C. R. Physique 12 (2011) 323–331

Contents lists available at ScienceDirect

Comptes Rendus Physique

www.sciencedirect.com

Quantum Hall effect / Effet Hall quantique

Introduction to the theory of the integer quantum Hall effect

Introduction à la théorie de l'effet Hall quantique entier

Benoît Douçot

Laboratoire de physique théorique et hautes energies, CNRS et université Pierre-et-Marie-Curie, boîte 126, 4, place Jussieu, 75252 Paris cedex 05, France

ARTICLE INFO

Article history:
Available online 13 May 2011

Keywords:
Integer quantum Hall effect
Topological arguments
Adiabatic transport

Mots-clés :
Effet Hall quantique entier
Arguments topologiques
Transport adiabatique

ABSTRACT

We review the various theoretical arguments which strongly suggest that integer quantization of the Hall conductance holds exactly for a non-relativistic two-dimensional electron system in the presence of a static disorder and electron–electron interactions, when the chemical potential lies in a mobility gap of the system. The semi-classical picture, the Laughlin gauge-invariance argument and the connection between the Kubo Hall conductance and Chern numbers are presented, and the relevance of these idealized geometries to real systems is discussed.

RÉSUMÉ

Cet article présente les différents arguments théoriques qui suggèrent fortement que la quantification de la conductance de Hall est exacte, pour un système d'électrons bidimensionnel, en présence de désordre statique et d'interactions entre électrons, pourvu que le potentiel chimique se trouve dans un gap de mobilité. L'image semi-classique, l'argument d'invariance de jauge de Laughlin et la connection entre la conductance de Hall au sens de Kubo et un nombre de Chern sont successivement passés en revue. La pertinence de ces géométries idéalisées pour les situations plus réalistes est ensuite discutée.

1. Introduction

Constructing a rigorous theory of the integer quantum Hall effect, discovered by K. von Klitzing in 1980 [1], has remained a challenge for theoretical physicists and mathematicians for many years, although very early on, R. Laughlin [2], D. Thouless and others [3] proposed various *topological* arguments showing why the Hall conductance has to be an integer in units of the quantum conductance e^2/h. Although topological phenomena received already a lot of attention in condensed matter and statistical physics during the seventies, under the form of quantized vortices and solitons, and more generally, in the classification of defects in ordered systems, the integer Hall effect has certainly produced a deep cultural change because topology suddenly appeared as the most direct way to understand such a beautiful and mysterious phenomenon. Berry phases [4], quantum pumps [5], fractional statistics [6], topological quantum computation [7], topological insulators [8] illustrate the widespread diffusion of these topological concepts in many areas of contemporary physics. However, the theoretical status of the integer quantization has remained to some extent controversial, because these beautiful and striking arguments could not be regarded as totally rigorous from the mathematical side, and also because they use some idealized

E-mail address: doucot@lpthe.jussieu.fr.

1631-0705/$ – see front matter
doi:10.1016/j.crhy.2011.02.003

geometries such as a cylinder or even a torus, together with a rather artificial-looking averaging procedure over all possible boundary conditions. Part of the difficulty comes also from the crucial role played by static disorder in the presence of plateaus in the Hall conductance. A dramatic effect of disorder is to fill the gaps between Landau levels, which complicates the use of topological arguments. Fortunately, these states are almost all localized, excepted for a discrete set of energies which are in one-to-one correspondence with the unperturbed Landau levels. Localization plays indeed a key role in the rigorous proof of integer Hall quantization constructed by Bellissard, for an infinite non-interacting but disordered electronic system [9].

The purpose of this short review is to present these various arguments, with an emphasis on physical intuition rather than mathematical rigor. It seems most likely that integer Hall conductance quantization holds for any two-dimensional electron system in the presence of a static random potential and of electron–electron interactions, whenever the Fermi level lies in a mobility gap of the system. Unfortunately, a rigorous mathematical proof is still missing for the interacting case. But physicists have developed a consistent picture within which there is no doubt that integer quantization remains exact, even after interactions are included. As we already discussed, topological arguments play a central role in this picture.

This article is organized as follows. Section 2 discusses the limit of smooth disorder potentials, when the magnetic length is much smaller than the typical scale of the potential energy variations. The classical motion can be described as a fast cyclotron rotation around a guiding center which drifts slowly around the equipotential lines of the smooth potential. In a semi-classical picture cyclotron orbits are quantized, which gives rise to Landau levels. The slow motion of the guiding center generates a Hall conductance which is exactly e^2/h per completely filled Landau level. Unfortunately, this picture breaks down for sharp scatterers and also for interacting electrons, which calls for more general considerations. Section 3 exposes the famous Laughlin argument, and Section 4 explains the connection between the Kubo formula for the Hall conductance on a torus and Chern numbers. The relevance of these idealized geometries to real situations is the subject of Section 5.

