WKB analysis of energy band structure of modulated systems

1991 J. Phys. A: Math. Gen. 24 4999

(http://iopscience.iop.org/0305-4470/24/21/014)

J. Phys. A: Math. Gen. 24 (1991) 4999–5010. Printed in the UK

WKB analysis of energy band structure of modulated systems

G I Watson
Department of Theoretical Physics, University of Oxford, Oxford, OX1 3NP, UK

Received 17 April 1991, in final form 23 July 1991

Abstract. A discrete WKB approximation is applied to Harper's equation describing the motion of an electron in a square lattice subject to a magnetic field. The treatment is valid when the modulation parameter (proportional to the magnetic field) is a rational number M/N, in the limit of large N. A uniform asymptotic expression is derived for the Floquet wavenumber which characterizes the dynamics and spectral properties. The case of energies near the classical separatrix, crucial to the properties of the spectrum near the transition between localized and extended states, is included. An application is a detailed calculation of the total width of the bands in the critical region, for $M = 1$ and 2. The results are consistent with Thouless' conjecture that the bandwidth obeys a universal scaling form, independent of M.

1. Introduction

The discrete eigenvalue equation

$$a_{n+1} + e_n a_n + a_{n-1} = \omega a_n \tag{1.1}$$

has been studied in connection with a variety of physical systems characterized by the presence of two independent length scales (for a review, see Sokoloff 1985). Here the sequence $\{a_n : n \text{ integer}\}$ represents the wavefunction of the system, ω is the energy eigenvalue, and the modulation e_n is a periodic or almost periodic function of its index. Systems of interest include a tight-binding model of electrons in a lattice with a uniform applied magnetic field (Harper 1955, Hofstadter 1976), a modulated spring model of the lattice dynamics of incommensurate crystals (Currat and Janssen 1988), and a model of spin fluctuations in longitudinally modulated magnetic materials (Lovesey 1988a). A further example occurs in a mean-field treatment of anyons on a square lattice (Hasegawa *et al* 1990).

The most widely studied case is a sinusoidal modulation,

$$e_n = 2\gamma \cos(2\pi\phi n + \Delta) \tag{1.2}$$

in which case the difference equation (1.1) is known as Harper's equation or the discrete Mathieu equation. This form is appropriate for the electron problem, where γ is unity for an isotropic lattice, ϕ is proportional to the magnetic field, and the wavefunction $\{a_n\}$ is the amplitude of a Wannier function centred at the lattice site n. The sites are labelled along an axis perpendicular to the field, whose direction depends on the choice of gauge for the vector potential.

0305-4470/91/214999+12$03.50

If the modulation parameter ϕ is a rational number, M/N, the system is periodic, and is exactly soluble by a generalization of the Bloch treatment usually presented for crystal electrons (Lovesey 1988a,b, Lovesey and Westhead 1990a,b, Lovesey *et al* 1991). The eigenvalues ω, allowing bounded solutions, form N bands. Dynamic response functions, densities of states and band edges may be written in closed form for any given periodicity. Modulations incommensurate with the crystal period, represented by irrational values of ϕ, may be studied using the large-N limit, since an irrational may be approximated arbitrarily well by a rational with sufficiently large denominator. Unfortunately, the exact solutions increase rapidly in complexity with N, and are of little practical use in the incommensurate limit. Qualitative features of the energy spectrum have been successfully studied by numerical techniques (for example, Hofstadter 1976) but knowledge of the structure of the eigenfunctions, crucial to the dynamics of these systems, is still incomplete.

In this paper we apply a WKB technique to the Harper equation, valid in the limit of small magnetic fields. The treatment complements work along similar lines by Harper (1955), Azbel' (1964), Wilkinson (1984a,b) and Wang *et al* (1987) by providing uniform asymptotic expressions for the eigenfunctions and the Floquet wavenumber which characterizes the dynamics and spectral properties. The results become exact in the large-N limit. The band edges in the energy spectrum are obtained as roots of a transcendental equation. An application is the investigation of scaling behaviour of the total bandwidth near the critical point $\gamma = 1$: a result of Thouless (1990) is derived by more direct means and extended to $M = 2$.

