Critical and bicritical properties of Harper's equation with next-nearest-neighbor coupling

J. H. Han and D. J. Thouless
Department of Physics, FM-15, University of Washington, Seattle, Washington 98195

H. Hiramoto
College of Humanities and Sciences, Nihon University, Sakurajoshui, Setagaya-ku, Tokyo 156, Japan

M. Kohmoto
Institute for Solid State Physics, University of Tokyo, 7-22-1 Roppongi, Minato-ku, Tokyo 106, Japan
(Received 25 May 1994)

We have exploited a variety of techniques to study the universality and stability of the scaling properties of Harper's equation, the equation for a particle moving on a tight-binding square lattice in the presence of a gauge field, when coupling to next-nearest sites is added. We find, from numerical and analytical studies, that the scaling behavior of the total width of the spectrum and the multifractal nature of the spectrum are unchanged, provided the next-nearest-neighbor coupling terms are below a certain threshold value. The full square symmetry of the Hamiltonian is not required for criticality, but the square diagonals should remain as reflection lines. A bicritical line is found at the boundary between the region in which the nearest-neighbor terms dominate and the region in which the next-nearest-neighbor terms dominate. On the bicritical line a different critical exponent for the width of the spectrum and different multifractal behavior are found. In the region in which the next-nearest-neighbor terms dominate, the behavior is still critical if the Hamiltonian is invariant under reflection in the directions parallel to the sides of the square, but a new length scale enters, and the behavior is no longer universal but shows strongly oscillatory behavior. For a flux per unit cell equal to $1/q$ the measure of the spectrum is proportional to $1/q$ in this region, but if it is a ratio of Fibonacci numbers the measure decreases with a rather higher inverse power of the denominator.

I. INTRODUCTION

Harper's equation can be derived as the equation for an electron in a strong two-dimensional periodic potential and a weak magnetic field, or for an electron in a strong magnetic field and a weak periodic potential. The Hamiltonian can be written in the form

$$H(p_x, x) = 2t_a \cos p_x + 2t_b \cos x \,, \tag{1.1}$$

where

$$p_x = -2\pi i \phi d/dx \tag{1.2}$$

is the variable conjugate to x. Azbel[1] showed that in the case $t_a = t_b$, which corresponds to a periodic potential with square symmetry, the spectrum forms a "devil's staircase" for irrational values of ϕ, and Hofstadter[2] generated computer drawings of the spectrum for rational values of ϕ, and discussed the self-similarity and scaling of the spectrum. Work by Aubry and André[3] exploited the symmetry of the problem under canonical transformations that interchange x and p_x, and showed that, for irrational values of ϕ, there is a localization length in the x direction, independent of energy,

$$L = 1/\ln(t_b/t_a) \quad \text{for} \quad t_b > t_a, \tag{1.3}$$

which diverges at this symmetry point $t_a = t_b$. Also the sum of the widths of all the energy bands, the measure of the spectrum, has the form

$$W = 4|t_b - t_a| \,, \tag{1.4}$$

which vanishes at the same point. These properties have suggested that this point is like a critical point of the system, that $|1 - t_a/t_b|$ represents the distance from the critical point, and that, where $\phi = p/q$ is a rational, the denominator q acts as a finite size in finite size scaling theory.[4]

Various methods have been used to study the Harper equation. There are some rigorous analytical results on the Lyapunov exponent (reciprocal of the localization length),[5] and on the measure of the spectrum (sum of the bandwidths).[6–11] Aubry duality gives information about the localization length.

If $H(p_x, x)$ defined in Eq. (1.1) is treated as a classical Hamiltonian there is an obvious interpretation of the critical point, since for $t_a > t_b$ the energy contours surrounding the minima of H are separated from the contours surrounding the maxima of H by a region of orbits open in the x direction, while for $t_a < t_b$ there are orbits open in the p_x direction. For the symmetric case $t_a = t_b$ there is a single energy $E = 0$ separating the two types of closed orbits. A more subtle semiclassical analysis,

based on scaling theory, must be introduced to explain why the behavior is not just singular at the energy of this separatrix, but is singular at all energies in the spectrum.

Much of the information we have about this problem comes from numerical studies of the spectrum for rational values of ϕ similar to the one carried out by Hofstadter. The vanishing of the measure of the spectrum at the critical point gives a one-parameter indicator of the critical point. Numerical studies on the Harper equation indicate that this measure is given by Eq. (1.4) for all irrational values of ϕ, and that for rational values $\phi = p/q$ the measure scales as[4]

$$W \sim (t_b - t_a) g[q(1 - t_a/t_b)] , \tag{1.5}$$

where the scaling function $g(s)$ behaves as $9.3299/s$ when its argument is small, and tends to ± 4 for large $|s|$. Only the corrections to scaling seem to be sensitive to the value of p.[8,12] The distribution of bandwidths within this spectrum gives a more detailed picture of the spectrum, but this is sensitive to the value of ϕ. The simplest forms might be expected if ϕ is an irrational solution of a quadratic equation, so that its continued fraction expansion repeats itself, or if it is a rational approximant obtained by truncating such a repeating continued fraction. The golden mean, which is the limit of the ratios of successive Fibonacci numbers, is the simplest case of this sort. In this case the spectrum has a multifractal structure that is self-similar at the critical point.[2,13–15]

Another method that has been used to get a number of important results takes its simplest form when ϕ is small, so that Eq. (1.1) can be treated semiclassically. It is then a second order difference equation with a slowly varying central term, for which the WKB equation can be used. For $t_b \approx t_a$ only those bands close to $E = 0$ have appreciable bandwidth, and the others all have widths that vanish exponentially as $1/\phi$ gets large, since the widths are determined by tunneling between successive minima or maxima of $2t_b \cos x$. The case $\phi = 1/q$ is particularly simple, and this has been used to derive an analytic form for the function $g(s)$ of Eq. (1.5).[6,8,9]

The critical properties of the Harper equation have been explored in great detail by various combinations of these methods. However, it is not clear to what extent the elegant properties of the Harper equation are special properties of that equation, and to what extent they are robust, and survive perturbations of the model. On the basis of the small ϕ behavior, Suslov[16] has argued that modifications of the x dependence of Eq. (1.1) which maintain the periodicity will lead to an energy dependence of the critical value of t_b. This behavior is supported by numerical calculations.[17–19] On the other hand, Helffer and Sjöstrand[20] have argued that variants of Eq. (1.1) which preserve the invariance under the canonical transformation $p_x \to -x$, $x \to p_x$ are critical at all energies. It is the purpose of this work to take these questions further by exploring in some detail a simple generalization of the Harper equation.

The equation we explore is the generalization of Eq. (1.1) which takes the form

$$H(p_x, x) = 2t_a \cos p_x + 2t_b \cos x + 2t_{a\bar{b}} \cos(p_x - x) + 2t_{ab} \cos(p_x + x) . \tag{1.6}$$

This could represent a tight-binding model in which electrons can tunnel to next-nearest neighbors as well as to nearest neighbors on a rectangular lattice, and, for $t_{ab} = 0$, $t_a = t_b = t_{a\bar{b}}$, it can represent an electron on a triangular lattice.[21] This form of the Hamiltonian is very convenient for numerical work, as it remains in the form of a three-term difference equation, and there have been a number of earlier studies of it.[4,21,22] It has enough parameters that we can study the effects of different features of the Hamiltonian such as the breaking of the symmetry, and the change in the nature of the constant energy contours. For $t_a = t_b$ and $t_{a\bar{b}} = t_{ab}$ the system has square symmetry, but there are two quite different regimes with this symmetry, according to whether t_a or t_{ab} is dominant. We show that there is actually an interesting bicritical point that separates the two different regimes.

In this paper, we have exploited most of the techniques mentioned above. In Sec. II, we discuss the main features of the energy contours of Eq. (1.6) and discuss the symmetries of the system. In Sec. III, we discuss the form of the characteristic equation whose roots give the eigenvalues, repeat some earlier results which were obtained by the use of Aubry duality, discuss the extension of finite size scaling arguments to this case, and extend the sum rule of Last and Wilkinson[10] to this case. In Sec. IV, we present an analysis in terms of multifractals for the case that ϕ is a ratio of neighboring Fibonacci numbers. In Sec. V, we give the results of calculations of the measure of the spectrum when ϕ is the ratio of Fibonacci numbers and also when it is a fraction such as $1/q$ or $p/(p^2+1)$ that represents a slow modulation of the diagonal term of the difference equation. These two cases complement one another, since for the ratio of Fibonacci numbers the spectrum is spread over a large number of bands, even close to the critical point, whereas for small values of ϕ the only bands with significant measure are close to zero energy. We find that the behavior is relatively simple in the region in which $t_a = t_b$ is dominant, but there is a bicritical region not only at $t_a = t_b = 2t_{a\bar{b}} = 2t_{ab}$, but also for $t_a = 2t_{a\bar{b}} = 2t_{ab} > t_b$. For $t_{a\bar{b}} = t_{ab}$ dominant the situation appears to be much more complicated, with some important oscillatory terms which confuse the analysis of numerical results.

In Sec. VI, we do what we can to explain the results we have from numerical analysis in terms of WKB theory and scaling theory. In some cases our understanding is reasonably complete, but in other cases we can do little more than explain why the problem is complicated. There is a concluding discussion in Sec. VII.

II. CLASSICAL ORBITS AND SYMMETRY

In this discussion we assume all the coefficients t_a, t_b, $t_{a\bar{b}}$, and t_{ab} are positive. For the case $t_a = t_b$, $t_{a\bar{b}} = t_{ab}$ the spectrum of the Hamiltonian given by Eq. (1.6) is invariant under the eight operations of the symmetry group of the square. The four proper rotations are generated

by $p_x \to -x$, $x \to p_x$, while the time reversal operation $p_x \to -p_x$ generates the improper rotations. The Hamiltonian is also invariant under the group of translations in phase space corresponding to a square lattice. The classical Hamiltonian has a maximum at $p_x = 0 = x$, where its value is $4(t_a + t_{ab})$, and at equivalent lattice points. For $t_a > 2t_{ab}$ it has a minimum at $p_x = \pi = x$, where its value is $-4(t_a - t_{ab})$. There are two saddle points in each unit cell, at $p_x = 0$, $x = \pi$, and at $p_x = \pi$, $x = 0$, where its value is $-4t_{ab}$. At this value of the energy there is a contour given by

$$\cos p_x = -\frac{2t_{ab} + t_a \cos x}{t_a + 2t_{ab}\cos x} , \tag{2.1}$$

which threads the system, separating contours that surround minima from those that surround maxima. For $t_a < 2t_{ab}$ the maximum is unchanged, but the points at $p_x = \pi = x$ become subsidiary maxima, while the points at $p_x = 0$, $x = \pi$, and at $p_x = \pi$, $x = 0$ become minima. Four more saddle points appear at $\cos p_x = -t_a/2t_{ab}$, $\cos x = -t_a/2t_{ab}$, where the energy is $-t_a^2/t_{ab}$. The contour joining the saddle points and separating contours surrounding minima from those surrounding maxima is now

$$\cos p_x = -t_a/2t_{ab} \quad \text{or} \quad \cos x = -t_a/2t_{ab} . \tag{2.2}$$

In this paper, we pay particular attention to the bicritical point $t_a = t_b = 2t_{a\bar{b}} = 2t_{ab}$, where all contours surround maxima except for the lines where the energy has its minimum value $-2t_a$. At the points $p_x = \pi = x$ the lowest nonvanishing partial derivatives of the energy are the fourth derivatives.