2. Physics in a strong magnetic field: semi-classical picture

2.1. Hamiltonian description of single electron motion

If we choose a vector potential $\mathbf{A}(\mathbf{r})$ for the magnetic field component B perpendicular to the plane, $B = \partial_x A_y - \partial_y A_x$, the classical Hamiltonian for the motion of an electron is simply:

$$H_0 = \frac{1}{2m}(\mathbf{p} + e\mathbf{A})^2 = \frac{\boldsymbol{\pi}^2}{2m} \tag{1}$$

where $\mathbf{p} = m\dot{\mathbf{r}} - e\mathbf{A}$ is the momentum conjugated to $\mathbf{r}$, and the quantities $\mathbf{p}$ and $\mathbf{r}$ obey the Poisson brackets ($i, j \in \{x, y\}$):

$$\{p_i, p_j\} = 0, \qquad \{r_i, r_j\} = 0, \qquad \{p_i, r_j\} = \delta_{ij} \tag{2}$$

In the above Hamiltonian, we have introduced the gauge-invariant dynamical momentum $\boldsymbol{\pi} = (\pi_x, \pi_y)$:

$$\boldsymbol{\pi} = m\dot{\mathbf{r}} = \mathbf{p} + e\mathbf{A} \tag{3}$$

whose components obey the Poisson brackets:

$$\{\pi_i, \pi_j\} = \epsilon_{ij} eB, \qquad \{\pi_i, r_j\} = \delta_{ij} \tag{4}$$

where ϵ_{ij} is the antisymmetric tensor $\epsilon_{xy} = -\epsilon_{yx} = 1$.

Classically, the trajectories of the free Hamiltonian H_0 are circles, along which the electron runs with a constant angular velocity $\omega_c = eB/m$. Note that this frequency is independent of the orbit size, in close analogy to the case of a harmonic oscillator. This analogy results immediately from the fact that H_0 is quadratic in π_x and π_y, which happen to behave as canonically conjugate variables. It is then natural to introduce the guiding center of the orbit $\mathbf{R} = (R_x, R_y)$ by the equation:

$$\mathbf{v} = \frac{\boldsymbol{\pi}}{m} = \omega_c \hat{z} \wedge (\mathbf{r} - \mathbf{R}) \tag{5}$$

which implies that:

$$\mathbf{R} = \mathbf{r} + \frac{\hat{z} \wedge \boldsymbol{\pi}}{eB} \tag{6}$$

Explicitly, the two components of $\mathbf{R}$ are:

$$R_x = x - \frac{1}{eB}\pi_y, \qquad R_y = y + \frac{1}{eB}\pi_x \tag{7}$$

It is immediate to check that $\{\pi_i, R_j\} = 0$, so that the vector $\mathbf{R}$ is indeed a constant of motion. Another remarkable fact is that the two components of $\mathbf{R}$ do not mutually commute, and their Poisson bracket is given by:

$$\{R_i, R_j\} = -\epsilon_{ij}\frac{1}{eB} \tag{8}$$

Another useful interpretation of R_x and R_y is motivated by their explicit definitions (7) where we see that R_y (resp. $-R_x$) generates the product of translation along the x-axis (resp. the y-axis), and of a gauge transformation. An essential feature of the physics under a uniform magnetic field is that, as shown by Eq. (8), these two generators no longer commute, as it is the case in the absence of a magnetic field. This fact plays an essential role in quantum Hall physics, as we shall see shortly.

To understand the physical meaning of the guiding center it is instructive to consider the motion of a charged particle in presence of an external potential $V(\mathbf{r})$, which is supposed to vary slowly ($|\partial_i\partial_j V| \ll m\omega_c^2$ for all i, j):

$$H = H_0 - eV(\mathbf{r}) \tag{9}$$

We are interested in the motion of the guiding center in presence of $V(\mathbf{r})$. If the radius of the cyclotron orbit is sufficiently small and the speed of rotation sufficiently fast so that the potential seen during a rotation is approximatively constant, we can average over time which amounts to replace the position $\mathbf{r}$ by the guiding center $\mathbf{R}$. In this approximation, the guiding center motion is given by:

$$\dot{\mathbf{R}} = \frac{\mathbf{B}\wedge\nabla V}{B^2} \tag{10}$$

The motion decomposes into a fast rotation around the cyclotron orbit and a slow motion of the guiding center along the equipotential lines of $V(\mathbf{r})$.

If the potential is smoothly varying, we can divide the equipotential lines into two kinds: Those located near the extrema of V which are closed, and those located near its mean value, which can wind a long way through the saddle points of V. In presence of an electric field $\mathbf{E}$ the potential seen by the electrons becomes $V(\mathbf{R}) - e\mathbf{E}.\mathbf{R}$, and according to whether the equipotential line we consider is closed or extended, the electron traveling along it is localized or not. We shall return to this point later.

A very important consequence of this effective dynamics for guiding centers in the high magnetic field limit is that area is preserved under the time-evolution. This is a special case of Liouville's theorem on the conservation of phase-space volumes for Hamiltonian systems. But here, very remarkably, phase-space has to be identified with the physical plane, since the two coordinates of the guiding center are canonically conjugated according to (8). Physically, this means that if the initial condition is such that the electronic density is constant inside a domain Ω_0, and zero outside, after the system has evolved according to the dynamics (10), it is still constant inside a deformed domain Ω_t of the *same* area as Ω_0 and zero outside. In other words, the electronic fluid is incompressible. As we shall see in Section 2.2, this property plays a crucial role in understanding the quantization of σ_{xy}.

For the quantum system, we simply have to replace Poisson brackets by commutators, according to the usual rule: $\{X, Y\} \to \frac{\hbar}{i}[X, Y]$. The dynamical momenta and the guiding center define two sets of operators which obey the commutation relations analogous to Eqs. (4), (8):

$$[\pi_i, \pi_j] = -i\hbar\epsilon_{ij}eB, \qquad [R_i, R_j] = i\hbar\epsilon_{ij}\frac{1}{eB}, \qquad [\pi_i, R_j] = 0 \tag{11}$$

From this, we immediately find the energy spectrum of H_0, namely:

$$E_n = \hbar\omega_c\left(n + \frac{1}{2}\right) \tag{12}$$

where n can be any non-negative integer. Each energy branch with a given value of n is called a Landau level.

