, $V(x) = \cos x$. If $\hbar$ is small, then for $E_s < E$ approximate eigenstates and eigenvalues are given by Bohr–Sommerfeld quantization, as in (2.7), and the degeneracy of these solutions is lifted by tunnelling through the classically forbidden regions.

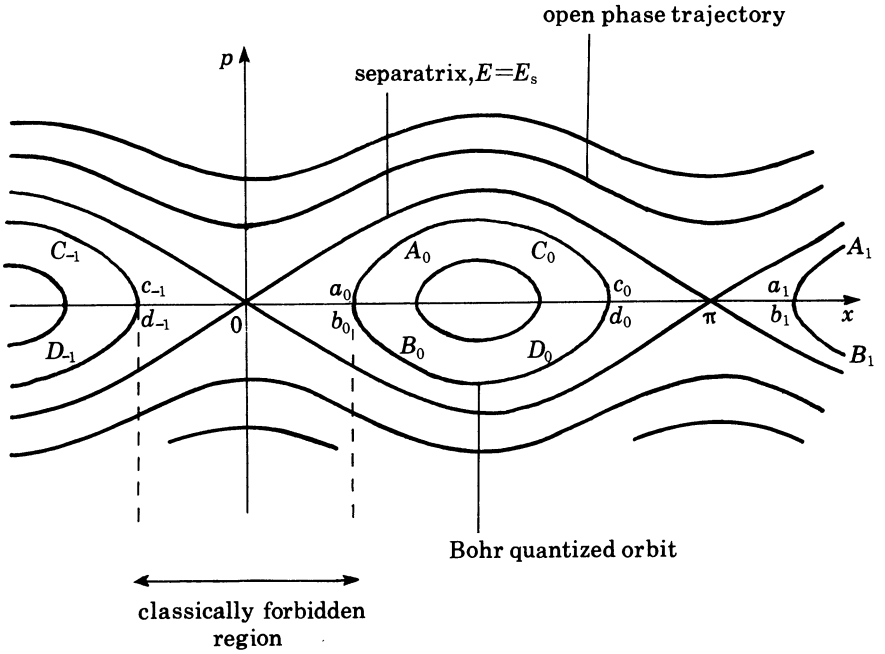

FIGURE 2. Phase trajectories of the Hamiltonian $\hat{H} = \hat{p}^2/2m + V\cos\hat{x}$.

By a well known result (Berry & Mount 1972), the magnitude t_x of the tunnelling coefficient to cross the barrier from x_1 to x_2 is given, in the semiclassical limit, by

$$\left.\begin{aligned} t_x &= \exp\left(-\frac{1}{\hbar}\int_{x_1}^{x_2} |p(x)|\,\mathrm{d}x\right), \\ p^2(x)/2m + V(x) &= E, \end{aligned}\right\} \tag{3.2}$$

and the tunnelling coefficient T_x and reflection coefficient R_x to pass from the branch C_0 of $p(x)$ to branches A_1 and D_0 respectively, are given by:

$$T_x = t_x, \quad R_x = -\mathrm{i}r_x, \quad r_x^2 + t_x^2 = 1. \tag{3.3}$$

Let the amplitudes of the W.K.B. wavefunctions at the classical turning points x_1 and x_2 be $a_0, b_0, a_1, b_1 \ldots$ for the branches $A_0, B_0, A_1, B_1 \ldots$ of $p(x)$ (see figure 2). Then it is easy to show (Arnold 1980) that these amplitudes are coupled by a transfer matrix of the form

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = \bar{T}_x \begin{bmatrix} c_n \\ d_n \end{bmatrix}, \quad \bar{T}_x = \begin{bmatrix} \dfrac{1}{T_x^*} & -\dfrac{R_x^*}{T_x^*} \\ -\dfrac{R_x}{T_x} & \dfrac{1}{T_x} \end{bmatrix}. \tag{3.4}$$

Now consider the propagation of an eigenfunction of energy $E < E_s$. The transfer matrix for crossing the classically allowed region between two turning points is given by

$$\left.\begin{aligned} \begin{bmatrix} c_n \\ d_n \end{bmatrix} = \tilde{F} \begin{bmatrix} a_n \\ b_n \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} e^{\frac{1}{2}i\phi} & 0 \\ 0 & e^{-\frac{1}{2}i\phi} \end{bmatrix}, \\ \phi = \frac{1}{\hbar} \oint p(x)\,\mathrm{d}x = \frac{1}{\hbar} S(E), \end{aligned}\right\} \tag{3.5}$$

the integral in (3.5) is evaluated around the closed orbit $H(x,p) = E$.

The overall transfer matrix is given by

$$\left.\begin{aligned} \begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = \tilde{g} \begin{bmatrix} a_n \\ b_n \end{bmatrix}, \quad \tilde{g} = \tilde{T}_x \tilde{F}, \\ \tilde{g} = \frac{1}{t_x} \begin{bmatrix} \mathrm{e}^{\frac{1}{2}i\phi} & -\mathrm{i}r_x \mathrm{e}^{\frac{1}{2}i\phi} \\ \mathrm{i}r_x \mathrm{e}^{\frac{1}{2}i\phi} & \mathrm{e}^{-\frac{1}{2}i\phi} \end{bmatrix}. \end{aligned}\right\} \tag{3.6}$$

Note that in all problems of this type

$$\det \tilde{g} = 1, \tag{3.7}$$

where $\tilde{g}$ can be any transfer matrix; this expresses the fact that the Wronskian of two linearly independent solutions is a constant.

The condition for eigenstates of (3.1) to satisfy the Bloch condition is that the eigenvalues of g should be $\mathrm{e}^{\pm ik}$; by (3.7) this requires

$$\cos k = \tfrac{1}{2}\mathrm{Tr}\,\tilde{g}. \tag{3.8}$$

In the semiclassical limit, in which t_x is small, the energy bands are found by writing

$$\left.\begin{aligned} \phi = 2\pi(m+\tfrac{1}{2}) + \chi, \quad \chi = \frac{1}{\hbar}\left(\frac{\partial S}{\partial E}\right)_{E_m} \Delta E, \\ \frac{1}{\hbar} S(E_m) = 2\pi(m+\tfrac{1}{2}), \quad E = E_m + \Delta E, \end{aligned}\right\} \tag{3.9}$$

where $m = 0, 1, 2, \ldots$, so that E is given by

$$E = E_m + 2C_m \cos k, \tag{3.10}$$

$$C_m = (-1)^{m+1} \hbar t_x / (\partial S/\partial E)_{E_m}. \tag{3.11}$$

Also, from (3.6) it can be shown that the amplitudes a_n satisfy the difference equation

$$C_m(a_{n+1} + a_{n-1}) + E_m a_n = E a_n \tag{3.12}$$

and that b_n is approximately equal to $-\mathrm{i}a_n$ for solutions that are eigenstates, so that the a_n alone define the solution in the nth potential well. The important point is that (3.12) is the equation of a tight-binding model for the mth band, with hopping matrix elements C_m. The W.K.B. method thus gives an asymptotic

formula for the hopping matrix element of a suitably defined tight-binding model, with Bohr-quantized states as a basis set.

When $E > E_s$, the analysis is much simpler; if weak reflections above the barrier are ignored, the transfer matrix is simply given by

$$\tilde{g} = \begin{bmatrix} e^{\frac{1}{2}i\phi} & 0 \\ 0 & e^{-\frac{1}{2}i\phi} \end{bmatrix}, \tag{3.13}$$

where now

$$\tfrac{1}{2}\phi = S(E)/\hbar, \quad S(E) = \int_{-\pi}^{\pi} p(x)\,dx; \tag{3.14}$$

in this approximation the spectrum is continuous for $E > E_s$ and the Bloch wavevector k, given by (3.8), is related to the energy E by

$$(2\pi m + k)\hbar = S(E), \quad m = 0, \pm 1, \pm 2. \tag{3.15}$$

(ii) *p-Tunnelling*

To illustrate the effect of tunnelling in the p-direction in the W.K.B. analysis of difference equations, consider a discrete Schrödinger equation

$$\psi(x+\hbar) + \psi(x-\hbar) + V(x)\psi(x) = E\psi(x), \tag{3.16}$$

when $V(x)$ has a single minimum, for example $V(x) = \frac{1}{2}kx^2$. Alternatively (3.16) can be written explicitly as a difference equation

$$\begin{aligned} &\psi_{n+1} + \psi_{n-1} + V_n\psi_n = E\psi_n, \\ &\psi_n = \psi(x_n), \quad V_n = V(x_n), \quad x_n = n\hbar + \Delta. \end{aligned} \tag{3.17}$$

Note that the eigenvalue E will be a function of the phase, Δ, of the solution (see figure 3). Because E depends on Δ, the solution of (3.16) must be of the form

$$\psi(x) = \sum_n a_n(\Delta)\,\delta(x - n\hbar - \Delta), \tag{3.18}$$

so that in the momentum representation the phase Δ becomes a Bloch wavevector:

$$\left.\begin{aligned} \psi(p) &= \int dx\, e^{ipx/\hbar}\psi(x), \\ &= e^{ip\Delta/\hbar}\sum_n a_n(\Delta)\, e^{inp}, \\ &= e^{ip\Delta/\hbar}\, U_\Delta(p), \end{aligned}\right\} \tag{3.19}$$

where $U_\Delta(p)$ is 2π-periodic.

The W.K.B. method can be used to obtain approximate solutions when $\hbar$ is small; (3.16) can be written

$$\hat{H}\psi(x) = (2\cos\hat{p} + V(x))\,\psi(x) = E\psi(x) \tag{3.20}$$

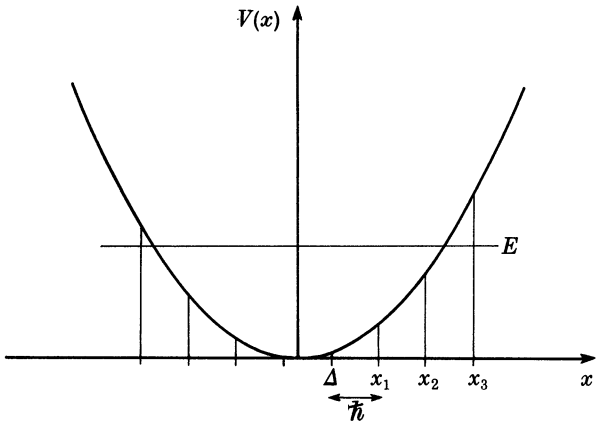

FIGURE 3. Illustration of the definition of the phase Δ, which parametrizes solutions of the difference equation $\psi(x+\hbar)+\psi(x-\hbar)+\frac{1}{2}Kx^2\psi(x)=E\psi(x)$, where $x_n = n\hbar+\Delta$ and $\psi_n = \psi(x_n)$.

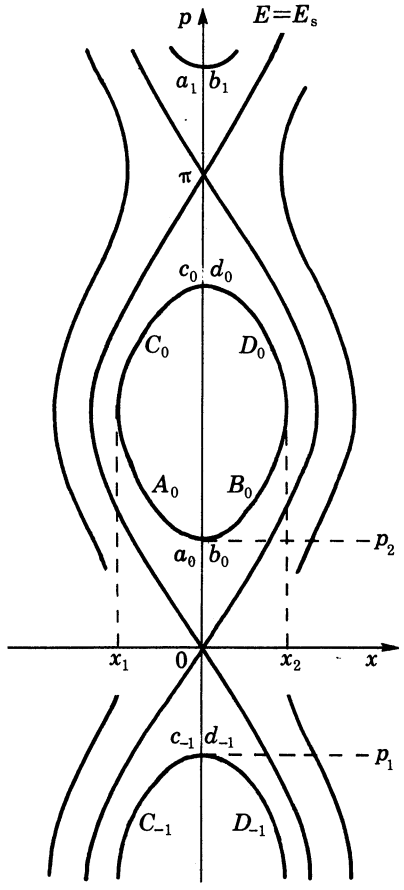

FIGURE 4. Phase trajectories of the Hamiltonian $\hat{H} = 2\cos\hat{p}+\frac{1}{2}K\hat{x}^2$. Capital letters label the branches of $p(x)$, and lower case letters label the amplitudes of the corresponding W.K.B. solutions.

so that the classical Hamiltonian corresponding to $\hat{H}$ is

$$H(x,p) = 2\cos p + V(x). \tag{3.21}$$

The contours of $H(x,p)$ are sketched in figure 4 for the case $V(x) = \frac{1}{2}kx^2$.

The method for obtaining the W.K.B. solutions is analogous to the previous problem. Consider first the closed orbits corresponding to $E < E_s$. The transfer matrix connecting solutions across the classically forbidden regions is written

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = \tilde{T}_p \begin{bmatrix} c_n \\ d_n \end{bmatrix}, \quad \tilde{T}_p = \begin{bmatrix} \dfrac{1}{T_p^*} & \dfrac{-R_p^*}{T_p^*} \\ \dfrac{-R_p}{T_p} & \dfrac{1}{T_p} \end{bmatrix}. \tag{3.22}$$

The transmission and reflection coefficients T_p, R_p are obtained in Appendix A; they are given by:

$$\begin{aligned} &T_p = \mathrm{i}t_p, \quad t_p = \exp\left(-\frac{1}{\hbar}\int_{p_1}^{p_2} |x(p)|\,\mathrm{d}p\right), \\ &R_p = r_p, \quad r_p^2 + t_p^2 = 1. \end{aligned} \tag{3.23}$$

Note that (3.23) has the same form as (3.2), (3.3), apart from a change in phase of the coefficients, and reversal of the roles of x and p.

Taking into account the fact that the reflection coefficients at the first order classical turning points at x_1 and x_2 are, by a standard result (Berry & Mount 1972), $-\mathrm{i}$, the transfer matrix for crossing the classically allowed region is

$$\begin{aligned} &\begin{bmatrix} c_n \\ d_n \end{bmatrix} = \tilde{F} \begin{bmatrix} a_n \\ b_n \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} -\mathrm{i}\,\mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} & 0 \\ 0 & \mathrm{i}\,\mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} \end{bmatrix}, \\ &\phi = \hbar S(E), \quad S(E) = \oint p(x)\,\mathrm{d}x. \end{aligned} \tag{3.24}$$

By working out the overall transfer matrix, $\tilde{g} = \tilde{T}_p \tilde{F}$, and applying (3.8) and (3.9),

$$E = E_m + 2C'_m \cos(2\pi\Delta/\hbar), \tag{3.25}$$

where $$S(E_m) = 2\pi(m+\tfrac{1}{2})\hbar \quad \text{and} \quad C'_m = (-1)^{m+1} t_p \hbar / (\partial S/\partial E)_{E_m}. \tag{3.26}$$

These equations show that the Δ-dependent correction to the Bohr–Sommerfeld quantization (2.7) is due to tunnelling in the p-direction, and varies sinusoidally with Δ. In the context of solving (1.1), this equation determines the variation in site energies for the approximate localized states obtained by Bohr-quantization. There are other, possibly larger, corrections to (2.7) obtained by expanding $\Sigma(x)$ in (2.5) to higher order in $\hbar$, but these corrections are independent of Δ and not important for the purposes of this paper.

For completeness, the result for open phase trajectories in figure 4 for $E > E_s$ is analogous to (3.15),

$$\left.\begin{aligned} 2\pi m\hbar + 2\pi\Delta &= S(E), \\ S(E) &= \int_{-\pi}^{\pi} x(p)\,\mathrm{d}p. \end{aligned}\right\} \tag{3.27}$$

So, for $E < E_s$ the eigenvalues E_m are determined by Bohr-quantization and have an exponentially small dependence on Δ, whereas for $E > E_s$ there is a continuous dependence of E on Δ, given by (3.27).

4. A RENORMALIZATION GROUP TRANSFORMATION

In this section an approximate renormalization group transformation is suggested, which transforms (1.1) into a new difference equation of the same form as (1.1), but with new values of α and β, describing the system on a longer length scale and a narrower energy scale.

(i) *An approximate renormalization group scheme*

The idea behind the method is as follows, given a Schrödinger equation in the form of a tight-binding model, such as (1.1).

(*a*) Approximate, localized, solutions are formed, having a localization length of L 'atomic spacings'. These approximate solutions will be termed 'quasi-eigenstates'.