2. General features of the eigenfunctions

We assume throughout that $\phi = M/N$, where M and N are co-prime integers. Since the eigenvalue equation (1.1) is linear, a general solution may be expressed in terms of two linearly independent polynomial solutions p_n and q_n, chosen to satisfy $p_0 = q_1 = 0$ and $p_1 = q_0 = 1$. All quantities of interest may be expressed in terms of these functions (Lovesey *et al* 1991), in particular the density of states, normalised to unity,

$$Z(\omega) = \begin{cases} (1/2\pi N)|\theta_N'(\omega)|(4-\theta_N^2)^{-1/2} & |\theta_N| \leqslant 2 \\ 0 & |\theta_N| > 2 \end{cases} \tag{2.1}$$

where

$$\theta_N = p_{N+1} + q_N \tag{2.2}$$

and the prime indicates the derivative with respect to ω. The band edges, for a particular value of the phase Δ, are thus the zeros of the polynomial $\theta_N^2 - 4$. For a given ω in a band, $|\theta_N| \leqslant 2$, there exist solutions of (1.1) obeying Floquet's theorem in the form

$$a_{n+N} = \mathrm{e}^{\mathrm{i}Nk} a_n \qquad \text{where} \qquad 2\cos(Nk) = \theta_N \tag{2.3}$$

and k is the Floquet wavenumber. Roots of $\theta_N = 2$ correspond to symmetric Floquet solutions, $k = 0$, while $\theta_N = -2$ yields antisymmetric solutions, $k = \pi/N$.

In the electron problem, the phase Δ is the y-component of the Bloch wavevector, and most quantities of interest are integrals over Δ. The full energy spectrum is therefore the union of eigenvalues for all values of Δ. The duality property of Harper's equation (Aubry and André 1980) implies that θ_N depends on Δ only through the term $-2\gamma^N \cos(N\Delta)$, and hence the band edges correspond to extrema of $2\cos(Nk) + 2\gamma^N \cos(N\Delta)$, namely

$$\begin{aligned} &\text{Symmetric roots:} && \theta_N(\Delta = 0) = 2 \\ &\text{Antisymmetric roots:} && \theta_N(\Delta = 0) = -2(1 + 2\gamma^N). \end{aligned} \tag{2.4}$$

Henceforth we set $\Delta = 0$ accordingly. The spectrum is confined to the interval $-2(1+\gamma) \leqslant \omega \leqslant 2(1+\gamma)$, is symmetrical about $\omega = 0$, and is invariant under the duality transformation $\gamma \to 1/\gamma$, $\omega \to \omega/\gamma$. A change of sign of γ may be absorbed in the phase Δ and has no effect on the spectrum. Hence no generality is lost in taking $0 \leqslant \gamma \leqslant 1$ and $\omega \geqslant 0$. For reasons explained later we also exclude the case $\gamma \approx 0$.

3. WKB solution

If the wavelength $2\pi/\phi$ is large, the sinusoidal modulation varies slowly compared with the wavefunction, and one may apply a continuum approximation to (1.1). Thus we are interested in the limit $\phi \to 0$, or equivalently $N \to \infty$ with fixed M. In physical terms, we suppose the radius of the electron's cyclotron orbit to be large compared with the lattice spacing. A similar treatment applicable when ϕ is close to a non-zero rational has been suggested by Sokoloff (1981); that topic is not pursued here.

The starting point for the analysis is a discrete form of the physical optics approximation of WKB theory (Bender and Orszag 1978),

$$a_n = \exp\left[NS_0(x) + S_1(x)\right] \tag{3.1}$$

where $x = n/N$ contains the order dependence. The functions S_0 and S_1, to be determined, are the leading order terms in a formal WKB expansion in powers of the small parameter $1/N$, which plays here the same role as $\hbar$ in semiclassical approximations in quantum mechanics. Functions of integer index will be written as functions of the continuous variable x; thus a_N and $a(x)$ will be used interchangeably. With this notation, $a(x)$ satisfies

$$a(x + 1/N) + a(x - 1/N) = 2\cos\left[p(x)\right] a(x) \tag{3.2}$$

where

$$p(x) = \cos^{-1}\left[\omega/2 - \gamma\cos(2\pi M x)\right] \tag{3.3}$$

is analogous to the classical momentum. Substituting (3.1) into this equation and equating powers of N, we find

$$\begin{aligned} &\cosh S_0' = \cos p \\ &2S_1' \sinh S_0' + S_0'' \cosh S_0' = 0 \end{aligned} \tag{3.4}$$

in which the primes denote differentiation with respect to x. Solving for S_0 and S_1 generates two linearly independent WKB solutions,

$$a(x) = [\sin p(x)]^{-1/2} \exp\left[\pm i N \int^x p(t)\,dt\right] \tag{3.5}$$

which form the basis for the subsequent analysis.

Near the classical turning points where $p(x) = 0$, the WKB approximation breaks down since the local wavelength $2\pi/Np$ is comparable with the modulation wavelength. Away from these regions, (3.5) yields the leading term in the asymptotic expansion of a_n as $N \to \infty$. Higher-order corrections, generated by extending the formal WKB expansion (3.1), are of order $1/N$.

