Hall resistance,

$$R_{\mathrm{H}} = \frac{V_{\mathrm{H}}}{I} = \frac{WE_y}{Wj_x} = -\frac{B_z}{ec\rho}, \tag{15.7}$$

is linearly related to the applied magnetic field, while the longitudinal resistance,

$$R_{\mathrm{L}} = \frac{V_{\mathrm{L}}}{I} = \frac{LE_x}{Wj_x} = \frac{L}{W}\frac{1}{\sigma_0}, \tag{15.8}$$

is independent of the magnetic field. Experimentally, these relationships agree well with Hall's measurements in 1879 (H1879). Note that R_{H} as we have calculated it here is independent of the details of the electron scattering processes. In general there is a weak dependence on such processes.

In 1980 von Klitzing and collaborators (KDP1980) noticed striking deviations from the resistances given by Eqs. (15.7) and (15.8) at sufficiently high magnetic fields and ultra-low temperatures. They performed their remarkable experiments on a 2d electron gas confined at the interface between SiO_2 and Si in an Si MOSFET. They found that the Hall voltage as the magnetic field increased exhibited distinct flat or plateau regions that were highly reproducible from sample to sample, as shown in Fig. 15.2. From the value of the Hall voltage at the plateau regions, they deduced that the Hall conductance (the inverse of the resistance) must be quantized in units of e^2/h. The quantization of the conductance in the form

$$\sigma_{\mathrm{H}} = -\frac{ne^2}{h}, \tag{15.9}$$

where n is an integer, was found to hold for one part in 10^7. Equally surprising was the behavior of the longitudinal resistance. They found that the longitudinal resistance vanished exactly at the Hall plateaus, indicating the onset of dissipationless transport. The presence of both quantization and perfect conduction indicates that something quite fundamental is at the heart of these experiments. At much higher magnetic fields and lower temperatures, Tsui, Störmer, and Gossard (TSG1982) found that the plateau regions in the Hall voltage were more plentiful than had been thought possible. They found that the plateau regions can occur when the conductance is a fractional multiple of e^2/h, indicating the presence of fractionally charged excitations. We will show that the integer Hall effect can be understood simply as a quantization of the edge current, whereas the fractional effect arises from a fundamentally new correlated many-body state, the Laughlin state (L1983).

15.2 Landau levels

To start, we solve for the wavefunctions describing an electron moving in a plane pierced by a perpendicular magnetic field. Following Landau's original treatment, we orient the

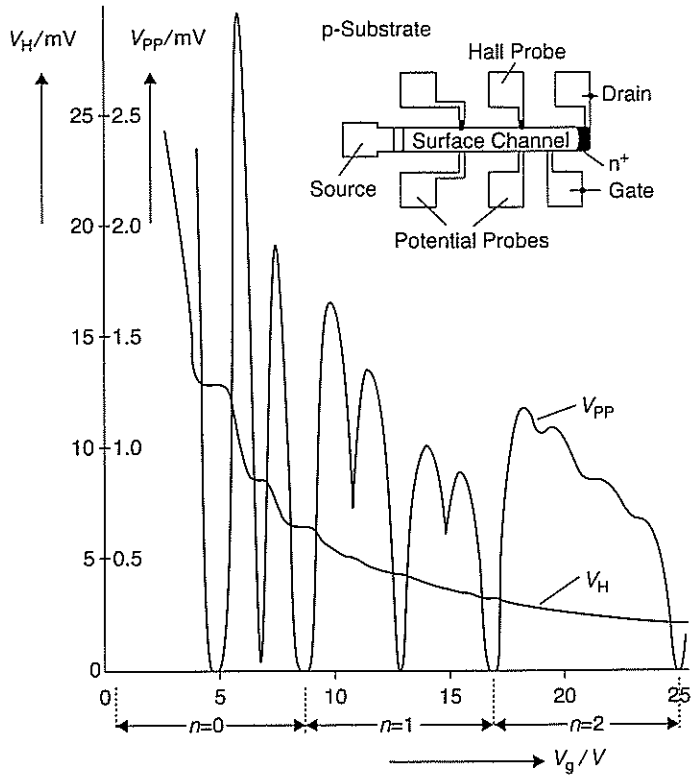

Fig. 15.2 Hall voltage, V_H, and the potential drop between the potential probes, V_{pp}, as a function of gate voltage V_g at $T = 1.5$ K. The magnetic field was held fixed at 18 T and the source drain current at 1 μA. Shown in the inset is a top view of the device with a length of $L = 400$ μm, a width of $W = 50$ μm, and a distance between the potential probes of $L_{pp} = 130$ μm. The Hall plateaus occur at integer values of the filling in each Landau level indicated with the index, n. At the plateaus in the Hall voltage, the longitudinal voltage goes to zero, indicating the presence of dissipationless transport.

