Topological aspects of the quantum Hall effect

1997 J. Phys.: Condens. Matter 9 2507

(http://iopscience.iop.org/0953-8984/9/12/003)

J. Phys.: Condens. Matter **9** (1997) 2507–2549. Printed in the UK PII: S0953-8984(97)64230-7

REVIEW ARTICLE

Topological aspects of the quantum Hall effect

Y Hatsugai†

Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Received 12 August 1996, in final form 30 December 1996

Abstract. The quantum Hall effect is a typical realization of topological effects in condensed matter physics. In this article, some of the topological aspects of the quantum Hall effect are reviewed. For the lattice fermions, the Hall conductance of the system is expressed in terms of two different topological invariants. One is the famous TKNN integer which is related to the bulk state. The other is the winding number of the edge state on the complex-energy surface which is generally a high-genus Riemann surface. We will describe them in detail.

Therefore we have two topological expressions for the Hall conductance. Actually these two expressions give the same integer, although they look quite different. This means that one can explain the quantum Hall effect by using either the edge states or the bulk states, that is, $\sigma_{xy}^{\mathrm{edge}} = \sigma_{xy}^{\mathrm{bulk}}$.

1. Introduction

In the past few decades, topology has played an important role in condensed matter physics [1]. In many cases, topological invariants have emerged as a consequence of the quantization of classical mechanics. Dirac's magnetic monopole [2, 3], the vortex in type II superconductors [4], and the Aharonov–Bohm effect [5] are old examples. In these examples, the quantizations can be discussed in a geometrical language. Once the observable is written as a discrete topological number, it is stable against a weak perturbation since small continuous changes cannot modify the physical result discretely.

Recently, the topological discussion of the quantum mechanical phase has again become a focus in several different fields. Since the discovery of Berry's phase in an adiabatic process [6], many studies have been carried out, and they are interpreted using the concept of the *geometrical phase* [1].

In the quantum Hall effect (QHE), the Hall conductance σ_{xy} is quantized with extremely high accuracy [7]. Therefore it is natural to look for a topological interpretation for the effect. In this article, we aim to review some of the topological aspects of the QHE.

The effect of the magnetic field on an electronic system is an old problem and several studies were carried out before the discovery of the QHE. We should mention the pioneering work by Zak where the topological aspect of the problem is discussed by using magnetic translation operators [8]. A magnetic field implies a non-local effect, since the physically important quantity is not the magnetic field but the vector potential that describes the magnetic field. This is also the origin of the Aharonov–Bohm effect, in which the existence of a non-zero vector potential gives a physical effect even if the magnetic field is zero [5]

† E-mail: hatsugai@coral.t.u-tokyo.ac.jp.

0953-8984/97/122507+43$19.50

Since the discovery of the QHE, important topological investigations have been performed. The famous gauge invariance argument of Laughlin [9] is fundamental to the phenomena, and its relationship with the edge states which are localized near the sample edges is stressed by Halperin [10]. One might think that the effect on the edges has only a secondary effect on the physics. However, this is not the case. It is essentially important in the QHE. One neglects the effect of the edges in a naive thermodynamic limit. Such a procedure cannot be justified here since the existence of the edges changes the topology of the system, and this gives completely different physics.

The problem acquires a richer structure when one considers the effect of the magnetic field on a lattice which can be a model for electrons with a periodic potential. This was discussed by Azbel [11] and later by Hofstadter [12], and Wannier [13]. The topological aspect of the QHE with a periodic potential was first discussed by Thouless, Kohmoto, Nightingale, and den Nijs [14]. In this famous work, the topological character of the system is described in terms of the bulk state, and a topological expression for the Hall conductance is given by the bulk state (the 'TKNN' integer). The Hall conductance is given by the Chern number over the magnetic Brillouin zone [14, 15]. This is a bulk theory. Recently, the topological aspect of the edge state on the lattice was also discussed by the present author, where the Hall conductance is given by using the edge states (the 'winding number') [16]. The work is motivated by the work [10] and a trial to find a topological meaning for the Hall conductance (in terms of edges). This is an edge theory. Thus we have two topological expressions for the Hall conductance. Here a natural question is that of whether there is a relationship between the two. This is clarified also. The two topological expressions in terms of the bulk states and the edge states, which look quite different, actually give the same integer [17, 18]. We try to explain this in this article.

Now, at the end of this introduction, we mention recent developments in algebraic aspects of the effects of the magnetic field [19–21]. In addition to the topological aspect, a mathematically new type of algebraic structure, the 'quantum group', is hidden in the problem. Generators of the quantum group have non-trivial commutation relations, related to the commutation relations of covariant translations. The quantum group has one parameter characterizing the algebraic structure which is given by $e^{i\phi/2}$, where ϕ is the flux per plaquette. In this sense, we may say that the geometrical phase of the system also characterizes the algebraic structure of the system.

The QHE can be divided into two cases: the integer case and the fractional case. The Hall conductance is quantized at integer values or fractional values, respectively. Although the discussion in this article is mainly of the integer case, we will briefly comment on the fractional case, using Jain's construction or adiabatic heuristic arguments (see references [22–24]). This argument is also topological, and all of the results for the integer QHE can be mapped onto the fractional QHE by using it [25].

In section 2, we discuss the symmetries of the system, focusing on the geometrical aspect of the electronic system in a magnetic field. In section 3, Laughlin's argument for the integer QHE is briefly reviewed. Also the importance of the edge state, first stressed by Halperin, is demonstrated. These can be understood as important topological aspects of the QHE. When one considers the system on the lattice, topological aspects of the QHE are clearly demonstrated in interesting ways. In section 4, the Hall conductance σ_{xy} of the bulk state on a lattice is reviewed and its topological meaning is discussed. Here σ_{xy} is given by the total vorticity defined as the phase of the Bloch state in the magnetic Brillouin zone. This is the famous topological invariant (the 'TKNN' integer). On the other hand, when one considers the effect of edges on the lattice, the Hall conductance of the edge state is also described by another topological invariant. This is defined on the complex-

energy surface of the problem which is a high-genus Riemann surface in general. On the surface, the edge states give zero points of the wavefunction and their winding number on the Riemann surface gives the Hall conductance. This topological description of the Hall conductance in terms of the edge states is given in section 5. Now the Hall conductance of the system is described by two topological invariants. One is related to the bulk states and the other is related to the edge states. Although the two topological invariants are apparently quite different, they are closely related, as will be discussed in section 6. In section 7, the adiabatic heuristic argument is briefly reviewed. We include several appendices to make the review self-contained.

2. Symmetry and the geometrical aspect of an electronic system with a magnetic field

Let us first discuss the symmetries of an electronic system in a magnetic field. The Hamiltonian of a system with a uniform magnetic field $\boldsymbol{B}$ is given by

$$H = \frac{1}{2m}(\Pi_x^2 + \Pi_y^2) \tag{1}$$

$$\Pi_\alpha = -\mathrm{i}\hbar\,\partial_\alpha + eA_\alpha \qquad (\alpha = x, y) \tag{2}$$

$$A = A_\alpha\ \mathrm{d}x_\alpha \tag{3}$$

$$\mathrm{d}A = B\ \mathrm{d}x \times \mathrm{d}y. \tag{4}$$

When the magnetic field is not zero, the momenta along the x- and y-directions, Π_x and Π_y, do not commute with each other:

$$[\Pi^x, \Pi^y] = -\mathrm{i}\frac{\hbar^2}{l^2} \tag{5}$$

$$l = \left(\frac{\hbar}{eB}\right)^{1/2}. \tag{6}$$

It is convenient to introduce a bosonic operator a ($[a, a^\dagger] = 1$) via

$$\begin{aligned}\Pi_x &= \frac{\hbar}{l}\frac{1}{\mathrm{i}\sqrt{2}}(a - a^\dagger)\\ \Pi_y &= \frac{\hbar}{l}\frac{1}{\sqrt{2}}(a + a^\dagger).\end{aligned} \tag{7}$$

Then the Hamiltonian (1) is expressed as the equation for a harmonic oscillator, given by

$$H = \hbar\omega_c\left(a^\dagger a + \frac{1}{2}\right) \tag{8}$$

$$\omega_c = \frac{eB}{m} \tag{9}$$

which describes the Landau level with Landau gap $\hbar\omega_c$. This two-dimensional Hamiltonian has other constants of the motion: *'guiding centre'* operators $R_{x,y}$, commuting with the Hamiltonian [26]:

$$[H, R_{x,y}] = 0 \tag{10}$$

$$R_x = x - \frac{l^2}{\hbar}\Pi_y \tag{11}$$

$$R_y = y + \frac{l^2}{\hbar}\Pi_x. \tag{12}$$

In general, each of the guiding centre operators $R_{x,y}$ operates non-trivially on the eigenstate of the Hamiltonian (1) which introduces the Landau degeneracy. Since the guiding centre operators R_x and R_x do not commute with each other, one can only diagonalize a linear combination of them. The convenient choice for the linear combination depends on an explicit gauge choice, and a suitable combination helps in a geometry-dependent discussion. When one discusses a cylindrical geometry with the Landau gauge, $A = B(0, x)$ (periodic in y and finite in x), it is useful to diagonalize R_x. On the other hand, when one uses a disc geometry with a symmetric gauge, $A = B(-y, x)/2$, it is useful to diagonalize $R^2 = R_x^2 + R_y^2$.

We focus on two important symmetries of the system. One is the local U(1) gauge symmetry given by

$$\begin{aligned} A' &= A + \mathrm{d}\chi \\ -\mathrm{i}\bar{h}\,\partial'_\alpha &= -\mathrm{i}\bar{h}\,\partial_\alpha - e\,\partial_\alpha\chi\,. \end{aligned} \tag{13}$$

The invariance of the Hamiltonian (1) is obvious.

Here let us define the covariant translation operators $T_{x,y}(a)$ [8]

$$T_\alpha(a) = \mathrm{e}^{\mathrm{i}a\Pi_\alpha} \tag{14}$$

which satisfy

$$T_x^{-1}(a)T_y^{-1}(b)T_y(b)T_x(a) = \mathrm{e}^{-\mathrm{i}ab/l^2}. \tag{15}$$

Hence, cyclic evolution of the covariant translation operator gives a non-trivial phase:

$$P\exp\left(\mathrm{i}\frac{1}{\bar{h}}\oint_C \mathrm{d}x_\alpha\ \Pi_\alpha\right) = \exp\left(-\mathrm{i}\frac{S_C}{l^2}\right) \tag{16}$$

where S_C is the area enclosed by the curve C and P is a path ordering [8]. When there is a non-zero magnetic field, a simple translation of a particle is non-trivial and causes a non-zero phase change of the wavefunction. This phase equation (16) is the geometrical phase which is the origin of the topological character of the problem.

The other important symmetry is given by the magnetic translation operators $\hat{T}_x$ and $\hat{T}_y$ defined as [8, 27]

$$\hat{T}_\alpha = \mathrm{e}^{\mathrm{i}a\hat{\Pi}_\alpha} \tag{17}$$

$$\hat{\Pi}_\alpha = \Pi_\alpha + (\boldsymbol{B}\times\boldsymbol{r})_\alpha \tag{18}$$

$$\hat{\Pi}_x = \Pi_x - By \tag{19}$$

$$\hat{\Pi}_y = \Pi_y + Bx. \tag{20}$$

Each of the $\hat{T}_x$ and $\hat{T}_y$ commutes with the Hamiltonian:

$$[H, \hat{T}_x] = [H, \hat{T}_y] = 0 \tag{21}$$

although $\hat{T}_x$ and $\hat{T}_y$ do not commute with each other: $[\hat{T}_x, \hat{T}_y] \neq 0$.

These two symmetries (U(1) and the magnetic translations) are the most important characteristics of the system. In the following, we try to reformulate the theory for the lattice while preserving the symmetries. We prefer a lattice theory since there is no ambiguity related to the cut-off. By taking an appropriate limit, one can always recover the continuum theory. Therefore, the lattice theory includes more information. Also, physically, the effect of the lattice can be understood as that of a periodic potential in a crystal. By using a recently developed experimental technique, these periodic structures can be realized in some realistic mesoscopic systems (for example, 'quantum dots' and 'super-atoms'). When one considers the lattice system, it has two fundamental areas. One is the area of the minimum lattice

plaquette and the other is the magnetic area l^2. The existence of these two quantities can bring out very rich structures when the ratio is not rational [28].

In a lattice system, electrons in a magnetic field can be described by the tight-binding Hamiltonian

$$H = T_x + T_y + T_x^\dagger + T_y^\dagger \tag{22}$$

where T_x and T_y are the *covariant translation* operators defined by

$$T_x = \sum_{m,n} c_{m+1,n}^\dagger c_{m,n} \mathrm{e}^{\mathrm{i}\theta_{m,n}^x} \tag{23}$$

$$T_y = \sum_{m,n} c_{m,n+1}^\dagger c_{m,n} \mathrm{e}^{\mathrm{i}\theta_{m,n}^y} \tag{24}$$

where $c_{m,n}$ is the annihilation operator for an electron at site (m, n). The phase factors $\theta_{m,n}^x$ and $\theta_{m,n}^y$ are related to the flux per plaquette $\phi_{m,n}$ at (m, n) by

$$\mathrm{rot}_{(m,n)}\theta = \Delta_x \theta_{m,n}^y - \Delta_y \theta_{m,n}^x = 2\pi\phi_{m,n} \tag{25}$$

where the difference operators Δ_x and Δ_y operate on a lattice function $f_{m,n}$ as follows:

$$\Delta_x f_{m,n} = f_{m+1,n} - f_{m,n} \tag{26}$$

$$\Delta_y f_{m,n} = f_{m,n+1} - f_{m,n}. \tag{27}$$

This lattice Hamiltonian H also has a local U(1) gauge symmetry, i.e., it is invariant under

$$c_i \to \Omega_i c_i \tag{28}$$

$$\mathrm{e}^{\mathrm{i}\theta_{ij}} \to \Omega_i \mathrm{e}^{\mathrm{i}\theta_{ij}} \Omega_j^{-1} \tag{29}$$

$$|\Omega_j| = 1 \qquad \forall j = (m, n). \tag{30}$$

The form of the Hamiltonian above is sometimes interpreted as an approximation, the so-called Peierls substitution. One might consequently be concerned about the accuracy of the approximation. However, since the model above has the same symmetry as the continuum model, the present lattice Hamiltonian has a fundamental importance beyond that of a simple approximation. In figure 1, the energy spectrum of the lattice system is shown.

This is the famous Hofstadter butterfly diagram which includes a rich structure [12]. In the weak-field limit $\phi \to 0$ and in the low-density limit, the spectrum appears to be composed of many straight lines. Actually it gives the Landau level in the continuous limit. (See appendix A.)

Next let us consider the symmetry which is related to the translations of the lattice system. (See reference [29].) The covariant translations T_x and T_y do not commute with each other. For a one-particle state $|\Psi_{m,n}\rangle = c_{m,n}^\dagger|0\rangle$ which is localized at the site (m, n), their operation is as follows:

$$T_y T_x |\Psi_{m,n}\rangle = \mathrm{e}^{\mathrm{i}2\pi\phi_{m,n}} T_x T_y |\Psi_{m,n}\rangle. \tag{31}$$

In the following, we choose a special gauge. In this case, there is an important symmetry of the Hamiltonian, described by the magnetic translation operators $\hat{T}_x$ and $\hat{T}_y$ which commute with the Hamiltonian:

$$[H, \hat{T}_x] = [H, \hat{T}_y] = 0. \tag{32}$$

They are explicitly given by

$$\hat{T}_x = \sum_{m,n} c_{m+1,n}^\dagger c_{m,n} \mathrm{e}^{\mathrm{i}\chi_{m,n}^x} \tag{33}$$

$$\hat{T}_y = \sum_{m,n} c_{m,n+1}^\dagger c_{m,n} \mathrm{e}^{\mathrm{i}\chi_{m,n}^y} \tag{34}$$

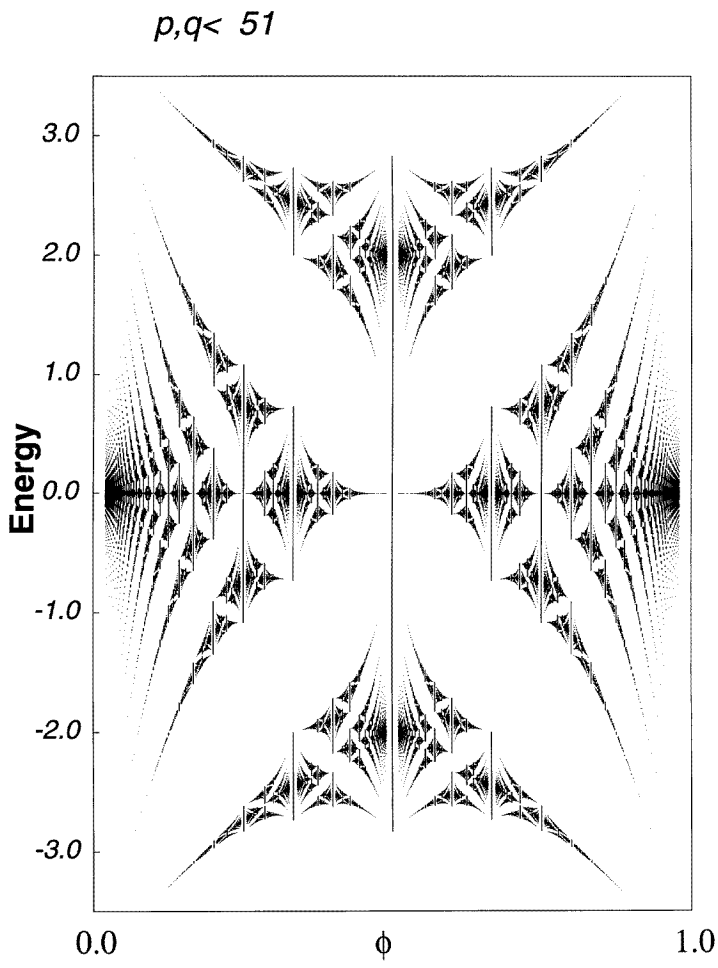


Figure 1. The energy spectrum of the lattice Hamiltonian with a uniform magnetic field as a function of the flux per plaquette.

where the phases $\chi^{x,y}_{m,n}$ satisfy the relations

$$\Delta_x \chi^x_{m,n} = \Delta_x \theta^x_{m,n} \tag{35}$$

$$\Delta_y \chi^x_{m,n} = \Delta_x \theta^y_{m,n} \ (= \Delta_y \theta^x_{m,n} + 2\pi\phi_{m,n}) \tag{36}$$

$$\Delta_x \chi^y_{m,n} = \Delta_y \theta^x_{m,n} \ (= \Delta_x \theta^y_{m,n} - 2\pi\phi_{m,n}) \tag{37}$$

$$\Delta_y \chi^y_{m,n} = \Delta_y \theta^y_{m,n}. \tag{38}$$

In the parentheses, we have used equation (26) and equation (27). We can easily solve them to give

$$\chi^x_{m,n} = \theta^x_{m,n} + 2\pi n\phi_{m,n} \tag{39}$$

$$\chi^y_{m,n} = \theta^y_{m,n} - 2\pi m\phi_{m,n} \tag{40}$$

which are gauge dependent.