When $\omega < 2(1-\gamma)$ there are no turning points and (3.5) is a uniform approximation for all x in a finite interval. From (2.3),

$$\theta_N = 2\cos\left[N\int_0^1 p(t)\,dt\right]. \tag{3.6}$$

Hence, to leading order, all positive ω less than $2(1-\gamma)$ are eigenvalues. The wavefunctions are oscillatory extended states, corresponding to classical phase trajectories open in the x-direction (Wilkinson 1984a).

Turning points arise for $\omega > 2(1-\gamma)$, corresponding to the existence of closed orbits in classical phase space. In this case, treated in section 3.1, WKB solutions in adjoining regions are matched using a local solution at the turning point. When $\omega \approx \pm 2(1-\gamma)$ or $\pm 2(1+\gamma)$, quadratic turning points occur, as described in section 3.2. In section 3.3 the various approximations are united into a single formula, valid uniformly in ω and γ.

3.1. First-order turning points

If $\omega > 2(1-\gamma)$ there are $2M$ turning points,

$$\frac{\beta}{M},\ \frac{1-\beta}{M},\ \frac{1+\beta}{M},\ \frac{2-\beta}{M},\ \ldots,\ \frac{M-1+\beta}{M},\ \frac{M-\beta}{M}$$

in the unit interval, where

$$\beta = (2\pi)^{-1}\cos^{-1}[(\omega/2-1)/\gamma]. \tag{3.7}$$

Consider initially the interval $0 \leqslant x \leqslant 1/M$, which contains two turning points. Defining

$$\begin{aligned}\eta(x) &= 2\phi^{-1}\int_x^{\beta} \cos^{-1}[\omega/2 - \gamma\cos(2\pi t)]\,dt \\ \sigma(x) &= 2\phi^{-1}\int_x^{1/2} \cosh^{-1}[\omega/2 - \gamma\cos(2\pi t)]\,dt\end{aligned} \tag{3.8}$$

the wavefunction for $x < \beta/M$ is proportional to $\exp[\pm i\eta(Mx)/2]$, while in the classically forbidden region $\beta/M < x < (1-\beta)/M$ it contains the exponential

factors $\exp[\pm\sigma(Mx)/2]$. The matching procedure is analogous to the case of a continuous variable: the solutions near the turning point are standard Airy functions,

$$\mathrm{Ai}[\alpha^{1/3}(x-\beta/M)] \qquad \text{and} \qquad \mathrm{Bi}[\alpha^{1/3}(x-\beta/M)] \tag{3.9}$$

in which $\alpha = 4\pi M N^2 \sin(2\pi\beta)$ is the local slope of the potential. A phase factor $\pi/4$ appears when exponential and oscillatory solutions are linked. A similar analysis applies at the second turning point, yielding the WKB solution

$$a(x) = \begin{cases} (\sin p)^{-1/2} \mathrm{e}^{-\sigma/2} \cos[\eta(Mx)/2 + \pi/4] & 0 \leqslant x < \beta/M \\ (\sinh \tilde{p})^{-1/2} \exp[-\sigma(Mx)/2] & \beta/M < x < (1-\beta)/M \\ 2(\sin p)^{-1/2} \mathrm{e}^{\sigma/2} \cos[\eta(1-Mx)/2 - \pi/4] & (1-\beta)/M < x \leqslant 1/M \end{cases} \tag{3.10}$$

where $\sigma = \sigma(\beta)$ and

$$\tilde{p}(x) = \cosh^{-1}[\omega/2 - \gamma\cos(2\pi M x)] \tag{3.11}$$

in the forbidden region. This compound solution, supplemented by Airy functions of the type (3.9) near the turning points, is uniformly asymptotic to a solution of (1.1) over the interval $0 \leqslant x \leqslant 1/M$. A second linearly independent solution is $a(1/M - x)$.

We are primarily interested in the change in $a(x)$ over a unit interval, requiring a chain of M such composite solutions. In the allowed region containing $x = j/M$, we write

$$a(x) = A_j a(x - j/M) + B_j a((j+1)/M - x). \tag{3.12}$$

Consistency between adjacent regions demands that the coefficients satisfy the linear equations

$$\begin{pmatrix} A_{j+1} \\ B_{j+1} \end{pmatrix} = \begin{pmatrix} 2\mathrm{e}^{\sigma}\cos\eta & -\sin\eta \\ \sin\eta & \frac{1}{2}\mathrm{e}^{-\sigma}\cos\eta \end{pmatrix} \begin{pmatrix} A_j \\ B_j \end{pmatrix} \equiv T \begin{pmatrix} A_j \\ B_j \end{pmatrix} \tag{3.13}$$

where $\eta = \eta(0)$. This completes the solution for the eigenvectors, which are generated from (3.12) by successive powers of the matrix T.

