Nobel Lecture: Fractional quantization*

R. B. Laughlin

Department of Physics, Stanford University, Stanford, California 94305

One of my favorite times in the academic year occurs in early spring when I give my class of extremely bright graduate students, who have mastered quantum mechanics but are otherwise unsuspecting and innocent, a take-home exam in which they are asked to deduce superfluidity from first principles. There is no doubt a special place in hell being reserved for me at this very moment for this mean trick, for the task is impossible. Superfluidity, like the fractional quantum Hall effect, is an emergent phenomenon—a low-energy collective effect of huge numbers of particles that cannot be deduced from the microscopic equations of motion in a rigorous way and that disappears completely when the system is taken apart (Anderson, 1972). There are prototypes for superfluids, of course, and students who memorize them have taken the first step down the long road to understanding the phenomenon, but these are all approximate and in the end not deductive at all, but fits to experiment. The students feel betrayed and hurt by this experience because they have been trained to think in reductionist terms and thus to believe that everything not amenable to such thinking is unimportant. But nature is much more heartless than I am, and those students who stay in physics long enough to seriously confront the experimental record eventually come to understand that the reductionist idea is wrong a great deal of the time, and perhaps always. One common response in the early stages of learning is that superconductivity and the quantum Hall effect are not fundamental and therefore not worth taking seriously. When this happens I just open up the AIP Handbook and show the disbeliever that the accepted values of e and h are defined by these effects, and that ends that. The world is full of things for which one's understanding, i.e., one's ability to predict what will happen in an experiment, is degraded by taking the system apart, including most delightfully the standard model of elementary particles itself. I myself have come to suspect most of the important outstanding problems in physics are emergent in nature, including particularly quantum gravity.

One of the things an emergent phenomenon can do is create new particles. When a large number of atoms condense into a crystal, the phonon, the elementary quantum of sound, becomes a perfectly legitimate particle at low energy scales. It propagates freely, does not decay, carries momentum and energy related to wavelength and frequency in the usual way, interacts by simple rules that may be verified experimentally, mediates the attractive interaction responsible for conventional superconductivity, and so forth, and none of these things depends in detail on the underlying equations of motion. They are generic properties of the crystalline state. The phonon ceases to have meaning when the crystal is taken apart, however, because sound makes no sense in an isolated atom. A somewhat more esoteric example, although a more apt one, is the Landau quasiparticle of a metal (Pines and Nozières, 1966). This is an excited quantum state that behaves like an extra electron added to a cold Fermi sea, but which is actually a complex motion of all the electrons in the metal. It is not possible to deduce the existence of quasiparticles from first principles. They exist instead as a generic feature of the metallic state and cease to exist if the state does. This problem is not limited to solids. Even the humble electron, the most elementary particle imaginable, carries a polarization of the Dirac sea with it as it travels from place to place and is thus itself a complex motion of all the electrons in the sea. In quantum physics there is no logical way to distinguish a real particle from an excited state of the system that behaves like one. We therefore use the same word for both.

Whenever one is confronted with unpredictable emergent phenomena there is need for a sound definition of sameness of two states of matter. The one most of us prefer is the existence of a reversible adiabatic map. One imagines slowly changing the underlying equations of motion, checking at each stage to make sure the ground state and low-lying excitations have evolved in a one-to-one way. There is actually no need to check this if the system contains only a small number of particles, for the mapping is then guaranteed to be one-to-one by virtue of the adiabatic principle. But if the system contains a thermodynamically large number of particles it can happen that a small change to the equations of motion results in a violent rearrangement of the ground state and low-lying excitations and a corresponding breakdown of the one-to-one mapping. This is a quantum phase transition. We say that two states are the same phase of matter if they can be slowly transformed into each other without encountering a quantum phase transition, and different phases of matter if they cannot. By this definition metals, insulators, and superconductors are all different phases of matter, but two metals with slightly different electron-electron repulsion strengths are the same. This definition of sameness is one of the most powerful ideas in physics, for it relieves us of the need to compute the properties of complex systems from first principles to understand them. Instead

we can just find a prototype that is easy to solve and then map our solution backward adiabatically.

SOLITONS

The idea that particles carrying parts of an elementary quantum number might occur as an emergent phenomenon is not new. Already in the late 1970s there was a serious theory literature on this subject stemming from a key paper of Jackiw and Rebbi (1976), but more generally from the rapid advances in field theory made in association with the solution of the strong interactions. The topological soliton or kink particle of Jackiw and Rebbi is conceptually similar to the 't Hooft–Polyakov monopole ('t Hooft, 1974; Polyakov, 1974) and the skyrmion (Skyrme, 1961), both of which were proposed as simplified models of real elementary particles. Solid-state physicists first became aware of these ideas through a paper by Su, Schrieffer, and Heeger (1979) in which it was proposed that solitons might be the charge carriers of the conducting polymer polyacetylene. I first heard about this through my colleague Gene Mele, who was at the time working on the theory of polyacetylene in collaboration with Michael Rice at the Xerox Webster Research Center (Rice *et al.*, 1976). Gene was obsessed with solitons and had gone to such lengths to make realistic models of them that he could convince any reasonable person of their existence, and quickly convinced me. This is, of course, a theorist's statement. People who were not experts in quantum mechanics tended to find the idea outrageous, particularly since the experimental evidence for the existence of solitons was always indirect. It eventually came to pass that enthusiasm for the soliton idea among funding agents waned, the particle theorists diverted their energies to strings, and solitons were effectively forgotten. It is difficult nowadays to find anyone younger than I am who knows anything about them. I have always found the history of solitons to be a poignant comment on the effect of fashion on scientific thought, for there was no doubt among well-informed physicists that the idea was right and of potentially great importance.

The basic idea of the soliton is illustrated in Fig. 1. Polyacetylene is a zigzag planar chain of CH units with every other bond contracted slightly. Because there are two equivalent ways of contracting the bonds, there is a possibility of domain walls between even-contracted regions and odd-contracted ones. These are the solitons. Realistic modeling reveals that these domain walls should be quite mobile, with an acceleration mass of about 10 times the electron mass, and that they should carry either a net electric charge of e and no net spin, or a spin of 1/2 and no charge. This breakup of the charge and spin quantum numbers of the electron is the important new effect, for an excitation carrying charge but no spin, or vice versa, cannot be adiabatically deformed into a free electron as a matter of principle. The excitations of a conventional insulator, in contrast, are necessarily deformable into free electrons, and thus always
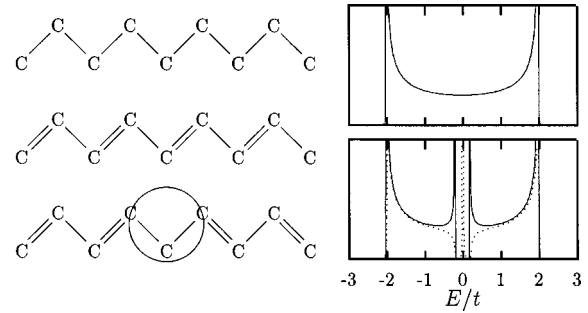

FIG. 1. Physics of polyacetylene. The symmetric configuration of the atoms shown in the upper left is absolutely unstable to a distortion that contracts every other bond. This causes the one-electron density of states, shown on the right, to acquire a gap. Formation of a domain wall results in the development of a state at mid gap (dotted curve on the right) and a net electric charge of $+e$ if this state is empty.

have charge e and spin 1/2. The alleged properties of solitons were therefore quite unprecedented and extraordinary.

The property of polyacetylene that causes solitons to exist is its discrete broken symmetry. If one wraps a molecule with an even number of segments into a ring, one discovers that it has two equivalent quantum-mechanical ground states, and that these transform into each other under clockwise rotation by one segment, a symmetry of the underlying equations of motion. A conventional insulator would have only one ground state and would transform under this operation into itself. These two states acquire classical integrity in the thermodynamic limit. When the ring is small, local perturbations, such as a force applied to one atom only, can mix the two ground states in arbitrary ways, and in particular can tunnel the system from the even-doubled state to the odd-doubled one. But this tunneling becomes exponentially suppressed as the size of the ring grows, and eventually becomes insignificant. Classical integrity and two-fold degeneracy together make the broken-symmetry state fundamentally different from the conventional insulating state; one cannot be deformed into the other in the thermodynamic limit without encountering a quantum phase transition.