vector potential

$$A_y = Bx, \quad A_x = 0, \tag{15.10}$$

along the y-direction and the single-particle Schrödinger equation takes the form

$$-\frac{\hbar^2}{2m}\left(\partial_x^2 + (\partial_y - ieBx/\hbar c)^2\right)\psi(x,y) = E\psi(x,y). \tag{15.11}$$

This choice of gauge is most convenient to describe transport in the integer quantum Hall effect. In the context of the fractional quantum Hall effect, however, we will find it expedient to work in the symmetric gauge in which

$$\mathbf{A} = \frac{B}{2}(y\hat{x} - x\hat{y}). \tag{15.12}$$

In the symmetric gauge, applying a magnetic field in the z-direction leads to a harmonic oscillator problem along both the x- and y-axes. Hence, this problem can easily be solved once the solution to the simpler problem described by Eq. (15.10) is obtained.

Translational invariance in the y-direction suggests that we write the wavefunction as

$$\psi_{n,k}(x,y) = \mathrm{e}^{iky} f_n(x). \tag{15.13}$$

Substitution of $\psi_{n,k}(x,y)$ into Eq. (15.11) reveals that $f_n(x)$ is a solution to a harmonic oscillator equation

$$\frac{\hbar\omega_c}{2}\left(-\ell^2\partial_x^2 + (x/\ell - \ell k)^2\right) f_n(x) = \epsilon_n f_n(x), \tag{15.14}$$

where, unlike the cyclotron frequency, the length scale

$$\ell \equiv \sqrt{\frac{\hbar c}{eB}} = \frac{250\text{Å}}{\sqrt{B}} \tag{15.15}$$

is independent of the effective mass and is changed entirely by varying the magnetic field. Known as the magnetic length, ℓ is roughly 250 Å for a field of $B = 1$ T. From the harmonic oscillator ground state, we generate a Gaussian family of wavefunctions

$$\psi_{n,k}(x,y) = \mathrm{e}^{iky} H_n(x/\ell - \ell k)\mathrm{e}^{-\frac{(x-x_k)^2}{2\ell^2}} \tag{15.16}$$

which are extended in the y-direction but localized in x and centered at $x_k = \ell^2 k$. In Eq. (15.16), H_n is a Hermite polynomial. Each state indexed by n is known as a Landau level. The energy of each Landau level is

$$\epsilon_n = \hbar\omega_c\left(n + \frac{1}{2}\right) \tag{15.17}$$

and hence is independent of k. As a result, several iso-energetic states compose each Landau level. For a field of $B = 1$ T, the zero-point energy is on the order of 10^{-4} eV, or 1.34 K. We will see the effects of quantization in the discrete Landau levels if the temperature is lower than that determined by the zero-point energy. Our estimate of 1.34 K is a bit in error as we have not used the semiconductor effective mass. For GaAs, $m^* = 0.06m$; hence, the zero-point energy increases by a factor of 16 as does the temperature at which quantization effects in Landau levels are experimentally observable.

As a result of the degeneracy, each Landau level can hold many electrons. The degeneracy is determined by the distinct number of k values that generate a state within the same Landau level. We note that the states comprising each Landau level are centered at $x_k = \ell^2 k$, where

k can take on a range of values consistent with the confinement of the system in the y-direction. Let L and W be the spatial extents of the sample in the x- and y-directions, respectively. If we write the wavevector k as

$$k_m = \frac{2\pi m}{W}, \tag{15.18}$$

with m an integer, the maximum number of states allowable in each Landau level is obtained by solving the condition $L = \ell^2 k_{N_{\max}}$ or, equivalently,

$$N_{\max} = \frac{LW}{2\pi\ell^2} = \frac{eBLW}{hc}. \tag{15.19}$$

The right-hand side of this expression has a simple physical interpretation. The total magnetic flux in each Landau level is a product of the magnetic field and the area of the sample, BLW. This quantity must be equal to the number of electrons in each level times the flux quantum, hc/e. We see then that $N_{\max}$ is also the number of electrons in each Landau level. Consequently, we associate with each Landau level

$$n_B = \frac{1}{2\pi\ell^2} = \frac{eB}{hc} \tag{15.20}$$

as the number of states per unit area. Physically, $1/n_B$ is the irreducible area each state occupies in a Landau level. For $B = 1$ T, the irreducible area corresponds to a square with sides of about 0.6×10^{-9} m, roughly ten times the Bohr radius. Note the area $1/n_B$ is invariant from one Landau level to the next. As a result, the total number of filled Landau levels is given by

$$\nu = \frac{\rho}{n_B} \tag{15.21}$$

where ρ is the number of electrons per unit area. In the integer quantum Hall effect, ν is an integer.