The function θ_N is readily evaluated from (2.3), yielding

$$\theta_N = \mathrm{Tr}\, T^M. \tag{3.14}$$

Since $\det T = 1$, the solution is conveniently expressed in terms of the Floquet wavenumber,

$$2\cos(k/\phi) = \mathrm{Tr}\, T = (2\mathrm{e}^{\sigma} + \tfrac{1}{2}\mathrm{e}^{-\sigma})\cos\eta. \tag{3.15}$$

The dominant term in θ_N is $(2\mathrm{e}^{\sigma}\cos\eta)^M$, so the bands are centred roughly at the zeros of $\cos\eta$ and their width, proportional to $\mathrm{e}^{-M\sigma}$, decreases very rapidly with ω and N. In summary, to a first approximation the positive energy bands fill the region $0 < \omega < 2(1-\gamma)$ and are very narrow for $\omega > 2(1-\gamma)$. Consequently the total width of the bands approaches $4(1-\gamma)$ in the incommensurate limit, as discovered empirically by Aubry and André (1980). A detailed calculation of the bandwidth is given in section 5.

3.2. Second-order turning points

The assumptions leading to (3.10) are not valid when the slope α is small, indicating quadratic behaviour of the potential near the turning point. The precise criterion for failure of the linear turning point solution is that the fractional change in p over one wavelength $2\pi/Np$, at the extrema of $\cos(2\pi Mx)$, be of order unity, and hence $|\omega \pm 2(1 \pm \gamma)| = O(1/N)$. An equivalent condition is that the wavelength is comparable to the distance between the turning points. The cases $\omega \approx \pm 2(1+\gamma)$ are of little interest as they correspond to the extreme edges of the spectrum, where the bands have negligible width. The regions $|\omega \pm 2(1-\gamma)| = O(1/N)$ however, although vanishingly small as $N \to \infty$, dominate the total bandwidth. They correspond to phase trajectories near the classical separatrix.

We apply local analysis at the extrema of $\cos(2\pi Mx)$. In a neighbourhood of the maximum $x = j/M$, the wavefunction satisfies the parabolic cylinder equation,

$$\frac{d^2b}{d\xi^2} + (\tfrac{1}{4}\xi^2 - \alpha_1)\, b(\xi) = 0 \tag{3.16}$$

where $b_n = (-1)^n a_n$ is written as a function of the local variable $\xi = (4\pi MN)^{1/2}\gamma^{1/4}(x - j/M)$. In the notation of Abramowitz and Stegun (1972), the solution is a linear combination of $W(\alpha_1, \xi)$ and $W(\alpha_1, -\xi)$; similarly at the minima of the cosine the local solutions are $W(\alpha_2, \pm\xi)$, where

$$\alpha_1 = -\frac{[\omega + 2(1-\gamma)]}{4\pi\phi\sqrt{\gamma}} \qquad \text{and} \qquad \alpha_2 = \frac{[\omega - 2(1-\gamma)]}{4\pi\phi\sqrt{\gamma}}. \tag{3.17}$$

Using the asymptotic form for parabolic cylinder functions, the local solutions may be matched with the WKB form in the intermediate regions, which are free of turning points. Writing

$$(-1)^n a_n = C_j W(\alpha_1, \xi) + D_j W(\alpha_1, -\xi) \tag{3.18}$$

the matrix transforming the coefficients is given by the product

$$\begin{pmatrix} 1 & 0 \\ 0 & e^{\kappa_1} \end{pmatrix} R[\tfrac{1}{2}\chi - \pi\phi'(j+1)] \begin{pmatrix} e^{\kappa_2} & 0 \\ 0 & e^{-\kappa_2} \end{pmatrix} R[\tfrac{1}{2}\chi + \pi\phi' j] \begin{pmatrix} e^{-\kappa_1} & 0 \\ 0 & 1 \end{pmatrix} \tag{3.19}$$

in which ϕ' is the fractional part of $1/\phi$, $R[\theta]$ is the 2×2 rotation matrix, and

$$\sinh \kappa_i = e^{\pi\alpha_i} \qquad i = 1, 2 \tag{3.20}$$

$$\chi = \eta + \alpha_1 \log|\alpha_1| - \alpha_1 - \arg\Gamma(\tfrac{1}{2} + i\alpha_1) - \alpha_2 \log|\alpha_2| + \alpha_2 + \arg\Gamma(\tfrac{1}{2} + i\alpha_2)\,. \tag{3.21}$$

The dependence of the transformation matrix on the index j, arising from matching the WKB phase with the $(-1)^n$ factor in (3.18), is a considerable complication. Its significance in terms of band clustering is discussed in section 4.