The fact that the guiding center coordinates R_x and R_y commute with H_0, but do not mutually commute implies that the energy spectrum is extremely degenerate. Because of the uncertainty principle $\Delta R_x \Delta R_y = \frac{\hbar}{eB}$, the physical plane can be thought of as divided into disjointed cells of area $2\pi l^2$ where the guiding center can be localized. The magnetic length $l = \sqrt{\frac{\hbar}{eB}}$ corresponds to the area threaded by one magnetic flux quantum $\Phi_0 = 2\pi\hbar/e$. The degeneracy per energy level and per unit area is therefore $1/2\pi l^2$ so that in an area Ω, the number of degenerate states is:

$$N_\Omega = \frac{\Omega}{2\pi l^2} \tag{13}$$

Imagine now that we continuously fill a bounded region of the plane with non-interacting electrons. Let n be the electron number density. We introduce the so-called filling factor as the number of electrons per cell:

$$\nu = n2\pi l^2 \tag{14}$$

Because of the Pauli principle, a cell can be occupied by one electron only per energy level. Therefore, each time the filling factor reaches an integer, an energy level gets filled and the next electron must be added to the next energy level. Thus, the energy per added electron (chemical potential) jumps by a quantity $\hbar\omega_c$. This is the signature of an incompressible state, which gives rise to the integer quantum Hall effect.

This very idealized model seems to indicate that the integer Hall effect should be observed only at the specific values of the magnetic field for which ν given by (14) is an integer. However, experimentally, the plateaus in the Hall conductance and the corresponding vanishing of the longitudinal conductance correspond to finite intervals of the transverse field B. To account for this discrepancy, it is essential to take into account the presence of a random potential resulting from the donor sites located in the vicinity of the 2D electron layer.

2.2. Semi-classical picture of electron dynamics

As suggested long ago by Prange and Joynt [10] and Trugman [11], it is very illuminating to consider the limit of an extremely strong magnetic field, so that the magnetic length $l = (\frac{\hbar}{eB})^{1/2}$ is much smaller than the typical length-scales associated to the spatial variations of the impurity potential $U_{\text{imp}}(\mathbf{r})$. Classically, we have seen that the guiding center $\mathbf{R}$ of classical orbits for a single electron obeys the following equations of motion:

$$\dot{\mathbf{R}} = \frac{\mathbf{B} \wedge \boldsymbol{\nabla}}{B^2}(V + U_{\text{imp}})(\mathbf{R}) \tag{15}$$

where $V(\mathbf{R})$ is the sum of the smooth confining potential and the potential induced by the (small) driving electric field. In particular, this implies that $W(\mathbf{R}) = (V + U_{\text{imp}})(\mathbf{R})$ is conserved, so the classical trajectories of guiding centers in the infinite B limit coincide with equipotential curves of the function $W(\mathbf{R})$. We expect that after quantization, single particle eigenstates are located along narrow strips of width l centered on these equipotential lines. As usual in semi-classical quantization, only a discrete set of classical orbits are allowed. An extension of the Bohr–Sommerfeld principle indicates that for closed classical orbits, only those which enclose an integer number of flux quanta give rise to quantum eigenstates. Let us denote by W_i the electrostatic potential values associated to these selected orbits. We get then the following semi-classical spectrum:

$$E_{i,n} = -eW_i + \hbar\omega_c\left(n + \frac{1}{2}\right) \tag{16}$$

where n is any non-negative integer corresponding to quantizing the fast cyclotron motion around the slow moving guiding center. For a fixed value of n, we may then speak of a generalized nth Landau level, although the degeneracy of this level is lifted by the joint effect of the driving electric field and the impurity potential. When such a level is completely filled, it induces a spatial density of electrons (after coarse-graining on a length-scale of the order of the corresponding cyclotron radius $n^{1/2}l$) equal to $\frac{eB}{h}$, mostly insensitive to the form of the effective potential $W(\mathbf{R})$.

Let us choose for our system the geometry of a horizontal strip. In a steady state close to equilibrium, the total electrostatic potential $W(\mathbf{R})$ is constant on both the upper and the lower edges, which will be assumed to be flat to simplify the notations. Specifically these edges correspond to $y = 0$ and $y = L_y$. The Hall voltage V_H is then equal to $W(x_0, L_y) - W(x_0, 0)$ for any vertical section of the wire corresponding to $x = x_0$. On average, we expect a global Hall current j_x in the horizontal direction. Each generalized Landau level produces a local current:

$$j_{x,n} = -e\left(\frac{eB}{h}\right)\theta\left(\mu + eW(\mathbf{R}) - \hbar\omega_c\left(n + \frac{1}{2}\right)\right)v_x(\mathbf{R}) \tag{17}$$

where $v_x(\mathbf{R}) = -\frac{1}{B}\frac{\partial W}{\partial y}(\mathbf{R})$, and μ is the chemical potential of the electronic system. The Heaviside step function $\theta(\mu + eW(\mathbf{R}) - \hbar\omega_c(n + \frac{1}{2}))$ is the limiting form of the Fermi–Dirac distribution for the semi-classical spectrum (16) at zero temperature. Let us now integrate this local current along a vertical section of the sample, at fixed $x = x_0$. This yields:

$$I_{x,n}(x_0) = \frac{e^2}{h}\int_0^{L_y} \mathrm{d}y\,\theta\left(\mu + eW(x_0, y) - \hbar\omega_c\left(n + \frac{1}{2}\right)\right)\frac{\partial W}{\partial y}(x_0, y) \tag{18}$$