(*b*) The matrix elements of the Hamiltonian are found using a basis set formed from these quasi-eigenstates.

(*c*) The quasi-eigenstates are partitioned into N bands according to their energy. The approximation that is made is that, for the purposes of examining solutions with energy near E, say, only the interaction of states in the band containing the energy E need be considered. The justification for making this approximation is that the localization properties, and the fine structure of the energy spectrum, depend more on resonances between widely spatially separated quasi-eigenstates of similar energy, than on interactions with nearby states with widely separated energies.

(*d*) To complete the r.g. transformation, two scaling operations are introduced: the energy scale, ΔE, of the band considered is expanded to match the energy scale of the original system, and the length scale is contracted by a factor L.

This procedure produces a new tight-binding equation, describing the system on a longer length scale, and with the energy scale magnified near the energy E.

A similar r.g. scheme has been proposed by Wegner (1976) to study scaling laws at the mobility edges in higher-dimensional Anderson models. This type of scheme is especially well suited to the present problem however: the partitioning of the quasi-eigenstates into bands according to their energies is not arbitrary, as it is in the Wegner problem; it will be shown that the energy bands appear in a natural way, and in the semiclassical case these bands are extremely narrow, so that the 'single band' approximation in step (*c*) is valid to a very high accuracy.

The following subsections describe how the W.K.B. method can be used to construct the quasi-eigenstates and the matrix elements describing their interactions, and so to produce an explicit r.g. recursion. Other methods could, in principle, be used when $\hbar$ is not small, but in practice the W.K.B. method gives qualitatively correct answers in most cases of interest.

(ii) *Approximate solutions used as a new basis*

Given the approximate solutions obtained in § 2, it is natural to use these, wherever possible, as the quasi-eigenstates localized over a range of 2π. By using the W.K.B. method, localized quasi-eigenstates can be constructed on phase trajectories that are either extended in p, or that are closed orbits. For reasons that will become apparent, phase trajectories extended in x are always associated with extended eigenstates.

In each interval of length 2π there are an average of $2\pi/\hbar = 1/\beta$ states. As an example of how a new basis set of quasi-eigenstates could be constructed, consider the important case when $\alpha = 1$, where all the phase trajectories except those at $E = E_s = 0$ are closed orbits (see figure 1). It is possible to construct N pairs of Bohr-quantized quasi-eigenstates (N states with $E > 0$, N with $E < 0$) for every interval of length 2π, where N is given by the condition

$$0 \leqslant 1/\beta - 2N < 3, \tag{4.1}$$

this leaves a residual density of $1/\beta'$ states per interval of length 2π, associated with open phase trajectories at the separatrix, $E_s = 0$,

$$1/\beta' = 1/\beta - 2N. \tag{4.2}$$

These states cannot be approximated by quasi-eigenstates localized over a range of 2π; they can, however, be approximated by quasi-eigenstates with a longer localization length, but this will not be done here.

The Bohr-quantized states are degenerate in the absence of tunnelling. When tunnelling is taken into account, these states are broadened into bands, and in the subsequent analysis the assumption will be made that matrix elements corresponding to inter-band transitions play no important role, and can be neglected – in fact they will not even be calculated. So, in this problem, there is a natural partition of the spectrum into energy bands.

It will be useful to have a convention for labelling the bands, corresponding to Bohr-quantized orbits in phase space, in the case $\alpha = 1$. The central band, corresponding to $E = E_s = 0$, will be labelled the $\nu = 0$ band, and the bands above and below will be labelled $\nu = +1, +2, \ldots$ and $\nu = -1, -2, \ldots$ respectively, in order of increasing and decreasing energy.

(iii) *An asymptotic form for the r.g. transformation*

Consider the case in which $\hbar$ is small, so that semiclassical results are applicable, and the states forming the νth band are all supported by closed orbits. In this case

the results of §§ 2 and 3 enable an explicit calculation of the r.g. transformation to be given, which is believed to be an asymptotic result for the semiclassical limit, $\hbar \to 0$.

The energy of a Bohr-quantized quasi-eigenstate centred on $x = 2\pi n$ is given by (3.25); it includes the energy given by the Bohr-Sommerfeld quantization rule, plus a correction for p-tunnelling. This tunnelling correction takes into account the fact that the energy has a small dependence on the phase, Δ_n, of the sampling points (separated by a distance $\hbar$) compared to the 2π-periodic modulation of the potential. The phase for the quasi-eigenstate centred on $x = 2\pi n$ is

$$\Delta_n = 2\pi n \bmod \hbar, \tag{4.3}$$

so that (3.25) shows that the energy of the quasi-eigenstates varies periodically

$$\left.\begin{aligned} E_n &= E^{(\nu)} + C_p^{(\nu)} \cos 2\pi\beta_1 n, \\ C_p^{(\nu)} &= (-1)^{m+1} t_p^{(\nu)} \hbar / (\partial S/\partial E)_{E^{(\nu)}}, \quad S(E^{(\nu)}) = 2\pi(m+\tfrac{1}{2})\hbar, \end{aligned}\right\} \tag{4.4}$$

where (ν) labels the energy band, and β_1 is given by

$$\beta_1 = 1/\beta - [1/\beta] \tag{4.5}$$

and $[x]$ means 'integer part of x', so that

$$\begin{aligned} 2\pi\Delta_n/\hbar = 4\pi^2 n/\hbar = 2\pi n/\beta = 2\pi\beta_1 n \bmod 2\pi, \\ \cos(2\pi\Delta_n/\hbar) = \cos 2\pi\beta_1 n. \end{aligned} \tag{4.6}$$

There is an obvious relation in (4.3)–(4.6) to the expansion of β as a continued fraction.

In § 3 it was also shown that the effect of x-tunnelling between adjacent closed orbits can be described by hopping matrix elements, $C_x^{(\nu)}$, given by (3.11),

$$C_x^{(\nu)} = (-1)^{m+1} t_x^{(\nu)} \hbar / (\partial S/\partial E)_{E^{(\nu)}}, \tag{4.7}$$

and, to first order in t_x, the matrix elements for hopping beyond the nearest-neighbour orbits are zero.

Applying the assumption contained in step (c) of the r.g. scheme, the eigenfunctions of energy close to $E^{(\nu)}$ are expanded as follows:

$$\psi(x) = \sum_n a_n \phi^{(\nu)}(x - 2\pi n), \tag{4.8}$$

where $\phi^{(\nu)}$ is the W.K.B. approximation for the eigenfunction for the νth band near $x = 0$. Only quasi-eigenstates in the νth band are included, so that matrix elements for inter-band transitions can be ignored.

Given the 'site energies' (4.4), and the 'hopping' matrix elements (4.7), the amplitudes a_n in (4.8) satisfy a tight-binding equation

$$C_x^{(\nu)}(a_{n+1} + a_{n-1}) + 2C_p^{(\nu)} \cos(2\pi\beta_1 n)\, a_n = (E - E^{(\nu)})\, a_n. \tag{4.9}$$

This equation, after dividing through by $C_x^{(\nu)}$, is of exactly the same form as (1.1), except that the parameters α and β have new values, and is the end result of the

first cycle of the r.g. transformation. The general form of the r.g. transformation is described in § 4 (iv).

An alternative derivation of this result, using W.K.B. theory throughout without introducing matrix elements, is described in Appendix C.

It is believed that the result (4.9) represents an asymptotic solution in the limit $\hbar \to 0$, since in this limit the width of the νth band, determined by the tunnelling formulae (4.4) and (4.7), becomes vanishingly small, which implies that the use of a single band approximation (4.8) is increasingly better justified. Since the formulae used for the site energies and hopping matrix elements (4.4) and (4.7) are asymptotic, and other matrix elements for hopping beyond nearest neighbour phase trajectories are of higher order in the transmission coefficients, the result (4.9) can probably be regarded as an asymptotic formula.

(iv) *The r.g. equations in general form*

Now the equations can be given for the iteration of the r.g. transformation described in § 3 (iii). The iteration number of the r.g. transformation will be denoted by N, and the indices of the bands at all previous stages of the r.g. transformation by $(\nu_N) = (\nu_0, \nu_1, \ldots, \nu_N)$ – meaning the ν_Nth sub-band of the ... ν_1th sub-band of the ν_0th band.

At the Nth stage, the approximation to the wavefunction is written by

$$\psi_N(x) = \sum_n \alpha_n \psi_{N-1}^{(\nu_{N-1})}(x_{N-1} - 2\pi n), \tag{4.10}$$

where $\psi_{N-1}^{(\nu_{N-1})}(x_{N-1})$ is the localized approximate solution obtained at the $(N-1)$th stage, for the ν_{N-1}th sub-band of the ν_{N-2}th sub-band ... of the ν_0th band.

The scaled coordinate x_{N-1} is given by

$$x_N = \beta_{N-1} x_{N-1} = \left(\prod_{k=0}^{N-1} \beta_k \right) x. \tag{4.11}$$

The coefficients a_n satisfy the difference equation

$$a_{n+1} + a_{n-1} + 2\alpha_N^{(\nu_{N-1})} \cos(2\pi\beta_N n)\, a_n = E_N a_n, \tag{4.12}$$

this eigenvalue problem can also be written

$$\left.\begin{aligned} &\hat{H}_N^{(\nu_{N-1})} \psi_N = E_N \psi_N, \\ &\hat{H}_N^{(\nu_{N-1})} = \cos \hat{x}_N + \alpha_N^{(\nu_{N-1})} \cos \hat{p}_N, \\ &\hat{p}_N = -\mathrm{i}\hbar_N \, \mathrm{d}/\mathrm{d}x_N, \end{aligned}\right\} \tag{4.13}$$

and is of exactly the same form as (1.1), except that α and β are replaced by $\alpha_N^{(\nu_{N-1})}$ and β_N.

By analogy with the results (4.3) to (4.9), $\alpha_N^{(\nu_{N-1})}$ and β_N are given by

$$\alpha_N^{(\nu_{N-1})} = C_p^{(\nu_{N-1})} / C_x^{(\nu_{N-1})} \tag{4.14}$$

and

$$\left.\begin{aligned} &\beta_N = 1/\beta_{N-1} - [1/\beta_{N-1}], \\ &\hbar_N = 2\pi\beta_N, \end{aligned}\right\} \tag{4.15}$$

where $C_p^{(\nu_{N-1})}$ and $C_x^{(\nu_{N-1})}$ are defined by analogy with (4.4) and (4.6), by using the classical Hamiltonian $H_{N-1}^{(\nu_{N-1})}$ and the Planck constant $\hbar_{N-1}$, to evaluate these formulae.

The scaled energy E_N is given by

$$E_n = [E_{N-1} - E_{N-1}^{(\nu_{N-1})}]/C_x^{(\nu_{N-1})}, \tag{4.16}$$

where $E_{N-1}^{(\nu_{N-1})}$ is the energy of the (ν_{N-1}th) sub-band, expressed in the scaled energy E_{N-1} for the $(N-1)$th stage of the r.g. iteration.

If $\hbar_N$ is small, a solution for the amplitudes a_n can be found by using the W.K.B. method. If this solution corresponds to a closed phase trajectory of the Hamiltonian $H_N^{(\nu_{N-1})}$, then the energy of this solution is given by the Bohr–Sommerfeld quantization rule. So

$$a_n = \phi^{(\nu_N)}(n\hbar_N), \tag{4.17}$$

where $\phi^{(\nu_N)}$ is the W.K.B. approximate wavefunction of energy $E_N^{(\nu_N)}$, given by

$$S(E_N^{(\nu_N)}) = \oint_{H_N = E_N} p_N(x_N)\,\mathrm{d}x_N = 2\pi(m+\tfrac{1}{2})\hbar_N. \tag{4.18}$$

The substitution of (4.17) into (4.10) gives a localized approximation for $\psi_N(x)$, so that equations (4.10)–(4.18) may be further iterated.

To summarize, assuming the expansion (4.10) for the wavefunction, the amplitudes a_n are related by the recursion relation (4.12), which is equivalent to the new Hamiltonian (4.13). The new values of α and β are given by (4.14) and (4.15) and the scalings of x and E by (4.11) and (4.16), respectively. If the phase trajectories of H_N are closed orbits, and $\hbar_N$ is small, an approximate W.K.B. solution is obtained for the amplitudes a_n (4.17), and the r.g. transformation can be further iterated.

Note that this semiclassical solution for the r.g. transformation applies only so long as the phase trajectories are closed, and the $\hbar_N$ are small.

(v) *The role of continued fractions*

The transformation (4.15) applied to β, the ratio of the two periodicities present in (1.1), is independent of that of all the other parameters:

$$\beta_{N+1} = 1/\beta_N - [1/\beta_N], \quad \beta_0 = \beta, \tag{4.19}$$

and is equivalent to the coding of the number β as a continued fraction

$$\beta = [n_1, n_2, n_3, \ldots] = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 \ldots}}}, \tag{4.20}$$

since

$$n_1 = [1/\beta], \quad n_2 = [1/\beta_1] \ldots n_{N+1} = [1/\beta_N]. \tag{4.21}$$

It will be useful to list a few assorted facts about the continued fraction representation of irrational numbers.

First, the sequence of rational numbers

$$\beta_N = p_N/q_N = [n_1, n_2, n_3, \ldots, n_N] = \cfrac{1}{n_1+\cfrac{1}{n_2+\cfrac{1}{n_3+\ldots \cfrac{1}{n_N}}}} \tag{4.22}$$

converges to β as $N \to \infty$ such that

$$|\beta - p_N/q_N| < C/q_N^2 \tag{4.23}$$

for some constant C.

Another important property is that, for almost all irrational numbers β, the coefficients n_N of the continued fraction form a statistically independent random sequence for large N, for which the probability distribution $P(n)$ of the coefficients is

$$P(n) = \frac{1}{\ln 2} \ln \frac{(n+1)^2}{n(n+2)} \tag{4.24}$$

(see Kac 1959), and satisfies

$$P(n) \underset{n\to\infty}{\to} \frac{1}{\ln 2} \frac{1}{n^2}. \tag{4.25}$$

By (4.21) the semiclassicality condition at the Nth stage of the r.g. transformation is $n_N \gg 1$. Equation (4.24) shows that, for almost all irrational numbers β, this condition does not hold for every stage of the r.g. transformation.

An important class of irrational numbers, which do not have the property (4.24) (and so are of zero measure), are the quadratic irrational numbers, for which the sequence of coefficients n_N is periodic for sufficiently large N. The simplest type of quadratic numbers are those of the form $\beta = [n, n, n, \ldots] = \frac{1}{2}[(n^2+1)^{\frac{1}{2}} - n]$; from (4.15) it can be seen that these are invariant under the transformation of β.

For a more detailed description of the continued fraction representation of numbers, see Khinchin (1964).

5. Properties of the renormalization group transformation

In this section the properties of the approximate r.g. transformation derived in § 4 will be described. The r.g. transformation predicts the hierarchical structure of the spectrum and eigenfunctions when $\alpha = 1$, and that a localization transition will occur, with all eigenstates extended for $\alpha < 1$ and all eigenstates localized for $\alpha > 1$. A criterion will be given for these remarkable properties to occur.

Although the r.g. transformation presented in § 4 is approximate, since it depends on a semiclassical argument, the results are well supported by numerical evidence.

Also, it will be seen that the extension of the semiclassical method described in Appendix B provides further evidence for the validity of the results.

(i) *The critical case,* $\alpha = 1$, *in the semiclassical limit*

Consider first the case in which $\alpha = 1$, so that almost all the phase trajectories are closed orbits, and the number $\beta = [n_1, n_2, n_3, \ldots]$ is chosen so that all of the β, defined by (4.15), are small; in this case W.K.B. theory is applicable.

At the lowest level of approximation the spectrum consists of $2N+1$ bands, where N is given by (4.1); note that $2N+1 \simeq n_1$, so that this is consistent with Azbel's statement (Azbel 1964*a*, *b*) that there are approximately n_1 bands.