In the following, let us assume that the magnetic field is uniform and rational, that is,

$$\phi_{m,n} = \phi = p/q \tag{41}$$

with mutually prime integers p and q. We select a Landau gauge:

$$\theta^x_{m,n} = 0 \qquad \theta^y_{m,n} = 2\pi\phi m. \tag{42}$$

These do not commute with each other. However, one can explicitly check that $\hat{T}_x^q$ and $\hat{T}_y$ commute with each other. Thus we can have simultaneous eigenstates $|k_x, k_y\rangle$ of H, $\hat{T}_x^q$ and $\hat{T}_y$, which are specified by the momentum in the magnetic Brillouin zone ($0 \leqslant k_x \leqslant 2\pi/q$, mod $2\pi/q$, $0 \leqslant k_y \leqslant 2\pi$, mod 2π) (the Bloch theorem), as follows:

$$H|k_x, k_y\rangle = E(k_x, k_y)|k_x, k_y\rangle \tag{43}$$

$$\hat{T}_x^q|k_x, k_y\rangle = \mathrm{e}^{\mathrm{i}qk_x}|k_x, k_y\rangle \tag{44}$$

$$\hat{T}_y|k_x, k_y\rangle = \mathrm{e}^{\mathrm{i}k_y}|k_x, k_y\rangle. \tag{45}$$

3. Quantization of the Hall conductance using the gauge invariance: Laughlin's theory and Halperin's edge states

The Hall conductance has a fundamental topological meaning, as first pointed out by Laughlin in his gauge invariance argument [9]. The effect of edges is essential in the argument which was stressed by Halperin [10]. We review these results briefly.

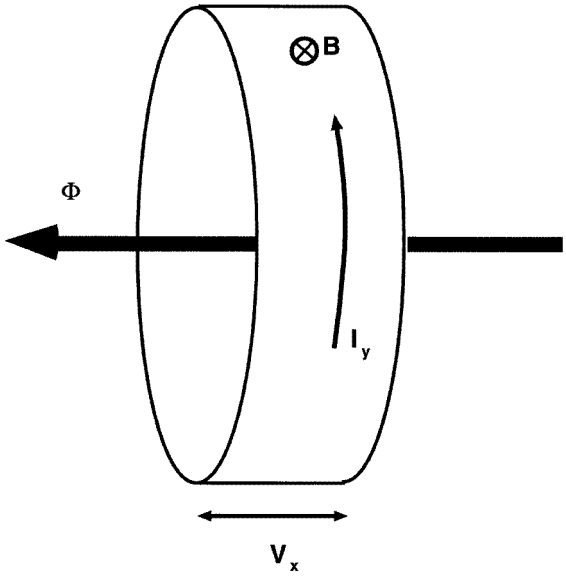


Figure 2. The cylindrical geometry for using the gauge invariance given by Laughlin. In addition to the magnetic field $\boldsymbol{B}$, there is an Aharonov–Bohm flux Φ through the hole.

In a cylindrical geometry (figure 2), the Hall current along the cylindrical direction I_y is given by

$$I_y = \frac{\partial E}{\partial \Phi} \tag{46}$$

where E is the total energy of the system and Φ is the Aharonov–Bohm (AB) flux through the hole of the cylinder [4]. This is known as a Byers–Yang formula. Here let us discretize

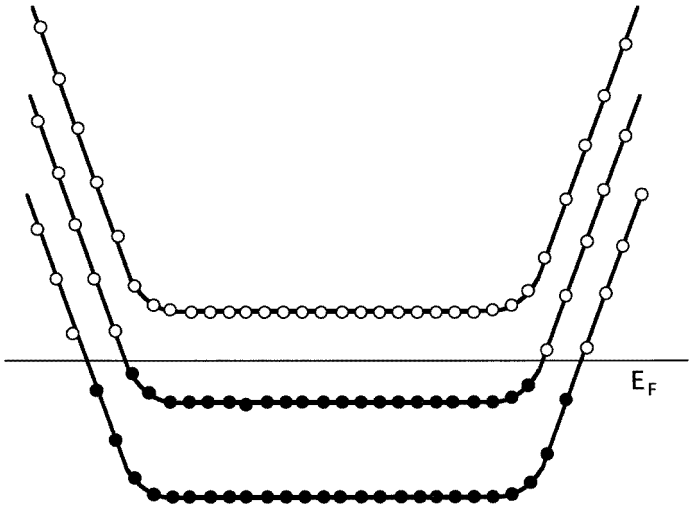


Figure 3. A schematic energy diagram for the Landau levels and edge states in a continuous theory.

this formula as

$$I_y = \frac{\Delta E}{\Delta \Phi} \tag{47}$$

and take the difference in Φ, $\Delta\Phi$, as the flux quantum $\Phi_0 = h/e$, where we set $c = 1$ (we use this convention throughout this article). In a physical situation, let us consider an *adiabatic* process to change the AB flux by just one flux quantum Φ_0 very slowly, to keep the state in the ground state of the snapshot Hamiltonian. Here we assume that the Fermi energy is in the energy gap or in the energy region of the localized states. When the change is adiabatically slow (the energy gap is sufficiently large), the systems before and after the process are gauge equivalent via the so-called 'large' gauge transformation (see appendix B). Therefore the electronic spectrum is unchanged. A possible effect of this process is that some electrons are carried from the left-hand (right-hand) edge to the right-hand (left-hand) edge. Such a number of electrons is, of course, an integer (we cannot specify the integral value, however) and we denote it as n. The change of the AB flux can be understood as the electric field along the transverse direction x, and we denote the voltage across the cylinder by V_x. Then the energy cost of the process is given by $\Delta E = neV_x$ ($\Delta\Phi = \Phi_0$). Therefore using equation (46), we have for the Hall conductance $\sigma_{xy} = -\sigma_{yx} = I_y/V_x$:

$$\sigma_{xy} = -n\frac{e^2}{h}. \tag{48}$$

By this beautiful argument, the integral quantization of the Hall conductance is demonstrated explicitly. We only use the gauge symmetry which is preserved even in the system with randomness. A necessary condition of the argument is the validity of the adiabatic assumption used above, which is justified when there is an energy gap between the ground state and the excited state. Usually in the integer QHE, this energy gap can be identified as the Landau gap. Therefore this argument is stable unless the randomness is strong enough to collapse the Landau gap.

This is beautiful and essential in the QHE. However, one cannot obtain the quantized value n from this gauge invariance argument, above. In a continuous system, this value is

fixed by Halperin as the integer n_{filled}, where n_{filled} is the number of bulk Landau levels which are below the Fermi energy [10]. This is explained as follows. When one takes the Landau gauge and diagonalizes the guiding centre along the x-direction, R_x, each one-particle state is specified by the momentum along the y-direction, k_y. In this gauge, $R_x = -l^2 k_y$ and the corresponding wavefunction is localized near $x = \langle R_x \rangle = -l^2 k_y$ in a stripe shape. For the bulk states (states with $\langle R_x \rangle$ away from the edges), different states with momentum k_y span a basis of the Landau level even when there is randomness. On the other hand, localized states near the edges feel an edge potential and generally have high energies compared with those of the bulk states. Therefore the Landau levels have finite slopes near the edges (see figure 3). If the cylinder has a finite length L_y along the y-direction, the momentum $k_y = 2\pi n_y / L_y$ is discretized and the AB flux also enters the problem in the form

$$k_y + 2\pi \frac{\Phi}{\Phi_0} \frac{1}{L_y} = 2\pi \frac{n_y + \Phi/\Phi_0}{L_y} \qquad n_y\text{: integer.} \tag{49}$$

Therefore, changing the AB flux by one flux quantum shifts the k_y to the neighbouring one, which corresponds to shifting the state spatially by $(2\pi/L_y)l^2$ in this gauge. (See the later discussion for the lattice system.) From figure 3, one can confirm that the number of states, which are carried form the left to the right during Laughlin's adiabatic process, is the number which are filled Landau levels, n_{filled}. Therefore,

$$\sigma_{xy} = n_{\text{filled}} \frac{e^2}{h}. \tag{50}$$

This is the explanation of the QHE, in terms of edge states in the continuous system.

This gauge-invariant, edge-dependent argument is one of the topological aspects of the QHE.

4. The topological meaning of the σ_{xy} in terms of bulk states: Chern numbers on the Brillouin zone

The importance of gauge invariance which was revealed by Laughlin [9] and Halperin [10] is one of the topological effects in the QHE. When one considers the system on a lattice, another topological meaning of the Hall conductance becomes clear which originates from the non-commutativity of the magnetic translation operators. Therefore we may say that the geometrical phase of the system gives another topological aspect of the QHE. This topological character of the Hall conductance was first established by Thouless, Kohmoto, Nightingale and den Nijs (TKNN) [14, 15, 30–33] for the bulk state [14], and later its meaning in relation to the edge states was discovered by the present author [16–18]. In this section, we will review the TKNN theory and discuss the topological meaning of the Hall conductance using the bulk state language.

Let us focus on the dc Hall conductance σ_{xy}, since we can clearly demonstrate the topological origin via the adiabatic approximation [30]. (Also we can derive a general expression for the conductance via the Kubo formula [34, 35]. It is, however, not topological in general as shown later.)

We impose a periodic boundary condition on an $L_x \times L_y$ lattice. The effect of an external electric field applied along the x-direction is included in the problem via $E_x = -\partial A_x / \partial t$. The AB flux, Φ, along the y-direction is also considered. Then, the Hamiltonian is time dependent and is given by

$$H(t, \Phi) = T_x \exp\left(-\mathrm{i}\, 2\pi \frac{E_x}{\Phi_0} t\right) + T_y \exp\left(\mathrm{i}\, 2\pi \frac{\Phi}{\Phi_0} \frac{1}{L_y}\right) + \text{HC} \tag{51}$$

where $\Phi_0 = h/e$ is the flux quantum.

The current operator along the y-direction is given by

$$I_y = \frac{\partial H}{\partial \Phi} = \mathrm{i}\frac{e}{\hbar}\frac{1}{L_y}T_y + \mathrm{HC}. \tag{52}$$

In the Landau gauge, $\theta^x_{mn} = 0, \theta^y_{mn} = 2\pi\phi m$, the Hamiltonian and the current operators are written in the momentum representation as

$$H(t, \Phi) = L_x L_y \int_{-\pi}^{\pi} \frac{\mathrm{d}k_x}{2\pi} \int_{-\pi}^{\pi} \frac{\mathrm{d}k_y}{2\pi}\ H(\boldsymbol{k}, t, \Phi) \tag{53}$$

$$\begin{aligned} H(\boldsymbol{k}, t, \Phi) = {} & \exp\left(\mathrm{i}\left(k_x + 2\pi\frac{E_x}{\Phi_0}t\right)\right)c^\dagger(\boldsymbol{k})c(\boldsymbol{k}) \\ & + \exp\left(-\mathrm{i}\left(k_y - 2\pi\frac{\Phi}{\Phi_0}\frac{1}{L_y}\right)\right)c^\dagger(\boldsymbol{k}+\Delta\boldsymbol{k})c(\boldsymbol{k}) + \mathrm{HC} \end{aligned} \tag{54}$$

where $\Delta\boldsymbol{k} = (2\pi\phi, 0)$. Here we have used the fact that there are two translational symmetries, in both the x- and y-directions. (This is not the case when there is an edge.) Since

$$\frac{\partial}{\partial\Phi}H(\boldsymbol{k}, t, \Phi) = -\frac{2\pi}{\Phi_0 L_y}\,\partial_{k_y}H(\boldsymbol{k}, t, \Phi) = -\frac{1}{L_y}\frac{e}{\hbar}\,\partial_{k_y}H(\boldsymbol{k}, t, \Phi) \tag{55}$$

we have another expression for the current operator:

$$I_y = -L_x\frac{e}{\hbar}\int_{-\pi}^{\pi}\frac{\mathrm{d}k_x}{2\pi}\int_{-\pi}^{\pi}\frac{\mathrm{d}k_y}{2\pi}\ \partial_{k_y}H(\boldsymbol{k}, t, \Phi). \tag{56}$$

To obtain the ground state, $|G(t)\rangle$, of the time-dependent Hamiltonian, $H(t, \Phi)$, using the adiabatic approximation, let us expand $|G(t)\rangle$ in a complete set of the eigenstates $|\alpha(t)\rangle$ of the snapshot Hamiltonian

$$H(t, \Phi)|\alpha(t)\rangle = E_\alpha(t)|\alpha(t)\rangle \tag{57}$$

as

$$|G(t)\rangle = \exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\ E_g(t')\right)\sum_\alpha a_\alpha(t)|\alpha(t)\rangle. \tag{58}$$

When $E_g(t)$ is energetically well separated from the rest of the spectrum of $H(t, \Phi)$, we can approximately integrate the Schrödinger equation to obtain the coefficients $a_\alpha(t)$. The result is written as

$$|G(t)\rangle = \mathrm{e}^{\mathrm{i}\gamma_{\mathrm{Berry}}}\exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\ E_g(t')\right)\left(|g(t)\rangle + \mathrm{i}\hbar\sum_{\alpha\neq g}\frac{|\alpha(t)\rangle\langle\alpha(t)|\partial_t g(t)\rangle}{E_\alpha(t) - E_g(t)}\right) \tag{59}$$

$$\mathrm{i}\gamma_{\mathrm{Berry}} = \int \mathrm{d}t\ \langle\alpha|\dot{\alpha}\rangle \tag{60}$$

where $g(t)$ is the lowest eigenstate of the snapshot Hamiltonian and γ_{Berry} is the Berry phase associated with the present adiabatic process [23]. (γ_{Berry} is not important in the following argument.)

To justify the adiabatic approximation, the existence of an energy gap is crucial. The ground state has to be separated for the rest of the eigenstates by a finite energy gap. Alternatively, the matrix element could be negligible, which is the case when the Fermi energy is in the energy region of the localized states. We assume that the flux per plaquette ϕ is rational, $\phi = p/q$, with coprime integers p and q. Then the energy spectrum consists of q energy bands. Each energy band corresponds to the Landau level in the continuous limit.

(See appendix A.) Then equation (57) is reduced to a $q \times q$ matrix eigenvalue equation with parameters $\boldsymbol{k} = (k_x, k_y)$:

$$\sum_{j=1}^{q} h_{i,j}(\boldsymbol{k}, t)\psi_j^l(\boldsymbol{k}, t) = \epsilon^l(\boldsymbol{k}, t)\psi_i^l(\boldsymbol{k}, t) \tag{61}$$

whose lth eigenvalue gives the energy of the lth band $\epsilon^l(\boldsymbol{k}, t)$. Now the parameters k_x and k_y run in $k_x \in [0, 2\pi/q]$ and $k_y \in [0, 2\pi]$, which defines the magnetic Brillouin zone T^2_{MBZ}. We assume that the Fermi energy is in the jth energy gap. Then the ground state $|g(t)\rangle$ is explicitly written as

$$|g(t)\rangle = \prod_{\boldsymbol{k}} \prod_{l=1}^{j} c^\dagger(\boldsymbol{k}, l)|0\rangle \tag{62}$$

$$c^\dagger(\boldsymbol{k}, l) = \sum_{m=1}^{q} \psi_j^l c^\dagger(k_x + 2\pi m/q, k_y) \tag{63}$$

where ψ_j^l is a coefficient of the eigenvector in equation (61).

The current along the y-direction in the ground state is evaluated within the adiabatic approximation as

$$\langle I_y\rangle = \langle G|I_y|G\rangle - \langle g|I_y|g\rangle \tag{64}$$

$$= L_x \mathrm{i}e \int_{T^2_{\mathrm{MBZ}}} \frac{\mathrm{d}^2k}{(2\pi)^2} \sum_{\alpha\neq g} \frac{\langle g|\partial_{k_y}H|\alpha\rangle\langle\alpha|\partial_t g\rangle - \langle\partial_t g|\alpha\rangle\langle\alpha|\partial_{k_y}H|g\rangle}{E_\alpha - E_g} \tag{65}$$

$$= \frac{E_x L_x \mathrm{i}e}{2\pi\Phi_0} \int_{T^2_{\mathrm{MBZ}}} \mathrm{d}^2k \sum_{\alpha\neq g} \frac{\langle g|\partial_{k_y}H|\alpha\rangle\langle\alpha|\partial_{k_x} g\rangle - \langle\partial_{k_x} g|\alpha\rangle\langle\alpha|\partial_{k_y}H|g\rangle}{E_\alpha - E_g} \tag{66}$$

$$= V_x \frac{e^2}{h} \frac{1}{2\pi\mathrm{i}} \int_{T^2_{\mathrm{MBZ}}} \mathrm{d}^2k \sum_{\alpha\neq g} \langle g|\partial_{k_y}\alpha\rangle\langle\alpha|\partial_{k_x} g\rangle + \langle\partial_{k_x} g|\alpha\rangle\langle\alpha|\partial_{k_y} g\rangle \tag{67}$$

$$= V_x \frac{e^2}{h} \frac{1}{2\pi\mathrm{i}} \int_{T^2_{\mathrm{MBZ}}} \mathrm{d}^2k \ (\langle\partial_{k_x} g|\partial_{k_y} g\rangle - \langle\partial_{k_y} g|\partial_{k_x} g\rangle) \tag{68}$$

where we have used the relations $V_x = E_x L_x$, $\partial_t H = -2\pi(E_x/\Phi_0)\,\partial_{k_x} H$ and $\langle\alpha|\partial_{k_y}H|\beta\rangle = (E_\beta - E_\alpha)\langle\alpha|\partial_{k_y}\beta\rangle$. Since the Hall conductance satisfies $\sigma_{xy} = -\sigma_{yx}$, we obtain

$$\sigma_{xy}^{j,\mathrm{bulk}} = -\frac{e^2}{h} \frac{1}{2\pi\mathrm{i}} \int_{T^2_{\mathrm{MBZ}}} \mathrm{d}k_i \times \mathrm{d}k_j \ \langle\partial_{k_i} g|\partial_{k_j} g\rangle \tag{69}$$

$$= -\frac{e^2}{h} \frac{1}{2\pi\mathrm{i}} \int_{T^2_{\mathrm{MBZ}}} \mathrm{d}k_x \ \mathrm{d}k_y \ (\boldsymbol{\nabla}_{\boldsymbol{k}} \times \langle g|\boldsymbol{\nabla}_{\boldsymbol{k}} g\rangle)_z. \tag{70}$$

Since the ground state $|g\rangle$ is a filled Fermi sea below E_F, it also can be written as

$$\sigma_{xy}^{j,\mathrm{bulk}} = \frac{e^2}{h} \sum_{l=1}^{j} C_l \tag{71}$$

$$C_l = -\frac{1}{2\pi\mathrm{i}} \int\int_{T^2_{\mathrm{MBZ}}} \mathrm{d}k_x \ \mathrm{d}k_y \ [\boldsymbol{\nabla}_k \times \boldsymbol{A}^l(\boldsymbol{k})]_z \tag{72}$$

with

$$\boldsymbol{A}^l(\boldsymbol{k}) = \sum_{m=1}^{q} \psi_m^{l\,*}(\boldsymbol{k})\boldsymbol{\nabla}_k \psi_m^l(\boldsymbol{k}). \tag{73}$$

This is the formula obtained by TKNN [14, 15, 30]. Now let us discuss the topological meaning of the above expression.