The peculiar quantum numbers of the soliton are caused by the formation of a mid-gap state in the electron spectrum. The N-electron model Hamiltonian (Su *et al.*, 1979)

$$\mathcal{H} = \sum_j^N \left\{ t\left(1 - \frac{x_{j+1} - x_j}{l}\right) \sum_\sigma (c_{j,\sigma}^\dagger c_{j+1,\sigma} + c_{j+1,\sigma}^\dagger c_{j,\sigma}) \right.$$
$$\left. - \frac{\hbar^2}{2M} \frac{\partial^2}{\partial x_j^2} + \frac{1}{2} k(x_{j+1} - x_j)^2 \right\} \tag{1}$$

is solved in the limit of large M by picking fixed displacements $x_{j+1} - x_j = \pm \delta l$ and then minimizing the expected energy per site,
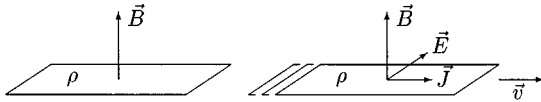
FIG. 2. When a disorder-free 2-d electron gas is placed in a magnetic field it can be Lorentz boosted to generate a current and a transverse electric field. The ratio of these then gives Eq. (3), which is inconsistent with all quantum Hall experiments. This shows that disorder is essential for plateau formation.

$$\frac{\langle \mathcal{H} \rangle}{N} = \frac{1}{2} k(l\delta)^2$$

$$- \frac{t}{2\pi} \int_0^{2\pi} \sqrt{(1+\delta^2)+(1-\delta^2)\cos(\theta)} \, d\theta. \quad (2)$$

A nonzero value of δ is obtained for any value of k, as the chain is absolutely unstable to symmetry breaking by virtue of the Peierls effect. The gapped electron spectrum shown in Fig. 2 is the density of states for $\delta = 0.1$. When the equation is solved again for the soliton one finds an extra state in the center of the gap. The soliton has charge $+e$ when this state is unoccupied. The remaining charge and spin states just reflect the various ways the mid-gap state can be populated with electrons.

The fact that soliton has charge $+e$ when unoccupied was first discovered numerically (Pople and Walmsley, 1962) and then subsequently explained by Schrieffer using his now-famous adiabatic winding argument (Su and Schrieffer, 1981). This idea is critical to the theory of the fractional quantum Hall effect, so I shall review it here briefly. One imagines adding a perturbation Hamiltonian that forces the order parameter to have a certain phase ϕ far from the origin. One then imagines locking the phase at the left end of the molecule to zero and then adiabatically advancing ϕ at the right end from 0 to π. While this is occurring the entire ground state at the right end is slowly sliding to the right, pumping electric charge out to infinity as it does so. The amount pumped by winding by π must be one electron, half the charge of the unit cell, since winding by twice this much just moves the unit cell over. But there is no net transfer of spin, as the system is a gapped singlet at every step in the operation. Since the operation also creates a soliton, we conclude that the soliton must have spin 0 and charge $+e$. It is deeply important that this argument is completely model-independent and relies only on the discrete broken symmetry of the bulk interior and the possibility of deforming the Hamiltonian into something simple at the ends of the sample without collapsing the gap.

There was an important variant of the soliton idea that foreshadowed the fractional quantum Hall discoveries, namely, when the polyacetylene was imagined to be so severely p-type doped that the number of electrons in the π band is $2N/3$ rather than N (Su and Schrieffer, 1981). The instability is then to contract every third bond, and there are two kinds of soliton—one winding forward by $2\pi/3$ and the other winding backward by the same amount—with charges $+2e/3$ and $+4e/3$ when all their mid-gap states are empty. So the separation of the spin and charge degrees of freedom of the electron in native polyacetylene is a special case of a more general effect in which the charge quantum number is fractionalized. The tendency of the Peierls instability to commensurate is actually so strong that fractionalization is expected to occur at other rational fillings as well, for example, $3N/4$ and $3N/5$. Indeed the only reason it would not occur at all rational fillings is that the gap for most of these states is small and thus susceptible to being overwhelmed by the ion kinetic energy (which we have neglected), finite temperature, or dirt. These effects all favor fractions with small denominators.

The fact that these beautiful and reasonable ideas of Schrieffer and Su never became widely accepted may be traced in the end to one key difficulty: Long-range bond-contraction order was never found in polyacetylene. It is perfectly possible for solitons to exist if the discrete symmetry breaking has not set in globally—for example, if it is interrupted every now and again by sample imperfections—and many experiments were done in highly defective and disordered polyacetylene with just this idea in mind. But the existence of solitons is *inescapable* only if the sample orders. This is a beautiful example of the special role ordering phenomena play in solid-state physics, for while not always necessary they are often sufficient to demonstrate the truth of a thing. The problem with polymers in this context is that they owe most of their unique properties to noncrystallinity and are intentionally designed to tangle and disorder, i.e., not to do the one thing that would prove the existence of solitons. So in this sense the hope of demonstrating fractional quantization in real polyacetylene conclusively was doomed from the start.

LOCALIZATION

The two-dimensional electron gas of a silicon field-effect transistor or a GaAs heterostructure, the venue of the integral and fractional quantum Hall effects, is notoriously imperfect. This is important because it immediately eliminates the possibility that microscopic details are responsible for Klaus von Klitzing's magnificent effect, the highly accurate quantization of the Hall conductance to integral multiples of e^2/h (von Klitzing *et al.*, 1980). The field-effect transistor, for example, is made by oxidizing the surface of a piece of silicon, an operation that always results in microscopic strain and bond disorder at the surface because the silicon and SiO_2 lattice parameters do not match. This problem is so troublesome that it is customary to oxidize in the presence of a small partial pressure of water so that hydrogen is available for tying up the occasional dangling silicon bond. GaAs heterostructures are better in this regard, as the interface between the GaAs and the $Al_xGa_{1-x}As$ alloy is nominally epitaxial, but the Al atoms are still substituted at random in the GaAs lattice and are thus scattering centers. In either system there is the problem of the dopant ions, which are always strong scatterers because they are stripped of their carriers and thus Coulombic. The technique of modulation doping,

invented by my co-winner Horst Störmer, mitigates this effect enormously but does not eliminate it. And of course there are chemical impurities gettered to the interface in unknown amounts. It is true that modern heterostructures have huge mobilities undreamed of in the old days, but they are not perfect.

Nor, as it turns out, would one want them to be, for imperfection is required for the quantum Hall effect to occur in the customary experimental configuration. Consider the situation, illustrated in Fig. 2, of a translationally invariant strip of charge density ρ in a normal magnetic field B. Because of translational invariance, flowing current along the strip is the same as Lorentz boosting by speed v, which gives a current $j = v\rho$, an electric field $E = vB/c$, and a Hall conductance

$$\sigma_{xy} = \frac{J}{E} = \frac{\rho c}{B}. \tag{3}$$

In a real field-effect transistor or heterostructure ρ is fixed by doping and the gate voltage and is *not* accurately quantized. Often it is even the variable against which the Hall plateaus are plotted. Thus this result is inconsistent with all quantum Hall experiments. Sample imperfection is implicated in the formation of plateaus because it is the only agent in the problem, other than sample ends, capable of destroying translational invariance.