Of these $2N+1$ bands, all but the central band can be treated with the r.g. transformation (4.10)–(4.18); each band is then described by an equation of the same form as (1.1), with α and β replaced by α_1 and β_1. Because of the symmetry of the classical Hamiltonian (2.4) between x and p when $\alpha = 1$, the magnitudes of the tunnelling coefficients T_x, T_p for x- and p-tunnelling are equal (see (A 7) and (A 17)); thus $\alpha_1 = \alpha = 1$. Since β_1 is assumed to be small, the W.K.B. method can be applied to each of these new difference equations describing each new band; because $\alpha_1 = 1$ there is again the same symmetry of the classical Hamiltonian and a splitting of each band into approximately n_2 sub-bands occurs, etc.

Thus the value $\alpha = 1$ is invariant under the r.g. transformation, and the r.g. transformation predicts a hierarchical splitting of the spectrum; there are (approximately) n_1 bands, each splitting into (approximately) n_2 sub-bands, etc. At each stage of this subdivision of the spectrum, the central band of any group is not described by the r.g. transformation (4.10)–(4.18) (see §4(ii)), so that these equations do not provide a complete description of the spectrum.

Fixed points of the r.g. transformation (4.10)–(4.18) are obtained when $\alpha = 1$ and β is a simple quadratic number; $\beta = [n_1, n_1, n_1, \ldots]$. In this case both α and β are invariant under the r.g. transformation, and for a nested sequence of sub-bands $(\nu) = (\nu_1, \nu_2, \nu_3, \ldots) = (\nu, \nu, \nu, \nu, \ldots)$ (i.e. the νth sub-band of the νth sub-band of ... the νth band), the scale change of the energy variable (4.16) is the same at each iteration of the transformation, so a fixed point of the r.g. transformation is obtained. Plainly an infinity of such fixed points exist (labelled by n and ν), and also periodic orbits in the space of Hamiltonians are possible, for example when $\alpha = 1$ and β is a quadratic number with a cyclic sequence of coefficients in its continued fraction expansion.

These results are illustrated in figure 5 by displaying a magnified section of the continuous spectrum obtained when $\alpha = 1$ and β is a rational number approximating a quadratic irrational as in (4.22). The way in which these computed results were obtained is described briefly in Appendix F.

The eigenstates, as well as the spectrum, exhibit a hierarchical structure when $\alpha = 1$. They can be described by using the recursive definition contained in the r.g. transformation (4.10)–(4.18); the functions $\psi_N(x)$ represent successive localized approximations to $\psi(x)$, exhibiting structure on successively larger length scales.

The first two approximations to $\psi(x)$ are

$$\left.\begin{aligned}\psi_0(x) &= \sum_n a_n \delta(x - n\hbar - \Delta), \\ a_n &= \phi^{(\nu_0)}(x_n), \quad x_n = n\hbar + \Delta,\end{aligned}\right\} \tag{5.1}$$

and

$$\left.\begin{aligned}\psi_1(x) &= \sum_n a_n \psi_0^{(\nu_0)}(x - 2\pi n), \\ a_n &= \phi^{(\nu_1)}(n\hbar_1),\end{aligned}\right\} \tag{5.2}$$

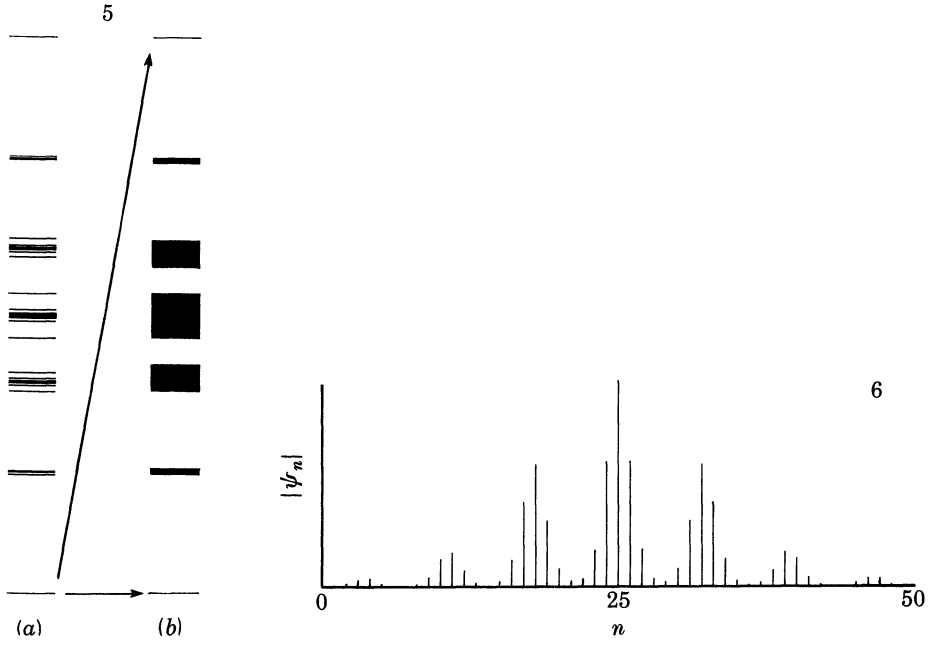

FIGURE 5. Illustration of the self similarity present in the spectrum when $\alpha = 1$ and $\beta = [n, n, n, \ldots]$ is a quadratic irrational number. (*a*) The spectrum for $\beta = [7, 7] = \frac{7}{50}$. (*b*) An enlargement of the lowest lying of the seven bands of (*a*), consisting of seven sub-bands.

FIGURE 6. Illustration of self similarity in the eigenstates when $\alpha = 1$ and β is a quadratic irrational. The state is the $k = 0$, $\Delta = 0$ eigenstate of the lowest of the 50 bands when $\beta = [7, 7] = \frac{7}{50}$ (see Appendix F). The envelope is the same over a range of 50 step lengths, as it is over a range of 7 step lengths.

where $\phi^{(\nu_0)}$ is the W.K.B. quasi-eigenstate for the ν_0th band, and $\phi^{(\nu_1)}$ is the W.K.B. quasi-eigenstate for the ν_1th sub-band of the ν_0th band. By using this notation, the general term of the recursion can be written down. By following the r.g. equations (4.10)–(4.17) this is

$$\left.\begin{aligned}\psi_N(x) &= \sum_n a_n \psi_{N-1}^{(\nu_{N-1})}(x_{N-1} - 2\pi n), \\ x_N = \beta_{N-1} x_{N-1}, &\quad \hbar_N = 2\pi\beta_N, \quad a_n = \phi^{(\nu_N)}(n\hbar_N).\end{aligned}\right\} \tag{5.3}$$

So to form the Nth approximation to the eigenstate, the $(N-1)$th approximation

is duplicated at sites a distance 2π apart in its scaled coordinate x_{N-1}, and the resulting 'comb' function multiplied by an envelope given by the Nth W.K.B. eigenfunction.

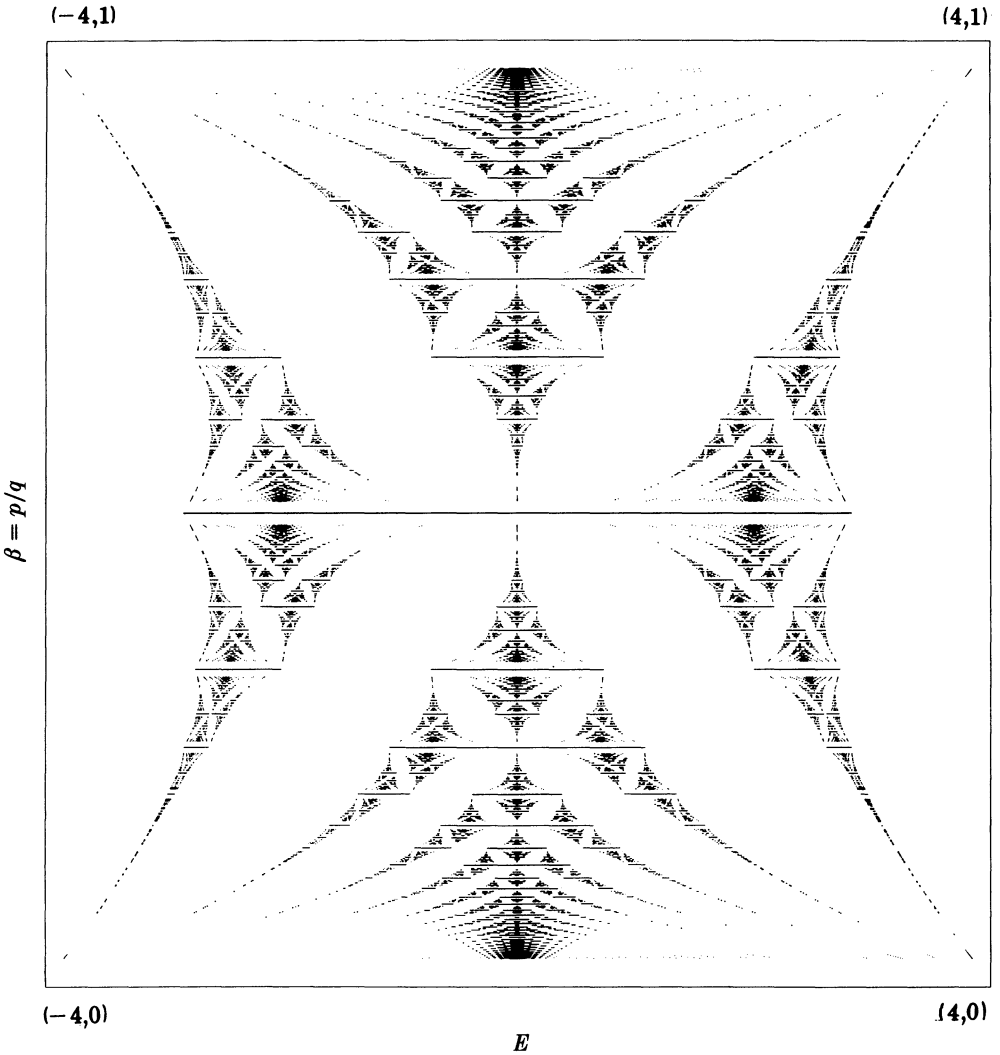

FIGURE 7. A plot of the spectrum of (1.1) when $\alpha = 1$ for every rational value $\beta = p/q$ with $q \leqslant 40$. This plot (invented by Hofstadter (1976)) gives a vivid illustration of the hierarchical properties of the spectrum.

An eigenstate is plotted in figure 6 to illustrate the hierarchical structure of the eigenstates when $\alpha = 1$. As in figure 5, the computed result is for a rational value of $\beta, \beta_N = p_N/q_N$; the structure of the eigenstates in this case exhibits the hierarchical structure described above only up to a length scale of q_N 'atomic spacings'; thereafter of course it is of Bloch form with period $q_N\hbar$.

(ii) *The critical case, $\alpha = 1$, away from the semiclassical limit*

Even when the coefficients n_N of the continued fraction expansion of β are not all large, it is found numerically that the r.g. transformation, (4.10)–(4.18), gives a qualitatively correct description of the spectrum. Hofstadter (1976) invented an interesting way of illustrating the hierarchical nature of the spectrum; his picture is reproduced in figure 7. It consists of a plot of the continuous spectrum of (1.1) for every rational value of β, $\beta = p/q$, with denominator q less some maximum value, against β. It gives a good impression of the hierarchical properties of the spectrum that are predicted for irrational values of β.

As mentioned earlier, the r.g. transformation derived in § 4 does not apply to the $\nu_0 = 0$ band (i.e. the central band), or to any nested sequence of sub-bands containing the central ($\nu = 0$) band of any set of sub-bands. Numerical work, supported by theoretical arguments, suggest that the $\nu_0 = 0$ band can be described by a similar r.g. transformation to that given in § 4, with $\alpha = 1$ invariant and β_1 replaced by β', defined by (4.2). Similar results seem to apply to the central sub-band of every set.

At first sight it may seem surprising that the approximate r.g. transformation given in § 4 gives a good description of the spectrum even when semiclassical conditions do not hold. The extension of the W.K.B. method described in Appendix B provides some insight into why this should be so. As proved in Appendix B, the phase trajectories are again almost all closed orbits when $\alpha = 1$, just as in the standard W.K.B. method. The arguments based on tunnelling between Bohr-quantized orbits can be adapted to this method to produce a difference equation analogous to (4.9), describing the interaction of these quasi-eigenstates, which will again be of the same form as (1.1), but with new parameters α_1^* and β_1^*. In this case also, $\alpha^* = 1$ is invariant under this new version of the r.g. transformation, and a hierarchical subdivision of the spectrum is again predicted. If β_1^* is itself sufficiently close to a rational number for the method of Appendix B to be applicable, this new r.g. transformation can be iterated. Since almost all β are sufficiently close to some rational number for this method to be applied with the semiclassicality condition satisfied to any desired accuracy, this supports evidence that the spectrum probably has a hierarchically clustered structure down to arbitrarily fine energy scales.

For a typical irrational number $\beta = [n_1, n_2, n_3, \ldots]$, the n_N form a statistically stationary random sequence. The predicted hierarchical clustering of the spectrum into ever narrower sub-bands is therefore a statistically self similar process, and the spectrum generated by this process will be a fractal Cantor set (Mandelbrot 1982). Because the distribution of the coefficients n_N takes the universal form (4.24) for almost all β, the fractal dimension, D_f, of the spectrum should have a universal value, D_f^*, for almost all β. It is conjectured that D_f^* is equal to zero; a simplified argument leading to this result is given in Appendix D. For particular values of β forming a set of zero measure the fractal dimension may have a non-zero value; for example, when β is rational D_f is unity and when β is a quadratic number D_f has a non-trivial value between zero and unity.

To summarize, the variable α has a fixed value under the r.g. transformation when $\alpha = 1$. In this case, the semiclassical theory predicts a hierarchical structure of the spectrum, which is, for almost all β, a fractal Cantor set. The eigenstates all have a hierarchical structure; they are extended and exhibit structure on all length scales. Numerical evidence shows that the semiclassical predictions are qualitatively correct even when the coefficients of the continued fraction expansion are not all large. For typical numbers β, the r.g. transformation has a chaotic property due to the 'randomness' of the continued fraction coefficients.

(iii) *Stability of the fixed value* $\alpha = 1$

Now consider what happens when $\alpha = 1 + \delta$, where δ is small, and assume that the semiclassicality condition, that $\beta_N \ll 1$ for all N, holds (the β_N are given by (4.15)).

By examining the equations describing how α is mapped under the r.g. transformation (4.10)–(4.18), (A 7), (A 17),

$$\left.\begin{aligned} \alpha_{N+1}^{(\nu_N)} &= C_p^{(\nu_N)}/C_x^{(\nu_N)} = t_p^{(\nu_N)}/t_x^{(\nu_N)}, \\ t_p^{(\nu_N)} &= \exp\left(-\frac{1}{\hbar_N}\int_{p_1}^{p_2} x_N(p_N)\,\mathrm{d}p_N\right), \\ t_x^{(\nu_N)} &= \exp\left(-\frac{1}{\hbar_N}\int_{x_1}^{x_2} p_N(x_N)\,\mathrm{d}x_N\right), \\ \cos p_N + 2\alpha_N^{(\nu_{N-1})}\cos x_N &= E_N, \end{aligned}\right\} \tag{5.4}$$

it can easily be shown that if $\alpha_N = 1 + \delta_N$, and $\alpha_{N+1} = 1 + \delta_{N+1}$, then the ratio $\kappa_{N+1} = \delta_{N+1}/\delta_N$ satisfies $\kappa_N > 1$, so that $\alpha = 1$ is an unstable fixed point of the variable α: if $\alpha > 1$, then the α_N form an increasing sequence and if $\alpha < 1$, the α_N form a decreasing sequence. The 'critical exponent' κ_N is estimated in Appendix E.

This result also applies if the extended semiclassical method of Appendix B is applied in the case in which not all of the β_N are small; again when α is not equal to unity, one of the tunnelling coefficients becomes progressively larger than the other one as the r.g. transformation is iterated. So, once again, the method developed in Appendix B shows that results derived in the semiclassical limit are qualitatively correct in the general case.

As α diverges away from its critical point at $\alpha = 1$, open phase trajectories appear in the phase plane, and the r.g. transformation (4.10)–(4.18) ceases to be applicable at those energies for which the phase trajectories are extended.