An important observation here is that the magnetic Brillouin zone $T^2_{\rm MBZ}$ is topologically a torus rather than a rectangle [15]. Since a torus does not have a boundary, an application of Stokes's theorem to equation (72) would give $\sigma_{xy} = 0$ if $\boldsymbol{A}^l(\boldsymbol{k})$ were well defined on the entire torus $T^2_{\rm MBZ}$. A possible non-zero value of σ_{xy} is a consequence of a non-trivial topology of $\boldsymbol{A}^l(\boldsymbol{k})$. In order to better understand the relevant topology, let us examine the effect of a phase transformation of the wavefunction:

$$\psi^l(\boldsymbol{k})' = \mathrm{e}^{\mathrm{i}f(\boldsymbol{k})}\psi^l(\boldsymbol{k}) \tag{74}$$

where $f(\boldsymbol{k})$ is an arbitrary smooth function of $\boldsymbol{k}$ over the Brillouin zone. The corresponding 'gauge' transformation for $\boldsymbol{A}^l(\boldsymbol{k})$ is

$$\boldsymbol{A}^l(\boldsymbol{k})' = \boldsymbol{A}^l(\boldsymbol{k}) + \mathrm{i}\boldsymbol{\nabla}_k f(\boldsymbol{k}) \tag{75}$$

which leaves σ_{xy} invariant. The non-trivial topology arises when the phase of the wavefunction cannot be determined uniquely and smoothly over the entire Brillouin zone. The gauge transformation defined above implies that the overall phase of the wavefunction can be chosen arbitrarily. It can be determined, for example, by demanding that the qth component of $|\boldsymbol{k}, l\rangle$ be real. However, this is not enough to fix the phase over the Brillouin zone, when $\psi^l_q(\boldsymbol{k})$ vanishes at some points. Let us denote these zeros by k^*_s with $s = 1, \ldots, N$, and define small regions around the zeros by $R^\epsilon_s = \{\boldsymbol{k} \in T^2_{\rm MBZ} \big| |\boldsymbol{k} - \boldsymbol{k}^*_s| < \epsilon,\ \Psi^j_q(\boldsymbol{k}^*_s) = 0\}$, $\epsilon > 0$. We may adopt a different phase convention in R_s so that another component, for example, $\psi^l_1(\boldsymbol{k})$, is real. (We denote it as $|\boldsymbol{k}, l\rangle'$.) Then the overall phase is uniquely determined over the entire Brillouin zone $T^2_{\rm MBZ}$. At the boundaries, ∂R_s, we have a phase mismatch:

$$|\boldsymbol{k}, l\rangle' = \mathrm{e}^{\mathrm{i}f(\boldsymbol{k})}|\boldsymbol{k}, l\rangle. \tag{76}$$

By using the above formulae for gauge transformation, equations (74), (75), we have [15]

$$\sigma^l_{xy} = -\sum_{s=1}^{N} n_s \tag{77}$$

$$n_s = \frac{1}{2\pi}\oint_{\partial R_s} \boldsymbol{\nabla} f(\boldsymbol{k}). \tag{78}$$

Here the n_ss must be integers since all of the state vectors must fit together when we complete full closed paths around each R_s. This implies that the zeros of a certain component of the Bloch function define vortices in the Brillouin zone, whose integral vorticities contribute to the Hall conductance. While the phase of the wavefunction depends on the phase convention (gauge choice), the total vorticity is a gauge-invariant quantity. In this way, in principle, counting the total vorticity of the U(1) phase of the Bloch wavefunction gives the bulk Hall conductance. We use the above expression later to discuss the relationship between the bulk Hall conductance $\sigma^{\rm bulk}_{xy}$ and the edge Hall conductance $\sigma^{\rm edge}_{xy}$.

Before closing this section, we comment on the frequency-dependent conductance, including the longitudinal one. General expressions for the conductance tensors can be obtained by using the Kubo formula. We simply cite the results here [35]. (Some of the results are also obtained in reference [34].) They are expressed as

$$\sigma_{xy}(\omega) = \sigma^{\rm top}_{xy} + \sigma^{S}_{xy}(\omega) \tag{79}$$

$$\sigma^{\rm top}_{xy} = \frac{1}{2\pi\mathrm{i}}\int_{\rm MBZ} \mathrm{d}^2k \ \sum_l f^l(\boldsymbol{k})\boldsymbol{\nabla} \times \langle \boldsymbol{k}, l|\boldsymbol{\nabla}|\boldsymbol{k}, l\rangle \tag{80}$$

$$\sigma_{xy}^{S}(\omega)=\frac{1}{2\pi\mathrm{i}}\int_{\mathrm{MBZ}}\mathrm{d}^2k\ \sum_{l\neq l'}v_x^{ll'}(\boldsymbol{k})v_y^{l'l}(\boldsymbol{k})\frac{f^{l'}(\boldsymbol{k})-f^{l}(\boldsymbol{k})}{(E^{l}(\boldsymbol{k})-E^{l'}(\boldsymbol{k}))^2}\frac{\hbar\omega}{\hbar\omega-\mathrm{i}0+E^{l'}(\boldsymbol{k})-E^{l}(\boldsymbol{k})}. \tag{81}$$

Since $\sigma_{xy}^{S}(\omega=0)=0$, the static Hall conductivity is given by $\sigma_{xy}^{\mathrm{top}}$ which is given by a topological invariant. The longitudinal conductivity $\sigma_{xx}(\omega)$ is given by

$$\sigma_{xx}(\omega)=\sigma_{xx}^{D}(\omega)+\sigma_{xx}^{S}(\omega) \tag{82}$$

$$\sigma_{xx}^{D}(\omega)=\frac{1}{2\pi\mathrm{i}}\int_{\mathrm{MBZ}}\mathrm{d}^2k\ \sum_{l}|v_x^{ll}(\boldsymbol{k})|^2\delta(E_F-E^{l}(\boldsymbol{k}))\frac{1}{\hbar\omega-\mathrm{i}0} \tag{83}$$

$$\sigma_{xx}^{S}(\omega)=\frac{1}{2\pi\mathrm{i}}\int_{\mathrm{MBZ}}\mathrm{d}^2k\ \sum_{l\neq l'}|v_x^{ll'}(\boldsymbol{k})|^2\frac{f^{l'}(\boldsymbol{k})-f^{l}(\boldsymbol{k})}{(E^{l}(\boldsymbol{k})-E^{l'}(\boldsymbol{k}))^2}\frac{\hbar\omega}{\hbar\omega-\mathrm{i}0+E^{l'}(\boldsymbol{k})-E^{l}(\boldsymbol{k})} \tag{84}$$

where $\sigma_{xx}^{D}(\omega)$ is the so-called Drude term and $\sigma_{xx}^{S}(\omega)$ is the contribution from interband scattering processes. The real parts are evaluated as

$$\mathrm{Re}\,\sigma_{xx}^{D}(\omega)=D(q)\delta(\hbar\omega) \tag{85}$$

$$\begin{aligned}D(q)&=\frac{1}{2}\int_{\mathrm{MBZ}}\mathrm{d}^2k\ \sum_{l}|v_x^{ll}(\boldsymbol{k})|^2\delta(E_F-E^{l}(\boldsymbol{k}))\\&=\begin{cases}\dfrac{1}{2}\displaystyle\oint_{E^m(k)=E_F}\mathrm{d}k\ |v_x^{mm}(\boldsymbol{k})|^2\frac{1}{|\boldsymbol{\nabla}E^{m}(k)|} & (E_F\ \text{is in the}\ m\text{th band})\\ 0 & (E_F\ \text{is in the energy gap}).\end{cases}\end{aligned} \tag{86}$$

and

$$\begin{aligned}\mathrm{Re}\,\sigma_{xx}^{S}(\omega)&=\frac{1}{2}\int_{\mathrm{MBZ}}\mathrm{d}^2k\ \sum_{l\neq l'}|v_x^{ll'}(\boldsymbol{k})|^2\{f^{l'}(\boldsymbol{k})-f^{l}(\boldsymbol{k})\}\delta(\hbar\omega+E^{l'}(\boldsymbol{k})-E^{l}(\boldsymbol{k}))\frac{1}{\hbar\omega}\\&=\frac{1}{2\hbar\omega}\sum_{l\neq l',E^{l}\geqslant E_F\geqslant E^{l'}}\oint_{E^{l}(k)-E^{l'}(k)=\hbar\omega}\mathrm{d}k\ |v_x^{ll'}(\boldsymbol{k})|^2\frac{1}{|\boldsymbol{\nabla}(E^{l}(k)-E^{l'}(k))|}\end{aligned} \tag{87}$$

where $f^{l}(\boldsymbol{k})=\theta(E_F-E^{l}(\boldsymbol{k}))$. Only the zero-frequency Hall conductance has any topological meaning.

5. The topological meaning of the σ_{xy} in terms of edge states: winding numbers on the complex-energy surface

The topological meaning of the Hall conductance of the Bloch states (bulk state) is clear, as discussed in the previous section. The Bloch state is defined in the periodic system. One can also define it when the system size is infinite. In any case, the system does not have a boundary. In this sense, the Bloch state is considered as a bulk state. In the discussion of the quantized Hall effect, however, the importance of edge states was pointed out at the beginning [10].

The Hall conductance can be expressed in terms of the edge states on the lattice system, as has been done by Halperin for the continuous system. Then the natural question is that of whether or not the Hall conductance of the edge state has topological meaning. In this section, we give the answer to this question. It *does* have a topological meaning, which is completely different from the bulk one (TKNN) [16].

5.1. Setting up the cylindrical system

To realize the system with edges on a lattice, let us consider a cylindrical system (figure 4). This geometry has been treated numerically by several authors [36–40]. We try to discuss the problem analytically as far as possible.

In this section, let us start from the tight-binding Hamiltonian written in the following form:

$$H = -\sum_{m=1}^{L_x-2}\sum_{n=1}^{L_y}[t_x c^\dagger_{m+1,n}c_{m,n} + t_y \mathrm{e}^{\mathrm{i}(2\pi\Phi/\Phi_0)/L_y} c^\dagger_{m,n+1}\mathrm{e}^{\mathrm{i}2\pi\phi m}c_{m,n}] + \mathrm{HC}. \quad (88)$$

We assume that the system size is L_y along the y-direction and impose a periodic boundary condition in the y-direction. The factor $\mathrm{e}^{\mathrm{i}(2\pi\Phi/\Phi_0)/L_y}$ represents the flux Φ (in units of flux quanta, $\Phi_0 = hc/e$) through the hole (figure 4).

We use a momentum representation in the y-direction:

$$c_{m,n} = \frac{1}{\sqrt{L_y}}\sum_{k_y}\mathrm{e}^{\mathrm{i}k_y n}c_m(k_y) \quad (89)$$

where k_y takes the discrete values $k_y = 2\pi n_y/L_y$, $n_y = 1, \ldots, L_y$. Then the Hamiltonian is given as

$$H = \sum_{k_y} H(k_y) \quad (90)$$

$$H(k_y) = -t_x\sum_{m=1}^{L_x-2}[c^\dagger_{m+1}(k_y)c_m(k_y) + c^\dagger_m(k_y)c_{m+1}(k_y)] - 2t_y\sum_{m=1}^{L_x-1}\cos\left(k_y - 2\pi\frac{\Phi}{\Phi_0 L_y} - 2\pi\phi m\right)c^\dagger_m(k_y)c_m(k_y). \quad (91)$$

For a one-particle state $|\Psi(k_y, \Phi)\rangle = \sum_m \Psi_m(k_y, \Phi)c^\dagger_m(k_y)|0\rangle$, the Schrödinger equation $H|\Psi\rangle = E|\Psi\rangle$ is reduced to that of a one-dimensional problem with parameters k_y and Φ:

$$-t_x\{\Psi_{m+1}(k_y, \Phi) + \Psi_{m-1}(k_y, \Phi)\} - 2t_y\cos\left(k_y - 2\pi\frac{\Phi}{\Phi_0 L_y} - 2\pi\phi m\right)\Psi_m(k_y, \Phi) = E\Psi_m(k_y, \Phi). \quad (92)$$

It is the Harper equation which has been discussed in various different contexts [41, 12, 13].

The boundary condition for realizing the cylindrical geometry is

$$\Psi_0 = \Psi_{L_x} = 0. \quad (93)$$

The energy spectrum of the system is determined by equation (92) and equation (93). We solve this eigenvalue equation using a transfer matrix. Let us write equation (92) in the following matrix form:

$$\begin{pmatrix}\Psi_{m+1}(\epsilon, k_y, \Phi)\\ \Psi_m(\epsilon, k_y, \Phi)\end{pmatrix} = \tilde{M}_m(\epsilon, k_y, \Phi)\begin{pmatrix}\Psi_m(\epsilon, k_y, \Phi)\\ \Psi_{m-1}(\epsilon, k_y, \Phi)\end{pmatrix} \quad (94)$$

$$\tilde{M}_m(\epsilon, k_y, \Phi) = \begin{pmatrix}-\epsilon - 2r\cos(k_y - 2\pi\Phi/\Phi_0 L_y - 2\pi\phi m) & -1\\ 1 & 0\end{pmatrix} \quad (95)$$

where $\epsilon = E/t_x$ is a reduced energy and $r = t_y/t_x$ represents the anisotropy of hopping. (We do not explicitly write in the k_y- and Φ-dependence in the following.) We assume that the system size along the x-direction is commensurate with the flux; that is, we assume

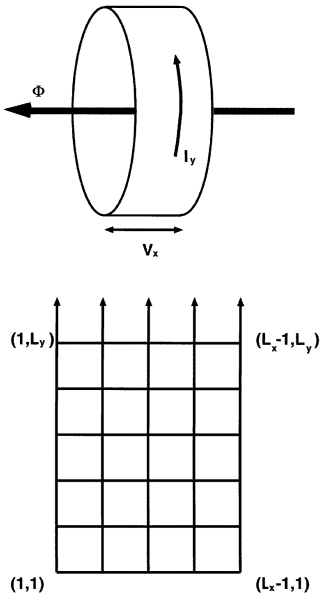


Figure 4. The Laughlin–Halperin geometry. We assume that the system is on a square lattice. The system is periodic along the y-direction and there are two edges, $x = 1$ and $x = L_x - 1$, along the x-direction.

$L_x = ql$ for some integer l. This assumption is needed just for technical reasons. Then, we get a reduced form of the transfer matrix:

$$\begin{pmatrix} \Psi_{L_x+1}(\epsilon) \\ \Psi_{L_x}(\epsilon) \end{pmatrix} = [M(\epsilon)]^l \begin{pmatrix} \Psi_1 \\ \Psi_0 \end{pmatrix} \tag{96}$$

$$M(\epsilon) = \tilde{M}_q(\epsilon)\tilde{M}_{q-1}(\epsilon)\cdots\tilde{M}_2(\epsilon)\tilde{M}_1(\epsilon) = \begin{pmatrix} M_{11}(\epsilon) & M_{12}(\epsilon) \\ M_{21}(\epsilon) & M_{22}(\epsilon) \end{pmatrix} \tag{97}$$

where $M_{11}(\epsilon)$, $M_{12}(\epsilon)$, $M_{21}(\epsilon)$ and $M_{22}(\epsilon)$ are polynomials in ϵ of degree q, $q-1$, $q-1$ and $q-2$, respectively.

All of the solutions of equation (92) with various boundary conditions are obtained for various choices of Ψ_0 and Ψ_1. The eigenvalue problem equation (92) is replaced by an algebraic equation using this transfer matrix. When the boundary conditions along the x-direction are $\Psi_0 = 0$ and $\Psi_{L_x} = 0$, we can choose $\Psi_0 = 0$ (and $\Psi_1 = 1$ to fix the normalization). Then, the energy eigenvalues are determined by $(M(\epsilon)^l)_{21} = 0$, since $\Psi_{L_x}(\epsilon) = (M(\epsilon)^l)_{21}$. From a direct calculation, we know that $(M(\epsilon)^l)_{21}$ is a polynomial in ϵ of degree $L_x - 1$ and has $L_x - 1$ real roots, since they are eigenvalues of the Hermitian Hamiltonian.

By this procedure, the problem of two-dimensional electrons in a uniform magnetic field is reduced to a one-dimensional problem with the parameters k_y and Φ. This is related to the discrete Hill equation, and several studies have been done in the context of non-linear lattice theory [42–45].

5.2. Analytic continuation of the energy: the energy band and the edge state

In this section, we investigate the spectrum of the one-dimensional problem given by the previous section, paying special attention to the edge states. The boundary condition of our problem is

$$\Psi_{L_x} = \Psi_0 = 0. \tag{98}$$

The spectrum of the problem is discrete and is given by the roots of the $(qL_l - 1)$th algebraic equation

$$(M(\epsilon)^l)_{21} = 0. \tag{99}$$

First we point out that solutions of

$$M_{21}(\epsilon) = 0 \tag{100}$$

satisfy equation (99), since a product of tri-diagonal matrices is also tri-diagonal. We write the solutions of equation (100) as μ_j $(\mu_i < \mu_j, i < j)$ [46], which are energies of edge states as we see below.

In fact, equation (100) determines the energies of the edge states. When we set $\Psi_0 = 0$ and $\Psi_1 = 1$ in equation (96), we get

$$\Psi_{qk+1}(\mu_j) = [M_{11}(\mu_j)]^k. \tag{101}$$

If we use the usual normalized wavefunction [47] ($\sum_{m=1,L_x-1} |\Psi_m|^2 = 1$) and the number of sites $L_x - 1$ is sufficiently large, this means that the state is exponentially localized at the edges:

$$|M_{11}(\mu_j)| < 1\text{: localized at } x \approx 1 \qquad \text{(left-hand edge)} \tag{102}$$

$$|M_{11}(\mu_j)| > 1\text{: localized at } x \approx L_x - 1 \qquad \text{(right-hand edge).} \tag{103}$$

This proves that μ_j is the energy of the edge state. (When $|M_{11}(\mu_j)| = 1$, the edge state is degenerate with the bulk state. However, this does not occur generally. This phenomenon is important in its relation to the bulk topological discussion.) Equation (100) means that the wavefunction satisfies

$$\Psi_q = 0 \qquad \Psi_0 = 0. \tag{104}$$

This is a boundary condition of a g-sites ($g = q - 1$) problem with equation (92). The eigenvalues of the $g \times g$ matrix determine the energies of the edge states in this way.

There is an important relationship between the spectrum of this fixed-boundary system and that of the infinite periodic system [42, 43]. If we consider the system as infinite along the x-direction instead of having edges, the one-dimensional Hamiltonian (92) has a translational invariance of period q. Then the Bloch (Floquet) theorem requires that the wavefunction of this infinite-size system satisfies

$$\Psi_{m+q}(\epsilon) = \rho(\epsilon)\Psi_m(\epsilon) \qquad |\rho(\epsilon)| = 1. \tag{105}$$

This should be compared with equation (104). By applying this equation for $m = 0$ and 1, one obtains that ρ is an eigenvalue of M and that ρ is a solution of

$$\rho^2 - \Delta(\epsilon)\rho + 1 = 0 \tag{106}$$

where

$$\Delta(\epsilon) = \mathrm{Tr}\, M = M_{11}(\epsilon) + M_{22}(\epsilon). \tag{107}$$

We have used the fact that

$$\det M(\epsilon) = M_{11}(\epsilon)M_{22}(\epsilon) - M_{12}(\epsilon)M_{21}(\epsilon) = 1 \tag{108}$$

since $\det \tilde{M}_m(\epsilon) = 1$ for all m.

