The fractional quantum Hall effect: Laughlin wave function, fractional charge and statistics.

The fractional QHE is evidently prima facie impossible to obtain within an independent-electron picture, since it would appear to require that the extended states be only partially occupied and this would immediately lead to a nonzero value of Σ_{xx}. What this suggests is that electron-electron interactions lead to some kind of gap in the spectrum of the extended states, analogous to the "cyclotron gap" $\hbar\omega_c$ for the IQHE, and that disorder then plays essentially the same role as in the integral case. It is convenient to concentrate on the limit of high magnetic field, so that the ratio α of the Landau-level spacing $\hbar\omega_c$ to the typical value of the Coulomb interaction $(e^2/\epsilon l_{\rm M})$ (which can be rewritten as $a_{\rm eff}/l_{\rm M}$, where $a_{\rm eff}$ is the effective Bohr radius $\frac{\hbar^2}{m^* e^2/\epsilon}$) is large compared to 1. Actually, for the systems in common use (Si MOSFET's and GaAs-GaAlAs heterostructures) we have $\alpha \equiv a_{\rm eff}/l_{\rm M} \sim$ 0.1-0.2 at 1 T, and since α increases only as $B^{1/2}$ we would need rather strong magnetic fields to reach this regime in practice; however, it is a useful simplification to consider it. If we indeed have $\alpha \gg 1$, then it is plausible that to a first approximation we need consider only the states in the last partially occupied Landau level (in particular, for $\nu < 1$ only the LLL). Then the kinetic energy $\hbar\omega_c$ falls right out of the problem and we have to worry only about the Coulomb energy and any impurity potentials. If the latter dominate, then presumably the relevant electron states are localized and the system is a conventional insulator. However, this need not be the case: in fact, since the length of the IQHE plateaux on the ν-axis is a measure of the width of the localized region of the band, it follows that when those lengths are small (i.e. the plateaux cover only a small fraction of the ν-axis) then the Coulomb interaction, while small compared to $\hbar\omega_c$, can still be $\gtrsim$ the impurity potentials.

Under these conditions the prime desideratum is to minimize the Coulomb energy. One obvious way to do this is to form a "Wigner crystal", that is a regular lattice of localized electrons, and the general belief is that for sufficiently small values of ν this is what the system actually does (though so far there has been no definitive observation of a Wigner-crystal state in the semiconductor systems - it has been seen in electrons on the surface of liquid He). However, fortunately both theory and experiment suggest that at values of ν not too small compared to 1 something much more interesting happens.

The ansatz originally written down by Laughlin for the ground state wave function (GSWF) of the FQHE state was based on an inspired guess, but it has subsequently been very well confirmed, at least for small number of electrons, by numerical solutions. I now give the essentials of Laughlin's argument, restricting myself for simplicity to the lowest Landau level and neglecting as usual spin or valley degeneracy.

We have seen (lecture 16) that a possible choice of eigenstates for the LLL is of the form (I omit the subscript "0" indicating that $n = 0$)

$$\psi_l(z) = z^l \exp -|z|^2/4l_{\rm M}^2 \quad (z \equiv x + iy,\ l_{\rm M}^2 \equiv \hbar/eB) \tag{1}$$

with $0 \leqslant l \leqslant N_\varphi$ ($N_\varphi \equiv A/\phi_o \equiv$ no. of flux quanta). It is clear that a general state of

the *many-body* system can be written in the form

$$\Psi(z_1, z_2 \dots z_N) = g(z_1, z_2 \dots z_N) \prod_i \exp -|z_i|^2/4l_{\rm M}^2 \tag{2}$$

with $\mathrm{g}(z_1, z_2 \dots z_N)$ a polynomial such that the maximum power of any z_i occurring is N_φ. A plausible choice which may satisfy this condition (cf. below) is the Jastrow form

$$g(z_1, z_2 \dots z_N) = \prod_{i<j} f(z_i - z_j) \tag{3}$$

with the condition, required for Fermi antisymmetry, that $f(z)$ is an odd function of z. This form is plausible as with an appropriate choice of f it will tend to keep the electrons apart and thereby reduce the Coulomb repulsion. Furthermore, in view of the rotational invariance of the Hamiltonian it is clear that we can, if we wish, choose the eigenfunctions to be simultaneously eigenfunctions of the total (canonical) angular momentum operator $-i\hbar \sum_i \partial/\partial\phi_i$. A little thought shows that this condition requires that $f(z)$ be a simple (odd) power of z, $f(z) = z^q$, (q odd). These arguments, if accepted, thus uniquely determine the form of the GSWF:

$$\boxed{\Psi(z_1, z_2 \dots z_N) = \prod_{j<k} (z_j - z_k)^q \exp - \sum_i |z_i|^2/4l_{\rm M}^2} \quad (q \text{ odd}) \tag{4}$$