The significance of this instability of the variable α is as follows: suppose that $\alpha > 1$, so that the α_N increase under the r.g. transformation at least up to the point at which the equations (4.10)–(4.18) applying to closed phase trajectories cease to be applicable. Then, if $\alpha > 1$, the variation in 'site energies' of the successive tight-binding models (4.12) (describing the system on longer length-scales) dominates over the 'hopping' matrix elements; this implies a tendency toward localization. Similarly, if $\alpha < 1$, the α_N decrease, so that the 'hopping' matrix elements dominate

over the variation in 'site energies', implying a tendency toward having extended eigenstates. Since the r.g. transformation (4.10)–(4.18) breaks down when the phase trajectories cease to be closed (and anyway is not exact), this is not a proof that a localization transition does occur, and in fact it has been proved that there is a set of values of β of zero measure for which no localization transition occurs (see Gordon (1976); Simon (1982)).

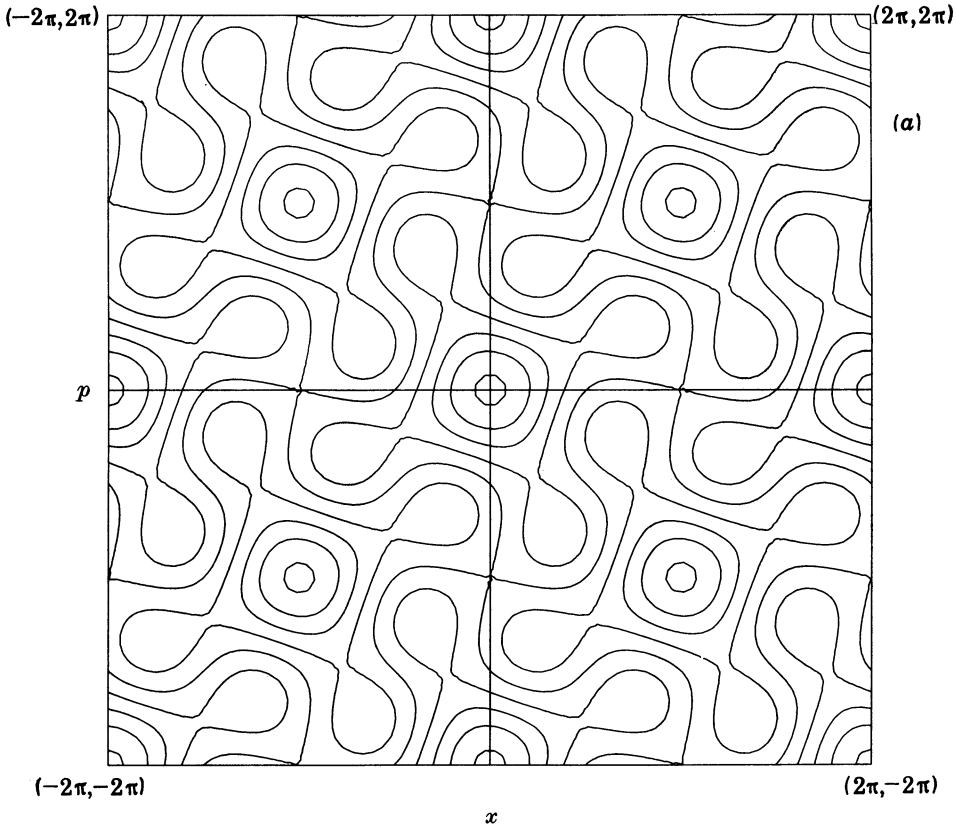

FIGURE 8. Illustration of the quantization of the Hamiltonian defined in (5.11), which has a fourfold symmetry in the phase plane. (*a*) Phase trajectories; (*b*) (opposite) a plot analogous to figure 7, showing hierarchies of structure in the spectrum.

This localization transition at $\alpha = 1$ was previously predicted by Aubry & André (1979) using a more or less rigorous argument based on the Thouless formula (Thouless 1972). The results obtained here are only approximate, but they do give a physically understandable picture of why there is an abrupt localization transition, independent of β and E, at $\alpha = 1$.

(iv) *Generalization to other difference equations*

A natural generalization of the Hamiltonian (2.4) is to consider the most general Hamiltonian that is periodic in x and p,

$$\left.\begin{aligned} H(x,p) &= \sum_{nn'} A_{nn'} \exp\{\mathrm{i}[n(ax+bp)+n'(cx+dp)]\}, \\ A_{nn'} &= A^*_{-n-n'}. \end{aligned}\right\} \tag{5.5}$$

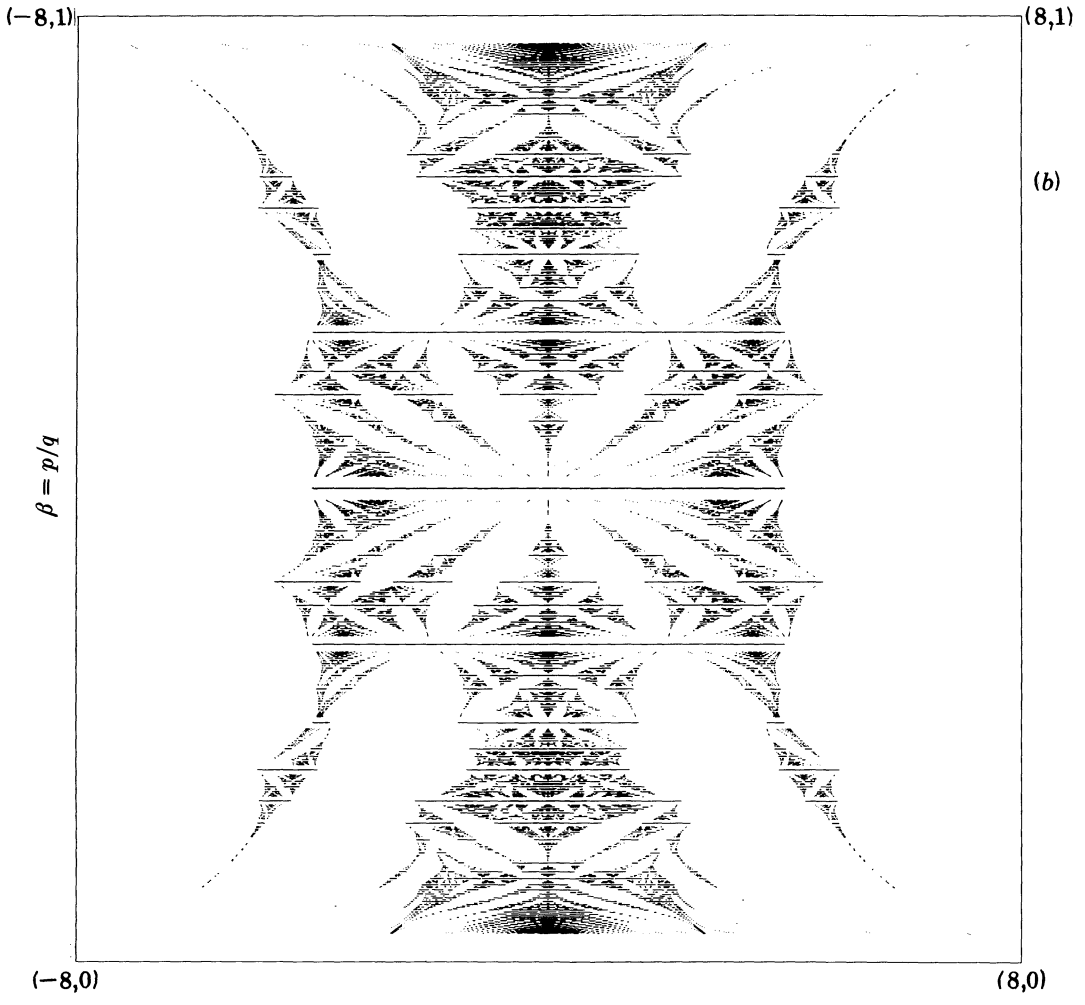

FIGURE 8 (*b*). For description see opposite.

By a suitable linear canonical (i.e. symplectic) transformation:

$$\begin{bmatrix} x \\ p \end{bmatrix} \to \begin{bmatrix} x' \\ p' \end{bmatrix} = \tilde{A} \begin{bmatrix} x \\ p \end{bmatrix}, \quad \det \tilde{A} = 1, \tag{5.6}$$

followed by a (non-canonical) rescaling of x and p; any such Hamiltonian can be written in the form $$H(x',p') = \sum_{nn'} A_{nn'} \exp[\mathrm{i}(nx'+n'p')]. \tag{5.7}$$

Since (5.5) is a multilinear operator and the canonical transformation (5.6) is symplectic, the classical Hamiltonians (5.5) and (5.7) are equivalent not only classically but also quantum mechanically. This can be proved in the Weyl–Wigner formalism (see Groenewold 1946), in which multilinear operators $\hat{f} = f(\hat{x},\hat{p})$ map to Weyl–Wigner representative functions $f(x,p)$ in the phase-plane, and the Wigner functions $W(x,p)$ of eigenstates $\psi(x)$, as well as the representative functions of the operators, map classically under a symplectic transformation of the phase plane (Ozorio de Almeida & Hannay 1982). The non-canonical change of scale can be taken up by rescaling $\hbar$.

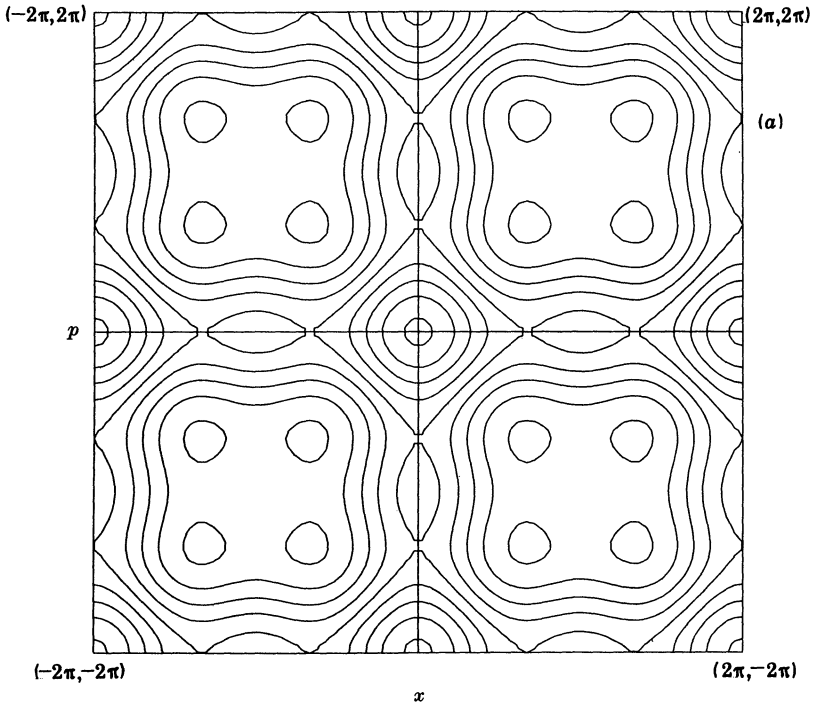

FIGURE 9. Quantization of $\hat{H} = 2(\cos\hat{x}+\cos\hat{p})+\cos 2\hat{p}+\cos 2\hat{x}$. (*a*) Phase trajectories; (*b*) (opposite) a plot of the spectrum, analogous to figure 7.

When (5.7) is quantized it yields a difference equation with periodic coefficients:

$$\begin{aligned}
\hat{H} &= H(\hat{x},\hat{p}) = \sum_{nn'} A_{nn'} \exp[\mathrm{i}(n\hat{x}+n'\hat{p})],\\
\hat{H}\psi(x) &= \sum_{n'}\Big(\sum_{n} A_{nn'} \mathrm{e}^{\mathrm{i}nx} \mathrm{e}^{-\frac{1}{2}nn'\hbar} \mathrm{e}^{\mathrm{i}n'\hbar}\Big) \mathrm{e}^{\mathrm{i}n'\hat{p}}\psi(x),\\
\hat{H}\psi(x) &= \sum_{n'}\Big(\sum_{n} A_{nn'} \mathrm{e}^{\mathrm{i}nx} \mathrm{e}^{-\frac{1}{2}nn'\hbar}\Big)\psi(x+n'\hbar) = E\psi(x)
\end{aligned} \tag{5.8}$$

(a Baker–Hausdorff relation has been used to evaluate $\exp[\mathrm{i}(n\hat{x}+n'\hat{p})]$).

Conversely, any Hermitian eigenvalue equation of the form of a difference equation with periodic coefficients is equivalent to a classical Hamiltonian of the form (5.7); the coefficients $A_{nn'}$ may be functions of the step length $\hbar$.

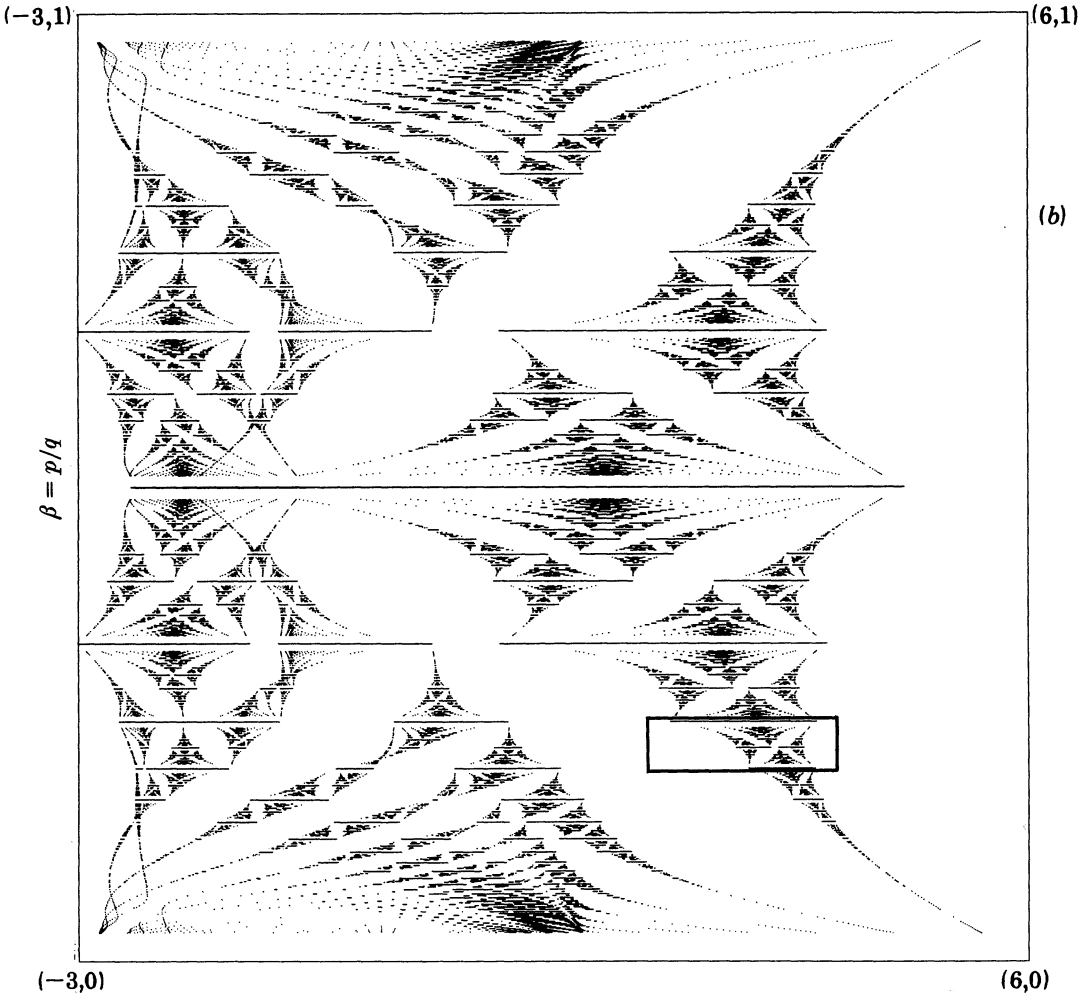

FIGURE 9 (*b*). For description see opposite.