For the case $t_a = t_b$, $t_{a\bar{b}} > t_{ab}$ the symmetry operation $p_x \leftrightarrow x$, a reflection symmetry in phase space, remains, as well as rotation by π. There is still a maximum at $p_x = 0 = x$, where the energy is $4t_a + 2t_{a\bar{b}} + 2t_{ab}$, and at equivalent lattice points. For $t_a > t_{a\bar{b}}$ there is a minimum at $p_x = \pi = x$, where its value is $-4t_a + 2t_{a\bar{b}} + 2t_{ab}$. For $t_a^2 > 4t_{a\bar{b}}t_{ab}$ there are two saddle points in each unit cell, at $p_x = 0$, $x = \pi$ and at $p_x = \pi$, $x = 0$, where its value is $-2(t_{a\bar{b}} + t_{ab})$. At this value of the energy there is a contour given by

$$2t_{ab}\cos[\tfrac{1}{2}(p_x + x)] = -\cos[\tfrac{1}{2}(p_x - x)] \times \left[t_a \pm \sqrt{t_a^2 - 4t_{a\bar{b}}t_{ab}}\right] , \tag{2.3}$$

which threads the system, provided $t_a > t_{a\bar{b}} + t_{ab}$. For $t_a < t_{a\bar{b}} + t_{ab}$ there is a range of energies in the neighborhood of $-2(t_{a\bar{b}} + t_{ab})$, with $\cos[\frac{1}{2}(p_x - x)]$ near ± 1, for which there are no values of $(p_x + x)$ that satisfy Eq. (2.3). In this case there are open orbits in the direction of constant $(p_x - x)$.

A special case of this sort is $t_a = t_b = t_{a\bar{b}}$, $t_{ab} = 0$, which has triangular symmetry, and is equivalent to the case worked out numerically by Claro and Wannier.[21]

For $t_a \neq t_b$, $t_{a\bar{b}} = t_{ab}$ the Hamiltonian is invariant under $p_x \to -p_x$. For $t_{a\bar{b}} = t_{ab} > t_b > t_a$ there are maxima at $p_x = 0 = x$ and at $p_x = \pi = x$, where the energy has the values $4t_{ab} \pm (t_a + t_b)$, and minima at $p_x = 0$, $x = \pi$ and at $p_x = \pi$, $x = 0$, where the energies are $-4t_{ab} \pm (t_a - t_b)$. There are saddle points given by

$$\cos p_x = -t_b/2t_{ab}, \quad \cos x = -t_a/2t_{ab} , \tag{2.4}$$

and the energy has the value $-t_a t_b/t_{ab}$ on the lines on which either of these conditions is satisfied. For $t_b > 2t_{a\bar{b}} = 2t_{ab} > t_a$ the four saddle points given by Eq. (2.4) disappear, and the new saddle points are at $p_x = \pi$, $x = 0$ and at $p_x = \pi = x$, where the energies are $-t_a \pm (2t_{ab} - t_b)$. There are orbits open in the p_x direction between these energies.

III. CHARACTERISTIC EQUATION AND AUBRY DUALITY

From Eq. (1.2) it can be seen that the operator $2\cos p_x$ is a displacement operator that displaces the coordinate by $2\pi\phi$. The eigenvalue problem for the Hamiltonian (1.6) takes the form of a set of finite difference problems

$$(t_a + t_{a\bar{b}}e^{2\pi i\phi(n-\frac{1}{2})+ik_2} + t_{ab}e^{-2\pi i\phi(n-\frac{1}{2})-ik_2})a_{n-1} + 2t_b\cos(2\pi\phi n + k_2)a_n + (t_a + t_{a\bar{b}}e^{-2\pi i\phi(n+\frac{1}{2})-ik_2} + t_{ab}e^{2\pi\phi(n+\frac{1}{2})+ik_2})a_{n+1} = Ea_n , \tag{3.1}$$

with the variable parameter k_2 determined by the initial value of x. When $\phi = p/q$ is rational this equation is periodic with period q, and solutions of the Floquet form

$$a_n = c_n e^{ik_1 n} , \tag{3.2}$$

with c_n periodic, can be found. This then yields a finite matrix problem, with the matrix tridiagonal apart from the top right and bottom left corners, for which the characteristic polynomial has as its only k_1 dependent term

$$(-1)^{q-1}e^{ik_1 q}\prod_{n=0}^{q-1}\left(t_a + t_{a\bar{b}}e^{-\frac{2\pi ip}{q}(n+\frac{1}{2})-ik_2} + t_{ab}e^{\frac{2\pi ip}{q}(n+\frac{1}{2})+ik_2}\right) + \text{c.c.} \tag{3.3}$$

This product can be expressed in terms of a Chebyshev polynomial T_q (see Appendix A), and this gives

$$(-1)^{q-1}4t^q\cos(qk_1)T_q\left(\frac{t_a}{2t}\right)$$
$$+(-1)^p2\{t_{a\bar{b}}^q\cos[q(k_1-k_2)]$$
$$+t_{ab}^q\cos[q(k_1+k_2)]\}\,, \tag{3.4}$$

where

$$t^2=t_{a\bar{b}}t_{ab}\,. \tag{3.5}$$

Since the whole spectrum is, from Eq. (1.6), clearly invariant under the interchange of p_x, x and t_a, t_b, there must also be a similar k_2 dependent term in the characteristic polynomial, so the characteristic polynomial can be written in the form

$$P(E)=P_0(E)-(-1)^q4t^q\left\{\cos(qk_1)T_q\left(\frac{t_a}{2t}\right)\right.$$
$$\left.+\cos(qk_2)T_q\left(\frac{t_b}{2t}\right)\right\}$$
$$+(-1)^p2\{t_{a\bar{b}}^q\cos[q(k_1-k_2)]$$
$$+t_{ab}^q\cos[q(k_1+k_2)]\}\,, \tag{3.6}$$

where $P_0(E)$ is independent of k_1, k_2. The energy bands are determined by the variation of the solutions of the characteristic equation as k_1, k_2 are varied.

Some simple analysis can show which of the terms in this expression will dominate in the limit of large q. For $x>1$ we have

$$[T_q(x)]^{1/q}\approx x+\sqrt{x^2-1}\,, \tag{3.7}$$

and so, for $t_b\geq t_a$, $t_{a\bar{b}}>t_{ab}$, the dominant term in Eq. (3.6) is of order

$$[\tfrac{1}{2}(t_b+\sqrt{t_b^2-4t^2})]^q \tag{3.8}$$

for $t_b>t_{a\bar{b}}+t_{ab}$, and is of order $t_{a\bar{b}}^q$ for $t_b<t_{a\bar{b}}+t_{ab}$. This transition from a regime where the bandwidths are dominated by the term depending on k_2 to a regime where the bandwidths are dominated by a term depending on k_1-k_2 occurs at the same values of the parameters as the changes in the nature of the classical orbits which we discussed in connection with Eqs. (2.3) and (2.4). Under these conditions only one term in Eq. (3.6) is relevant for large q, or two terms if $t_a=t_b$. Since this expression is independent of the value of the energy, one should expect the critical values of the parameters to be energy independent, whereas the classical orbits only give information about the behavior at the singular value of the energy.

For $2t_{a\bar{b}}=2t_{ab}>t_b\geq t_a$ the situation is very different, since the Chebyshev polynomials are now of order unity, and all four terms in Eq. (3.6) would seem to be marginal for large q. In particular one should expect the bandwidths to depend on

$$\cos\left[q\arccos\left(\frac{t_a}{2t_{ab}}\right)\right]\text{ and }\cos\left[q\arccos\left(\frac{t_b}{2t_{ab}}\right)\right], \tag{3.9}$$

so the bandwidths should display nearly periodic behavior in q. In fact we found such behavior in the numerical studies reported in this paper before we had realized that they ought to be found.

In an earlier work[4] it was shown how the argument of Aubry and André can be adapted to this situation. There are three parts to this argument. First, they state that the element of the Green function connecting the two ends of a tridiagonal matrix can be expressed as the product of the next-to-diagonal matrix elements divided by the characteristic polynomial. The product of off-diagonal matrix elements is just the coefficient of $\cos p_x$ in Eq. (3.6). Second, they compare this expression with the expression for the dual problem obtained by interchanging p_x and x. The characteristic polynomial is unchanged, and the ratio of the products of next-to-diagonal elements can be used to generate an expression for the difference between the Lyapunov exponents in the x and p_x directions, in the form

$$\lambda_x-\lambda_p=\lim_{q\to\infty}\frac{1}{q}\ln\left[\frac{4t^q|T_q(\frac{t_b}{2t})|+2t_{a\bar{b}}^q+2t_{ab}^q}{4t^q|T_q(\frac{t_a}{2t})|+2t_{a\bar{b}}^q+2t_{ab}^q}\right]. \tag{3.10}$$

This is zero for $t_{a\bar{b}}+t_{ab}\geq t_b\geq t_a$. For $t_b\geq t_a\geq t_{a\bar{b}}+t_{ab}$ it gives

$$\lambda_x-\lambda_p=\ln\left[\frac{\frac{1}{2}t_b+\frac{1}{2}\sqrt{t_b^2-4t^2}}{\frac{1}{2}t_a+\frac{1}{2}\sqrt{t_a^2-4t^2}}\right], \tag{3.11}$$

and for $t_b\geq t_{a\bar{b}}+t_{ab}\geq t_a$, $t_{a\bar{b}}\geq t_{ab}$ it gives

$$\lambda_x-\lambda_p=\ln\left[\frac{\frac{1}{2}t_b+\frac{1}{2}\sqrt{t_b^2-4t^2}}{t_{a\bar{b}}}\right]. \tag{3.12}$$

The third part of the argument, which we find to be rather more subtle, says that if λ_x is positive then λ_p must be zero, so Eqs. (3.11) and (3.12) are actually equations for λ_x rather than for $\lambda_x-\lambda_p$. Eigenstates for the Hamiltonian (1.6) in p_x space can be found from eigenstates in x space by Fourier transformation. These eigenstates in x space have their support on a lattice of points, so their Fourier transforms are periodic. If they are localized in space their Fourier transforms are smooth periodic functions. Functions of this sort corresponding to the same value of the energy cannot be superposed to give localization in p_x space as well as localization in x space.

These results show that states are localized in x space, independent of energy, for generic irrational ϕ provided t_b is greater than both t_a and $t_{a\bar{b}}+t_{ab}$. This condition, now independent of energy, is the same as the condition for the existence of open orbits extended in the p_x direction given in the discussion of Eq. (2.4).

The finite size scaling argument that leads to Eq. (1.5) can be generalized to deal with the critical properties of Eq. (1.6). For $t_b > t_a > t_{a\bar{b}} + t_{ab}$ the width for irrational ϕ is still given[4] by Eq. (1.4), but the scaling length is now given by Eq. (3.11) so that we have

$$W \sim (t_b - t_a) g\left(q \ln \left[\frac{\frac{1}{2}t_b + \frac{1}{2}\sqrt{t_b^2 - 4t^2}}{\frac{1}{2}t_a + \frac{1}{2}\sqrt{t_a^2 - 4t^2}} \right] \right)$$
$$\approx (t_b - t_a) g\left(q \frac{t_b - t_a}{\sqrt{t_b^2 - 4t^2}} \right) . \tag{3.13}$$

In particular, this results in the prediction that the measure of the spectrum should scale as

$$W \sim \frac{9.3299}{q} \sqrt{t_b^2 - 4t^2} \tag{3.14}$$

for $t_a = t_b \geq t_{a\bar{b}} + t_{ab}$. A special case of this is the triangular lattice with $t_a = t_b = t_{a\bar{b}}$, $t_{ab} = 0$, where the measure of the spectrum is the same as it is for the square lattice.

For the case $t_b > 2t_{a\bar{b}} = 2t_{ab} \geq t_a$ the measure of the spectrum for ϕ irrational is $4t_b - 8t_{ab}$ and the scaling length is given by Eq. (3.12), so the use of the same argument would lead to a finite size scaling expression of the form

$$W \sim (t_b - 2t) G\left(q\sqrt{t_b - 2t_{ab}} \frac{\sqrt{t_b + 2t_{ab}} + \sqrt{t_b - 2t_{ab}}}{2t_{ab}} \right) . \tag{3.15}$$

The scaling function $G(s)$ must diverge as s^{-2} at the origin, so that at the point $t_b = 2t_{a\bar{b}} = 2t_{ab}$ the measure of the spectrum is finite, and goes to zero like q^{-2} for large q. However, this argument does not take account of the second length scale introduced in Eq. (3.9), so we should expect the function G in Eq. (3.15), and the coefficient of the limiting q^{-2} behavior, to depend on t_a according to the form given in Eq. (3.9), which is periodic or nearly periodic in q.