The Floquet wavenumber once again satisfies a trace condition, which may be written

$$\theta_N = \mathrm{Tr}(W_{M-1} U W_{M-2} U \ldots W_0 U) \tag{3.22}$$

where

$$W_j = \begin{pmatrix} -\sinh\kappa_1\cos(2\pi\phi' j) + \cosh\kappa_1 & -\sinh\kappa_1\sin(2\pi\phi' j) \\ -\sinh\kappa_1\sin(2\pi\phi' j) & \sinh\kappa_1\cos(2\pi\phi' j) + \cosh\kappa_1 \end{pmatrix} \tag{3.23}$$

and

$$U = \begin{pmatrix} \cosh\kappa_2\cos\chi + \sinh\kappa_2 & -\cosh\kappa_2\sin\chi \\ \cosh\kappa_2\sin\chi & \cosh\kappa_2\cos\chi - \sinh\kappa_2 \end{pmatrix} \tag{3.24}$$

The first four cases are

$$\begin{aligned}
M = 1 \qquad & \theta_N = 2\cosh\kappa_1\cosh\kappa_2\cos\chi - 2\sinh\kappa_1\sinh\kappa_2 \\
M = 2 \qquad & \theta_N = 2\cosh^2\kappa_1\cosh^2\kappa_2\cos 2\chi - 2\sinh^2\kappa_1\cosh^2\kappa_2 + 2\sinh^2\kappa_2 \\
M = 3 \qquad & \theta_N = 2\cosh^3\kappa_1\cosh^3\kappa_2\cos 3\chi - 2\sinh^3\kappa_1\sinh^3\kappa_2 \\
& \quad + 6\cosh\kappa_1\cosh\kappa_2[\sinh^2\kappa_1\cos(\chi + 2\pi\phi') \\
& \quad + \sinh^2\kappa_2\cos\chi] \\
M = 4 \qquad & \theta_N = 2\cosh^4\kappa_1\cosh^4\kappa_2\cos 4\chi - 4\sinh^4\kappa_1 - 2\cosh^4\kappa_1\cosh^4\kappa_2 \\
& \quad + 2 + 16\cosh^2\kappa_1\cosh^2\kappa_2\cos\chi \\
& \quad \times [\sinh^2\kappa_1\cos(\chi + 2\pi\phi') + \sinh^2\kappa_2\cos\chi].
\end{aligned} \tag{3.25}$$

Results for larger M are readily generated but are somewhat unwieldy.

3.3. Uniform approximation

In this section we regard θ_N as a function of the scaled energy variable α_2, defined in (3.17), representing the deviation from the breakdown region $\omega \approx 2(1-\gamma)$. The asymptotic expressions (3.6) and (3.15) may be combined in the form

$$2\cos(k/\phi) = 2e^{\sigma}\cos\eta \qquad |\alpha_2| \gg 1 \tag{3.26}$$

where $\beta = 1/2$ for $\omega < 2(1-\gamma)$. The third case, (3.22), is valid for $|\alpha_2| \ll N$. Furthermore, we have the asymptotic identities

$$\begin{aligned}
&\chi = \eta \qquad && |\alpha_2| \gg 1 \\
&\sigma = \pi\alpha_2 \qquad && |\alpha_2| \ll N.
\end{aligned} \tag{3.27}$$

In the overlap region $1 \ll |\alpha_2| \ll N$, we have $\sigma \gg 1$ and so the various expressions for θ_N do indeed agree asymptotically. Equation (3.22) encompasses both definitions, neglecting sub-dominant terms, if we take κ_2 to be defined by

$$\sinh\kappa_2 = \begin{cases} e^{\sigma} & \alpha_2 \geqslant 0 \\ e^{\pi\alpha_2} & \alpha_2 < 0. \end{cases} \tag{3.28}$$

With this modification, (3.22), including the particular instances (3.25), is a uniform asymptotic approximation to θ_N over the entire range $0 \leqslant \omega \leqslant 2(1+\gamma) - \epsilon$, for any ϵ.