The above integral is easily computed, and the result distinguishes between four cases:

1) $-eW(x_0, y) + \hbar\omega_c(n + \frac{1}{2}) \equiv E_n(x_0, y) > \mu$ for both $y = 0$ and $y = L_y$.
   This means that the nth generalized Landau level is unoccupied in the presence of the external driving voltage, in the limit where the impurity potential vanishes. In this case, $I_{x,n}(x_0) = 0$. Note that this value is independent of the strength of the local impurity potential. In particular, if $U_{\text{imp}}(\mathbf{R})$ has deep local minima, it may happen that $E_n(\mathbf{R}) < \mu$ in some finite areas, meaning that there are occupied bound states in the nth Landau level localized near impurities. But in this very large field limit, we see that such localized states do not contribute to the global Hall current.
2) $E_n(x_0, y) < \mu$ for both $y = 0$ and $y = L_y$.
   In the limit of vanishing impurity potential, the corresponding Landau level is then fully occupied. We obtain:

$$I_{x,n}(x_0) = \frac{e^2}{h}\big(W(x_0, L_y) - W(x_0, 0)\big) = \frac{e^2}{h}V_H \tag{19}$$

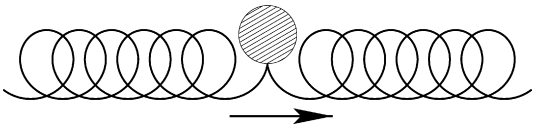

Fig. 1. Illustration of cyclotron motion for a classical charged particle in the presence of uniform perpendicular magnetic and electric fields. A hard circular point scatterer is depicted as a dashed–filled circle.

Again, this result is independent of the strength of the impurity potential. Such a fully occupied level provides therefore a contribution equal to $\frac{e^2}{h}$ to the total Hall conductance.

3) $E_n(x_0, 0) > \mu$ and $E_n(x_0, L_y) < \mu$.
Then:

$$I_{x,n}(x_0) = \frac{e^2}{h} \frac{\mu - E_n(x_0, L_y)}{e} \tag{20}$$

4) $E_n(x_0, 0) < \mu$ and $E_n(x_0, L_y) > \mu$.
Then:

$$I_{x,n}(x_0) = \frac{e^2}{h} \frac{E_n(x_0, 0) - \mu}{e} \tag{21}$$

These last two cases correspond to Landau levels which are partially filled in the absence of impurity potential, but in the presence of the driving field. They destroy the quantization of σ_{xy}. In order to avoid them, one has to fix the chemical potential in a gap of the unperturbed Landau level spectrum, and impose a weak enough driving electric field, typically such that $e|V(x_0, L_y) - V(x_0, 0)| < \hbar\omega_c$. If these conditions are satisfied, we have an integer number p of filled Landau levels which contribute to the Hall current, so that:

$$I_x(x_0) = p \frac{e^2}{h} V_H \tag{22}$$

in perfect agreement with:

$$\sigma_{xy} = p \frac{e^2}{h} \tag{23}$$

When does this simple and appealing picture break down? Clearly, it is problematic when the typical scale of the impurity potential becomes comparable to the magnetic length l. In this case, it is no longer possible to preserve such a simple description of quantum energy eigenstates. Nevertheless, as shown in Fig. 1, we may expect that a strong scatterer (modeled as an impenetrable disk of radius a) does not disturb the shape of a strip-like eigenstate excepted in its immediate vicinity, even when a is small compared to the cyclotron radius. In particular, the overall direction of propagation of a wave-packet is not modified by the presence of such impurities. These qualitative expectations are confirmed by more detailed perturbative calculations [10].

More serious problems arise when strong localized scatterers are densely packed, namely with an average nearest-neighbor spacing of the order of l. In this case, classical trajectories become very complicated. Quantum-mechanically, we expect that such a strong potential induces strong mixing between different Landau levels, and therefore, a perturbative analysis in the disorder is not very helpful. The key physical concept at this point is Anderson localization [12]. In two dimensions (and in the absence of a magnetic field), it is very likely that all energy eigenstates are spatially localized [13]. However, there are theoretical arguments [14,15] and substantial numerical evidences [16] that in the presence of a uniform magnetic field, some delocalized eigenstates exist for a discrete set of energies, in one-to-one correspondence with the original Landau levels. These states can be viewed as the continuation to strong disorder of the semi-classical orbits which percolate throughout the whole sample [11]. These are the states which are responsible for the Hall conductance. A review on these magnetic-field induced delocalization transitions may be found in a paper by Kramer et al. [17]. The main problem is now to show that the quantization of σ_{xy} still holds, even when the disorder cannot be considered as weak. A key insight explaining why this occurs, has been provided by Laughlin [2]. Its power stems from its non-perturbative nature. Indeed, it relies only on the gauge-invariance of quantum mechanics!

