Instantons and the spectrum of Bloch electrons in a magnetic field

Denise Freed*
Jadwin Hall, Princeton University, Princeton, New Jersey 08544

Jeffrey A. Harvey†
Enrico Fermi Institute, University of Chicago, 5640 Ellis Avenue, Chicago, Illinois 60637-1433

(Received 16 January 1990)

A semiclassical path-integral treatment of the spectrum of a charged particle moving in a two-dimensional periodic potential in the presence of a transverse magnetic field is presented. Complex instanton solutions to the Euclidean equations of motion are identified, and it is shown that the dilute-instanton-gas sum reproduces features of the spectrum studied by previous authors. Comparison of the path-integral results to a perturbative calculation of the splitting of the lowest Landau level is used to relate the self-similar nature of the spectrum to a type of duality in the parameter α, the number of flux quanta per unit cell. As a byproduct of the instanton calculation the generating functional for a sum over lattice paths of given length and algebraic area modulo a fundamental area determined by the magnetic field is evaluated.

I. INTRODUCTION

The quantum-mechanical system of an electron moving in a two-dimensional periodic potential in a magnetic field transverse to the plane of motion has attracted a great deal of interest. It provided an early example of a physical system whose properties depend crucially on the rationality or irrationality of a physical parameter, in this case the ratio of the flux through a unit cell of the lattice to the flux quantum, $\alpha = BA/\Phi_0$ with $\Phi_0 = hc/e$ and A the area of a unit cell. When this ratio is irrational it also provides a simple model of an incommensurate system with the competing periodicities being the lattice periodicity and a periodicity determined by a lattice whose unit cell encloses one flux quantum. Recently it has attracted renewed interest because of its connections to the quantized Hall effect[1] and to the mean-field theory of the Hubbard model.[2] Our own interest in this model was triggered by similarities between instantons in this model and instantons in gauge theory in $2+1$ dimensions with a Chern-Simons term. In both cases the Euclidean functional integral becomes complex and one must search for complex saddle points. In addition, the T violating term in each case modifies the behavior of the solutions at large distances, making perturbation theory suspect.

This problem has been studied using a variety of techniques in different physical regimes. Although instanton techniques have been used to study simpler problems involving tunneling in the presence of a magnetic field,[16] to our knowledge they have not been applied to this general class of problems involving a magnetic field and a periodic potential. We will try to show that path-integral techniques can be successfully applied to this problem and that they shed some new light on the intricate recursive structure of the spectrum.

In the process of working out the instanton sum in this theory we have had to evaluate a sum over lattice paths of the form

$$g(S_x, S_y, e^{-i\epsilon B}) = \sum_{\substack{\text{All 2D lattice paths} \\ \text{with fixed endpoints}}} S_x^{\,l_x} S_y^{\,l_y} e^{-i\epsilon BA}, \qquad (1.1)$$

with l_x, l_y the number of steps in the x and y directions respectively, and A the algebraic area of the path. We expect sums of this form to have other applications in problems of this type, for example, in studying the properties of a crystal of fractional statistics particles.

The outline of this paper is as follows. In Sec. II we will briefly summarize previous work on this problem. In Sec. III we will discuss the path-integral formulation of the problem and the nature of the complex saddle points or instantons. In the limits of zero or infinite magnetic field we give analytic solutions of the instanton equations. In Sec. IV we discuss the dilute-instanton-gas sum. It is here that the rational properties of the magnetic field play a crucial role. We derive an expression for the dilute-instanton-gas sum in terms of a sum over 2D lattice paths weighted by the length of the path and weighted by a phase depending on the area of the path. We show that this can be reduced to a sum over one-dimensional lattice paths and then show that this sum reproduces properties of the spectrum known from previous analyses. In the final section we compare the path integral calculation with a perturbative calculation of the splitting of the lowest Landau level in a regime where both are valid and thereby demonstrate the self-similar nature of the spectrum. We then discuss an alternate path-integral calculation, which treats the potential as a perturbation and gives a Euclidean path-integral version of Pippard's orbit lattice approach.[3] For a particular choice of the potential, a comparison between this description and the original path-integral calculation reveals a duality symmetry between a system with α units of flux per unit cell and one with $1/\alpha$ units of flux. We relate this duality to the self-similar nature of the spectrum and to a path-integral version of Wilkinson's renormalization-group flow for the spectrum.[4] In the

Appendix we compare our results on the structure of the spectrum to those of Wilkinson.

II. SUMMARY OF PREVIOUS WORK

The Hamiltonian we will be concerned with is given by

$$H=\frac{1}{2m}\left[-i\hbar\nabla-\frac{e}{c}\mathbf{A}\right]^2+V(\mathbf{x}) \tag{2.1}$$

with $B=\partial_x A_y-\partial_y A_x$ constant. V is assumed to periodic with $V(\mathbf{x}+\mathbf{l})=V(\mathbf{x})$ where $\mathbf{l}=n\mathbf{e}_1+m\mathbf{e}_2$ and $\mathbf{e}_1,\mathbf{e}_2$ form a basis for the lattice. We will mainly restrict ourselves to a square lattice determined by the potential

$$V=-V_0[\cos(2\pi x/a)+\cos(2\pi y/a)-2]$$

with $V_0\geq 0$ so that $\mathbf{e}_1=a\hat{\mathbf{x}},\mathbf{e}_2=a\hat{\mathbf{y}}$. This choice of potential gives a particularly simple form for the spectrum when calculated perturbatively in V. In the final section we will discuss the extension of our results to more general potentials.

It will be useful in what follows to introduce the dimensionless parameters $\alpha=a^2Be/hc$ and $g^2=(aBe)^2/V_0mc^2$. g is the ratio of the cyclotron frequency to the frequency of oscillation about a minimum of V. In terms of the dimensionless variables $\bar{x}=x/a$ and $\bar{y}=y/a$ the Hamiltonian takes the form

$$H=g^2H_0+V \tag{2.2}$$

with

$$\begin{aligned} H_0/V_0&=-\frac{1}{8\pi^2\alpha^2}(\partial_{\bar{x}}^2+\partial_{\bar{y}}^2)+\frac{\bar{y}^2}{2}-\frac{i}{2\pi\alpha}\bar{y}\partial_{\bar{x}}\ , \\ V/V_0&=-[\cos(2\pi\bar{x})+\cos(2\pi\bar{y})-2]\ , \end{aligned} \tag{2.3}$$

where we have chosen Landau gauge $\mathbf{A}=(-By,0)$. We would like to know the spectrum of H as a function of α and g^2.

For vanishing magnetic fields the answer of course is well known. The Hamiltonian then commutes with the translation operators which generate translations by the basis vectors of the lattice. The spectrum consists of Bloch bands parametrized by the eigenvalues (k_x,k_y) of the translation operators.

For nonzero magnetic fields the situation is more subtle. The basic problem arises from the fact that although V and B are invariant under lattice translations, $\mathbf{A}$ is not, but rather changes by a gauge transformation. In order to construct translation operators that commute with the Hamiltonian one takes[5–7]

$$U_A(\mathbf{l})=\exp\left[\frac{i}{\hbar}\mathbf{l}\cdot\left[\mathbf{p}-\frac{e}{c}\mathbf{A}+\frac{e}{c}\mathbf{B}\times\mathbf{r}\right]\right]\ , \tag{2.4}$$

which satisfies $[U_A(\mathbf{l}),H]=0$. However, the translation operators no longer commute with each other. Instead, there is a net phase upon translation around a unit cell of the lattice:

$$U_A(\mathbf{e}_1)U_A(\mathbf{e}_2)U_A^{-1}(\mathbf{e}_1)U_A^{-1}(\mathbf{e}_2)=e^{2\pi i\alpha}\ . \tag{2.5}$$

For later use we also define $\epsilon=2\pi/\Phi_0=e/\hbar c$ so that the phase factor is $e^{2\pi i\alpha}\equiv e^{i\epsilon a^2B}$. From a mathematical point of view, the presence of a constant magnetic field results in a projective representation of the translation group.

The importance of rational α is now clear. If $\alpha=p/q$ with p,q integer, then we can find a larger magnetic lattice whose unit cell is q times as large and contains p units of flux. The translation operators that generate this lattice then commute with each other as well as with the Hamiltonian. For example, we could take $U_A(q\mathbf{e}_1)$ and $U_A(\mathbf{e}_2)$ as the generators of the magnetic translation group. We can then construct Bloch wave functions on the magnetic lattice, which are eigenfunctions of the Hamiltonian as well as of the magnetic translation operators. It is easy to show that the eigenvalues of $U_A(q\mathbf{e}_1)$ and $U_A(\mathbf{e}_2)$ are given by e^{iqk_1a} and e^{ik_2a} respectively. k_1 and k_2 can be restricted to a magnetic Brillouin zone: $0\leq k_1\leq 2\pi/qa$, $0\leq k_2\leq 2\pi/a$.

We can write the eigenfunctions of the magnetic translation operators in Bloch form as

$$\psi_{k_1,k_2}(x,y)=e^{ik_1x+ik_2y}u_{k_1,k_2}(x,y)\ . \tag{2.6}$$

Working in Landau gauge with $\mathbf{A}=(-By,0)$ the translation operators then take the form

$$\begin{aligned} U_A(q\mathbf{e}_1)&=e^{iqap_x/\hbar}\ , \\ U_A(\mathbf{e}_2)&=e^{iap_y/\hbar+i\epsilon Bax}\ , \end{aligned} \tag{2.7}$$

and the u_{k_1,k_2} must then satisfy the generalized Bloch conditions

$$\begin{aligned} u_{k_1,k_2}(x+qa,y)&=u_{k_1,k_2}(x,y)\ , \\ u_{k_1,k_2}(x,y+a)&=e^{-i\epsilon aBx}u_{k_1,k_2}(x,y)\ . \end{aligned} \tag{2.8}$$

Since $[H,U_A(\mathbf{e}_1)]=0$, the states

$$U_A(n\mathbf{e}_1)\psi_{k_1,k_2}(x,y)=\psi_{k_1,k_2-2\pi\alpha n/qa}(x,y)\ , \qquad n=1\ldots q-1$$

are degenerate in energy.

One simple example of this formalism arises when we consider an "empty" lattice, that is we impose the preceding periodicity conditions but take the potential to be identically zero. The eigenfunctions of the Hamiltonian are then the Landau wave functions given by

$$\psi_{k_x}^{(n)}(x,y)=e^{ik_xx}\phi^{(n)}(y-y_0) \tag{2.9}$$

with energy $E_n=(n+1/2)\hbar\omega_c$. They are also eigenfunctions of $U_A(q\mathbf{e}_1)$ with eigenvalue e^{ik_xqa}. The $\phi^{(n)}$ are the usual harmonic-oscillator wave functions given in terms of Hermite polynomials. Explicitly we have

$$\phi^{(n)}(y-y_0)=N_ne^{-\epsilon B(y-y_0)^2/2}H_n(\sqrt{\epsilon B}\,(y-y_0)) \tag{2.10}$$

with H_n the nth Hermite polynomial, $y_0=-k_x/\epsilon B$, and N_n a normalization factor.

We can construct simultaneous eigenfunctions of $U_A(q\mathbf{e}_1)$, H, and $U_A(\mathbf{e}_2)$ as linear combinations of these Landau functions,

$$\phi_{k_x,k_y}^{(n)}(x,y)=\sum_m e^{-imak_y}U_A(m\mathbf{e}_2)\psi_{k_x}^{(n)}(x,y)\ . \tag{2.11}$$

Writing this out explicitly we have

$$\phi^{(n)}_{k_x,k_y}(x,y)=N_n\exp(ik_x x)\sum_m \exp[ima(\epsilon Bx-k_y)-\epsilon B(y-y_0+ma)^2/2]H_n(\sqrt{\epsilon B}\,(y-y_0+ma)) \ . \tag{2.12}$$

For the lowest Landau level this can be written in terms of the Jacobi Θ function

$$\Theta(z|\tau)=\sum_n e^{i\pi n^2\tau+2\pi inz} \ , \tag{2.13}$$

as

$$\phi^{(0)}_{k_x,k_y}(x,y)=N_0 e^{ik_x x}e^{-\epsilon B(y-y_0)^2/2}\Theta(z|\tau) \tag{2.14}$$

with

$$\begin{aligned}&\tau=i\alpha \ ,\\ &z=\frac{\alpha}{a}[(x-x_0)+i(y-y_0)] \ ,\end{aligned} \tag{2.15}$$

and $x_0=k_y/\epsilon B$.