The argument of Last and Wilkinson,[10] which provides a lower bound to the spectrum for the critical case, can be generalized to deal with the situations we consider in this work. The simplest case is given by $t_b > t_a \geq t_{a\bar{b}} + t_{ab}$, where the result of Avron, Mouche, and Simon[7] that the intersection spectrum (the intersection over k_2 of the spectra for fixed k_2) has measure $4(t_b - t_a)$ remains valid, as is shown in Appendix B. The intersection spectrum is defined, as can be seen from Eq. (3.6), as the set of values of E for which $P_0(E)$ lies in the range

$$\pm 4t^q \left\{ T_q\left(\frac{t_b}{2t}\right) - T_q\left(\frac{t_a}{2t}\right) \right\}$$
$$-(-1)^p 2(t_{a\bar{b}}^q + t_{ab}^q) . \tag{3.16}$$

As t_b approaches t_a this range approaches zero as

$$4(t_b - t_a) t^{q-1} T_q'\left(\frac{t_a}{2t}\right) , \tag{3.17}$$

and each band centered on E_α has a width approximately equal to this range divided by $|P_0'(E_\alpha)|$, so that the sum rule for the derivatives of P at the points of the intersection spectrum at the critical point $t_a = t_b \geq t_{a\bar{b}} + t_{ab}$ is

$$\sum_{\alpha=1}^{q} \frac{1}{|P'(E_\alpha)|} = \frac{1}{t^{q-1} T_q'(\frac{t_a}{2t})} . \tag{3.18}$$

For q large and $t_b = t_a > t_{a\bar{b}} + t_{ab}$ this gives

$$\sum_{\alpha=1}^{q} \frac{1}{|P'(E_\alpha)|} \approx \frac{\sqrt{t_a^2 - 4t^2}}{2qt^q T_q(\frac{t_a}{2t})} \approx \frac{2^q \sqrt{t_a^2 - 4t^2}}{q\left(t_a + \sqrt{t_a^2 - 4t^2}\right)^q} . \tag{3.19}$$

For $t_b = t_a = t_{a\bar{b}} + t_{ab}$ it gives

$$\sum_{\alpha=1}^{q} \frac{1}{|P'(E_\alpha)|} = \frac{t_{a\bar{b}} - t_{ab}}{q(t_{a\bar{b}}^q - t_{ab}^q)} \tag{3.20}$$

for all q. For the bicritical case $t_b = t_a = 2t_{a\bar{b}} = 2t_{ab}$ it gives

$$\sum_{\alpha=1}^{q} \frac{1}{|P'(E_\alpha)|} = \frac{1}{q^2 t_{ab}^{q-1}} . \tag{3.21}$$

For $t_b > t_{a\bar{b}} + t_{ab} > t_a$ there is no intersection spectrum, since the ranges of the constant in Eq. (3.6) are nonoverlapping for $qk_2 = 0$ and $qk_2 = \pi$. However, we can get an exact expression for the intersection over k_1 of the spectra for fixed k_1. This would be the intersection spectrum for the dual problem with t_a, t_b interchanged. This spectrum is the set of values of E for which $P_0(E)$ lies in the range

$$\pm \left[4t^q T_q\left(\frac{t_b}{2t}\right) - 2t_{a\bar{b}}^q - 2t_{ab}^q \right]$$
$$-(-1)^p 4t^q T_q\left(\frac{t_a}{2t}\right) , \tag{3.22}$$

and it is shown in Appendix B that the measure of this spectrum is exactly $4(t_b - t_{a\bar{b}} - t_{ab})$. The same argument that led to Eq. (3.18) leads without approximation to Eq. (3.20) in the case $t_b = t_{a\bar{b}} + t_{ab} > t_a$, and to Eq. (3.21) in the bicritical case $t_b = 2t_{a\bar{b}} = 2t_{ab} > t_a$.

These sum rules can be used both to generate rough estimates for the measure of the spectrum (the union over k_2 of the spectrum for fixed k_2) and to get rigorous lower and upper bounds, by repeating the arguments used by Last and Wilkinson[10] and Last.[11] The rough estimate of this sum of bandwidths is obtained by multiplying the range of the constant term in the expression (3.6) for $P(E)$ by the appropriate expressions for $\sum 1/|P'|$ in Eqs. (3.19)–(3.21). For $t_b = t_a \geq t_{a\bar{b}} + t_{ab}$ this gives

$$W \approx 8\sqrt{t_a^2 - 4t^2}/q . \tag{3.23}$$

For $t_b = t_{a\bar{b}} + t_{ab} \geq t_a$ it gives

$$W \approx 8|t_{a\bar{b}} - t_{ab}|/q \,, \tag{3.24}$$

while for $t_b = 2t_{a\bar{b}} = 2t_{ab} \geq t_a$ it gives

$$W \approx 8t_b/q^2 \,. \tag{3.25}$$

The argument given by Last[11] for the upper bound needs no modifications for the case we are considering. The result he obtained is that if the spectrum is defined as the set of values of E for which

$$-b_1 \leq P_0(E) \leq b_2 \,, \tag{3.26}$$

where b_1, b_2 are positive, with $P_0(E)$ a polynomial whose zeros E_α are all real and distinct, and for which the zeros of the derivative all lie outside the bands, then the sum of the widths of the bands W satisfies

$$W < e(b_1 + b_2) \sum_\alpha \frac{1}{|P_0'(E_\alpha)|} \,. \tag{3.27}$$

The argument for the lower bound needs some modification because $b_1 \neq b_2$, but a straightforward extension of Last's argument gives

$$W > \frac{1}{2} \min\left(b_1 + \sqrt{b_1^2 + 4b_1b_2}, b_2 + \sqrt{b_2^2 + 4b_1b_2}\right) \times \sum_\alpha \frac{1}{|P_0'(E_\alpha)|} \,. \tag{3.28}$$

The parameters b_1, b_2 are given by the differences between the value of the expression in Eq. (3.16) or Eq. (3.22) that defines the intersection spectrum, and the two similar expressions that define the band edge. For $t_b = t_a > t_{a\bar{b}} + t_{ab}$ we get

$$b_{1,2} = 8t^q T_q\left(\frac{t_a}{2t}\right) \pm 4(t_{a\bar{b}}^q + t_{ab}^q) \,, \tag{3.29}$$

and the second term in this expression becomes negligible for large enough q. The bounds are, therefore,

$$8e\sqrt{t_a^2 - 4t^2}/q > W > 2(\sqrt{5}+1)\sqrt{t_a^2 - 4t^2}/q \,. \tag{3.30}$$

For $t_b = t_{a\bar{b}} + t_{ab} \geq t_a$ the parameters are

$$b_{1,2} = 4t_{a\bar{b}}^q + 4t_{ab}^q \pm 8t^q T_q\left(\frac{t_a}{2t}\right) \,. \tag{3.31}$$

For $t_{a\bar{b}} \neq t_{ab}$ only one of these terms is relevant in the large q limit, so that the bounds obtained from Eqs. (3.20), (3.27), (3.28), and (3.31) are

$$8e|t_{a\bar{b}} - t_{ab}|/q > W > 2(\sqrt{5}+1)|t_{a\bar{b}} - t_{ab}|/q \,. \tag{3.32}$$

For $t_b = 2t_{a\bar{b}} = 2t_{ab} \geq t_a$ the t_a dependent term in Eq. (3.31) is also relevant, and Eqs. (3.21), (3.27), (3.28), and (3.31) give

$$\frac{8et_b}{q^2} > W > \frac{2t_b}{q^2}\sqrt{1-\gamma}[\sqrt{5+3\gamma} + \sqrt{1-\gamma}] \,, \tag{3.33}$$

where

$$\gamma = \left|T_q\left(\frac{t_a}{2t_{ab}}\right)\right| \,. \tag{3.34}$$

This is an example of the importance of the second relevant length scale mentioned in connection with Eq. (3.9).

For the case $t_a = t_b = 2t_{a\bar{b}} = 2t_{ab}$ this line of argument gives us no useful lower bound, since all our lower bounds reduce to zero. For $t_{a\bar{b}} + t_{ab} > t_b \geq t_a$ we have not succeeded in finding an exact expression for the intersection spectrum or some equivalent spectrum.

IV. MULTIFRACTAL ANALYSIS

In this section, we perform numerical scaling analyses for the energy spectrum when $\phi = p/q$ approaches the quadratic irrational number $1/\tau \equiv (\sqrt{5}-1)/2$: the inverse of the golden mean. We take p/q to be F_{n-1}/F_n, where F_n is the nth Fibonacci number defined recursively by $F_n = F_{n-1} + F_{n-2}$ and $F_0 = F_1 = 1$. Note that $q = F_n \sim \tau^n$ for large n. In order to obtain a spectrum for a rational approximant, the Bloch theorem is applied first to Eq. (3.1), which is periodic with period F_n. Then the system becomes effectively finite and the spectrum is obtained by a numerical diagonalization.

To discuss localization of the eigenstates of Eq. (3.1) we examine its spectrum for fixed k_2. When $\phi = p/q$, the spectrum consists of q bands whose widths are denoted by Δ_i $(i = 1, \ldots, q)$. Since each band has the same number of states, we assign a probability measure 1/q to each band. With increasing q, each band splits into many subbands. In order to understand the scaling of the spectrum, we introduce a scaling index α by

$$1/q \sim \Delta_i^{\alpha_i} \,. \tag{4.1}$$

If states are localized in the limit of $q \to \infty$, the bandwidths Δ decrease exponentially with q, so that α goes to zero. This corresponds to a point spectrum. If states are extended, on the other hand, Δ scales as $1/q$, so α is 1. This corresponds to an absolutely continuous spectrum. In the critical case, α is expected to take values between 0 and 1. The values of α have a distribution on the whole spectrum. This situation corresponds to a singular continuous spectrum.

It is clear from the analysis of the difference between the union spectrum and the intersection spectrum given in Sec. III (Ref. 10) that the value of k_2 only plays a crucial role for the discrete spectrum, when the eigenstates are localized. For the absolutely continuous spectrum there is only an exponentially small dependence on the value of k_2. For the critical case, since the intersection spectrum of zero measure divides each subband of the union spectrum into two halves, it is clear that to take fixed k_2 gives a spectrum whose measure is roughly one half that of the union spectrum. In the subsequent discussion we will avoid the region of localized states and consider the spectrum found by taking the union over all values of k_2.

For a systematic analysis of systems with such complex scaling behavior, it is convenient to use the multifractal

technique developed by Halsey *et al.*[23] They have introduced the spectrum of singularity $f(\alpha)$ defined by

$$\Omega(\alpha) \sim \langle \Delta \rangle^{-f(\alpha)} , \tag{4.2}$$

where $\Omega(\alpha)d\alpha$ is a number of bands whose scaling index α lies between α and $\alpha + d\alpha$, and $\langle \Delta \rangle$ is a representative value of Δ which was not specified clearly in Ref. 23. We first explain the multifractal technique as reformulated by Kohmoto.[24]

A. Formulation

First introduce a scaling index ϵ by

$$\epsilon_i = -\frac{1}{n} \ln \Delta_i . \tag{4.3}$$

It is related to α by

$$\alpha\epsilon = \ln \tau. \tag{4.4}$$

We also define an "entropy function" $S(\epsilon)$ by

$$S(\epsilon) = \frac{1}{n} \ln \Omega'(\epsilon), \tag{4.5}$$

where $\Omega'(\epsilon)d\epsilon$ is the number of bands whose scaling index lies between ϵ and $\epsilon + d\epsilon$, namely $\Omega'(\epsilon) = \Omega(\alpha)|d\alpha/d\epsilon|$. Here it is important to notice that Δ_i and $\Omega'(\epsilon)$ depend exponentially on n. A band at the nth level splits into many bands at a higher level and may thus yield a number of different values of the scaling indices ϵ. However, we expect that the entropy function which represents the distribution of ϵ will converge to a smooth limiting form as n tends to infinity, and give complete information about the scaling behavior. As in the formalism of statistical mechanics, it is convenient to introduce a "partition function" and a "free energy" which are defined by

$$Z_n(\beta) = \sum_{i=1}^{q} \Delta_i^{\beta} \tag{4.6}$$

and

$$F(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln Z_n(\beta). \tag{4.7}$$

Once the free energy is calculated, the entropy function is obtained by a Legendre transformation,

$$S(\epsilon) = F(\beta) + \beta\epsilon, \tag{4.8}$$

$$\epsilon = -\frac{dF(\beta)}{d\beta}. \tag{4.9}$$

Thus by changing "temperature" β one can pick a value of ϵ and then the corresponding $S(\epsilon)$ is calculated. On the other hand, β can be written in terms of ϵ as

$$\beta = \frac{dS(\epsilon)}{d\epsilon}. \tag{4.10}$$

Usually $S(\epsilon)$ is defined on an interval $[\epsilon_{\min}, \epsilon_{\max}]$ and there is no scaling behavior corresponding to ϵ which is outside the interval and $S(\epsilon) = 0$. However, $F(\beta)$ is still defined there and from (4.8) it is given by $F(\beta) = -\epsilon_{\max}\beta$ for $\beta > \beta_{\min}$ and $F(\beta) = -\epsilon_{\min}\beta$ for $\beta < \beta_{\max}$. Thus useful information is only contained in $F(\beta)$ for the region between $\beta_{\min}$ and $\beta_{\max}$ where it is not linear.