3. The Laughlin argument

Let us consider the same strip as before, defined by $0 \leqslant y \leqslant L_y$, but let us fold it into a cylinder by identifying points (x, y) and $(x + L_x, y)$. The magnetic field $\mathbf{B}$ is still normal to this finite domain, and a driving electric field is still applied along the y-direction. To evaluate the current $j_x(\mathbf{r})$ in the quantum mechanical ground-state of this system, Laughlin uses the following exact relation:

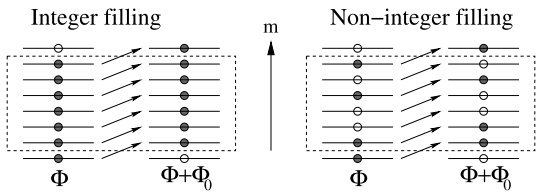

Fig. 2. Illustration of the change in the spatial distribution of occupied energy states within a single Landau level, as flux Φ is changed into $\Phi + \Phi_0$. For integer filling, no change occurs in the bulk of the system (depicted by the dashed parallelogram), and the global effect is to transfer one electron from the lower to the upper boundary. For non-integer filling factor, this adiabatic process also implies a change of the level occupancy pattern in the bulk of the system.

$$j_x(\mathbf{r}) = -\frac{\partial \langle H \rangle}{\partial A_x(\mathbf{r})} \tag{24}$$

where $\mathbf{A}(\mathbf{r})$ is the external magnetic vector potential. Let us now impose spacial variations $\delta\mathbf{A}(\mathbf{r})$ of the form: $\delta A_x(\mathbf{r}) = \frac{\delta\Phi}{L_x}$ and $\delta A_y(\mathbf{r}) = 0$. Such variations do *not* modify the gauge-invariant electric and magnetic fields, but they introduce an Aharonov–Bohm flux through any closed path winding once around the cylinder in the positive x-direction. The corresponding infinitesimal variation of the system average energy is:

$$\delta\langle H \rangle = -\frac{\delta\Phi}{L_x} \int_0^{L_x} \mathrm{d}x \int_0^{L_y} \mathrm{d}y \; j_x(\mathbf{r}) = -\delta\Phi \, I_x \tag{25}$$

So the Hall current I_x is simply expressed as:

$$I_x = -\frac{\mathrm{d}\langle H \rangle}{\mathrm{d}\Phi} \tag{26}$$

Now Laughlin assumes that as Φ varies, the ground-state wave-function $|\Psi_0(\Phi)\rangle$ undergoes a smooth evolution, Φ being considered as an external parameter of the system Hamiltonian. A sufficient condition for this to occur is when the ground-state is unique, and well separated by a finite energy gap from excited states created in the *bulk* of the system. This happens for instance for non-interacting electrons with an integer filling factor in the limit of a weak impurity potential.

Let us now vary Φ by a *finite* quantity $\Phi_0 = h/e$. Note that the effect of changing Φ is simply the same as changing the periodic boundary condition along the x-direction, so its effect on a macroscopic system is expected to be small. One way to see this is to consider single particle eigenstates in the Landau gauge ($A_x = -By$, $A_y = 0$). These states are localized in narrow strips centered around horizontal lines such that $y_m = \frac{2\pi l^2}{L_x}(m + \Phi/\Phi_0)$, m integer. So changing Φ into $\Phi + \Phi_0$ amounts simply to changing m into $m + 1$ and the single electron spectrum is invariant in this operation. This in fact expresses the gauge-invariance of quantum-mechanics, as first emphasized by Aharonov and Bohm. In particular, this periodicity of the spectrum as a function of Φ with period Φ_0 holds also for interacting electron systems. Denoting by $\Delta\langle H \rangle$ the variation of the system energy during such process, Laughlin assumes that we may still write:

$$I_x = -\frac{\Delta\langle H \rangle}{\Phi_0} \tag{27}$$

The replacement of a derivative with respect to flux by a finite difference amounts to averaging the Hall conductance over all possible boundary conditions around the cylinder. As shown by Niu and Thouless [18], this is expected to induce only an exponentially small error in the ratios L_x/l, L_y/l. We shall return to this point in Section 5 below.

Suppose now that the chemical potential is such that the ground-state is well separated from excited states by an energy gap, at least when the driving electric field vanishes. Again, this is the case for non-interacting electrons with an integer filling factor in the limit of weak impurity potential. Then upon changing Φ into $\Phi + \Phi_0$, we cannot modify the wave-function in the bulk of the system. However, as the example of non-interacting electrons suggests (see Fig. 2), we may still transfer an integer number p of electrons (since the quantum number m is shifted into $m + 1$) from the lower edge to the upper edge. More precisely, for non-interacting electrons with an integer filling factor ν, then $p = \nu$. In this situation, the energy variation $\Delta\langle H \rangle$ during the shift from Φ to $\Phi + \Phi_0$ is of purely electrostatic origin, so that:

$$\Delta\langle H \rangle = -peV_H \tag{28}$$

From Eq. (27), this yields:

$$I_x = p\frac{e^2}{h}V_H \tag{29}$$

or equivalently, $\sigma_{xy} = p\frac{e^2}{h}$.

Note that disorder and interactions are both allowed in the Laughlin argument. However, the use of an adiabatic approximation seems to require a finite spectral gap between the ground-state and any excited state of the many-electron system. But such a true spectral gap is likely to disappear in the presence of a large disorder. In fact, only extended states are expected to contribute to the charge flow induced by the adiabatic change in the Aharonov–Bohm flux. Therefore, it seems that this argument remains valid under the weaker assumption that the Fermi level lies in a mobility gap of the spectrum.