This is the celebrated Laughlin ansatz for the GSWF of the FQHE; as already mentioned, it has received considerable support from numerical studies, which show that for small N the overlap with numerically computed GSWF is $> 99\%$. Note that for the special case $q = 1$, the Laughlin wave function is just an alternative way of writing the noninteracting GSWF we have had for the IQHE.

One crucial point to note about the Laughlin wave function is that *the filling factor is not arbitrary* but is uniquely fixed by the odd integer q. To see this, imagine that we write out the algebraic factor $\prod_{j<k}(z_j - z_k)^q$ explicitly as a sum of powers of the z_j. It is clear that the maximum power of any given z_j which can occur is just Nq where N is the total number of particles. But the power of z_j corresponds, in the radial-gauge choice of basis, to the l-value of the orbit occupied by the j-th particle, which we have seen enclosed l quanta of flux, and the maximum[1] of l is A/ϕ_0 where A is the area of the sample (assumed circular for simplicity). Hence we have $Nq = A\phi_0$, i.e. a filling factor $\nu \equiv N/\phi_0 A = 1/q$: or in words, q flux quanta per electron.

A second important feature of the Laughlin wave function is that despite the occurrence of the factor $\exp -|z_i|^2/4l_{\rm M}^2$, which appears to pick out the origin as special, the single-particle probability density which it describes is nearly uniform. Perhaps the easiest way to see this is to rewrite the argument of the exponent as

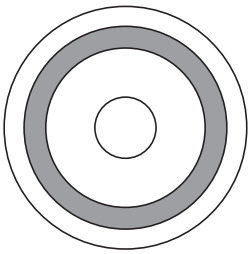
$$\sum_i |z_i|^2 \equiv \frac{1}{N} \sum_{i<j} |z_i - z_j|^2 + N|Z|^2 \tag{5}$$

[1] That is, the maximum value which allows the guiding center to remain within the sample radius. Of course we expect the states to be modified near the edge.

where $Z \equiv N^{-1}\sum_i z_i$ is the (complex) coordinate of the COM. It is then clear that apart from the factor $\exp -N|Z|^2/4l_{\rm M}^2$ the wave function, and hence the probability density, is a function only of the *relative* coordinates, so the single-particle probability density should be nearly uniform − in fact, just as uniform as it is for the uncorrelated wave function which describes the IQHE (which also contains the factor $\exp -N|Z|^2/4l_{\rm M}^2$, which is necessary to ensure that the density is zero for $\mid z_i \mid$ well outside the radius of the disk).

The most obvious question regarding the Laughlin state is: Does it yield the experimental result that the Hall conductance is quantized in units of $\nu e^2/h$, where $\nu = 1/q$ is the filling factor? The argument proceeds in parallel with that for the IQHE: Consider, as in the last lecture, a Corbino-disk geometry with an "impurity" region shielded by impurity-free rings, with a uniform magnetic field $\boldsymbol{B}$ plus a variable AB flux $\Phi_{\rm AB}(t)$ applied through the hole. Imagine that we slowly vary Φ_{AB}, over a time T, not by one but by q flux quanta. In this process, q orbits will have moved out through the outer edge of the disk and q in through the inner edge, and we will have restored the original situation. However, because the filling factor is $1/q$ rather than 1, a single electron will have left at the outer edge and entered at the inner one. Hence the current $I = e/T$ is related to the voltage $V = q\Phi_0/T$ by a Hall conductance $\Sigma_{\rm H} \equiv I/V$ given by

$$\Sigma_{\rm H} = e^2/hq \equiv \nu e^2/h \tag{6}$$



precisely as observed experimentally. The explanation of the finite length of the plateaux is then essentially as in the IQHE.

