Conditions for the quantum Hall effect

Robert Joynt
TCM Group, Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, England

R. E. Prange
Department of Physics and Astronomy and Institute for Physical Sciences and Technology, University of Maryland, College Park, Maryland 20742

(Received 7 November 1983)

The quantum Hall effect occurs in imperfect two-dimensional systems of electrons subjected to strong perpendicular magnetic fields. We determine the conditions on the random imperfection potential, relative to the magnetic field, under which the quantum Hall effect may be expected to occur, if electron-electron interactions are neglected. It is found that all physically reasonable disordered potentials should exhibit the effect in the limit of large field (with electron density increased proportional to the field). This is done by examining the behavior of the eigenstates in various kinds of potentials and collecting the results. Earlier work is reviewed in this context. The consequences for localization in two dimensions in a strong magnetic field are discussed. The mechanism of localization turns out to be qualitatively different from that of Anderson localization.

I. INTRODUCTION

The quantum Hall effect[1–3] is one of the most exciting discoveries in physics of the past decade. Although much of the theoretical interest has shifted to the anomalous quantum Hall effect,[4] there are important questions remaining in the normal case. In this paper we discuss the theory of the normal quantum Hall effect. Our main conclusions were published earlier in a brief communication.[5]

The Hall and longitudinal conductances, σ_{yx} and σ_{xx}, of a pure, two-dimensional noninteracting electron gas in a perpendicular magnetic field can be calculated exactly if an integral number of Landau levels are precisely filled. Then σ_{xx} (and the resistance ρ_{xx}) vanish, since the electrons drift perpendicular to both the electric and magnetic fields, while $\sigma_{xy}=ne/B=N_L e^2/h$, where N_L is the (integral) number of filled Landau levels, and h is Planck's constant. This result relies on the fact that a single filled Landau level has density $n=eB/h$.

In actual experiments in inversion layers, at sufficiently strong fields, low temperatures, and high mobilities, these ideal values are found, indeed to an astonishing precision. What is more, they persist over a finite interval as the density or magnetic field is varied, giving a step structure to the Hall conductance versus electronic density plot. This is illustrated in Fig. 1. This aspect cannot be explained without invoking impurity effects.

These step values are observed to be quantized to an accuracy of almost 0.1 ppm. Since the fine-structure constant $\alpha=e^2/hc=2\sigma_{yx}/N_L c$ has been previously determined only to about this accuracy (and since c is known to nine significant figures), the two-dimensional electron gas offers a way to improve our knowledge of this fundamental physical quantity.[3,6]

The conductance of a two-dimensional electron gas in a strong perpendicular field is thus a case in which impurities, imperfections, and finite temperature do not necessarily have a big dissipative effect. In that respect the situation is similar to the Meissner effect in superconductors, or for that matter, similar to the effect of a small number of impurities on an insulator. Obviously, the existence of an energy gap is the common denominator in these cases. Then temperature effects, for example, have the possibility of being exponentially small.

The effect of the interelectron Coulomb interactions is somewhat complicated and is the main interest in the case of the anomalous effect.[4,7] On the one hand, the basic quantum Hall effect phenomena and the gap (between Landau levels) already exist for the noninteracting electron model, whereas superconductivity exists only by virtue of the electron-electron interaction. On the other hand, the anomalous quantum Hall effect corresponds to fractionally filled Landau levels, e.g., $N_L=\frac{1}{3}, \frac{2}{5}, \frac{2}{3}$, etc.

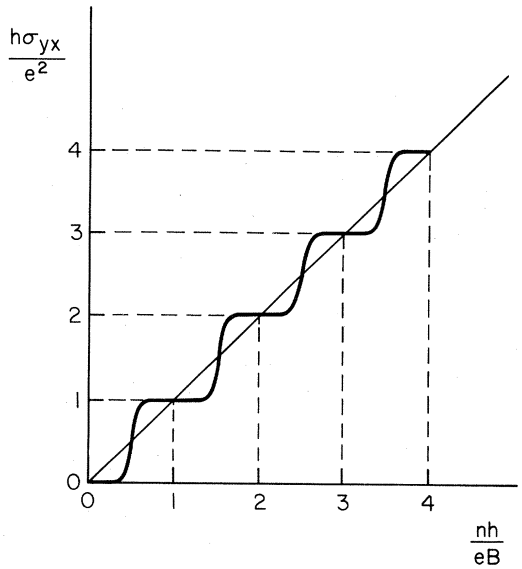

FIG. 1. Hall conductance vs density showing the step structure found in clean two-dimensional systems. The straight line is the conductance in the absence of impurities.

This can only be explained on the basis of electron interactions, in which the Coulomb repulsion favors a state with an energy gap at these rational fractional fillings. This situation is not unlike the case of the insulator, where in the simplest cases the gap is a consequence of the nontrivial external potential of the substrate ions. There are, however, (Mott) insulators for which the Coulomb interaction plays an essential role.

In this paper we shall consider the two-dimensional noninteracting electron gas in the presence of random substrate potentials in a strong magnetic field. We thus concentrate on the effects of imperfections, which are *also* essential to the problem. This is the case because the Hall steps are seen for a *range* of densities, not just those corresponding to integer Landau level occupation. In fact, if there are *no* imperfections, but with any interactions and at any temperature, it is easy to show (by going to the frame of reference $c\vec{E}\times\vec{B}/B^2$) that $\sigma_{xy}=nec/B$, with n the two-dimensional number density of electrons and this density is not quantized.

The general picture which is developed is that there is a subset of electrons (in extended states) which carry a net current and form a sort of fluid of effectively the ideal (perfect Landau-level) density. The local velocity, actual density, and current density of this fluid varies from point to point, depending on the potential in which it finds itself, but the *total* current is the ideal current. This type of behavior is seen in experiments using special probes to measure the internal currents of a sample.[8] The remaining electrons, or holes, are localized by the random potential. Extra electrons can be accommodated by the localized states, up to a point, which account for the plateau structure. The electrons in the mobile fluid cannot undergo inelastic dissipative scatterings when all these states are well below the Fermi level (and the temperature is sufficiently low), i.e., there is an energy gap against changing the state of the fluid. The gap is between bands of extended (fluid) states, which are separated by the localized states. This applies when the density is in the middle of a step. Also, in order that the Fermi level remain between steps, there must be states to be occupied, else the Fermi level will jump with a small change in density.

When the density is between steps, the Fermi level is in the middle of the fluid electron states. Then there can be dissipation, σ_{xx} will be positive, and σ_{yx} will not have an ideal value.

This picture presumably applies to the anomalous quantum Hall effect as well. However, it has yet to be worked out for that case. The convenient independent electron concepts of localized and extended states are not manifestly applicable there.

In this work we shall treat cases of strong disorder. This is thus of interest because of its connection to the general question of localization in two dimensions. It is known from scaling arguments that all states, in the absence of magnetic field, are localized, if only logarithmically.[9] Both the power and the shortcoming of these arguments, however, is that the finer details of the electronic states in the random potential are irrelevant to their validity. Not too much needs to be known about the wave functions, but little can be concluded about them either. This is especially true of the extended states. In the strong-field case treated here we believe we can say a certain amount about the properties of the states, and in a quantitative way. It turns out that the process by which localization in a strong magnetic field is produced by disorder is quite different from Anderson localization. This lends credence to the view[10,11] that even weak magnetic fields have an essential effect on localization in two dimensions, since that case can be considered as intermediate between Anderson localization and the type of localization considered here. We will not attempt to link all these together in the present work, but only to explicate the nature of the states in the strong field.

The plan of the paper is as follows. Section II is a discussion of the pure system, some special cases of impurity potentials, and the qualitative nature of the disordered potential in real systems. Sections III—V each analyze a different kind of disordered potential. In each case the quantum Hall effect is derived and the criteria for validity of the derivation are discussed. The three sorts of potential are as follows: Sec. III, small potentials; Sec. IV, scattering potentials; Sec. V, smooth potentials. Section VI combines the results of the previous sections, and deduces as much as possible about the Hall effect and localization. Some remarks about finite temperature and interactions are added. Section VII is the conclusion.

We will find that the quantum Hall effect occurs for all nonpathological disordered potentials in sufficiently strong field. The phenomenon is *not* peculiar to any particular model. This conclusion has been stated before,[12] but without detailed discussion of its microscopic justification.

II. PURE SYSTEM AND SPECIAL CASES

The system is a noninteracting gas of electrons confined to the x-y plane and subjected to a magnetic field in the z direction. We consider only one spin direction and employ units in which m, the effective mass of the electron, $\hbar\omega_c$, the cyclotron energy, and the magnetic length $l\equiv(\hbar c/eB)^{1/2}$ are all unity. (The length l is about 50—100 Å.) The Hamiltonian is

$$H=\frac{1}{2}\left[-i\frac{\partial}{\partial x}+A_x\right]^2+\frac{1}{2}\left[-i\frac{\partial}{\partial y}+A_y\right]^2 -\frac{1}{2}+U(x,y)\,. \tag{2.1}$$

$\vec{A}$ is the vector potential and U is the disordered potential V plus the applied electric field. If $U=0$ the energy levels are $E_n=n$ and each is infinitely degenerate. It is possible to choose linear combinations such that the energy eigenfunctions are either localized or extended. For example, if we choose to diagonalize the z component of angular momentum, the solutions in the gauge $\vec{A}=\frac{1}{2}(y,\ -x)$ are

$$f(\rho,\phi)=[2\pi\Gamma(\tfrac{1}{2}\,|\,m\,|\,+1)]^{-1/2}e^{im\phi}e^{-\rho^2/4}(\tfrac{1}{2}\rho^2)^{|\,m\,|/2}\,, \tag{2.2}$$

with $m=-\infty,\ldots,-1,0,1,\ldots,\infty$). These functions

form rings around the origin of approximate radius $\rho=(2\,|\,m\,|\,)^{1/2}$ and width l. They are thus localized. On the other hand, we can choose to diagonalize the y component of momentum. The Landau gauge $\vec{A}=(0,-x)$ is convenient and the solutions in the lowest level are

$$g(x,y)=(L\sqrt{\pi})^{-1/2}e^{ipy}e^{-(x+p)^2/2}\,. \tag{2.3}$$

Here $p=2\pi k/L$, with k integer. L is the length of the system in the y direction, and $|\,p\,|<W/2$, where W is the length of the system in the x direction. These states are extended and form strips along the line $x=$const, again of width l.