Using Eq. (4.4) and identifying $\langle \Delta \rangle = \exp(-n\epsilon)$ [see (4.1)], $f(\alpha)$ is related to the entropy function by

$$f(\alpha) = \frac{S(\epsilon)}{\epsilon}. \tag{4.11}$$

$f(\alpha)$ is defined on an interval $[\alpha_{\min}, \alpha_{\max}]$ where $\alpha_{\min} = \ln\tau/\epsilon_{\max}$ and $\alpha_{\max} = \ln\tau/\epsilon_{\min}$. The maximum value of $f(\alpha)$ gives the Hausdorff dimension of the spectrum.

B. Numerical results for $f(\alpha)$

In all our numerical work we have taken $t_{a\bar{b}} = t_{ab}$, so we refer just to the value of t_{ab} in the rest of this section. Since we are studying the union of the spectrum over all values of k_2, the system is symmetric with respect to an interchange of t_a and t_b. Thus we carry out numerical calculations only for $t_b \leq t_a$. The system is characterized by two parameters t_b/t_a and t_{ab}/t_a.

Before going to the numerical results we notice that the results of Sec. III for the Lyapunov exponents [in particular Eq. (3.11) and the discussions below it] determine types of spectra in some parts of the parameter space. In region I in Fig. 1 ($t_{ab}/t_a < 1/2$), the eigenstates are extended in the x direction for irrational ϕ. Thus the spectrum is absolutely continuous and the minimum of α is $\alpha_{\min} = 1$ and $f(\alpha_{\min}) = 1$. On line BC ($t_a = t_b$), Eq. (3.1) is self-dual and the Lyapunov exponents for both x and p_x directions are zero. Then the spectrum is expected to be singular continuous. The properties of region II have not been determined by the analysis of the Lyapunov exponents in Sec. III, but it will be shown nu-

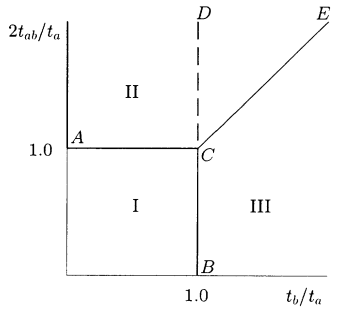

FIG. 1. Phase diagram of the different regions of the parameters of the Hamiltonian, for $t_{ab} = t_{a\bar{b}}$. In region I states are pure extended, and in region III the states are purely localized states. On the line BC the Hamiltonian has both diagonals as reflection lines, and the behavior is critical, very similar to that at the Harper point B. Bicritical behavior is found not only at the point C, but along the lines AC and CE.

merically that the spectrum is singular continuous in the whole region II.

The energy spectra for $t_a = t_b$ (self-dual line BD) and $n = 10$ are shown in Fig. 2(a). As is well known, the spectrum for $t_{ab} = 0$ (the critical point of pure Harper's equation) has a self-similar structure. The whole spectrum has three main bands, each main band has three subbands, and so on. This structure remains unchanged for $0 < 2t_{ab} < t_a$, but at $2t_{ab} = t_a$ it changes. For example, band edges a and b in pure Harper's equation in Fig. 2(a) continuously change to a' and b' as t_{ab} is increased. They are still band edges for $2t_{ab} < t_a$. When $2t_{ab} \to t_a$, however, the two points tend to the same point c and are no longer band edges [see Fig. 2(b)]. For $2t_{ab} > t_a$, the topological structure changes further.

In order to see the global scaling behavior, we have performed numerical calculations of $f(\alpha)$, which is defined in the limit of $n \to \infty$, by extrapolating the numerical data up to $n = 19$ ($q = F_{19} = 6765$).

Figure 2 is a plot of $f(\alpha)$ for $t_{ab} = 0.4$. This plot is identical, within the precision of the plot, to the plot obtained for the case $t_{ab} = 0$ (pure Harper's equation). As was pointed out by Tang and Kohmoto,[14] $f(\alpha)$ is defined on the interval between $\alpha_{\min} \simeq 0.421$ and $\alpha_{\max} \simeq 0.547$, and takes the maximum value $f = 0.5$ at $\alpha = 0.5$. Calculations for a variety of other values of t_{ab}/t_a in the range $0 < t_{ab}/t_a < 0.5$ give results that look identical to Fig. 3, and we conclude that $f(\alpha)$ is universal in this range. At the critical point of the pure Harper's equation, $\alpha_{\min}$ is the scaling index of the edges of the spectrum (and the edges of each subband), whereas $\alpha_{\max}$ is the scaling index of the center of the spectrum (and the center of each subband). Even if t_{ab} is nonzero, the situation remains the same as long as $2t_{ab} < t_a$. More specifically, the scaling index of each band is identical to that of the topologically corresponding band of the pure Harper's equation.

At the bicritical point C ($2t_{ab} = t_a = t_b$) the shape of $f(\alpha)$ suddenly changes. The nonzero values of f are found in the range between $\alpha_{\min} \simeq 0.272$, $\alpha_{\max} \simeq 0.421$, and the maximum value of f is about 0.33 at $\alpha \simeq 0.33$. Note that $\alpha_{\max}$ at this point is identical to $\alpha_{\min}$ for $2t_{ab} < t_a$. At $2t_{ab} = t_a$, however, the topo-

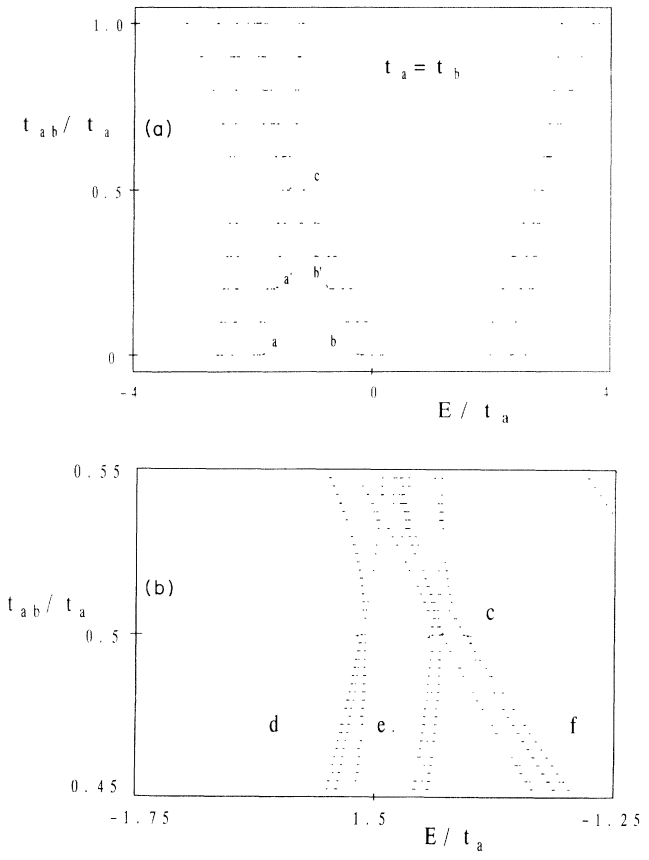

FIG. 2. (a) Spectra for various points on line BD in Fig. 1 ($t_a = t_b$). (b) Enlarged version of (a).

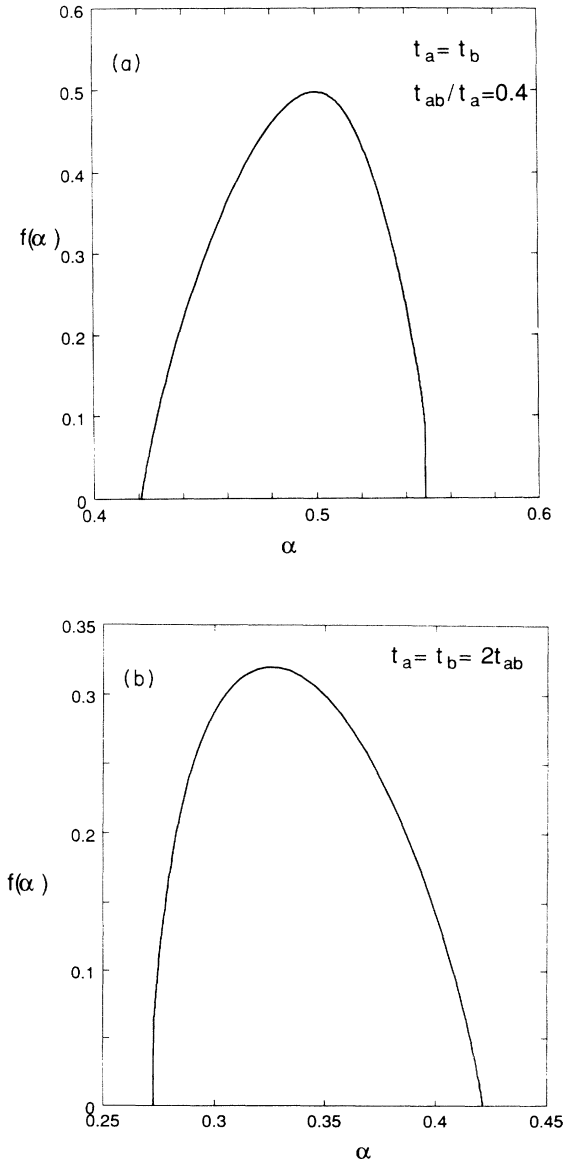

FIG. 3. (a) Plot of $f(\alpha)$ for $t_{ab} = 0.4t_a$, $t_a = t_b$. Other points on the line BC of Fig. 1 give identical results. (b) Plot of $f(\alpha)$ for the bicritical point C where $t_{ab} = 0.5t_a$, $t_a = t_b$.

TABLE I. Table of the extremal values of α and the position of the maximum in the multifractal analysis for various values of the parameters of the Hamiltonian.

t_{ab}/t_a	t_b/t_a	$\alpha_{\min}$	$\alpha_{\max}$	α_0	$f(\alpha_0)$
0,0.2,0.4	1.0	0.421	0.547	0.500	0.4980
0.5	1.0	0.272	0.421	0.326	0.3200
0.5	0.75	0.282	0.381	0.35	0.34
0.5	0.5	0.281	0.366	0.332	0.3294
0.5	0.25	0.281	0.370	0.33	0.32
1.0	1.0	0.300	0.650	0.43	0.41
2.0	1.0	0.305	0.640	0.42	0.41
3.0	1.0	0.313	0.628	0.44	0.42
0.6	0.5	0.29	0.41	0.36	0.35

logical structure of the spectrum and the scaling indices are different. For example, the scaling index of the bands coming from the centers of the subbands of pure Harper's equation is 0.303. It is 0.289 at the edges of bands [e.g., d, e, and f in Fig. 2(b)]. On the other hand, the index of the edges for $2t_{ab} < t_a$ remains 0.421 at c in Fig. 2 and becomes $\alpha_{\max}$.

Plots of $f(\alpha)$ on line CD ($2t_{ab} > t_a = t_b$) were obtained for some higher values of t_{ab}/t_a. Although $f(\alpha)$ is not universal on line CD, there is some similarity between the curves we found, in that the maximum points are nearly at $\alpha = 0.4$ and the maximum values are also about 0.4. The values of $\alpha_{\min}$, $\alpha_{\max}$, α_0, the position of the maximum of f, and $f(\alpha_0)$, the maximum value, are shown in Table I.