4. The Kubo approach and Chern numbers

Let us further modify the geometry used for the Laughlin argument by gluing together the lower ($y = 0$) and upper ($y = L_y$) edges of the cylinder, thus forming a torus. The small external driving field is still uniform, directed along $\hat{\mathbf{y}}$, and we shall still measure the current density $\langle j_x \rangle$ along the $\hat{\mathbf{x}}$-direction. As we wish to apply linear response theory, it is convenient to work with a time-dependent electric field:

$$E_y(t) = \int \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega t} \tilde{E}_y(\omega) \tag{30}$$

The frequency-dependent Hall conductivity $\sigma_{xy}(\omega)$ is defined by:

$$\tilde{j}_x(\omega) = \sigma_{xy}(\omega) \tilde{E}_y(\omega) \tag{31}$$

Quantum-mechanically, the simplest way to introduce an electric field is through a time-dependent vector-potential $\delta\mathbf{A}(t)$ such that $E_y(t) = -\frac{\partial \delta A_y(t)}{\partial t}$, or equivalently: $\tilde{E}_y(\omega) = i\omega \delta \tilde{A}_y(\omega)$. If we define $K_{xy}(\omega)$ to be such that $\tilde{j}_x(\omega) = K_{xy}(\omega)\delta\tilde{A}_y(\omega)$, then: $\sigma_{xy}(\omega) = \frac{K_{xy}(\omega)}{i\omega}$. For a uniform electric field, δA_y is also spatially uniform. In real space and time, the response kernel K_{xy} is given by the standard Kubo linear response formula:

$$K_{xy}(\mathbf{r}, t; t') = \frac{i}{\hbar} \left\langle \left[\frac{\delta H}{\delta A_x(\mathbf{r})}, \int \frac{\delta H}{\delta A_y(\mathbf{r}')} \, \mathrm{d}^2\mathbf{r}' \right] \right\rangle \tag{32}$$

where we have used again $j_x(\mathbf{r}) = -\frac{\delta H}{\delta A_x(\mathbf{r})}$, and the fact that δA_y is uniform. The quantum-mechanical expectation values are taken in the ground-state of the system, since we are assuming a very low temperature. In the absence of impurities, we expect a uniform current, but if impurities are present, only the total current $I_x(x) = \int_0^{L_y} \mathrm{d}y \, j_x(x, y)$ is independent of x (because of current conservation) in the static limit. It is therefore natural to average the above response function over the "probe" position $\mathbf{r}$. Introducing fluxes Φ_x and Φ_y as in the previous section, but now along the two main directions of the torus, we may write this space-averaged response function as:

$$K_{xy}(\mathbf{r}, t; t') = \frac{i}{\hbar} \frac{1}{L_x L_y} \left\langle \left[\int \frac{\delta H}{\delta A_x(\mathbf{r})} \, \mathrm{d}^2\,\mathbf{r}, \int \frac{\delta H}{\delta A_y(\mathbf{r}')} \, \mathrm{d}^2\,\mathbf{r}' \right] \right\rangle = \frac{i}{\hbar} \left\langle \left[\frac{\partial H}{\partial \Phi_x}, \frac{\partial H}{\partial \Phi_y} \right] \right\rangle \tag{33}$$

Transforming to Fourier-space, we now obtain:

$$\sigma_{xy}(\omega) = \frac{i}{\hbar\omega} \sum_{\alpha} \left\{ \frac{\langle 0|\partial_x H|\alpha\rangle\langle\alpha|\partial_y H|0\rangle}{\omega - \omega_{\alpha 0}} - \frac{\langle 0|\partial_y H|\alpha\rangle\langle\alpha|\partial_x H|0\rangle}{\omega + \omega_{\alpha 0}} \right\} \tag{34}$$

where $|0\rangle$ is the ground-state and $|\alpha\rangle$ denotes a complete orthonormal basis of energy eigenstates of H, with energies E_α. The Bohr frequencies $\omega_{\alpha 0}$ are equal to $(E_\alpha - E_0)/\hbar$. In this expression, H is the *full* Hamiltonian of the system in the absence of driving electric field. It may therefore include both impurity potentials and interaction effects. To simplify notations, $\partial_x H$ and $\partial_x H$ stand respectively for $\frac{\partial H}{\partial \Phi_x}$ and $\frac{\partial H}{\partial \Phi_y}$. Gauge-invariance requires that the current vanishes when a static uniform vector potential is applied. This enables us to replace the above expression by:

$$\sigma_{xy}(\omega) = \frac{i}{\hbar} \sum_{\alpha} \left\{ \frac{\langle 0|\partial_x H|\alpha\rangle\langle\alpha|\partial_y H|0\rangle}{\omega_{\alpha 0}(\omega - \omega_{\alpha 0})} + \frac{\langle 0|\partial_y H|\alpha\rangle\langle\alpha|\partial_x H|0\rangle}{\omega_{\alpha 0}(\omega + \omega_{\alpha 0})} \right\} \tag{35}$$

which has a well-defined static limit $\omega \to 0$ provided the system has a finite energy gap, so that denominators are not vanishing in this limit. We may then write the static Hall conductance as:

$$\sigma_{xy} = \frac{\hbar}{i} \sum_{\alpha} \left\{ \frac{\langle 0|\partial_x H|\alpha\rangle\langle\alpha|\partial_y H|0\rangle}{(E_0 - E_\alpha)^2} - \frac{\langle 0|\partial_y H|\alpha\rangle\langle\alpha|\partial_x H|0\rangle}{(E_0 - E_\alpha)^2} \right\} \tag{36}$$