If $U\neq 0$ then the degeneracy is broken, and the eigenstates are uniquely specified. Their localization characteristics are no longer ambiguous. An exactly solvable example is $U=C/(x^2+y^2)$. Let $\kappa=\frac{1}{2}(m^2+C^2)^{1/2}$. Then the energies are $F_{nm}=n+\kappa+m/2$ and the eigenstates of the $n=0$ level are

$$\psi_{om}=[2\pi\Gamma(\kappa)]^{1/2}e^{im\phi}e^{-\rho^2/4}2^{-\kappa}\rho^{2\kappa}\,. \tag{2.4}$$

The degeneracy is completely broken, and we have true localization. In fact, as we shall see below, *all* states are localized for a potential of infinite range in the absence of an external electric field. (Even a potential of finite range will bind all states, though in this case the binding energy will be exponentially small for most of them. For practical purposes the number of bound states is finite.) This holds regardless of whether the potential is attractive or repulsive.

If instead of an impurity potential we apply an electric field in the $-x$ direction, then $U=-vx$, where $v=E/B$ is the classical drift velocity. E is the applied electric field and B is the magnetic field. The y component of momentum is still a good quantum number. The states in the lowest Landau level are

$$g_v(x,y)=(L\sqrt{\pi})^{-1/2}e^{i(p+v)y}e^{-(x+p+v)^2/2}\,, \tag{2.5}$$

where p satisfies the same conditions as in Eq. (2.3). The energies of these states are $E_{op}=pv$ so that the degeneracy is completely lifted. Again the wave functions follow the contours of constant potential, which are now the lines $x=const$. These lines escape off to infinite distance so that the stationary states are extended.

The question of localization in this sytem can be seen as a competition between the impurities and the electric field. The former produce closed-contour lines and favor localization, whereas the latter carry some of these lines off to infinity, giving extended states. There is one more exactly solvable case[13] which combines an applied field and an impurity potential: $U=-vx+\lambda\delta(\vec{r})$. For a sufficiently weak applied field ($v<\lambda/10$), a condition which is always satisfied in practice (v is of order 10^{-4}), there is one bound state per Landau level and the remaining states are extended.

To understand the Hall effect we need the current of the stationary states in the y direction. Each of the states (2.5) can be seen by direct application of the current operator to have an expectation value for the current of $(0,ev)$. The total number of states in one level is $WL/2\pi$. Thus the total current in the y direction contributed by one level is $J_y=evWL/2\pi$, which, reverting to normal units, gives a Hall conductance $\sigma_{yx}=J_y/WLE=e^2/h$, the quantum of conductance. This value persists even if we add the δ-function impurity. The extended states carry additional current such that they compensate for the zero current carried by the bound states. The total conductance of one Landau level is precisely e^2/h for this case. We will show that similar picture holds more generally: There are a number of bound states which carry no Hall current, but the same potential which binds these electrons accelerates others, producing the ideal quantized conductance.

As a final characterization of the pure system and for future reference we calculate the retarded Green's function in the presence of an applied field. Then the complete set of eigenfunctions is

$$\phi_{np}=(2^n n!L\sqrt{\pi})^{-1/2}e^{i(p+v)y}H_n(x+p)e^{-(x+p+v)^2/2}\,. \tag{2.6}$$

Here H_n is the nth Hermite polynomial and again we work in the Landau gauge. The eigenenergies are $E_{np}=n+pv$ so that the degeneracy is completely lifted. Each level has broadened into a band and we assume that v is sufficiently small that the widths of the bands are smaller than their spacing. The expression for the Green's function is

$$G_0(\vec{r},\vec{r}',E)=\sum_{n=0}^{\infty}\frac{L}{2\pi}\int dp\frac{\phi_{np}(\vec{r})\phi^*_{np}(\vec{r}')}{E-E_{np}+i\delta}\,,$$

which after some algebraic manipulation becomes

$$G_0=\frac{X}{2\pi^{3/2}}\sum_{n=0}^{\infty}2^{-n}(n!)^{-1}\times\int dp\frac{H_n(x+p)H_n(x'+p)e^{-(p+\alpha/2)^2}}{E-n-pv+i\delta}\,. \tag{2.7}$$

Here X is defined as $\exp\{i[v-(x+x')/2](y-y')-(r-r')^2/4\}$ and $\alpha=x+x'-i(y-y')$. Let us examine the expression (2.7) first in the case where E lies in a gap, i.e., $E\neq n+pv$ for any n,p. Then the term pv in the denominator of the integrand can safely be ignored. The integral can be performed, leading to

$$\begin{aligned}G&=(X/2\pi)\sum_n(E-n)^{-1}L_n((\vec{r}-\vec{r}')^2/2)\,,\\&=(1/2\pi)X\Gamma(-E)\Psi(-E,1;(\vec{r}-\vec{r}')^2/2)\,.\end{aligned} \tag{2.8}$$

Ψ is a confluent hypergeometric function of the second kind. To obtain the last expression we have used the generating function for the Laguerre polynomials. Ψ behaves for these values of its parameters asymptotically as $|\,r-r'\,|^{2E}$. Together with the definition for X we find

$$|G(\vec{r},\vec{r}',E)| \sim |\vec{r}-\vec{r}'|^{2E} e^{(\vec{r}-\vec{r}')/4}, \quad |\vec{r}-\vec{r}'| \to \infty . \quad (2.9)$$

Calculation to higher order in v does not change exponentially decaying behavior of G.

Now take the case where E lies in the Mth band, say $E=M+p_0v$. We will not bother to get the exact form for the Green's function, but only the asymptotic behavior. The terms in the sum with $n\neq M$ give exponential decay as above. The integral can be conveniently performed around an infinite rectangular contour with the real axis as one edge and the line $\mathrm{Im}p=(y-y')/2$ as the other. This extracts the asymptotic form as

$$G_0(\vec{r},\vec{r}',E)=\frac{i}{\sqrt{\pi v}}\Theta(y-y')\exp[-\tfrac{1}{2}(p_0+x)^2-\tfrac{1}{2}(p_0+x')^2+i(p_0+v)(y-y')]H_M(x+p_0)H_M(x'+p_0) . \quad (2.10)$$

Thus the Green's function vanishes unless $x\sim x'\sim p_0$. We will explore the implications of this later on.

We intend to examine the behavior of the electronic states in several different types of disordered potentials, but first it would be nice to know what the potential looks like in a real system, at least qualitatively. This is fairly well known for the zero-field case. A comprehensive review has been done by Ando, Fowler, and Stern.[14]

A chief difference in the two-dimensional, as opposed to the three-dimensional, case is that an average fluctuation V between two states is

$$\int f_1^*(x,y)\zeta_0^*(z)V(x,y,z)f_2(x,y)\zeta_0(z)dx\,dy\,dz . \quad (2.11)$$

The f's are the two-dimensional wave functions, and the ζ_0's are the lowest energy eigenfunctions of the one-dimensional Hamiltonian of the z direction. These last are the same for all states of interest. The effective potential to be used in the two-dimensional Schrödinger equation is

$$V_e(x,y)=\int |\zeta_0(z)|^2 V(x,y,z)dz . \quad (2.12)$$

For example, consider an unscreened point charge Qe at the origin. The effective potential would be

$$V_e(x,y)=-\frac{Qe^2}{2\kappa}b^3\int_0^\infty z^2 e^{-bz}(r^2+z^2)^{-1/2}dz , \quad (2.13)$$

where we have substituted the approximate form of Fang and Howard[15] for the wave function of the third dimension. $1/b$ is the extension in the z direction, about 50–100 Å. The Coulomb form for the effective potential holds for distances greater than this, but is cut off at small distances. In the above, κ is the dielectric constant, which is assumed to be the same in the semiconductor and insulator. Otherwise, image charge effects should be included in (2.13).

The overall effect is a smoothing on the scale b, which is, unfortunately, of the same order of magnitude as l. The same will hold for impurities located far away from the interface, both those in the oxide and those in the semiconductor. There will be many such defects, producing a potential which is disordered, but still quite smooth.

Screening[16] will be important also, unless all Landau levels are either completely full or completely empty and virtual inter-Landau-level transitions can be neglected. Its effect will be complementary to the finite-thickness smoothing, since it weakens the long-wavelength components of the potential. However, the detailed effects of screening in the strong-field case have not been worked out. The exchange correlation effects are even less clearly understood. In particular, the characteristic screening length is not known. However, most of the basic lengths in the problem, such as b, l, the Coulomb length l_c defined by $e^2/\kappa l_c=\hbar\omega_c$, and $n^{-1/2}$ are comparable in practice.

A further argument goes as follows: Suppose a conductivity is defined, which can be done by averaging over a sufficiently large length scale. Suppose σ_{xx} although small is positive, which will be true at finite temperature. Then $(\partial^2/\partial x^2+\partial^2/\partial y^2)\Phi(x,y)=0$, since the current is divergence free, where Φ is the two-dimensional electrochemical potential. Thus Φ has no maximum or minimum in the sample interior. Now Φ is to first approximation just a smoothed $V(x,y)$. If the scale over which one must average to define conductivity is l, this suggests that the fluctuating potential V has its important Fourier components in the range of wavelengths about l but the components at much larger or smaller wavelengths are not so important. However, we shall not impose the condition that the long-wavelength components of V have no interior maximum or minimum since we do not know very well the scale of averaging required for this to be true. Still, we must show that even quite small potentials have no effect on the total current, since we are interested in very precise answers.