It should now be clear that if the number of electrons in the system is an integral multiple of n_B, then the conductance is quantized. Under such conditions, the electron density $\rho = nn_B$, where n is an integer. As the reciprocal of the Hall resistance, the Hall conductance is given by $\sigma_{\mathrm{H}} = -ec\rho/B = ecnn_B/B = -ne^2/h$.

We can formulate a more penetrating argument for the quantization by appealing to the vanishing of the Lorentz force. If the system is translationally invariant, then the vanishing of the Lorentz force signifies that we can switch to a reference frame which moves at a velocity $\mathbf{v}$ relative to the laboratory frame such that $\mathbf{v} \times \mathbf{B} = -c\mathbf{E}$. In this reference frame, the velocity is given by $v_i = cE_j/B_k\epsilon_{ijk}$ where ϵ_{ijk} is the totally antisymmetric unit tensor defined by

$$\begin{aligned} \epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1, &\quad \text{even permutation,} \\ \epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1, &\quad \text{odd permutation.} \end{aligned} \tag{15.22}$$

The total current along the i-axis is given by Qev_i, where Q is the total charge in the system. If n Landau levels are occupied with $N_{\max}$ electrons in each, then $Q = nN_{\max}$. Hence, the current density is given by

$$j_i = \frac{eQv_i}{LW} = \frac{ecQE_j}{BLW}\epsilon_{ij} = \sigma_{ij}E_j, \tag{15.23}$$

where σ_{ij}, the coefficient of E_j, is the transverse current. This current is antisymmetric with respect to permutation of the indices x and y. Recall that the total magnetic flux $BLW = N_{\max}hc/e$. As a consequence,

$$\sigma_{xy} = \frac{ecQ}{BLW} = \frac{ecnN_{\max}}{N_{\max}hc/e} = \frac{ne^2}{h}. \tag{15.24}$$

Because the transverse current is antisymmetric, $\sigma_{xy} = -\sigma_{yx}$. In the moving reference frame, the diagonal conductance, $\sigma_{xx} = 0$ as a result of the vanishing of the longitudinal electric field. There is a fundamental physical reason for the vanishing of σ_{xx}, however. Because the Fermi level lies in the gap between the highest-occupied and the lowest-unoccupied Landau levels, σ_{xx} vanishes. Alternatively, the allowable phase space for scattering states vanishes when the Fermi level lies in a gap; hence $\rho_{xx} = 0$ as well. We see then that the conductance in the quantum Hall regime is a purely off-diagonal tensor with elements

$$\sigma = \begin{bmatrix} 0 & \frac{ne^2}{h} \\ -\frac{ne^2}{h} & 0 \end{bmatrix}. \tag{15.25}$$

Whether σ_{xy} or σ_{yx} are identified as the proper Hall conductance simply depends on the axis system used to orient the electric and magnetic fields.

15.3 The role of disorder

While the preceding argument is simple, it does not apply to dirty systems in which translational invariance is broken. Further, it cannot explain the origin of fractional values of the conductance. In fact, it is easy to see that, without disorder, we cannot account for the plateau nature of the quantum Hall effect. In a translationally invariant system, all the electronic states are extended. If an integral number of Landau levels are occupied, then the Fermi level lies in the gap between the highest-occupied and lowest-unoccupied Landau levels. As the magnetic field decreases, the Fermi level remains constant until the next Landau level is filled, at which point it jumps discontinuously. This would suggest that the Hall conductance should decrease monotonically as a function of magnetic field as in the classical case. From whence then do the plateaus come?

It turns out that disorder saves us. As we showed in Chapter 13, disorder changes both the spatial extent and the energy of electronic states. Hence, the degenerate band of states comprising each Landau level can be thought of as being broadened into a band of states that we describe approximately as having a Lorentzian lineshape centered at the unperturbed

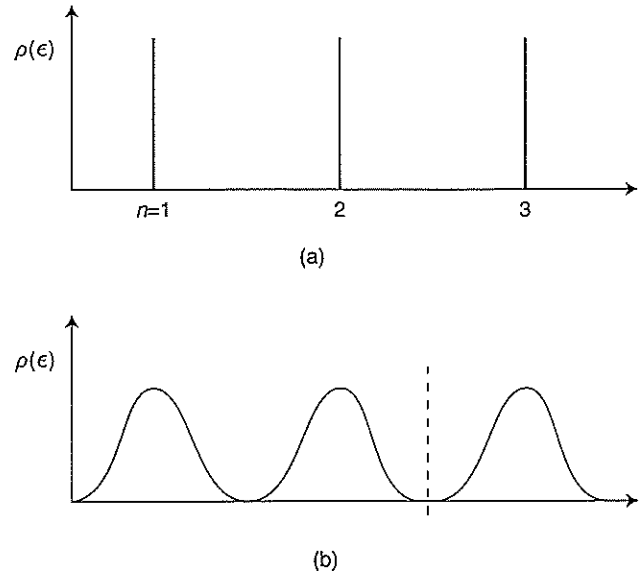

FIG 15.3 (a) Density of states of Landau levels in a magnetic field. (b) Broadening of the Landau levels as a result of disorder. The dashed line shows the position of the Fermi level.