The critical behaviour that occurs in (1.1) when $\alpha = 1$, in which hierarchical structures occur in the spectrum and eigenstates for all values of E and β, can also occur for other Hamiltonians of the form (5.7). From the arguments in §§ 3, 4 and 5, it can be seen that the r.g. transformation normally maps the Hamiltonian toward having extended phase trajectories along some 'preferred' direction in phase space, along which tunnelling occurs most easily. When $\alpha = 1$, and the Hamiltonian (2.4) has points of fourfold symmetry in the phase plane, this tendency is frustrated and the critical behaviour described in §§ 5(i), 5(ii) occurs. These considerations lead

to the proposal of an at least sufficient criterion for this type of critical behaviour in a difference equation of the type (5.8), formulated in terms of the corresponding classical Hamiltonian (5.7). If the classical Hamiltonian is such that, under the effect of a suitable symplectic transformation, every direction in phase space can be transformed into at least one distinct direction, while leaving the classical Hamiltonian invariant, then the critical behaviour will occur (i.e. the spectrum and eigenstates of the difference operator exhibit hierarchical structures for all β and all E).

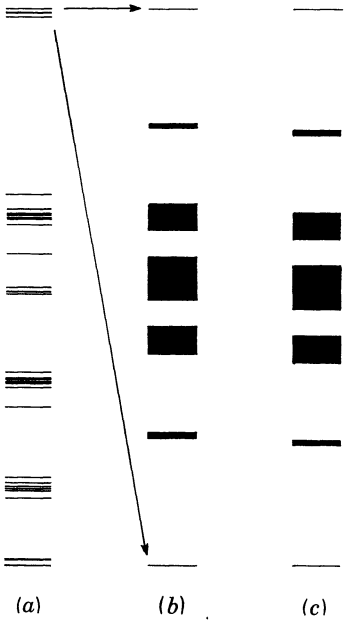

FIGURE 10. Illustration of the universality of the Hamiltonian $\hat{H} = \cos\hat{x} + \cos\hat{p}$, in the semiclassical limit. (*a*) Spectrum of $\hat{H} = 2(\cos\hat{p} + \cos\hat{x}) + \cos 2\hat{p} + \cos 2\hat{x}$, for $\beta = [7, 7] = \frac{7}{50}$, which may be considered small. (*b*) A band of this spectrum is enlarged, showing seven sub-bands. (*c*) The spectrum of $\hat{H} = \cos\hat{x} + \cos\hat{p}$ for $\beta = [7] = \frac{1}{7}$, showing a strong resemblance to (*b*).

This criterion can be rephrased as follows. If the classical Hamiltonian $H(x, p)$, periodic in x and p, is such that, by a suitable symplectic transformation, it is transformed into a form with centres of 3, 4 or 6-fold symmetry in the phase plane, then the 'critical' behaviour results. This result, which shows that a rotational symmetry in the phase plane can have a physical significance, is surprising, but it is well supported by numerical evidence.

For classical Hamiltonians with a fourfold symmetry, expressed by the relation

$$A_{nn'} = A_{n'-n} \tag{5.9}$$

between the Fourier coefficients in (5.7), the Hamiltonian

$$H = \cos x + \cos p \tag{5.10}$$

(i.e. the case $\alpha = 1$ in (2.4)) represents a universal Hamiltonian in the semiclassical

limit. This is to be understood in the following sense. Consider the simplest case, in which there is only one Bohr-quantized orbit per unit cell at the energy considered. Then the W.K.B. analysis can be followed through in just the same way as for (1.1),

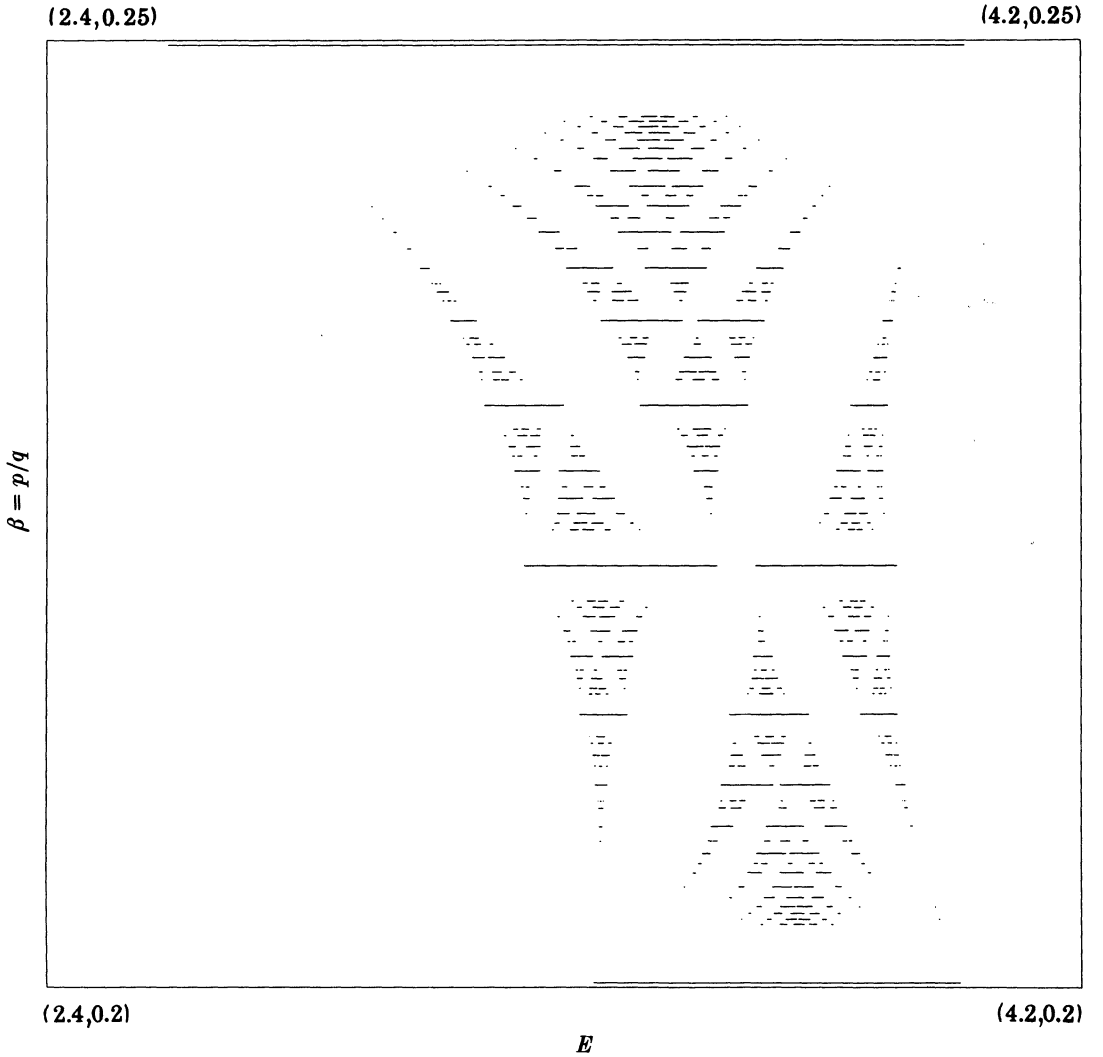

FIGURE 11. Further illustration of the universality of the Hamiltonian $\hat{H} = \cos\hat{x} + \cos\hat{p}$. Apart from a distortion, this enlargement of the region in the small rectangle in figure 9 (*b*) shows a strong similarity to figure 7.

and the tunnelling results show that the W.K.B. quasi-eigenstates have a sinusoidal variation in site energies, and constant nearest neighbour hopping matrix element, which, by symmetry, is equal in magnitude to the amplitude of the variation in site energies. So, in the semiclassical case, the quasi-eigenstates are coupled by a difference equation corresponding to the 'fixed point' Hamiltonian (5.10).

Examples of the quantization of Hamiltonians with fourfold symmetry are illustrated in figures 8 and 9, which show contour plots, and diagrams of the spectrum analogous to figure 7, for the Hamiltonians

$$H = 2\left[\cos x + \cos p + \cos\left(2x - p\right) + \cos\left(x + 2p\right)\right] \tag{5.11}$$

and

$$H = 2\left(\cos p + \cos x\right) + \cos 2p + \cos 2x \tag{5.12}$$

respectively. These plots illustrate the hierarchical structures present in the spectrum.

The universality properties of the Hamiltonian (5.10) are illustrated in figure 10; some of the fine structure present in the spectrum of (5.12) for a rational value of β is compared with a comparable spectrum obtained from the universal Hamiltonian (5.10); a strong similarity is observed. Also, in figure 11, an enlargement of part of figure 9 is shown, which bears a strong resemblance to figure 7, which illustrates the spectrum of (5.10).

An important point is that the Hamiltonian must normally have an exact fourfold symmetry to exhibit the hierarchical properties described here. It is not sufficient for (5.9) to hold only in the limit $\hbar \to 0$, as suggested by Azbel (1964*b*).

Hamiltonians with other than fourfold symmetries have been investigated. Classical Hamiltonians with either centres of threefold or sixfold symmetry belong to the same universality class, which is distinct from that for fourfold symmetry described here. The eigenstates and spectrum also exhibit remarkable hierarchical properties, which can be understood by a W.K.B. analysis similar to that presented here. It is found that the universal Hamiltonian, analogous to (5.10), is

$$H = \cos x + \cos \tfrac{1}{2}(x + \sqrt{3}p) + \cos \tfrac{1}{2}(x - \sqrt{3}p), \tag{5.13}$$

which has centres of sixfold symmetry. The quantized Hamiltonian corresponding to (5.13) describes the eigenstates and spectrum of these systems on large spatial scales and small energy scales. This Hamiltonian has previously been studied numerically by Claro & Wannier (1979).

6. Possible applications in solid state physics

Apart from the intrinsic interest of the equation (1.1), there are two areas of solid state physics where these results may be relevant.

(i) *Electrons in incommensurate structures*

There is an obvious application to the study of electrons and phonons in 'one-dimensional' systems where there are two periodicities present, for example for a quasi-one-dimensional solid with an incommensurate soliton lattice.

In these systems a single-band tight-binding model for the electron Hamiltonian would have the form of a difference equation with periodic coefficients of the form (5.8). For instance, (1.1) would arise from a tight-binding model with constant nearest neighbour hopping matrix elements and a sinusoidally modulated site energy.

The critical behaviour described in § 5 (iv), in which the classical Hamiltonian (5.7) corresponding to these difference equations (5.8) has a fourfold symmetry in the phase plane, does not occur generically in this problem since a symmetry in the phase plane has no physical significance in this context. In general there will not be a global localization transition as a parameter of the model is varied, but a movement of a 'mobility edge' through the spectrum.

When the ratio, β, of the atomic spacing to the period of modulation is small, the semiclassical method can give useful information, however. From the results already discussed, it is apparent that in the semiclassical case contours of the classical Hamiltonian that are open in the x-direction are associated with extended eigenstates, and contours open in the p-direction usually correspond to localized states: closed contours may be associated with either extended or localized eigenstates, depending on which of the magnitudes of the x-tunnelling or the p-tunnelling coefficients is larger, respectively. Thus an approximate criterion for the mobility edge can be given in the semiclassical case, $\beta \ll 1$. The magnitudes $t_x(E)$, $t_p(E)$ of the x- and p-tunnelling coefficients between closed phase trajectories are calculated as a function of the energy, E. Following the discussion in §§ 4 and 5 (iv), the quantity

$$\alpha(E) = t_p(E)/t_x(E), \tag{6.1}$$

which is analogous to α_1 obtained in (4.14), characterizes an equation of the form (4.12) coupling these closed orbits. Then the discussion in § 5 (iii) shows that the condition for the 'mobility edge' is $\alpha(E) = 1$, with extended eigenstates for $\alpha(E) < 1$ and localized eigenstates for $\alpha(E) > 1$. (These results need to be modified slightly according to the type of lattice formed by the closed phase trajectories at energy E.)

(ii) *Bloch electrons in a magnetic field*

Eigenvalue equations of the form (5.8), i.e. difference equations with periodic coefficients, also arise in the problem of solving the Schrödinger equation in two dimensions for a periodic potential with a magnetic field applied perpendicular to the plane of the motion.

In the low magnetic field limit, a difference equation of this type is usually obtained through the use of the Peierls substitution (Peierls 1933; Harper 1955; Hofstadter 1976), which ignores the effect of inter-band transitions introduced by the magnetic field. In the high magnetic field limit, difference equations of this type arise through considering the matrix elements of the Hamiltonian in a basis of Landau states (Rauh 1974, 1975; Schellnhuber *et al.* 1981); in this case the difference equation is obtained by ignoring the matrix elements between Landau states of

different quantum number. So, in both cases, the difference equation arises through the neglect of inter-band effects.

In both of these cases, it is found that the classical Hamiltonian corresponding to the difference equation obtained has the same symmetry (which may, in some cases, only be revealed by making a suitable symplectic transformation) as the periodic potential. This fact is almost self-evident for the Peierls substitution (see Hofstadter (1976) for a simple derivation), and in the high magnetic field limit the result follows from the work of Schellnhuber *et al.* (1981), which obtain the matrix elements coupling Landau states for a general periodic potential.

So the possibility arises, as first hinted at by Azbel (1964*a*), that the hierarchical properties of the spectrum and eigenstates found for 'critical' cases of these difference equations may represent properties of real systems, i.e. the Bloch electrons in a crystal with a magnetic field applied along an axis of 3-, 4-, or 6-fold symmetry; the symmetry in phase space required for the criticality condition is provided here by a physical symmetry of the crystal lattice.

It remains to be seen whether these hierarchical structures really do have a physical significance. Apart from questions of their relevance to many electron systems in real materials, and of the difficulty of experimental observations (see Hofstadter 1976), it still remains to be seen whether these structures occur in exact solutions of the problem when inter-band transitions are not ignored. This question will be discussed in a future paper.

7. Concluding Remarks

The results presented here show how the critical behaviour of (1.1), which gives rise to hierarchical structures in the solutions throughout the spectrum when $\alpha = 1$, may be understood in terms of an approximate renormalization group transformation. This renormalization group transformation also explains the localization transition predicted by Aubry & André (1979).

The results are probably somewhat stronger than they may at first appear, since the renormalization group transformation is obtained in what is probably an asymptotic form, and the small parameter $\hbar$ can usually be made arbitrarily small by using the extended semiclassical method of Appendix B.

An exact renormalization group transformation, based on multiplying strings of transfer matrices, has also been found, and an approximate fixed point has been determined numerically for $\alpha = 1$, $\beta = [1, 1, 1, \ldots] = \frac{1}{2}(\sqrt{5}-1)$. These results may be described in a later paper. Some work along these lines has also been done by Ostlund *et al.* (1983), and Kohmoto *et al.* (1983), who studied an equation related to (1.1) for which it is particularly simple to set up an exact renormalization group recursion.

The critical behaviour described in this paper is strongly related to the symmetry properties of the underlying classical Hamiltonian; the universality class of Hamiltonians of fourfold symmetry has been described in some detail in this paper.

The other universality class, that of Hamiltonians of threefold and sixfold symmetry, will be described in a future paper.

In the application of these results to the problem of Bloch electrons in a magnetic field, these symmetries in phase space, which determine the critical properties, may be provided by a geometrical symmetry of the crystal lattice, in models obtained using a single-band approximation. The question of whether these hierarchical structures also occur in exact solutions is also being investigated.

The method of Appendix B, as well as being important for the argument presented here, has an independent interest: it leads to an observable consequence of the non-single-valued nature of wavefunctions under continuation in a parameter space related to that discussed by Berry (1983), expressed as a new form of Bohr–Sommerfeld quantization. This effect also has applications in calculating corrections to the Bohr–Sommerfeld quantization condition for the motion of Bloch electrons in a weak magnetic field, and in other problems. An extended discussion of this method will also be given in a later paper.

While this paper was being prepared for publication I learned of papers by Thouless & Niu (1983) and Thouless (1983), which contain conclusions that overlap with those of this paper, although they use different methods. As this paper goes to press I have learnt of a paper in which Suslov (1982) arrived independently at many of the same results. I have also learnt of a paper by Ostlund & Pandit (1983) that describes an exact r.g. transformation for (1.1), based on multiplying transfer matrices.

I would like to thank Professor M. V. Berry, F.R.S., and Dr J. H. Hannay for reading the manuscript and making useful suggestions.