The rms magnitude of V can presumably be fairly large (of order $\hbar\omega_c$) in dirty systems, and considerably smaller in clean ones. The maximum magnitude of V can probably be on this scale (for example, near a charged impurity) even in a clean system, but this will occur rather rarely.

A further source of scattering which cannot be treated on quite the same footing is surface roughness. The interface is really a rough wall into which the inversion layer may not penetrate. It gives rise to an effective potential with a correlation length estimated[17] to be about 15 Å. The latter value is open to some doubt because it is deduced from theoretical fits to conductivity data and the agreement still leaves something to be desired. It is thought that some sort of more short-range scattering may be necessary for a full explanation of the data, especially since the conductivity in weak magnetic fields seems to require such an assumption.[14]

In summary, the effective two-dimensional disordered potential in a strong magnetic field probably has its dominant wavelengths at just the most difficult range. It is not even proven that a single-particle potential is well defined, or is a smooth function of density and/or field, in view of the anomalous Hall effect. Nevertheless, we can make considerable progress by studying the effects of the potential under various assumptions.

III. GAUGE ARGUMENTS

Here we review the gauge arguments of Laughlin,[12] in the geometry of Halperin.[18] Consider a flat annular sample in the x-y plane. There is a uniform field B in the z direction, and, in addition, a magnetic flux Φ threading the hole. There is an applied dc electric field directed towards the origin and consequently a Hall current flows azimuthally. The setup is shown in Fig. 2. Now let Φ be changed adiabatically by one flux quantum from 0 to h/e. In this process the energy levels of extended states will move because the phase of a wave function which wraps around the origin must change by an additional $e\Phi/h$ in order to rejoin itself. (See Fig. 3.) Energy levels of localized states will not change since a state localized at R suffer only an overall *constant* phase change, with the phase being proportional to the value of the gauge function at R. When the value of Φ reaches h/e the energy levels are the same as when $\Phi=0$, since the phase advance is 2π, and the problem has returned to its original boundary conditions. This fact is the basis of the Byers-Yang theorem on the periodicity in Φ of the free energy of a superconducting ring.[19] Also, see Ref. 19 for a formal statement of the argument.

In spite of the circumstance that the *levels* remain unaffected by the flux change, the *total energy* of the system may change since some electrons may have changed their adiabatic occupation. What we must investigate is the adiabatic movement of the levels as Φ changes. Figure 3 shows the energy-level structure of the system, where it is assumed that there is an inner and outer "guard" ring which is free of imperfections. In this region the eigenstates are localized in the areas between the lines, i.e., they lie in circular annuli. As the threading flux changes, the eigenstates adiabatically transform one step inward, and if there are electrons filling the states, there will be a corresponding current. Assuming that there are no imperfections at all, the total change in energy is eEW, since the net result is the transfer of one electron from one edge to the other. This can be thought of as a transient Hall current due to the azimuthally directed induced electric field. Now we set the change in electrostatic energy equal to the change in the total electronic energy to obtain $I_H\Delta\Phi=eEW$ or $I_H=e^2EW/h$, which gives a Hall conductance of $\sigma_{yx}=e^2/h$. If N_L levels are filled, then N electrons are transferred and the result is $\sigma_{yx}=N_Le^2/h$.

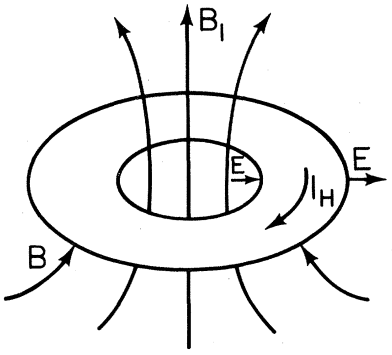

FIG. 2. Geometry of the Laughlin argument in the version given by Halperin. B and E are the dc fields. B_1 is the auxiliary field which threads the annulus. The change in the system as B_1 is increased is considered.

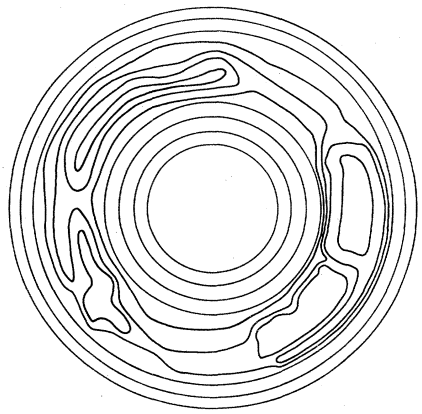

FIG. 3. Approximate boundaries of states of a system whose imperfections are confined to a central ring. As the threading flux changes by one quantum, each extended state, which is relatively large all around the ring, passes over adiabatically into its neighbor. The localized states are not affected by the threading flux.

Next we consider the impure system. We have pictured the eigenstate boundaries in the central region of Fig. 3. As we shall see, each state occupies an area approximately $2\pi l^2$. Some states are localized; others extend all the way around the ring. The localized states are unaffected by the threading flux, but the extended states transform adiabatically into their next neighbor. Note that there is a natural geometric succession of states for this transformation. After one quantum of flux has threaded the loop, an electron has moved from the outer to the inner edge, and consequently the energy is changed by the difference in the electronic energy at the two edges. This is just the electrochemical potential difference (not simply the electric potential difference). That, of course, is the potential which is measured, so again the ideal Hall conductivity is obtained.

We believe this is the correct physical picture of the states, which shows how they organize in a regular way. However, an argument of Halperin[18] gives some conditions which, if satisfied, guarantee extended states and an ideal quantum Hall effect, and which do not rely on details of the eigenstate structure.

Consider that, as the threading flux is changed, that the outer guard ring "injects" an electron into the central region while another is "extracted" from the impure ring into the inner guard ring. This process must be able to go on at low energy, since the flux is changed adiabatically. There thus must be *some* extended states in the central ring since only they are affected by the change of the flux. If these states are all well away from the Fermi level, there is no other possibility but that they somehow be organized to pass the charge smoothly along through the central ring. For if not, charge would accumulate, costing much energy.

If the mobile states in the impure region go up to the Fermi level, then there is the possibility that charge is not transferred through the central region. The simplest possibility is that the bulk of the imperfect region is insulat-

ing, so that there are two new edges (at the junction of the guard rings and the central region). Then at these new edges (just as at the sample edges[18]), there must be extended states going up to the Fermi level. The density of states for these two situations is illustrated in Fig. 4.

Thus we must find the conditions under which there are no *extended* states at all energies between two Landau levels. There must be *some* localized states, however, outside the energy region occupied by the extended states, in order that the Fermi level can lie in that region for an extended range of densities. In fact, we shall generally find that the extended states occupy a narrow range of energies, at least in clean situations.

An easy sufficiency condition that there be no extended states between Landau levels (indeed that there be no states at all) was stated independently by Halperin[18] and by Thouless.[20] Namely, assume that $|V(r)|$ is everywhere less than $\hbar\omega_c/2$, where V is the impurity potential. In this case, the potential $V(r)$ does not move the original Landau levels by as much as half the distance to the next level.

The fact that no level moves by more than $V_{\max}$ can be shown very simply:[21] Let $H_\lambda = H_0 + \lambda V$ and ψ_λ satisfy $H_\lambda\psi_\lambda = E_\lambda\psi_\lambda$. E_1 is the actual energy in the potential and E_0 is the unperturbed energy. Then the Feynman-Hellmann theorem states that

$$\frac{dE_\lambda}{d\lambda} = \langle \psi_\lambda | V | \psi_\lambda \rangle \ . \tag{3.4}$$

Integrating this over λ gives

$$E_1 - E_0 \le \int_0^1 d\lambda \int V_{\max} |\psi_\lambda(\vec{r})|^2 d^2r \le \hbar\omega_c/2 \ . \tag{3.5}$$

It is worth recording a slightly simpler version of this argument which avoids the microscopic question of how the system changes its energy, and focuses on the total charge transported as the flux changes. Using the same geometry as in Fig. 2, draw a line in the sample which winds once around the origin. Let this loop have line element $d\vec{L}$ and the enclosed area have element $d\vec{S}$. Take the applied electric field to be 0 and change Φ from 0 to h/e. Then

$$\frac{d\Phi}{dt} = \frac{d}{dt}\int \vec{B}\cdot d\vec{S} = -\int \vec{E}\cdot d\vec{L} \ . \tag{3.1}$$

Integrating this over time yields

$$\frac{h}{e} = -\int dt \int d\vec{L}\cdot\overleftrightarrow{\rho}\cdot\vec{J} \ , \tag{3.2}$$

where J is the current across a line of unit length and ρ is the resistivity tensor. Let ρ_{xx} vanish. Finally we obtain

$$h/e = \rho_{yx}\Delta Q \ , \tag{3.3}$$

where the ΔQ is the total charge transferred from one side of the loop L to the other. The assumption is that this is one electron charge per Landau level. Then $\sigma_{yx} = 1/\rho_{yx}$ (true if $\rho_{xx} = 0$) is correctly quantized. If σ_{xx} is not strictly 0 then there will be corrections to σ_{yx} of order $(e^2/h)(\sigma_{xx}^2/\sigma_{yx}^2)$, but these corrections can be made unobservably small by the proper choice of experimental parameters.