It is now convenient to view the Aharonov–Bohm fluxes Φ_x and Φ_y as external parameters. The ground-state $|0\rangle$ becomes a function of $(\Phi_x, \Phi_y) \equiv \boldsymbol{\Phi}$ and we shall denote it by $|\boldsymbol{\Phi}\rangle$. This allows us to recast the previous equation as:

$$\sigma_{xy}(\boldsymbol{\Phi}) = \frac{\hbar}{i} \left(\frac{\partial\langle\boldsymbol{\Phi}|}{\partial\Phi_x} \frac{\partial|\boldsymbol{\Phi}\rangle}{\partial\Phi_y} - \frac{\partial\langle\boldsymbol{\Phi}|}{\partial\Phi_y} \frac{\partial|\boldsymbol{\Phi}\rangle}{\partial\Phi_x} \right). \tag{37}$$

This may be regarded as the curl of a two-dimensional vector:

$$\frac{1}{i}\left(\langle \boldsymbol{\Phi}|\frac{\partial|\boldsymbol{\Phi}\rangle}{\partial \Phi_x}, \langle \boldsymbol{\Phi}|\frac{\partial|\boldsymbol{\Phi}\rangle}{\partial \Phi_y}\right) \tag{38}$$

Since $\langle \boldsymbol{\Phi}|\boldsymbol{\Phi}\rangle = 1$, this vector has purely real components. The above expression depends on a choice of two Aharonov–Bohm fluxes (Φ_x, Φ_y), which by gauge transformations is equivalent to choosing the following boundary conditions for the wave-functions (for simplicity of notation, we consider just one electron here, since generalization to N electrons is obvious):

$$\Psi(x + L_x, y) = e^{i2\pi \frac{\Phi_x}{\Phi_0}} \Psi(x, y) \tag{39}$$

$$\Psi(x, y + L_y) = e^{i2\pi \frac{\Phi_y}{\Phi_0}} \Psi(x, y) \tag{40}$$

This is of course connected to the idea that all physical quantities, like $\sigma_{xy}(\boldsymbol{\Phi})$ are periodic functions of both Φ_x and Φ_y with period Φ_0. So the $\boldsymbol{\Phi}$-plane may be folded onto a two-dimensional torus.

Let us now make the assumption that $\sigma_{xy}(\boldsymbol{\Phi})$ is only very weakly modified upon changing these boundary conditions. In the case where the ground-state is well separated from excited states by a finite energy gap, arguments have been given to show that $\sigma_{xy}(\boldsymbol{\Phi})$ becomes constant for a large system, up to corrections which are exponentially small in L_x/l, L_y/l [18]. This point will be further discussed in Section 5 below. We may therefore replace $\sigma_{xy}(\boldsymbol{\Phi})$ by its average over the $\boldsymbol{\Phi}$-torus and then transform the two-dimensional integral of a curl into a line-integral along the boundary of the square $[0, \Phi_0] \times [0, \Phi_0]$:

$$\sigma_{xy} = \frac{e^2}{h}\frac{1}{2\pi i}\int_{\square} \langle \boldsymbol{\Phi}|\,\mathrm{d}|\boldsymbol{\Phi}\rangle \tag{41}$$

But as can easily be shown, the quantity $\frac{1}{2\pi i}\int_{\square}\langle \boldsymbol{\Phi}|\,d|\boldsymbol{\Phi}\rangle$ is equal to $2\pi n_{\mathrm{C}}$ where n_{C} is an integer called a Chern number. Finally, we get [19,20]:

$$\sigma_{xy} = n_{\mathrm{C}}\frac{e^2}{h} \tag{42}$$

As for the Laughlin argument, the main advantage of this line of approach is that it can be applied to disordered and interacting systems as well, under the assumption of a finite spectral gap. The weaker assumption of a mobility gap should be sufficient to describe the Hall conductance as a Chern number. Indeed, at least for non-interacting electrons in the presence of disorder, it has been shown that only extended states can generate a non-zero Chern number [21].

5. Accuracy of the Hall conductivity

To which extent are these beautiful arguments on idealized systems relevant to real situations, which involve finite systems with boundaries and contacts with electronic reservoirs? In fact, this question subdivides at least into two issues. We have emphasized that, even for the Laughlin argument and for the torus geometry, some kind of averaging over a continuous family of boundary conditions is involved, in order to get a quantized Hall conductance. This is obvious in the case of the torus, but already on the cylinder considered by Laughlin, we had to replace, in the evaluation of the transverse current, the derivative of the total energy with respect to flux by a finite difference, see Eq. (27). So the main issue here is to show that a change in boundary condition, far away from the region where the Hall voltage is measured, does not change the Hall conductance. The second issue is the effect of edges. As emphasized by many authors [22–24], quenching of the electronic kinetic energy under the form of the fast cyclotron motion allows for propagating states along the edges of a finite system, which are weakly affected by disorder, because scattering occurs mostly in the forward direction, as suggested already in the semi-classical approach (see Fig. 1). The key physical features to address both issues is the mobility gap for states in the bulk, together with the chiral nature of the propagating states along edges. As shown by Niu and Thouless [18], these two properties imply that the Hall conductivity at a given point in the system depends only on its neighborhood, over a length-scale of the order of the localization length ξ. Any modification affecting the system at a distance $L \gg \xi$ away from this point produces only an exponentially small correction (as a function of L/ξ) to the local Hall conductivity. This explains the insensibility of the macroscopic Hall conductance to both the sample geometry and the choice of boundary conditions.