energy of each Landau level. This is illustrated in Fig. 15.3. Intuitively, the further an electronic state moves away from the unperturbed energy of each Landau level, the more affected it is by the disorder and hence the more it has a tendency to be localized. This can be seen by treating the disorder perturbatively. Hence, we arrive at the simple picture that the states close to the center of the Landau level are less localized than those at the edge of the Lorentzian distribution. We showed in Chapter 13, however, that current-carrying states do not survive for even an infinitesimal amount of disorder in $d = 2$. If this state of affairs persists in the presence of a magnetic field, we arrive at the conclusion that the conductance should vanish in quantum Hall systems as well. However, a magnetic field is present. As we showed in Chapter 12, magnetic fields break time-reversal symmetry and hence disrupt the phase coherence needed to localize electronic states. In $d = 2$, field theoretic (P1984) as well as numerical studies (AA1981; P1981; T1983) show that the scaling theory of localization does in fact break down and current-carrying states obtain. As expected, they remain clustered at the unperturbed energy of each Landau level. All other states are localized. A sharp mobility edge demarcates the separation in energy between the extended and localized states.

From the simple picture that extended states form only at the center of each Landau level and all the other states are localized, we can explain the origin of the quantum Hall plateaus. Because the current is carried only by the states at the center of each Landau level, the current should jump discontinuously as the Fermi level is tuned through the center of each Landau level. Further, the current should remain constant if the occupation of the extended states remains unchanged. That is, although increasing the magnetic field causes the chemical potential to move away from the magical place where the extended states are

located, the conductance does not change because the chemical potential now resides in a region where the states are localized. The plateaus correspond to the range of magnetic fields for which the population in the extended states is fixed. The presence of precisely flat steps in the Hall voltage attests to the extreme localization of all states in a Landau level except for the narrow region of extended states located at the center. It is for this reason that the quantum Hall effect is fundamentally rooted in disorder. Paradoxically, disorder does not affect the value of the Hall conductance. Specifically, Aoki and Ando (AA1981) and Prange (P1981; P1987) showed to the lowest order in the drift velocity, $v_x = cE_x/B_z$, that although an isolated δ-function impurity binds an electron state, the extended states carry just enough extra current to compensate for the loss.

15.4 Currents at the edge

Thus far, we have argued that disorder localizes all electronic states except for those in a narrow window around the unperturbed energy of each Landau level. Once the chemical potential moves into this region, increasing the field further has no effect on the conductance because all other states are localized. Quantum Hall plateaus originate then from the separation in energy between extended and localized states. The only possible deviations from perfectly flat plateaus might originate from thermally activated transport from a localized state to an extended state at the center of a Landau level. At low temperatures, such processes contribute negligibly to the transport.

We have yet to explain, however, why the quantization of the conductance in integral multiples of e^2/h is so precise. As noted by Laughlin (L1980), the precise quantization of the conductance suggests that the quantum Hall effect must be due to a fundamental principle devoid of any material parameters, such as geometry. To this end, Laughlin formulated a gauge principle to explain the quantization of the Hall conductance.

To understand the essence of this argument, we first make a general observation regarding the current in 2d systems in a magnetic field. As stated earlier, electrons in a magnetic field move in circular orbits as a result of the Lorentz force. As illustrated in Fig. 15.4, in the bulk of a sample, clockwise and counter-clockwise pieces of neighboring cyclotron orbits overlap, leading to a vanishing of the current in the bulk. The situation is quite different at the edges of the sample, however. At the edge, the orbits are truncated in response to the confining potential created by the boundary. Once an electron is reflected by the boundary, it still attempts to move in a circular orbit. This induces a skipping-type motion of an electron at the boundary of the sample, as shown in Fig. 15.4. Such motion generates an edge current that flows in the clockwise direction for a magnetic field oriented along the positive z-direction. The chirality of the edge current is determined then by the direction of the magnetic field. While the above argument is valid strictly when the magnetic length much exceeds the wavelength of the electron, $\ell \gg \hbar/p_F$, the chirality of the edge current can be established quite generally from the presence of an induction field at the boundary (W1990). That the current arises from states at the edge is a truly novel feature of quantum Hall systems. The chirality of the current at the edge is also at the heart of why the edge states remain extended. As we learned in Chapter 12, backscattering is essential