We can also try to adapt the Thouless topological argument given in the last lecture to the case of fractional ν (though Thouless himself does not do this). The argument proceeds as in lecture 17 up to the equation (eqn. (12) of lecture 17, where we integrate up to $2\pi q$ and thus divide by q)

$$\bar{\Sigma}_{\rm H} = \frac{ie^2}{2\pi hq}\int_0^{2\pi} d\phi_J \int_0^{2\pi q} d\phi_V \Big\{ \int d^N\mathbf{r}\big(\partial\Psi_0^*/\partial\phi_V\,\partial\Psi_0/\partial\phi_J - \partial\Psi_0^*/\partial\phi_J\,\partial\Psi_0/\partial\phi_V\big)\Big\} \tag{7}$$

However, at this point we need to postulate that the GSWF does *not* return to itself when $\phi_J \to \phi_J + 2\pi$, $\phi_V \to \phi_V + 2\pi$ but only when $\phi_J \to \phi_J + 2\pi$, $\phi_V \to \phi_V + 2\pi q$. This is actually a delicate point, since there is a general theorem (the Byers–Yang theorem) which tells us that both the energy levels and the wave functions of the many-body system must be invariant under the former transformation (as well as, a fortiori, under the latter one). The solution is that the "unwanted" ground states may not be accessible over any reasonable timescale (cf. Thouless and Gefen, PRL **66**, 806 (1991), and lecture 19). In that case all we can conclude is that the double integral has the value $-2\pi in$ where n is integral; hence, we conclude

$$\bar{\Sigma}_{\rm H} = \frac{ne^2}{hq} = n\nu e^2/h \tag{8}$$

It is clear clear that this argument raises some issues additional to the ones which occur in the original Thouless argument, so it perhaps cannot claim (even) the same degree of rigor.

Returning now to our earlier Corbino-disk argument, this raises the obvious question: What if we were to decide to change the AB flux not by q but only by *one* flux quantum? Before answering this question, let's address a related one, namely: Suppose we take a uniform circular disk and apply a magnetic field $\boldsymbol{B}$ such that the filling is exactly $\nu = 1/q$ where q is integral; then we expect that the groundstate of the system is described to a good approximation by the Laughlin wave function. Now imagine that we increase $\boldsymbol{B}$ by a small amount as to introduce exactly one extra flux quantum in the one of area of the disk. How is the many-body wave function (MBWF) modified? It is clear that the l-value of the outermost Landau orbit has increased by 1. It is plausible (though not perhaps totally self-evident) that the system will respond so as to keep the occupation of this and neighboring orbits unchanged. However, this requires that the maximum power of z_j in the MBWF is no longer $l_{\max}$ $(= Nq)$ but rather $l_{\max} + 1 = Nq + 1$. Thus, we must take the original Laughlin state and multiply it by the (symmetric) function $\prod_i z_i$:

$$\Psi' = \left[\prod_{i=1}^{N} z_i\right] \Psi_0(z_1, z_2 \ldots z_N) \tag{9}$$

where Ψ_0 is the Laughlin wave function.[2]

What have we done? As mentioned above, despite appearances the Laughlin wave function does not really pick out the origin as "special", in fact the one-particle density is constant over the disk up to within $\sim l_{\mathrm{M}}$ of the edges. However, the extra factor in (9) clearly reduces to zero the probability of finding an electron at the origin, and (when combined with the usual exponential factor) depresses it over a region of dimension $\sim l_{\mathrm{M}}^2$ around the origin; in effect, we have achieved the desired result, namely that the "extra" flux quantum has no electron associated with it. We have created a "hole"! It is clear that there is nothing special about the origin, and we would equally well have created a hole at the point z_0 by generalizing (9) to

$$\Psi'(z_0) \equiv \left[\prod_{i=1}^{N} (z_i - z_0)\right] \Psi_0(z_1, z_2 \ldots z_N) \tag{10}$$

Provided z_0 is not within $\sim l_{\mathrm{M}}$ of the edge of the disk, this leaves the occupation of the states at the edge essentially unchanged.