We have also calculated the form of f on the line AC ($2t_{ab}/t_a = 1, t_a > t_b$). As in the previous case on line CD, the curves are somewhat similar to one another in the sense that the maximum points are at $0.3 < \alpha < 0.35$ and the maximum values are also between 0.3 and 0.35. The main features of these results are also given in Table I.

Finally, we have calculated $f(\alpha)$ in region II, and one example is given in Table I. In this region, the convergence of the extrapolation $n \to \infty$ from the numerical data for finite n is not so good as the previous cases. Thus we cannot obtain results for $f(\alpha)$ reliable enough to establish or disprove universal features. Even though the errors are rather large, however, it appears that $f(\alpha)$ is a continuous function defined on a finite range of α. Thus we conclude that the spectrum is multifractal and singular continuous in region II. It is possible that the position α_0 and height $f(\alpha_0)$ of the maximum are universal in the region II, and also along the line AC (including the point C).

V. NUMERICAL RESULTS FOR BANDWIDTHS

A. Total bandwidth for Fibonacci sequence

In the previous section, we have obtained the following: (1) for $t_a = t_b$ and $2t_{ab} < t_a$ (line BC in Fig. 1), the scaling behavior of the spectrum is universal, that is, $f(\alpha)$ is completely identical to that of the $t_{ab} = 0$ case (pure Harper's equation); (2) at the bicritical point of $t_a = t_b$ and $2t_{ab} = t_a$ (point C in Fig. 1), the scaling behavior suddenly changes; (3) when $t_a = t_b$ and $2t_{ab} > t_a$ (line CD in Fig. 1), the scaling behavior is clearly different from those of (1) and (2). Although it is not completely universal, the changes are small for increasing t_{ab}/t_a; (4) when $t_a > t_b$ and $2t_{ab} = t_a$ (line AC in Fig. 1), the scaling behavior is similar but not identical to that of point C ($t_a = t_b$ and $2t_{ab} = t_a$); (5) when $2t_{ab} > t_a$ and $t_a > t_b$ (region II in Fig. 1), it is difficult to estimate $f(\alpha)$ but it is certain that the spectrum is also multifractal and singular continuous.

To make the above statements more concrete, we investigate scaling of the total bandwidths. Recall that in the pure Harper's equation in the critical case ($t_a = t_b, t_{ab} = 0$), the total bandwidth W scales as $W \sim 1/q$ for large q [see Eq. (1.5)]. We know from the analytic results of Sec. III that this scaling also holds all along the line BC ($t_a = t_b > 2t_{ab}$) except at C. The numerical results show that for the Fibonacci sequence the result

$$qW \sim 9.3299\sqrt{t_a^2 - 4t_{ab}^2} \tag{5.1}$$

holds accurately, in agreement with Eq. (3.14).

On the bicritical line AC ($2t_{ab} = t_a \geq t_b$), which separates the region of t_a dominant from the region of t_{ab} dominant, we know from the results of Sec. III that the sum of the bandwidths should scale as q^{-2}. Figure 4 shows the results for q^2W plotted against n ($\sim \ln q/\ln\tau$) for different values of t_b/t_a on this line. In this subsection, all the plots of W are in the units such that $t_a = 1$. It is seen that the scaling index δ for the global behavior of W which is defined by

$$W \sim (1/q)^{\delta}\,, \tag{5.2}$$

is 2. For $t_a = t_b = 2t_{ab}$, the quantity WF_n^2/t_a tends rapidly to 6.4911. In addition to this power-law decrease, an oscillatory behavior is observed for $t_b \neq t_a$. The oscillation has a period 4 against n in the case $2t_b = t_a$, which is probably related to the period 8 of the Fibonacci sequence modulo 3.

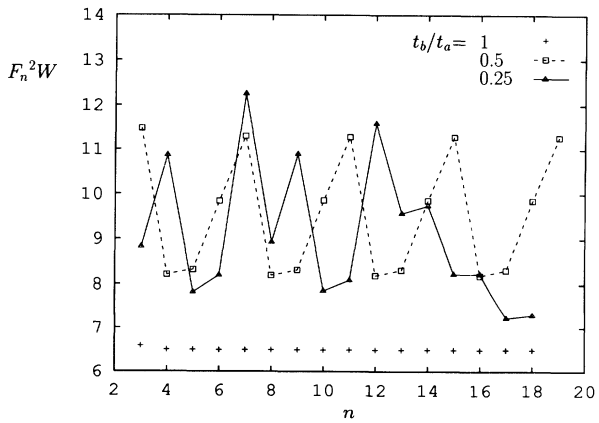

FIG. 4. Plot of F_n^2W versus n for $t_{ab}/t_a = 0.5$ and various values of t_b/t_a. A period 4 oscillation as a function of n for the case $t_b/t_a = 0.5$ can be seen clearly superposed on the $1/F_n^2$ dependence of W.

In the region $2t_{ab} > t_a \geq t_b$ we find two different types of behavior, but in all cases the analysis is complicated by oscillatory terms superposed on a general power-law dependence on q. Plots of $\ln W$ against $\ln F_n$ (which is proportional to n) appear to lie close to a line of slope -1.25 for the case $2t_{ab} > t_a = t_b$ (the line CD in Fig. 1), whereas they cluster around a line of slope -1.56 for the case $2t_{ab} > t_a > t_b$ (the interior of the region II in Fig. 1). Figure 5 shows plots of $\ln(F_n^{1.25}W)$ against n for various examples of $2t_{ab} > t_a = t_b$. There is a period 4 oscillation in the case $t_{ab} = t_a = t_b$, similar to the oscillation of period 4 that shows up in Fig. 4 for the case $t_{ab} = t_a/2 = t_b$. There is also an oscillation of period 6 for the case $t_{ab} = t_a/\sqrt{2}$. In other cases the oscillation around the general horizontal trend is irregular, and shows no signs of diminishing as n increases. An accurate estimate of the exponent δ can be made when the period is short, but we can only make a rough estimate when there is no period within the range of n we can use. There appears to be a slight difference, of order 0.01, between the values we get for δ with $t_{ab} = t_a = t_b$ and with $t_{ab} = t_a/\sqrt{2} = t_b/\sqrt{2}$.

Figure 6 shows plots of $\ln(F_n^{1.56}W)$ against n for various examples of $2t_{ab} > t_a > t_b$. Again, there are large oscillations about the general horizontal trend of the plots, and a simple period, 6 in this case, can be seen clearly for $t_{ab} = t_a/\sqrt{2}$, $t_b = 0$. For reasons that are discussed later in this paper, we think that this period is related to the period 12 of the Fibonacci sequence modulo 4. The value of $\delta = 1.56$ is not so easy to estimate for these examples, but it is clearly intermediate between the value we found on the line CD of Fig. 1 and the value of 2 which is known to be correct for the bicritical line AC.

The results in this subsection are summarized as follows. The scaling index δ for the total bandwidth is 1 on the critical line BC in Fig. 1. Not only at point C but also on line AC, δ is 2, i.e., the bicritical behavior is

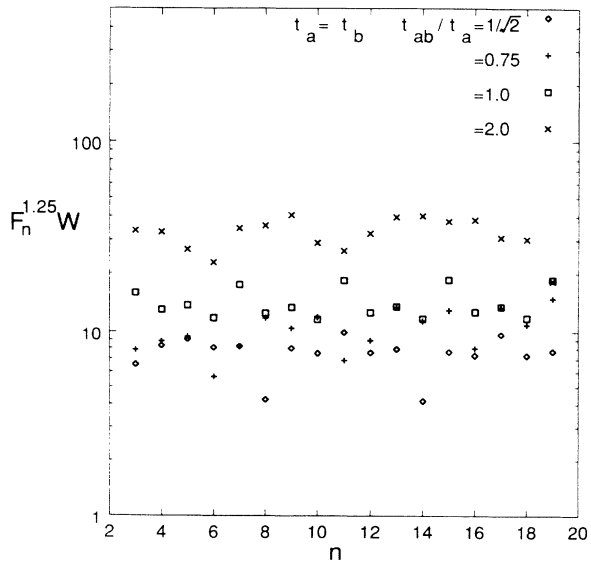

FIG. 5. Plots of $F_n^{1.25}W$ on a logarithmic scale versus n for $t_a = t_b$ and various values of t_{ab}/t_a. Period 6 can be seen for $t_{ab}/t_a = 1/\sqrt{2}$, and period 4 for $t_{ab}/t_a = 1$.

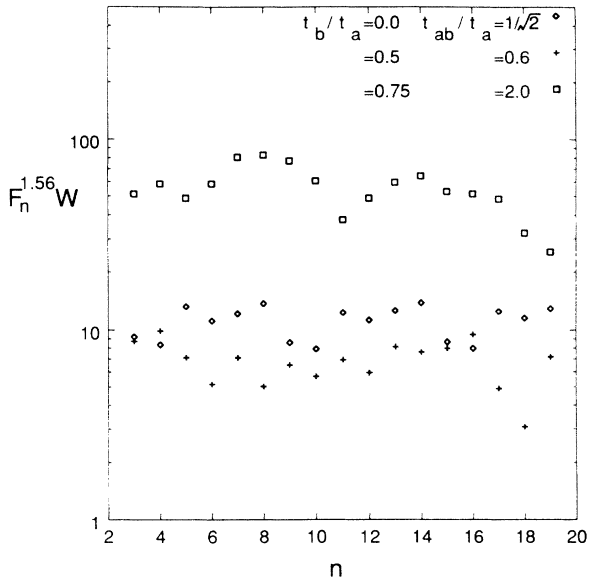

FIG. 6. Plots of $F_n^{1.56}W$ on a logarithmic scale versus n for various values of t_{ab}/t_a and $t_a/t_b \neq 1$. For $t_{ab}/t_a = 1/\sqrt{2}$ a period of 6 can be clearly seen.

found. On line CD, we cannot find a significant dependence of δ on t_{ab}/t_a, and it is about 1.25. Also in region II, the situation is similar and the value of δ is about 1.56.

In the limit of $t_{ab}/t_a \to \infty$, the problem reduces to the case only with nearest-neighbor couplings. Thus the δ must be 1 in this limit. Although we investigated δ for quite large t_{ab}/t_a (up to $t_{ab}/t_a = 10$) in region II and on the line CD, we could not find a tendency that δ decreases and approaches 1.

B. Scaling for $\phi = 1/q$

For rational approximants to the golden mean the total bandwidth is spread over a large number of bands in different energy ranges. For large denominator rational approximants to a small denominator rational, say p_0/q_0, the total bandwidth is concentrated in narrow ranges about the q_0 values of the energy where there is a logarithmic singularity in the density of states for the case $\phi = p_0/q_0$.[10,12] For sequences such as $\phi = 1/q$ or $2/q$ the limit of the sequence gives $p_0 = 0$, $q_0 = 1$, and all the width comes from one singular energy at or near the center of the band. To a considerable extent the results for this case can be understood in terms of the WKB analysis presented in Sec. VI, and many, but not all, of the results appear to generalize to more complicated sequences of fractional values of ϕ.

We have made some numerical checks of the scaling relations (1.5) and (3.13) for the case $t_b > t_a > 2t_{ab}$, and of the scaling relation (3.15) for $t_b > 2t_{ab} > t_a$, but these have not been extensive, and we have not found particularly interesting features. We have concentrated on three critical or bicritical cases. For $t_b = t_a > 2t_{ab}$ the bounds (3.30) show that the width must scale as $1/q$,

and finite size scaling theory suggests that it should have the limiting form given by Eq. (3.14). For the bicritical case $t_b = 2t_{ab} \geq t_a$ the bounds Eq. (3.33) show that the width must scale as $1/q^2$ (except possibly in the special case $t_a = t_b$), but Eq. (3.15) does not tell us much more, since we expect the additional length given by Eq. (3.9) to be involved. Finally, there is the critical case $2t_{a\bar{b}} = 2t_{ab} > t_b \geq t_a$, which one might expect to be analogous to the other critical case, since the dominant terms are essentially the same, but rotated through an angle $\pi/4$ in the p_x, x plane, with one half the size of unit cell. However, we know no rigorous bounds in this case, have derived no useful finite size scaling relations, and expect the influence of the additional lengths given by Eq. (3.9) to be important.