There have been many investigations of the spectrum of this periodic problem and we know that it consists of the q energy bands (continuous spectrum) [14, 30, 11–13, 48, 49]

$$\epsilon \in [\lambda_1, \lambda_2], \ldots, [\lambda_{2j-1}, \lambda_{2j}], \ldots, [\lambda_{2q-1}, \lambda_{2q}] \qquad (\lambda_i \leqslant \lambda_j, i < j). \quad (109)$$

These energy bands are determined by the condition

$$[\Delta(\epsilon)]^2 \leqslant 4. \quad (110)$$

On the other hand, the energies of the edge states μ_j satisfy $M_{11}M_{22} = 1$ by equation (100) and equation (108). Then we have, by equation (107),

$$[\Delta(\mu_j)]^2 = \left(M_{11} + \frac{1}{M_{11}}\right)^2 \geqslant 4. \quad (111)$$

This means that μ_j lies in the energy gaps or at the band edges. Furthermore, we can show that each gap has only one edge state [42, 43]:

$$\mu_j \in [\lambda_{2j}, \lambda_{2j+1}] \qquad j = 1, \ldots, g \ (= q - 1) \quad (112)$$

where $[\lambda_{2j}, \lambda_{2j+1}]$ is the jth energy gap from below. (See appendix C.)

Notice that the boundary condition of our problem is not equation (104), but equation (98). Equation (98) has many solutions other than the solutions of equation (100). But we can show that these extra solutions are in the energy band regions. Therefore they are not the edge states, but are bulk states if the system is sufficiently large.

In the above discussion, we considered the infinite system with period q. But it is also possible to consider the infinite system with period L_x $(= lq)$. In this picture, the spectrum is composed of L_x bands [50]:

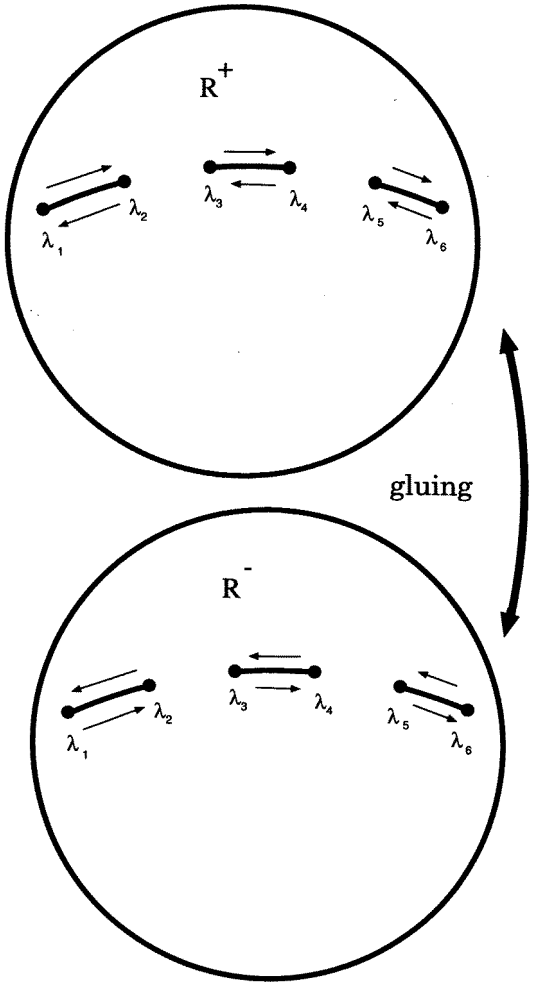
$$[\tilde{\lambda}_1, \tilde{\lambda}_2], \ldots, [\tilde{\lambda}_{2j-1}, \tilde{\lambda}_{2j}], \ldots, [\tilde{\lambda}_{2L_x-1}, \tilde{\lambda}_{2L_x}]. \quad (113)$$

However, the whole spectrum should remain unchanged. Thus, succeeding l-bands touch each other and compose one band:

$$\begin{aligned} &\tilde{\lambda}_{2l(i-1)+2} = \tilde{\lambda}_{2l(i-1)+3} \\ &\vdots \\ &\tilde{\lambda}_{2l(i-1)+2j} = \tilde{\lambda}_{2l(i-1)+2j+1} \\ &\vdots \\ &\tilde{\lambda}_{2l(i-1)+2l-2} = \tilde{\lambda}_{2l(i-1)+2l-1} \\ &i = 1, \ldots, q \text{ and } j = 1, \ldots, l-1. \end{aligned} \quad (114)$$

Using the same argument for equation (112), the energies of our boundary condition, equation (98), are given by these degenerate energies $\tilde{\lambda}_{2l(i-1)+2j}$, $i = 1, \ldots, q$, $j = 1, \ldots, l-1$ (which are in the energy bands and of the bulk states) and the energies of the edge states, μ_j, $j = 1, \ldots, g$ $(= q - 1)$. Counting the number of roots, the above are all of the solutions of our boundary condition, equation (98), and all of the edge states are given by μ_j, $j = 1, \ldots, g$.

When L_x is sufficiently large, the discrete spectrum for the fixed boundary condition converges to a continuous energy band plus the edge states. Thus the spectrum is asymptotically given by the energy band equation (109) and the edges states μ_j in the $L_x \to \infty$ limit.



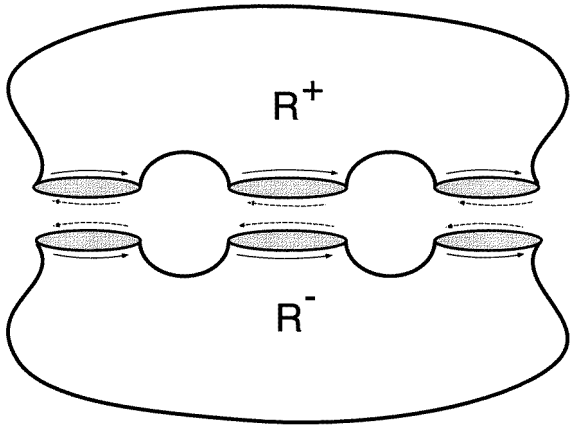
(a)

Figure 5. (a) Two Riemann sheets (Riemann spheres) with $q = g + 1$ cuts which correspond to the energy bands of the system. The Riemann surface of the Bloch function is obtained by gluing the two spheres along the arrows near the cuts. (b) Gluing two Riemann spheres with $q = g + 1$. (c) The Riemann surface of the Bloch function under the rational flux $\phi = p/q$. The number of gaps, g, is the genus of the Riemann surface. α_j and β_j are the canonical loops (generators of the fundamental group) on the Riemann surface.

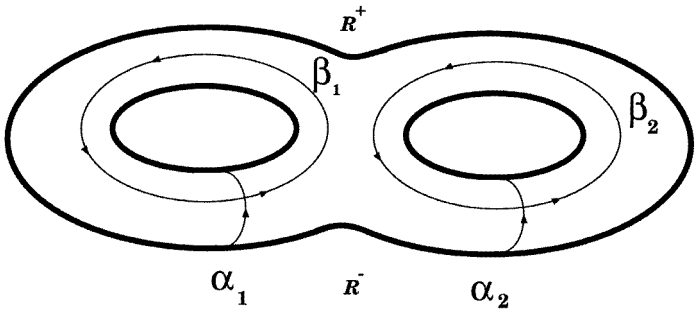
5.3. The winding number of the edge state on the Riemann surface (complex-energy surface)

Here let us consider the Bloch function at site q. The Bloch function is obtained via a different choice of Ψ_0 and Ψ_1 from those for the fixed boundary condition discussed above. For the Bloch function, ${}^t(\Psi_0, \Psi_1)$ is an eigenvector of M with the eigenvalue ρ (see equation (105)). From equation (106), we get $(z = \epsilon)$

$$\rho(z) = \frac{1}{2}(\Delta(z) - \sqrt{\Delta(z)^2 - 4}). \tag{115}$$



(b)



(c)

Figure 5. (Continued)

When we set $\Psi_1 = 1$, the Bloch function at site q is given by

$$\Psi_q(z) = -\frac{M_{11}(z) + M_{22}(z) - \sqrt{\Delta(z)^2 - 4}}{-M_{11}(z) + M_{22}(z) + \sqrt{\Delta(z)^2 - 4}} M_{21}(z). \tag{116}$$

Here we use a Riemann surface of an analytical function

$$\omega = \sqrt{\Delta(z)^2 - 4} = \sqrt{(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_{2q-1})(z - \lambda_{2q})} \tag{117}$$

or a hyperelliptic curve, $\omega^2 = \Delta(z)^2 - 4$, to define the square root consistently. In the above expression, we write the energy ϵ as a complex variable z. In this way, the Bloch wavefunction is defined on the genus-g $(= q - 1)$ Riemann surface (which is a complex-energy surface).

The branch cuts of this function are specified by the q energy bands (equation (109)), and we have to use two sheets, or Riemann spheres (R^+ and R^-), to define the Riemann surface. (The Riemann spheres are obtained by compactifying $z = \infty$ to one point.)

The Riemann surface is obtained by gluing the two sheets at these q cuts along the arrows (see figure 5). The genus of the Riemann surface is $g = q - 1$, which is the number of energy gaps.

The branch of the function is specified by requiring

$$\sqrt{\Delta(z)^2 - 4} > 0 \qquad (z \to -\infty \text{ on the real axis of } R^+). \tag{118}$$

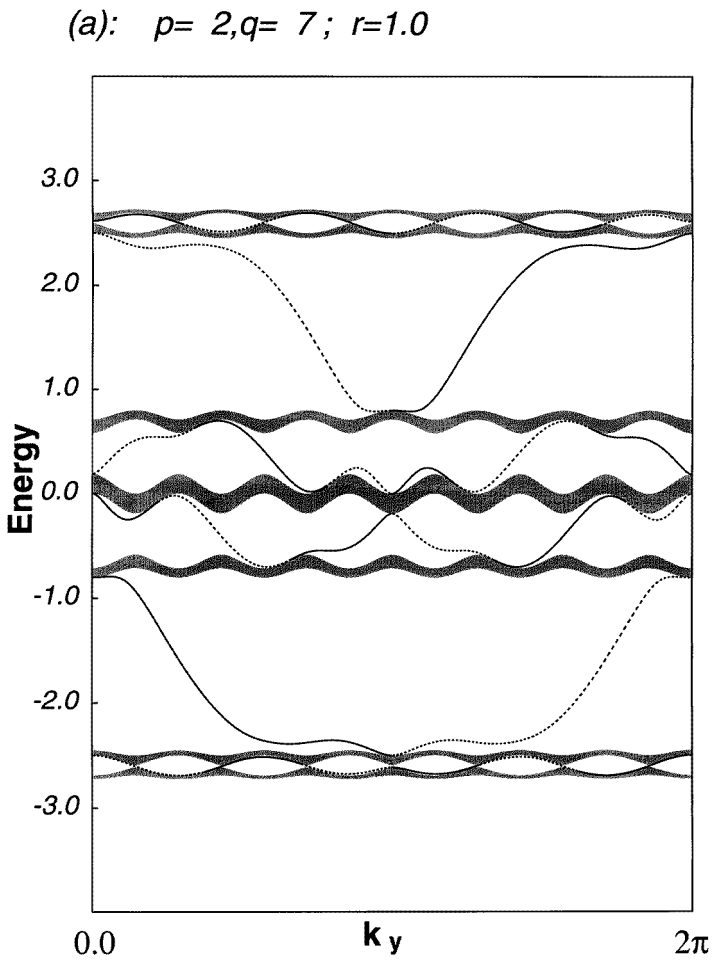


Figure 6. Asymptotic energy spectra of the two-dimensional tight-binding electrons with edges under the rational flux $\phi = p/q$. The number of sites along the x-direction is $L_x - 1 \approx \infty$ but we assume that the size is commensurate with the flux ϕ: $L_x = ql$, l: integer. The shaded areas indicate the energy bands (asymptotically, $L_x \to \infty$) and the lines indicate the spectra of the edge states. A solid line indicates that the energy of the edge state is on the upper Riemann surface R^+ (the state is localized near $x \approx 1$) and a dotted line indicates that it is on the lower Riemann surface R^- (the state is localized near $x \approx L_x - 1$). (a) $\phi = 2/7, r = 1.0$; $C(\mu_1) \sim \beta_1^{-3}$, $C(\mu_2) \sim \beta_2^1$, $C(\mu_3) \sim \beta_3^{-2}$, $C(\mu_4) \sim \beta_4^2$, $C(\mu_5) \sim \beta_5^{-1}$, $C(\mu_6) \sim \beta_6^3$; (b) $\phi = 3/7, r = 1.0$; $C(\mu_1) \sim \beta_1^{-2}$, $C(\mu_2) \sim \beta_2^3$, $C(\mu_3) \sim \beta_3^1$, $C(\mu_4) \sim \beta_4^{-1}$, $C(\mu_5) \sim \beta_5^{-3}$, $C(\mu_6) \sim \beta_6^2$; (c) $\phi = 2/5, r = 1.0$; $C(\mu_1) \sim \beta_1^{-2}$, $C(\mu_2) \sim \beta_2^1$, $C(\mu_3) \sim \beta_3^{-1}$, $C(\mu_4) \sim \beta_4^2$; (d) $\phi = 1/6, r = 1.0$; $C(\mu_1) \sim \beta_1^1$, $C(\mu_2) \sim \beta_2^2$, $C(\mu_3) \sim \beta_3^x, x = ?$, $C(\mu_4) \sim \beta_4^{-2}$, $C(\mu_5) \sim \beta_5^{-1}$ (in the third gap, the gap closes at some values of k_y; then the topology of the Riemann surface changes and we cannot define the winding number without ambiguity); (e) $\phi = 3/7, r = 0.5$; compare with (b).

Then if z lies in the jth gap from below on the real axis (note that there are two real axes),

$$\alpha(-1)^j\sqrt{\Delta(z)^2 - 4} \geqslant 0 \qquad z\text{: real on } R^\alpha \ (\alpha = +, -). \tag{119}$$

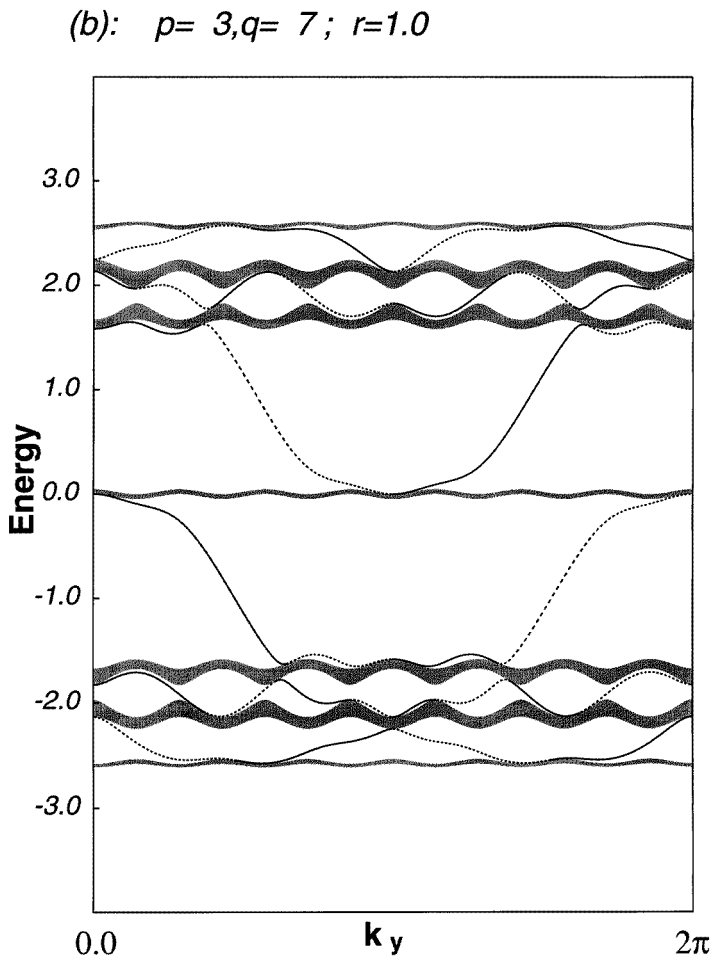


Figure 6. (Continued)

At the energies of the edge states μ_j, $M_{11}M_{22} = 1 + M_{12}M_{21} = 1$ by equation (100) and $\Delta^2 - 4 = (M_{11} + M_{22})^2 - 4 = (M_{11} - M_{22})^2$. By equation (119), this means that

$$\sqrt{\Delta(\mu_j)^2 - 4} = \alpha(-1)^j |M_{11}(\mu_j) - M_{22}(\mu_j)| \qquad (\mu_j \in R^\alpha, \alpha = +, -). \tag{120}$$

By equations (116), (119) and (120), we get

$$\Psi_q(\mu_j + \delta) \approx -\frac{M_{11} + M_{22} - \alpha(-1)^j |M_{11} - M_{22}|}{-M_{11} + M_{22} + \alpha(-1)^j |M_{11} - M_{22}|} M_{21} \qquad (\mu_j \in R^\alpha) \tag{121}$$

where $|\delta| \ll 1$.

From a simple calculation, we can also show that

$$\Delta(\epsilon) = M_{11}(\epsilon) + M_{22}(\epsilon) \begin{cases} \leqslant -2 & j\text{: odd} \\ & \epsilon \in [\lambda_{2j}, \lambda_{2j+1}] \text{ (on } R^\pm) \\ \geqslant 2 & j\text{: even} \end{cases} \tag{122}$$

when the energy ϵ is in the jth gap. From equations (102), (103), (121) and (122), we get important results.

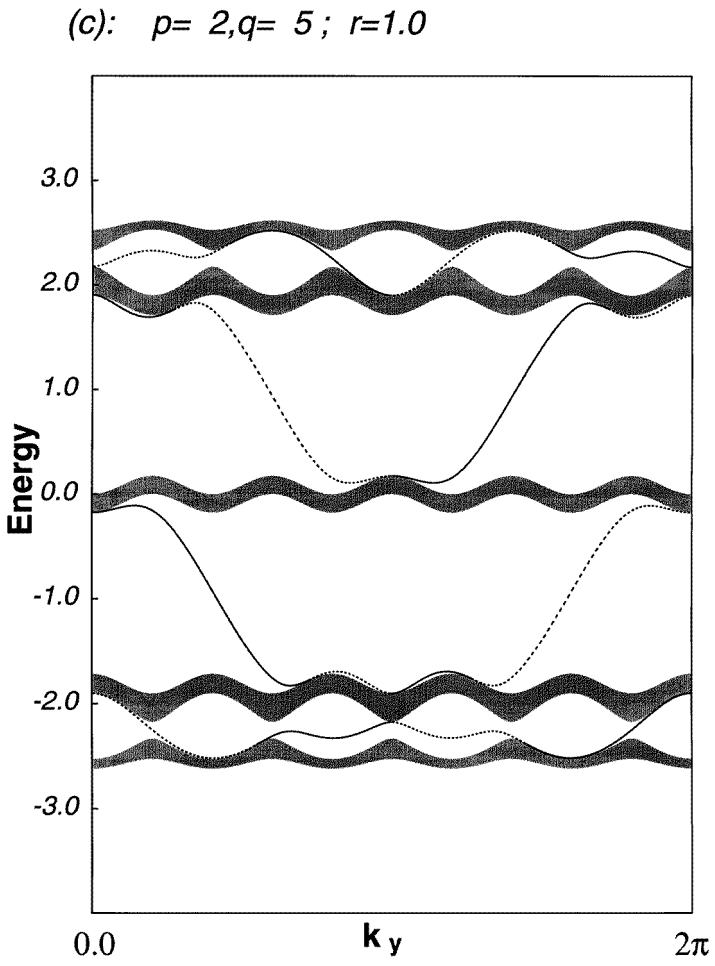


Figure 6. (Continued)

The energy of the edge state μ_j gives a zero point of the Bloch function on the genus-g ($\phi = p/q, g = q - 1$) Riemann surface. When the zero is on the upper sheet of the Riemann surface, the edge state is localized near the left-hand, $x \approx 1$, edge. When the zero is on the lower sheet of the Riemann surface, the edge state is localized near the right-hand, $x \approx L_x - 1$, edge [51].