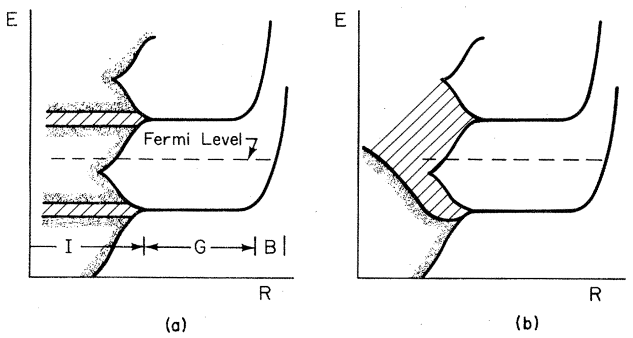

FIG. 4. Spectrum of the system in the guard ring G, the central impurity region I, and their junction. The system boundary is at B. The Fermi level is given by the dashed line. In the left diagram, the hypothesis is made that there exist gaps in the extended states (slashed region), separated by regions of localized states (cross-hatched boundaries). This case would give the ideal quantum Hall effect. To the right, it is assumed that some extended states near the junction rise up to the Fermi level.

IV. SCATTERING POTENTIALS

In this section we examine the effects of a potential which is due to a random array of scatterers. More precisely, consider a function $V_{sc}(x,y)$ which is 0 except in isolated regions, each such region being separated from every other by distances greater than l. Then the Hamiltonian has the form $H = H_0 + V_{sc}$, where

$$H_0 = -\frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{1}{2}\left[-i\frac{\partial}{\partial y} + x\right]^2 - \frac{1}{2} - vx \tag{4.1}$$

so that the applied field has been included in H_0. The eigenfunctions of H_0 have been written down in Eq. (2.6). Each eigenfunction belongs to a *distinct* energy. The spectrum of H_0 is completely nondegenerate. Owing to the magnetic field, even degeneracy from time-reversal invariance is lacking. This means that elastic scattering cannot change the direction of propagation of a wave packet, since its mean energy must be conserved. There is forward scattering only. We shall see that only elastic scattering is allowed. These facts are the physical origin of the "vanishing of the scattering rate" invoked by von Klitzing[3] to explain the original data. Formally it may be seen as follows. The Lippmann-Schwinger equation governs the scattering:

$$\psi(\vec{r},E) = \psi_0(\vec{r},E) + \int d^2r' G_0(\vec{r},\vec{r}',E) V_{sc}(\vec{r}')\psi(\vec{r}',E) \ , \tag{4.2}$$

where $\psi(r,E)$ is the scattering state of energy E, $\psi_0(r,E)$ is the state in the absence of scattering. The free retarded Green's function G_0 has been calculated in Eq. (2.8). Comparison with Eq. (2.6) shows that

$$G_0(\vec{r},\vec{r}',E) \sim i\phi(\vec{r},E)\phi^*(\vec{r}',E) \tag{4.3}$$

in the asymptotic region $|\vec{r}-\vec{r}'| \gg l$, $y \gg y'$. This factorization property follows directly from the nondegeneracy of the energy levels. Substituting the asymptotic form of G_0 into Eq. (4.2) gives

$$\psi(\vec{r},E)\sim e^{i\delta(E)}\psi_0(\vec{r},E) \tag{4.4}$$

in the asymptotic region. This means that a wave packet moving along the line $x=$const at speed v will suffer distortion and acceleration in the neighborhood of the scattering center but will ultimately recover its original speed and path. It does so after traveling a distance of order l away from the scatterer. This is unlike normal scattering where the length scale of the approach to the asymptotic region is set by the size of the potential. The normal free Hamiltonian contains no parameters having the dimensions of length, but that is not so here.

The most important aspect of the scattering of a wave packet for the purposes of the quantum Hall effect is the acceleration. Let the wave packet, before it reaches the scattering center, have the form

$$W(\vec{r},t)=\int dE\,F(E)\psi_0(\vec{r},E)^{-iEt}\,, \tag{4.5}$$

where F is some envelope function strongly peaked around E_0. The form for W long after it passes the site is

$$W(\vec{r},t)=\int dE\,F(E)\psi_0(\vec{r},E)e^{i\delta(E)-iEt}\cong\int dE\,F(E)\psi_0(\vec{r},E)\exp\left[i\left[\delta(E_0)+(E-E_0)\frac{d\delta}{dE}\bigg|_{E_0}-Et\right]\right]$$

$$=\exp\left[i\left[\delta(E_0)-E_0\frac{d\delta}{dE}\bigg|_{E_0}\right]\right]W_0(\vec{r},t+t_A)\,, \tag{4.6}$$

where $t_A=-d\delta/dE\mid_{E_0}$ and W_0 is the form for the wave packet in the absence of scattering. The packet is thus advanced from its unscattered position if $d\delta/dE\mid_{E_0}<0$, and retarded if $d\delta/dE\mid_{E_0}>0$. We shall see that the former circumstance is typical when the potential has bound states. This acceleration makes the extended states carry an extra current which compensates for the zero current carried by any localized states associated with the scattering center. Let us now see how this compensation can be shown to be exact.

Consider the Hamiltonian

$$H(\eta)=\frac{1}{2}\left[\frac{\partial^2}{\partial x^2}+\left[-i\frac{\partial}{\partial y}+x+\eta\right]^2-1\right]-vx+V_{\rm sc}(x,y)\,. \tag{4.7}$$

Direct differentiation gives $dH/d\eta=-j_y/e$, an operator equation. The Feynman-Hellmann theorem states that

$$\left\langle\psi_\alpha(\eta)\left|\frac{\partial H}{\partial\eta}\right|\psi_\alpha(\eta)\right\rangle=\frac{\partial E_\alpha}{\partial\eta}\,, \tag{4.8}$$

where $\psi_\alpha(\eta)$ are the normalized eigenvectors of $H(\eta)$ and $E_\alpha(\eta)$ their eigenenergies. Thus

$$\partial E_\alpha/\partial\eta=j_y^\alpha(\eta)\,. \tag{4.9}$$

The variation of η can be considered either as a gauge transformation $A=(0,-x)\to A'=A+f$, with $f=-\eta y$, or as a change in boundary conditions:

$$\psi(y=0)=\psi(y=L)\to\psi(y=0)=e^{i\lambda L}\psi(L)\,.$$

Equation (4.8) is closely related to Laughlin's argument. In that method the derivative on the right-hand side is replaced by a differential change of η from 0 to $2\pi/L$. We can then verify the statement concerning the invariance of energy levels by noting the invariance of the boundary conditions under such a change. The present derivation differs from that of Laughlin in that Eq. (4.7) is not summed over α, so that we can use it to find out which states carry what current. It also makes explicit what was implicit in that proof: that the quantized conductance holds also in simply connected geometries and without resorting to auxiliary fields. What we wish to do now is to combine Eq. (4.7) with the peculiar scattering theory of the system to give a microscopic picture of the Hall current.

For simplicity take the case of a single scattering center represented by $U(x,y)$ and choose coordinates such that the region of support of U is symmetrically placed with respect to the y axis, i.e., $U(x,y)=0$ for $|x|>x_0$. The change in the Hall current that the potential produces can be deduced from its effect on the energy spectrum of $H_0(\eta)=H(\eta)-U$ as shown by (4.7).

Since the observed current arises from the interplay of bound and extended states, an appropriate mathematical tool for the problem is Fredholm theory, which treats the two on an equal footing.[22] In this formulation we find the exact eigenenergies of the system by locating the zeros of the function

$$D(E)=\det\frac{E-H}{E-H_0}=\det[1-G_0(E)U]\,. \tag{4.10}$$

The behavior of these energies under variation of η can also be established. We relegate the details of the calculation to Appendix A and state the conclusions.

The states can be divided into three classes.

(i) Bound states, N_b in number, characterized by $dE/d\eta=0$, which naturally carry no current.

(ii) Extended states having $|E/v|-x_0\gg l$. These states are unaffected by the potential since they are out of its range. They have $dE/d\eta=v$, as in the pure case, and therefore their current is unperturbed.

(iii) Extended states which are scattered by U, and carry additional current. We will devote the remainder of the section to the analysis of the properties of these states.

Define a function $\Delta(E)$ which is the energy shift $E_\alpha-E_\alpha^0$ at $E=E_\alpha^0$ and the E_α are the eigenvalues of the extended states, paired with the E_α^0 in order from lowest to highest. Then the analytic properties of $D(E)$ yield the following behavior for $\Delta(E)$ over one Landau level: It is continuous and has limiting values $\Delta(E_{\rm min})=0$, $\Delta(E_{\rm max})=2\pi vN_b/L$. $E_{\rm min}$ and $E_{\rm max}$ are the edges of the

unperturbed band. The entire change in $\Delta(E)$ comes in the range $-vx_0 < E < vx_0$, i.e., from the states of class (iii).

$\Delta(E)$ is in fact related to the phase shifts of the scattering. In the region far from the impurity site the perturbed states have the form

$$\psi_\alpha(x,y)=C_\alpha(x)\exp[i\delta(E_\alpha)+iE_\alpha y/v] . \tag{4.11}$$

The energies satisfy

$$\frac{E_\alpha}{v}L+\delta(E_\alpha)=2\pi k_\alpha-\lambda L , \tag{4.12}$$

where k_α is an integer and $\delta(E_\alpha)=-L\Delta(E_\alpha)/v$. The behaviors of $\Delta(E)$ and $\delta(E)$ are thereby seen to be the form of Levinson's theorem[23] appropriate to the system. We are now ready to apply Eq. (4.8) to the scattering states. From Eq. (4.12) we see that as η runs from 0 to $2\pi/L$, $k_\alpha \to k_\alpha+1$. Equation (4.8) therefore becomes

$$\frac{-j_y^\alpha}{e}=\frac{L}{2\pi}(E_{\alpha+1}-E_\alpha)=v+\frac{L}{2\pi}\frac{d\Delta}{dE} . \tag{4.13}$$

Summing this over α:

$$\frac{-j_y^{\text{total}}}{e}=(N-N_b)v+\frac{L}{2\pi}[\Delta(E_{\max})-\Delta(E_{\min})]$$

$$=Nv , \tag{4.14}$$

where N is the number of states in a single Landau level.