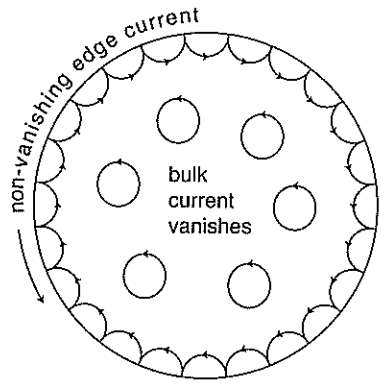

Fig. 15.4 Cyclotron orbits in a quantum Hall system. The circular orbits are caused by the Lorentz force. In the bulk of the sample, clockwise and counter-clockwise pieces of cyclotron orbits overlap and cancel, leading to a vanishing of the current in the bulk. At the edges, the orbits are truncated and give rise to an edge current.

for localization to obtain. There can be no backscattering for a chiral edge state. Hence, they resist localization by a random potential.

Since the current is carried entirely by the edges, the geometry of our system cannot matter. This simple realization already implies that the current in the quantum Hall effect is a topological invariant. We consider then a quantum Hall disk with a hole punched into the center, as depicted in Fig. 15.5. This geometry is equivalent to the one used by Halperin (H1982) in his reformulation of the original Laughlin argument. The disk is pierced with a uniform magnetic field in the positive z-direction. Truncation of the cyclotron orbits at the outer rim of the disk leads to an edge current that flows in the clockwise direction. However, at the inner radius, confinement leads to a current in the opposite direction. Clearly then, if the outer and inner edges of our annulus are at the same chemical potential, no net current will flow in the system. Let's assume that the Fermi levels of the inner and outer edges differ by an amount eE_0. This difference might be due to asymmetries in the confinement potentials at the inner and outer edges as well as to any electro-chemical potentials that might be present. If n Landau levels are occupied, then the total potential drop is neE_0. The identical argument leading to Eq. (15.23) can now be invoked and we obtain immediately that the net current between the inner and outer edges is quantized in units of e^2/h.

However, with a little more effort, we can extract the same result from a different way. We consider now the second geometry shown in Fig. 15.5(b). In formulating the gauge argument, we will find it easier to work with this geometry as the Landau gauge used previously is directly applicable. Our coordinate system is chosen so that the y-coordinate runs around the ribbon. As before, the current is carried by the edge states only. These states encircle the ribbon preserving phase coherence as they return to the origin. For the wavefunctions characterizing the edge states to be single-valued, the flux enclosed upon one trek around the disk must be an integral multiple of 2π. Hence, whatever change we make in the vector potential should satisfy the condition that

$$A = \frac{nhc}{eL}, \tag{15.26}$$

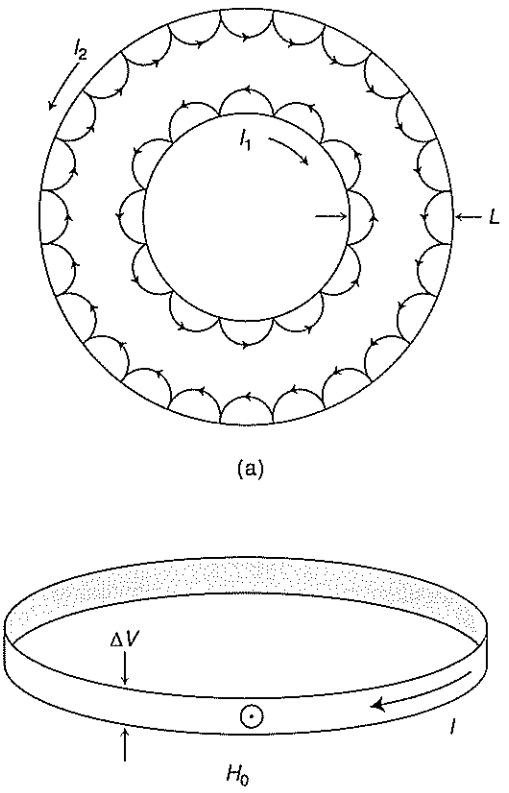

Fig 15.5 (a) A quantum Hall disk pierced by a magnetic field pointing out of the page. The arrows indicate the direction of the edge currents. (b) A quantum Hall ribbon in which the magnetic field is everywhere perpendicular to the surface. This geometry is mathematically equivalent to a rectangle with periodic boundary conditions in one direction. The circumference of the ribbon is L and its width is W. The Hall voltage is $\Delta V = E_0 W$.