The above considerations are all for fixed k_y and Φ. As seen from equation (92), the spectrum is a function of $k_y - 2\pi\Phi/L_y$. Actually, the allowed values of k_y are discrete since our system is finite in the y-direction. But we can change it almost continuously when L_y is sufficiently large. Even if L_y is small, we can extrapolate between different k_ys by changing Φ. Keeping this point in mind, we consider k_y as a continuous variable for the time being.

In figures 6, we show asymptotic energy spectra of the two-dimensional tight-binding electrons with two edges under the rational flux $\phi = p/q$. The shaded areas indicate the asymptotic ($L_x \to \infty$) energy band regions and the lines indicate the spectra of the edge states. A solid line indicates that the edge state is localized at $x \approx 1$ and a dotted line

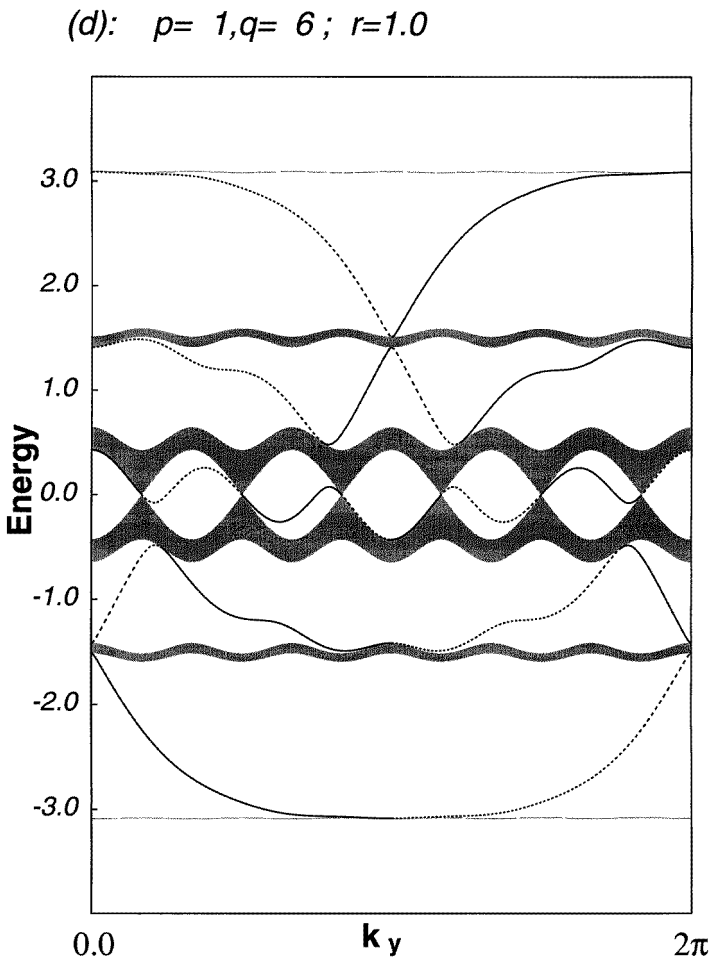
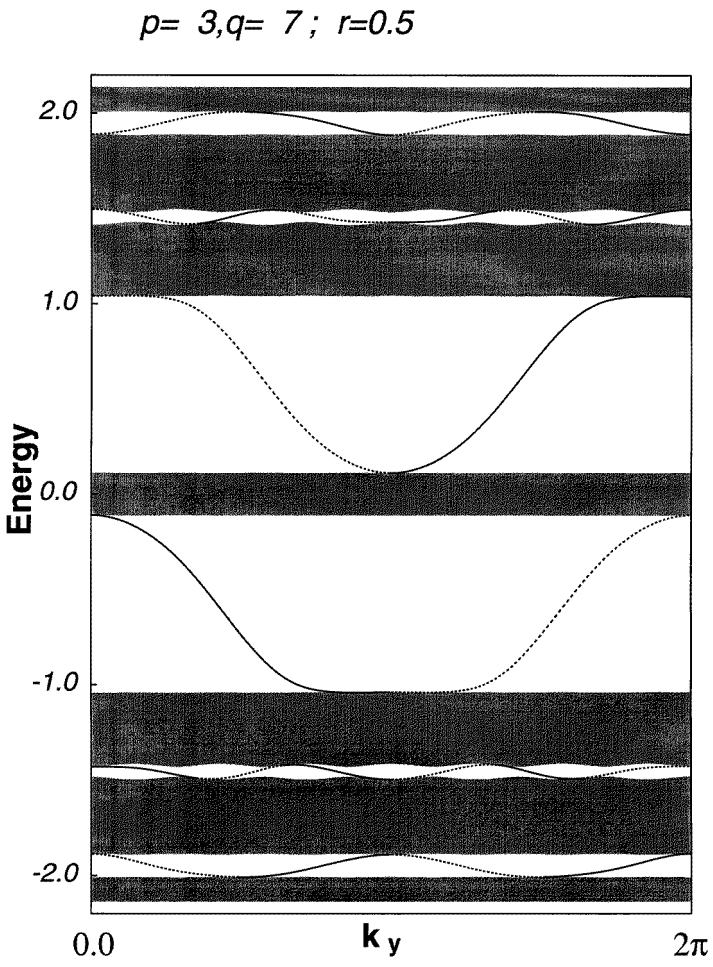


Figure 6. (Continued)

indicates that it is localized at $x \approx L_x - 1$ (equation (102) and equation (103)).

The Riemann surface of the Bloch function is given by the fixed k_y. By changing the k_y, let us consider a family of Riemann surfaces. The surface can be modified by changing k_y; however, the topology of the Riemann surface cannot be changed unless the energy gaps close. In other words, the topology of the Riemann surface does not change if there exist $g = q - 1$ energy gaps in the two-dimensional problem of tight-binding electrons in a magnetic field *without boundaries*. We know that there is a degeneracy at the zero mode in the $q =$ even case [31, 30, 48, 49]. Furthermore, the gap-closing phenomenon occurs when we include nearest-neighbour hoppings [49]. In these cases, the topology of the Riemann surface changes at some value of k_y, and that introduces an ambiguity in the quantized Hall conductance σ_{xy}, since σ_{xy} is given by a topological number on this Riemann surface as we show in the following. For example, we show the result for the $q = 6$ case in figure 6(d). It clearly shows the degeneracies at zero energy [48, 30, 49]. At these degenerate points, one of the holes of the Riemann surface collapses and the topology of the Riemann surface changes. We will discuss the effect of this topology change in relation to σ_{xy} later.



(e)

Figure 6. (Continued)

First of all, the spectrum is a periodic function of k_y with the period 2π. This implies that the zero point of Ψ_q, that is, the energy of the edge state, μ_j, forms a closed loop $C(\mu_j)$ on the Riemann surface on changing k_y from 0 to 2π. When μ_j moves to a different sheet of the Riemann surface, we get $M_{11}(\mu_j) = M_{22}(\mu_j) = \pm 1$, that is, μ_j has to be at the band edge. Using the above discussions, we can follow the movement of μ_j on the Riemann surface using figures 6. The interesting fact is that this movement on the Riemann surface is not always monotonic (for example, see $C(\mu_3)$ in figures 6(a) and 6(b).)

On the genus-g Riemann surface, all kinds of closed loop (the first homotopy group) are generated by the $2g$ canonical loops (generators), α_j and β_j, $j = 1, \ldots, g$, with the defining relation $\prod_{j=1}^{g}(\alpha_j\beta_j\alpha_j^{-1}\beta_j^{-1}) = \hat{1}$. Here $\hat{1}$ means that the loop can be collapsed by continuous change. See figure 5(c). The intersection number of these curves (including directions; see figure 8) [42, 44] $I(\alpha_j, \beta_k)$ is given by

$$I(\alpha_j, \beta_k) = \delta_{jk}. \tag{123}$$

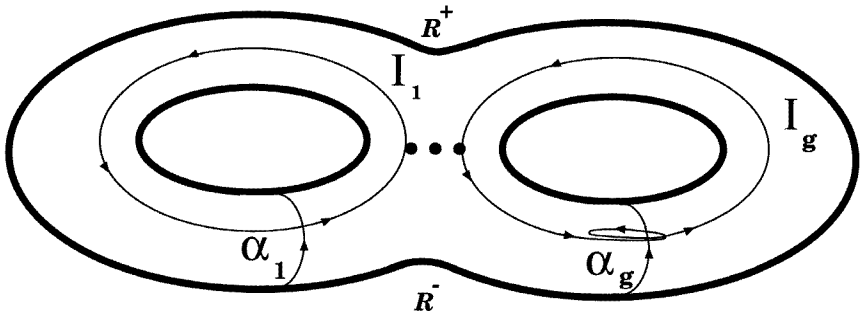
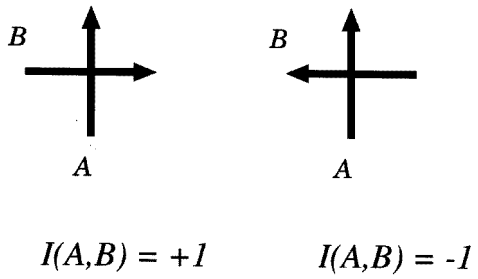


Figure 7. The edge state energy moves on the Riemann surface. The movement is not necessarily monotonic.



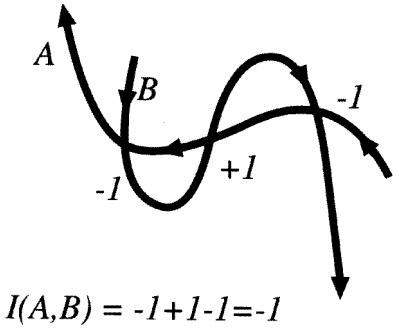


Figure 8. The intersection number $I(A, B)$ of two curves A and B. The intersection points contribute $+1$ or -1 according to the direction.

All of the closed loops on the surface are spanned by α_j and β_j. We can observe that μ_j moves t times around the jth hole for some integer t, that is, homotopically

$$C(\mu_j) \cong \beta_j^t \qquad (t\text{: integer}). \tag{124}$$

This means that

$$I(\alpha_k, C(\mu_j)) = t\delta_{kj} \tag{125}$$

even if the movement of μ_j is not monotonic. Clearly this intersection number is a topological number on the Riemann surface [44]. It is also a winding number of the edge state around the jth hole.

Here also, we get interesting results.

The winding number of the edge state μ_j, which is given by the number of intersections $I(\alpha_j, C(\mu_j))$ between the canonical loop α_j on the Riemann surface and the trace of μ_j, $C(\mu_j)$, gives the quantized Hall conductance σ_{xy} when the Fermi energy lies in the jth gap:

$$\sigma_{xy} = -\frac{e^2}{h} I(\alpha_j, C(\mu_j)). \tag{126}$$

The quantized Hall conductance is expressed in terms of a new topological number on the Riemann surface. This will be understood from the following example using the Laughlin–Halperin geometry [9, 10, 52]. Let us imagine that the Fermi energy lies in a gap as shown in figure 9.

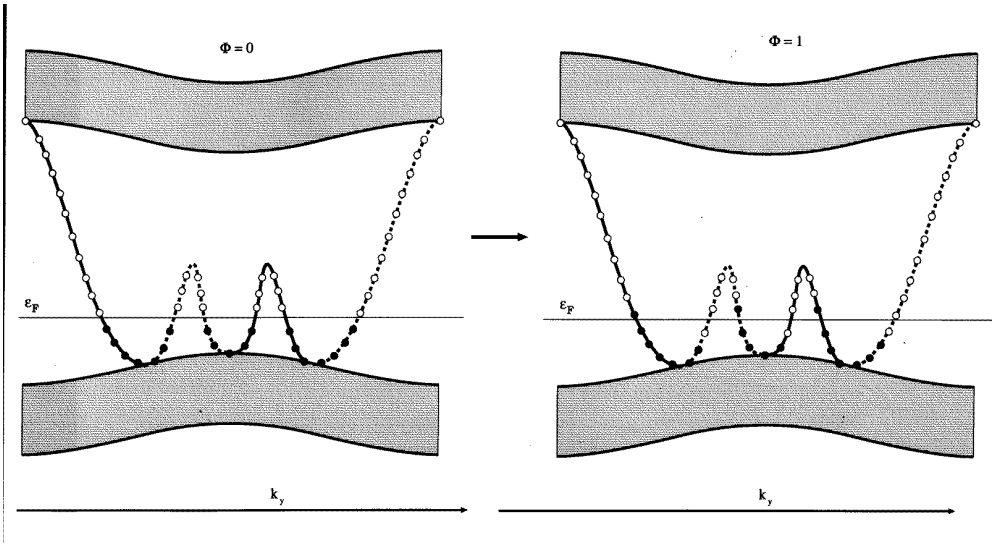


Figure 9. A schematic diagram of the edge states in the energy gap. The circles in the figure denote allowed k_y-points of the edge states. In the left-hand panel, the state is in the ground state. On changing the central flux threading the hole of the system adiabatically from 0 to 1, the ground state is connected to the excited state as shown in the right-hand panel. The solid lines indicate where the state is localized near the left-hand edge $x \approx 1$ and the dotted lines indicate where it is localized near the right-hand edge $x \approx L_x - 1$. We can count how many states are transported during the process by investigating the figure. This number is also related to the intersection number (winding number) of the curves (see the text).

In the following, we consider the discreteness of the allowed k_y explicitly. In the ground state, some of the edge states are occupied (black circles) and some of them are not occupied (white circles) (see the left-hand panel of figure 9). Then let us increase the central flux Φ in the system adiabatically from 0 to 1. All of the states, including edge states and bulk states, are labelled with their values of k_y. During the adiabatic process, the state k_y is shifted to the state $k_y - 2\pi\Phi/\Phi_0 L_y$. For the initial and final states, the spectrum is the same due to the (large) gauge invariance. However, the state has not necessarily returned to the original state. In fact, the bulk states do return to the original states since the Fermi

energy is in the energy gap. But the edge states do not return to the original state as seen from the right-hand panel of figure 9.

In this example, we see two states carried from the right-hand edge to the left-hand edge, but one state carried from the left to the right. As a result, just one state is carried from the right to the left. (The winding number or the intersection number $I(\alpha_j, C(\mu_j))$ in this example is 1.) From this example, it is evident how one can reach a conclusion for general cases. In general, when the Fermi energy lies in the jth energy gap, the *net result* is that $I(\alpha_j, C(\mu_j))$ states are carried from the right-hand edge ($x = L_x - 1$) to the left-hand edge ($x = 1$). Thus the energy of the process is $\Delta E = (-I(\alpha_j, C(\mu_j))(-e)V_x$ where V_x is a voltage along the x-direction. By the Byers and Yang formula, the Hall current I_y is given by [4, 9, 10]

$$I_y = \frac{\Delta E}{\Delta \Phi} = \sigma_{yx} V_x = -\sigma_{xy} V_x. \tag{127}$$

Thus we get the expression for σ_{xy} given as equation (126).

The Hall conductance of the bulk system σ_{xy} is given by the first Chern number of the U(1) fibre bundle on a torus (the magnetic Brillouin zone) [15] and it is a topological invariant [31]. Here we can express σ_{xy} as a different topological number, $I(\alpha_j, C(\mu_j))$, on the Riemann surface of the Bloch function by investigating the edge states.

Now let us study several examples numerically. For an infinite system, we know that σ_{xy} is given by the solution of the following Diophantine equation for an integer t:

$$j \equiv tp \qquad |t| \leqslant q/2 \ (\mathrm{mod}\ q). \tag{128}$$

In this equation, we assume that the Fermi energy lies in the jth gap [53, 54]. Then the Hall conductance is quantized as $-t(e^2/h)$ [14, 30]. Here we have obtained another method for calculating σ_{xy}, by calculating the winding number of the edge state. This means that $I(\alpha_j, C(\mu_j))$ satisfies the Diophantine equation

$$j \equiv I(\alpha_j, C(\mu_j))p \qquad (\mathrm{mod}\ q) \tag{129}$$

(see table 1).

We have performed extensive calculations for many cases ($q < 13$) and confirmed that the integers that we obtained from counting the intersection numbers are the same as those given by the Diophantine equation. The Diophantine equation was originally derived in the anisotropic limit using perturbation theory [14, 30]. For this anisotropic case, we also performed several numerical calculations. In figure 6(e), we show one example for the anisotropic case. Comparing it with the isotropic case (b), we see that the winding number is the same. Furthermore, we see that the movement of the zero point on the Riemann surface is monotonic. The Diophantine equation is also derived in the edge state picture [16].

5.4. Gap closing and Dirac fermions

Here, we comment on a gap closing which occurs in some situations [30, 48, 49]. One example is the $\phi = 1/6$ case (see figure 6(d)). Except for the third gap, we can define the winding number without ambiguity. For the third gap, however, one of the g ($= 5$) holes of the Riemann surface collapses at the degenerate points. At these points, the topology of the Riemann surface changes and the winding number is not well defined. (See figure 10.)

This corresponds to the ambiguity of the σ_{xy}. At the degenerate points, two energy bands are degenerate and the energies of the edge state are pinched by the energy bands and also degenerate [55]. If we include next-nearest-neighbour hopping, the degeneracies are removed and we get a well defined winding number for this case [56, 49, 57, 58].

Table 1. Solutions of the Diophantine equation corresponding to figures 6. Compare these with the winding numbers in figures 6. $j = sq + tp$.