I also wish to thank the U.K. Science and Engineering Research Council for a postgraduate studentship.

Appendix A

This appendix describes how the tunnelling coefficients connecting the various branches of the Hamiltonian (2.4) can be obtained. The method used is the 'complex method' of W.K.B. theory, which is well reviewed in a book by Heading (1962).

Consider the function $p(z)$ defined by

$$H = E = 2\cos p(z) + 2\alpha \cos z, \tag{A 1}$$

this is singular whenever $\cos p(z) = \pm 1$, which leads to turning points z_0 satisfying

$$\cos z_0 = (\tfrac{1}{2}E \pm 1)/\alpha. \tag{A 2}$$

Note that the turning points are normally of first order, for example near points where $\cos p = +1$:

$$\left.\begin{aligned} &\cos p(z) \simeq 1 - \tfrac{1}{2}p^2, \\ &\cos p(z) = \tfrac{1}{2}[E - \alpha\cos(z - z_0)] \simeq 1 + \tfrac{1}{2}\alpha \sin z_0 (z - z_0) + \ldots, \end{aligned}\right\} \tag{A 3}$$

so that in the neighbourhood of a turning point

$$p^2 = \alpha \sin z_0 (z - z_0) + \ldots, \tag{A 4}$$

thus the turning points are of first order, except when $\sin z_0 = 0$, corresponding to values of E for which the turning point is a saddle point of $H(x,p)$. The following results describe the treatment of first order turning points only; the solution near saddle points is given by Azbel (1964*a*, *b*), who only treated this case.

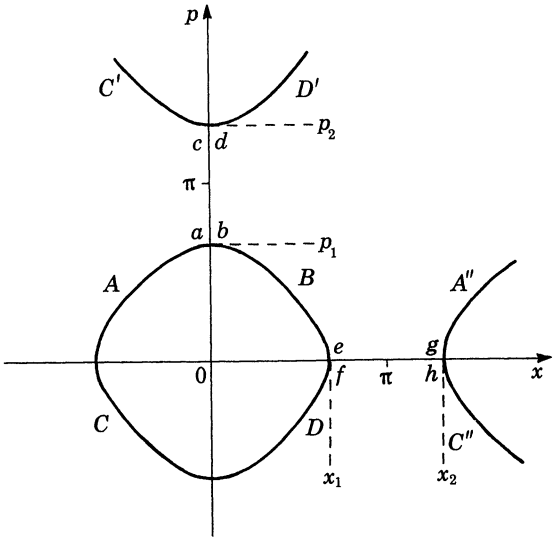

FIGURE 12. Phase trajectories of $H = 2\cos p + 2\alpha \cos x$, when $\alpha = 1$, defining the branches of $p(x)$ (capital letters) and amplitudes of solutions (lower case letters) referred to in Appendix A.

Consider the branches of $p(x)$ labelled $A, B, \ldots$ in figure 12. The solution $\psi(z)$ is a linear combination of solutions

$$\left.\begin{aligned} \psi_X(z, z_0) &= (\sin p_X(z))^{-\frac{1}{2}} \exp\left[\frac{\mathrm{i}}{\hbar} S_X(z, z_0)\right], \\ S_X(z, z_0) &= \int_{z_0}^{z} p_X(z')\,\mathrm{d}z', \end{aligned}\right\} \tag{A 5}$$

where z_0 is the phase reference level, and X labels a branch of $p(z)$. These solutions are combined with amplitudes labelled $a, b, c, \ldots$ at the points indicated in figure 12.

By a well known result (Arnold 1980), the amplitudes e, f, g, h are related by the formula

$$\begin{bmatrix} g \\ h \end{bmatrix} = \tilde{T}_x \begin{bmatrix} e \\ f \end{bmatrix}, \quad \tilde{T}_x = \begin{bmatrix} \dfrac{1}{T_x^*} & \dfrac{-R_x^*}{T_x^*} \\ \dfrac{-R_x}{T_x} & \dfrac{1}{T_x} \end{bmatrix}, \tag{A 6}$$

where T_x is the tunnelling coefficient to go from branch B to branch A'', and R_x is the corresponding reflection coefficient to go from branch B to branch D.

The transmission coefficient is easily shown to be given by the usual formula

$$T_x = \exp\left[\frac{\mathrm{i}}{\hbar}\int_{x_1}^{x_2} p(x')\,\mathrm{d}x'\right], \tag{A 7}$$

where the branch of $p(x)$ between the real turning points x_1 and x_2 is chosen to make T_x small; note that T_x is real. Also, treating the turning point at $x = x_1$ in the usual way gives

$$R_x = -\mathrm{i}r_x, \tag{A 8}$$

where

$$r_x = |R_x|, \quad t_x = |T_x|, \quad r_x^2 + t_x^2 = 1. \tag{A 9}$$

Now consider the more interesting problem of the relation between the amplitudes a, b, c, d, which will be written

$$\begin{bmatrix} c \\ d \end{bmatrix} = \tilde{T}_p \begin{bmatrix} a \\ b \end{bmatrix}, \quad \tilde{T}_p = \begin{bmatrix} \dfrac{1}{T_p^*} & \dfrac{-R_p^*}{T_p^*} \\ \dfrac{-R_p}{T_p} & \dfrac{1}{T_p} \end{bmatrix}. \tag{A 10}$$

The transfer matrix $\tilde{T}_p$ describes tunnelling in the p-direction; T_p is the coefficient for tunnelling from branch A to branch C', and R_p the coefficient for reflection from branch A onto branch B. In more conventional terms, considering motion along the x-axis, T_p is the reflection coefficient for the weak reflection caused by complex turning points of $p(z)$ at $z_0 = \pm \mathrm{i}a$ (given by A 2), and R_p is the corresponding transmission coefficient.

To calculate T_p, require that a solution in the region $x > 0$ has no contribution from $\psi_{D'}$; by using the connection formulae the solution in the region $x < 0$, and therefore T_p, can be found.

Take the phase reference level, z_0, for the solutions (A 5) to be the turning point at $-\mathrm{i}a$ in the lower half plane. The action integrals

$$S_X(z, z_0) = \int_{z_0}^{z} p_X(z')\,\mathrm{d}z' \tag{A 11}$$

can be split into a regular part $(z - z_0)p_0$, where p_0 is the momentum $p(z_0)$ at the turning point, and a singular part

$$S_X^*(z, z_0) = \int_{z_0}^{z} (p_X(z) - p_0)\,\mathrm{d}z. \tag{A 12}$$

The pattern of Stokes lines, defined by $\operatorname{Im} S^* = 0$, and antiStokes lines, defined by $\operatorname{Re} S^* = 0$, associated with the pair of turning points at $z_0 = \pm \mathrm{i}a$ is sketched in figure 13 (see Heading (1962) for a review of the Stokes phenomenon).

The solution $\psi_B(z, -\mathrm{i}a)$ is dominant over $\psi_{D'}(z, -\mathrm{i}a)$ in domain 1 in figure 13, which includes the positive real axis. To verify this, consider the quantity

$$\exp\left[\frac{\mathrm{i}}{\hbar}\int_{-\mathrm{i}a}^{0} p_X(z)\,\mathrm{d}z\right], \tag{A 13}$$

which can be seen, by inspection of figure 14, to be smaller for the branch $p_K(z)$ connecting to the point $p = p_2$, than for $p_L(z)$ connecting to $p = p_1$. Similarly, in domain 2 in figure 13, ψ_A is dominant over $\psi_{C'}$.

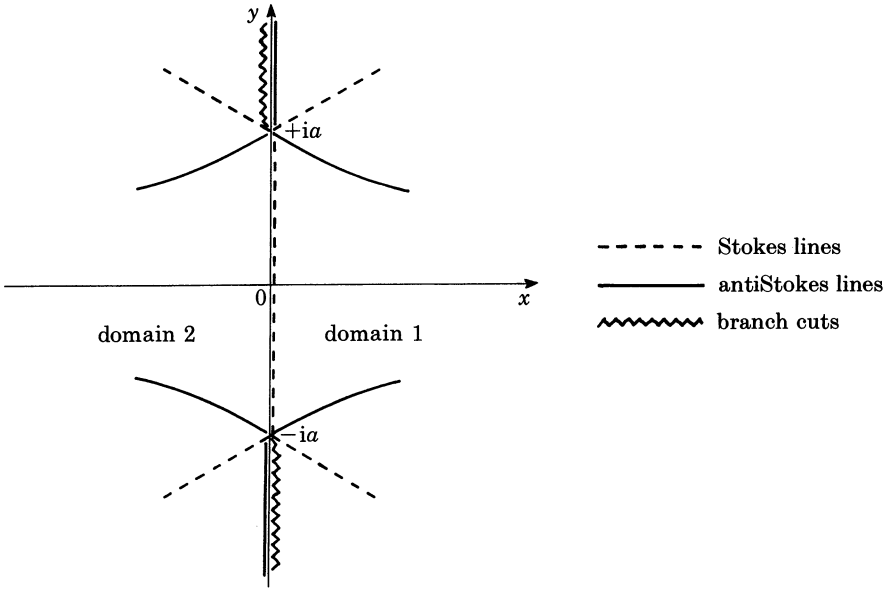

FIGURE 13. Sketch of Stokes lines and antiStokes lines associated with the pair of complex turning points of $p(z)$ at $x = 0$.

On crossing a Stokes line in the positive (anticlockwise) sense, Heading (1962) shows that the Stokes constant is $+\mathrm{i}$. So the connection formula for going from domain 1 to domain 2 in figure 13 is

$$\psi_A(z,-\mathrm{i}a)+\mathrm{i}\psi_{C'}(z,-\mathrm{i}a) \overset{\text{domain 2}}{\leftarrow} \psi(z) \overset{\text{domain 1}}{\rightarrow} \psi_B(z,-\mathrm{i}a). \tag{A 14}$$

Now make the origin, $z = 0$, the phase reference level

$$\psi_A(x,0) \overset{x<0}{+}\mathrm{i}\exp\left\{\frac{1}{\hbar}\int_0^{-\mathrm{i}a}[p_L(z)-p_K(z)]\,\mathrm{d}z\right\}\psi_{C'}(x,0) \leftarrow \psi(z) \rightarrow \psi_B(x,0) \overset{x>0}{}, \tag{A 15}$$

so that the reflection coefficient is, in this approximation,

$$T_p = \mathrm{i}\exp\left\{\frac{\mathrm{i}}{\hbar}\int_0^{-\mathrm{i}a}[p_L(z)-p_K(z)]\,\mathrm{d}z\right\} = \mathrm{i}\exp\left(-\frac{1}{\hbar}W\right), \tag{A 16}$$

where W is the area shaded in figure 14. The important point is that T_p can also be written as

$$T_p = \mathrm{i}\exp\left[\frac{\mathrm{i}}{\hbar}\int_{p_1}^{p_2}x(p)\,\mathrm{d}p\right] = \exp\left[-\frac{1}{\hbar}\int_{p_1}^{p_2}|x(p)|\,\mathrm{d}p\right], \tag{A 17}$$

as is obvious from inspection of figure 14. This formula is analogous to (A 7), and allows the interpretation of T_p as a tunnelling coefficient in the p-direction. The magnitude of T_p, $|T_p| = e^{-W/\hbar}$, could be deduced immediately since this must be invariant under the canonical transformation $x \to x' = p$, $p \to p' = -x$. The phase of T_p is not invariant, however, since transmission in the p-direction does not involve a classical turning point of motion along the x-axis.

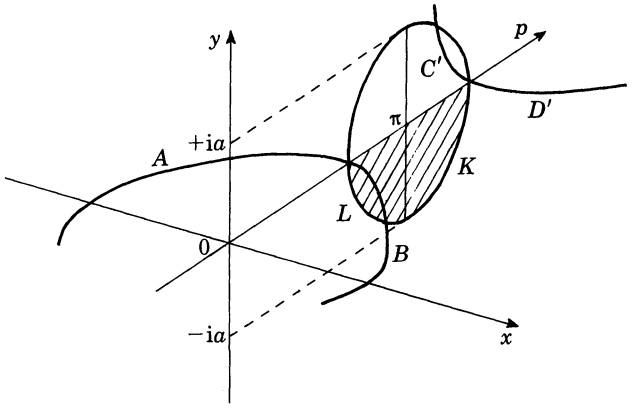

FIGURE 14. Branches of $p(z)$, $z = x+iy$, are labelled A, B, C... for real z, as in figure 12, and K, L, ... for imaginary z.

The simplified calculation described here does not determine R_p directly; by the usual considerations R_p is given by

$$R_p = r_p, \quad r_p^2 + t_p^2 = 1. \tag{A 18}$$

Finally, although the results here apply specifically to the phase trajectories of (2.3), these methods can be used to analyse the tunnelling problem for all the Hamiltonians considered in this paper.

APPENDIX B

Sokoloff (1981) has argued that W.K.B. methods can be applied to difference equations with periodic coefficients, such as (1.1), even when the ratio, β, of the periodicities is not small. This appendix describes some points about this method of particular relevance to this paper.

Suppose that β is a rational number; $\beta = p/q$, where p and q are coprime integers. Then (1.1) takes the form

$$\psi_{n+1} + \psi_{n-1} + 2\alpha \cos(2\pi p n/q + \Delta)\psi_n = E\psi_n, \tag{B 1}$$

which has a solution in Bloch form

$$\psi_n = e^{ikn} \sum_{m=0}^{q-1} a_m(\Delta) \exp(2\pi i p m n/q). \tag{B 2}$$

If β is very close to the rational number p/q, then the solution should locally be very similar to the Bloch solution (B 2), with a phase Δ that varies slowly and linearly with position. By writing

$$\left.\begin{aligned} \beta &= p/q + \Delta\beta, \\ \hbar &= 2\pi\Delta\beta, \\ \psi_n &= \psi(x_n), \quad x_n = n\hbar, \end{aligned}\right\} \tag{B 3}$$

where $\Delta\beta$ is small, (1.1) can be written

$$\psi(x+\hbar) + \psi(x-\hbar) + 2\alpha\cos(2\pi px/q\hbar + x)\,\psi(x) = E\psi(x), \tag{B 4}$$

so that the phase Δ in (B 1) varies slowly with the coordinate x in (B 4). Since, locally, the solution of (B 4) is close to the Bloch form (B 2), an approximate solution takes the form

$$\psi(x) = A(x)\exp[\mathrm{i}S(x)/\hbar]\sum_{m=0}^{q-1} a_m(x)\exp(2\pi\mathrm{i}pmx/q\hbar), \tag{B 5}$$

i.e. a slowly varying amplitude $A(x)$ multiplying a solution in Bloch form. The potential is almost periodic over a range $q\hbar$,and the Bloch wavevector is $S'q$.

By using (B 5) as a trial solution, it is found that, to lowest order in $\hbar$, the slowly varying Fourier coefficients $a_m(x)$ satisfy the difference equation

$$\alpha\,\mathrm{e}^{-\mathrm{i}x}a_{m+1} + \alpha\,\mathrm{e}^{\mathrm{i}x}a_{m-1} + 2\cos(2\pi pm/q + S')\,a_m = E\,a_m. \tag{B 6}$$

The eigenvalue E, and amplitudes a_m in (B 6), depend on the two parameters x and S'; $a_m(x)$ and $S'(x)$ are found by solving (B 6) at constant energy E.

The contours of $E(x, S')$ in the (x, S') plane will be termed phase trajectories, since they play the same role as the phase trajectories introduced in § 2, i.e. $E(x, S')$ is an effective classical Hamiltonian. Note that $E(x, S')$ is a q-valued function, since (B 6) can be solved by diagonalizing a $q \times q$ matrix.

Now consider some properties of the phase trajectories of (B 6). Note that (B 6) is of the same form as the equation for the site amplitudes ψ_n in Bloch form (cf. (B 1)),

$$\begin{aligned} &\psi_n = A_n\,\mathrm{e}^{\mathrm{i}kn} \\ &\mathrm{e}^{\mathrm{i}k}A_{n+1} + \mathrm{e}^{-\mathrm{i}k}A_{n-1} + 2\cos(2\pi pn/q + x)\,A_n = EA_n. \end{aligned} \tag{B 7}$$

This equivalence can be expressed by a duality transformation (previously considered by Aubry & André (1979)) between (B 6) and (B 7); given the equation (B 6) for the a_ms, the equation for the A_ns is obtained by the substitutions

$$\left.\begin{aligned} &\alpha \to 1/\alpha, \quad E \to E/\alpha, \\ &x \to -k, \quad S' \to x. \end{aligned}\right\} \tag{B 8}$$

So (B 6) is invariant under the canonical transformation $(x, S') \to (-k,\ x)$ when $\alpha = 1$, so that the phase trajectories of (B 6) also have a fourfold symmetry in the phase plane when $\alpha = 1$.