When $V\equiv 0$ the spectrum is infinitely degenerate. From a group-theoretical point of view this is a consequence of the fact that the operator which generates translations in the y direction commutes with the Hamiltonian but not with the generator of translations in the x direction, so that the Landau wave functions are necessarily mapped onto other degenerate Landau functions by the action of the y translation operator.

The analysis of this problem when both the potential and magnetic field are nonvanishing has occupied a great deal of effort over the last 50 years. Although in most cases of physical interest the shift in energies associated with the magnetic field is small compared with the Fermi energy or the energy gap between bands, it is not possible to solve the problem by doing simple perturbation theory in the magnetic field because a magnetic field of any size changes the qualitative long-range behavior of the wave functions.

Historically the first approach to this problem was based on a tight-binding approximation using the Peierls' substitution.[8] What this amounts to is considering a tight-binding form for the Bloch energy function:

$$E(\mathbf{k})=2E_0[\cos(k_x a)+\cos(k_y a)] \ , \tag{2.16}$$

and then replacing $\hbar\mathbf{k}$ by the operator $\mathbf{p}-e\,\mathbf{A}/c$ to create an operator which is treated as an effective single-band Hamiltonian. This approximation ignores interband mixing and is thus valid only if $\hbar\omega_c \ll E_g$ where E_g is a typical gap energy so that magnetic breakdown can be ignored.[9] In our model this corresponds to assuming that $g^2\ll 2\pi\alpha$. With this substitution the Schrödinger equation for this effective Hamiltonian can be turned into a difference equation. In Landau gauge with $A_x=-By$ this gives

$$\begin{aligned}E_0[&\psi(x,y+a)+\psi(x,y-a)+e^{+i\epsilon Bay}\psi(x+a,y)\\ &+e^{-i\epsilon Bay}\psi(x-a,y)]=E\psi(x,y) \ .\end{aligned} \tag{2.17}$$

With the substitutions

$$x=ma, \quad y=na, \quad \varepsilon=E/E_0 \ , \tag{2.18}$$

and writing

$$\psi(ma,na)=e^{ivm}g(n) \ , \tag{2.19}$$

the Schrödinger equation reduces to a one-dimensional difference equation:

$$g(n+1)+g(n-1)+2\cos(2\pi n\alpha+\nu)g(n)=\varepsilon g(n) \ , \tag{2.20}$$

known as Harper's equation or the almost Mathieu equation. Comparing with the earlier discussion we see that ψ is an eigenfunction of $U_A(q\mathbf{e}_1)$ with $\nu=k_1 a$ but not an eigenfunction of $U_A(\mathbf{e}_2)$. The spectrum of this equation was analyzed by Azbel'[10] using continued fractions and by Hofstadter[11] and Wannier[12] using transfer-matrix techniques. Hofstadter's work emphasized the self-similar structure of the resulting spectrum and made famous the graph of the spectrum of Harper's equation as a function of α reproduced in Fig. 1. One of the remarkable facts about this spectrum is its recursive or self-similar nature. For example, Fig. 2 gives a plot of the p lowest bands in the spectrum for $\alpha=p/q$ between $\frac{1}{4}$ and $\frac{1}{3}$ and denominators q up to 200. Figure 2 clearly resembles Fig. 1 up to an α dependent rescaling. In Sec. V we will discuss the physics behind this self-similar structure.

Remarkably, Harper's equation with the parameter α replaced by $1/\alpha$ arises in a different approach to this problem based on studying the perturbation of the lowest Landau level induced by a weak periodic potential.[13] In the limit of large g^2 the potential term is much smaller than the H_0 term in (2.2) and one can study the spectrum by doing standard first-order degenerate perturbation theory. Writing the energy as

$$E=g^2E_{-1}+E_0+g^{-2}E_1+\cdots \tag{2.21}$$

we have

$$E_{-1}=\hbar\omega_c/2g^2=V_0/2\pi\alpha \ .$$

To calculate E_0 we have to evaluate the matrix elements of

$$\cos(2\pi x/a)+\cos(2\pi y/a)-2$$

between the degenerate Landau wave functions (2.9) labeled by k_x. For a fixed value of k_x, say k_x^0, let C_m be the coefficient of the Landau wave function with $k_x=k_x^0+2\pi m/a$. The $\cos(2\pi y/a)$ term is clearly diagonal in k_x. Its matrix element is proportional to

$$\cos(2\pi y_0/a)=\cos(2\pi m/\alpha+k_x a/\alpha) \ .$$

The $\cos(2\pi x/a)$ term couples Landau wave functions with values of k_x differing by $\pm 2\pi/a$ and thus couples $C_{m\pm 1}$ to C_m. The resulting equations for the coupled Landau wave functions are just Harper's equation (2.20) with the parameter α replaced by $1/\alpha$ and with the parameter ν determined by the initial value of k_x. Varying

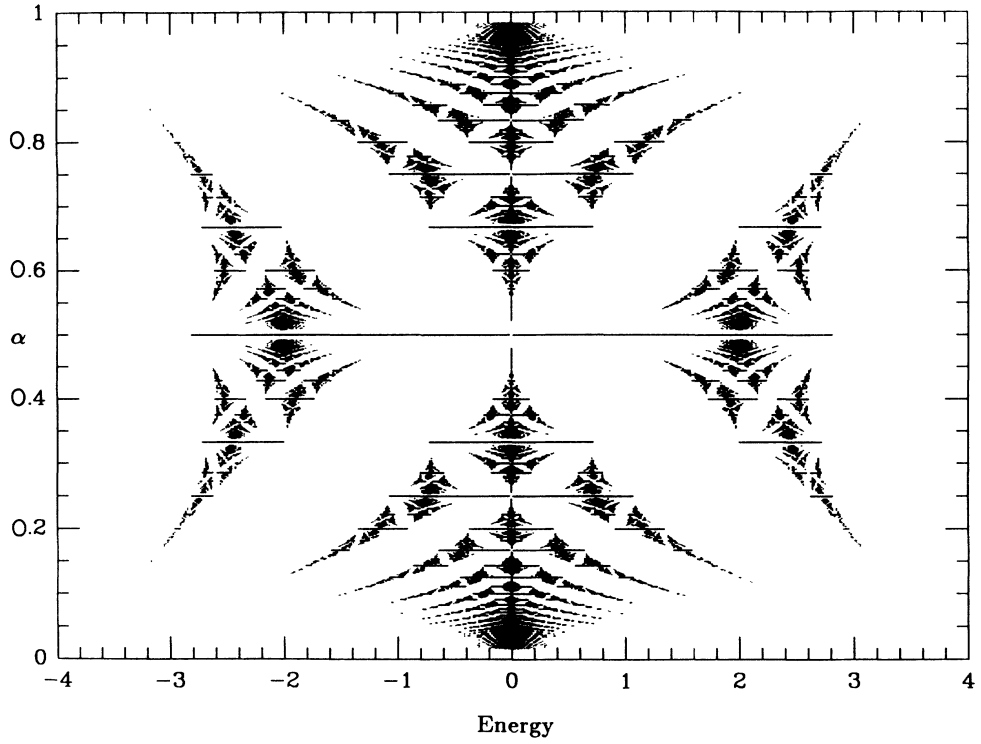

FIG. 1. Spectrum of Harper's equation as a function of α, the number of flux quanta per unit cell, for rational $\alpha=p/q$ and denominators q up to 60.

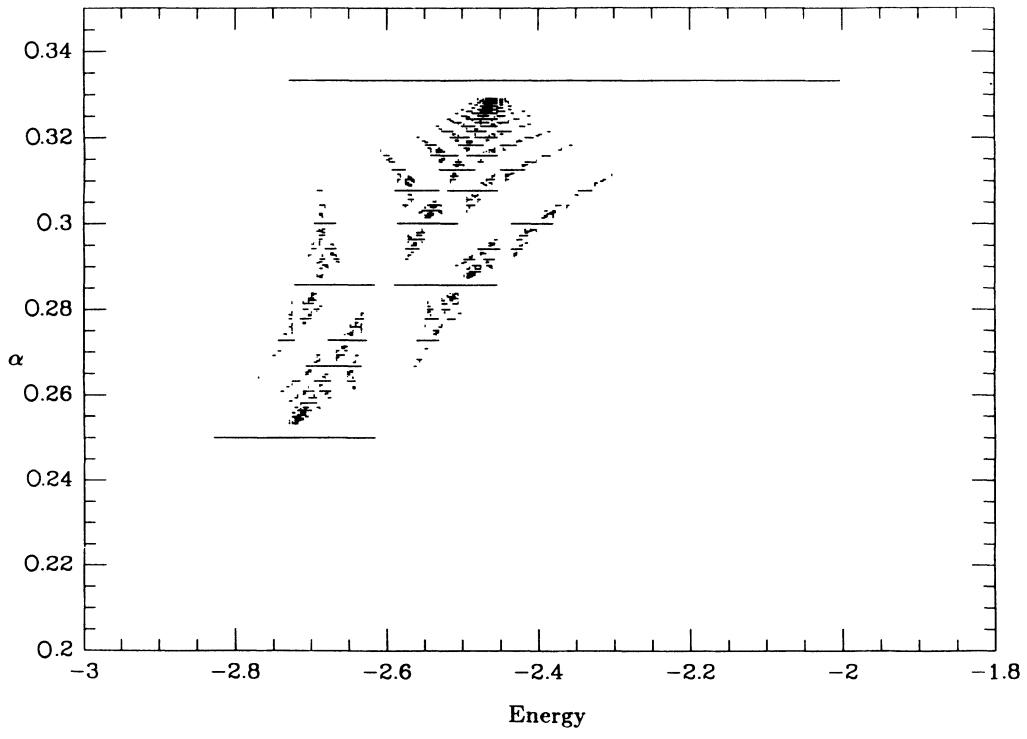

FIG. 2. The p lowest bands in the spectrum of Harper's equation for $\alpha=p/q$ between $\frac{1}{4}$ and $\frac{1}{3}$ and denominators q up to 200.

k_x over the allowed range and evaluating the matrix elements gives

$$E_0^{\text{pert}} = V_0[2 + \tfrac{1}{2} e^{-\pi/2\alpha} \mathcal{H}(1/\alpha)] , \tag{2.22}$$

where $\mathcal{H}(1/\alpha)$ means the spectrum of Harper's equation with parameter $1/\alpha$. In Fig. 3 we plot E_0^{pert} for $1/\alpha$ between 0 and 1 as obtained by appropriately rescaling Fig. 1. The region in the dashed box containing the lowest bands for $1/\alpha$ between $\frac{1}{4}$ and $\frac{1}{3}$ is shown in expanded form in the right-hand plot of Fig. 5 and will be compared in Sec. V to an instanton calculation of the same portion of the spectrum.

It is striking that the same equation with the parameter α replaced by $1/\alpha$ occurs in two different treatments of this problem which apply to different physical regimes. It suggests that there is some sort of duality in the problem and is also reminiscent of the self-similar structure of the spectrum. It is mainly this second point that we wish to explore. We will do this by showing that an instanton calculation of the spectrum at large g^2 (where the tight-binding calculation is not valid) gives a spectrum proportional to $\mathcal{H}(\alpha)$ which is valid for $\alpha = p/q \gg 1$. Demanding that this agree with the perturbative calculation which is also valid for large g^2 and given in terms of $\mathcal{H}(1/\alpha)$ will allow us to determine analytically how the spectrum is mapped into itself. Our analysis is closest to that of Wilkinson[4] who used semiclassical techniques to directly analyze Harper's equation following earlier work by Azbel'.[14] However, our techniques and point of view are somewhat different. We next turn to a discussion of instanton techniques in this problem.