The increased current of the scattered states precisely compensates that not carried by the bound states. This result is nonperturbative and completely independent of the magnitude of U. Since the compensating current comes from states for which the scattering is strong it is not surprising that treatments based on the Born approximation do not give this result.

If U_{sc} consists of many centers, it can be represented schematically as in Fig. 5(a). The multiple scattering analysis is trivial, however, since the stochastic element normally so difficult to deal with is not present. The scattering states simply follow the lines of constant potential in the region where U_{sc} is zero. At each encounter with a scattering center they suffer a phase shift, but these shifts simply add. It is easy to see that the conclusion is the same as for a single scattering center: The current flowing through any region is the same as if no impurities were present. The Hall conductance remains quantized.

What potentials can be treated as scattering potentials in this sense? For scattering theory to be valid we must be able to specify the boundary condition that the wave function approaches its asymptotic form to some desired degree of accuracy. Since $G_0(r,0,E)\exp(-r^2/l^2)$ we require that the scatterers be separated by a distance $R=sl$, where s is a numerical factor of order 1. If relative errors are to be of order 10^{-6}, the accuracy of the Hall-effect measurements, then $s\sim 4$. Circular impurities of radius d of increasing concentration are shown in Figs. 3(a) and 3(b). As the concentration increases, the minimum width of the potential-tree network decreases. At some critical concentration n_c this width will become less than R. Conversely, at n_c the scattering centers, now considered to be of radius $d+R/2$, themselves form a continuous network, a barrier to the passage of the Hall current. The alternative views are shown in Figs. 3(a) and 3(b). At n_c the scattering theory argument for the quantized Hall conductance ceases to be valid.

This percolation problem is equivalent to that of placing conducting disks on an insulating plane. At a concentration n_c of disks the plane becomes conducting. n_c has been determined numerically by Pike and Seager.[26] Their result, translated into our notation, is

$$n_c=\frac{0.35}{(d+R/2)^2} .$$

The numerical coefficient is accurate to within about 1%. If d is assumed to be short compared to the magnetic

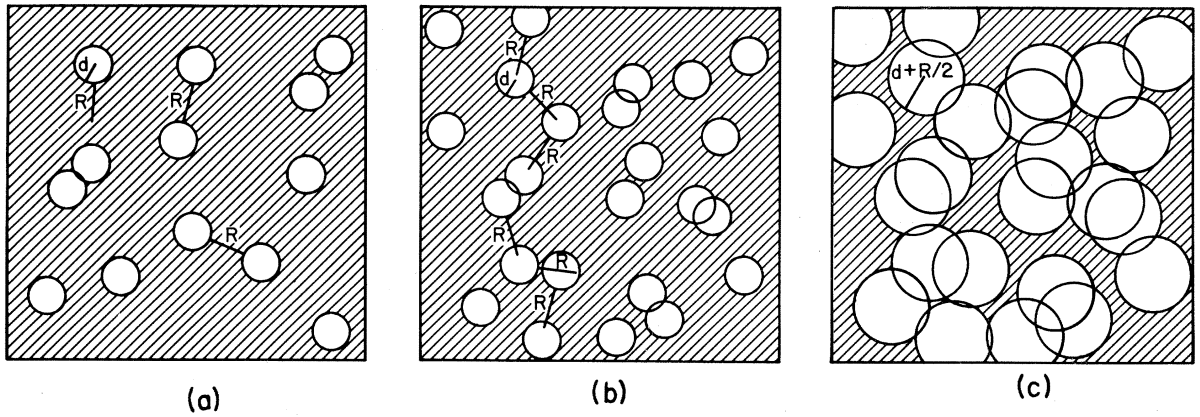

FIG. 5. (a) Scattering potential at less than the critical concentration. The shaded potential-free region contains rivers which are everywhere wider than R. If the sample were infinite, this region would form an infinite connected random network of such rivers. (b) Scattering potential at a concentration higher than the critical one. The infinite network of potential-free regions has been cut by a barrier of impurities. If the potential has large-scale homogeneity, then the barriers form an infinite network. The rivers of width greater than R are now isolated lakes within this network. (c) Alternative view of Fig. 3(b) showing the analogy to disc conduction. At the concentration shown, the discs will conduct because of the continuous path going up the left center of the picture. In an infinite system there will be an infinite number of such paths in any direction.

length, then this can be turned into a condition on the magnetic field

$$B_c=\frac{s^2}{1.40}\frac{\hbar cn}{e}\ .$$

Using a typical value $n=3\times10^{11}/\text{cm}^2$ this becomes (in units of T)

$$B_c=1.3s^2\ .$$

This is roughly the field strength at which the quantum Hall effect is seen.

If the impurities are not distributed at random, the picture changes slightly. A cluster of impurities can be grouped into a single center for the scattering analysis. This reduces the effective value of the impurity density and hence also the critical value of the magnetic field.

V. SMOOTH POTENTIALS

In this section we consider the case where both an applied electric field and a disordered potential are presented but the derivatives of the potential are small (in a sense which will be specified). What is required is the quantum equivalent of the guiding center approximation of classical electromagnetic theory.[24] In this approximation the motion of the particle is split into two parts: the fast motion of the particle around its circular orbit and the slow drift of the center of that orbit in the potential gradient. The center moves so as to cancel the time-averaged sum of electric and magnetic forces on the particle. Thus it gains no energy over a period, and the long-term motion is along a line of constant electric potential.

The expression for the Green's function in the presence of a potential U, of arbitrary magnitude, is

$$G(\vec{r},\vec{r}',t-t')=\int D(\vec{r}(\tau))\exp\left[\frac{i}{\hbar}\int_{t'}^{t}\left[M\frac{\dot{\vec{r}}^2}{2}-eB\dot{x}\dot{y}-U(\vec{r})\right]d\tau\right]\ .$$

We have returned to conventional units in order to facilitate comparison of relative magnitudes for large B. Classically the orbit center with coordinates (X,Y) moves according to the equations of motion:

$$eB\dot{X}=\frac{\partial U}{\partial Y}\ ,\quad eB\dot{Y}=-\frac{\partial U}{\partial X}\ . \tag{5.2}$$

Immediate consequences of this are

$$\dot{\vec{R}}\perp\nabla U\ ,\quad |\dot{\vec{R}}|=\frac{e}{B}|\nabla U|\ . \tag{5.3}$$

Our procedure will be to expand the paths in the functional integral about the guiding center path $(X(\tau),Y(\tau))$. This path in fact minimizes the sum of the last two terms in the integral. We write $x=X+u$, $y=Y+v$, and choose initial conditions $X(t')=x(t')$, $Y(t')=y(t')$. Then we can rewrite G as

$$G(\vec{r},\vec{r}',t-t')=\exp\left[\frac{i}{h}\int\left[\frac{M}{2}\dot{\vec{R}}^2-eB\dot{X}Y-U(X,Y)\right]d\tau\right]G'\ . \tag{5.4}$$

Here

$$G'(u_f,v_f,0,0,t-t')=\int D(u(\tau))D(v(\tau))\exp\left[\frac{i}{h}\int d\tau\left[\frac{M}{2}(\dot{u}^2+\dot{v}^2)-eB\dot{u}v-M\ddot{X}u-M\ddot{Y}u\right.\right.$$
$$\left.\left.-\frac{1}{2}u^2\frac{\partial^2U}{\partial X^2}-uv\frac{\partial^2U}{\partial X\,\partial Y}-\frac{1}{2}v^2\frac{\partial^2U}{\partial Y^2}\right]\right]\ . \tag{5.5}$$

In the expression for G' we have kept only terms up to quadratic in u and v and have integrated by parts where appropriate. $\partial^2u/\partial X^2$ is to be interpreted as $\partial^2u/\partial x^2|_{x=X(\tau)}$ and similarly for other such terms. u_f and v_f are determined by the guiding center equations of motion:

$$u_f=x-X(t)\ ,\quad v_f=y-Y(t)\ .$$

Some remarks about formula (5.5): (1) Since in contrast to the usual stationary phase approximation we did not choose the Euler-Langrange path about which to minimize, cross terms involving quadratic products of X, u, etc., appear in the action for G'. (2) Because we have subtracted out the motion of the orbit center, G' is the propagator for a particle in a magnetic field in the rest frame of the center of motion. The particle therefore sees a potential which is effectively time dependent since in this frame the (actually time-independent) potential is moving by it. (3) The first factor is just the exponential of the "action" of a particle whose differential equation of motion is (5.2). Since U is independent of time this factor is a function of the time difference $t-t'$ only. Hence this must be the case for G' as well.

In Appendix B the effects of the last five terms in the action for G' are analyzed. If B is large or the potential is smooth, they can be treated as perturbations. The precise conditions are

$$|\nabla^2 U|\,|\vec{\nabla} U| \ll \frac{\omega_c^2}{l^3} \propto B^{7/2}$$

and

$$|\nabla^2 U| \ll \frac{\omega_c}{l^2} \propto B^2$$

for the first two terms involving the derivatives of U to be small. The conclusion of Appendix B is that these terms just serve to shift and distort, in classical language, the circular motion of the orbit. The Green's function of energy E stills falls off exponentially on the scale of l away from the equipotential line of that energy. Its qualitative behavior is not altered in this respect from the Green's function of Eq. (2.10). Thus there will be an equipotential line associated with each state. The overall correspondence of the contour map of the potential and the electronic configuration is preserved if the potential is smooth. In this section we ignore the higher-order terms, and the conclusions stated below should be understood in that context.