It follows immediately from (B 6) that

$$E(x, S') = E(x + 2\pi p/q, S') = E(x + 2\pi, S'),$$

so that, since p and q are coprime, E is periodic in S' with period $2\pi/q$. Also, applying the duality transform (B 8), the same results apply to the x-coordinate, so that E is periodic with period $2\pi/q$ in both x and S'.

In fact it can be shown that $E(x, S')$ takes the form

$$E(x, S') = f_{\alpha, j}(\cos qS' + \alpha^q \cos qx), \tag{B 9}$$

where $f_{\alpha, j}$ is a function that depends on the parameter α and has q branches (labelled by j). This result can be proved as follows. Consider (B 7) in the form

$$\left.\begin{aligned} \begin{bmatrix} \psi_{n+1} \\ \psi_n \end{bmatrix} &= \tilde{M}(x_n, E) \begin{bmatrix} \psi_n \\ \psi_{n-1} \end{bmatrix}, \\ \tilde{M}(x, E) &= \begin{bmatrix} E - 2\alpha \cos x & -1 \\ 1 & 0 \end{bmatrix}, \\ x_{n+1} &= x_n + 2\pi p/q, \quad x_0 = x, \end{aligned}\right\} \tag{B 10}$$

so that, at the position x,

$$\begin{bmatrix} \psi_{(n+1)q+1} \\ \psi_{(n+1)q} \end{bmatrix} = \tilde{T}_\alpha(x, E) \begin{bmatrix} \psi_{nq+1} \\ \psi_{nq} \end{bmatrix}, \tag{B 11}$$

$\tilde{T}_\alpha(x, E) = \tilde{M}(x + 2\pi(q-1)p/q, E) \ldots \tilde{M}(x + 2\pi p/q, E)\, \tilde{M}(x, E)$.

Now, for the ψ_n to form a Bloch solution with Bloch wavevector $S'q$ requires that, since $\tilde{T}$ is unimodular,

$$\cos S'q = \tfrac{1}{2}\mathrm{Tr}\, \tilde{T}_\alpha(x, E). \tag{B 12}$$

By using the duality transformation (B 8) this condition can also be written

$$\cos xq = \mathrm{Tr}\, \tilde{T}_{1/\alpha}(-S', E/\alpha). \tag{B 13}$$

Now (B 12) and (B 13) are two equivalent eigenvalue conditions in the form of qth order polynomials in E, so that their coefficients must be proportional to each other. By comparing (B 12) and (B 14), it can be seen that only the constant term can be a function of x or S', and that this term must contain only a linear combination of $\cos xq$ and $\cos S'q$. Then from (B 11) and (B 12), it follows that the eigenvalue condition takes the form

$$P_\alpha(E) = \cos S'q + \alpha^q \cos xq, \tag{B 14}$$

where $P_\alpha(E)$ is a qth order polynomial in E with coefficients that depend on the parameter α. Since the phase trajectories can be found by diagonalizing a $q \times q$ complex Hermitian matrix, the classical effective Hamiltonian has q branches. In the generic case, no degeneracies between the q sheets of $E(x, S')$ are expected to occur, since three parameters must be varied to degenerate the eigenvalues of a generic complex Hermitian matrix (see Von Neumann & Wigner (1929); Berry (1983)). Because of the special form of (B 6), there are sometimes degeneracies where

sheets of the classical Hamiltonian $E(x, S')$ meet at points in the phase plane. These degeneracies are of a type recently studied by Berry & Wilkinson (1983). From these remarks it is clear that (B 14) can be inverted to give the phase trajectories in the form (B 19). Equation (B 14) has also been obtained by Bellissard & Simon (1982) in another context.

Given the effective classical Hamiltonian $E(x, S')$, solutions supported by its phase trajectories can be constructed, as in standard W.K.B. theory. The analysis of turning points, Bohr–Sommerfeld quantization, tunnelling between phase trajectories, etc. will not be pursued here; the results are similar in most respects to those of standard W.K.B. theory.

The importance of this extended W.K.B. method is as follows. When β is very close to a rational number p/q, the W.K.B. method can still be used, and from (B 9) the phase trajectories have the same combinations of open and closed curves for any given α as occur in the standard W.K.B. method. This means that many of the results of §§ 4, 5, which depend only on the form of the phase trajectories, carry over to the general case in which β is not small.

Now consider the condition for the validity of the extended W.K.B. method proposed here. This is that the period, $2\pi/q$, of the classical Hamiltonian $E(x, S')$ in the phase plane should be much greater than the period, $q\hbar$, of the underlying potential, i.e.

$$q\hbar \ll 2\pi/q \tag{B 15}$$

or

$$|\beta - p/q|\, q^2/2\pi \ll 1. \tag{B 16}$$

By some elementary properties of the continued fraction representation of irrational numbers (see Khinchin 1964) (see also § 4 (iv)) it follows that, for almost all β, there exist numbers p, q such that the left side of (B 16) can be made arbitrarily small, so that the conditions for the W.K.B. method to be applicable can be satisfied to any desired accuracy.

Finally, a very interesting point arises in connection with the Bohr–Sommerfeld quantization condition in this scheme. The Fourier coefficients, a_m, which can be obtained by diagonalizing a $q \times q$ complex Hermitian matrix, cannot be given as single-valued and continuous functions of position in the phase plane; when they are transported in the appropriate way around a closed orbit, they are multiplied by a complex phase factor when they return to their starting point. This phase change is related to an effect that occurs in adiabatic theory, discussed by Berry (1983). The noval feature in this application is that the phase change is directly observable as a correction to the Bohr–Sommerfeld quantization formula. This correction to the Bohr-Sommerfeld formula has wider applications, and will be discussed in detail in a future paper.

Appendix C

This appendix gives an alternative derivation of the renormalization group transformation in the semiclassical limit, by using a W.K.B. method throughout. This was the method originally used by Azbel to study this problem.

The transfer matrix describing the coupling of a general rectangular lattice of closed phase trajectories is given, together with an example of how this is used to derive (4.9) in the semiclassical limit. These results are more general than those given by Azbel (1964*a*, *b*), in that Azbel only considered the case of exact fourfold symmetry, and only considered orbits close to the separatices, where the assumptions underlying the method start to break down.

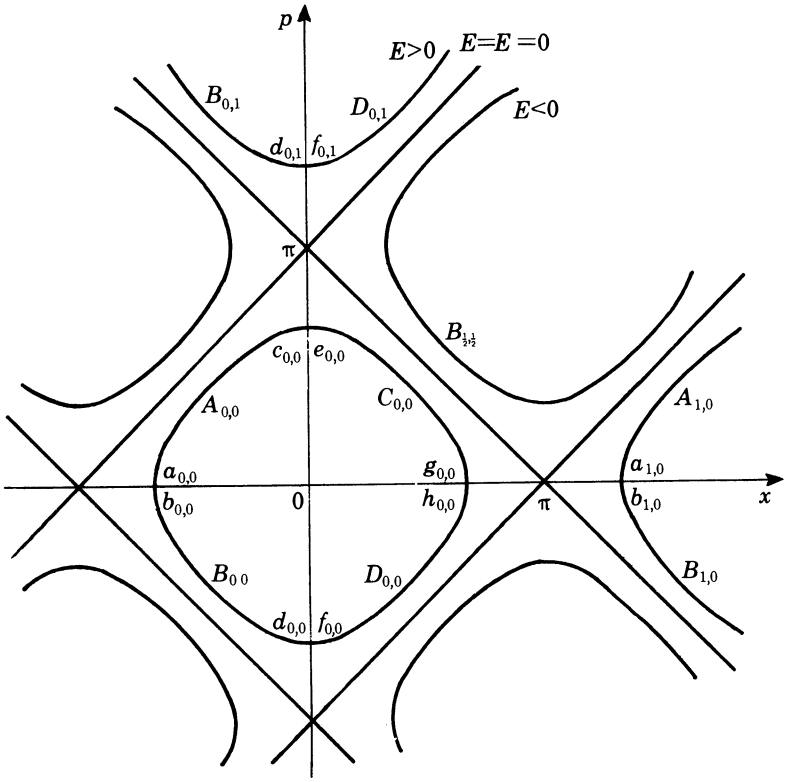

FIGURE 15. Phase trajectories of $H = 2\cos p + 2\alpha\cos x$ when $\alpha = 1$, defining branches of $p(x)$, A_{nm}, B_{nm}, ... and amplitudes a_{nm}, b_{nm}, ... referred to in Appendix C.

The labelling of the branches of $p(x)$ is illustrated in figure 15; the orbits centred on $(2\pi n, 2\pi m)$ are labelled by integers and half-integers n and m, and the branches of $p(x)$ by the letters $A, B, C, \ldots$. Lower case letters $a_{n,m}, b_{n,m}, \ldots$, indicate the amplitudes of the solution at various positions and on various branches of $p(x)$.

Following the results of § 3 (ii) to construct a solution of the difference equation (1.1) out of the approximate W.K.B. solutions in the region of $x = 2\pi n$, the

W.K.B. solutions supported by the n, m orbit are combined with a phase factor z_n^m, where z_n is given by

$$z_n = \mathrm{e}^{2\pi \mathrm{i} \Delta n/\hbar} = \mathrm{e}^{4\pi^2 n \mathrm{i}/\hbar}. \tag{C 1}$$

So the amplitudes $a_{n,m}, b_{n,m}, \ldots$, are related by the Bloch condition as follows:

$$\mu_{n,m+1} = z_n \mu_{n,m}, \tag{C 2}$$

where μ is any one of $a, b, c, \ldots$. This constraint leaves two undetermined coefficients at any value of x, as are required for the solution of a second order equation.

A transfer matrix, $\tilde{g}_n$, will be obtained relating the amplitudes a_n, b_n in successive unit cells

$$\begin{bmatrix} a_{n+1,0} \\ b_{n+1,0} \end{bmatrix} = \tilde{g}_n \begin{bmatrix} a_{n,0} \\ b_{n,0} \end{bmatrix}. \tag{C 3}$$

It is assumed that the transfer matrices $\tilde{T}_x$, $\tilde{T}_p$ describing tunnelling in the x- and p-directions, are known:

$$\left.\begin{aligned} \begin{bmatrix} a_{n+1,m} \\ b_{n+1,m} \end{bmatrix} &= \tilde{T}_x \begin{bmatrix} g_{n,m} \\ h_{n,m} \end{bmatrix}, \quad \tilde{T}_x = \begin{bmatrix} \dfrac{1}{T_x^*} & \dfrac{-R_x^*}{T_x^*} \\ \dfrac{-R_x}{T_x} & \dfrac{1}{T_x} \end{bmatrix}, \\ \begin{bmatrix} d_{n,m+1} \\ f_{n,m+1} \end{bmatrix} &= \tilde{T}_p \begin{bmatrix} c_{n,m} \\ e_{n,m} \end{bmatrix}, \quad \tilde{T}_p = \begin{bmatrix} \dfrac{1}{T_p^*} & \dfrac{-R_p^*}{T_p^*} \\ \dfrac{-R_p}{T_p} & \dfrac{1}{T_p} \end{bmatrix}. \end{aligned}\right\} \tag{C 4}$$

The calculation of T_x, T_p, R_x, R_p in the semiclassical limit is described in Appendix A. By using (C 2) and (C 4), it is found that

$$\left.\begin{aligned} \begin{bmatrix} d_{n,1} \\ f_{n,1} \end{bmatrix} &= z_n \begin{bmatrix} d_{n,0} \\ f_{n,0} \end{bmatrix} = \tilde{T}_p \begin{bmatrix} c_{n,0} \\ e_{n,0} \end{bmatrix}, \\ \begin{bmatrix} e_{n,0} \\ f_{n,0} \end{bmatrix} &= \frac{1}{R_p^*} \begin{bmatrix} 1 & -T_p^* z_n \\ T_p^* z_n^{-1} & \dfrac{-T_p^*}{T_p} \end{bmatrix} \begin{bmatrix} c_{n,0} \\ d_{n,0} \end{bmatrix} \end{aligned}\right\} \tag{C 5}$$

and that

$$\begin{bmatrix} g_{n,0} \\ h_{n,0} \end{bmatrix} = \frac{1}{R_p^*} \begin{bmatrix} \mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} & -T_p^* z_n \\ T_p^* z_n^{-1} & \dfrac{-T_p^* \mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi}}{T_p} \end{bmatrix} \begin{bmatrix} a_{n,0} \\ b_{n,0} \end{bmatrix}, \tag{C 6}$$

where

$$\phi = \frac{1}{\hbar} S(E), \quad S(E) = \oint_{H=E} p(x)\, \mathrm{d}x, \tag{C 7}$$

so that $\tilde{g}_n$ is given by

$$\tilde{g}_n = \frac{1}{R_p^*} \begin{bmatrix} \dfrac{1}{T_x^*} \mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} - \dfrac{R_x^* T_p^*}{T_x^*} z_n^{-1} & \dfrac{R_x^* T_p^*}{T_x^* T_p} \mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - \dfrac{T_p^*}{T_x^*} z_n \\ \dfrac{-R_x}{T_x} \mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} + \dfrac{T_p^*}{T_x} z_n^{-1} & \dfrac{-T_p^*}{T_x T_p} \mathrm{e}^{-\frac{1}{2}\phi} + \dfrac{R_x T_p^*}{T_x} z_n \end{bmatrix}. \tag{C 8}$$

This result generalizes that given by Azbel (1964*a*, *b*). It applies to the coupling of W.K.B. solutions that are supported by closed phase trajectories and satisfy the Bloch condition (C 2).

To illustrate the significance of (C 8), it will be used to rederive the renormalization group transformation in the semiclassical limit $\hbar \to 0$, by using the formulae for the connection coefficients derived in Appendix A

$$\begin{aligned} &T_x \simeq t_x, \quad T_p \simeq \mathrm{i}t_p, \\ &R_x \simeq -\mathrm{i}, \quad R_p \simeq 1, \\ &t_x = \exp\left(-\frac{1}{\hbar}\int_{x_1}^{x_2} |p(x)|\,\mathrm{d}x\right), \quad t_p = \exp\left(-\frac{1}{\hbar}\int_{p_1}^{p_2} |x(p)|\,\mathrm{d}p\right), \end{aligned} \tag{C 9}$$

so that $\tilde{g}_n$ is approximated by

$$\tilde{g}_n = \frac{1}{t_x}\begin{bmatrix} \mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} - t_p z_n^{-1} & -\mathrm{i}(\mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_n) \\ \mathrm{i}(\mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} - t_p z_n^{-1}) & \mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_n \end{bmatrix}, \tag{C 10}$$

which is the general formula for the transfer matrix in the semiclassical limit. Because of the approximations $|R_x| = |R_p| = 1$, (C 10) no longer satisfies the condition $\det \tilde{g}_n = 1$, but this does not matter for the subsequent analysis presented here.