The most obvious thing for disorder to do is cause Anderson localization (Anderson, 1958). Localization is the underlying cause of the insulating state that results when an idealized noninteracting metal is subjected to a sufficiently large random potential. It simply means that all the eigenstates of the one-electron Hamiltonian below a certain energy have finite spatial extent, so that occupying them with electrons contributes nothing to the zero-frequency conductivity. Real metals, in which the electrons interact, have a similar metal-insulator transition, and it is believed that the two states of matter in question are adiabatically continuable to their noninteracting counterparts. Already at the time of the quantum Hall discovery there was a large experimental literature on localization in 2-d metals, particularly those occurring in semiconductors, from which it was known that localization effects were so strong in the absence of a magnetic field that only the insulating state should exist at zero temperature (Abrahams *et al.*, 1979; Dolan and Osheroff, 1979; Bishop *et al.*, 1980; Bergman, 1982). So there were many reasons to suspect the quantum Hall effect of being an emergent phenomenon that, like the Anderson insulator, could be understood solely in terms of one-electron quantum mechanics and localization.

There are two exact results that are particularly important for developing this idea. The first is the solution of the trivial model

$$\mathcal{H} = \sum_{j}^{N} \left\{ \frac{1}{2m_e} \left[\frac{\hbar}{i} \vec{\nabla}_j - \frac{e}{c} \vec{A}(\vec{r}_j) \right]^2 + Eey_j \right\}, \tag{4}$$

where

$$\vec{A}(\vec{r}) = By\hat{x}. \tag{5}$$

With lengths measured in multiples of

$$l = \sqrt{\frac{\hbar}{m\omega_c}} = \sqrt{\frac{\hbar c}{eB}} \quad \left(\omega_c = \frac{eB}{mc} \right), \tag{6}$$

one finds this to be a Slater determinant of the orbitals

$$\psi_{k,n}(x,y) = \frac{1}{\sqrt{2^n n! \sqrt{\pi} L_x}} e^{ikx} e^{(y+y_0-k)^2/2}$$
$$\times \left(\frac{\partial}{\partial y} \right)^n e^{-(y+y_0-k)^2}, \tag{7}$$

where $y_0 = eEl/\hbar\omega_c$, the energies of which are

$$E_{k,n} = (n+1/2)\hbar\omega_c + \hbar c k \left(\frac{E}{B} \right) - \frac{mc^2}{2} \left(\frac{E}{B} \right)^2. \tag{8}$$

If the electric field E is small, so there is a large gap between Landau levels n and $n+1$, and the chemical potential is adjusted to lie in this gap, then the number of electrons in the sample is $N = nL_xL_y/2\pi l^2$, the charge density is $\rho = ne/2\pi l^2$, and Eq. (3) becomes $\sigma_{xy} = \rho c/B = ne^2/h$. As a check on this result we note that

$$\frac{1}{m} \int \int \psi^*_{k,n}(x,y) \left(\frac{\hbar}{i} \frac{\partial}{\partial x} - \frac{e}{c} A_x \right) \psi_{k,n}(x,y) dx dy = c\frac{E}{B}. \tag{9}$$

Thus the current carried by each orbital is e times classical drift velocity cE/B. Adding these up and and dividing by the sample area, we find that

$$J = \frac{ne^2}{\hbar} E. \tag{10}$$

The second important result is the exact solution of the Hamiltonian

$$\mathcal{H}' = \mathcal{H} - V_0 l^2 \sum_{j}^{N} \delta^2(\vec{r}_j), \tag{11}$$

originally worked out by R. E. Prange (1981). The δ-function impurity potential is found to bind a single localized state down from each Landau level, and this state does not carry electric current, consistent with one's intuition. However, the remaining delocalized states turn out to be carrying too much, and the sum of the excess just exactly cancels the loss of a state! I have always found this result astonishing, for it is as though the remaining delocalized states understood that one of their comrades had been killed and were pulling harder to make up for its loss.

In the light of hindsight it is possible to prove that the Prange effect occurs for a broad class of impurity potentials. For this we must invoke some more general principles, for it is scarcely practical to diagonalize all possible impurity Hamiltonians and compute the current carried by their thermodynamically large number of orbitals one by one. Instead we notice that the current operator is formally the derivative of the Hamiltonian with respect to vector potential (Laughlin, 1981). That is, if we let
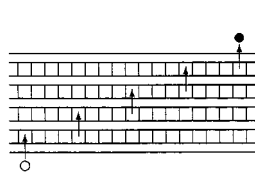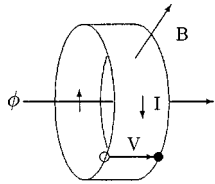
FIG. 3. Illustration of thought experiment identifying quantum of Hall conductance with the electron charge. As magnetic flux $\Delta\phi = hc/e$ is adiabatically forced through the loop, one electron per Landau level is transferred from one edge to the other. In the absence of disorder the transfer is accomplished by a mechanical shift of the one-body wave functions that evolves each wave function into its neighbor. When small amounts of disorder are present the number of electrons transferred must be exactly the same, but the mechanism of transfer is not, for the wave functions are violently distorted by even the smallest perturbation.

$$\vec{A} \rightarrow \vec{A} + A_0 \hat{x}, \tag{12}$$

where A_0 is a constant, then

$$\frac{e}{m_e} \sum_j \left[\frac{\hbar}{i} \frac{\partial}{\partial x} - \frac{e}{c} A_x(\vec{r}_j) \right] = -c \frac{\partial\mathcal{H}}{\partial A_0}. \tag{13}$$

This is actually true when the particles interact as well, a fact that will prove useful later. Normally this relation is of little help, for the addition of a constant vector potential is simply a gauge transformation, which has no physical meaning. But if the sample is wrapped into a loop, as shown in Fig. 3, then A_0 acquires the physical meaning of a magnetic flux $\phi = A_0 L_x$ forced through the loop. Let us now imagine picking a sequence of flux values and solving the problem

$$\mathcal{H}_\phi |\Psi_\phi\rangle = E_\phi |\Psi_\phi\rangle \tag{14}$$

for each one, so that the many-body ground state $|\Psi_\phi\rangle$ and corresponding energy eigenvalue E_ϕ are tabulated functions of ϕ. Then by the Hellman-Feynman theorem we have

$$\left\langle \Psi_\phi \left| \frac{\partial\mathcal{H}_\phi}{\partial\phi} \right| \Psi_\phi \right\rangle = \frac{\partial}{\partial\phi} \langle\Psi_\phi|\mathcal{H}_\phi|\Psi_\phi\rangle = \frac{\partial E_\phi}{\partial\phi}. \tag{15}$$

That is, the total current at any given value of ϕ is just the adiabatic derivative of the total energy with respect to ϕ. This is actually not surprising, for slowly changing ϕ creates an electromotive force around the loop, which does work on the system if current is flowing. If there is no dissipation then the energy eigenvalue must increase accordingly. The effect is just Faraday's law of induction. Now if the loop is large, so that Aharonov-Bohm oscillations are suppressed and the current changes negligibly during the insertion process, the adiabatic derivative may be replaced by a differential

$$I = c \frac{\Delta E}{\Delta\phi}, \tag{16}$$

where the denominator is the flux quantum hc/e. The advantage of this is that $\mathcal{H}_{\Delta\phi}$ is then exactly equal to $\mathcal{H}_0$ up to a gauge transformation. This means that the energy can have increased only through repopulation of
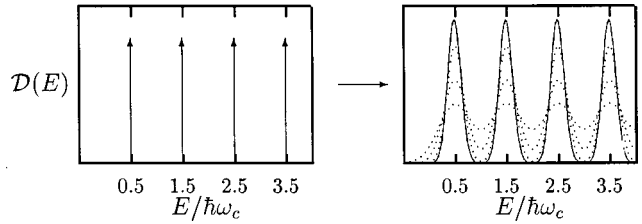

FIG. 4. Effect of disorder on the one-body density of states. Landau levels at $(n + 1/2)\hbar\omega_c$, which are highly degenerate, first broaden slightly without filling in the gap, thus maintaining the quantum Hall sum rule without the need to invoke localization. The states in the tails are localized, however, so the Fermi level may be moved down into them with no ill effects. Additional disorder broadens the spectrum further until the tails grow together. The quantization is still exact, however, because all the states at the Fermi surface are localized. This enables the quantum Hall effect to occur even when there are no true gaps.