The overall form for G allows us to characterize the states and settle questions of localization versus extension for this sort of potential. A contour map for a smooth potential which includes an applied field looks schematically as in Fig. 6(a). The entire area may be divided into two parts: a connected region of open-contour lines and a disconnected region of closed-contour lines. The ratio of the areas of these two regions is a measure of the amount of disorder in the system and in fact determines the ratio of localized to extended states. We can see this by the following arguments.

1. Localized states (closed contours). Take two electrons in the same Landau level and in neighboring orbits. The difference in their energies, ΔE, is given by the Bohr frequency condition as h/T, with T the period of motion around the contour. The Bohr condition is justifiable for potentials varying slowly on a scale of l, since one state occupies an area l^2, so that the quantum numbers involved are large. The area between the orbits is

$$\Delta S = \int \frac{\Delta E}{|\Delta U|} v\, dt\ ,$$

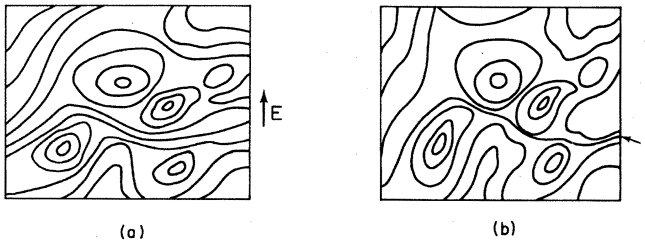

FIG. 6. (a) Equipotential map of a smooth potential with a nonzero applied field in the direction shown. A finite area is associated with the open lines of the potential. This area will percolate from one end to the other and will therefore contain extended states. (b) Smooth potential without an applied field. In the infinite-area limit, all contour lines will be closed, except for possible isolated lines passing through saddle points. One such line is indicated.

where the integral is taken around the orbit. By using now Eq. (5.2) to find v, $\Delta S = 2\pi l^2$. Hence in a region of area A of closed contours there are $A/2\pi l^2 + n_0$ states, with n_0 a constant. n_0 can be shown to be of order 1 by looking at the properties of the functions (2.2). Thus in the present approximation, the density of electronic charge remains essentially uniform in the neighborhood of impurities (if all the states are full) and is equal to its unperturbed value. This contrasts with the usual situation since the strength of the potential plays no role. It is a consequence of the fact that the energy $\hbar\omega_c$ determining the length scale l and thus the density is much greater than the energy $|l^2\nabla^2 U|$ which would prefer a different density. The relative shift in density of a full Landau level is locally about $\Delta n/n \approx l^2 |\nabla^2 U|/\hbar\omega_c$. This small change in density does not change the net current, however.

2. Extended states (open contours). A similar argument, with similar results, can be applied to these states. The separation of neighboring energy levels is determined by the periodic boundary conditions since the function higher in energy must have one more oscillation traversing the system. To apply this condition we need to look more closely at the first factor of G in Eq. (5.4):

$$\exp\left[\frac{i}{\hbar}\int d\tau [eB\dot{Y}X - U(X,Y)]\right].$$

The first term is the result of a gauge change and the kinetic energy of the slow motion of the orbit center has been dropped. (It is easy to show that it is negligible for large B.) Recalling that U is constant along the guiding center path and using

$$\int d\tau\, \dot{Y}X = \int X\, dY\ ,$$

we find an exponential factor $\exp[-iU(t-t')+iBSe]$, where S is the area between the equipotential line U and the y axis. (A change of gauge changes the axis relative to which S is measured.) The eigenfunction expansion for G is

$$\sum_n \psi_n(\vec{r})\psi_n^*(\vec{r}')\exp[-iE_n(t-t')]\ ,$$

so we see that the phase advance is proportional to the area S. Hence the phase difference between two neighboring extended states is just the area between them divided by l^2. (This quantity is gauge invariant, as of course it must be.) Setting this equal to 2π gives $\Delta S = 2\pi l^2$. The conclusion is that the total density in the open-contour region remains also equal to the unperturbed value. Therefore the ratio of localized to extended states is equal to the ratio of the areas of closed- and open-contour lines.

We can now use this approximation to get the Hall current. As we have seen, the essential coordinate dependence of G comes from the exponential factor, so that we can calculate the current from it alone. Details can be found in Appendix B. From Eq. (5.3) it is evident that charge is transported at a rate $c|\Delta U|/B$ in the direction $\Delta U \times \vec{B}$. The charge crossing a line $y = \text{const}$ per unit time is given by

$$-n_e \int_{-W/2}^{W/2} \dot{Y} dx\ .$$

There is no contribution from regions of closed-contour lines since

$$\int \dot{Y}\,dx = -\Delta U/eB = 0$$

across such a region. The open-lined regions yield

$$n_e c \left[U\left[\frac{W}{2}\right] - U\left[-\frac{W}{2}\right] \right] \Big/ B \ ,$$

where n_e is the electron density in these regions. If all extended states are filled (not the same as the entire Landau level being filled), we insert our earlier results to obtain a current of $eEW/2\pi l^2 B$, corresponding to a Hall conductance e^2/h. This shows that the Hall conductance is not affected by a slowly-varying potential. It also verifies the conjecture of Tsui and Allen[25] that the impurities "punch out holes" of zero conductivity but the remaining material retains the same conductivity. Since the conductance is independent of area it is unchanged. Note that the present approximation, by exactly the same reasoning as above, gives zero transverse conductivity σ_{xx}.

Although these arguments give a good picture of the type of states encountered in the case of a smooth potential, a question remains. This is the question of the magnitude of the corrections coming because of the finiteness of the potential gradients. The actual arguments given would suggest that the corrections are ordinary perturbations, and thus of the order of some power of the ∇U. However, experience in the guiding center problem suggests that the ripples in the potential can be regarded as *adiabatic* perturbations. This sort of approximation usually has exponentially small corrections in the regime where it is applicable. In the present context that would mean that the current calculated above is correct except for unobservably small deviations as long as the conditions (5.6) hold. This arises physically because the time scale of the perturbations in the moving frame is $v\nabla U/U$, whereas the underlying time scale is $1/\omega_c$. More precisely, there are ordinary perturbation corrections to the guiding center trajectory, or in this case, to the wave function, but certain overall quantities suffer corrections which are exponentially small, i.e., of order $\exp(-\hbar\omega_c/l\,|\,\nabla U\,|\,)$. Thus the wave packets move along lines which are not quite the equipotential lines, the density is not quite the ideal density, etc., but the current is unaffected. However, to prove this by directly resuming the perturbation theory in ∇U is cumbersome. We shall therefore resort to an indirect argument below to reach this conclusion more rigorously.

Before doing this we must consider whether there are any extended states at all in the system. For the case of a scattering potential this is assured becuase the "flat" regions will contain some such states. The smooth potential, on the other hand, can be nonzero almost everywhere in the system. A picture of such a potential is shown in Figs. 6(a) and 6(b) for different situations. If there is no applied field, then one can expect *all* states to be localized. This is easiest to see in Fig. 6(b), where a part of an infinite system is shown. All contours must be closed, except for isolated ones which pass through saddle points. The contour map for an infinite system of this type has been analyzed in general by Zallen and Scher,[26] and in the specific context of the quantum Hall effect by Iordansky[27] and by Kazarinov and Luryi.[28] They conclude that in the absence of an applied field only states at a single critical energy can percolate to infinite distance. The actual argument is closely related to that of the preceding section. Suppose there were two equipotential lines belonging to different energies which percolate. In the absence of an applied electric field there is no difference between the x and y directions, so the two percolating lines of different energy would have to cross, which is impossible by their definition.

There are no percolating states in the absence of applied field because the line occupies an area of measure 0, and the number of states is proportional to the area. In finite applied field the situation is different. Now the x direction is different than the y direction and the percolating line widens into a strip of finite area as shown in Fig. 6(a). In other words, the electric field delocalizes some of the states. The number of such states is proportional to the electric field, for small field.

We now see that the general line of reasoning which gives the ideal values for the Hall conductivity applies to the case of smoothly-varying potentials. Namely, the only states which can carry current are those which percolate, and these are in a very narrow band of energies. All the other states are localized. Exactly what this percolation energy is, and exactly where the percolating states lie is a perturbation problem, given the fluctuating potential $U(x,y)$, although to first approximation, the problem can be solved by consideration of the equipotential lines of U. As long as the Fermi level is well away from this percolation energy, however, the Laughlin-Halperin argument has force, and the ideal Hall conductivity is obtained.

Good samples at low temperature may have U's which are smooth. Then the mobile states should lie in a narrow energy band, which implies that the transverse conductivity should have wide plateaus, when the Fermi level is away from this band, and narrow regions between plateaus (where also $\sigma_{xx} \neq 0$) when the Fermi level is in the midst of the band of states. Since the width of this band is a monotonic function of applied electric field, this would suggest that the steepness of the Hall steps should increase as the field is made smaller, and that the conductivity there should be non-Ohmic. The latter effect has been observed.[29]

VI. THE GENERAL CASE

In this section we use the results of the previous three sections to analyze the general question of which kinds of disordered potentials will produce the Hall effect and which will not. First take the case of a scattered potential superimposed on a smooth potential so that the Hamiltonian is

$$H = H_0 + V_{\rm sm} + V_{\rm sc} \ , \tag{6.1}$$

where H_0 includes the applied electric field, $V_{\rm sm}$ is the smooth potential, and $V_{\rm sc}$ is the sum of isolated scattering center potentials. The Green's function for $H_0 + V_{\rm sm}$, $G_{\rm sm}(\vec{r},\vec{r}',t-t')$, is that of Sec. V. It practically vanishes unless $\vec{r}(t)$ is close to the solution of the guiding-center

equation of motion with initial condition that the particle is at point $\vec{r}'$ at time t'. In particular G_{sm} is exponentially small unless $\vec{r}$ and $\vec{r}'$ lie on the same equipotential line. This quasi-one-dimensionality arises from the same cause as in the pure system—the lack of degeneracy. Now consider the scattering problem posed by the Hamiltonian (6.1). A particle is traveling along a possibly curved equipotential line when it encounters a scattering site. The one-dimensionality of G_{sm} will restrict scattering to be in the forward direction, i.e., along the equipotential. The asymptotic states may be classified in precisely the same way as in Sec. IV and the same phase-shift analysis applied locally. The total current flowing through the region containing the scattering potential will be the same as if that potential were absent. This remains true even if the scattering center binds some electrons. So even if both types of potential are present the quantized conductance may be expected.