The case $t_a = t_b > 2t_{ab}$ is critical and the measure W scales like $\sqrt{t_a^2 - 4t_{ab}^2}/q$ as predicted by the finite size scaling theory of Eq. (3.14). We have done a numerical check for numerator 2 and q odd, where we have evaluated $qW/\sqrt{t_a^2 - 4t_{ab}^2}$ for t_{ab}=0, 0.2, 0.4 and found that they very rapidly converge to a common value. The difference between values of $qW/\sqrt{t_a^2 - 4t_{ab}^2}$ for different values of t_{ab} is less than one part in 10^5 for q greater than 41. This implies that the energy scale is reduced by a factor $\sqrt{1 - 4t_{ab}^2/t_a^2}$ and that there are no logarithmic corrections in the case of p=2 as we increase t_{ab}.

For numerator equal to unity, the scaling limit remains the same, but there are large corrections to scaling, which are shown in Fig. 7. These show an interesting cusplike oscillation of $qW/\sqrt{1 - 4t_{ab}^2/t_a^2}$ for nonzero t_{ab}. The periodicity of this oscillation can be related to the integral of the classical momentum over the length of the system, as we discuss in Sec. VI, while the magnitude of the oscillation is bounded by two curves which are followed by odd and even values of q with t_{ab}=0.[8]

The $2t_{ab} > \max(t_a, t_b)$ region is analogous to the $t_a = t_b$ dominant region just considered in that, without t_a, t_b, the problem reduces to that of the Harper's equation with twice as much flux per unit cell, and that t_a, t_b may be regarded as perturbations to the critical Harper problem. It is probably not important that the perturbations generally break the square symmetry of the system, since, as we discussed in Sec. II, the symmetry under $p_x \to -p_x$ remains, as well as rotation of phase space by π. In the cases we have examined with $\phi = 1/q$ the measure scales like $1/q$ for $\max(t_a, t_b) < 2t_{ab}$. As was mentioned in Sec. III, qW does not seem to tend to a limit, but oscillates in the t_{ab} dominant regime.

To study the periodic nature of the measure, it is convenient to choose the set of parameters $t_a/2t_{ab} = \cos(\pi p_1/q_1)$, $t_b/2t_{ab} = \cos(\pi p_2/q_2)$, and characterize the system with a set of fractions $(p_1/q_1, p_2/q_2)$. Several graphs of qW as a function of q are shown in Fig. 8. For $p_1 = p_2 = 1$, we have found strong peaks at multiples of $q_1 \times q_2$ (primary peaks) and much less strong peaks at multiples of q_1 (secondary peaks) if q_1 is a small integer such as two or three. For q_1 and q_2 both large, close, and relatively prime, for example, five and six or five and seven, the secondary peaks do not seem to occur at definite multiples of either number. The primary peaks always occur at multiples of $q_1 \times q_2$. If $q_1 = q_2$ or they share a common factor, then the secondary peaks occur at integer multiples of their lowest common multiple, and the values of qW at these points nearly match those of the primary peaks.

In the bicritical case with $2t_{ab} = t_b > t_a$ we know from the inequalities (3.32) that the measure of the spectrum

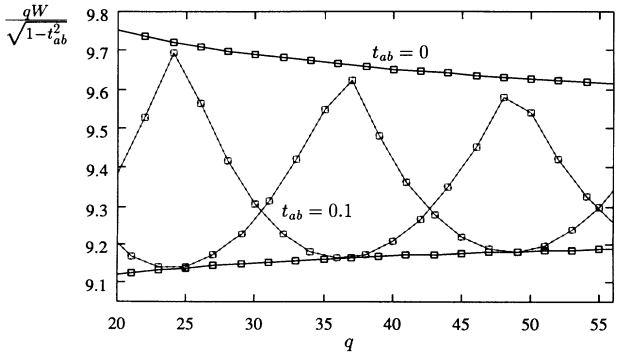

FIG. 7. Plots of $qW/\sqrt{1 - 4t_{ab}^2}$ as a function of q for t_{ab}=0, 0.1 with $t_a = t_b$=1. The squares representing t_{ab}=0 form envelopes (upper and lower bounds) for other values of $t_{ab} < 0.5$.

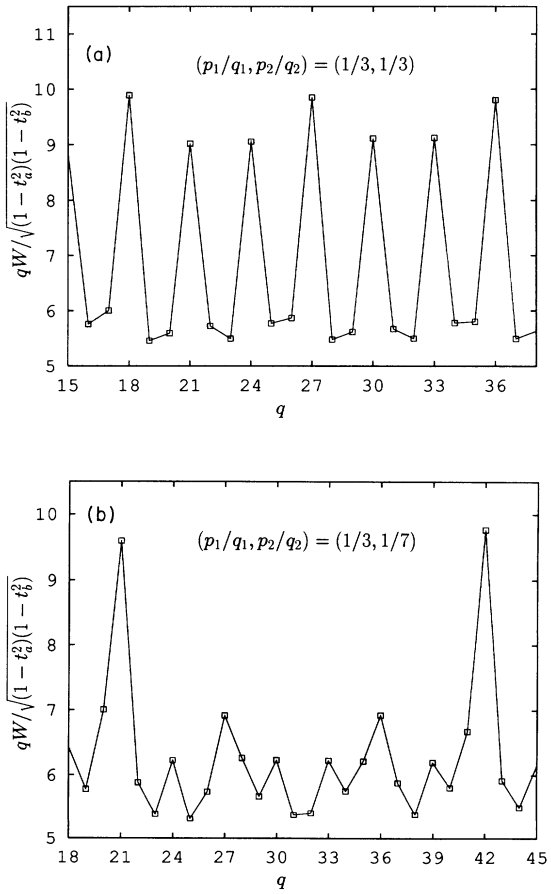

FIG. 8. Plots of $qW/\sqrt{(1 - t_a^2)(1 - t_b^2)}$ as a function of q in the critical regime $2t_{ab} = 1 > t_a, t_b$ for several values of $(t_a, t_b) = [\cos(\pi p_1/q_1), \cos(\pi p_2/q_2)]$. Each graph is characterized by a set of fractions, $(p_1/q_1, p_2/q_2)$.

must be proportional to q^{-2}, and there is one length scale remaining which is given by Eq. (3.9), so we might expect a scaling form

$$q^2W = g_{bc}[q \arccos(t_a/2t_{ab})] . \quad (5.3)$$

Figure 9 shows $g_{bc}(x)$, with $x = q\arccos(t_a/2t_{ab})$, for $t_a/2t_{ab}$=1/3 and 0.99. In the limit $t_a = 2t_{ab}$, $g_{bc}(x)$ becomes a slowly increasing function of x, and we have not determined its limiting value; this slow convergence may be a special feature of particular fractional forms of ϕ, as we found no sign of it for the Fibonacci sequence. Figure 9(b) transparently displays a cusplike form of the scaling function, whose variation lies well within the bounds given by Eq. (3.33). Earlier in this subsection, such cusps were reported when we considered the $t_a = t_b$ dominant region. It turns out that a function of the form $g_{bc}(x) = A - B\ln[1 + |\sin(\pi(x - \delta))|]$, where A, B, and δ are constants chosen to fit, describes the actual scaling function rather well. The motivation for this form of $g_{bc}(x)$ came from the idea that x must be a dimensionless quantity and that such a quantity can be obtained by multiplying q on both sides of Eq. (3.10) *before* one takes the limit $q \to \infty$. Since $t_{a\bar{b}} = t_{ab}$, this gives us $x = q\lambda_x = \ln 2 - \ln(1 + |T_q(t_a/2t_{ab})|)$ which seems to contain essential features of $g_{bc}(x)$. Based on this hypothesis, the cusplike behavior can be understood as the reflection of the importance of the absolute value of the Chebyshev polynomial.

C. Scaling for $\phi = p/(p^2 \pm 1)$

The sequences $\phi = p/(p^2 \pm 1)$ are intermediate between $\phi = 1/q$ and $\phi = F_n/F_{n-1}$. The continued fraction expansion has only two terms in it, rather than the $n/2$ terms for the Fibonacci sequence, and it represents a slow modulation of the diagonal term, and so could be treated by WKB methods (although we have not succeeded in carrying out such an analysis in this case). It is, however, more complicated than the $\phi = 1/q$ case in that the significant contributions to the bandwidth come not only from bands in the immediate vicinity of the singular energy $-t_at_b/t_{ab}$, but also from the centers of neighboring clusters of p subbands.

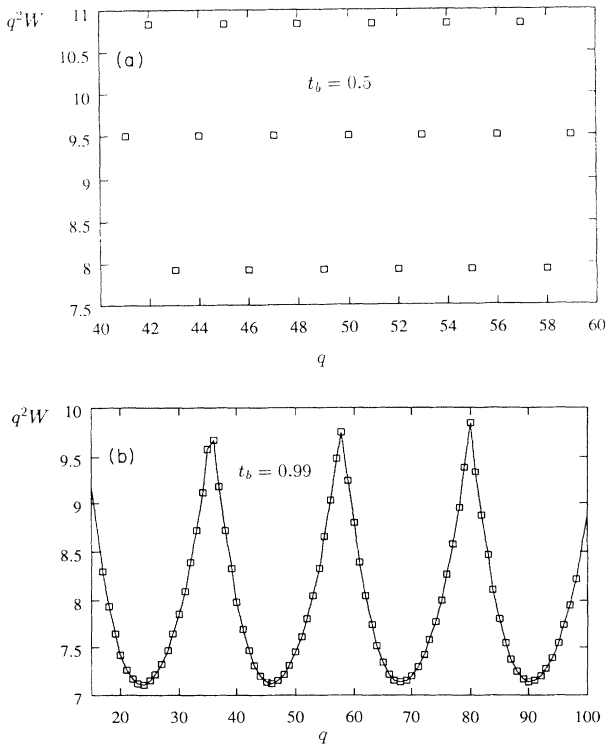

FIG. 9. Plots of q^2W as a function of q at the bicritical point $2t_{ab}$=1=t_a for t_b=0.5 and 0.99, respectively.

We have studied the bicritical case $t_b = 2t_{ab}$=1 and $t_a = \cos(p_1\pi/q_1)$, where q_1 was kept small (≤ 7). With this choice of t_a we expect behavior periodic in q in the limiting values of q^2W coming from the constant term Eq. (3.6), but there may also be some dependence on the numerator p coming from other terms in the characteristic equation. We have confined ourselves to numerators not exceeding 20, and denominators up to 401. As in the previous case of a simple fraction of ϕ, the scaling function showed periodicity (up to corrections to scaling) with periods equal to q_1 for both $p^2 \pm 1$. The pattern of graphs are quite different in two cases and values of maxima and minima as well as where the maxima and minima occur do not agree in general. The convergence to the limit was much slower here than it was with $\phi = 1/q$. This might have to do with the fact that, for the fraction $\phi = p/(p^2 \pm 1) \sim 1/p$ with $p \leq 20$, the corrections are not yet completely negligible. For $p_1/q_1 = 1/4$, we should expect a period of 2 because $p^2 \pm 1 \pmod 4$ alternates between 2 (0) and 1 (3) for odd and even p but instead we observe a period of 4. Apparently the scaling function here is more complicated than what was the case if the relevant variable was simply the ratio of two length scales of the system. This shows conclusively that there is periodic dependence on the value of the numerator as well as on the value of the denominator.