the original states. How this comes about in the translationally invariant case is shown in Fig. 3. The orbitals in the presence of nonzero ϕ are

$$\psi_{k,n}(x,y) = \frac{1}{\sqrt{2^n n! \sqrt{\pi} L_x}} e^{ikx} e^{(y+y_0-k-\alpha)^2/2}$$

$$\times \left(\frac{\partial}{\partial y} \right)^n e^{-(y+y_0-k-\alpha)^2}, \tag{17}$$

where $\alpha = e l \phi_0 / \hbar c L_x$. As ϕ is advanced from 0 to $\Delta\phi$ they simply slide over like a shift register, the net result being to transfer one state per Landau level from the left side of the sample to the right, i.e., to transfer one electron across the sample per occupied Landau level. If the potential difference between the two sides is V we thus have

$$I = c \frac{neV}{\Delta\phi} = n \frac{e^2}{h}. \tag{18}$$

Now we can imagine turning on a small impurity potential in the interior of the ribbon. If the potential is sufficiently small that the gaps between Landau levels remain clean then there can be no change in the outcome of this thought experiment, for adiabatic evolution of ϕ stuffs exactly one state per Landau level into the disordered region on the left side and pulls exactly one out on the right. Conservation of states requires that they get through somehow. This applies not only to random potentials but to nonrandom ones as well, including the δ function used by Prange. Thus we have shown that any one-electron Hamiltonian that can be adiabatically evolved into ideal Landau levels without having states cross the Fermi level has an exactly quantized Hall conductance.

This previous argument can be significantly strengthened if the potential is random. In Fig. 4 we plot the one-electron density of states of the Hamiltonian,

$$\mathcal{H} = \sum_j^N \left\{ \frac{1}{2m_e} \left[\frac{\hbar}{i} \vec{\nabla}_j - \frac{e}{c} \vec{A}(\vec{r}_j) \right]^2 + V_{\text{random}}(\vec{r}_j) \right\} \tag{19}$$

for various strengths of V_{random}. The exact shape of this density of states is model-dependent, but the basic features are not. When V_{random} is small its primary effect is to break the degeneracy of the Landau level. The states in the tails of the distribution that results are localized. This is easiest to see in the limit that V_{random} is slowly varying and smooth, for then the one-body eigenstates are well approximated as racetracks that travel along equipotentials. The analogy commonly drawn is with a mountainous landscape filled up to a certain height with water (Trugman, 1983). If the water level is low one gets small, isolated lakes, which become increasingly isolated as the water level drops because the deepest valleys occur rarely. If it is high one gets small, isolated islands, which become increasingly isolated as the water level rises because the tallest mountains occur rarely. Somewhere in the middle is a percolation point dominated at long length scales by a vast, crenulated shoreline which cannot distinguish islands from lakes. The significance of localization in the tails is that localized states are "not there" in the flux-winding thought experiment, in the sense that a state confined to one side of the loop cannot tell that the vector potential being added adiabatically is associated with flux ϕ through the loop. It thinks that we are doing a gauge transformation, so it simply changes its phase. It does not move in response to addition of ϕ, and its energy does not change. This means that it contributes nothing to the sum rule, and that whether it is occupied or not is irrelevant to the thought experiment. Thus the theorem of exact quantization is true not only if there are no states at the Fermi level, but also if all the states at the Fermi level are localized.

The ability of the exact quantization theorem to be extended to cases where the density of states does not have true gaps is crucial for accounting for real quantum Hall experiments. A field-effect transistor is a capacitor. It stores charge in response to the application of a gate voltage by the rule $Q = CV_g$, where C is determined by the oxide thickness, sample area, and little else. Accordingly, a sweep of the gate voltage is really a sweep of Q, not of chemical potential. The chemical potential in a real experiment simply adjusts itself to whatever is required to fix the charge at Q. Thus, if it were not for localization, the chemical potential would always be pinned in a Landau level, this Landau level would always be only partially occupied, and the conditions for observing the quantum Hall effect would never be achieved. But if the Fermi level lies in a region of localized states then the chemical potential can move about freely, populating or depopulating the localized states at will with no effect whatsoever on the Hall conductance. Another aspect of the experiments nicely accounted for by localization is lack of parallel resistance. In the limit that the electrons do not interact the parallel conductance σ_{xx} is due to electric dipole transitions from states just below the Fermi level to just above. If the states in question are localized, σ_{xx} must be zero. Localization causes insulation. However, the resistivity and conductivity tensors are inverses of each other, so that if $\sigma_{xy} = ne^2/h$ we also have

$$\begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}^{-1} = \frac{h}{ne^2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (20)$$

Note that experiments that tune the magnetic-field strength instead of V_g are not fundamentally different, as they simply vary the length scale against which the charge density is measured. This still requires the chemical potential to adjust itself to keep Q fixed.

The flux-winding experiment also demonstrates that small amounts of disorder cannot completely localize all the states in a Landau level. There must be extended states, for if there were not the whole Landau level would be "not there" and therefore incapable of carrying current. This is a very nontrivial result, for the arguments that all states should be localized in two dimensions in the absence of a magnetic field are sound. Just exactly where the extended states reside in the weak-disorder limit is a matter of some debate. Most experiments are consistent with the idea, originally proposed by Levine, Libby, and Pruisken (1983), that there is a scaling theory of localization for this problem, although a different one from the field-free case, and that the extended states occur at one and only one energy somewhere near the center of the broadened Landau level. The strong-disorder limit is not controversial at all, however, for it is clear that complete localization must occur in this case. Another important implication of the flux-winding result therefore is that extended-state bands cannot simply disappear but must "float" through the Fermi surface as the disorder potential is increased, for only when extended states appear at the Fermi energy does the argument fail. The floating effect, which was predicted simultaneously by David Khmel'nitzkii (1984) and me (Laughlin, 1984), was eventually observed experimentally (Glozman *et al.*, 1995).

The larger idea underlying these arguments is that the quantum Hall effect is an emergent phenomenon characterized by the ability of the matter in question to pump an integral number of electrons across the sample in a flux-winding experiment. The noninteracting models we have been discussing are nothing but prototypes; having understood them we identify the real experiments as their adiabatic continuations. This is not a very radical idea, for except for the presence of a magnetic field the ideal Landau level is not that different from a band insulator, a state of matter known to be continuable to its noninteracting prototype or, more precisely, defined by this property. The integrity of the low-lying excitations in either case is protected by the existence of the gap, which assures there are no other states into which a low-energy excitation can decay. This does not prove that other states are impossible—which is fortunate since the discovery of the fractional quantum Hall effect proves otherwise—but it does show that it makes sense, particularly since there are plenty of experimental instances of adiabatic continuability from an interacting system to a noninteracting one in solid-state physics. If the state in question does map to a noninteracting prototype then a flux-winding experiment either dissipates or pumps an integral number of electrons across the

sample. When the latter occurs, the Hall conductance is accurately quantized. The reason is that it measures the charge of the particle being pumped, in this case the charge of the electron.

FRACTIONAL QUANTUM HALL STATE

The fractional quantum Hall state is *not* adiabatically deformable to any noninteracting electron state. I am always astonished at how upset people get over this statement, for with a proper definition of a state of matter and a full understanding of the integral quantum Hall effect there is no other possible conclusion. The Hall conductance would necessarily be quantized to an integer because it is conserved by the adiabatic map and is an integer in the noninteracting limit by virtue of gauge invariance and the discreteness of the electron charge. So the fractional quantum Hall state is something unprecedented—a new state of matter.

Its phenomenology, however, is the same as that of the integral quantum Hall state in almost every detail (Tsui *et al.*, 1982). There is a plateau. The Hall conductance in the plateau is accurately a pure number times e^2/h. The parallel resistance and conductance are both zero in the plateau. Finite-temperature deviations from exact quantization are activated or obey the Mott variable-range hopping law, depending on the temperature. The only qualitative difference between the two effects is the quantum of Hall conductance.