III. PATH INTEGRALS AND INSTANTONS IN A MAGNETIC FIELD

In the following two sections we will evaluate the Euclidean path-integral in the dilute-gas approximation to find the low-lying energy levels. We will first discuss the interpretation of the path integral and then present the form of the instanton solutions. The following section contains the explicit evaluation of the dilute-instanton-gas sum.

Recall the derivation of the path integral for the energies of the lowest band in a one-dimensional periodic potential.[15] Suppose the potential has minima at $x = Na$, N integer, and let $|N\rangle$ be an approximate eigenstate of the Hamiltonian centered around $x = Na$. We construct eigenstates of the translation operator $U = e^{ipa/\hbar}$ as

$$|\theta\rangle = \sum_N e^{-iN\theta} |N\rangle = \sum_N e^{-iN\theta} U^N |0\rangle . \tag{3.1}$$

The states $|\theta\rangle$ will to leading order be simultaneous eigenstates of U and the Hamiltonian H. We thus have

$$\langle \theta | e^{-HT/\hbar} | \theta' \rangle = 2\pi\delta(\theta - \theta') e^{-E_\theta T/\hbar} , \tag{3.2}$$

with θ defined mod2π. In the limit $T \to \infty$ only the energy E_θ of the lowest band will survive.

To write a path-integral expression for the matrix element (3.2) we first use (3.1) to write it as

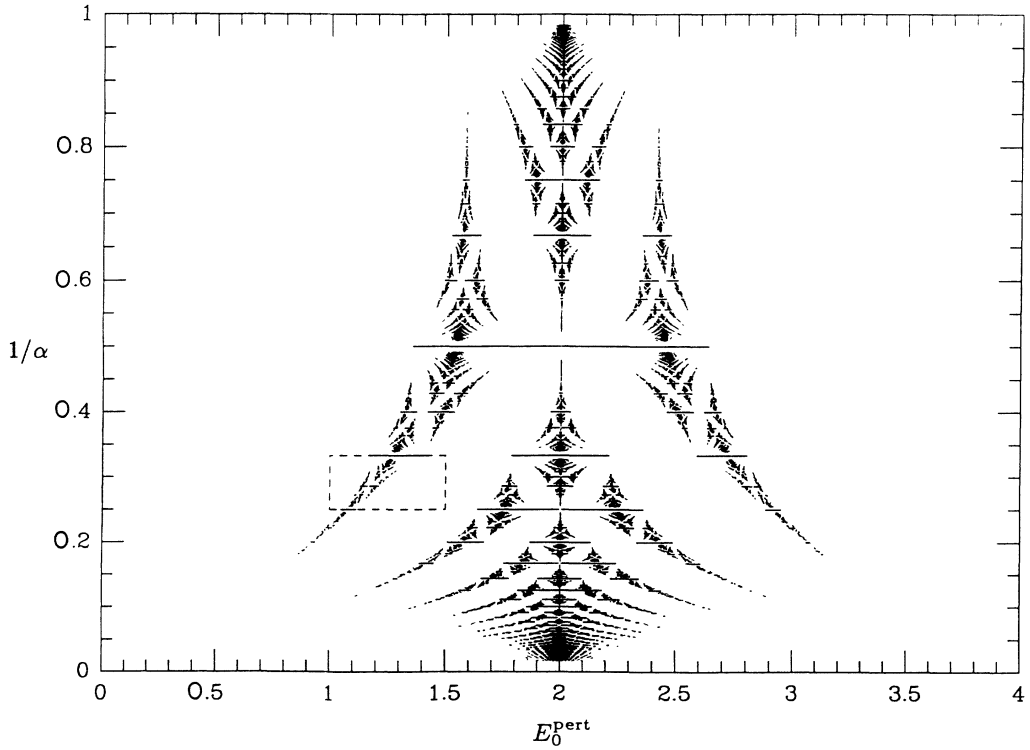

FIG. 3. E_0^{pert} as a function of $1/\alpha$ as obtained by rescaling Fig. 1 according to Eq. (2.22).

$$\sum_N \sum_M e^{-i(N\theta - M\theta')} \langle M|e^{-HT/\hbar}|N\rangle = \sum_{Q_+} \sum_{Q_-} e^{-iQ_+(\theta-\theta')} e^{iQ_-(\theta+\theta')} \langle 2Q_-|e^{-HT/\hbar}|0\rangle \tag{3.3}$$

with $Q_\pm = (M \pm N)/2$. Using

$$\sum_{Q_+} e^{iQ_+(\theta-\theta')} = 2\pi\delta(\theta-\theta')$$

and the usual path-integral expression for the matrix elements $\langle Q|e^{-HT/\hbar}|0\rangle$ we then have

$$\lim_{T\to\infty} e^{-E_\theta T/\hbar} = \lim_{T\to\infty} \sum_Q e^{iQ\theta} \int [\mathcal{D}X]_Q e^{-S_E/\hbar} , \tag{3.4}$$

where $[\mathcal{D}X]_Q$ indicates the path integral over paths with $X(-T/2)=0$ and $X(+T/2)=Qa$. We can then write this as

$$\lim_{T\to\infty} \int \mathcal{D}X e^{-S_E/\hbar + iQ\theta} , \tag{3.5}$$

where the integral now sums over paths starting at 0 and ending at any minima and where $Q=(1/a)\int \dot{x}\, dt$. The semiclassical instanton calculation of (3.5) is given by an unrestricted sum over instantons and anti-instantons in the presence of the "topological" term θQ. Note that we have neglected to insert factors for the initial and final wave functions. This is equivalent to assuming that the states $|N\rangle$ are sufficiently narrow that they can be approximated by position eigenstates.

We now wish to extend this analysis to the case of a two-dimensional periodic potential with a magnetic field to get a path-integral expression for the matrix elements

$$\langle \theta|e^{-HT/\hbar}|\theta'\rangle \tag{3.6}$$

with $|\theta\rangle$ an eigenstate of the translation operators $U_A(q\mathbf{e}_1)$ and $U_A(\mathbf{e}_2)$ with eigenvalues $e^{i\theta_x}$ and $e^{i\theta_y}$ respectively. In the previous notation $\theta_x = qk_1a$ and $\theta_y = k_2a$.

In the absence of tunneling we have approximate position eigenstates localized about the minima which we will denote by pairs of integers, $|\mathbf{n}\rangle \equiv |n_x, n_y\rangle$. We can actually construct q different eigenstates of $U_A(q\mathbf{e}_1)$ and $U_A(\mathbf{e}_2)$ as

$$|\theta, j\rangle = \sum_{N_x, N_y} e^{-i\mathbf{N}\cdot\theta} U_A(N_y\mathbf{e}_2) U_A(N_x q\mathbf{e}_1)|j,0\rangle , \qquad j = 0 \cdots q-1 . \tag{3.7}$$

Now consider the matrix element

$$\langle \theta, j|e^{-HT/\hbar}|\theta', j'\rangle . \tag{3.8}$$

Since H commutes with $U_A(q\mathbf{e}_1)$ and $U_A(\mathbf{e}_2)$ this matrix element will be diagonal in θ, but we expect that because of tunneling there will be mixing between states labeled by different j. The true energy eigenstates will then be of the form

$$|E_i, \theta\rangle = \sum_j |\theta, j\rangle \langle \theta, j|E_i, \theta\rangle , \tag{3.9}$$

and we can thus write (3.8) in the form

$$\sum_i \langle \theta, j|E_i, \theta\rangle \langle E_i, \theta|\theta', j'\rangle e^{-E_i(\theta)T/\hbar} . \tag{3.10}$$

For large T this will be dominated by the q lowest energies consistent with the given values of θ. Multiplying (3.10) by $\langle \theta', j'|E_k, \theta'\rangle$ and summing over j' projects out the energy of the kth band.

To derive a path-integral expression for (3.8) we use the explicit forms for $U_A(q\mathbf{e}_1)$ and $U_A(\mathbf{e}_2)$ to write the right-hand side of (3.7) as

$$\sum_{\mathbf{N}} e^{-i\mathbf{N}\cdot\theta} e^{2\pi i\alpha N_y j} |j + qN_x, N_y\rangle . \tag{3.11}$$

Using this we can write (3.8) in the form

$$\sum_{\mathbf{N}} \sum_{\mathbf{N}'} e^{i\mathbf{N}\cdot\theta - i\mathbf{N}'\cdot\theta'} e^{-2\pi i\alpha(jN_y - j'N_y')} \langle j + qN_x, N_y|e^{-HT/\hbar}|j' + qN_x', N_y'\rangle . \tag{3.12}$$

The matrix element appearing in (3.12) can be written as

$$\langle j + qN_x, N_y|U_A((j'+qN_x')\mathbf{e}_1) U_A(N_y'\mathbf{e}_2) e^{-HT/\hbar}|0,0\rangle = e^{2\pi i\alpha N_y'(j-j')} \langle j - j' + q(N_x - N_x'), N_y - N_y'|e^{-HT/\hbar}|0,0\rangle . \tag{3.13}$$

Rewriting the sums in (3.12) in terms of $\mathbf{Q}=\mathbf{N}-\mathbf{N}'$ and $\mathbf{M}=\mathbf{N}+\mathbf{N}'$ and performing the sum over $\mathbf{M}$ then gives

$$(2\pi)^2\delta^2(\theta-\theta') \sum_{\mathbf{Q}} \exp[iQ_x\theta_x + iQ_y(\theta_y - 2\pi\alpha j)] \langle j - j' + qQ_x, Q_y|e^{-HT/\hbar}|0,0\rangle . \tag{3.14}$$

If we set $j=0$ and sum over j' we can then rewrite this as a path-integral summed over paths starting at (0,0) and ending at any minimum $\mathbf{n}$, with the addition of a "topological" term $\theta\cdot(1/a)\int \dot{\mathbf{x}} dt$ to the action:

$$\sum_{\mathbf{Q}} \int_{\substack{\mathbf{x}(-T/2)=(0,0) \\ \mathbf{x}(T/2)=a\mathbf{Q}}} \mathcal{D}x\, \mathcal{D}y\, e^{-S_E/\hbar + iQ\cdot\theta} . \tag{3.15}$$

In the limit of large T this path integral should by this analysis give (3.10) summed over j' and with j set equal to zero. We thus expect q different energy bands labeled by θ to emerge from the path-integral analysis.

In order to carry out the evaluation of (3.15) we need to determine the form of the instanton solutions and then perform the dilute-instanton-gas sum. The evaluation of the sum is somewhat technical and is presented in the following section. In the rest of this section we discuss the form of the instanton solutions. The Euclidean action is given by

$$S_E=\int dt\left[\frac{m}{2}(\dot{x}^2+\dot{y}^2)+i\frac{eB}{c}\dot{x}y+V(x,y)\right], \tag{3.16}$$

where $V(x,y)$ satisfies

$$V(x+a,y)=V(x,y) \text{ and } V(x,y+a)=V(x,y) . \tag{3.17}$$

In order to discuss the explicit form of the Euclidean action, we will consider the potential

$$V(x,y)=-V_0[\cos(2\pi x/a)+\cos(2\pi y/a)-2] \tag{3.18}$$

with minima at $x=n_x a, y=n_y a$ for n_x and n_y integers. However, the general form of the instanton gas sum only depends on the fact that the potential is periodic in x and y and that, except for the Aharonov-Bohm phase, the action due to the instanton solution is real.

Call the cyclotron frequency $\omega_c=eB/mc$ and the frequency due to the potential $\omega_0=(2\pi/a)\sqrt{V_0/m}$. To write the equations in dimensionless form we first rescale $x/a\to x, y/a\to y$ and introduce a dimensionless time coordinate $\xi=t/\tau$ with $\tau=ma^2\omega_c/V_0=\alpha h/V_0$. The Euclidean action and equations of motion then take a simple form in terms of the dimensionless quantities α and $g^2=\omega_c\tau=(2\pi\omega_c/\omega_0)^2$:

$$\frac{S_E}{\hbar}=2\pi\alpha\int d\xi\left\{\frac{1}{2g^2}\left[\left[\frac{dx}{d\xi}\right]^2+\left[\frac{dy}{d\xi}\right]^2\right]+iy\frac{dx}{d\xi}-[\cos(2\pi x)+\cos(2\pi y)-2]\right\}, \tag{3.19}$$

from which follows the instanton equations of motion

$$\frac{d^2x}{d\xi^2}=g^2\left[-i\frac{dy}{d\xi}+2\pi\sin(2\pi x)\right], \tag{3.20}$$

$$\frac{d^2y}{d\xi^2}=g^2\left[i\frac{dx}{d\xi}+2\pi\sin(2\pi y)\right] . \tag{3.21}$$

Thus the form of the instanton solution depends only on g^2. The Euclidean action depends on both α and g^2 but the α dependence is trivial, just giving an overall rescaling of the action.