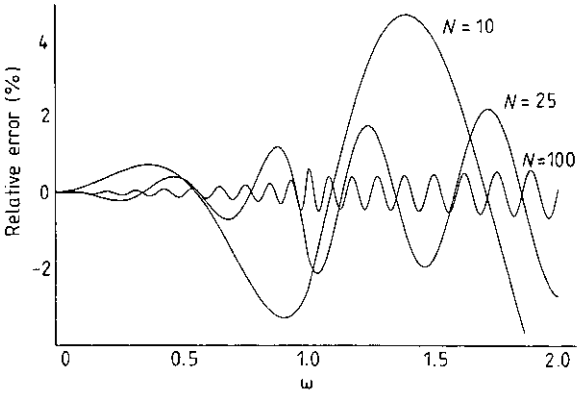

Figure 1. The relative error $(2e^{\sigma})^{-M}\Delta\theta_N$ is plotted as a function of the energy ω, for $\gamma = 0.5$, $M = 1$ and $N = 10, 25, 100$.

The relative error of the asymptotic expression for θ_N is plotted in figure 1 as a function of ω for selected values of N. The quantity $(2e^{\sigma})^{-M}\Delta\theta_N$, where $\Delta\theta_N$ is the absolute error, is used as a measure of relative error which avoids divergence at the zeros of θ_N. Exact values of θ_N were calculated by direct iteration of the eigenvalue equation (1.1). In accordance with the remarks above, the error is seen to be uniformly bounded over the interval of interest and to decrease as $1/N$ for large N. Encouragingly, the error is tolerable even for modest values of N.

4. Hierarchical clustering

Clustering rules were first constructed empirically by Hofstadter (1976) based on the peculiar self-similar nature of the 'butterfly' diagram, a plot of the spectrum as a function of the modulation parameter ϕ. These rules state that for a given ϕ, the spectrum consists of a central band and two side bands, each further split into sub-bands according to transformed values of ϕ, namely

$$\phi' = \begin{cases} \{1/\phi\} & \text{for the side bands} \\ \{\phi/(1-2\phi)\} & \text{for the centre band} \end{cases} \tag{4.1}$$

where $\{\ldots\}$ denotes the fractional part. An analytical derivation of the rules, using perturbation theory, was given by Stinchcombe and Bell (1987).

Some features of this clustered structure are present in the asymptotic solutions obtained in section 3. In the large-α_2 region, bands are concentrated at the points where $\cos\eta \approx 0$, or more precisely where the local energy coordinate

$$\omega' = e^{\sigma}\cos\eta \tag{4.2}$$

is small. It follows from (3.26) that θ_N behaves locally like $2\cos(M\omega')$ or $2\sin(M\omega')$, which describes a cluster of M bands separated by large gaps from neighbouring clusters. These are precisely the side bands of (4.1), associated with modulation parameter $\phi' = N/M$. Secondary clustering effects, from $\{1/\{1/\phi\}\}$ for instance, do not occur at this level of approximation.

By contrast, the quadratic turning point solution (3.22), in the breakdown region $\omega \approx 2(1-\gamma)$, does display these higher-order effects. Too see this, note that θ_N is the trace of a product of M matrices, each containing factors of the form $\cos(2\pi\phi' j)$. The condition that ω lies in a band is equivalent to a boundedness restriction on the solutions of an M-periodic difference equation. Explicitly, if

$$\begin{pmatrix} c_{j+1} \\ d_{j+1} \end{pmatrix} = W_j U \begin{pmatrix} c_j \\ d_j \end{pmatrix} = \begin{pmatrix} A_j^{11} & A_j^{12} \\ A_j^{21} & A_j^{22} \end{pmatrix} \begin{pmatrix} c_j \\ d_j \end{pmatrix} \tag{4.3}$$

where $\det(W_j U) = 1$, then

$$A_j^{12} c_{j-1} - (A_j^{11} A_{j-1}^{12} + A_j^{12} A_{j-1}^{22}) c_j + A_{j-1}^{12} c_{j+1} = 0. \tag{4.4}$$

The coefficients in this equation are modulated according to the transformed parameter $\phi' = \{1/\phi\}$. Although (4.4) is different from (1.1), in that all three terms are modulated non-sinusoidally, it is reasonable to expect further splitting to accord with (4.1); indeed Stinchcombe and Bell (1987) have argued on general grounds that the same rules apply universally to a generic class of almost periodic difference equations, including (4.4). We conclude that all orders of the clustering hierarchy are manifested in the asymptotic solution in the breakdown region.

A similar scheme has been used by Wilkinson (1984a) to construct an approximate renormalisation group transformation for Harper's equation.