Given these facts the simplest and most obvious explanation, indeed the only conceivable one, is that the new state is adiabatically deformable into something physically similar to a filled Landau level except with fractionally charged excitations. Adiabatic winding of a flux quantum—which returns the Hamiltonian back to itself exactly—must transfer an integral number of these objects across the sample. Localization of the objects must account for the existence of the plateau. All the arguments about deformability and exactness of the quantization must go through as before. As is commonly the case with new emergent phenomena, it is the experiments that tell us these things must be true, not theories. Theories can help us better understand the experiments, in particular by providing a tangible prototype vacuum, but the deeper reason to accept these conclusions is that the experiments give us no alternative.

My prototype ground state for the original 1/3 effect discovered by Tsui, Störmer, and Gossard is (Laughlin, 1983a)

$$\Psi_m(z_1,\ldots,z_N)=\prod_{j\langle k\rangle}(z_j-z_k)^m \exp\left[-\frac{1}{4l^2}\sum_j^N |z_j|^2\right],$$
(21)

where m is an odd integer, in this case 3, and $z_j=x_j+iy_j$ is the location of the jth particle expressed as a complex number. Horst likes to joke that his whole effect fits in one tiny equation, and I am deeply flattered every time he makes this joke in public, but the truth is that the equation is simple only because he and Dan had

the good fortune to find the 1/3 state first. Most of the other 30-odd fractional quantum Hall vacua that have now been discovered do not have prototypes this simple, and I would be remiss in not pointing out that there are now reasonable alternatives (Jain, 1989). These ground states are all adiabatically deformable into each other, however, and are in this sense fundamentally the same. My wave function was originally conceived as a variational ground state for the model Hamiltonian

$$\mathcal{H}=\sum_j^N \left\{\frac{1}{2m_e}\left[\frac{\hbar}{i}\vec{\nabla}_j-\frac{e}{c}\vec{A}(\vec{r}_j)\right]^2+V_{\text{ion}}(\vec{r}_j)\right\}$$
$$+\sum_{j<k}^N v(\vec{r}_j-\vec{r}_k),$$
(22)

where

$$\vec{A}(\vec{r})=\frac{B}{2}(x\hat{y}-y\hat{x}),$$

$$V_{\text{ion}}(\vec{r})=-\rho\int_{\text{sample}} v(\vec{r}-\vec{r}')d\vec{r}'.$$
(23)

The ion potential is present only because one of the electron-electron repulsions we wish to consider is the Coulomb interaction $v(r)=e^2/r$, which must be neutralized by a background charge density ρ for the system to be stable. This wave function was subsequently shown by Duncan Haldane (1983) to be an exact ground state of a class of Hamiltonians with nonlocal potentials.

The most important feature of this wave function is that it locks the electron density at exactly $1/2\pi ml^2$ in the limit that N becomes thermodynamically large. We know this to be true because the square of the wave function is equivalent to the probability distribution function of a classical one-component plasma. Letting

$$|\Psi(z_1,\ldots,z_N)|^2=e^{-\beta\Phi(z_1,\ldots,z_N)},$$
(24)

and choosing $\beta=1/m$ to make the analogy transparent, we obtain

$$\Phi(z_1,\ldots,z_N)=-2m^2\sum_{j<k}^N \ln|z_j-z_k|+\frac{m}{2l^2}\sum_j^N |z_j|^2.$$
(25)

This is the potential energy of particles of "charge" m repelling each other logarithmically—the natural Coulomb potential in two dimensions—and being attracted to the origin a uniform "charge" density $1/2\pi l^2$. In order to have local electrical neutrality, which is essential in a plasma, the particles must have density $\rho=1/2\pi ml^2$.

It is also very important that the state is not crystalline when m is small. This is obvious for the case of $m=1$, for then the wave function is just a full Landau level, but for other values of m one must appeal to the extensive literature on the classical one-component plasma (Caillol *et al.*, 1982). Numerical studies have shown that crystallinity—or, more precisely, power-law correlations, for true crystallinity is impossible in two dimensions at finite temperature—occurs when the thermal coupling

constant $\Gamma = 2m$ is about 140. Thus we are deeply in the liquid range at $m=3$ and generally when m is a small odd integer. It was partially on the basis of this that I predicted the existence of a fractional quantum Hall state at 1/5, which was eventually found experimentally (Chang *et al.*, 1984).

Another important feature of this state is the gap in its excitation spectrum, which is to say its preferred density. This gap is indicated by the experiments and also by the functional form of Ψ_m, which gives only densities $1/2\pi ml^2$, but was demonstrated rigorously to exist only when Haldane and Rezayi (1985) numerically diagonalized Eq. (22) on a small sphere. Exactly the same value for this gap—about $0.08e^2/l$ for the Coulombic case at $m=3$—was obtained by Girvin, MacDonald, and Platzman (1985) using hydrodynamic arguments. This is important because it identifies the lowest-energy excitation to be a quantum of compressional sound. In most quantum fluids the density operator ρ_q, appropriately projected, has a large amplitude to create a phonon. If one assumes this to be the only significant amplitude then one can exploit a sum rule to express the excitation energy E_q in terms of ground-state properties solely (Feynman, 1972). Thus letting

$$\hat{\rho}_q = P\left[\sum_j^N e^{i\vec{q}\cdot\vec{r}_j}\right]P, \qquad (26)$$

where P is the projector onto the lowest Landau level, and denoting by $|x\rangle$ an arbitrary excited state of the Hamiltonian of energy E_x above the ground state, we have

$$E_q = \frac{\Sigma_x E_x |\langle x|\hat{\rho}_q|\Psi_m\rangle|^2}{\Sigma_x |\langle x|\hat{\rho}_q|\Psi_m\rangle|^2}$$

$$= \frac{1}{2}\frac{\langle\Psi_m|[\hat{\rho}_{-q},[\mathcal{H},\hat{\rho}_q]]|\Psi_m\rangle}{\langle\Psi_m|\hat{\rho}_{-q}\hat{\rho}_q|\Psi_m\rangle}. \qquad (27)$$

The rigidity of the classical plasma implicit in Ψ_m causes the denominator of this expression to vanish unusually rapidly as $\vec{q}\to 0$, the result being that E_q comes in at a finite constant. The dispersion relation has a shallow minimum at $q=1.4/l$ for $m=3$, a value close to the ordering wave vector $q=1.56/l$ of the competing Wigner crystal. This minimum is aptly analogous to the roton of liquid ^{4}He. E_q is quite similar to the dispersion relation of the conventional exciton in a filled Landau level (Kallin and Halperin, 1984).

The existence of an energy gap is sufficient to prove that the state has elementary excitations that carry a fraction of the electron charge. One imagines a thought experiment, illustrated in Fig. 5, in which the sample is poked with a thin magnetic solenoid through which magnetic flux ϕ is adiabatically inserted. What happens near the solenoid in this process is complicated and model dependent, but far away from the solenoid the effect is simply to translate each Landau orbit inward as the phase is advanced, just as occurred in the integral quantum Hall loop experiment. Since the Hamiltonian is returned to its original state by the advance of ϕ from 0
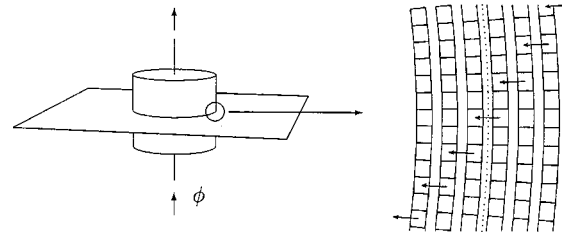

FIG. 5. Illustration of thought experiment demonstrating the existence of fractionally charged excitations. A fractional quantum Hall ground state is pierced with an infinitely thin magnetic solenoid, through which magnetic flux ϕ is inserted adiabatically. The effect far away from the solenoid is to move states toward the solenoid. Advancing the phase all the way to $\Delta\phi = hc/e$ returns the Hamiltonian to its original state, up to an unimportant gauge transformation. This means that exactly one state must be transported through the surface of a large Gauss's law pillbox centered at the solenoid, and also that the solenoid may be removed at the end with no ill effects. The net result is the creation of an eigenstate of the original Hamiltonian with charge e/m, the average charge per state at infinity.

to $\Delta\phi = hc/e$, up to an unimportant gauge transformation, the net effect of this advance must be to draw exactly one state per Landau level through the surface of an imaginary Gauss's law pillbox enclosing the solenoid. This draws in charge e/m, the average charge per state at infinity, and piles it up somewhere in the vicinity of the solenoid. The solenoid may then be removed, leaving behind an exact excited state of the original Hamiltonian carrying charge e/m, the charge *density* of the ground state.