We are looking for solutions to the equations of motion that start at the bottom of one well, $(x,y)=(n_x,n_y)$ and travel to the bottom of an adjacent well, where $(x,y)=(n_x\pm1,n_y)$ or $(n_x,n_y\pm1)$. Since the action for these are the same up to a gauge-dependent phase, which for our choice of gauge is acquired only by an instanton traveling in the x direction, we will analyze instantons in the x direction. The magnetic-field terms in equations (3.20) and (3.21) are imaginary, so we cannot find real solutions to the equations of motion. However, for any potential that is even around its minima, like the cosine potential we are considering here, we can look for solutions of the form $x=n_x+x_R(\xi)$ and $y=y_R+iy_I(\xi)$, where $x_R(\xi)$ and $y_I(\xi)$ are both real functions of ξ; $y_R=n_y$; $x_R(\xi)$ runs from 0 to ±1; and $y_I(\xi)$ begins and ends at 0. The instanton equations are then purely real:

$$\frac{d^2x_R}{d\xi^2}=g^2\left[\frac{dy_I}{d\xi}+2\pi\sin(2\pi x_R)\right], \tag{3.22}$$

$$\frac{d^2y_I}{d\xi^2}=g^2\left[\frac{dx_R}{d\xi}+2\pi\sinh(2\pi y_I)\right] . \tag{3.23}$$

The imaginary part of the Euclidean action for such an instanton corresponds to the Aharonov-Bohm phase and reflects the fact that the action is not translationally invariant. It is given by

$$\mathrm{Im}S_E/\hbar=2\pi\alpha y_R\Delta x_R . \tag{3.24}$$

The real part of the action is

$$\mathrm{Re}S_E/\hbar=2\pi\alpha\int_{-\infty}^{+\infty}d\xi\frac{1}{2g^2}\left[\frac{dx_R^2}{d\xi^2}-\frac{dy_I^2}{d\xi^2}\right]-\frac{dx_R}{d\xi}y_I -[\cos(2\pi x_R)+\cosh(2\pi y_I)-2] . \tag{3.25}$$

There are two limits in which it is possible to find analytic solutions to the instanton equations. The first is in the limit of vanishing magnetic field. This corresponds to the limit $\alpha\to0$ and $g^2\to0$ with the ratio α/g held fixed. In terms of the original variables we can take $y_I=0$ and use energy conservation to find

$$\left[\frac{dx}{dt}\right]^2=\frac{2V_0}{m}[1-\cos(2\pi x/a)] \tag{3.26}$$

with the solution

$$x(t)=\frac{2a}{\pi}\tan^{-1}\left[e^{2\pi\sqrt{V_0/ma^2}(t-t_0)}\right], \tag{3.27}$$

where t_0 is the arbitrary midpoint time of the instanton. In this limit the Euclidean action is given by $S_E/\hbar=4(V_0ma^2)^{1/2}/\pi\hbar$.

In Sec. V, we will require the explicit form of the action in the limit that the magnetic field is large compared to the strength of the potential, or as $g^2\to\infty$.[16] The instanton equations in this limit are given by

$$\frac{dy_I}{d\xi}=-2\pi\sin(2\pi x_R) , \tag{3.28}$$

$$\frac{dx_R}{d\xi}=-2\pi\sinh(2\pi y_I) . \tag{3.29}$$

These imply that the potential evaluated along the complex instanton path is constant (and equal to zero for our choice of potential). The real part of the Euclidean ac-

tion is then given by

$$\mathrm{Re}S_E/\hbar=-2\pi\alpha\int dt\frac{dx_R}{dt}y_I=-2\pi\alpha\int_0^1 dx_R y_I(x_R)\ . \tag{3.30}$$

Using

$$\cosh(2\pi y_I)=2-\cos(2\pi x_R)$$

we find

$$\frac{dx_R}{d\xi}=-2\pi[3-4\cos(2\pi x_R)+\cos^2(2\pi x_R)]^{1/2}\ , \tag{3.31}$$

which can be integrated to give the solution

$$\begin{aligned}2\pi x_R&=2\tan^{-1}(\sqrt{2}\sinh 4\pi^2\xi)+\pi\ ,\\ 2\pi y_I&=-\cosh^{-1}(1+2\,\mathrm{sech}8\pi^2\xi)\ .\end{aligned} \tag{3.32}$$

The real part of the Euclidean action is

$$\mathrm{Re}S_E/\hbar=\alpha\int_0^1\cosh^{-1}[2-\cos(2\pi x_R)]dx_R\ . \tag{3.33}$$

Numerical evaluation of the integral gives

$$\mathrm{Re}S_E/\hbar\simeq(\alpha/2\pi)7.327\,72\ .$$

For intermediate values of g^2 it is necessary to use numerical methods to find the instanton solutions. We have done this using relaxation methods, the results are presented in Fig. 4 which gives the real part of the Euclidean action divided by $2\pi\alpha$ as a function of g^2. Note that the action takes its minimum value as g^2 approaches ∞ and increases rapidly for $g^2<1$. Thus the semiclassical limit is valid at large α for all values of g^2.

For any value of the parameters and a general periodic potential, each instanton in the $\pm x$ direction contributes

$$\int dt\, S_x e^{-i2\pi\alpha y_R\Delta x_R} \tag{3.34}$$

and each instanton in the $\pm y$ direction contributes

$$\int dt\, S_y \tag{3.35}$$

where $S_x=K_x e^{-(1/\hbar)S_{Ex}}$ and $S_y=K_y e^{-(1/\hbar)S_{Ey}}$, and S_{Ex} and S_{Ey} are the real parts of the classical Euclidean actions for instantons in the x and y directions, respectively. The factors K_x and K_y are due to the Gaussian integrals obtained by expanding the action around the classical instanton solutions in the x and y directions. More specifically, K_x is proportional to

$$\left[\frac{\det'[D+V''(\mathbf{x}_{\mathrm{inst}})]}{\det[D+V''(0)]}\right]^{-1/2}\ , \tag{3.36}$$

where the prime means that the determinant has been divided by its lowest eigenvalue, and where the determinants are evaluated over functions that vanish at $\pm T/2$. The operators D and $V''(\mathbf{x})$ are given by

$$D=\begin{bmatrix}-\frac{m}{2}\frac{\partial^2}{\partial t^2} & -i\frac{m\omega_c}{2}\frac{\partial}{\partial t}\\ i\frac{m\omega_c}{2}\frac{\partial}{\partial t} & -\frac{m}{2}\frac{\partial^2}{\partial t^2}\end{bmatrix} \tag{3.37}$$

and

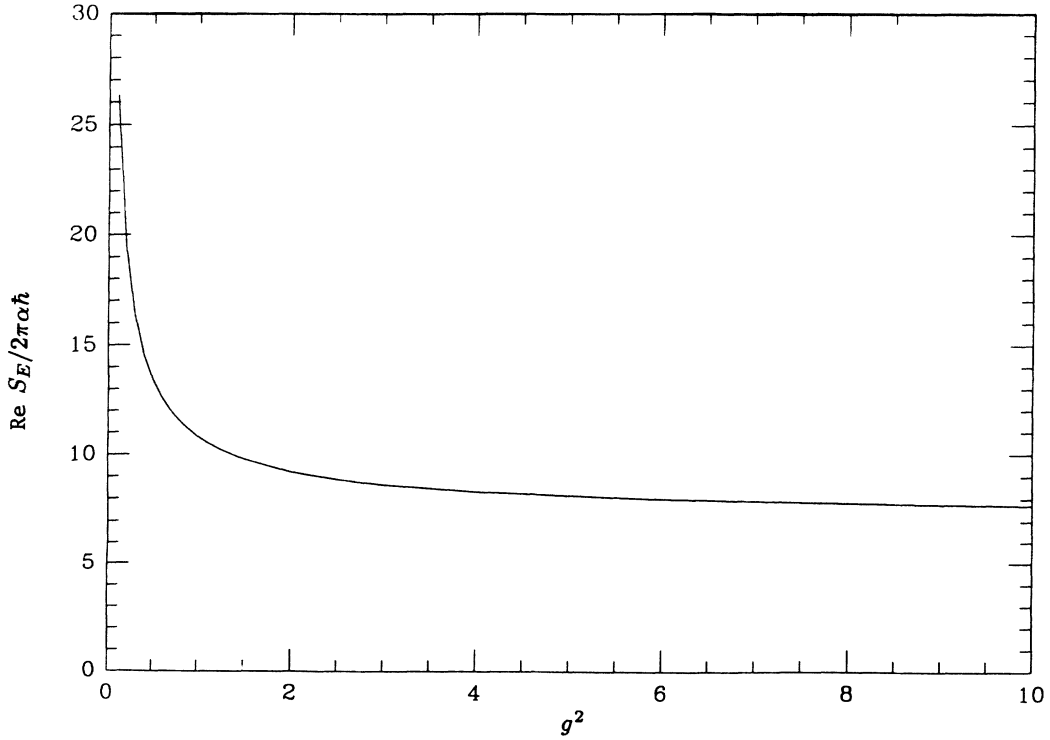

FIG. 4. Euclidean action divided by $2\pi\alpha\hbar$ as a function of g^2 for instantons in a sinusoidal potential.

$$V''(\mathbf{x})=\frac{1}{2}\begin{bmatrix} \dfrac{\partial^2 V}{\partial x^2} & \dfrac{\partial^2 V}{\partial x\,\partial y} \\ \dfrac{\partial^2 V}{\partial y\,\partial x} & \dfrac{\partial^2 V}{\partial y^2} \end{bmatrix}.$$

The constant of proportionality must correct for changing the integral over the zero mode to the integral over time. Once again, the eigenvalue equations, $(D+V'')\mathbf{x}(t)=\lambda\mathbf{x}(t)$, are complex, but for potentials like the simple cosine potential in Eq. (3.18), if we let one of $x(t)$ and $y(t)$ be purely real and the other purely imaginary, then we obtain real symmetric equations. The real determinants can then be solved for numerically by a generalization of the method used for one-dimensional determinants in Refs. 17 and 18.

Because the instanton solutions are complex, one ought to make sure that for each t it is possible to deform the integrals over $x(t)$ and $y(t)$ into the complex plane so that the contours pass through the classical solution, $x_{\text{inst}}(t)$ and $y_{\text{inst}}(t)$ along the direction of steepest descent. In addition, it should be checked that the Gaussian determinant does not pick up any additional minus signs or phases due to the direction of integration. We will not pursue these questions in this paper.

IV. THE DILUTE-GAS SUM

We are now ready to evaluate the path integral for the propagator in the dilute-gas approximation. This is accomplished by summing over all classical paths in which the particle travels from the bottom of one well to the bottom of another along adjoining instanton paths. We assume that the only significant contribution is due to instantons traveling between adjacent wells. Each of these instantons contributes a factor given by equation (3.34) or (3.35). The dilute-gas sum is then

$$\sum_{j'}\langle\boldsymbol{\theta},0|e^{-HT/\hbar}|\boldsymbol{\theta}',j'\rangle=(2\pi)^2\delta^2(\boldsymbol{\theta}-\boldsymbol{\theta}')f(T)$$
$$\times\sum\int_{-T/2}^{T/2}dt_1\int_{-T/2}^{t_1}dt_2\cdots\int_{-T/2}^{t_{M+N-1}}dt_{M+N}\,e^{i\mathbf{Q}\cdot\boldsymbol{\theta}}S_x^M S_y^N\exp\left[-i\epsilon B\sum_{k=1}^{M+N}y_k\Delta x_k\right]. \tag{4.1}$$

The sum is over all instanton paths starting at $\mathbf{x}_0=(0,0)$ and ending at all possible minima $\mathbf{x}_f=a\mathbf{Q}$. For each path, M and N are the total number of instantons in the $\pm x$ and $\pm y$ directions, respectively. $f(T)\propto\det[D+V''(0)]$ gives the correct energy when there is no tunneling. For the potential in Eq. (3.18), $f(T)$ is proportional to $\exp[-\frac{1}{2}(\omega_c^2+4\omega_0^2)^{1/2}T]$ for large T. Lastly, y_k denotes the initial y coordinate of the kth instanton and Δx_k is the distance it travels in the x direction. The integrals over time can immediately be evaluated and give a factor of $T^{M+N}/(M+N)!$.