We now raise the crucial question: What is the "effective charge" of the hole we have added? The most direct way of answering this question would be to evaluate the single-particle density near the origin in the original Laughlin state and in the state Ψ', multiply by the electron charge, subtract the former from the latter and integrate over a large ($\gg l_{\mathrm{M}}$) region around z_0. This is straightforward but tedious. A more intuitive

[2] Or more precisely the Laughlin wave function adjusted for the slightly changed value of l_{M} (an effect of order N^{-1}).

way of getting the result is to imagine that we have introduced q such holes at the same point and at the same time increased B so that q flux quanta are added, so that the MBWF is

$$\Psi_{q\,\mathrm{holes}} = \prod_{i=1}^{N}(z_i - z_0)^q \Psi_0(z_1, z_2 \ldots z_N) \tag{11}$$

Now suppose that we introduce an $N+1$-th electron, assign to it a wave function $\delta(z_{N+1} - z_0)\exp -|z_{N+1}|^2/4l_{\mathrm{M}}^2$, and form the $N+1$-electron wave function by integrating over z_0, i.e.

$$\Psi_{0\,(N+1)} \equiv \int dz_0 \prod_{i=1}^{N}(z_i - z_0)^q \Psi_{0\,(N)}(z_1, z_2 \ldots z_N)\delta(z_{N+1} - z_0)\exp -|z_{N+1}|^2/4l_{\mathrm{M}}^2 \tag{12}$$

It is easy to see that the explicit form of $\Psi_{0,N+1}(z_1, z_2 \ldots z_N)$ is

$$\Psi_0(z_1, z_2 \ldots z_{N+1}) = \prod_{i=1}^{N+1}(z_i - z_0)^q \exp - \sum_{i=1}^{N+1}|z_i|^2/4l_{\mathrm{M}}^2 \tag{13}$$

that is, it is exactly of the form of the Laughlin wave function for $N+1$ electrons (and $q(N+1)$ flux quanta, since we recall we added q quanta). In other words, the addition of the extra electron has "cancelled" q added holes. Consequently, we conclude that the charge e^* of the hole is given by

$$e^* = -e/q \tag{14}$$

-fractional charge!

The physical interpretation is that the MBWF has changed in such a way that the average probability of finding an electron near the origin has been reduced by 1/3 (and this is, of course, confirmed by the quantitative calculation of $\rho(r)$).

Where has the missing probability density gone? It has not gone, as one might perhaps guess, to the outer edge of the disk, since we constructed our trial wave function precisely so as to leave the occupation of the states (when labeled, say, by their distance from the edge) near this edge unchanged. Rather, is has been delocalized over a wide range ($\sim R$) of r, since the probability density has shifted slightly far the states with l of order (but not equal to) $l_{\max}$. Of course, as in the case of the "fractional charge" associated with $(CH_2)_x$, the question arises whether this is a "sharp" variable. I do not know of any paper which explicitly addresses this question, but I would bet that the answer is similar to that given by Rajaraman and Bell[3] for $(CH_2)_x$, namely that to make the charge variable "sharp" (i.e. to ensure that quantum fluctuations are negligible compared to the average value) requires one to weight, in the definition of total "charge", the density with some function $f(z)$:

$$Q = \int \hat{\Psi}^+(z)\hat{\Psi}(z)f(z - z_o)dz, \int f(z)dz = 1 \tag{15}$$

[3]Phys. Lett. **116B**, 151 (1982): cf. also Kivelson and Schrieffer, Phys. Rev B **25**, 6447 (1982)

where $f(z)$ varies significantly over a distance $\gg l_M$ but $\ll$ the disk radius.

So far, we have seen how to modify the MBWF so as to accommodate an *increase* of the number of flux quanta by one. What if we *decrease* it by 1? Then, intuitively, we have to introduce an object with fractional charge $+e/q$, or alternatively to *increase* the probability density near the origin (or near an arbitrary point z_0). Following an argument similar to the above one, we need to *decrease* the maximum power of z_j by one. If the extra "negative charge" is to be created at the origin, the obvious way to do this is to operate on the polynomial part of the Laughlin wave function (but not the exponential) with $\partial/\partial z_i$ for each i, i.e.

$$\Psi'_- \sim \prod_{i=1}^{N} \left\{ \exp -|z_i|^2/4l_{\mathrm{M}}^2 \, \frac{\partial}{\partial z_i} (\exp +|z_i|^2/4l_{\mathrm{M}}^2) \right\} \Psi_0(z_1, z_2 \ldots z_N) \tag{16}$$

(the Ψ_0 obviously occurs only once - there is a slight problem of notation here!)