A slight modification of this line of reasoning shows that the quantization of the fractional charge is exact. Let us imagine the situation illustrated in Fig. 6, in which the bare electron mass changes slowly across the sample, interpolating between a realistic value in region A to one in region C so small that Eq. (21) becomes exact, all the while maintaining the integrity of the gap. We now perform the flux-winding experiment as before, only this time we make the Gauss's law pillbox so big that it cuts region C. What happens in region A is complicated and impossible to predict accurately from first principles, but
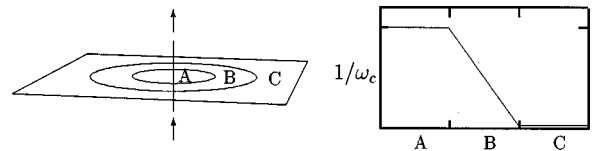

FIG. 6. Illustration of thought experiment showing that the fractional charge is exact. One imagines a Hamiltonian parameter, such as the bare electron mass, that varies slowly in space so as to continuously deform a test system A into an ideal one at infinity C. A flux-winding experiment performed in A then causes charge e/m to flow through a Gauss's law pillbox that cuts region C, regardless of the details, and this charge has no choice but to accumulate near the solenoid. In this way region A inherits exact properties from region C by virtue of integrity of the gap across region B.

in region C exactly e/m of electric charge must be drawn in from infinity per the previous arguments. But there is no place for this charge to go other than the solenoid, at least if region A is large. Thus the operation must have made an excitation in region A with charge exactly e/m, regardless of microscopic details. We have actually proved something stronger in this example, for a second Gauss's law box placed inside region A shows that this same charge must be related in a fundamental way to the charge density in region A. Thus the latter is also quantized to the ideal value, even though Eq. (21) is not exact in this region. These arguments are quite general and apply equally well to Hamiltonian parameters other than the bare mass we may wish to make nonideal. The charge is conserved by any adiabatic modification of the Hamiltonian that maintains the gap and is therefore characteristic of the entire phase of matter in question, not simply a particular prototype.

Wave functions that reasonably approximate these fractionally charged excitations, which I gave the unfortunate name "quasiparticles," are

$$\Psi_{z_0}^+(z_1,\ldots,z_N) = \exp\left[-\frac{1}{4l^2}\sum_j^N |z_j|^2\right]$$
$$\times \prod_j^N (z_j-z_0) \prod_{j<k}^N (z_j-z_k)^m, \quad (28)$$

for the positive excitation at location z_0, and

$$\Psi_{z_0}^-(z_1,\ldots,z_N) = \exp\left[-\frac{1}{4l^2}\sum_j^N |z_j|^2\right]$$
$$\times \prod_j^N \left(2l^2\frac{\partial}{\partial z_j}-z_0^*\right)\prod_{j<k}^N (z_j-z_k)^m,$$
$$(29)$$

for the negative one. The excitation energies for these I originally estimated were $0.022\,e^2/l$ and $0.025\,e^2/l$, respectively, for Coulomb interactions at $m=3$ (Laughlin, 1983a). Subsequent numerical work by Haldane and Rezayi (1985) and Morf and Halperin (1986) found the better numbers $0.026\,e^2/l$ and $0.073\,e^2/l$. These improved energies sum to a value slightly higher than the roton gap, as expected if the collective mode is viewed as a quasiparticle-quasihole bound state, i.e., as an exciton. Charge transport gaps of this size are found experimentally (Boebinger et al., 1985) but are about half the theoretical value corrected for finite thickness of the 2-d electron gas. The remaining error is not of great concern because disorder is expected to severely lower the experimental gap. It is easy to see that the quasihole wave function describes a charge-1/3 excitation. Following the procedure of Eq. (24) we obtain

$$\Phi(z_1,\ldots,z_N) = -2\sum_j^N \ln|z_j-z_0|$$
$$-2m^2\sum_{j<k}^N \ln|z_j-z_k| + \frac{m}{2l^2}\sum_j^N |z_j|^2.$$
$$(30)$$

The plasma particles now see a phantom of "charge" 1 at location z_0 and arrange themselves on average to accumulate equal and opposite "charge" near z_0. It is somewhat more involved to demonstrate that the quasielectron wave function also describes a charge-1/3 excitation, but the reasoning is similar (Laughlin, 1987).

There are now a number of experimental papers which have reported the direct observation of charge $e/3$ quasiparticles. The most recent and widely cited are the shot-noise measurements of Saminadayar et al. (1997) and de-Picciotto et al. (1997), which detect the charge through fluctuations in the current leaking across a narrow neck in a Hall bar. These very impressive experiments are more subtle than they appear at first glance because the relevant tunneling processes occur between sample edges, the excitation spectra of which are not gapped as they are in the bulk interior but rather the gapless spectra of chiral Luttinger liquids (Wen, 1990). The carriers in this strange one-dimensional metal carry a charge $e/3$ inherited from the bulk but are somewhat different physically from the quasiparticles in the interior and might even be construed as a different phenomenon. The shot noise expected from the tunneling of these excitations turns out to have the classical form with the electron charge reduced to $e/3$ (Kane and Fisher, 1994). Somewhat more controversial, but in my view quite sound, are the resonant tunneling experiments of Goldman and Su (1995; Goldman, 1996), which were well-controlled versions of a an older experiment by Simmons et al. (1989). These experiments also involve transport across a narrow neck of a Hall bar but measure zero-frequency transport as a function of carrier density and magnetic field instead of time-dependent current fluctuations (Jain et al., 1993). I should also mention the famous experiments of Clark (Clark et al., 1988), which reported the observation of fractional charge in the high-temperature intercept of the activated parallel conductivity of the Hall plateau. These measurements were startling but controversial because no simple theoretical basis for the effect could be found. My own view, however, has continued to be that the best spectroscopy of this charge is the quantized Hall conductance itself, particularly in the limit that the sample is so dirty that arguments based on idealized edges make no sense, for the flux-winding sum rule measures the charge of the object transported across the sample, regardless of details.

FRACTIONAL STATISTICS

Fractional quantum Hall quasiparticles exert a long-range velocity-dependent force on each other—a gauge force—which is unique in the physics literature in having neither a progenitor in the underlying equations of motion nor an associated continuous broken symmetry. It arises spontaneously along with the charge fractionalization and is an essential part of the effect, in that the quantum states of the quasiparticles would not count up properly if it were absent. This force, which is called fractional statistics (Leinaas and Myrheim, 1977; Wilc-

zek, 1982), has a measurable consequence, namely, the values of the subsidiary fractions 2/5 and 2/7 and their daughters in the fractional quantum Hall hierarchy.

An isolated quasiparticle behaves physically like an ordinary electron or hole in a Landau level, except that its magnetic length is effectively $\sqrt{m}l$. This follows from its degeneracy in z_0, the functional form of the overlap matrix

$$\frac{\langle \Psi_{z_0}^+ | \Psi_{z_0'}^+ \rangle}{\sqrt{\langle \Psi_{z_0}^+ | \Psi_{z_0}^+ \rangle \langle \Psi_{z_0'}^+ | \Psi_{z_0'}^+ \rangle}}$$

$$= \exp\left[-\frac{1}{4ml^2}(|z_0|^2 - 2z_0^* z_0' + |z_0'|^2) \right], \qquad (31)$$

and the equivalence of the wave function at $m=1$ to an ordinary hole in the orbital

$$\phi_{z_0}(z) = \exp\left[-\frac{1}{4l^2}(|z|^2 + |z_0|^2 + 2z_0^* z) \right]. \qquad (32)$$

This mapping is both one-to-one and physically apt.