$\phi = p/q$	jth gap	t: $\sigma_{xy} = (e^2/h)t$	s
$\phi = 2/7$	1	−3	1
	2	1	0
	3	−2	1
	4	2	0
	5	−1	1
	6	3	0
$\phi = 3/7$	1	−2	1
	2	3	−1
	3	1	0
	4	−1	1
	5	−3	2
	6	2	0
$\phi = 2/5$	1	−2	1
	2	1	0
	3	−1	1
	4	2	0
$\phi = 1/6$	1	1	0
	2	2	0
	3	?	?
	4	−2	1
	5	−1	1

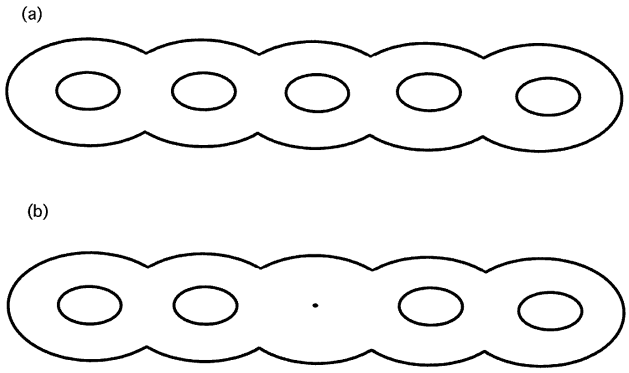


Figure 10. The Riemann surface of the Bloch function for $\phi = 1/6$; (a) at a general k_y and (b) at $k_y = 2\pi(n + 1/2)/6$, $n = 0, 1, 2, 3, 4,$ and 5, where the third gap collapses at the momenta and the corresponding hole of the Riemann surface closes. Therefore the winding number of the edge state in the third gap (around the third hole) is ambiguous when changing k_y from 0 to 2π.

In general, when the gap closes, two-dimensional Dirac fermions appear near the gap-closing points. This is discussed in detail in another article [58].

5.5. Effects of randomness

As is well known, the existence of randomness is necessary for explaining the QHE [9, 59, 10, 16, 60]. However, it is very difficult to include the effects of the randomness completely in a large system. Here we rather artificially introduce *one-dimensional*

randomness in a two-dimensional system. The Hamiltonian is given by

$$\text{equation (88)} + \sum_{n,m} V(m) c^{\dagger}_{m,n} c_{m,n} \tag{130}$$

where we assume that $V(m)$ is a uniform random number in the interval $[-V_{rnd}, V_{rnd}]$. This potential is random along the x-direction but uniform in the y-direction. Bulk properties such as the localization depend crucially on the dimensionality. However, as far as the edge states are concerned, we hope that the artificial one-dimensional randomness still includes some effects of the true two-dimensional randomness. In this case, we can perform a Fourier transformation along the y-direction, and the spectrum of the system is obtained by diagonalizing the one-dimensional system with the parameter k_y.

We show the spectrum of the system for $p = 2, q = 5$, and $L_x = 125$, with randomness $V_{rnd} = 0.5t_x$, in figure 11. (With realistic randomness, our naive speculation is that we lose the quantum number k_y and we cannot specify the state by giving k_y, but that the spectrum and the many-body wavefunction remain similar.)

Let us compare figure 11 with figure 6(c). It seems that there still exist edge states in this random system. For the second and third gaps, we can clearly distinguish the edge states from the bulk states. When the Fermi energy lies in these gaps, we can repeat the argument of the previous section and obtain the quantized value of the Hall conductance.

In the first and the fourth energy gaps, however, there are several level crossings. Thus we cannot perform the adiabatic process discussed in the previous section. This means that σ_{xy} is not well quantized when the Fermi energy is in these energy gaps. By this argument, we believe that the quantization is more accurate when the Fermi energy lies in the larger energy gap.

The essential point of the above argument is that we have to treat one (macroscopic) Schrödinger equation even if there is randomness in the system. The usual averaging procedures for randomness are not suitable for the quantized Hall conductance, especially for the edge state effects.

The bulk state argument discussed in the previous section is only applicable for an ideal clean system and it is not easy to justify the argument for a realistic system. The occurrence of the Anderson localization makes the situation quite subtle. On the other hand, the edge states cannot localize in a macroscopic system, since the overlaps of the two edge states at two different edges are exponentially small even in a random system. Therefore, one anticipates that the edge states argument is stable against the effect of the randomness.

Another important comment is that two energy levels of the edge states at different edges can cross (not repel) in suitable geometry (cylindrical) when one considers a spectral flow in a macroscopic system. This can be understood as a realization of a new symmetry in a random macroscopic system.

6. The relationship between the two topological numbers

As discussed in sections 4 and 5, the quantization of the Hall conductance can be explained topologically in two ways. One is in terms of the bulk states and the other is in terms of the edge states. Therefore we have two topological expressions for the Hall conductance. Their topological meanings are apparently quite different. However, there should be some relationship between the two. This relationship was made clear in [17, 18]. In this section, we will explain it, following these references.

The two topological expressions for the Hall conductance, when the Fermi energy is in

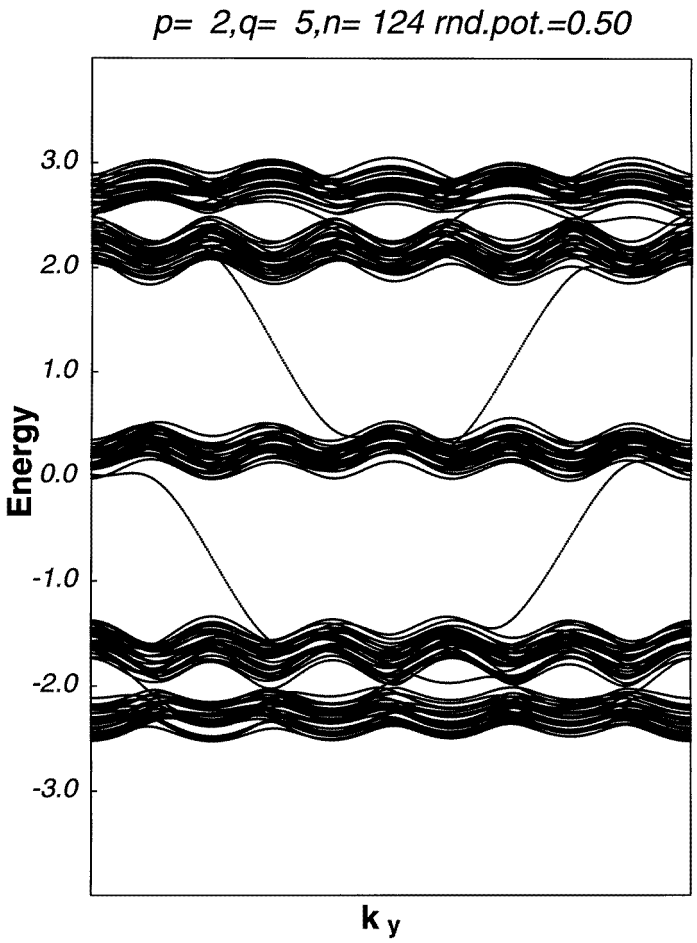


Figure 11. The energy spectrum of the two-dimensional tight-binding electrons with two edges under the rational flux $\phi = p/q$ where we include the effect of a 'one-dimensional' random potential ($t_x = t_y = 1$, $V_{rnd} = 0.5$, $p = 2, q = 5$, and $L_x = 125$). See the text, and compare with figure 6(c).

the jth energy gap, are

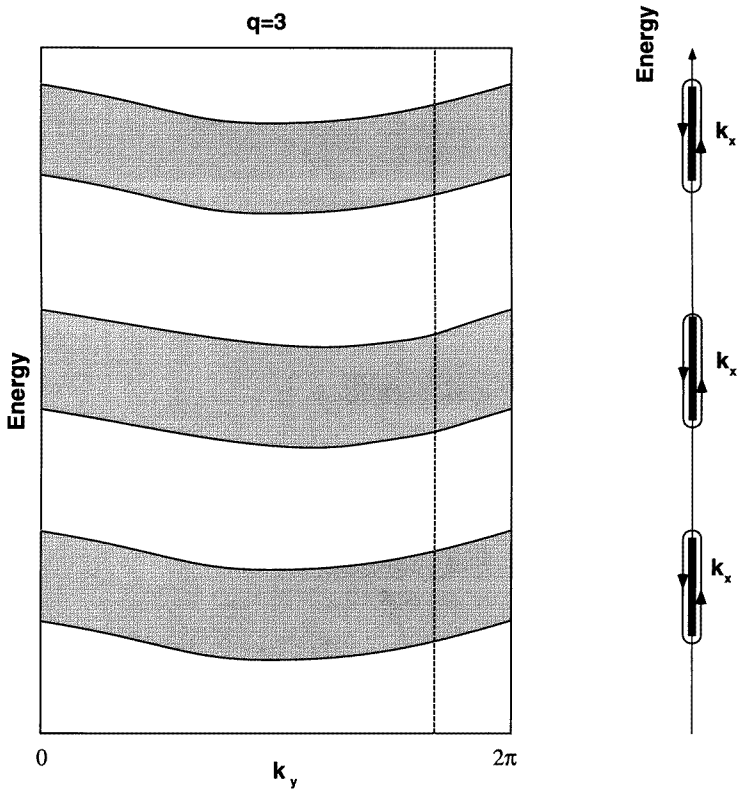
$$\sigma_{xy}^{\text{bulk},j} = -\frac{e^2}{h}\sum_{l=1}^{j} C_l \tag{131}$$

$$C_l = \frac{1}{2\pi \mathrm{i}}\int\int_{T^2_{\text{MBZ}}} \mathrm{d}k_x\ \mathrm{d}k_y\ [\nabla_k \times \boldsymbol{A}^l(\boldsymbol{k})]_z \tag{132}$$

and

$$\sigma_{xy}^{\text{edge},j} = -\frac{e^2}{h} I(\alpha_j, C(\mu_j)). \tag{133}$$

The bulk expression is given by the integral over the magnetic Brillouin zone and the edge expression is given by the winding number of the zero of the edge state energy on the Riemann surface. To discuss the relationship between the two, we need to clarify the



(a)

Figure 12. (a) A schematic energy spectrum. The shaded areas indicate the energies of the bulk state and the lines indicate those of the edge states. (b) The shaded areas in (a) actually indicate tubes parametrized by k_x. (c) The tubes in (b) are actually q-tori which constitute the magnetic Brillouin zone for q energy bands.

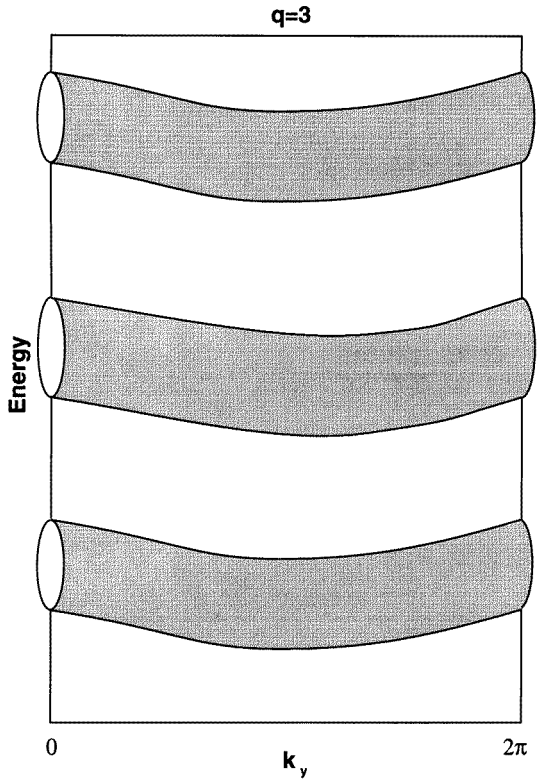
relationship between the magnetic Brillouin zone and the Riemann surface (a complex-energy surface).

Let us consider two different systems, with (cylindrical) and without (infinite or periodic) edges. In both cases, let the system be periodic in the y-direction. Therefore, by Fourier transforming along the y-direction, one obtains a one-dimensional Schrödinger equation:

$$-t\{\Psi_{m+1}(k_y) + \Psi_{m-1}(k_y)\} - 2t\cos(k_y - 2\pi\phi m)\Psi_m(k_y) = E(k_y)\Psi_m(k_y). \tag{134}$$

Without edges, that is, if the system is infinite, this one-dimensional system has a period q when $\phi = p/q$, and the Bloch theorem tells us that the function Ψ_m satisfies $\Psi_m(k_x, k_y) = \mathrm{e}^{\mathrm{i}k_x m}\tilde{u}_m(k_x, k_y)$ where $\tilde{u}_{m+q}(\boldsymbol{k}) = \tilde{u}_m(\boldsymbol{k})$ and $k_x \in [0, 2\pi/q]$.

In this case, the vector potential in the magnetic Brillouin zone, $\boldsymbol{A}^l(\boldsymbol{k})$ is given by $\boldsymbol{A}^l_u(\boldsymbol{k}) = \langle u^l(\boldsymbol{k})|\boldsymbol{\nabla}_k|u^l(\boldsymbol{k})\rangle = \{u^l_m(\boldsymbol{k})\}^*(\boldsymbol{k})\,\boldsymbol{\nabla}_k u^l_m(\boldsymbol{k})$ and u^l_m is obtained from $\tilde{u}^l_m$ via the normalization $\langle u^l(\boldsymbol{k})|u^l(\boldsymbol{k})\rangle = 1$. When the Fermi energy lies in the jth gap, the Hall conductance of the system is given by $\sum_{l=1}^{j}\sigma^l_{xy}$. The wavefunction $|u^j(\boldsymbol{k})\rangle$ forms a U(1) fibre bundle on the magnetic Brillouin zone and the integral equation (132) is the Chern number which is a topological invariant of the U(1) bundle [14, 15, 30]. This gives a topological meaning to the Hall conductance in the infinite system (without



(b)

Figure 12. (Continued)

edges). In equation (132), there is a gauge freedom which arises from the *phase ambiguity* of $|u(\boldsymbol{k})\rangle$. One can use another gauge to calculate σ_{xy}^{j} via the gauge transformation $|v^{j}(\boldsymbol{k})\rangle = \mathrm{e}^{\mathrm{i}f(\boldsymbol{k})}|u^{j}(\boldsymbol{k})\rangle$ (f: real).

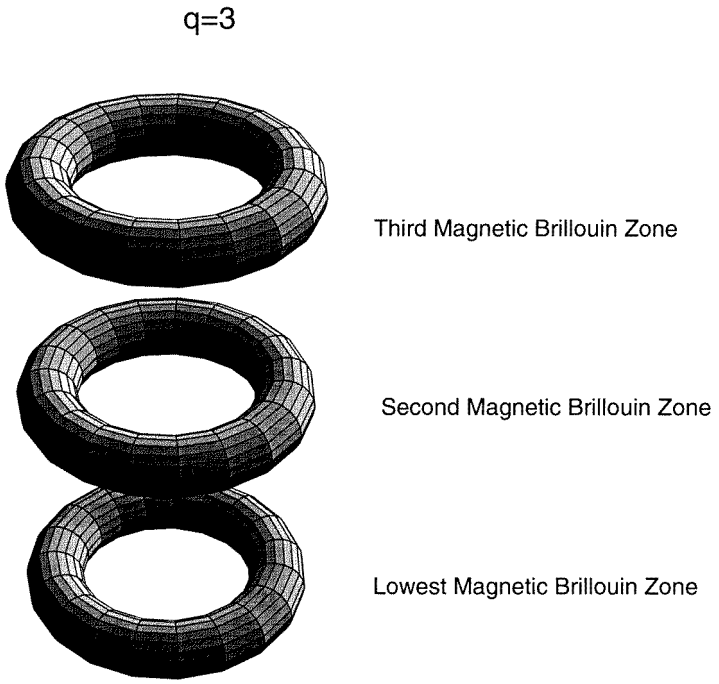
As discussed in section 4, we divide T^2_{MBZ} into several regions. We require that $\Psi_m^j(\boldsymbol{k})$ be an analytical function of k_x and k_y in the magnetic Brillouin zone and choose $\Psi_1^j = 1$; i.e., $\tilde{u}_1^j = \mathrm{e}^{-\mathrm{i}k_x}$. By using a geometry with edges, we give below an explicit construction for $\Psi_m^j(\boldsymbol{k})$ that satisfies these requirements . This convention is not compatible with the periodicity of the magnetic Brillouin zone. However, we avoid this difficulty in the following by using the freedom of the gauge transformation.

We denote the zero points of the $\Psi_q^j(\boldsymbol{k})$ in the magnetic Brillouin zone by $\boldsymbol{k} = \boldsymbol{k}_1^*, \ldots, \boldsymbol{k}_{N_j}^*$. At these points, $u_q(\boldsymbol{k}) = 0$ also. Then we divide the whole magnetic Brillouin zone T^2_{MBZ} into $N_j + 1$ parts as follows:

$$R_s^{\epsilon} = \{\boldsymbol{k} \in T^2_{\mathrm{MBZ}} \,|\, |\boldsymbol{k} - \boldsymbol{k}_s^*| < \epsilon,\ \Psi_q^j(\boldsymbol{k}_s^*) = 0\} \tag{135}$$

$$R_0 = T^2_{\mathrm{MBZ}} \Big\backslash \bigcup_{s=1}^{N_j} R_s^{\epsilon} \tag{136}$$

where $s = 1, \ldots, N_j$, and $\epsilon \ll 1$. In the local regions R_s^{ϵ}, we use a phase convention for $|U^j(\boldsymbol{k})\rangle$ to calculate $\boldsymbol{A}_U(\boldsymbol{k})$: that $U_m^j(\boldsymbol{k}) = u_m^j(\boldsymbol{k})$. In the region R_0, we use another phase convention for $|V^j(\boldsymbol{k})\rangle$ to calculate $\boldsymbol{A}_V^j(\boldsymbol{k})$: that $V_q^j(\boldsymbol{k})$ is real and positive. This convention



(c)

Figure 12. (Continued)

for R_0 is compatible with the periodicity of the magnetic Brillouin zone, provided that we choose $V_m^j(\boldsymbol{k}) = u_m^j(\boldsymbol{k})\mathrm{e}^{-\mathrm{i}\xi(\boldsymbol{k})}$, where $\xi(\boldsymbol{k}) = \mathrm{Im}\ln u_q^j(\boldsymbol{k}) = \mathrm{Im}\ln\tilde{u}_q^j(\boldsymbol{k})$. This is well defined since $\Psi_q^j(\boldsymbol{k}) = \mathrm{e}^{\mathrm{i}qk_x}\tilde{u}_q(\boldsymbol{k}) \neq 0$ in R_s^ϵ. Then using Stokes's theorem in each region, we get from equation (132)

$$\sigma_{xy}^j = -\sum_{s=1}^{N_j}\frac{1}{2\pi\mathrm{i}}\oint_{\partial R_s}\mathrm{d}\boldsymbol{k}\cdot(\boldsymbol{A}_U^j(\boldsymbol{k}) - \boldsymbol{A}_V^j(\boldsymbol{k}))$$
$$= -\sum_{s=1}^{N_j}\frac{1}{2\pi}\oint_{\partial R_s}\mathrm{d}\boldsymbol{k}\cdot\boldsymbol{\nabla}\xi(\boldsymbol{k}) = -\sum_{s=1}^{N_j}\frac{1}{2\pi}\oint_{\partial R_s}\mathrm{d}\boldsymbol{k}\cdot\boldsymbol{\nabla}\,\mathrm{Im}\ln\Psi_q(\boldsymbol{k}) \qquad (137)$$

where we use $\oint_{\partial R_s}\mathrm{d}\boldsymbol{k}\cdot\boldsymbol{\nabla}\,\mathrm{Im}\ln\mathrm{e}^{\mathrm{i}qk_x} = 0$ in the last line. Since the zero of the wavefunction at $\boldsymbol{k} = \boldsymbol{k}_s^*$ gives a vortex-like structure, σ_{xy}^j is obtained as the sum of the vorticities in the magnetic Brillouin zone [15].