The main difference between this sum and the usual sum without a magnetic field is that in this case we must keep track not only of the total length of each path, which is proportional to the number of instantons M and N, but also of the total area, $\sum_{i=1}^{M+N}y_i\Delta x_i$, enclosed by each path.

In order to evaluate the dilute gas sum, we must first parametrize the paths. This can be done by dividing the "time" into $N+1$ intervals labeled from 0 to N. Each interval, except the first one, begins with the particle taking one step in the y direction. We define the set of numbers $(n_1,\ldots,n_N)$ with $n_i=\pm1$ so that $an_i\hat{\mathbf{y}}$ denotes the total distance traveled during the ith such step. During the ith interval, the particle takes M_i steps in the x direction. To describe each of these M_i steps we require the set of numbers $(m_{i1},\ldots,m_im_{iM_i})$ with $m_{ij}=\pm1$. Then $am_{ij}\hat{\mathbf{x}}$ is the total distance traveled on the jth step during the ith interval. In terms of these parameters, we can write the total number of steps taken in the x direction as

$$M=\sum_{i=0}^{N}M_i\,, \tag{4.2a}$$

the total distance traveled in the x direction as

$$x_f-x_0=a\sum_{i=0}^{N}\sum_{j=1}^{M_i}m_{ij}\,, \tag{4.2b}$$

the total distance traveled in the y direction as

$$y_f-y_0=a\sum_{i=1}^{N}n_i\,, \tag{4.2c}$$

the total distance traveled in the x direction during the ith interval as

$$\Delta x_i=a\sum_{k=1}^{M_i}m_{ik}\,, \tag{4.2d}$$

and the y component of the particle's position during the ith interval as

$$y_i = y_0 + a\sum_{j=1}^{i} n_j \ . \tag{4.2e}$$

The total area enclosed by the path is given by

$$\sum_{i=0}^{N} y_i \Delta x_i = a^2 \sum_{i=0}^{N}\sum_{k=1}^{M_i} m_{ik}\left(\sum_{j=1}^{i} n_j\right) , \tag{4.3}$$

where $\sum_{j=1}^{0} n_j$ is defined to be 0. We will need the expression $\exp(-i\epsilon B\sum y_i \Delta x_i)$, which we can write as

$$\prod_{i=0}^{N}\prod_{k=1}^{M_i} \exp\left[-i2\pi\alpha m_{ik}\sum_{j=1}^{i} n_j\right] . \tag{4.4}$$

With this parametrization, the instanton-gas sum becomes

$$f(T)\sum_{N=0}^{\infty}\sum_{M_0=0}^{\infty}\cdots\sum_{M_N=0}^{\infty}\sum_{\substack{(n_1,\ldots,n_N)\\ n_i=\pm 1}}\prod_{t=0}^{N}\prod_{l=0}^{M_t}\left[\sum_{m_{tl}=\pm 1}\right]\left[\exp\left[i\theta_x\sum_{ik} m_{ik} + i\theta_y\sum_{j} n_j\right]\right.$$
$$\left.\times S_x^M S_y^N \frac{T^{M+N}}{(M+N)!}\prod_{i=0}^{N}\prod_{k=1}^{M_i}\exp\left[-i2\pi\alpha m_{ik}\sum_{j=1}^{i} n_j\right]\right] . \tag{4.5}$$

To facilitate doing this sum, we introduce a δ function in the form

$$1 = \sum_{L=0}^{\infty}\delta_{L,M+N} = \frac{1}{2\pi}\sum_{L=0}^{\infty}\int_0^{2\pi} d\nu \exp\left[i\nu\left(L - N - \sum_i M_i\right)\right] . \tag{4.6}$$

Collecting all the terms that depend on N and M_i only in the combination $L = N + \sum_i M_i$, we can sum $\exp(i\nu L)T^L/L!$ over L to obtain the following form for the propagator:

$$f(T)\frac{1}{2\pi}\int_0^{2\pi} d\nu\, e^{Te^{i\nu}} G(\theta_x,\theta_y,\nu) , \tag{4.7}$$

where

$$G(\theta_x,\theta_y,\nu) = \sum_{N=0}^{\infty}\sum_{M_0=0}^{\infty}\cdots\sum_{\substack{M_N=0\\ n_i=\pm 1}}^{\infty}\sum_{(n_1,\ldots,n_N)}\prod_{t=0}^{N}\prod_{l=0}^{M_t}\left[\sum_{m_{tl}=\pm 1}\right]\left\{\exp\left[i\theta_y\sum_j n_j\right]\right.$$
$$\times\exp\left[-i\nu\left[N+\sum_i M_i\right]\right]S_x^M S_y^N$$
$$\left.\times\prod_{i=0}^{N}\prod_{k=1}^{M_i}\exp\left[-im_{ik}\left[2\pi\alpha\sum_{j=1}^{i} n_j - \theta_x\right]\right]\right\}$$
$$= \sum_{\mathbf{x}_f}\exp\left[i\boldsymbol{\theta}\cdot\frac{(\mathbf{x}_f-\mathbf{x}_0)}{a}\right]\sum_{\substack{\text{All 2D lattice}\\ \text{paths with}\\ \text{endpoints } \mathbf{x}_0,\mathbf{x}_f}} e^{-i\nu M}S_x^M e^{-i\nu N}S_y^N e^{-i\epsilon BA} , \tag{4.8}$$

where $\mathbf{x}_0 = (0,0)$. Now the sums over the m_{ik} and M_i can be easily performed. The first just give $\cos(2\pi\alpha\sum_{j=1}^{i} n_j - \theta_x)$ and then the second is a geometric series. We are then left only with the sums over N and $(n_1,\ldots,n_N)$, so what remains is really a sum over one-dimensional random walks. This sum is

$$G(\theta_x,\theta_y,\nu) = \sum_{N=0}^{\infty}\sum_{\substack{(n_1,\ldots,n_N)\\ n_i=\pm 1}} e^{-i\nu N}S_y^N \exp\left[i\theta_y\sum_{i=1}^{N} n_i\right]\prod_{i=0}^{N}\left[1 - 2S_x e^{-i\nu}\cos\left[2\pi\alpha\sum_{j=1}^{i} n_j - \theta_x\right]\right]^{-1} . \tag{4.9}$$

At this point it is important to remember that the argument of the cosine, $2\pi\alpha\sum_{j=1}^{i} n_j - \theta_x$, is defined only modulo 2π. When $2\pi\alpha = 2\pi p/q$, this implies that we only need to know the values of $\sum_{j=1}^{i} n_j \bmod q$. The fact that $\sum_{j=1}^{i} n_j$ takes on only the q distinct values of $j = 0, 1, \ldots, q-1 \pmod q$ will lead to the existence of the q different energy bands.

When α is irrational, $\sum_{j=1}^{i} n_j$ can be any integer from zero to infinity, which is equivalent to setting $q \to \infty$.

$\sum_{j=1}^{i} n_j$ is the net distance traveled after the first i steps of the one-dimensional walk. If we number each site of the one-dimensional lattice consecutively, starting with zero at the particle's initial position, then for each path of N steps we can define l_j as the number of times the particle lands on a site that is congruent to j mod$q$, and the net distance traveled as

$$\Delta = \sum_{i=1}^{N} n_i \ . \tag{4.10}$$

With these definitions, the total number of steps is $N = \sum_{j=0}^{q-1} l_j$ and the sum can be expressed as

$$G = \sum_{N=0}^{\infty} \sum_{(n_1,\ldots,n_N)} \frac{e^{i\theta_y \Delta}}{1-2S_x e^{-i\nu}\cos\theta_x} \prod_{j=0}^{q-1} \left[\frac{e^{-i\nu l_j} S_y^{l_j}}{[1-2S_x e^{-i\nu}\cos(2\pi\alpha j - \theta_x)]^{l_j}} \right] . \tag{4.11}$$

to simplify the equation for G, we set

$$t_j = \frac{e^{-i\nu} S_y}{1-2S_x e^{-i\nu}\cos(2\pi\alpha j - \theta_x)} \ , \tag{4.12}$$

so that G has the form

$$G(\theta_x,\theta_y,\nu) = \frac{1}{1-2S_x e^{-i\nu}\cos\theta_x} \sum_{N=0}^{\infty} \sum_{\text{paths}} e^{i\theta_y \Delta} \prod_{j=0}^{q-1} t_j^{l_j} \ , \tag{4.13}$$

where the second sum is over all random walks of length N starting at zero. Thus, we want to sum the expression $e^{i\theta_y \Delta} \prod_{j=0}^{q-1} t_j^{l_j}$ over all paths.

For a particular path, the summand, $e^{i\theta_y \Delta} \prod_{j=0}^{q-1} t_j^{l_j}$, can be found recursively. For example, suppose that on the $(M+1)$st step the particle lands on a site that is numbered k mod$q$. This is possible only if after the Mth step it was on a $(k-1)$ mod$q$ site or a $(k+1)$ mod$q$ site. In the first case, the particle moves forward during the $(M+1)$st step, so the distance traveled, Δ, increases by one unit. As a result, the summand gains a factor of $e^{i\theta_y}$. In addition, because the particle has landed on a k mod$q$ site, l_k increases by one, so the summand also picks up a factor of t_k. Similarly, in the second case the summand gains a factor of $e^{-i\theta_y} t_k$.

More generally, we define the q-dimensional vector $\tilde{F}(N)$ as

$$F_j(N) = \sum_{\text{paths}} \prod_{k=0}^{q-1} (t_k)^{l_k} (e^{i\theta_y})^{\Delta}, \quad j = 0,1,\ldots,q-1 \ , \tag{4.14}$$

where the sum is over all paths of length N starting on site zero and ending on a j mod$q$ site. Then $F_k(N+1)$ is obtained as follows:

$$F_k(N+1) = t_k e^{-i\theta_y} F_{k+1}(N) + t_k e^{i\theta_y} F_{k-1}(N)$$

$$\equiv A_{ki} F_i(N) \ . \tag{4.15}$$

Because this expression is valid for any N, we can write the vector $\tilde{F}(N)$ as

$$\tilde{F}(N) = A^N \tilde{F}(0) \ , \tag{4.16}$$

where $\tilde{F}(0)$ describes the starting configuration. In the case we are studying, the particle starts on the zero mod$q$ site with an accompanying factor of $1/(1-2S_x e^{-i\nu}\cos\theta_x)$. Thus, $\tilde{F}(0)$ is given by

$$\tilde{F}(0) = \frac{1}{1-2S_x e^{-i\nu}\cos\theta_x} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} . \tag{4.17}$$

The $q \times q$ matrix A is given by

$$A_{jk} = t_j e^{-i\theta_y} \delta_{j+1,k} + t_j e^{i\theta_y} \delta_{j-1,k} \ , \tag{4.18}$$

where the argument of the δ function must be evaluated mod$q$. We want the sum over all paths starting at zero, not just over paths with fixed length and fixed end points, so we must sum $F_j(N)$ over all N and j. Thus we obtain the following expression for G:

$$G = \sum_{N=0}^{\infty} \sum_{j=0}^{q-1} F_j(N) = \sum_{N=0}^{\infty} (1,1,\ldots,1) A^N \times \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \frac{1}{1-2S_x e^{-i\nu}\cos\theta_x} \ , \tag{4.19}$$

or, after performing the sum,

$$G = \frac{1}{1-2S_x e^{-i\nu}\cos\theta_x} (1,1,\ldots,1)(I-A)^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} . \tag{4.20}$$