5. Total bandwidth

The behaviour of the spectrum near the self-dual point $\gamma = 1$ is of considerable interest, because the eigenfunctions undergo a transition from localized to extended behaviour there (Aubry and André 1980). As an application of the uniform asymptotic solution, we calculate the total measure of the energy spectrum in this critical region. From (2.4) this bandwidth, W, is equal to the total width of the N intervals in which

$$-2(1+2\gamma^N) \leqslant \theta_N \leqslant 2. \tag{5.1}$$

As remarked earlier, the widths of individual bands decreases extremely rapidly with ω, so that negligible measure is contained in the region $\alpha_2 \gg 1$; this statement may be justified rigorously. In the large-N limit, all significant width is concentrated in the region $\omega \approx 0$, in which (3.22) is valid.

Restricting attention for the moment to $M = 1$ and introducing the variables

$$s = \frac{1}{2}\pi(\alpha_1 + \alpha_2) = -\frac{(1-\gamma)}{2\phi\sqrt{\gamma}} \qquad u = \frac{1}{2}\pi(\alpha_2 - \alpha_1) = \frac{\omega}{4\phi\sqrt{\gamma}} \tag{5.2}$$

we have, for large N,

$$\theta_N = 4\mathrm{e}^s(\cosh^2 s + \sinh^2 u)^{1/2} \cos\chi - 2\mathrm{e}^{2s}. \tag{5.3}$$

If γ is sufficiently close to the self-dual point so that $s = \mathrm{O}(1)$, then $\gamma^N = \mathrm{e}^{2Ms}$ and the band edges satisfy

$$(\cosh^2 s + \sinh^2 u)^{1/2} \cos\chi = \pm \cosh s. \tag{5.4}$$

Subsequent analysis depends on the fact that χ varies very rapidly compared with u: this follows from the small α_2 expansion,

$$\eta = \eta_0 + 2\alpha_2[\log(\pi\phi/4) - 1] + (\alpha_1 + \alpha_2)\log|\alpha_1 + \alpha_2| - \alpha_1 \log|\alpha_1| + \alpha_2 \log|\alpha_2| \tag{5.5}$$

which implies that

$$\frac{\mathrm{d}\chi}{\mathrm{d}u} = 2\log(\pi\phi/4) + \Re\left[\psi\left(\tfrac{1}{2} + \mathrm{i}\alpha_1\right) + \psi\left(\tfrac{1}{2} + \mathrm{i}\alpha_2\right)\right] \tag{5.6}$$

is of order $\log N$. It follows that to leading order, u is constant over a given band. Zeros of $\cos\chi$ fall within the bands; if $\delta\chi$ represents the deviation from such a zero, the width of a single band is

$$\delta\omega = \frac{\mathrm{d}\omega}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}\chi}\delta\chi = \frac{8}{N}\frac{\mathrm{d}u}{\mathrm{d}\chi}\tan^{-1}\left(\frac{\cosh s}{\sinh u}\right) \tag{5.7}$$

neglecting quantities of order $(N\log N)^{-1}$. The total bandwidth, including bands at negative energies, is then a sum over all the zeros, which may be approximated by an integral,

$$W = 2\sum\delta\omega = \frac{16}{\pi N}\int_0^\infty \tan^{-1}\left(\frac{\cosh s}{\sinh u}\right)\mathrm{d}u \qquad 1 - \gamma = \mathrm{O}(1/N). \tag{5.8}$$

This formula describes completely the bandwidth in the vicinity of the self-dual point. After some algebra, it may be written in the alternative form

$$W = 4(1-\gamma) + 32(\pi N)^{-1}\int_0^{\exp[-N(1-\gamma)/2]} t^{-1}\tan^{-1} t\,\mathrm{d}t \tag{5.9}$$

in agreement with Thouless (1990). At the self-dual point, $s = 0$,

$$W = \frac{32}{\pi N}\,(0.915\,965\,59\ldots) \qquad M = 1 \tag{5.10}$$

where the number in parentheses is Catalan's constant (Abramowitz and Stegun 1972). Considering terms neglected in the derivation, higher-order corrections to the bandwidth are likely to be of order $(N\log N)^{-1}$.

The analysis for $M = 2$ follows similar lines, yielding

$$W = (32/\pi N)\int_0^\infty \left\{\cos^{-1}\left[\tfrac{1}{2}\mathrm{e}^u(\cosh^2 s + \sinh^2 u)^{-1/2}\right] - \sin^{-1}\left[\tfrac{1}{2}\mathrm{e}^{-u}(\cosh^2 s + \sinh^2 u)^{-1/2}\right]\right\}\mathrm{d}u. \tag{5.11}$$

This integral may be reduced to standard algebraic integrals by differentiation with respect to s, followed by the substitution $y = \mathrm{e}^{2u}$; the bandwidth is then obtained by integrating with respect to s. As the algebra is rather tedious we simply state the result: W is given by the same formula, (5.9), as found for $M = 1$, reducing at the

self-dual point to (5.10) involving Catalan's constant. Thus the bandwidth near the self-dual point has precisely the same large-N behaviour in both cases $M = 1$ and 2.