From (C 10), a second order difference equation will be found, which the amplitudes $\alpha_n = a_{n,0}$ satisfy. Given

$$\tilde{g}_n = \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix}, \quad \det \tilde{g}_n = 1, \tag{C 11}$$

the amplitudes α_n satisfy the equation

$$B_{n-1}\alpha_{n+1} + B_n \alpha_{n-1} = (A_n B_{n-1} + B_n D_{n-1})\,\alpha_n. \tag{C 12}$$

For the present purposes only the terms of lowest order in the transmission coefficients will be obtained; using the form (C 10) for $\tilde{g}_n$ gives

$$\begin{aligned} &(\mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_{n-1})\,\alpha_{n+1} + (\mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_n)\,\alpha_{n-1} \\ &\quad = \frac{1}{t_x}[(\mathrm{e}^{\frac{1}{2}\mathrm{i}\phi} - t_p z_n^{-1})(\mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_{n-1}) + (\mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_{n-1})(\mathrm{e}^{-\frac{1}{2}\mathrm{i}\phi} - t_p z_n)]\,\alpha_n. \end{aligned} \tag{C 13}$$

Making the substitutions

$$\phi = 2\pi(m + \tfrac{1}{2}) + \chi, \quad \frac{1}{\hbar}S(E_m) = 2\pi(m + \tfrac{1}{2}), \quad \chi = \frac{1}{\hbar}\left(\frac{\partial S}{\partial E}\right)_{E_m}\Delta E \tag{C 14}$$

and retaining only lowest order terms,

$$(-1)^{m+1}(\alpha_{n+1} + \alpha_{n-1}) + (-1)^{m+1} t_p / t_x^{-1}(z_n + z_n^{-1})\,\alpha_n = \frac{1}{\hbar}\left(\frac{\partial S}{\partial E}\right)_{E_m}\Delta E\,\alpha_n. \tag{C 15}$$

Now, by writing

$$z_n = \mathrm{e}^{4\pi^2 \mathrm{i} n/\hbar} = \mathrm{e}^{2\pi \mathrm{i}\beta_1 n}, \quad \beta_1 = 1/\beta - [1/\beta], \tag{C 16}$$

where $[x]$ means 'interger part of x', (C 15) reads

$$C_p^{(m)}(\alpha_{n+1} + \alpha_{n-1}) + 2C_x^{(m)} \cos(2\pi\beta_1 n)\,\alpha_n = \Delta E \alpha_n, \tag{C 17}$$

where

$$C_p^{(m)} = (-1)^{m+1}\hbar t_p/(\partial S/\partial E)_{E_m}, \quad C_x^{(m)} = (-1)^{m+1}\hbar t_x/(\partial S/\partial E)_{E_m}. \qquad (C\,18)$$

These equations (C 17), (C 18) are equivalent to (4.9); the amplitudes α_n describe the amplitudes of Bohr-quantized quasi-eigenstates centred on $x = 2\pi n$, and (C 17) is the Hamiltonian equation describing the interaction of these states.

One final point concerns the application of the Bloch condition (3.19) in the p-direction, which expresses the fact that the superposition of W.K.B. solutions corresponds to the solution of a difference equation. In (C 1), (C 2) a different Bloch wavevector is applied to each row of closed orbits along the p-direction. This is because the approximate wavefunctions are confined to different regions of the x-axis, separated by classically forbidden regions, and they are therefore treated separately, with different phases Δ_n. This approximation will break down for states at the separatrix where the semiclassical W.K.B. solutions are not well separated. The exact, global solution of the difference equation (1.1) will, of course, have a single Bloch wavevector associated with it in the p-representation, in just the same way as has (3.16). This fact is of little use in analysing (1.1) to determine the form of its solutions, however.

Appendix D

This appendix gives a simplified discussion of the fractal dimension of the spectrum of (1.1) when $\alpha = 1$, for almost all β.

The discussion of the r.g. transformation in § 5(ii) predicts that, when $\alpha = 1$, the spectrum is a fractal Cantor set (see Mandelbrot 1982), provided the coefficients n_N of the continued fraction expansion of β form a statistically stationary sequence.

It is natural, therefore, to try to characterize the spectrum by its fractal dimension, especially as this would, by the discussion in § 5(iv), be expected to be a universal quantity for all Hamiltonians with given symmetry, for almost all β.

For the purposes of this paper the fractal dimension, D_f, will be defined by:

$$D_f = \lim_{\rho \to 0} -\frac{\ln \mathcal{N}(\rho)}{\ln \rho}, \qquad (D\,1)$$

where $\mathcal{N}(\rho)$ is the minimum number of intervals of length ρ required to cover the set. This definition of dimension (due to L. Pontrjagrin and L. Schnirelman) is easier to use than the definition preferred by Mandelbrot (1982) (due to F. Hausdorff and A. S. Besicovitch), and in most cases gives the same answer.

Consider the following simplified model of the spectrum. Assume that the spectrum can be generated recursively as follows. Take an interval and split it into n_1 equal subintervals of length $\rho(n_1)$. Split each of these subintervals into n_2 equal subintervals of length $\rho(n_2)$... etc. The numbers n_i are taken to be the coefficients of the continued fraction expansion of β.

In the semiclassical limit the fraction of the width of the ν_Nth sub-band occupied

by the ν_{n+1}th sub-band is given by the energy scaling factor for the r.g. transformation (4.1),

$$C_x^{(\nu_N)} = \exp\left[-\frac{1}{\hbar_N}\int_{x_1}^{x_2}|p(x)|\,\mathrm{d}x\right]\hbar_N/(\partial S/\partial E)_{E^{(\nu_N)}}, \tag{D 2}$$

which depends on the sub-band considered as well as on the value of $\hbar_N$. The fact that all the sub-bands are of different width is very difficult to deal with, and is ignored in this present discussion. The function $\rho(n_N)$ used in this model only takes account of the variation of (D 2) with $\hbar_N$;

$$\rho(n_N) = A\hbar_N\,\mathrm{e}^{-B/\hbar_N}, \tag{D 3}$$

so that, using the fact that $\hbar_N \simeq 2\pi/n_N$, $\rho(n)$ is taken as

$$\rho(n_N) = \frac{A'}{n_N}\exp(-B'n_N) \tag{D 4}$$

in the semiclassical limit, where A' and B' are constants.

Now consider how to calculate the fractal dimension of this set. After N stages of the recursion generating the set, there are

$$\mathcal{N}_N = \prod_{i=1}^{N} n_i \tag{D 5}$$

intervals, each of length

$$\rho_N = \prod_{i=1}^{N}\rho(n_i), \tag{D 6}$$

and it is obvious that the fractal set can be covered by $\mathcal{N}_N$ intervals of length ρ_N (in fact this is usually the most economical covering, but this fact will not be needed). Thus

$$D_{\mathrm{f}} \leqslant \lim_{N\to\infty} -\sum_{i=1}^{N}\ln n_i \Big/ \sum_{i=1}^{N}\ln\rho(n_i), \tag{D 7}$$

or

$$D_{\mathrm{f}} \leqslant \langle\ln n\rangle/\langle\ln\rho(n)\rangle, \tag{D 8}$$

where the angle brackets denote an average over the distribution $P(n)$ of the n_is. Now for almost all β, $P(n)$ takes the universal form (4.24), and using the limiting form (4.25) for large n, and the semiclassical result (D 4) for $\rho(n)$, it is easily found that the denominator in (D 8) diverges to $-\infty$. The numerator, however, has the value $\ln 2.686$. So (D 8) gives

$$D_{\mathrm{f}} = 0 \tag{D 9}$$

for almost all β.

The use of the semiclassical form for $\rho(n)$ is justified because the semiclassical results are asymptotic for large n (small $\hbar$), and this is just where the divergence of the mean of $\ln\rho(n)$ comes from. The result (D 9) is found to be unchanged when the different sizes of the sub-bands formed at each stage of the r.g. recursion are taken into account; the calculation is lengthy, however, because of the difficulty of finding the asymptotic behaviour of $\mathcal{N}(\rho)$; some of the results of this calculation have an independent interest and the author hopes to publish them separately.

Appendix E

The appendix gives details of the divergence of α from its fixed point, $\alpha = 1$, and estimates the critical exponent of this divergence in the semiclassical limit.

The r.g. recursion relation for α is given by

$$\alpha_{N+1} = \frac{t_p^{(\nu_N)}}{t_x^{(\nu_N)}} = \frac{\exp\left[\frac{-1}{\hbar_N}\Sigma_p^{(\nu_N)}\right]}{\exp\left[\frac{-1}{\hbar_N}\Sigma_x^{(\nu_N)}\right]} = \exp\left[\frac{-1}{\hbar_N}(\Sigma_p^{(\nu_N)} - \Sigma_x^{(\nu_N)})\right], \tag{E 1}$$

where the action integrals $\Sigma_x^{(\nu_N)}, \Sigma_p^{(\nu_N)}$ over the classically forbidden regions of the x- and p-axes are found through

$$\left.\begin{aligned} \Sigma_x = \int_{x_1}^{x_2} |p(x)|\,\mathrm{d}x, \quad \Sigma_p = \int_{p_1}^{p_2} |x(p)|\,\mathrm{d}p, \\ E = H_N = 2\cos p + 2\alpha_N \cos x, \\ p(x_1) = p(x_2) = x(p_1) = x(p_2) = 0. \end{aligned}\right\} \tag{E 2}$$

So Σ_x and Σ_p are given by

$$\left.\begin{aligned} \Sigma_x &= \int_{x_1}^{2\pi - x_1} \operatorname{arc\,cosh}\left[\tfrac{1}{2}E - (1+\delta_N)\cos u\right] du, \\ \Sigma_p &= \int_{p_1}^{2\pi - p_1} \operatorname{arc\,cosh}\left[\tfrac{1}{2}E - (1+\delta_N)^{-1}\cos u\right] du, \\ x_1 &= \arccos\left[(\tfrac{1}{2}E - 1)(1+\delta_N)^{-1}\right], \quad p_1 = \arccos\left[(\tfrac{1}{2}E - 1)(1+\delta_N)\right] \end{aligned}\right\} \tag{E 3}$$

and

$$\alpha_N = 1 + \delta_N$$

provided that, at the energy considered, the phase trajectories are closed.

When δ is small, its mapping under the r.g. transformation is given by

$$\left.\begin{aligned} \alpha_{N+1} = 1 + \delta_{N+1} &= \exp\left[\frac{-1}{\hbar_N}\left(\frac{\partial \Sigma_p}{\partial \delta} - \frac{\partial \Sigma_x}{\partial \delta}\right)\delta_N\right] \\ &\simeq 1 + \frac{1}{\hbar_N}\left(\frac{\partial \Sigma_p}{\partial \delta} - \frac{\partial \Sigma_x}{\partial \delta}\right)\Big|_{\delta=0} \delta_N. \end{aligned}\right\} \tag{E 4}$$

Therefore the ratio δ_{N+1}/δ_N is given by

$$\delta_{N+1}/\delta_N = \frac{1}{\hbar_N}\sigma(E), \tag{E 5}$$

where

$$\sigma(E) = \frac{\partial}{\partial \delta}(\Sigma_p - \Sigma_x)|_{\delta=0} = \int_{x^*}^{2\pi - x^*} \frac{-2\cos u\,\mathrm{d}u}{[(\tfrac{1}{2}E - \cos u)^2 - 1]^{\frac{1}{2}}}, \tag{E 6}$$

$$x^* = \arccos(\tfrac{1}{2}E - 1).$$

When E is small, it is easily shown that

$$\sigma(E) \simeq 2\pi, \quad E \ll 1, \tag{E 7}$$

and since, by symmetry, $\sigma(E)$ varies quadratically at $E = 0$, (E 7) is a fair approximation to $\sigma(E)$ for most of the energy range.

When $\beta = \hbar/2\pi$ is a quadratic irrational number and there are fixed points of the r.g. transformation, then

$$\kappa_{N+1} = \frac{\delta_{N+1}}{\delta_N} = \frac{1}{2\pi\beta}\sigma(E) \tag{E 8}$$

is a critical exponent of the r.g. transformation, describing the instability ($\kappa > 1$) of a fixed point in the semiclassical limit, $\beta \ll 1$.

Appendix F

This appendix gives a brief explanation of how the figures showing the spectra and eigenstates were prepared.

Even though this paper is primarily concerned with the properties of (1.1), and similar equations, when the ratio β of the periodicities is irrational, all the figures show solutions calculated for rational values of β. The exact spectra for an irrational β cannot, of course, be calculated since they are pathological sets, but the continuous spectrum for a nearby rational value of β will show the first stages of the hierarchical properties of the Cantor set spectrum.

The spectra of (1.1) were computed as follows. If β is a rational number, $\beta = p/q$, then (1.1) can be solved in Bloch form

$$e^{ik}a_{n+1} + e^{-ik}a_{n-1} + 2\alpha\cos(2\pi pn/q + \Delta)a_n = E a_n, \quad \psi_n = e^{ikn}a_n, \tag{F 1}$$

where the a_n form a periodic sequence of period q; the a_n and E are periodic functions of period $2\pi/q$ in both k and Δ. The plots of the spectra show the range between the maximum and minimum values of E in the (k, Δ) plane, found by finding the eigenvalues E of the $q \times q$ Hermitian matrix equation equivalent to (F 1). The other equations considered were solved in the same way.

References

Arnold, V. I. 1980 *Chapitres Supplémentaires de la Théorie des Equations Différentielles Ordinaires*. Moscow: MIR.

Aubry, S. & André, G. 1979 *Ann. Israel phys. Soc.* **3**, 133–164.

Azbel, M. Ya. 1964*a* *Zh. éksp. teor. Fiz.* **46**, 929 (*transl. Soviet Phys. JETP* **19**, 634–645).

Azbel, M. Ya. 1964*b* *Dokl. Akad. Nauk SSSR* **159**, 703 (*transl. Soviet Math. Dokl.*, **5**, 1549–1552).

Azbel, M. Ya. 1979 *Phys. Rev. Lett.* **43**, 1954–1957.

Bellisard, J. & Simon, B. 1982 *J. Funct. Anal.* **48**, 408–419.

Berry, M. V. 1983 Submitted to *Proc. R. Soc. Lond.*

Berry, M. V. & Mount, K. E. 1972 *Rep. Prog. Phys.* **35**, 315–397.

Berry, M. V. & Wilkinson, M. 1984 *Proc. R. Soc. Lond.* (In the press.)

Claro, F. & Wannier, G. H. 1979 *Phys. Rev.* B **19**, 6068–6074.

Craig, W. 1983 *Commun. Math. Phys.* **88**, 113–131.

Dinaburg, E. I. & Sinai, Ya. G. 1976 *Funct. Anal. Applications* **9**, 279–289.

Gordon, Ya. 1976 *Usp. math. Nauk.* **31**, 257–258.
Groenewold, H. J. 1946 *Physica* **12**, 405–460.
Harper, P. G. 1955 *Proc. phys. Soc.* A **68**, 874–892.
Heading, J. 1962 *An introduction to phase integral methods.* London: Methuen.
Hofstadter, D. R. 1976 *Phys. Rev.* B **14**, 2239–2249.
Kac, M. 1959 *Statistical independence in probability, analysis, and number theory*, Carus Mathematical Monographs no. 12. Mathematical Association of America.
Khinchin, A. Ya. 1964 *Continued Fractions.* University of Chicago Press.
Kohmoto, M., Kadanoff, L. P. & Teng, C. 1983 *Phys. Rev. Lett.* **50**, 1870–1872.
Landau, L. D. & Lifshitz, I. M. 1958 *Quantum Mechanics*, ch. 7. Oxford: Pergamon.
Mandelbrot, B. B. 1982 *The fractal geometry of Nature.* San Francisco: Freeman.
Ostlund, S. & Pandit, R. 1983 (Submitted.)
Ostlund, S., Pandit, R., Rand, D., Schellnhuber, H. J. & Siggia, E. D. 1983 *Phys. Rev. Lett.* **50**, 1873–1876.
Ozorio de Almeida, A. M. & Hannay, J. H. 1982 *Ann. Phys.* **138**, 115–154.
Peierls, R. 1933 *Z. Phys.* **80**, 763–791.
Rauh, A. 1974 *Phys. Status. Solidi.* B **65**, K131–K135.
Rauh, A. 1975 *Phys. Status. Solidi.* B **69**, K9–K13.
Schellnhuber, H. J., Obermair, G. M. & Rauh, A. 1981 *Phys. Rev.* B**23**, 5191–5202.
Simon, B. 1982 *Adv. appl. Maths* **3**, 463–490.
Sokoloff, J. B. 1981 *Phys. Rev.* B **23**, 2039–2041.
Suslov, I. M. 1982 *Soviet Phys. JETP* **56**, 612–617.
Thouless, D. J. 1972 *J. Phys.* C **5**, 77–81.
Thouless, D. J. & Niu, Q. 1983 *J. Phys.* A **16**, 1911–1919.
Thouless, D. J. 1983 (Submitted.)
Von Neumann, J. & Wigner, E. P. 1929 *Phys. Z.* **30**, 467–470.
Wegner, F. 1976 *Z. Phys.* B **25**, 327–337.