It will be convenient to write the t_j in terms of $z = e^{i\nu}$ and to write $(I-A)^{-1}$ in terms of a Hermitian matrix H which is independent of z. Thus, we define

$$r_j = 2S_x \cos(2\pi\alpha j - \theta_x) \ ,$$

so that

$$t_j=\frac{S_y}{z-r_j} . \tag{4.21}$$

In addition, let

$$Iz-H=\begin{bmatrix} z-r_0 & & & \\ & z-r_1 & & \\ & & \ddots & \\ & & & z-r_{q-1} \end{bmatrix}[I-A] . \tag{4.22}$$

Then

$$H_{jk}=r_j\delta_{jk}+S_y e^{-i\theta_y}\delta_{j+1,k}+S_y e^{i\theta_y}\delta_{j-1,k}$$

is the usual matrix for the energy eigenvalues in the tight-binding model. Using the standard expression for the inverse of a matrix, we can write

$$(1,1,\ldots,1)(I-A)^{-1}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}=\sum_j (I-A)^{-1}_{j0}$$
$$=\frac{\sum_j(-1)^j\det(I-A)^0_j}{\det(I-A)} , \tag{4.23}$$

where M^i_j denotes the matrix obtained by deleting the ith row and jth column of M. It follows from this identity that

$$\sum_{j'}\langle\theta',0|e^{-HT/\hbar}|\theta',j'\rangle=(2\pi)^2\delta^2(\theta-\theta')f(T)\sum_{\substack{z_p \text{ such that}\\ \det(Iz_p-H)=0}} e^{Tz_p}\frac{\sum_{j=0}^{q-1}\det(Iz_p-H)^0_j}{\partial_z\det(Iz_p-H)} . \tag{4.28}$$

Comparing with Eq. (3.10), we find that the energy eigenvalues are labeled by θ_x, θ_y, and p, and that the shift in the energy eigenvalues due to the tunneling is given by

$$E=-\hbar z_p, \quad \text{where } \det(Iz_p-H)=0 . \tag{4.29}$$

Thus, the energy levels are the eigenvalues of the matrix $-H$. As we will see in the following section, this is essentially Harper's equation.

We can also use the sum performed here to obtain an expression for the generating function

$$g=\sum_{\substack{\text{All 2D lattice paths}\\ \text{with fixed endpoints}}} S_x^M S_y^N e^{-i\epsilon BA} , \tag{4.30}$$

where aM and aN are the path's total length in the x and y directions, $A=\sum_i y_i\Delta x_i$ is the area enclosed by each path on the lattice, and $\mathbf{x}_0$ and $\mathbf{x}_f$ are the initial and final endpoints. In fact, $G(\theta_x,\theta_y,0)$, as defined in Eq. (4.8), is the Fourier transform of g, so g can be written as

$$\sum_j (I-A)^{-1}_{j0}=\frac{1}{z-r_0}\sum_j (Iz-H)^{-1}_{j0} . \tag{4.24}$$

Therefore, in terms of H and z, G is given by

$$G=z\sum_j (Iz-H)^{-1}_{j0} . \tag{4.25}$$

Substituting this expression for G and the integral over ν into the equation for the propagator, we obtain

$$\sum_{j'}\langle\theta,0|e^{-HT/\hbar}|\theta',j'\rangle=(2\pi)^2\delta^2(\theta-\theta')f(T)\times\int_0^{2\pi}d\nu\frac{e^{i\nu}}{2\pi}e^{Te^{i\nu}}\times\sum_{j=0}^{q-1}(Ie^{i\nu}-H)^{-1}_{j0} . \tag{4.26}$$

Next we write $z=e^{i\nu}$ and convert the integral over ν into the following contour integral:

$$\frac{1}{2\pi i}\oint_{\text{unit circle}}dz\, e^{Tz}\sum_j (Iz-H)^{-1}_{j0} . \tag{4.27}$$

The poles in the integrand are the poles of

$$\sum_j (Iz-H)^{-1}_{j0}=\left[\sum_{J=0}^{q-1}(-1)^j\det(Iz-H)^0_j\right]/\det(Iz-H) ,$$

which is a ratio of polynomials in z. Therefore, they occur at the z_p for which $\det(Iz_p-H)=0$. Summing the residues over all the poles, we obtain the final form of the propagator.

$$g(S_x,S_y,e^{-i\epsilon B})=\int_0^{2\pi}d\theta_x\int_0^{2\pi}d\theta_y\, A(\mathbf{x}_f,\mathbf{x}_0)G(\theta_x,\theta_y,0) , \tag{4.31}$$

where

$$A(\mathbf{x}_f,\mathbf{x}_0)=\frac{1}{4\pi^2}\exp -i\left[\theta_x\frac{(x_f-x_0)}{a}+\theta_y\frac{(y_f-y_0)}{a}+\epsilon B y_0(x_f-x_0)\right] . \tag{4.32}$$

Taking $\nu=0$ in the final equation (4.26) for G, we find the following expression for g:

$$g=\int_0^{2\pi}d\theta_x\int_0^{2\pi}d\theta_y\, A(\mathbf{x}_f,\mathbf{x}_0)\sum_j (I-H)^{-1}_{j0} . \tag{4.33}$$

Thus we have found an expression for g, which is the generating function for the number of paths on the lattice which have fixed end points, which enclose a fixed area modulo $2\pi/\epsilon B$, and which have fixed length in the x and y directions.

V. SELF-SIMILARITY AND RENORMALIZATION GROUP FLOW

To summarize the results of the preceding section, the energy levels found in the dilute-instanton-gas approximation are given by

$$E=E_c+\mathcal{H}(S_x,S_y,\alpha)\ , \tag{5.1}$$

where E_c is the ground-state energy of the particle in the absence of tunneling and $\mathcal{H}(S_x,S_y,\alpha)$ is the shift in the energy due to the tunneling. For the symmetric cosine potential in Eq. (3.18), E_c is given by

$$E_c=\frac{\hbar}{2}(\omega_c^2+4\omega_0^2)^{1/2}+\Delta E_c\ , \tag{5.2}$$

where ΔE_c is the shift in energy obtained by treating

$$V(x,y)-[m\omega_0^2(x^2+y^2)/2-2V_0]$$

as a perturbation to the harmonic well. The first-order term, proportional to V_0, in ΔE_c is

$$-2V_0[\exp(-\pi/\alpha)-1+\pi/\alpha]\ .$$

$\mathcal{H}(S_x,S_y,\alpha)$ stands for the eigenvalues of the equation

$$S_y e^{-i\theta_y}g(m+1)+S_y e^{i\theta_y}g(m-1) +2S_x\cos(2\pi\alpha m-\theta_x)g(m)=Eg(m)\ , \tag{5.3}$$

with $g(m+q)=g(m)$. In going from the determinant equation (4.29) for the eigenvalues to this difference equation, we have used the fact that

$$\mathcal{H}(S_x,S_y,\alpha)=-\mathcal{H}(S_x,S_y,\alpha)$$

when θ_y and θ_x are suitably redefined. If we let $h(m)=e^{-i\theta_y m}g(m)$, then the eigenvalue equation becomes

$$S_y h(m+1)+S_y h(m-1)+2S_x\cos(2\pi\alpha m-\theta_x)h(m) =Eh(m)\ , \tag{5.4}$$

with the new boundary conditions

$$h(m+q)=e^{-iq\theta_y}h(m)\ . \tag{5.5}$$

Equation (5.4) is the usual form of Harper's equation. For a symmetric potential, $S_x=S_y$, and the energy shift is proportional to $S_x=\hbar Ke^{-S_E/\hbar}$, so we will write

$$\mathcal{H}(S_x,S_y,\alpha)=\hbar Ke^{-S_E/\hbar}\mathcal{H}(\alpha)\ .$$

In what follows, we will restrict ourselves to the symmetric potential of Eq. (3.18).

In the limit $g^2\to\infty$ we expect the answer obtained using the action in Eq. (3.33) to be correct to zeroth order in $1/g^2$. In addition, the Euclidean action is proportional to α, so in the limit of large α, the semiclassical approximation is valid. Ignoring the effect of the phases, we expect the dilute gas approximation to be valid when the reciprocal of the density of instantons, given by $1/(Ke^{-S_E/\hbar})$, is much greater than the time it takes for an instanton to tunnel, or when

$$1/(Ke^{-S_E/\hbar})\gg\tau=\frac{\alpha h}{V_0}\ . \tag{5.6}$$

Assuming that K is not an exponential in α, we see that $1\gg\tau Ke^{-S_e/\hbar}$ is automatically satisfied for large α, since $S_E\approx 7.3\alpha/2\pi$. Therefore, for large α and g^2, the path-integral calculation should be valid to zeroth order in $1/g^2$. In this limit, the energy levels are given by

$$E^{\text{inst}}=\frac{\hbar}{2}\omega_c\left(1+2\frac{\omega_0^2}{\omega_c^2}\right)+\hbar Ke^{-\alpha\gamma/2\pi}\mathcal{H}(\alpha) =V_0\left[\frac{g^2}{4\pi\alpha}+\frac{2\pi}{\alpha}\right]+\hbar Ke^{-\alpha\gamma/2\pi}\mathcal{H}(\alpha)\ , \tag{5.7}$$

where we have set

$$\gamma=\int_0^{2\pi}\cosh^{-1}[2-\cos(x)]dx\simeq 7.3277\ . \tag{5.8}$$

However, the calculations described in Sec. II which treated the potential as a perturbation of the lowest Landau level and led to (2.22) are also valid for large α and g^2, to the same order in g^2. These calculations gave energies

$$E^{\text{pert}}=\frac{\hbar}{2}\omega_c+2V_0+\frac{V_0}{2}e^{-\pi/2\alpha}\mathcal{H}\left[\frac{1}{\alpha}\right]\ . \tag{5.9}$$

These two expressions for the energy levels do not immediately appear to be consistent. One predicts energy splittings that depend on α and therefore appear to be nonperturbative in $\hbar$ since $\alpha\propto 1/\hbar$. The other depends on $1/\alpha$ and so appears to be a perturbative result. Because the solutions to $\mathcal{H}(\alpha)$ are found by diagonalizing a q by q matrix when $\alpha=p/q$, in the first case we expect q bands and in the second case we expect p bands. We can reconcile these two results by noting, as others have done,[19,10,11] that in the region where $\alpha>1$, we also have $p>q$, so the instanton sum could be calculating the q lowest bands in the perturbation calculation. This is further supported by the facts that the leading term to the energy for both spectra is $(\hbar/2)\omega_c$ and that the width of the splitting in the instanton calculation, $\hbar Ke^{-\alpha\gamma/2\pi}$, is exponentially smaller than the width of the splitting in the perturbation calculation, $(V_0/2)e^{-\pi/2\alpha}$, so the first spectrum could be contained inside the second. Inspection of Fig. 1 now viewed as the perturbative spectrum as a function of $1/\alpha$ also shows that the q lowest bands of the spectrum as $1/\alpha$ varies between say $1/N$ and $1/(N+1)$ contains a rescaled version of the whole spectrum which is consistent with the fact that α varies from N to $N+1$ in this range. In an Appendix we demonstrate that this picture is justified and that our answer agrees with a previous analysis of the spectrum by Wilkinson.[4] The fact that (5.7) and (5.9) must agree for the q lowest bands also gives an analytical formula for how the spec-

trum is mapped into itself. We will return to this question momentarily.

We can also use path-integral techniques to see that there are dual pictures of the system when $g^2 \gg 1$ and $\alpha \gg 1$. It can either be viewed as in Sec. IV as a particle tunneling between potential wells spaced a apart with tunneling time $\approx \tau$, or as a particle in Landau orbits with period $2\pi/\omega_c$ whose centers jump from site to site in a lattice with spacing a/α.