Now let us consider a system in a strip geometry with edges, i.e., the system size is infinite in the y-direction and finite in the x-direction [16]. We apply the transfer-matrix method of the previous section. The boundary condition for the strip geometry is

$$\Psi_0^{(e)}(k_y) = \Psi_{L_x}^{(e)}(k_y) = 0 \qquad \text{(with edges)}. \qquad (138)$$

We assume that L_x is commensurate with the flux ϕ, that is, $L_x = ql$ with some integer l. Mathematically, the difference between systems with and without edges corresponds to a difference in boundary conditions. For the system without edges, the condition follows

from the Bloch theorem:

$$\Psi^{(b)}_{m+q}(k_y) = \rho\Psi^{(b)}_m(k_y) \qquad \rho = \mathrm{e}^{\mathrm{i}qk_x} \qquad (k_x \in [0, 2\pi/q]) \qquad \text{(without edges)}. \tag{139}$$

As discussed before, the energy of the edge state μ_j $(j = 1, \ldots, g)$ is determined by a boundary condition of the q-site problem [16]:

$$M_{21}(\mu_j) = \Psi^{(e)}_q(\mu_j) = 0. \tag{140}$$

On the other hand, the Bloch function (without edges) satisfies equation (139). Therefore Ψ_1 and Ψ_0 form an eigenvector of M with the eigenvalue ρ:

$$M(\epsilon)\begin{pmatrix}\Psi^{(b)}_1\\ \Psi^{(b)}_0\end{pmatrix} = \begin{pmatrix}\Psi^{(b)}_{q+1}(\epsilon)\\ \Psi^{(b)}_q(\epsilon)\end{pmatrix} = \rho(\epsilon)\begin{pmatrix}\Psi^{(b)}_1\\ \Psi^{(b)}_0\end{pmatrix}. \tag{141}$$

In equation (141), the energy ϵ represents a real variable and we can analytically continue ϵ to a complex energy z in order to discuss the wavefunction of the edge state.

We get from equation (141)

$$\rho(z) = \frac{1}{2}(\Delta(z) - \sqrt{\Delta(z)^2 - 4}) \tag{142}$$

$$\Psi^{(b)}_q(z) = -\frac{\rho M_{21}(z)}{M_{22} - \rho} = \frac{1}{M_{12}}(\rho^2 - \rho M_{11}) \tag{143}$$

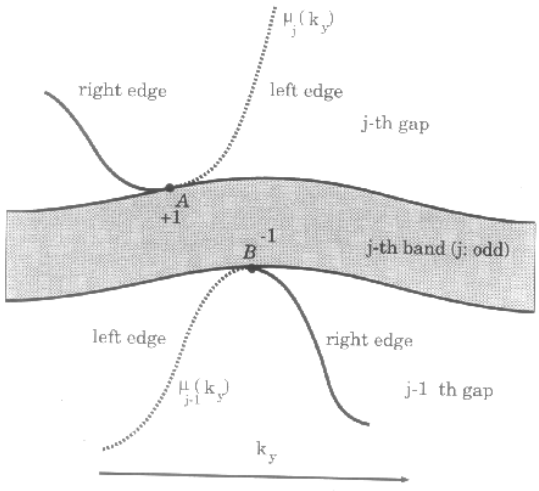
where $\Delta(z) = \mathrm{Tr}\, M(z)$ and we impose an initial condition $\Psi^{(b)}_1 = 1$ which is compatible with the above requirements. To consider the wavefunction as a single-valued function of the energy, we need to use the Riemann surface $\Sigma_g(k_y)$ of $\omega(z) = \sqrt{\Delta(z)^2 - 4}$.

Now we calculate the Hall conductance in equation (137) by using the explicit wavefunction given as equation (143). When we restrict the complex energy to the real axis, we obtain the Bloch function for the jth band $\Psi^j_q(\epsilon, k_y)$ in the magnetic Brillouin zone. The energy ϵ of the energy band lies on the q circles S^1_j, $j = 1, \ldots, q$, which are explicitly defined on $\Sigma_q(k_y)$ (see figure 5). This is a kind of on-shell condition. These circles are parametrized by k_x through $\rho = \mathrm{e}^{\mathrm{i}qk_x}$. Using the relation $-2\mathrm{i}\sin(qk_x) = \sqrt{\Delta^2 - 4}$ and taking into consideration the branch convention for $\sqrt{\Delta^2 - 4}$, we can get an explicit parametrization of the circles $S^1_j(k_x)$. This parametrization of $S^1_j(k_x)$ is crucial for the discussion. Therefore, shaded areas in figure 12(a) are actually q-tubes, as shown in figure 12(b). Since $k_y = 0$ and $k_y = 2\pi$ should be identified, we have to glue the boundaries of the tubes together. After that they become q-tori (figure 12(c)).

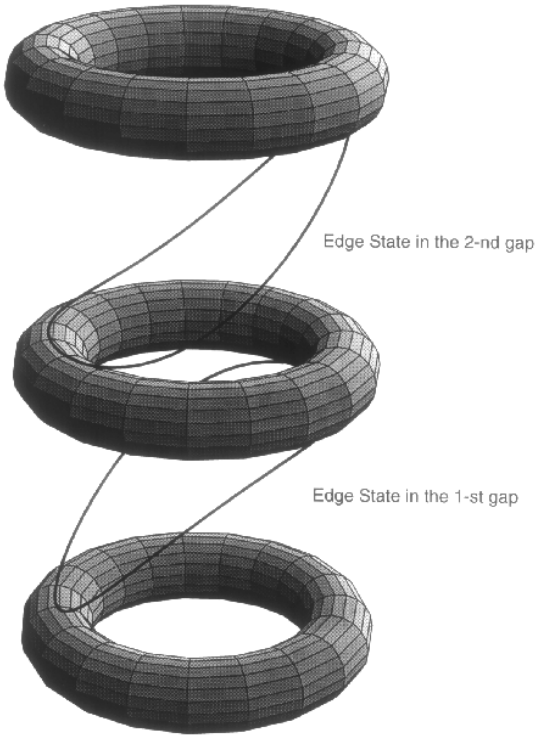
The momentum k_y also lies on the circle $S^1(k_y)$ and the product $S^1_j(k_x) \times S^1(k_y)$ is the magnetic Brillouin zone for each energy band j. The Bloch function $\Psi^j_q(\epsilon, k_y)$ is analytical in ϵ and k_y, and ϵ is analytical in k_x and k_y. Therefore $\Psi^j_q(k_x, k_y)$ is also an analytical function of both k_x, and k_y and $\Psi^j_1 = 1$. This Bloch function satisfies the requirements discussed earlier for calculating σ_{xy}.

Moreover one can see that the zeros of the Bloch function $\Psi^j_q(k_x, k_y)$ are given by the zeros of M_{21}, since ρ is always non-zero in the magnetic Brillouin zone and M_{12} is a polynomial in the energy variable. The zero of $M_{21}(\epsilon)$ is given by the energy of the edge state and always lies in energy gaps or at band edges. *Therefore the zeros of* $\Psi^{(b)}_q(k_x, k_y)$ *are given by the points where the edge state is degenerate with the bulk state at the band edges.* (See figures 13.)

Near the degenerate point $\boldsymbol{k} = (k^*_x, k^*_y)$, we expand the Ψ^j_q, up to linear order in $\delta k_x = k_x - k^*_x$ and $\delta k_y = k_y - k^*_y$, to calculate σ_{xy}. For example, consider the contribution



(a)



(b)

Figure 13. (a) A typical example for the degeneracy of the edge state with the bulk state. (b) A schematic picture of the edge states and the magnetic Brillouin zones for each energy band.

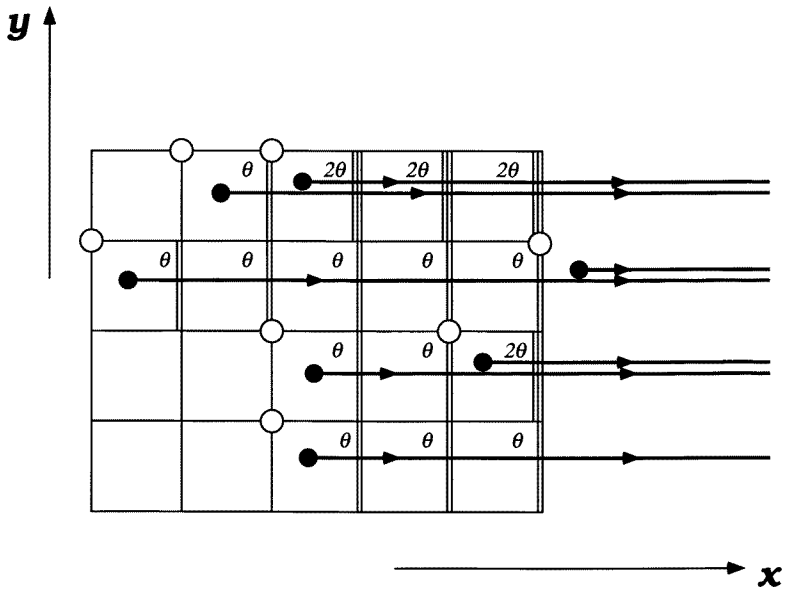


Figure 14. Anyons on the lattice. A θ-anyon is represented by a hard-core boson carrying a string with strength θ. (The boson carries a θ flux tube.) When another anyon cuts the string (in the right-pointing direction) from below, it gives a phase $e^{i\theta}$ to the system.

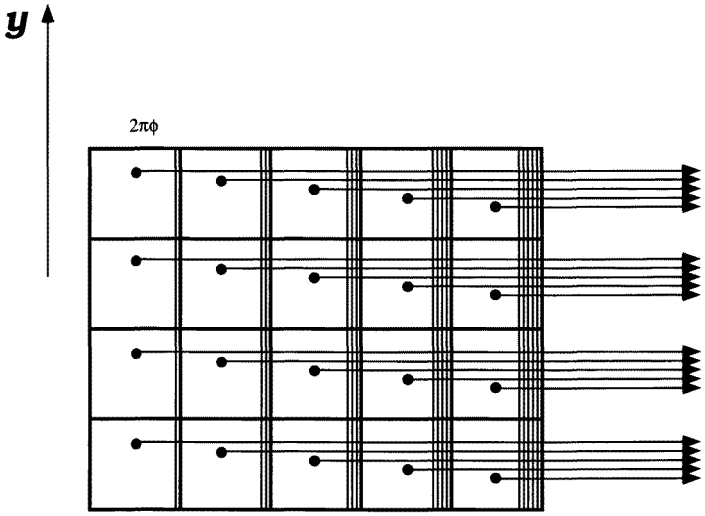


Figure 15. Hard-core bosons with the uniform magnetic field ϕ on the lattice. The effect of the magnetic field is expressed as strings on each plaquette.

from the jth energy band (j: odd) and focus on the degenerate point at the band top A as shown in figure 13. Near A, μ_j moves from the upper Riemann surface R^+ to the lower Riemann surface R^-. In the jth energy gap for odd j, one can show that $M_{11}(\mu_j(k_y)) < 0$ and $k_x^* = \pi/q$ [16]. Therefore we obtain $-1 < M_{11}(\mu_j(k_y))$ when $k_y < k_y^*$ (on the solid

line in the gap), and $M_{11}(\mu_j(k_y^*)) = -1$ and $M_{11}(\mu_j(k_y)) < -1$ when $k_y^* < k_y$ (on the dotted line). (See figures 6 and 13.)

Near the band edge, $k_x^* = \pi/q$, the energy deviation in δk_x is proportional to $(\delta k_x)^2$. Therefore $M_{11} \approx -1 - C\delta k_y$, $C > 0$ and $\rho \approx -1 - \mathrm{i}q\,\delta k_x$ up to linear terms in δk_x and δk_y. Then from equation (143), $\Psi_q^j(k_x, k_y)$ is given as $\Psi_q^j(k_x, k_y) \approx C'(-C\,\delta k_y + \mathrm{i}q\,\delta k_x)$. From the expression, we obtain $(1/2\pi)\oint_A \mathrm{d}\boldsymbol{k} \cdot \boldsymbol{\nabla}\,\mathrm{Im}\,\ln\Psi_q = +1$. When μ_j moves from the lower Riemann surface R^- to the upper Riemann surface R^+, the vorticity is -1. On the other hand, when the degeneracy is at the band bottom (point B in figure 13), one can obtain similarly $(1/2\pi)\oint_B \mathrm{d}\boldsymbol{k} \cdot \boldsymbol{\nabla}\,\mathrm{Im}\,\ln\Psi_q = -1$. Near B, μ_j moves from R^- to R^+. When μ_j moves from R^+ to R^-, it gives $+1$. This consideration is for the odd-j case, and a similar approach is also applied to the even-j case.

In this way, we can derive a formula for the general case. The sum of the vorticities at the band top is written as $I(C_j)$ by using the winding number of the edge state, and that at the band bottom is $-I(C_{j-1})$. Therefore the Hall conductance for the filled jth band is

$$\sigma_{xy}^{j,\mathrm{bulk}} = -\frac{e^2}{h}[I(C_j) - I(C_{j-1})] \equiv \sigma_{xy}^{j,\mathrm{edge}} \tag{144}$$

because $I(C_0) = 0$.

This expression shows the relationship between the TKNN number for the bulk state and the winding number of the edge state. It clarifies the relationship between the two interpretations of the Hall conductance with and without edges. The non-trivial contribution to the Hall conductance in the TKNN formula comes from the degeneracy of the edge states with the bulk state. We establish a direct connection between the Hall conductance in the bulk system and that due to the edge states.

One can say that the TKNN integer is a total vorticity in the magnetic Brillouin zone and that the vortex lines are given by the edge states. This edge state (vortex line) can be understood as an analogue of the Dirac strings in the discussion of the quantization of magnetic monopoles [2, 3]. (See figure 13(b).)

7. Topological aspects of the adiabatic heuristic argument: the fractional QHE from composite fermions

We have discussed several topological aspects of the integer QHE. All of the discussion can be extended to the fractional QHE by using Jain's construction of the fractional quantum Hall state (composite fermions). The main idea has been clearly summarized by Wilczek as an adiabatic heuristic principle. Finally, in this section, we will briefly discuss the adiabatic heuristic principle and Jain's construction. We discuss them for the lattice.

First we need to construct anyons on the lattice. This construction has been investigated by several groups [61–63]. Our preference is to view the anyon as a hard-core boson carrying a 'string' [63]. For an anyon obeying a θ-statistics, the phase change of the system acquires a phase factor $\mathrm{e}^{\mathrm{i}\theta}$ when two anyons are interchanged. We set the strength of the string as θ in this case. When another anyon cuts a string with a right-pointing arrow from below, for example, the system gains a phase $\mathrm{e}^{\mathrm{i}\theta}$. (See figure 14.)

As is easily checked from the figure, when anyon A goes around anyon B, anyon A cuts the string of anyon B and anyon B also cuts the string of anyon A, both in the same direction. This means that the process involving the anyon pair gives a phase change of $\mathrm{e}^{2\mathrm{i}\theta}$.

Now let us introduce a mean-field approximation for anyons, which is the key idea in the following discussion. Under this approximation, the phase change of the many-anyon

system under a certain process is replaced by the geometrical phase from the magnetic field ϕ_{av} (the flux per plaquette) of the hard-core boson.

When one anyon goes around the area S measured in units of the plaquette area, the phase change of the system is approximately given by

$$2\theta S\rho \tag{145}$$

where ρ is the anyon density and $S\rho$ is the average number of anyons inside area S. When the flux of the uniform magnetic field (the mean field) is ϕ_{av} per plaquette, the geometrical phase of the process is $2\pi\phi_{av}S$. If we try to replace the phase of the anyon system by that of the geometrical phase, we have $2\pi\phi_{av}S = 2\theta S\rho$. Therefore we have

$$\phi_{av}(\rho,\theta) = \frac{\theta}{\pi}\rho. \tag{146}$$

The original system of the anyons is mapped onto a system of hard-core bosons in a uniform magnetic field ϕ_{av}. The system in a uniform magnetic field is also represented by strings, by assigning a string to each plaquette. (See figure 15.) We restate the mean-field condition given as equation (146): *the total strength of strings going out from the total system should not be changed by the approximation*. In other words, the long-range behaviour of the vector potential should be the same after the approximation.

Now let us express the adiabatic heuristic principle [24]. When anyons with density ρ are in a real uniform magnetic field ϕ (per plaquette), the claim is that the low-energy physics of the system cannot change under an adiabatic process obeying

$$\Delta\phi_{\text{total}} = 0 \tag{147}$$

$$\phi_{\text{total}} = \phi + \phi_{av}. \tag{148}$$

From these expressions, one can relate the fractional QHE to the integer QHE [64, 22, 24]. Here let us discuss this for the lattice system. If equation (147) is integrated, it gives a series of adiabatically connected systems characterized by the integration constant. However, this adiabatic assumption has to be be checked; if the energy gap remains stable during the process, it can be justified adiabatically. When the integration constant is set to be zero, we have

$$\frac{\theta}{\pi} = -\frac{\phi}{\rho}. \tag{149}$$

This series includes important systems. It starts from $\theta = 0$ and $\phi \to 0$. It is a free hard-core boson. On taking the limits $\phi \to 0$ and $\rho \to 0$ keeping $\nu = \rho/\phi$ constant (see appendix A), we obtain

$$\frac{\theta}{\pi} = -\frac{1}{\nu}. \tag{150}$$

This is the complete filled Landau level $\nu = 1$ with $\theta = -\pi$ which is adiabatically connected to the free hard-core bosons. Since $\theta = -\pi$ simply means that the particles are the usual fermions, it is the integer quantum Hall state. This series also includes $\nu = 1/m$ with an odd integer m, $\theta = -m\pi$. Since the $\theta = -m\pi$ anyon is again just a usual fermion, this is the fractional quantum Hall state. The restriction to integer m arises from the fact that the particles are fermions ($m\pi \equiv \pi, \bmod 2\pi$). By using the arguments, the fractional QHE with the filling factor $1/m$ is adiabatically connected to the integer QHE (and also to the free hard-core boson). Here the fermion is represented by the hard-core boson with $m\pi$ flux tubes.

Acknowledgments

The author thanks A Zee, M Kohmoto, B I Halperin, Y S Wu, X G Wen, P A Lee, H Aoki, and T Ando for helpful discussions. He also thanks D Lidsky and Y Morita for careful reading of the manuscript. The work was supported in part by a Grant-in-Aid for Scientific Research from Ministry of Education, Science and Culture, Japan.

Appendix A. The recovery of the Landau level in the Hofstadter diagram

Here we comment on how the Landau level of the continuous theory is recovered from the lattice description.