We have seen that if the mobile states belonging to a given Landau level are distinct in energy from those belonging to other levels, then the Laughlin-Halperin argument holds. Clearly this condition holds for potentials characterized by Eq. (6.1).

Let us next consider the case $V=V_L+V_{sc}$, where V_L is considerably less in magnitude than $\hbar\omega/2$, e.g., $\max|V_L|=f\hbar\omega_c/2$. The problem is that we know little in detail about the effects of V_L. However, the limitation on its magnitude will enable us to bound perturbation theory to estimate certain quantities.

Consider the Green's function $G(\vec{r},\vec{r}',E)$ for $|\vec{r}-\vec{r}'|\equiv R$ very large. If G drops off exponentially with increasing R then there are only localized states at this energy. (Note that we are *not* considering an *average* G which could become small because of random phase cancellations.) Consider doing perturbation theory for G in powers of V_L. Then

$$G(\vec{r},\vec{r}',E)=G_{sc}(\vec{r},\vec{r}',E)+\int d\vec{r}_1 G_{sc}(\vec{r},\vec{r}_1,E)V_L(\vec{r}_1)\times G_{sc}(\vec{r}_1,\vec{r}',E)+\cdots +G_{sc}V_L G_{sc}V_L G_{sc}V_L\cdots V_L G_{sc}+\cdots , \tag{6.2}$$

and we have used a matrix notation to indicate the general term. Take $E=\frac{1}{2}=\hbar\omega_c/2$, halfway between the energy of the extended states of the unperturbed problem. At this energy G_{sc} vanishes exponentially as a function of separation R, because there are no extended states at this energy. In order that G not be exponentially small therefore, there must be a term in the series (6.2) which has integrations over a chain of intermediate positions $\vec{r}_1$, $\vec{r}_2,\ldots,\vec{r}_n$, such that $|\vec{r}_j-\vec{r}_{j-1}|\approx l$. To be a bit more precise, since the potential inside the scattering regions is arbitrary, we may assume that V_L vanishes there and is only nonvanishing outside the scattering regions. There will be a minimum distance D *between* scattering regions which is necessary to traverse to get from $\vec{r}$ to $\vec{r}'$ with $D\gg l$. Thus there will have to be at least about $n=D/l$ factors of $V_L G_{sc}$ in the terms of (6.2) that could contribute to a long range for G.

The Green's functions will have their arguments outside the scattering regions. They therefore may be approximated by expressions of the form

$$G_{sc}(r,r',E)=\sum{}'\phi_\alpha(r)\phi_\alpha^*(r')/(\hbar\omega_c/2) ,$$

where the sum is over extended states of essentially zero energy. It follows at once that $G\sim f^{D/l}$, and since D will increase on average proportionately to R, G is exponentially decaying for large R. Therefore, the states of this energy, halfway between the energies of the extended states, are localized, and the Hall-effect argument is applicable.

This argument can be extended in a straightforward way to the general case in which a smooth potential is added. Then this smoothness guarantees that, except for isolated scattering regions, the spatial region near the percolating energy contour will be the support of eigenstates with energies near the percolation energy, and that this region will be wide compared with l. Again, we do perturbation theory in powers of V_L for an energy equal to the percolation energy plus $\frac{1}{2}$. Again, the unperturbed Green's functions will decay exponentially, and this can only be counteracted by looking at high-order terms in the expansion. Again, the chain of intermediate states in these high-order terms will have to cross the region about the percolation path many times, and many factors of $V_L G_{sc}$ will be encountered, each of which will be bounded by $f/(\Delta E/2)$. Here ΔE may be somewhat less than $\frac{1}{2}$, because the energy difference between E and the states in the band can be somewhat smaller than $\frac{1}{2}$. However, if we make the bound on V_L somewhat stronger, we shall again have proven that the states sufficiently far from the percolating states remain localized, which suffices to prove that the Hall effect is ideally quantized.

We have thus found that the structure of the states in a potential which consists of arbitrary but isolated scattering centers connected by smoothly-varying regions is resistant to arbitrary perturbations, provided these are sufficiently weak. This structure is that the extended states lie in very narrow bands of energy corresponding to the Landau levels, and that their wave functions are large only in definite linear regions which percolate across the sample. The wave functions which are large close to these regions will have energies close to those of the extended states. Wave functions which correspond to energies well away from the extended states are also spatially isolated from the extended-state wave functions. Because energetic isolation correlates with physical isolation, a weak but otherwise arbitrary potential cannot generate hopping of long range at energies well away from the original extended wave functions.

It is not yet clear that adding a potential V_L *will* generate new extended states close in energy to the original narrow band of such states. Indeed, experience on the nonmagnetic problem suggests rather that a random potential tends to localize states which were originally extended. Here we know that this result is not possible, or at least that there must remain some extended states, even though it is no longer clear how narrow the band of energies which they fall in is. On the other hand, nothing in the nonmagnetic problem suggests a strong spatial correla-

tion of localized and extended states.

Presumably, a sufficiently strong potential V_L will destroy the Hall effect. At intermediate strengths it must do this by creating extended states midway between Landau levels, else the ideal Hall effect cannot disappear, by Halperin's argument. (Imagine that the potential in going from the guard ring to the interior is turned on slowly.) For extremely strong potentials, it is likely that the states are completely localized, since this case corresponds to the low-field, low-density limit, and it is difficult to imagine that the states in the Lifshitz tails are delocalized by a weak magnetic field. (Indeed, the Lifshitz tail states are spatially well separated.) The detailed evolution of the states as V_L is increased in magnitude remains an open question, however.

Thus if (at very low temperature) there is no quantum Hall effect, the potential could be such that (1) all states are localized, or (2) the bands of extended states are broadened sufficiently that there are no mobility gaps. The former case seems to have been observed in low-lying Landau levels. There are no plateaus and the transverse conductance is activated.

The localization of electrons in the strong magnetic field case differs from Anderson localization in two ways. The first is the spatial isolation of the localized states, as we have discussed. The second is the lack of stochasticity in the scattering processes. With a strong field since the sequence of scattering centers visited is linear and predictable with certainty, a wave packet always escapes from an array of random scatterers in the end. On the other hand it will not escape from its equipotential line. If that line is closed the electron will be localized. The overall scale of a localized state is not given by a characteristic exponential envelope but will depend on the contour map of the smooth background potential. On the other hand, the wave function is confined exponentially (indeed as a Gaussian) to the region to which it belongs.

Finite temperatures will of course destroy the quantum Hall effect when $k_B T$ is of the order of the Landau-level spacing, because the preceding arguments depend on the assumption of elastic scattering. If inelastic processes of sufficient energy to excite extended states to the Fermi level are available, then the ideal Hall effect must disappear.

VII. CONCLUSION

Insofar as the independent electron model is realistic, we conclude that for a given potential which is not pathological, there exists a magnetic field sufficiently strong so that the potential is weak in the sense that it is a sum of isolated scattering centers, a potential smooth on the scale of the magnetic length, and a potential of magnitude less than $\hbar\omega_c/2$. Estimates of what the potential is like in realistic systems indicate that at presently available fields, the potential is close to being weak. The fact that in good samples the effect clearly exists, while in poor ones it does not, is *a posteriori* evidence for this.

The nature of the states, and of their localization, extension, and scattering properties is quite clear and is also rather different from the zero-field case. Basically, the magnetic field controls one dimension of the wave function, and makes the states to be rather linear, lying near lines not too distorted from the equipotential lines of the potential. Thus states associated with closed loops of equipotential lines are localized, and states associated with percolating equipotential lines are extended. Potentials with large curvatures can distort this picture, but not completely destroy it, provided the regions of large curvature are isolated, or the net amplitude of the fluctuating potential is sufficiently small. The former fails to destroy the picture, because the isolated regions can only cause a forward scattering of the electrons. The potential of small magnitude cannot destroy the picture because it generates a convergent perturbation theory which cannot delocalize states at energies halfway between Landau levels.

Presumably, many of the same arguments can be used in the interacting case. Much more remains to be done in that direction, however. Even in the noninteracting model, we do not completely understand how to find the width and other details of the steps as a function of field, impurities, temperature, and density. There is also evidence that the electron-phonon interaction is strong.[30,31] These questions are connected with the details of how the quantum Hall-effect phenomena break down as the potential becomes stronger.

ACKNOWLEDGMENTS

We would like to acknowledge useful and interesting conversations with S. Das Sarma, V. Korenman, S. Fishman, and S. Sigur. We thank G. Kote and N. P. Ong for a copy of their work in advance of publication. This work was supported in part by National Science Foundation Grant No. DMR 7908819.