To obtain the second picture we treat the potential

$$V_x \cos(2\pi/a) + V_y \cos(2\pi/a)$$

as a perturbation to the lowest Landau level using path integral methods. In performing this calculation we find a sum similar to the dilute gas sum in Eq. (4.1), but with α replaced by $1/\alpha$. To obtain such a sum, we first expand out the exponential of the potential in the propagator:

$$\exp\frac{1}{\hbar}\int_{-T/2}^{T/2}\left[V_x\cos\frac{2\pi x}{a}+V_y\cos\frac{2\pi y}{a}\right]dt=\sum_{n=0}^{\infty}\frac{1}{n!}\left[\int_{-T/2}^{T/2}\left[\frac{V_x}{\hbar}\cos\frac{2\pi x(t_j)}{a}+\frac{V_y}{\hbar}\cos\frac{2\pi y(t_j)}{a}\right]dt_j\right]^n . \tag{5.10}$$

Because the cosine is a sum of exponentials, we can express it as a contribution to the action:

$$\int_{-T/2}^{T/2}\frac{V_x}{\hbar}\cos\frac{2\pi x(t_j)}{a}dt_j=\sum_{s=0}^{1}\int_{-T/2}^{T/2}\frac{V_x}{2\hbar}\exp\left[\frac{1}{\hbar}\int i\frac{2\pi\hbar}{a}x(t)(-1)^s\delta(t-t_j)dt\right]dt_j . \tag{5.11}$$

Thus, the potential acts like a linear source term in which the particle interacts with a δ-function current. The general form for the action is

$$S=\int[L_0(t)+\mathbf{J}(t)\cdot\mathbf{x}(t)]dt , \tag{5.12}$$

where

$$L_0(t)=\frac{m}{2}(\dot{x}^2+\dot{y}^2)+\frac{ieB}{c}y\dot{x} \tag{5.13}$$

and

$$\mathbf{J}(t)=i\sum_{j=1}^{n}(-1)^{s_j}\frac{2\pi\hbar}{a_j}\hat{\mathbf{x}}_j(t)\delta(t-t_j) . \tag{5.14}$$

The propagator is then

$$G=e^{-T(V_x+V_y)/\hbar}\sum_J\int Dx\,Dy e^{-S/\hbar} , \tag{5.15}$$

where

$$\sum_J=\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\left[\prod_{j=0}^{n_1+n_2}\sum_{s_j=0}^{1}\right]\int_{-T/2}^{T/2}dt_1\cdots\int_{-T/2}^{T/2}dt_{n_1+n_2}\left[\frac{V_x}{2\hbar}\right]^{n_1}\left[\frac{V_y}{2\hbar}\right]^{n_2}/(n_1+n_2)! \tag{5.16}$$

with $n=n_1+n_2$. The sum over currents J is like a sum over two-dimensional random walks with steps of size one taken at arbitrary times and weighted by a factor of $V_x/2\hbar$ for a step in the x direction and $V_y/2\hbar$ for one in the y direction.

The equations of motion for the Lagrangian in Eq. (5.12) are

$$m\ddot{x}+i\frac{eB}{c}\dot{y}=J_x(t)$$
$$=0 \quad \text{unless } t=t_j \text{ and } \hat{\mathbf{x}}_j=\hat{\mathbf{x}} , \tag{5.17}$$

$$m\ddot{y}-i\frac{eB}{c}\dot{x}=J_y(t)$$
$$=0 \quad \text{unless } t=t_j \text{ and } \hat{\mathbf{x}}_j=\hat{\mathbf{y}} . \tag{5.18}$$

From these it follows that, except when $t=t_j$, we have two constants of motion,

$$y-\frac{i}{\omega_c}\dot{x}=y_0 \quad \text{and} \quad x+\frac{i}{\omega_c}\dot{y}=x_0 . \tag{5.19}$$

When $t=t_j$, either x_0 or y_0 changes by an amount found by integrating the equations of motion over a small interval around t_j. If $\hat{\mathbf{x}}_j=\hat{\mathbf{y}}$, this change is

$$\Delta x_0|_{t=t_j}=(-1)^{s_j+1}\frac{2\pi\hbar c}{eBa}=(-1)^{s_j+1}\frac{a}{\alpha} , \tag{5.20}$$

and if $\hat{\mathbf{x}}_j=\hat{\mathbf{x}}$, it is

$$\Delta y_0|_{t=t_j}=(-1)^{s_j}\frac{2\pi\hbar c}{eBa}=(-1)^{s_j}\frac{a}{\alpha} . \tag{5.21}$$

In Minkowski space, the equations of motion imply that the particle is in a Landau orbit with frequency ω_c centered around $(x_0,y_0)=((a/\alpha)n,(a/\alpha)m)$. When the potential acts via the delta function in J, the center of the orbit hops to an adjacent site in the lattice

$(a/\alpha)n\hat{x}+(a/\alpha)m\hat{y}$. This is the picture used by Pippard.[3] In Euclidean space this corresponds to the particle sitting on the lattice site

$$(x_0,y_0)=((a/\alpha)n,(a/\alpha)m)$$

except for periods of time on the order of $1/\omega_c$ when the particle tunnels to an adjacent site. Therefore, we can treat this as an instanton problem again, but now the particle tunnels to sites spaced a/α apart instead of a apart. The action for such an instanton in a square lattice is

$$\frac{S_E}{\hbar}=i\frac{2\pi}{\alpha}\left[\frac{\alpha^2}{a^2}x_j\Delta y_j\right]+\frac{\pi}{2\alpha}\ . \tag{5.22}$$

Each time the potential "acts," the $e^{-S/\hbar}$ picks up another factor of $e^{-S_E/\hbar}$, so the expression for the energy levels arising from the propagator in Eq. (5.15) looks identical to the dilute-gas sum, Eq. (4.1), except that α is replaced by $1/\alpha$ and x and y are interchanged. This sum produces the same spectrum as the one obtained by perturbing the lowest Landau level in the Schrödinger equation. [However, the full expression for the propagator, which takes into account the gaussian integrals around S_E and the form of the wave functions, is much more complicated than Eq. (4.1).]

In this way we see that there are dual path-integral pictures of this system. In the first picture the centers of the orbits lie in a lattice whose unit cell encloses $1/\alpha$ units of flux and in the second picture the minima of the wells form a lattice whose unit cell encloses α units of flux. We have argued that the path-integral calculation of the second picture produces a spectrum which is a subset of the spectrum from the first. This calculation is the first step in deriving part of the self-similar structure of the spectrum and Wilkinson's renormalization group for Harper's equation.

We start with the spectrum in Eq. (5.9)

$$E^{\text{pert}}=\frac{\hbar}{2}\omega_{c0}+2V_0+\frac{V_0}{2}e^{-\pi/2\alpha_0}\mathcal{H}\left[\frac{1}{\alpha_0}\right]\ , \tag{5.23}$$

which we obtained from the Euclidean Lagrangian

$$L_E/\hbar=\frac{\pi\alpha_0}{\omega_{c0}}(\dot{x}^2+\dot{y}^2)+i\alpha_0\dot{x}y-(V_0/\hbar)\cos2\pi x -(V_0/\hbar)\cos2\pi y+2V_0/\hbar \tag{5.24}$$

by looking at tunneling between the lowest Landau orbits. However, we can also calculate the instanton tunneling for this Lagrangian to obtain the lowest bands in E^{pert}, E^{pert}, which is given by the spectrum in Eq. (5.7):

$$E^{\text{pert}}=E^{\text{inst}}=\frac{\hbar}{2}\omega_{c0}+V_0\frac{2\pi}{\alpha_0}+\frac{V_0}{\sqrt{\alpha_0}}e^{\alpha_0\gamma/2\pi}\mathcal{H}(\alpha_0)\ . \tag{5.25}$$

In the Appendix, we have drawn on Wilkinson's results to show that for large α $\hbar K\propto V_0/\sqrt{\alpha_0}$. Now we use the fact that $\mathcal{H}(\alpha)=\mathcal{H}(\alpha-[\alpha])$, where $[\alpha]$ is the greatest integer part of α. If $\alpha_0\neq[\alpha_0]$, we can define

$$\alpha_1=\frac{1}{\alpha_0-[\alpha_0]}\ , \tag{5.26}$$

$$V_1=\frac{2V_0}{\sqrt{\alpha_0}}e^{\alpha_0\gamma/2\pi}\ , \tag{5.27}$$

and

$$\omega_{c1}=V_0\frac{4\pi}{\hbar\alpha_0}\ . \tag{5.28}$$

Then, comparing Eqs. (5.25), (5.23), and (5.24), we see that $E^{\text{inst}}-(\hbar/2)\omega_{c0}$ is also the energy spectrum for the Lagrangian

$$L/\hbar=\frac{\pi\alpha_1}{\omega_{c1}}(\dot{x}^2+\dot{y}^2)+i\alpha_1\dot{x}y -(V_1/\hbar)\cos2\pi x-(V_1/\hbar)\cos2\pi y+(2V_1/\hbar) \tag{5.29}$$

if one looks at tunneling between the lowest Landau orbits whose centers now lie on a lattice with unit cells containing $1/\alpha_1$ units of flux. Because $\alpha_1>1$ and

$$g_1^2=\omega_{c1}\tau_1=4\pi^2\alpha_1(\alpha_0)^{-1/2}e^{\alpha_0\gamma/2\pi}\gg1\ , \tag{5.30}$$

this interpretation should be valid. As long as α_1 is large enough, we can repeat the instanton calculation with this new Lagrangian to obtain the lowest-lying levels of the spectrum $E^{\text{inst}}-(\hbar/2)\omega_{c0}$. The general transformations of the parameters are then

$$\alpha_{n+1}=\frac{1}{\alpha_n-[\alpha_n]}\ , \tag{5.31}$$

and, to the leading order in $1/\alpha_n$,

$$V_{n+1}=\frac{2V_n}{\sqrt{\alpha_n}}e^{-\alpha_n\gamma/2\pi}\ , \tag{5.32}$$

$$\omega_{cn+1}=\frac{4\pi V_n}{\hbar\alpha_n}\ . \tag{5.33}$$

From these transformations, we can find the values of τ_n and g_n^2:

$$\tau_{n+1}=\frac{\alpha_{n+1}}{2\sqrt{\alpha_n}}e^{\alpha_n\gamma/2\pi}\tau_n\gg\tau_n \tag{5.34}$$

and

$$g_n^2=4\pi^2\alpha_n(\alpha_{n-1})^{-1/2}e^{\alpha_{n-1}\gamma/2\pi}\quad\text{for } n\geq1\ . \tag{5.35}$$

Therefore, Wilkinson's renormalization group for the Hofstadter spectrum also defines a renormalization group for the Lagrangian in Eq. (5.24) or (3.16). At each step, the unit of time, τ, is rescaled by $e^{\alpha_n\gamma/2\pi}\alpha_n/2>1$ and the energy is given on a finer scale. Note that (at least after the first iteration) there is a fixed point whenever $\alpha_{n+1}=\alpha_n$ for all n and that when α is irrational the RG flow is determined by the expansion of α as a continued fraction. For any energy E that is an eigenvalue of Harper's equation, $-E$ is also an eigenvalue, so, with minor alterations, these renormalization transformations

also apply to the highest band in the spectrum. Thus we obtain a considerable amount of information about the self-similarity of the spectrum from these transformations. It should be emphasized that this particular form for the RG flow is only valid in the limit of large α. For intermediate values of α for which the dilute-instanton-gas approximation is still valid a similar procedure can be used, but the expression for $\hbar K$ in terms of $V_0/\sqrt{\alpha_0}$ will receive corrections. In addition, we must include the ΔE_c term in the expression for E^{inst}. As an example of the RG flow which also illustrates the size of these corrections we have plotted $E-\hbar\omega_c/2$ for the q lowest bands for $1/\alpha=q/p$ between $\frac{1}{4}$ and $\frac{1}{3}$ in Fig. 5. The left-hand plot in Fig. 5 shows the instanton prediction for these bands and was obtained from Eq. (5.7) with $\hbar K=V_0/\sqrt{\alpha}$ by subtracting $\hbar\omega_c/2$ and adding in the perturbative correction ΔE_c. The right-hand plot in Fig. 5 shows the perturbative prediction for these bands as given by E_0^{pert} in Eq. (2.22). The difference in darkness of the two plots is due only to the difference in the number of points included.