Let us consider a system of $N = L_x \times L_y$ lattice sites with M electrons (the electron density is $\rho = M/N$) and assume that the flux per plaquette is $\phi = 1/q$ ($q \to \infty$). This corresponds to the weak-field limit in the Hofstadter diagram given as figure 1. For simplicity we also assume that L_x and L_y are multiples of q and sufficiently large. Then energy eigenvalues are grouped into q groups which converge to q energy bands when $L_{x,y} \to \infty$. Here let us take it that the band width of the jth energy band is B_j. The jth group has $N_\phi = N/q$ states and the energy separation is B_j around E_j ($j = 1, \ldots, q$). From figure 1, we can see that $E_j \approx C + \omega j\phi$ when $j \ll q$ for some constant C and ω. (The spectrum is linear in ϕ when $j \ll q$ and $\phi \approx 0$.) The continuum limit is recovered by taking the limits

$$\phi \to 0 \tag{A1}$$

$$\rho \to 0 \tag{A2}$$

and keeping

$$\nu = M/N_\phi = \rho/\phi = \text{constant} \tag{A3}$$

which is the filling factor of the Landau level in the continuum limit. The energy spectrum is scaled as $E_j/\phi = \text{constant} + \omega j$ where ω gives the Landau gap. Since $B_j/\phi = q B_j \to 0$ ($q \to \infty$), the band width goes to zero [65]. This gives the Landau degeneracy in the continuum limit.

Appendix B. The large gauge transformation

For simplicity, let us consider the Hamiltonian $H(\{\theta_m\})$ of a one-dimensional ring with N sites:

$$H(\{\theta_m\}) = \sum_{m=1}^{N-1} c_{m+1}^\dagger \mathrm{e}^{\mathrm{i}\theta_m} c_m + c_1^\dagger \mathrm{e}^{\mathrm{i}\theta_N} c_N + \text{HC.} \tag{B1}$$

Now let us consider a gauge transformation $\hat{\mathcal{G}}(\{\chi_m\})$ given by

$$c_m \to \bar{c}_m = \Omega_m c_m$$
$$\Omega_m = \mathrm{e}^{\mathrm{i}2\pi\chi_m}. \tag{B2}$$

Then the Hamiltonian is transformed as follows:

$$H(\{\theta_m\}) \to H(\{\bar{\theta}_m\}) \tag{B3}$$

where

$$\bar{\theta}_m \equiv \theta_m - 2\pi(\chi_{m+1} - \chi_m) \pmod{2\pi} \qquad (1 \leqslant m \leqslant N-1)$$
$$\bar{\theta}_N \equiv \theta_N - 2\pi(\chi_1 - \chi_N) \pmod{2\pi}.$$

Since we have $\sum_{m=1}^{N}\theta_m \equiv \sum_{m=1}^{N}\bar{\theta}_m \pmod{2\pi}$, let us set

$$2\pi n = \sum_{m=1}^{N}\theta_m - \sum_{m=1}^{N}\bar{\theta}_m \tag{B4}$$

for some integer n (the winding number). Therefore we can rewrite $\bar{\theta}_m$ as

$$\begin{aligned} \bar{\theta}_m &= \theta_m - 2\pi(\chi_{m+1}-\chi_m) \qquad (1 \leqslant m \leqslant N-1) \\ \bar{\theta}_N &= \theta_N - 2\pi(\chi_1-\chi_N) - 2\pi n \end{aligned} \tag{B5}$$

where $\bar{\theta}_m = \theta_m - 2\pi(\chi_{m+1}-\chi_m) - 2\pi l_m$ $(m = 1,\ldots,N)$ for some integers l_m, $(\sum_{m=1}^{N} l_m = n)$, and we define $\bar{\chi}_m = \chi_m + \sum_{k=1}^{m-1} l_k$.

The gauge transformation $\mathcal{G}(\{\chi_m\}, n)$ is characterized by $\{\chi_m\}$ and the winding number n. When $n = 0$, $\mathcal{G}(\{\chi_m\}, n = 0)$, which we call a 'local' or 'small' gauge transformation, is continuously connected to the identity operator. On the other hand, when n is not a multiple of N, one cannot connect the gauge transformation to the identity operator via a continuous change. This is the so-called a 'large' or 'global' gauge transformation. When $n = Ns$ for some integer s, the effect of the winding number n is absorbed by defining $\bar{\chi}_m = \chi_m + ms$. (Note that changing χ_m is defined mod 1.) Therefore $\mathcal{G}(\{\chi_m\}, n = Ns)$ is also a small gauge transformation. The origin of the global character is the non-trivial topological structure of the system (the system is a 'ring').

Note that since the gauge transformation $\mathcal{G}$ is just a unitary transformation, the Hamiltonian remains unchanged. Therefore the spectrum is the same. In general, however, an eigenstate is transformed into a different eigenstate by the large gauge transformation. (The set of all of the eigenstates is the same.) In contrast to the large gauge transformation, the small gauge transformation preserves the eigenstate, since each operator c_m remains unchanged.

Appendix C. Each energy gap has one edge state

We used the fact that each energy gap has one and only one edge state when the lattice size along the x-direction is commensurate with the flux per plaquette. In this appendix, we will give an idea of the proof. Let us consider two sets of wavefunctions $\{\Psi_n^{\mathrm{I,II}}\}$ generated by the transfer matrix (95) with initial conditions

$$\text{I: } \Psi_0^{\mathrm{I}} = 0,\ \Psi_1^{\mathrm{I}} = 1 \tag{C1}$$

$$\text{II: } \Psi_0^{\mathrm{II}} = 1,\ \Psi_1^{\mathrm{II}} = 0. \tag{C2}$$

From (96) and (97),

$$M_{11}(\epsilon) = \Psi_{q+1}^{\mathrm{I}}(\epsilon) \tag{C3}$$

$$M_{21}(\epsilon) = \Psi_{q}^{\mathrm{I}}(\epsilon) \tag{C4}$$

$$M_{12}(\epsilon) = \Psi_{q+1}^{\mathrm{II}}(\epsilon) \tag{C5}$$

$$M_{22}(\epsilon) = \Psi_{q}^{\mathrm{II}}(\epsilon). \tag{C6}$$

Next we write down a Schrödinger equation (92) with two different energies as

$$-\Psi_{n+1} - \Psi_{n-1} - a_n\Psi_n = \epsilon\Psi_n \tag{C7}$$

$$-\tilde{\Psi}_{n+1} - \tilde{\Psi}_{n-1} - a_n\tilde{\Psi}_n = \tilde{\epsilon}\tilde{\Psi}_n \tag{C8}$$

$$a_n = 2r\cos\left(k_y - 2\pi\frac{\Phi}{L_y} - 2\pi\phi n\right). \tag{C9}$$

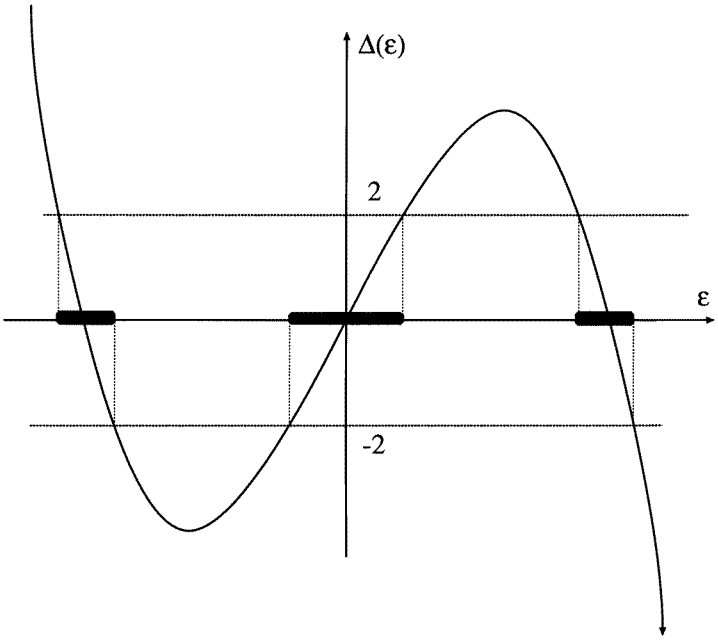


Figure A1. The behaviour of the function $\Delta(\epsilon)$. The thick lines denote the energy band regions.

Multiplying (C7) by $\tilde{\Psi}_n$ and (C8) by Ψ_n and taking the difference, we have

$$-\Psi_{q+1}\tilde{\Psi}_q + \Psi_1\tilde{\Psi}_0 + \tilde{\Psi}_{q+1}\Psi_q - \tilde{\Psi}_1\Psi_0 = (\epsilon - \tilde{\epsilon})\sum_{n=1}^{q}\Psi_n\tilde{\Psi}_n \tag{C10}$$

where summation over n is performed. Then by taking the limit $\tilde{\epsilon} \to \epsilon$ and setting Ψ_n and $\tilde{\Psi}_n$ to be one of the $\Psi_n^{\mathrm{I,II}}$s, we obtain the following four equations:

$$M_{11}M'_{21} - M_{21}M'_{11} = \sum_{n=1}^{q}|\Psi_n^{\mathrm{I}}|^2 \tag{C11}$$

$$M_{12}M'_{22} - M_{22}M'_{12} = \sum_{n=1}^{q}|\Psi_n^{\mathrm{II}}|^2 \tag{C12}$$

$$M_{11}M'_{22} - M_{21}M'_{12} = \sum_{n=1}^{q}\Psi_n^{\mathrm{I}}\Psi_n^{\mathrm{II}} \tag{C13}$$

$$M_{12}M'_{21} - M_{22}M'_{11} = \sum_{n=1}^{q}\Psi_n^{\mathrm{I}}\Psi_n^{\mathrm{II}} \tag{C14}$$

where we have used (C2), (C6) and $\det M = M_{11}M_{22} - M_{12}M_{21} = 1$. Finally calculating the sum (C11) $\times (-M_{12}) +$ (C12) $\times (M_{21}) +$ (C13) $\times (-M_{22}) +$ (C14) $\times (M_{11})$, we get

$$\frac{\mathrm{d}\Delta}{\mathrm{d}\epsilon} = -(M'_{11} + M'_{22}) = \sum_{n=1}^{q}(\Psi_n^{\mathrm{I}}, \Psi_n^{\mathrm{II}})\,\Xi\begin{pmatrix}\Psi_n^{\mathrm{I}}\\ \Psi_n^{\mathrm{II}}\end{pmatrix} \tag{C15}$$

$$\Xi = \begin{pmatrix} M_{12} & (M_{22} - M_{11})/2 \\ (M_{22} - M_{11})/2 & -M_{21}. \end{pmatrix}. \tag{C16}$$

By diagonalizing Ξ, we get

$$\frac{\mathrm{d}\Delta}{\mathrm{d}\epsilon} = \xi_1 \sum_{n=1}^{q} |\bar{\Psi}_n^{\mathrm{I}}|^2 + \xi_2 \sum_{n=1}^{q} |\bar{\Psi}_n^{\mathrm{II}}|^2 \tag{C17}$$

$$\xi_1 \xi_2 = \det \Xi = -\frac{\Delta^2 - 4}{4} \tag{C18}$$

$$\xi_1 + \xi_2 = \mathrm{Tr}\, \Xi = M_{12} - M_{21}. \tag{C19}$$

Also we mention the relation

$$M_{12} M_{21} = \frac{1}{4}(\Delta^2 - 4 - (M_{11} - M_{22})^2) \tag{C20}$$

($\Delta = M_{11} + M_{22}$). As discussed in the text, when the energy ϵ is in the energy band regions B, $\epsilon \in B$ satisfies $\Delta(\epsilon)^2 - 4 \leqslant 0$. Therefore in the region B, $\xi_1 \xi_2 > 0$ and sign $\xi_{1,2} = -\mathrm{sign}\ M_{21}$. Now we have

$$\mathrm{sgn}\ \frac{\mathrm{d}\Delta}{\mathrm{d}\epsilon} = -\mathrm{sgn}\ M_{21}. \tag{C21}$$

Since $\Delta(\epsilon)$ is a monotonic function in the energy band region B (counting the degree of ϵ in $\Delta(\epsilon)$), we see that M_{21} has to change its sign in the energy gap region. Counting the degree of $M_{21}(\epsilon)$, we see that M_{21} changes sign only once in each energy gap region. (See figure A1.) This means that each energy gap has one and only one edge state (i.e. the zero of M_{21}).

References

[1] See for example
Shapere A and Wilczek F (ed) 1989 *Geometric Phases in Physics* (Singapore: World Scientific)
[2] Dirac P A M 1931 *Proc. R. Soc.* A **133** 60
[3] Wu T T and Yang C N 1975 *Phys. Rev.* D **12** 3845
[4] Byers N and Yang C N 1961 *Phys. Rev. Lett.* **7** 46
[5] Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 485
[6] Berry M V 1984 *Proc. R. Soc.* A **392** 45
[7] von Klitzing K, Dorda G and Pepper M 1980 *Phys. Rev. Lett.* **45** 494
[8] Zak J 1964 *Phys. Rev.* **134** A1602
Zak J 1964 *Phys. Rev.* **134** A1607
[9] Laughlin R B 1981 *Phys. Rev.* B **23** 5632
[10] Halperin B I 1982 *Phys. Rev.* B **25** 2185
[11] Azbel M Ya 1964 *Zh. Eksp. Teor. Fiz.* **46** 929 (Engl. Transl. 1964 *Sov. Phys.–JETP* **19** 634)
[12] Hofstadter D R 1976 *Phys. Rev.* B **14** 2239
[13] Wannier G H 1978 *Phys. Status Solidi* b **88** 757
Wannier G H, Obermair G M and Ray R 1979 *Phys. Status Solidi* b **93** 337
[14] Thouless D J, Kohmoto M, Nightingale P and den Nijs M 1982 *Phys. Rev. Lett.* **49** 405
[15] Kohmoto M 1985 *Ann. Phys., NY* **160** 355
[16] Hatsugai Y 1993 *Phys. Rev.* B **48** 11 851
[17] Hatsugai Y 1993 *Phys. Rev. Lett.* **71** 3697
[18] Hatsugai Y 1996 *Quantum Coherence and Decoherence* ed K Fujikawa and Y A Ono (Amsterdam: North-Holland) p 167
[19] Wiegmann P B and Zabrodin A V 1994 *Phys. Rev. Lett.* **72** 1890
[20] Fadeev L D and Kashaev R M 1993 *University of Helsinki Report No* HU-TFT-93-63
[21] Hatsugai Y, Kohmoto M and Wu Y S 1994 *Phys. Rev. Lett.* **73** 1134
[22] Jain J K 1989 *Phys. Rev. Lett.* **63** 199
[23] M Stone (ed) 1992 *Quantum Hall Effect* (Singapore: World Scientific)
[24] Wilczek F (ed) 1990 *Fractional Statistics and Anyon Superconductivity* (Singapore: World Scientific)
[25] Hatsugai Y 1997 in preparation

[26] See
Haldane F D M 1987 *The Quantum Hall Effect* ed R E Prange and S M Girvin (Berlin: Springer)
[27] Wu Y-S, private communication
[28] Hiramoto H and Kohmoto M 1992 *Int. J. Mod. Phys.* B **6** 281
[29] Hatsugai Y, Kohmoto M and Wu Y-S 1996 *Phys. Rev.* B **53** 9697
[30] Kohmoto M 1989 *Phys. Rev.* B **39** 11 943
[31] Avron J, Seiler R and Simon B 1983 *Phys. Rev. Lett.* **51** 51
[32] Niu Q, Thouless D J and Wu Y S 1985 *Phys. Rev.* B **31** 3372
[33] Chang M-C and Niu Q 1996 *Phys. Rev.* B **53** 7010
[34] Douçot B and Stamp P C E 1991 *Phys. Rev. Lett.* **66** 2503
[35] Takahashi M, Hatsugai Y and Kohmoto M 1996 *J. Phys. Soc. Japan* **65** 529
[36] Rammal R, Toulouse G, Jaekel M T and Halperin B I 1983 *Phys. Rev.* B **27** 5142
[37] Schweitzer L, Kramer B and MacKinnon A 1984 *J. Phys. C: Solid State Phys.* **17** 4111
Johnston R, Kramer B, MacKinnon A and Schweitzer L 1986 *Surf. Sci.* **170** 256
[38] MacDonald A H 1984 *Phys. Rev.* B **29** 6563
[39] MacDonald A H and Středa P 1984 *Phys. Rev.* B **29** 1616
[40] Sun S N and Ralston J P 1991 *Phys. Rev.* B **44** 13 603
[41] Harper P G 1955 *Proc. Phys. Soc.* A **68** 874
[42] Toda M 1981 *Theory of Nonlinear Lattices* (Berlin: Springer) and references therein
[43] Date E and Tanaka S 1976 *Prog. Theor. Phys.* **55** 457
Date E and Tanaka S 1976 *Prog. Theor. Phys. Suppl.* **59** 107
[44] Mumford D 1983 *Tata Lectures on Theta I, II* (Boston: Birkhäuser)
[45] The theory of the non-linear lattice is applied to the problem of a one-dimensional quasi-crystal by Iguchi K 1992 *J. Math. Phys.* **33** 3938
[46] All roots of $M_{21}(\epsilon)$ are simple, as shown in appendix C.
[47] $\Psi_{qk+k'}, k' = 1, \ldots, q$, are also of the same order as Ψ_{qk+1} in general.
[48] Wen X G and Zee A 1989 *Nucl. Phys.* B **316** 641
[49] Hatsugai Y and Kohmoto M 1990 *Phys. Rev.* B **42** 8282
[50] These energies $\tilde{\lambda}_j$ are determined from $[\mathrm{Tr}\, M(\tilde{\lambda}_j)^l]^2 = 4$, which is similar to equation (110).
[51] When $\Psi_q(\mu_j) = 0$ for $\mu_j \in R^\alpha$, $\Psi_q(\mu_j) \neq 0$ for $\mu_j \in R^{-\alpha}$ in general, because $M_{21}(z)$ and the denominator of equation (121) vanish linearly at the point on this $R^{-\alpha}$ Riemann sheet [46]. In the following, we use the convention that μ_j denotes one of two $\mu_j \in R^\pm$ on the two Riemann sheets which gives the zero point of Ψ_q.
[52] Charge transport is also discussed in the following references:
Tešanović Z, Axel F and Halperin B I 1989 *Phys. Rev.* B **39** 8525
Kunz H 1986 *Phys. Rev. Lett.* **57** 1095
Thouless D J 1983 *Phys. Rev.* B **27** 6083
MacDonald A H 1983 *Phys. Rev.* B **28** 6713
[53] Widom A 1982 *Phys. Lett.* **90A** 474
[54] Středa P 1982 *J. Phys. C: Solid State Phys.* **15** L717
[55] These degenerate zero modes in the infinite system were investigated using topological methods by Wen and Zee [48] in detail.
[56] Wen X G, Wilczek F and Zee A 1989 *Phys. Rev.* B **39** 11 413
[57] Even if we include the next-nearest-neighbour hoppings, it is still possible to reduce the problem to a one-dimensional one [49] and to define the Riemann surface of the Bloch function.
[58] Hatsugai Y, Kohmoto M and Wu Y-S 1996 *Phys. Rev.* B **54** 4896
[59] Aoki H and Ando T 1981 *Solid State Commun.* **38** 1079
[60] Aoki H and Ando T 1986 *Phys. Rev. Lett.* **57** 3093
[61] Canright G S, Girvin S M and Brass A 1989 *Phys. Rev. Lett.* **63** 2291
[62] Wen X G, Dagotto E and Fradkin E 1990 *Phys. Rev.* B **42** 6110
[63] Hatsugai Y, Kohmoto M and Wu Y-S 1991 *Phys. Rev.* B **43** 2661
[64] Girvin S and MacDonald A H 1987 *Phys. Rev. Lett.* **58** 1252
[65] Thouless D J 1983 *Phys. Rev.* B **28** 4272