APPENDIX A

Here we prove the statements made in the text concerning the behavior of $\Delta(E)$. $D(E)$ has been defined earlier as

$$D(E)=\prod_{\alpha}\frac{E-E_\alpha}{E-E_\alpha^0},$$

where E_α^0 are the unperturbed eigenenergies and E_α are the energies of the full Hamiltonian. The number of factors in the numerator equals the number in the denominator since states can neither bifurcate nor disappear as V_{sc} is changed. (This can be proved rigorously.[32] The precise conditions are that the original spectrum be discrete and V_{sc} be a bounded operator.) We will examine the structure of D in a single Landau level, so we restrict the E_α^0 to belong to this level and define the E_α to be the energies of the states which develop from this level adiabatically as V_{sc} is turned on. The generalization to many levels is straightforward.

Among the energies E_α, N_b will correspond to bound states of V_{sc}. Let us separate these out from the product, as well as an equal number from the denominator:

$$D(E+i\epsilon)=\prod_{\beta=1}^{N_b}\frac{E-E_\beta+i\epsilon}{E-E_\beta^0+i\epsilon}\prod_{\alpha=1}^{N_L-N_b}\frac{E-E_\alpha+i\epsilon}{E-E_\alpha^0+i\epsilon}.$$

ϵ is a positive infinitesimal. The E_β will normally lie outside the continuum when they correspond to normalizable states of physically realistic potentials.[33] The choice of the E_β is discussed below. The second factor is now a function with $N_L - N_b$ zeros on the real axis. Each of these zeros is paired with a zero of the denominator which lies within a distance of order $(1/L)$. In the thermodynamic limit, that is, $L \to \infty$ followed by $\epsilon \to 0$:

$$D(E+i\epsilon)=\prod_{\beta=1}^{N_b}\frac{E-E_\beta+i\epsilon}{E-E_\beta^0+i\epsilon}\prod_\alpha\left[1-\frac{E_\alpha-E_\alpha^0}{E-E_\alpha^0+i\epsilon}\right]$$
$$=\frac{E-E_\beta+i\epsilon}{E-E_\beta^0+i\epsilon}\left[\exp\left[\frac{L}{2\pi v}\right]\int\frac{\Delta(E')dE'}{E'-E+i\epsilon}\right].$$

$\Delta(E')$ is defined as in the text: $\Delta(E')=E_\alpha-E_\alpha^0$ at $E_\alpha^0=E'$. To make the definition precise we must specify how the new and old energies are to be paired. It is convenient to take the E_β^0 to be the N_L-N_b highest of the unperturbed eigenvalues and to pair them with the E_β, the bound-state energies, in order from lowest to highest. The E_α^0, E_α pairs are also made in order from lowest to highest. Then as $L\to\infty$, the E_β^0 are all equal to the maximum value of the original energies, for example, $E_{\max}$, and $\Delta(E)$ is a continuous function which is 0 at $E=E_{\min}$, the lower end of the unperturbed spectrum. $D(E)$ is analytic except for a branch cut on the real axis where H_0 has its eigenvalues for the Landau-level question. Across the cut D has a phase change but $|D|$ is continuous.

We now wish to find $\Delta(E_{\max})$ since that will yield the information we want about the total current. To do this, the continuity of $|D(E_{\max})|$ is used. We can write this function as

$$D(E)=\left[\prod_\beta|E-E_\beta|\right]|E-E_{\max}|^{-N_b}$$
$$\times\exp\left[\frac{L}{2\pi v}\mathrm{P}\int_{E_{\min}}^{E_{\max}}\frac{\Delta(E')dE'}{E'-E}\right].$$

The singularities at $E=E_{\max}$ in the two factors must cancel out by the continuity condition. The second term in the exponent does not contribute because the derivatives of Δ vanish at $E=E_{\max}$. Continuity then implies

$$\exp\left[\frac{L}{2\pi v}\Delta(E_{\max})\ln|E-E_{\max}|\right]\sim|E-E_{\max}|^{N_b}$$

or

$$\Delta(E_{\max})=\frac{2\pi v}{L}N_b\ .$$

$\Delta(E)$ increases from 0 at the lower edge of the Landau level to $2\pi v N_b/L$ at the upper edge.

APPENDIX B

Here we address the question of when derivatives of $V_{\rm sm}$ higher than the second can be neglected. The action for G' in Eq. (5.5) in principle contains terms involving all such derivatives. Define $G_0'(u_f,v_f,0,0,t-t')$ to be the propagator for the first three terms of the action of G'. The time Fourier transform of this function has already been calculated in Eq. (2.8) and falls off as $\exp[(u_f^2+v_f^2)/4]$. Perturbative corrections to physical quantities will be given by expressions involving the integral of powers of G_0' times other terms in G'. For example, any integral involving the third term in G' will be proportional to

$$m\ddot{X}l=ml\frac{c}{eB}\frac{d}{dt}\frac{\partial V_{\rm sm}}{\partial Y}$$
$$=ml\frac{c}{eB}(\dot{\vec{R}}\cdot\vec{\nabla}_{\vec{R}})\frac{\partial V_{\rm sm}}{\partial Y}$$
$$=ml\frac{c^2}{e^2B^2}\left[\frac{\partial^2V_{\rm sm}}{\partial X\partial Y}\frac{\partial V_{\rm sm}}{\partial Y}-\frac{\partial^2V_{\rm sm}}{\partial Y^2}\frac{\partial V_{\rm sm}}{\partial X}\right]$$

A similar analysis may be made for the other terms. The magnitudes of the third and fourth terms relative to the zero-order terms are

$$|\nabla^2V_{\rm sm}||\vec{\nabla}V_{\rm sm}|l^3/\omega_c^2$$

and

$$|\nabla^2V_{\rm sm}|l^2/\omega_c\ ,$$

respectively. For perturbation theory to hold these must be much less than unity everywhere. If that is so then $V_{\rm sm}$ is really smooth. The main point, however, is not that the perturbations are very small, but that the electrons still follow the equipotential lines even if there are perturbations. If we write the Hamiltonian in terms of the moving coordinates

$$\vec{\rho}=\vec{r}-\vec{R}(t)\ ,$$

we find

$$H(\vec{\rho})=-\frac{\hbar^2}{2m}\nabla_\rho^2+\frac{eB}{emc}L_z(\vec{\rho})+\frac{e^2B^2}{8mc^2}(\rho_x^2+\rho_y^2)$$
$$+m\ddot{\vec{R}}(t)\cdot\vec{\rho}+\tfrac{1}{2}(\vec{\rho}\cdot\vec{\nabla}_R)^2V_{\rm sm}\ .$$

$L_z(\vec{\rho})$ is the z component of the angular momentum operator. The third term is an oscillator potential. The last has its gradients evaluated at $R(t)$. The perturbations are simply adiabatic changes in the parameters of a harmonic oscillator, which means that the density will undergo slow variations in the moving frame, but will remain as a bound state in that frame. In the laboratory frame, a wave packet which is once localized on an equipotential line will remain so forever.

[1]K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2]D. C. Tsui and A. C. Gossard, Appl. Phys. Lett. 38, 550 (1981).
[3]K. von Klitzing, in *Festkoerperprobleme,* Vol. XXI of *Advances in Solid State Physics,* edited by J. Treusch (Vieweg, Braunschweig, 1981), p. 1.
[4]D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. 48, 1559 (1982).
[5]Some of this work has appeared in R. E. Prange and R. Joynt, Phys. Rev. B 25, 2943 (1982). See also R. Joynt, Ph.D. thesis, University of Maryland, College Park, 1982 (unpublished).
[6]T. Kinoshita and W. B. Lindquist, Phys. Rev. Lett. 47, 1573 (1981).
[7]R. B. Laughlin, Phys. Rev. Lett. 50, 1395 (1983).
[8]D. C. Tsui (unpublished).
[9]E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. 42, 672 (1979).
[10]B. L. Altshuler, K. Khmel'nitzkii, A. I. Larkin, and P. A. Lee, Phys. Rev. B 22, 5142 (1980).
[11]M. J. Uren, R. A. Davies, and M. Pepper, J. Phys. C 14, L395 (1980).
[12]R. B. Laughlin, Phys. Rev. B 23, 5632 (1981).
[13]R. E. Prange, Phys. Rev. B 23, 4802 (1981).
[14]T. Ando, A. Fowler, and F. Stern, Rev. Mod. Phys. 54, 437 (1982).
[15]F. F. Fang and W. E. Howard, Phys. Rev. Lett. 16, 797 (1966).
[16]S. Das Sarma, Solid State Commun. 36, 357 (1980).
[17]T. Ando, J. Phys. Soc. Jpn. 43, 1616 (1973).
[18]B. I. Halperin, Phys. Rev. B 25, 2185 (1982).
[19]N. Byers and C. N. Yang, Phys. Rev. Lett. 7, 46 (1961).
[20]D. J. Thouless, J. Phys. C 14, 3475 (1981).
[21]We are indebted to S. Fishman for this proof.
[22]M. Baker, Ann. Phys. (N.Y.) 4, 27 (1958).
[23]G. E. Pike and C. H. Seager, Phys. Rev. B 10, 1421 (1974).
[24]T. G. Northrop, *The Adiabatic Motion of Charged Particles* (Interscience, New York, 1963).
[25]D. C. Tsui and S. J. Allen, Jr., Phys. Rev. B 24, 4082 (1981).
[26]R. Zallen and H. Scher, Phys. Rev. B 4, 4471 (1971).
[27]S. V. Iordansky, Solid State Commun. 43, 1 (1982).
[28]R. F. Kazarinov and S. Luryi, Phys. Rev. B 25, 5566 (1982).
[29]G. Kote and N. P. Ong (unpublished).
[30]T. Ando, Surf. Sci. 73, 1 (1978).
[31]H. Ekawa, S. Kawaji, and K. Nakamura, Jpn. J. Appl. Phys. 13, 126 (1974).; S. Das Sarma and R. Joynt (unpublished).
[32]K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (American Mathematic Society, Providence, 1965).
[33]M. Reed and B. Simon, *Analysis of Operators* (Academic, New York, 1978).