where L is the circumference of the ribbon. For small changes in the vector potential, the current in our system is gauge invariant. Consider now the energy

$$\epsilon_\alpha = \langle \Psi_\alpha | H | \Psi_\alpha \rangle \tag{15.27}$$

of a particular single-particle state, Ψ_α, where H is given by the left-hand side of Eq. (15.11). For short-hand notation, we have defined $\alpha = (n, k)$. We are interested in the derivative of the ϵ_α with respect to A. To simplify this derivative, we use the Hellman–Feynman (H1937) theorem

$$\frac{\partial E(\lambda)}{\partial \lambda} = \langle \Psi_\alpha | \frac{\partial H(\lambda)}{\partial \lambda} | \Psi_\alpha \rangle, \tag{15.28}$$

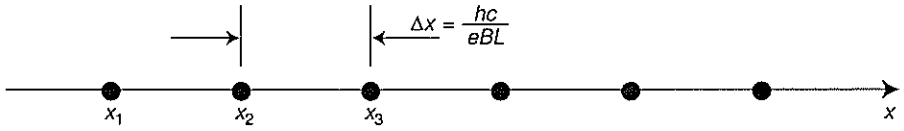

Fig. 15.6 The location of the centers for the electron states that comprise each Landau level. All states are assumed to be degenerate, with their centers given by $\ell^2 k_m = 2\pi \hbar c m/eBL$. The difference between two centers is hc/eBL. Threading the sample with one flux quantum transforms the mth state into the $(m-1)$st. This results in the transfer of charge from one edge of the sample to the other.

where λ is simply some variable in the Hamiltonian. Varying the single-particle energy with respect to A,

$$\frac{\partial \epsilon_\alpha}{\partial A} = \frac{-e}{mc}\langle \Psi_\alpha | \mathbf{p} - \frac{eA}{c} | \Psi_\alpha \rangle = L\frac{I_\alpha}{c}, \tag{15.29}$$

defines the current density carried by the state α as it traverses the disk. To evaluate the derivative, we note that in the presence of an electric field, the single-particle energies scale linearly with $eE_0 x_k$, where x_k locates the center of the states comprising each Landau level. If we modify the gauge term in the Hamiltonian such that $\mathbf{A} = Bx\hat{y} \to Bx\hat{y} + \Delta A\hat{y}$, then the location of each center is shifted by $x_k \to x_k - \Delta A/B$. Consequently, the single-particle energies are translated to $\epsilon_{n,k} \to \epsilon_{n,k} - eE_0 \Delta A/B$, and

$$\frac{\partial \epsilon_\alpha}{\partial A} = -\frac{eE_0}{B} \tag{15.30}$$

is independent of the state index. As illustrated in Fig. 15.6, Laughlin's gauge principle (L1980) follows from the fact that the difference between the location of the centers of two neighboring states in the same Landau level,

$$\Delta x_k = x_k^{m+1} - x_k^m = \ell^2 (k_{m+1} - k_m) = \frac{2\pi \ell^2}{L} = \frac{hc}{eBL} = \frac{\Delta A}{B}, \tag{15.31}$$

is directly related to the change in the vector potential. Consequently, if A is changed by a single flux quantum, the location of the mth center,

$$x_k^m \to x_k^m - \frac{\Delta A}{B} = x_k^m - \frac{hc}{eBL} = x_k^{m-1}, \tag{15.32}$$

is now coincident with the location of the $(m-1)$st center. Hence, when one flux quantum is threaded through the ribbon, the states in a Landau level all shift over by one, leading

to the net transfer of a single electron (per Landau level) from one edge of the sample to the other. Quantization of the gauge leads to quantization in the charge transfer! This is the Laughlin gauge principle. It illustrates beautifully the topological (T1982) nature of charge transport in the quantum Hall effect. To calculate the current, we substitute Eq. (15.30) into Eq. (15.29) and sum over all occupied Landau levels. We obtain immediately that the net current between the edges,

$$\begin{aligned} I &= \sum_{n,k} I_{n,k} \\ &= -\frac{ecN_e E_0}{LB} = -\frac{ecnn_B V_H}{B} = -\frac{ne^2 V_H}{h}, \end{aligned} \tag{15.33}$$

is an integer multiple of e^2/h with $V_H = E_0 W$, the Hall voltage and $N_e = nn_B LW$. The negative sign in the current corresponds to counter-clockwise motion on the ribbon. The Laughlin argument lays plain that the quantization of the Hall current or Hall conductance arises primarily from the restriction that the extended states must be single-valued as they traverse the edge of the sample. For a system in a magnetic field, the single-valuedness of the eigenstates manifests itself as a condition on allowable gauge transformations. It is this condition coupled with the integer filling of Landau levels that leads to the quantization. It is also paramount that the gauge transformation be carried adiabatically, so that the system remains in its ground state as the flux penetrating the system is changed.