The generalization to arbitrary positions z_0 of the quasiparticle is a little more tricky: evidently the expression has to depend on z_0, but since Ψ_0 does not contain z_0 the obvious choice, $\partial/\partial z_0$, is not an option. The next simplest choice is to replace $\partial/\partial z$ by $\partial/\partial z - f(z_0)$, and the simplest choice of $f(z_0)$ which has the correct dimensions is const. z_0^*/l_{M}^2. In fact, a detailed analysis (cf. Yoshioka section 2.2.4) of the Landau states in terms of an annihilation- and creation-operator formalism indicates that the correct procedure is to replace $\partial/\partial z_i$ by $\partial/\partial z_i - \frac{1}{2} z_0^*/l_M^2$; the form in which the one-quasiparticle MBWF is usually written (in terms of a dimensionless variable z_i measured in units of l_{M}) is therefore up to normalization

$$\Psi'_-(z_0) = \prod_{i=1}^{N} \left(\exp -|z_i|^2/4 \, (2\partial/\partial z_i - z_0^*) \, \exp -|z_i|^2/4 \right) \Psi_0(z_1, z_2 \ldots z_N) \tag{17}$$

with effective charge $e^* = +e/q$ localized near z_0 (and an equal spread-out negative charge).

The original Laughlin wave function was an ansatz designed to minimize repulsive Coulomb energy, and it achieves this, by keeping the *average* density constant while introducing repulsive correlations. The introduction of a quasihole or quasiparticle spoils this result and could be expected to cost some extra energy, which on dimensional grounds would be expected to be of the order of $e^2/\epsilon l_{\mathrm{M}}$ (in cgs units). This can be evaluated by calculating the appropriate averages of $(r_i - r_j)^{-1}$ etc., with the result that it is $\sim 0.025\, e^2/\epsilon l$ for a quasihole and $\sim 0.075\, e^2/\epsilon l$ for a quasiparticle. For experimentally realistic parameters this may be small compared to $\hbar\omega_c$. Since *any* "compression" of the system by increasing or decreasing ν away from the commensurate value $1/q$ (either by supplying/taking away electrons or by changing the magnetic field or both) requires the creation of quasiparticle or quasihole, it follows that there is a finite energy gap for such a change, and the fractional quantum Hall state is therefore said to be "incompressible". This is true within the naive model we have used so far: when we come to study the behavior at the edges of the sample we will see that in a sense a nonzero compressibility is realized there.

We are now in a position to answer our original question, namely, what happens when, in a Corbino-disk geometry, we adiabatically increase the AB flux by *one* flux quantum? The first point to make is that after such an increase one can perform a gauge transformation so as to return the Hamiltonian to exactly its original form, so that the true ground state must be unchanged. However, it certainly does not follow that under such and adiabatic change the system will automatically attain its true groundstate, any more than it follows for a superconducting ring under the same operation. What in fact may well happen is that the original groundstate evolves, under the adiabatic perturbation, into an excited state, and it is easy to guess what this will be, if the edges of the disk are open-circuted: we will produce a "fractional hole" at the inner edge, and the resultant state, although not the true groundstate, may be very metastable. (Cf. the Thouless-Gefen paper cited above).

What about density fluctuations around the Laughlin state? Unlike the case of a Fermi liquid where we can produce quasiparticle-quasihole pairs with arbitrary small energy, the only way to create such a fluctuation is to produce a *fractional* quasiparticle-quasihole pair, and the sum of the relevant energies is, as we have seen, always nonzero. However, one can think of forming an "exciton" out of a nearby qp-qh pair, and one would think that the Coulomb attraction between the qp and qh should lower the energy. This turns out to be true, but detailed calculation shows that it is never lowered to zero. In fact, for $r \gg l_{\rm M}$ one can argue phenomenologically as follows (cf. Yoshioka section 4.5.2): Suppose the quasielectron and quasihole circle one another at a definite separation r with velocity v. Balancing the Lorentz force $e^* vB$ with the Coulomb attraction $e^{*2}/4\pi\epsilon\epsilon_0 r^2$ gives $v = (1/4\pi\epsilon\epsilon_0)e^*/Br^2$. To obtain agreement with the result that $\partial E/\partial p \equiv \partial E/\partial(\hbar k) = v$, where