VI. DISCUSSION

We have shown that instanton techniques can be applied to study the spectrum of an electron in a periodic potential and a transverse magnetic field. By comparing the instanton approach with perturbation theory in the potential V we have shown how the recursive structure of the spectrum can be understood in terms of dual path integral pictures.

Although we have considered only symmetric potentials, the asymmetric case is also of great interest in the study of the localization transition in incommensurate systems. The RG flow in this case is discussed in detail in Ref. 4 and could also be discussed by the methods developed here.

For simplicity we have restricted ourselves to a simple sinusoidal potential. Clearly the instanton techniques can be applied to a general periodic potential, and at least for potentials with rectangular symmetry, will lead to (5.1) for the shift in the energy levels. On the other hand, for more general potentials, perturbation of the lowest Landau level will not lead to Harper's equation but to some variant of it. However, the arguments presented here show that the substructure of the energy levels will again be described by Eq. (5.3) so in this sense the sinusoidal potential is universal in that all potentials of this general symmetry will flow to it under the action of the RG flow.

We have also emphasized the instanton calculations for large g^2 because of their connection to the self-similarity of the spectrum through the comparison with perturbation theory. It is clear, however, that the semiclassical approach is valid over a much larger range of parameters. In particular, it is valid at large α for all g^2 and at small α as long as g^2 is sufficiently small. When both g and g/α are small, it reproduces the usual tight-binding calculation and gives a definite expression for the transfer integral in this limit.

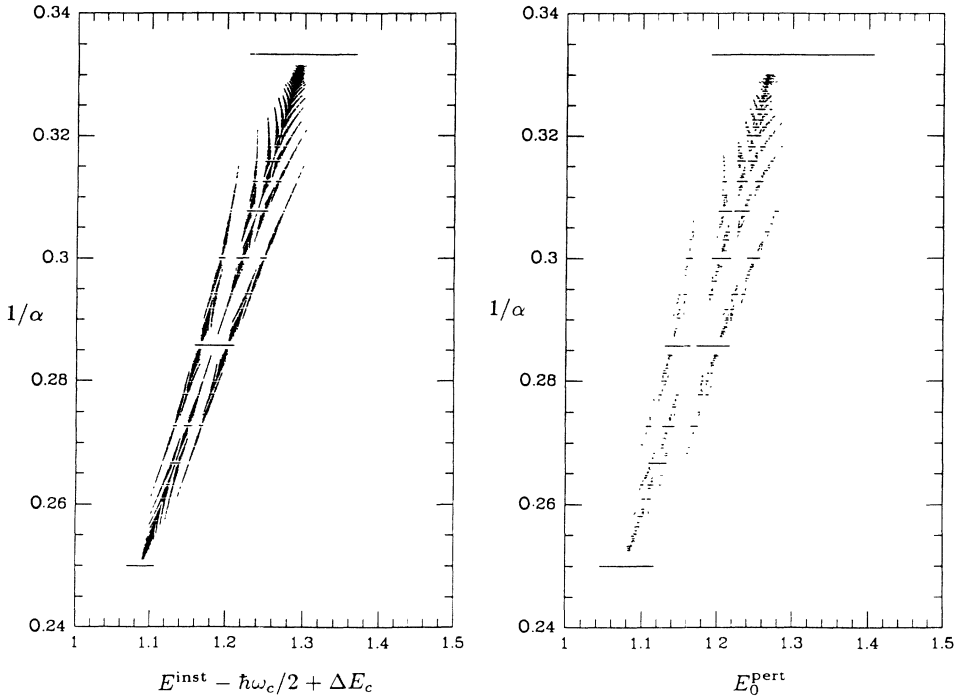

FIG. 5. Structure of the q lowest bands for $1/\alpha=q/p$ between $\frac{1}{3}$ and $\frac{1}{4}$ as predicted by the perturbative and instanton calculations.

Although we have considered only an idealized system, we hope that some of the techniques we have discussed will be applicable to the study of more realistic two-dimensional systems with T violating interactions. Also, direct experimental study of the spectrum discussed here no longer seems outside the realm of possibility.[20]

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation (NSF) Grant No. PHY80-19 754. D.F. was also supported by the NSF and the U.S. Department of Education. J.H. acknowledges the support of NSF Grant No. PHY86-58 033 and the Alfred P. Sloan Foundation.

APPENDIX A: COMPARISON WITH WILKINSONS' APPROACH

Drawing on the results of Wilkinson[4] we can show that the instanton calculation of the q lowest bands of the spectrum agrees with the q bands making up the lowest "subband" of the perturbative spectrum. Wilkinson writes Harper's equation as

$$H(x)\psi(x)=W_0(2\cos\hat{\mathbf{p}}+2\cos\hat{\mathbf{x}})\psi(x)=E\psi(x)\ , \tag{A1}$$

where

$$\hat{\mathbf{p}}=-i\tilde{h}d/dx \quad \text{with } \tilde{h}=\frac{2\pi}{\alpha}\ . \tag{A2}$$

To see the connection of this form to our analysis note that in the limit $g^2\rightarrow\infty$ the Lagrangian (3.19) yields essentially the above Hamiltonian since in this limit y becomes canonically conjugate to x.

Wilkinson derives an approximate renormalization group by using transfer matrices to describe tunneling between the degenerate, semiclassical states which are centered around the minima of $H(x)$. When $\alpha\gg1$, for the equation

$$Eg(m)=W_0\left[g(m+1)+g(m-1)+2\cos\left[2\pi m\frac{1}{\alpha}-\kappa\right]g(m)\right] \tag{A3}$$

he finds that the energy of the νth subband is given by the eigenvalue equation

$$E_\nu g(m)=E_n g(m)+W_1[g(m+1)+g(m-1)+2\cos(2\pi m\alpha-\kappa)g(m)]\ , \tag{A4}$$

where E_n is the energy of the nth state in one of the potential wells of $H(x)$. For the lowest state, W_1 is given by

$$W_1=-\frac{2}{\alpha}t_x W_0\ , \tag{A5}$$

where t_x is the magnitude of the tunneling coefficient for going from one well to the adjacent well:

$$t_x=\exp\left[-\frac{1}{\tilde{h}}\int_{x_1}^{x_2}|p(x)|dx\right] \tag{A6}$$

where $p(x)$ is given by

$$W_0[2\cos p(x)+2\cos x]=E_n \tag{A7}$$

and x_1 and x_2 are the classical turning points.

For the lower energy levels, x and p are near $(2n+1)\pi$, so we will redefine them by $x\rightarrow x-\pi, p\rightarrow p-\pi$ and expand the cos$x$ and cos$p$. The resulting equation for E_n is

$$W_0(-2+p^2-2+x^2)\psi(x)=E_n\psi(x)\ . \tag{A8}$$

This has eigenvalues

$$E_n=-4W_0+\tilde{h}2W_0(n+\tfrac{1}{2})\ , \tag{A9}$$

so the energy of the lowest state is

$$E_0=W_0(-4+\tilde{h})=W_0\left[-4+\frac{2\pi}{\alpha}\right]\ . \tag{A10}$$

For a state centered around $x=0$, $p=0$ with this energy, the classical turning points occur when $x=\pm\sqrt{\tilde{h}}$. In the tunneling region, between $x_1=\sqrt{\tilde{h}}$ and $x_2=2\pi-\sqrt{\tilde{h}}$, p must be imaginary, $p=ip_I$, so the equation for $p_I(x)$ is

$$\cosh p_I(x)=2-\frac{\tilde{h}}{2}-\cos x \tag{A11}$$

and the tunneling integral is given by

$$t_x=\exp\left[-\frac{1}{\tilde{h}}\int_{\sqrt{\tilde{h}}}^{2\pi-\sqrt{\tilde{h}}}\left|\cosh^{-1}\left[2-\frac{\tilde{h}}{2}-\cos x\right]\right|dx\right]$$
$$=c\tilde{h}^{-1/2}e^{-\gamma/\tilde{h}}[1+O(\tilde{h})]\ , \tag{A12}$$

where γ is defined in Eq. (5.8), and where c is a number independent of all the parameters. In what follows, and in Sec. V, we ignore the factors of $c/\sqrt{2\pi}$. Remembering that $\tilde{h}=2\pi/\alpha$, we obtain the rescaling of the energy,

$$W_1=-W_0\frac{2}{\sqrt{\alpha}}e^{-\alpha\gamma/2\pi}\ . \tag{A13}$$

Taking into account that E^{pert}, the energy spectrum in Eq. (2.22) found by the perturbation calculation, is the sum of $(\hbar/2)\omega_c+2V_0$ and the solutions to Harper's equation scaled by $W_0=(V_0/2)e^{-\pi/2\alpha}$, we find that the energy of the lowest subband is given by

$$E=\frac{\hbar}{2}\omega_c+2V_0+\frac{V_0}{2}e^{-\pi/2\alpha}\left[-4+\frac{2\pi}{\alpha}\right]$$
$$-\frac{V_0}{\sqrt{\alpha}}e^{-\alpha\gamma/2\pi}\mathcal{H}(\alpha) \tag{A14}$$

or

$$E=\frac{\hbar}{2}\omega_c+V_0\frac{2\pi}{\alpha}+\frac{V_0}{\sqrt{\alpha}}e^{-\alpha\gamma/2\pi}\mathcal{H}(\alpha)\ , \tag{A15}$$

where we have kept only terms to first order in $1/\alpha$ and we are making use of the fact that the spectrum of Harper's equation is symmetric about $E=0$. Comparing

this with the instanton calculation which gave (5.7) we see that in the limit of large α we have $\hbar K = V_0/\sqrt{\alpha}$ and this expression for the spectrum is then identical to the instanton calculation. It should be emphasized that this expression for K is only valid in the limit of large α. For intermediate values of α there are corrections to both $\hbar K$ and $E_n = E_c$ which must be included to obtain agreement between the instanton calculation and the perturbative calculation.

---

*Internet address: def@pupthy.princeton.edu.

†Internet address: harvey@control.uchicago.edu. Permanent address: Jadwin Hall, Princeton University, Princeton, NJ 08544.

[1]D. J. Thouless, S. Girvin, and A. McDonald, *The Quantum Hall Effect,* and references therein.

[2]I. Affleck and J. B. Marston, Phys. Rev. B **37**, 3774 (1988).

[3]A. B. Pippard, Proc. R. Soc. London, Ser. A **270**, 1 (1962); Philos. Trans. R. Soc. London **256**, 317 (1964).

[4]M. Wilkinson, Proc. R. Soc. London, Ser. A **391**, 305 (1984).

[5]E. Brown, Phys. Rev. **133**, A1038 (1964).

[6]J. Zak, Phys. Rev. **134**, A1602 (1964).

[7]W. Opechowski and W. G. Tam, Physica **42**, 529 (1969).

[8]R. E. Peierls, Z. Phys. **80**, 763 (1933).

[9]M. H. Cohen and L. M. Falicov, Phys. Rev. Lett. **7**, 231 (1961).

[10]M. Ya. Azbel', Zh. Eksp. Teor. Fiz. **46**, 939 (1964) [Sov. Phys.—JETP **19**, 634 (1964)].

[11]D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).

[12]G. H. Wannier, Phys. Status Solidi B **88**, 757 (1978).

[13]A. Rauh, Phys. Status Solidi B **65**, K131 (1974); **69**, K9 (1975).

[14]M. Ya. Azbel, Phys. Rev. Lett. **43**, 1954 (1979).

[15]See, e.g., R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).

[16]S. Kivelson, C. Kallin, D. P. Arovas, and J. R. Schrieffer, Phys. Rev. Lett. **56**, 813 (1986); Phys. Rev. B **36**, 1620 (1987); S. M. Girvin and T. Jach, *ibid.* **29**, 5617 (1984); J. K. Jain and S. Kivelson, *ibid.* **37**, 4111 (1988).

[17]See, e.g., S. Coleman, *Aspects of Symmetry, Selected Erice Lectures* (Cambridge University Press, Cambridge, 1985).

[18]L. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

[19]D. Langbein, Phys. Rev. **180**, 633 (1969).

[20]P. Chaikin (private communication).