A pair of quasiparticles, on the other hand, behaves like a pair of ordinary electrons or holes in a Landau level carrying magnetic solenoids containing a fraction of a flux quantum (Arovas *et al.*, 1984). The two-quasiparticle wave function

$$\Psi_{z_A z_B}^{++}(z_1, \ldots, z_N)$$

$$= \exp\left[-\frac{1}{4l^2} \sum_j^N |z_j|^2 \right] \prod_j^N (z_j - z_A)(z_j - z_B)$$

$$\times \prod_{j<k}^N (z_j - z_k)^m \qquad (33)$$

is equivalent at $m=1$ to a pair of ordinary holes in an otherwise full Landau level. It is obviously symmetric under interchange of the quantum numbers z_A and z_B and would thus appear to be the wave function of a pair of bosons. This is not correct, however, for it is uniquely the case in two dimensions that fermions have Bose representations and vice versa. We must therefore use the more sophisticated concept of a Berry phase to determine whether the particles are physically fermions or bosons. Suppose the Hamiltonian is modified slightly so as to stabilize a pair of quasiparticles at locations z_A and z_B. This might be accomplished, for example, by adding shallow potential wells at these locations. If the Hamiltonian parameter z_A is then evolved around adiabatically in a closed loop P, the wave function returns to itself up to the phase

$$\phi = \oint_P \vec{A} \cdot d\vec{s},$$

$$\vec{A}(z_A) = \lim_{z_A' \to z_A} \vec{\nabla}_A \langle \Psi^{++} z_A' z_B | \Psi^{++} z_A z_B \rangle. \qquad (34)$$

This expression gives

$$\vec{A}(\vec{r}) = \frac{1}{2m}(\vec{r} \times \hat{z}) - \frac{1}{m} \frac{(\vec{r} - \vec{r}_B) \times \hat{z}}{|\vec{r} - \vec{r}_B|^2} \qquad (35)$$

for $\vec{r}$ far from $\vec{r}_B$. The curl of this vector potential is just 2π times the local charge density, so the extra solenoidal component is simply a reflection of the missing charge associated with the presence of the quasihole at z_B. In the case of $m=1$ the extra Berry phase incurred in moving A around B is just 2π. This means that the phase incurred in going around halfway, so as to exchange the particles, is π. The particles in this case are fermions. Had we picked a Fermi representation in which to do the calculation, for example by multiplying the wave function by $z_A - z_B$, this extra phase would have come out to be zero, but the end result would still have been that the wave function returned to minus itself when the particles were exchanged. When $m \neq 1$, however, neither the Fermi nor the Bose representation gets rid of this solenoidal component completely. In this case it is not an artifact of the choice of representation but a real velocity-dependent force.

The important experimental effect of the fractional statistics is to change the way quasiparticles pack. If quasiholes were fermions, for example, so that the analogy with a Landau level were exact, then occupying every available state would result in a uniform charge density of $e/2\pi m^2 l^2$. This is correct for $m=1$ but for no other case, for the wave function

$$\Psi(z_1, \ldots, z_N) = \exp\left[-\frac{1}{4l^2} \sum_j^N |z_j|^2 \right]$$

$$\times \prod_j^N z_j^M \prod_{j<k}^N (z_j - z_k)^m, \qquad (36)$$

which describes M quasiholes packed as tightly together near the origin as possible, simply pushes the fluid back from a disc of area $2\pi M l^2$, thereby creating a fluid of uniform charge density $e/2\pi m l^2$. Bosons, of course, would pack at any density they liked, so the actual behavior of quasiparticles is somewhere in between. It was on the basis of such observations that Halperin (1984) first realized that the packing effect would account nicely for the observed subsidiary fractions 2/5 and 2/7 if the quasiparticles themselves were condensing into an analogue of the $1/m$ state. The Bose-representation wave function for condensing quasiparticles into the $\pm 1/3$ fractional-statistics analogue of the 1/3 state is

$$\Psi_\pm(\eta_1, \ldots, \eta_N) = \prod_{j\langle k\rangle}^N (\eta_j - \eta_k)^2 |\eta_j - \eta_k|^{\pm 1/3}$$

$$\times \exp\left[-\frac{1}{4l^2} \sum_j^N |z_j|^2 \right]. \qquad (37)$$

The corresponding charge densities are

$$\rho = \frac{e}{2\pi l^2} \left[\frac{1}{3} \mp \frac{1}{9(2 \pm 1/3)} \right]$$

$$= \frac{1}{2\pi l^2} \left[\begin{array}{ll} 2/7 & \text{(quasiholes)} \\ 2/5 & \text{(quasielectrons)} \end{array} \right]. \qquad (38)$$

Repeating this argument hierarchically, Halperin was able to predict a sequence of fractional quantum Hall states which agreed with experiment and also with Haldane's more algebraic derivation of the sequence (Haldane, 1983). It was subsequently discovered by Jain (1989) that the sequence of fractional quantum Hall ground states could be constructed by a method that did not employ quasiparticles at all, and thus the obvious conclusion that the occurrence of these fractions *proves* the existence of fractional statistics was called into question. However, it should not have been. The quasiparticles are quite far apart—about $3l$—in the 2/7 and 2/5 states, and the gap to make them is large, so to assume that they simply vanish when these subsidiary condensates form makes no sense. Had the quasiparticles been fermions these densities would have been $10/27 = 0.370$ rather than $2/5 = 0.40$ and $8/27 = 0.296$ rather than $2/7 = 0.286$. The effect of the fractional statistics is therefore small but measurable, about 5% of the observed condensation fraction.

REMARKS

The fractional quantum Hall effect is fascinating for a long list of reasons, but it is important, in my view, primarily for one: It establishes experimentally that both particles carrying an exact fraction of the electron charge e and powerful gauge forces between these particles, two central postulates of the standard model of elementary particles, can arise spontaneously as emergent phenomena. Other important aspects of the standard model, such as free fermions, relativity, renormalizability, spontaneous symmetry breaking, and the Higgs mechanism, already have apt solid-state analogues and in some cases were even modeled after them (Peskin, 1995), but fractional quantum numbers and gauge fields were thought to be fundamental, meaning that one had to postulate them. This is evidently not true. I have no idea whether the properties of the universe as we know it are fundamental or emergent, but I believe the mere possibility of the latter must give string theorists pause for it would imply that more than one set of microscopic equations is consistent with experiment—so that we are blind to the microscopic equations until better experiments are designed—and also that the true nature of these equations is irrelevant to our world. So the challenge to conventional thinking about the universe posed by these small-science discoveries is actually troubling and very deep.

Fractional quantum Hall quasiparticles are the elementary excitations of a distinct state of matter that cannot be deformed into noninteracting electrons without crossing a phase boundary. That means they are different from electrons in the only sensible way we have of defining different, and in particular are not adiabatic images of electrons the way quasiparticle excitations of metals and band insulators are. Some composite fermion enthusiasts claim otherwise—that these particles are nothing more than screened electrons (Jain, 1989)—but this is incorrect. The alleged screening process always runs afoul of a phase boundary at some point, in the process doing some great violence to the ground state and low-lying excitations. I emphasize these things because there is a regrettable tendency in solid-state physics to equate an understanding of nature with an ability to model, an attitude that sometimes leads to overlooking or misinterpreting the higher organizing principle actually responsible for an effect. In the case of the integral or fractional quantum Hall effects the essential thing is the accuracy of quantization. No amount of modeling done on any computer existing or contemplated will ever explain this accuracy by itself. Only a thermodynamic principle can do this. The idea that the quasiparticle is only a screened electron is unfortunately incompatible with the key principle at work in these experiments. If carefully analyzed it leads to the false conclusion that the Hall conductance is integrally quantized.