$$E = \epsilon_{\rm qp} + \epsilon_{\rm qh} - e^{*2}/4\pi\epsilon\epsilon_0 r \tag{18}$$

we set $\hbar k = e^* Br$; then, using the definitions of $l_{\rm M}$ and ν, we find

$$E(k) = \epsilon_{\rm qp} + \epsilon_{\rm qh} - e^{*2}\nu^3/4\pi\epsilon\epsilon_0 k l_{\rm M}^2 \tag{19}$$

One cannot take this seriously for $kl_{\rm M} \lesssim 1$, and more detailed calculations show that for $k \to 0$ the energy is finite and of order $\epsilon_{\rm qp} + \epsilon_{\rm qh}$. Nevertheless $E(k)$ does appear to have a minimum at $k \sim l_{\rm M}^{-1}$ – the "magnetoroton".

In the last part of this lecture I would like to address what is probably the most intriguing theoretical prediction concerning the Laughlin quasiparticles occurring in the FQHE, namely that they obey *fractional statistics*; this is perhaps the property that most fundamentally reflects the two-dimensionality of the underlying physical system.

Let's start with the observation first made in a seminal 1977 paper by Leinaas and Myrheim[4]: For particles moving in a *strictly* 2D physical space, the standard argument which leads, in 3D, to the necessity of either Bose or Fermi statistics does not apply. Let us review that argument briefly (cf. lecture 1): Consider two identical particles specified by coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$ (I neglect spin for simplicity). Since no physical quantity can

[4]Nuovo Cimento, **37B**, 132 (1977).

depend on whether it is particle 1 which is at $\mathbf{r}_1$, and 2 at $\mathbf{r}_2$ or vice versa, it follows (inter alia) that the probability $p(\mathbf{r}_1, \mathbf{r}_2) \equiv p(\mathbf{r}_2, \mathbf{r}_1)$, i.e.

$$|\Psi(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\Psi(\mathbf{r}_2, \mathbf{r}_1)|^2 \tag{20}$$

so that the probability amplitude (wave function) must satisfy

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \exp(i\alpha)\, \Psi(\mathbf{r}_2, \mathbf{r}_1) \tag{21}$$

where α may be any real number including zero. Call the argument leading to eqn. (21) step 1. Next, we observe that the operation of exchanging particles 1 and 2 *twice* (with the same sense) is equivalent, apart from a translation which is irrelevant in the present constant, to moving 1 (clockwise or anticlockwise, depending on the "sense" of the exchange path) around 2 and back to its original position: thus, under this operation we have, if $\Psi_{\rm rot}(\mathbf{r}_1, \mathbf{r}_2)$ denotes the final state so achieved,

$$\Psi_{\rm rot}(\mathbf{r}_1, \mathbf{r}_2) = \exp(2i\alpha)\, \Psi(\mathbf{r}_1, \mathbf{r}_2) \tag{22}$$

(step 2). Finally, we note that in 3 (or more) spatial dimensions such an "encirclement" operation can always be progressively deformed into the identity *without passing through the relative origin* $\mathbf{r}_1 = \mathbf{r}_2$, and since α is property only of the particles, not of the paths, we must thus have (step 3)

$$\Psi_{\rm rot}(\mathbf{r}_1, \mathbf{r}_2) \equiv \Psi(\mathbf{r}_1, \mathbf{r}_2) \tag{23}$$

Putting together eqn.s (21-23), we conclude that

$$\alpha = 0 \text{ or } \pi \tag{24}$$

the two possibilities corresponding to the familiar Bose and Fermi "statistics". Thus in 3 or more dimensions these are the *only* possible behaviors of the wave function of identical particles under exchange.

What Leinaas and Myrheim in effect pointed out is that for a system which is *strictly* 2-dimensional, while step 1 and 2 in the above argument are still valid, *step 3 need not be* (since in 2D it is not possible to deform the "encirclement" into the identity without passing through the relative origin). Thus, the "exchange phase" α which occurs in eqn. (21) can be any (real) number. Subsequently, particles having a value of α different from 0 or π were christened "anyons", and have been widely studied for their own sake in the mathematical-physics literature.