In the t_{ab} dominant critical region we have found instances of an anomalous (noninteger) scaling exponent. Figure 10 shows a plot of $\log(qW)$ against $\log q$ for the case $2t_{ab} = 1, t_a = 0$, $t_b = \cos(\pi/4)$, $\phi = p/(p^2 + 1)$. For numerators equal to 2 mod 4 we see the points lying on a line with a negative slope, whereas linear scaling should give no slope at all, as we see for the other values of p. The slope both of this plot, and of the

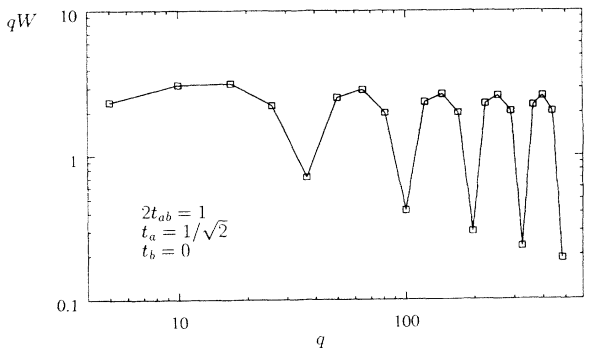

FIG. 10. Plot of qW on a logarithmic scale against $\log q$ for $\phi = p/q = p/(p^2 + 1)$. Points for $p = 2 \bmod 4$ are clearly seen to lie on a line with a slope -0.50.

very similar plot for $\phi = p/(p^2-1)$, is found to be within 0.3% of $-1/2$, which implies that, for the sequence $p/q = 2(2n+1)/[4(2n+1)^2 \pm 1]$, the scaling is like $1/q^{3/2}$ rather than $1/q$. For $t_a = 0, t_b = \cos(\pi/3)$, the plot alternates between points with exponent close to 1, if $p = 0, \pm 1 \pmod 6$, and those with exponent close to 1.5 if $p = 2, 3, 4 \pmod 6$ for both types of denominator. It is much harder to extract the critical exponents here because we have to increase p by six instead of four to arrive at points lying on the same line. For $0 < t_a = t_b < 2t_{ab}$, the evidence for noninteger exponent is far less obvious and we are not sure if the exponent is significantly different from one. The noninteger exponents found in our studies of the Fibonacci sequence are likely to be related to these results.

VI. WKB THEORY

The saddle point value of the energy contour E_s corresponds to quantum mechanical states that can thread the system without attenuation and there exists an interesting relation between the integral of the classical momentum p_x given by Eqs. (2.1)–(2.4) between turning points with periodicities in the scaling functions.

In Ref. 6 it was shown that for $\phi = 1/q$ the sum of the bandwidths could be related to the Green function at the turning points. For $t_a = t_b > t_{a\bar{b}} + t_{ab}$, energy close to $-2(t_{a\bar{b}} + t_{ab})$ and $x = \phi/2 + \xi$, where ξ is small, the continuum approximation for Eqs. (1.6) and (3.1) takes the form

$$H + 2(t_{a\bar{b}} + t_{ab}) \approx 4\pi^2\phi^2(t_a + t_{a\bar{b}} + t_{ab})\frac{d^2}{d\xi^2} + (t_a - t_{a\bar{b}} - t_{ab})\xi^2 + 2\pi i\phi(t_{a\bar{b}} - t_{ab})\left(\xi\frac{d}{d\xi} + \frac{d}{d\xi}\xi\right) . \tag{6.1}$$

A very similar expression is obtained near the other turning point with x close to zero and p_x close to π. This can be diagonalized by a canonical transformation, and the energy scale it yields is

$$4\pi\phi\sqrt{t_a^2 - 4t_{a\bar{b}}t_{ab}} \tag{6.2}$$

in both cases. The earlier arguments[6,8,12,10] applied to this case give the scaling result quoted in Eq. (3.14).

The corrections to this scaling form can be calculated by using the WKB approximation to get a more precise approximation to the bandwidths,[8] which involves the connection between the turning points as well as the behavior at the turning points. We have not worked this case out in detail, but we know that for $2t_{a\bar{b}} = 2t_{ab}$ the phase change around an orbit close to the critical orbit is

$$\frac{1}{2\pi\phi}\oint p_x dx \approx \frac{2}{\pi\phi}\int_0^\pi dx \arccos\left(-\frac{2t_{ab} + t_a\cos x}{t_a + 2t_{ab}\cos x}\right) . \tag{6.3}$$

Differentiation of the right-hand side of this equation with respect to t_{ab} gives an integral that can be evaluated explicitly, and reintegration of this result gives

$$\frac{q}{2\pi}\oint p_x dx \approx q\pi - \frac{4q}{\pi}\int_0^{\tanh^{-1}(2t_{ab}/t_a)} \frac{\theta}{\sinh\theta} d\theta . \tag{6.4}$$

The first term in this expression gives rise to the correction to scaling that alternates with the parity of q even for $t_{ab} = 0$, while the second term gives a period in q that goes like $\pi^2 t_a/4t_{ab}$ for small t_{ab}. For $t_{ab}/t_a = 0.1$ it gives 24.5 as the period, which is in good agreement with the numerical results shown in Fig. 7. In the limit $2t_{ab} \to t_a$ the second term is πq and cancels the periodicity due to the first term.

For $2t_{a\bar{b}} = 2t_{ab}$ dominant the problem is somewhat different. The classical contours through the saddle points are given by

$$\frac{1}{t_{ab}}(t_b + 2t_{ab}\cos p_x)(t_a + 2t_{ab}\cos x) = 0 . \tag{6.5}$$

The quadratic approximation to the Hamiltonian near one of the saddle points is given by

$$H + \frac{t_a t_b}{t_{ab}} \approx \pm\frac{2t_{ab}}{2\pi\phi} i\sqrt{\left(1 - \frac{t_a^2}{4t_{ab}^2}\right)\left(1 - \frac{t_b^2}{4t_{ab}^2}\right)} \times \left(\xi\frac{d}{d\xi} + \frac{d}{d\xi}\xi\right) , \tag{6.6}$$

and diagonalization of this by a canonical transformation gives an energy scale

$$4\pi\phi t_{ab}\sqrt{\left(1 - \frac{t_a^2}{4t_{ab}^2}\right)\left(1 - \frac{t_b^2}{4t_{ab}^2}\right)} . \tag{6.7}$$

This quantity gives a good account of the relative sizes of the energy scales of the bandwidths shown in Fig. 8. However, as we showed in the study of the characteristic equation in Sec. III, the lengths given in Eq. (3.9) are certainly relevant, and there must be terms periodic or nearly periodic in q. These should come from the rectangular contours given in Eq. (6.5), whose areas are

$$4\arccos\left(\pm\frac{t_a}{2t_{ab}}\right)\arccos\left(\pm\frac{t_b}{2t_{ab}}\right) , \tag{6.8}$$

and it is the ratio of these areas to the quantum of action $4\pi^2\phi$ given by Eq. (1.2) that determines this periodicity.

Qualitatively this accounts for the periods in q of 9 and 21 which show up in Figs. 8(a) and (b), since the smallest areas given by Eq. (6.8) are $4\pi^2/9$ and $4\pi^2/21$ in the two cases, and the larger rectangles are multiples of these. To understand these results in more detail we need to make a more careful study of the way the phases affect the dynamics.

The bicritical case $t_b = 2t_{ab}$ is simpler, since one of the arccosines in Eq. (6.8) is equal to π. For $t_a = t_{ab}$ the two rectangles have areas $4\pi^2/3$ and $8\pi^2/3$, so the period 3 in q which can be seen in Fig. 9(a) should be expected, while for $t_a = 0.495t_{ab}$, $\pi/\arccos(t_a/2t_{ab})$ is

equal to 22.2, which agrees well with the period shown in Fig. 9(b).

For a large denominator fraction p/q approximating a small denominator rational p_0/q_0, the commutator $[x, p_x]$ remains finite and the WKB approach does not directly apply. Wilkinson[25] has shown that at $t_a = t_b, t_{ab} = t_{a\bar{b}} = 0$, each one of q_0 clusters of bands are described by an effective Hamiltonian obtained from quantizing the inverse of the characteristic polynomial (3.6) for the nth band,

$$E_n(k_1, k_2) = P_{0,n}^{-1}[2t_a^{q_0}\cos(q_0 k_1) + 2t_b^{q_0}\cos(q_0 k_2)] \ . \quad (6.9)$$

The new flux for each cluster depends on the Chern number for that particular band, but is in general of the order of the difference $p/q - p_0/q_0$, therefore, small. A more direct way[12] is to regard the problem as having $\phi = p_0/q_0$ with the wave vector k_2 slowly modulated with a period $q_s q_0/|q_s p_0 - p_s q_0|$. The k_2 in (3.6) becomes $k_{2,n} = k_2 + 2\pi n[(p_s q_0 - p_0 q_s)/q_0 q_s]$ and k_1 is also site dependent. This approach generalizes to $t_{ab} \neq 0$ with the result that singular energies are the q_0 roots of the equation

$$P_0(E) - 4(-1)^{p_0} t_{ab}^{q_0} = 0 \quad (6.10)$$

for $t_a = t_b$ dominant and of

$$P_0(E) - 4(-1)^{p_0} t_{ab}^{q_0} T_{q_0}\left(\frac{t_a}{2t_{ab}}\right) T_{q_0}\left(\frac{t_b}{2t_{ab}}\right) = 0 \quad (6.11)$$

for $t_{ab} = t_{a\bar{b}}$ dominant regime. We have done some numerical checks for $p_0 = 1, q_0 = 2$ and found that the singular energies come at $\pm 2t_a$ and $\pm 2\sqrt{t_a^2 + t_b^2 - t_a^2 t_b^2}$, respectively, in agreement with predictions given by Eqs. (6.10) and (6.11).

VII. DISCUSSION

Our studies have shown that the characteristic critical properties of Harper's equation persist when the symmetry is broken by terms that couple sites to their next-nearest neighbors, provided that the reflection symmetry in the diagonals is preserved, and provided that the next-nearest-neighbor terms satisfy the inequality $t_{ab} + t_{a\bar{b}} < t_a = t_b$. In this region, the multifractal analysis gives a universal result, strict bounds for the width W of the spectrum show that it must go to zero with the reciprocal of the denominator q, and numerical results and semiclassical analysis show that qW has a limit that is rescaled by the next-nearest-neighbor coupling, but which is independent of the numerator p. It is only in the corrections to scaling that any oscillatory behavior shows up.

For the bicritical line $t_{ab} = t_{a\bar{b}} = \max(t_a, t_b)/2$, which divides the region dominated by the nearest-neighbor terms from that dominated by the next-nearest-neighbor terms, the multifractal behavior is quite different, and the width W of the spectrum is known to be proportional to $1/q^2$, as we know from the rigorous bounds,[10,11] as well as from numerical work. In this case an oscillatory behavior superposed on the $1/q^2$ dependence shows up, whose periodicity is at least partially understood from the WKB analysis of Sec. VI.

The region for which we have least understanding is the region dominated by the next-nearest-neighbor terms, with $t_{ab} = t_{a\bar{b}} > \max(t_a, t_b)/2$. Although the case $t_a = 0 = t_b$ is equivalent to the original Harper equation, with the axes turned by $\pi/4$ and the unit cell doubled in area, for any nonzero values of the nearest-neighbor coupling their strengths t_a, t_b remain relevant, as can be seen clearly from Eq. (3.6). The spectrum has a multifractal form, but the multifractal analysis gives different ranges of the exponent for different values of the parameters of the Hamiltonian. There seems to be a marked difference between the behavior for $t_a = t_b$ and for $t_a \neq t_b$. When the width W of the spectrum is studied the results are quite different for the ratio of Fibonacci numbers and for $1/q$ or similar sequences of fractions with fixed numerator. For $\phi = 1/q$ we find W going to zero like $1/q$, with an oscillatory coefficient, while for the Fibonacci sequence the oscillatory behavior is similar, but the dependence on the denominator is of the form $1/q^{\delta}$, where δ seems to be about 1.25 for $t_a = t_b$, and 1.56 for $t_a \neq t_b$.

We believe a partial understanding of this behavior can be obtained from our results for ϕ of the form $1/(q_1 + 1/q_2)$ which were discussed in Sec. V C. For large values of the numbers in the continued fraction expansion of ϕ we get relatively simple scaling behavior which can be studied by using WKB theory, although we have not carried out the analysis in detail yet. Because of the periodic or nearly periodic behavior as the q_j are varied, each stage of the scaling may carry one arbitrarily close to the bicritical boundary of the critical region, and we found examples, one of which is shown in Fig. 10, where for certain values of q_1 the dependence of W on q_2 at the next stage of scaling was the $1/q_2^2$ typical of the bicritical boundary. Approximants of the golden mean, which are the ratios of two Fibonacci numbers, have continued fraction expansions in which the terms are particularly small, so one is very far from the simple scaling behavior expected for large q_j, so the scaling behavior at each stage may be intermediate between the critical and the bicritical behavior.