As the plateau transitions are driven by changing the magnetic field or the filling, they constitute a genuine quantum phase transition of the kind studied in the previous chapter. In fact, because disorder is central to the story of the integer quantum Hall effect, the underlying plateau transitions represent one of the clearest examples of quantum critical phenomena in a disordered system. Progress in understanding the underlying field-theoretical description of the plateau transitions is based largely on the Chalker–Coddington tunneling network model (CC1988; MT1999; Z1999). Numerical simulations of this model (LWK1993; LC1994) have yielded a correlation length exponent of $\nu = 2.3$ which is consistent with experimental observations. However, this exponent is yet to be predicted by a rigorous theoretical account.

15.5 Topological insulators

In the quantum Hall effect, the current is carried by the edge. That it must be an integer multiple of e^2/h is determined entirely by the fact that when the electromagnetic gauge is changed by a single flux quantum, a single charge is transported across the sample. Since the edge states must be single-valued and their single-valuedness places a constraint on the possible changes in the electromagnetic gauge, the current in the quantum Hall system is robust. The topology of the sample is the only determining factor since σ_{xy} is entirely an edge effect. While the edge of the sample conducts, the bulk is entirely insulating. Hence, quantum Hall samples are topological insulators. Because time-reversal symmetry is broken, the backward- and forward-moving edge states are spatially separated. Hence, if

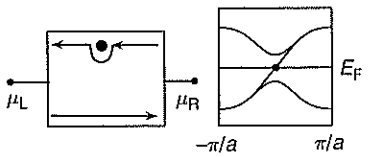

Fig. 15.7 (a) A quantum Hall bar with chiral edge states. (b) Corresponding schematic band structure for a quantum Hall system with a single edge state crossing the band gap. In the absence of time-reversal symmetry breaking, the edge states of a particular chirality cross the Fermi level once.

we plot the band structure of a quantum Hall system near an edge, it would look something like Fig. 15.7. Shown here is the band gap between two Landau levels and a single edge state traversing the gap. The key observation in quantum Hall systems is that $2 = 1 + 1$. That is, there are two edge states but only one at each edge as a result of the breaking of time-reversal symmetry. As pointed out previously, elastic scattering arising from impurities at each edge will not result in backscattering since the states at each edge propagate in only a single direction. There is a possibility that elastic scattering from one edge could give rise to a counter-propagating state. But this would require an electron jumping across the sample to the other edge. The probability for such an event is exponentially small, however. Hence, the only option for an electron encountering an elastic impurity is simply to go around it and continue moving in the same direction.

Consider now turning off the magnetic field. Is it still possible to have edge states that are impervious to backscattering? It turns out the answer is yes. It is this simple realization that spawned the field of topological insulators (KM2005a; KM2005b; TFK2008; HK2010; BHZ2006; R2009; MB2007; X2009; Z2009; QHZ2008; QZ2010; K2007; BZ2006). The key point here is that $4 = 2 + 2$ (QZ2010). That is, when time-reversal symmetry is preserved, there must be two propagating states at each edge. Time-reversal symmetry at each edge ensures that the forward and backward propagating states have opposite spin. Of course an electron can move in the opposite direction if it scatters to the opposite side of the sample. Here again, this is an exponentially small process. Hence, as long as the impurities are featureless in that the interactions arising from them cannot mix spin, there is no way to backscatter an electron. However, this argument, as formulated, fails if spin–orbit scattering is present. To our rescue comes a remarkable property of time-reversal for spin-1/2 particles. In quantum mechanics, the time-reversal operator for spin-1/2 particles is governed by the anti-unitary operator,

$$T = e^{i\frac{\pi}{2}\sigma_y}K, \tag{15.34}$$

where σ_y is the y-Pauli matrix defined in Chapter 7 and K performs complex conjugation. Using the fact that $\sigma_y^2 = 1$, one can expand the exponential in the time-reversal operator

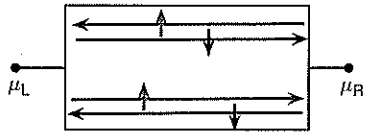

Fig. 15.8 A quantum spin Hall system with two propagating states per edge. Kramers' theorem implies that such states must be doubly degenerate. The up and down arrows indicate the spin.

and establish that for an arbitrary angle α,

$$e^{i\alpha/2\sigma_y} = i\alpha\sigma_y \sin\alpha/2 + \cos\alpha/2. \tag{15.35}$$

Consequently, rotating a spin-1/2 particle by 2π does not return the particle to its original state but to its negative and $T^2 = -1$. This is the origin of the Kramers doubling. Namely, all eigenstates of a time-reversal invariant (TRI) Hamiltonian must be doubly degenerate. Hence, the edge states have to come in pairs, one for each spin, as depicted in Fig. 15.8.