The work for which the three of us have been awarded the Nobel Prize was a collaborative effort of many excellent people in the most respected traditions of science. I join my colleagues in regretting that Art Gossard could not have shared in the Prize, as everyone in solid-state physics understands that materials are the soul of our science and that no significant intellectual progress is ever possible without them. I gratefully acknowledge the numerical work of Duncan Haldane and Ed Rezayi (Haldane and Rezayi, 1985), which was crucial in cementing the case that the energy gap existed and in calibrating the quasiparticle creation energies. I similarly acknowledge Bert Halperin's many outstanding contributions, including particularly his discovery that quasiparticles obey fractional statistics (Halperin, 1984). The list of fundamentally important contributions to the subject other than my own is so long that I cannot begin to do it justice. There are the numerous papers Steve Girvin and Alan MacDonald wrote, including particularly their obtaining, with Phil Platzman, the first accurate estimate of the energy gap (Girvin *et al.*, 1985). There is Ad Pruisken's work on localization in a magnetic field and his proposition of the first appropriately modified scaling theory (Levine *et al.*, 1983). There is Xiao-Gang Wen's work on chiral edge excitations (Wen, 1990) and the follow-on work of Charles Kane and Matthew Fisher (1994) describing quasiparticle tunneling through mesoscopic necks. There was the magnificent global phase diagram of the fractional quantum Hall effect proposed by Steve Kivelson, Dung-Hai Lee, and Shoucheng Zhang (1992). And of course there is the discovery of the strange Fermi surface at half-filling and its explanation in terms of composite fermions by Bert Halperin, Patrick Lee, and Nick Read (1993) that is now defining the intellectual frontier in this field. I hope all my colleagues who have been involved with this subject over the years, both those I have mentioned and those I have not, will accept my gratitude and appreciation for all they have done and my humble acknowledgement that the theory of the fractional quantum Hall effect, like all good science, is the work of many hands.

REFERENCES

Abrahams, E., P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, 1979, Phys. Rev. Lett. **42**, 673.

Anderson, P. W., 1958, Phys. Rev. **112**, 1900.

Anderson, P. W., 1972, Science **177**, 393.

Arovas, D., F. Wilczek, and J. R. Schrieffer, 1984, Phys. Rev. Lett. **53**, 722.

Bergman, G., 1982, Phys. Rev. Lett. **48**, 1046.

Bishop, D. J., D. C. Tsui, and R. C. Dynes, 1980, Phys. Rev. Lett. **44**, 1153.

Boebinger, G. S., A. M. Chang, H. L. Stormer, and D. C. Tsui, 1985, Phys. Rev. Lett. **55**, 1606.

Caillol, J. M., D. Levesque, J. J. Weis, and J. P. Hansen, 1982, J. Stat. Phys. **28**, 325.

Chang, A. M., P. Berglund, D. C. Tsui, H. L. Stormer, and J. C. M. Hwang, 1984, Phys. Rev. Lett. **53**, 997.

Clark, R. G., J. R. Mallett, S. R. Haynes, J. J. Harris, and C. T. Foxon, 1988, Phys. Rev. Lett. **60**, 1747.

de-Picciotto, R., M. Reznikov, M. Heiblum, V. Umansky, G. Bunin, and D. Mahalu, 1997, Nature (London) **389**, 162.

Dolan, G. J., and D. D. Osheroff, 1979, Phys. Rev. Lett. **43**, 721.

Feynman, R. P., 1972, *Statistical Mechanics* (Benjamin, Reading, Massachusetts).

Girvin, S. M., A. H. MacDonald, and P. M. Platzman, 1985, Phys. Rev. Lett. **54**, 581.

Glozman, I., C. E. Johson, and H. W. Jiang, 1995, Phys. Rev. Lett. **74**, 594.

Goldman, V. J., 1996, Surf. Sci. **361**, 1.

Goldman, V. J., and B. Su, 1995, Science **267**, 1010.

Haldane, F. D. M., 1983, Phys. Rev. Lett. **51**, 605.

Haldane, F. D. M., and E. H. Rezayi, 1985, Phys. Rev. Lett. **54**, 237.

Halperin, B. I., 1984, Phys. Rev. Lett. **52**, 1583.

Halperin, B. I., P. A. Lee, and N. Read, 1993, Phys. Rev. B **47**, 7312.

Jackiw, R., and C. Rebbi, 1976, Phys. Rev. D **13**, 3398.

Jain, J. K., 1989, Phys. Rev. Lett. **63**, 199.

Jain, J. K., A. Kivelson, and D. J. Thouless, 1993, Phys. Rev. Lett. **71**, 3003.

Kallin, C., and B. I. Halperin, 1984, Phys. Rev. B **30**, 5655.

Kane, C. L., and M. P. A. Fisher, 1994, Phys. Rev. Lett. **72**, 724.

Khmel'nitzkii, D. E., 1984, Phys. Lett. A **106**, 182.

Kivelson, S., D.-H. Lee, and S. Zhang, 1992, Phys. Rev. B **46**, 2223.

Laughlin, R. B., 1981, Phys. Rev. B **23**, 5632.

Laughlin, R. B., 1983a, Phys. Rev. Lett. **50**, 1395.

Laughlin, R. B., 1983b, Phys. Rev. B **27**, 3383.

Laughlin, R. B., 1984, Phys. Rev. Lett. **52**, 2034.

Laughlin, R. B., 1987, in *The Quantum Hall Effect*, edited by R. E. Prange and S. M. Girvin (Springer, Heidelberg), p. 233.

Leinaas, J. M., and J. Myrheim, 1977, Nuovo Cimento B **37**, 1.

Levine, H., S. B. Libby, and A. M. M. Pruisken, 1983, Phys. Rev. Lett. **51**, 1915.

Morf, R., and B. I. Halperin, 1986, Phys. Rev. B **33**, 1133.

Peskin, M. E., and D. V. Schroeder, 1995, *Introduction to Quantum Field Theory* (Addison-Wesley, Reading, Massachusetts).

Pines, D., and P. Nozières, 1966, *The Theory of Fermi Liquids* (Benjamin, New York).

Polyakov, A. M., 1974, Piz'ma Zh. Eksp. Teor. Fiz. **20**, 430 [JETP Lett. **20**, 1974].

Pople, J. A., and J. H. Walmsley, 1962, Mol. Phys. **5**, 15.

Prange, R. E., 1981, Phys. Rev. B **23**, 4802.

Rice, M. J., A. R. Bishop, J. A. Krumhansl, and S. E. Trullinger, 1976, Phys. Rev. Lett. **36**, 432.

Saminadayar, L., D. C. Glattli, Y. Jin, and B. Etienne, 1997, Phys. Rev. Lett. **79**, 2526.

Simmons, J. A., H. P. Wei, L. W. Engel, D. C. Tsui, and M. Shayegan, 1989, Phys. Rev. Lett. **63**, 1731.

Skyrme, T. H. R., 1961, Proc. R. Soc. London, Ser. A **262**, 233.

Su, W. P., J. R. Schrieffer, and A. J. Heeger, 1979, Phys. Rev. Lett. **42**, 1698.

Su, W. P., and J. R. Schrieffer, 1981, Phys. Rev. Lett. **46**, 738.

't Hooft, G., 1974, Nucl. Phys. B **79**, 276.

Trugman, S. A., 1983, Phys. Rev. B **27**, 7539.

Tsui, D. C., H. L. Stormer, and A. C. Gossard, 1982, Phys. Rev. Lett. **48**, 1559.

von Klitzing, K., G. Dorda, and M. Pepper, 1980, Phys. Rev. Lett. **45**, 494.

Wen, X.-G., 1990, Phys. Rev. Lett. **64**, 2206.

Wilczek, F., 1982, Phys. Rev. Lett. **48**, 957.