Note added in proof. Equations (4.6)–(4.11) can be used to show that the exponent δ defined in Eq. (5.2) is given by the value of $-f'$ at the point where $1 - f = -\alpha f'$, so that it is minus the slope of the tangent to the curve $f(\alpha)$ through the point $\alpha = 0, f = 1$. This implies that the maximum of the curve cannot be exactly at the point $\alpha = 0.5, f = 0.5$ given by Tang and Kohmoto[14] for the pure Harper equation, since we know that $\delta = 1$. More careful analysis of our numerical results gives the maximum at $\alpha = 0.500, f = 0.498$, in agreement with results for the Hausdorff dimension recently obtained by Wilkinson and Austin.[26] If this is used to deduce δ from our plots of f against α it gives very good results in those cases where there are no incommensurate oscillations (the first, second, and fourth rows of Table I). The results given in the third row of Table I are incompatible with the known result $\delta = 2$, but extrapolation errors are of order 0.02 for these cases with large oscillatory dependence on n.

ACKNOWLEDGMENTS

We are particularly grateful to Yoram Last for illuminating discussions and for pointing out that the upper and lower bounds obtained in Ref. 11 are almost immediately applicable to this problem.

APPENDIX A: EVALUATION OF PRODUCT OF OFF-DIAGONAL TERMS

Evaluation of the product in Eq. (3.3) is required to obtain Eq. (3.6). The expression

$$P(k_2)=\prod_{n=0}^{q-1}\left(t_a+t_{a\bar{b}}e^{-\frac{2\pi ip}{q}(n+\frac{1}{2})-ik_2}+t_{ab}e^{\frac{2\pi ip}{q}(n+\frac{1}{2})+ik_2}\right) \tag{A1}$$

must be periodic in k_2 with period $2\pi/q$, since the addition of $2\pi/q$ to k_2 yields a permutation of the same factors in the product. This tells us that the only terms that survive in the product are those with q factors of $\exp(-ik_2)$, those with q factors of $\exp(ik_2)$, and those with equal numbers of factors of $\exp(-ik_2)$ and $\exp(ik_2)$. The first two cases give

$$t_{a\bar{b}}^q e^{-\pi ipq-iqk_2}+t_{ab}^q e^{\pi ipq+iqk_2}\ , \tag{A2}$$

while the k_2 independent terms can be written as

$$Q=\prod_{n=0}^{q-1}\left(t_a+te^{-\frac{2\pi ip}{q}(n+\frac{1}{2})-ik_2}+te^{\frac{2\pi ip}{q}(n+\frac{1}{2})+ik_2}\right)-(-1)^{pq}2t^q\cos(qk_2)\ , \tag{A3}$$

where $t^2=t_{a\bar{b}}t_{ab}$; here we have subtracted off the k_2 dependent terms using Eq. (A2). The expression obtained from Eq. (A3) by setting $k_2=0$,

$$Q/t^q+(-1)^{pq}2=\prod_{n=0}^{q-1}\left(t_a/t+e^{-\frac{2\pi ip}{q}(n+\frac{1}{2})}+e^{\frac{2\pi ip}{q}(n+\frac{1}{2})}\right)\ , \tag{A4}$$

is a polynomial in t_a/t whose zeros are given by

$$\frac{t_a}{t}=-2\cos\left[\frac{2\pi p}{q}\left(n+\frac{1}{2}\right)\right]\ . \tag{A5}$$

These are the zeros of the equation

$$\cos\left(q\arccos\frac{t_a}{2t}\right)\equiv T_q\left(\frac{t_a}{2t}\right)=(-1)^{p-q}=-(-1)^{pq}\ , \tag{A6}$$

where T_q is the Chebyshev polynomial of order q. Since the coefficient of x^q in $T_q(x)$ is 2^{q-1}, this gives

$$Q/t^q+(-1)^{pq}2=2T_q(t_a/2t)+(-1)^{pq}2\ . \tag{A7}$$

Combination of this with Eq. (A2) in Eq. (A1) gives

$$\begin{aligned}P(k_2)&=\prod_{n=0}^{q-1}\left(t_a+t_{a\bar{b}}e^{-\frac{2\pi ip}{q}(n+\frac{1}{2})-ik_2}+t_{ab}e^{\frac{2\pi ip}{q}(n+\frac{1}{2})+ik_2}\right)\\&=(-1)^{p-q-1}[t_{a\bar{b}}^q e^{-iqk_2}+t_{ab}^q e^{iqk_2}]+2(t_{a\bar{b}}t_{ab})^{q/2}T_q\left(\frac{t_a}{2\sqrt{t_{a\bar{b}}t_{ab}}}\right)\ ,\end{aligned} \tag{A8}$$

which is the result used to derive Eq. (3.6).

APPENDIX B: EXACT EXPRESSIONS FOR THE INTERSECTION SPECTRUM

Avron, Mouche, and Simon[7] have shown that for the case $t_{a\bar{b}}=0=t_{ab}$ the intersection spectrum has measure $4|t_b-t_a|$. By exploiting the techniques used in earlier work[4,12] we can generalize this result to the case $t_b\geq t_a>t_{a\bar{b}}+t_{ab}$. This argument depends in its details on the parities of p and q, so we will give the argument explicitly for the case of p, q odd. For $qk_2=\pi$ and qk_1 zero or π we exploit the symmetry of Eq. (3.1) about the points $n=0$ and $n=q/2$, which allows the problem to be reduced to eigenvalue problems for tridiagonal matrices of order $(q\pm1)/2$. The eigenvalues E_m^{++} and E_m^{--} which correspond to eigenstates of Eqs. (3.1) and (3.2) that are even about both these symmetry points or odd about both of them give solutions with $qk_1=0$, while E_m^{+-} and E_m^{-+} give solutions with $qk_1=\pi$. We number the eigenvalues from highest to lowest, starting with zero for the solutions E_0^{++} and E_0^{+-} which are even about $n=0$, and starting with unity for those that are odd about this point. The highest value of m is $(q-1)/2$ in all four cases. The matrices corresponding to odd and even boundary conditions about the point $n=0$ differ only in whether the 01 element is zero or not, and the trace formula gives

$$2t_b-E_0^{+-}+\sum_m(E_m^{--}-E_m^{+-})=0\ , \tag{B1}$$

where each term in the sum is positive. The matrices corresponding to odd and even boundary conditions about $n=q/2$ differ in having an extra term $\pm(t_a-t_{a\bar{b}}-t_{ab})$ in the lowest diagonal element, so the trace formula gives

$$\sum_m(E_m^{-+}-E_m^{--})=2(t_a-t_{a\bar{b}}-t_{ab})\ , \tag{B2}$$

with each term in the sum positive. Addition of these two equations gives

$$\sum_m(E_m^{-+}-E_m^{+-})=E_0^{+-}-2(t_b-t_a+t_{a\bar{b}}+t_{ab})\ . \tag{B3}$$

The terms in the sum are all positive, and are equal to the band gaps with $qk_2=0$, $qk_1=\pi$ in the intersection spectrum, while the eigenvalue E_0^{+-} is the highest band edge of the intersection spectrum. A similar argu-

ment can be constructed for the solutions with $qk_2 = \pi$, $qk_1 = 0$, where we number the eigenvalues $\mathcal{E}_m^{++}$, $\mathcal{E}_m^{--}$ in increasing order. This gives

$$\sum_m (\mathcal{E}_m^{++} - \mathcal{E}_m^{--}) = -\mathcal{E}_0^{++} - 2(t_b - t_a - t_{a\bar{b}} - t_{ab}) \ . \quad \text{(B4)}$$

Again, all the terms in the sum are positive and give band gaps of the intersection spectrum, while $\mathcal{E}_0^{++}$ is the lowest band edge of this spectrum. Addition of Eqs. (B3) and (B4) gives the measure of the intersection spectrum as

$$E_0^{+-} - \mathcal{E}_0^{++} - \sum_m (E_m^{-+} - E_m^{+-}) - \sum_m (\mathcal{E}_m^{++} - \mathcal{E}_m^{--}) = 4(t_b - t_a) \ . \quad \text{(B5)}$$

For the case $t_b \geq t_{a\bar{b}} + t_{ab} > t_a$ there is no change in the argument that leads to Eq. (B4), but Eqs. (B1)–(B3) must be replaced by

$$2t_b - E_0^{++} + \sum_m (E_m^{-+} - E_m^{++}) = 0 \ , \quad \text{(B6)}$$

$$\sum_m (E_m^{--} - E_m^{-+}) = 2(t_{a\bar{b}} + t_{ab} - t_a) \ , \quad \text{(B7)}$$

and

$$\sum_m (E_m^{--} - E_m^{++}) = E_0^{++} - 2(t_b + t_a - t_{a\bar{b}} - t_{ab}) \ , \quad \text{(B8)}$$

with all terms in the sums positive. Addition of this to Eq. (B4) gives the measure of the spectrum for $qk_1 = 0$ as

$$E_0^{++} - \mathcal{E}_0^{++} - \sum_m (E_m^{--} - E_m^{++}) - \sum_m (\mathcal{E}_m^{++} - \mathcal{E}_m^{--}) = 4(t_b - t_{a\bar{b}} - t_{ab}) \ , \quad \text{(B9)}$$

and this gives the generalization of the result for the intersection spectrum for the case $t_b \geq t_{a\bar{b}} + t_{ab} \geq t_a$.

[1] M. Ya. Azbel, Zh. Eksp. Teor. Fiz. **46**, 929 (1964) [Sov. Phys. JETP **19**, 634 (1964)].
[2] D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).
[3] S. Aubry and G. André, Ann. Isr. Phys. Soc. **3**, 133 (1980).
[4] D. J. Thouless, Phys. Rev. B **28**, 4272 (1983).
[5] See B. Simon and B. Souillard, *Random Media* (Springer-Verlag, New York, 1987), IMA Vol. 7, and the references therein for a rigorous discussion of the relation between the Lyapunov exponent and the localization of eigenfunctions.
[6] D. J. Thouless, Commun. Math. Phys. **127**, 187 (1990).
[7] J. Avron, P. H. M. v. Mouche, and B. Simon, Commun. Math. Phys. **132**, 103 (1990).
[8] D. J. Thouless and Y. Tan, J. Phys. A **24**, 4055 (1991).
[9] G. I. Watson, J. Phys. A **24**, 4999 (1991).
[10] Y. Last and M. Wilkinson, J. Phys. A **25**, 6123 (1992).
[11] Y. Last, Commun. Math. Phys. (to be published).
[12] D. J. Thouless and Y. Tan, Physica **A177**, 567 (1991).
[13] M. Kohmoto, Phys. Rev. Lett. **51**, 1198 (1983).
[14] C. Tang and M. Kohmoto, Phys. Rev. B **34**, 2041 (1986).
[15] S. C. Bell and R. B. Stinchcombe, J. Phys. A **22**, 717 (1989).
[16] I. M. Suslov, Zh. Eksp. Teor. Fiz. **83**, 1079 (1982) [Sov. Phys. JETP **56**, 612 (1982)].
[17] C. M. Soukouklis and E. N. Economou, Phys. Rev. Lett. **48**, 1043 (1982).
[18] H. Hiramoto and M. Kohmoto, Phys. Rev. Lett. **62**, 2714 (1989).
[19] H. Hiramoto and M. Kohmoto, Phys. Rev. B **40**, 8225 (1989).
[20] B. Helffer and J. Sjöstrand, Mem. Soc. Math. France **40**, 139 (1989).
[21] F. H. Claro and G. H. Wannier, Phys. Rev. B **19**, 6068 (1979).
[22] Y. Hatsugai and M. Kohmoto, Phys. Rev. B **42**, 8282 (1990).
[23] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, L. Procaccia, and B. I. Shraiman, Phys. Rev. A **33**, 1141 (1986).
[24] M. Kohmoto, Phys. Rev. A **37**, 1345 (1988).
[25] M. Wilkinson, J. Phys. A **20**, 4337 (1987).
[26] M. Wilkinson and E. J. Austin, Phys. Rev. B **50**, 1420 (1994).