Numerical evidence supports Thouless' (1990) conjecture that the bandwidth formula (5.10) is universal, holding for all M even or odd. We have calculated band edges and widths using standard numerical diagonalisation routines applied to the $N \times N$ matrix corresponding to (1.1). The asymptotic expressions of section 3 provide a convenient check on the accuracy of the results. In all cases except $M = 1$, NW converged extremely rapidly to its limiting value, (5.10). For $M = 1$, the convergence was much slower, consistent with logarithmic approach to the limit as remarked above. The same limit is obtained even when M and N increase together in such a way that M/N approaches a non-zero limit, rational or irrational.

The results of this section suggest a stronger universality, namely that the bandwidth tends to the large-N limit (5.9), independent of M, in the entire critical region. To our knowledge, no analytical proof of this universality exists. Unfortunately, the above derivation does not readily extend to higher values of M due to the complicated algebraic form of θ_N, (3.25). The case where M/N approaches a non-zero limit is outside the scope of the present analysis.

6. Conclusions

We have applied a fresh approach, based on a discrete WKB approximation, to Harper's equation describing the motion of an electron in a square lattice subject to a uniform magnetic field. The treatment is valid when the parameter ϕ, proportional to the field, is a rational number M/N, and provides a uniform asymptotic approximation in the large-N limit. We cover the case, (3.22), omitted by Wilkinson (1984a), of trajectories near the classical separatrix. Although this corresponds to a very narrow energy range, it contains the dominant contribution to the total measure of the spectrum, W, in the critical region near the transition between localized and extended wavefunctions. Our detailed calculation of this bandwidth confirms, using a more transparent approach, the result, (5.9), of Thouless (1990) for $M = 1$. We have extended the calculation to $M = 2$, showing that W approaches a limiting form identical to that for $M = 1$, throughout the critical region. The conjecture that the large-N limit is independent of M, confirmed by numerical work, remains open.

Acknowledgments

I would like to thank Professor Stephen Lovesey for his invaluable guidance throughout the course of this work, and for helpful suggestions on the manuscript. This research was supported by a Hackett Studentship from the University of Western Australia.

References

Abramowitz M and Stegun I A 1972 *Handbook of Mathematical Functions* (New York: Dover)
Aubry S and André G 1980 *Ann. Israel Phys. Soc.* vol 3 (Bristol: Adam Hilger) p 133
Azbel' M Ya 1964 *Zh. Eksp. Teor. Fiz.* **46** 939 (Engl. transl. 1964 *Sov. Phys.–JETP* **19** 634)

Bender C M and Orszag S A 1978 *Advanced Mathematical Methods for Scientists and Engineers* (New York: McGraw-Hill)
Currat R and Janssen T 1988 *Solid State Phys.* **41** 201
Harper P G 1955 *Proc. Phys. Soc.* A **68** 874
Hasegawa Y, Hatsugai Y, Kohmoto M and Montambaux G 1990 *Phys. Rev.* B **41** 9174
Hofstadter D R 1976 *Phys. Rev.* B **14** 2239
Lovesey S W 1988a *J. Phys. C: Solid State Phys.* **21** 2805
—— 1988b *J. Phys. C: Solid State Phys.* **21** 4967
Lovesey S W and Westhead D R 1990a *J. Phys.: Condens. Matter* **2** 7407
—— 1990b *J. Phys.: Condens. Matter* **2** 7425
Lovesey S W, Watson G I and Westhead D R 1991 *Int. J. Mod. Phys.* B **5** 1
Sokoloff J B 1981 *Phys. Rev.* B **23** 2039
—— 1985 *Phys. Rep.* **126** 189
Stinchcombe R B and Bell S C 1987 *J. Phys. A: Math. Gen.* **20** L739
Thouless D J 1983 *Phys. Rev.* B **28** 4272
—— 1990 *Commun. Math. Phys.* **127** 187
Wang Y Y, Pannetier B and Rammal R 1987 *J. Physique* **48** 2067
Wilkinson M 1984a *Proc. R. Soc.* A **391** 305
—— 1984b *J. Phys. A: Math. Gen.* **17** 3459