That $T^2 = -1$ protects the surface states from opening a gap in the presence of spin–orbit scattering can be illustrated in one of two ways. Consider the consequences of elastic scattering. In a semiclassical picture, a forward-moving electron encountering an impurity which can have spin–orbit scattering can be reflected backwards. However, since the only channel for backward motion at the edge in question has the opposite spin, the electron must flip its spin. In this process, the spin picks up a phase of π. Running this process backward in time, TRI implies that there is an equivalent process in which a forward-moving electron is scattered backward, but picks up a phase change of $-\pi$. (The two processes can be thought of as anticlockwise and clockwise rotations of spin.) The magnitudes of these amplitudes are identical as long as TRI is present. However, the phase change between these two paths is 2π and hence the wavefunction changes by -1. Both paths add to yield a vanishing scattering amplitude from an impurity; that is, there is perfect destructive interference. Consequently, as long as TRI is present, there is no way to backscatter off an impurity, and the surface states must cross the gap, thereby giving rise to a net longitudinal current. Alternatively, we can establish the robustness of the surface states from a band argument. TRI implies that the surface states must come in pairs with equal and opposite momenta. Which momentum is assigned spin up or spin down is irrelevant. Further, the edge states must have the same energy at the TRI points in the Brillouin zone which for one-dimensional edge states are $k = 0$ and $k = \pi/a$. Note, the $k = -\pi/a$ point is identical to π/a by TRI. Away from the TRI points, the spin–orbit interaction can lift the degeneracy between the edge states. The question is how is the degeneracy lifted. If the pair of edge states joins the degenerate points as in Fig. 15.9(a), then the surface states are not robust. Figure 15.9(a) is topologically identical to Fig. 15.9(b) as the chemical potential can be moved up and down in the gap. Note, it is irrelevant that the edges states cross the Fermi energy twice. As long as they cross an even number of times, the analog of Fig. 15.9(b) can always be constructed. Consequently, the edge states acquire a mass. However, this is

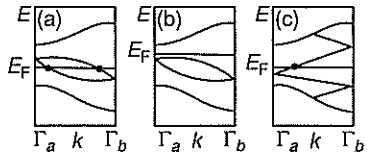

Fig. 15.9 Electronic dispersion between two Kramers degenerate (or time-reversal-invariance) points Γ_a and Γ_b. For a 2d sample, the degenerate points are at $\Gamma_a = 0$ and $\Gamma_b = \pi/a$. (a) Evolution of the surface states between the two Kramers points when the number of Fermi crossing points is even. (b) A distortion of (a) which gaps the surface states. Such a deformation of (a) is always possible if the edge states cross the chemical potential an even number of times. Consequently (a) and (b) are topologically indistinguishable. In (c), there is only a single crossing point in which case any deformation of the sample maintains the integrity of the edge states.

not the only option. The degeneracy can be lifted such that there is only a single crossing away from the TRI points, as illustrated in Fig. 15.9(c). In this case, the sample cannot be deformed (cut, for example) in any way to get rid of the single crossing point. That is, such states are protected by topology. We can distinguish these two cases by defining the index (TFK2008; KM2005a; KM2005b)

$$N_K = m \bmod 2, \tag{15.36}$$

where N_K is the number of Kramers pairs of edge states that cross the Fermi energy. For clarity, by m mod 2 we simply mean $m + 2p$, where p is any integer. If N_K is even, then $m = 0$, whereas $m = 1$ corresponds to N_K odd. Since there are only two possible values for m, we can think of m as being a Z_2 invariant. Z_2 is the group with two elements, namely 1 and 0, and hence is the simplest non-trivial group.

There is something subtle going on here which can be laid plain by considering spinful free electrons in two dimensions. If time-reversal symmetry is present, the simplest Hamiltonian that can be written for electrons propagating on the edge of a $(2+1)$-dimensional system is

$$H = p\sigma^z. \tag{15.37}$$

We ask a simple question: how can such edge states be gapped? The simplest augmentation is to add a term of the form $m_y\sigma^y + m_z\sigma^z$. However, since the mass remains invariant under $t \to -t$, this kind of coupling breaks TRI and hence is not allowed. To open a gap, we need to consider more than a single pair of edge states. Consider two copies of the system

$$H^{(2)} = p\mathbf{I} \otimes \sigma^z = \begin{pmatrix} p\sigma^z & 0 \\ 0 & p\sigma^z \end{pmatrix}.$$