At this point, we should emphasize that there is a mathematically rigorous proof of the Hall conductance integer quantization for a non-interacting system, in the presence of an arbitrary disorder, whose statistical distribution is translation-invariant [9]. This proof deals with an infinite system, within the framework of the Kubo linear response theory. Using extensively ideas from non-commutative geometry, Bellissard succeeded to relate the Kubo conductance to an integer, whose meaning has been clarified by Avron et al. [25], in a way which establishes a clear connection with the Laughlin argument. This number is defined as the relative index between two one particle operators: the projectors onto all occupied single particle levels for the system in the presence and in the absence of a singular flux singularity which carries one flux quantum. The index can be interpreted as the number of electrons which are injected from infinity towards the flux tube as the flux is changed adiabatically from 0 to Φ_0. Because the total number of electrons inside the Fermi sea is infinite for an infinite

system, some sophisticated mathematical apparatus is needed in order to define precisely its finite variation. But this is the price to pay to get rid of the somehow artificial cylindrical geometry. Unfortunately, extensions of these mathematically rigorous proofs of integer quantization to interacting electron systems have not been achieved so far.

6. Conclusion

Within the current theoretical understanding briefly reviewed here, it seems that the integer Hall conductance quantization is a very robust phenomenon. Powerful general arguments suggest that it holds for any two-dimensional electron system in the presence of a static random potential and of electron–electron interactions, whenever the Fermi level lies in a mobility gap of the system. We have seen that going from the ideal geometries involved in the Laughlin argument or in the rewriting of the Kubo linear response formula as a Chern number, to realistic geometries with open boundaries is likely to bring corrections to the Hall conductance which are exponentially small in the ratio of the system size to the localization length. As suggested recently [26], the only way to produce a finite correction to the quantized Hall conductance in the thermodynamic limit is to couple the relativistic electronic system to a quantized electromagnetic field. Penin has shown that the presence of the static magnetic field modifies the amplitude of virtual excitation for electron–positron pairs (vacuum polarization), and therefore, the effective coupling between electrons and the transverse electric field which induces the Hall current. These effects are however extremely small, because they involve the quantity $(\Phi_c/\Phi_0)^2$, where $\Phi_0 = h/e$ is the flux quantum, and Φ_c is the external magnetic flux on a loop whose linear size is the Compton wave-length of the electron. This corresponds to a magnetic field scale of the order of 4.41×10^9 T, which is huge compared to the experimental situations. The relative correction to the Hall conductance induced by vacuum polarization is then of the order of 10^{-20}, which is far below the present accuracy of the measurements.

References

[1] K.v. Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. 45 (1980) 494.
[2] R.B. Laughlin, Phys. Rev. B 23 (1981) 5632.
[3] D.J. Thouless, M. Kohmoto, M.P. Nightingale, M. den Nijs, Phys. Rev. Lett. 49 (1982) 405.
[4] M.V. Berry, Proc. R. Soc. Lond. Ser. A 392 (1984) 45.
[5] Q. Niu, Phys. Rev. Lett. 64 (1990) 1812.
[6] D. Arovas, J.R. Schrieffer, F. Wilczek, Phys. Rev. Lett. 53 (1984) 722.
[7] A. Kitaev, Ann. Phys. 321 (2006) 2.
[8] Liang Fu, C.L. Kane, E.J. Mele, Phys. Rev. Lett. 98 (2007) 106803.
[9] J. Bellissard, A. van Elst, H. Schulz-Baldes, J. Math. Phys. 35 (1994) 5373.
[10] R.E. Prange, R. Joynt, Phys. Rev. B 25 (2002) 2943.
[11] S.A. Trugman, Phys. Rev. B 27 (2003) 7539.
[12] P.W. Anderson, Phys. Rev. 109 (1958) 1492.
[13] E. Abrahams, P.W. Anderson, D.C. Licciardello, T.V. Ramakrishnan, Phys. Rev. Lett. 42 (1979) 673.
[14] D.E. Khmelnitskii, Pis'ma Zh. Eksp. Teor. Fiz. 38 (1983) 454;
D.E. Khmelnitskii, JETP Lett. 38 (1983) 552.
[15] H. Levine, S.B. Libby, A.M.M. Pruisken, Phys. Rev. Lett. 51 (1983) 1915.
[16] B. Huckestein, Rev. Modern Phys. 67 (1995) 357.
[17] B. Kramer, T. Ohtsuki, S. Kettemann, Phys. Rep. 417 (2005) 211.
[18] Q. Niu, D.J. Thouless, Phys. Rev. B 35 (1987) 2188.
[19] Q. Niu, D.J. Thouless, Y.-S. Wu, Phys. Rev. B 31 (1985) 3372.
[20] J.E. Avron, R. Seiler, Phys. Rev. Lett. 54 (1985) 259.
[21] D.P. Arovas, R.N. Bhatt, F.D.M. Haldane, P.B. Littlewood, R. Rammal, Phys. Rev. Lett. 60 (1988) 619.
[22] B.I. Halperin, Phys. Rev. B 25 (1982) 2185.
[23] M. Büttiker, Phys. Rev. B 38 (1988) 9375.
[24] X.G. Wen, Internat. J. Modern Phys. B 6 (1992) 1711.
[25] J.E. Avron, R. Seiler, B. Simon, Comm. Math. Phys. 159 (1994) 399.
[26] A.A. Penin, Phys. Rev. B 79 (2009) 113303.