Consider now two Laughlin quasiholes in the FQHE, say for definiteness with $\nu = 1/3$, located at positions w_1 and w_2. The relevant MBWF, which perfectly satisfies the condition of Fermi antisymmetry for the *electrons* (which of course really "live" in 3D!) is

$$\Psi_{2h} = \prod_{i=1}^{N} \prod_{j=1}^{N} (z_i - w_1)(z_j - w_2) \Psi_0(z_1, z_2 \ldots z_N) \tag{25}$$

where $\Psi_0(z_1, z_2 \ldots z_N)$ is the Laughlin wave function for the groundstate. The wave function (25) is, trivially, symmetric under the exchange of the quasihole coordinates w_1

and w_2, so at first sight the holes are just bosons. However, let us now ask the following question: Suppose we "pin" the two holes in some way with some kind of external control[5] and use the latter to exchange their positions *adiabatically* (in practice, over a timescale long compared to $\hbar/E_{\rm min}$ where $E_{\rm min}$ is the minimum excitation energy of the system, which we may estimate as of the order of the quasihole formation energy and thus $\sim e^2/\epsilon l_{\rm M}$). The MBWF will then evolve adiabatically (with no transitions to excited states, so that in particular the electrons never have a chance to move "into the third dimension") and thus must satisfy (cf. step (1) of the above argument)

$$\Psi_{\rm f}(z_1, z_2 \ldots z_N; w_1, w_2) = \exp i\alpha \, \Psi_{\rm in}(z_1, z_2 \ldots z_N; w_2, w_1) \tag{26}$$

(where $\Psi_{\rm in}$, $\Psi_{\rm f}$ denote respectively the wave function before and after exchange). The crucial question is: what is α? According to the understanding of this problem which is by now more or less standard, the answer is that for any simple Laughlin state we have for (say) clockwise exchange

$$\alpha = \nu\pi \tag{27}$$

so for the specific case considered ($\nu = 1/3$), $\alpha = \pi/3$. Thus a Laughlin quasihole (and also a quasiparticle) is indeed an "anyon". Note that if α is $+\pi/3$ for clockwise exchange, it must be necessarily be $-\pi/3$ for anticlockwise exchange, since the sequential product of the two processes is the identity. That the system "knows the difference" between clockwise and anticlockwise processes is not too surprising, since the magnetic field necessary to stabilize the QHE breaks parity.

To understand the result (27), it may be helpful to consider an analogous problem involving only two particles, which in turn can be understand by looking at two effects which can be illustrated in a simple one-particle example. The first effect, which we have already met, is "fractional charge" (or probability): to illustrate this notion we consider a single charged particle restricted to tunnel between the groundstate of two neighboring potential wells. Suppose we make a "projective measurement" of the charge Q_R to the right of the line $z = 0$ at some definite time t; then of course we will always find $Q_R = 1$ or 0. However, suppose we allow the system to relax to its groundstate, namely

$$\psi_0 = \cos\theta|R\rangle + \sin\theta|L\rangle, \quad \tan\theta \equiv \epsilon/\Delta \tag{28}$$

If we then perform a series of "weak" measurements over time so as to establish the *average* value of Q_R, then we will find a nonintegral result, $\langle Q_R\rangle = \cos^2\theta$ (which is of course what we would calculate). This does not seem particularly mysterious.

The second effect we need has to do with Berry's phase. Suppose that a QM system, say for definiteness a single particle, is subject to some classical control parameter λ (in general of the nature of a *vector* or something similar), so that its energy eigenfunctions, and in particular the GSWF, is a function of λ:

$$\Psi_0 = \Psi_0(\mathbf{r}, \lambda) \tag{29}$$

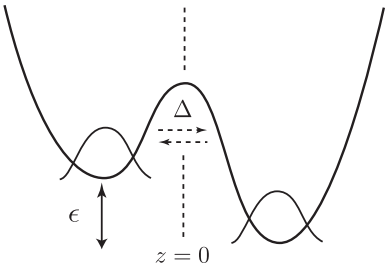
[5] e.g. an STM tip.

Suppose now that the parameter λ is varied *slowly* in time (i.e. over timescales long compared to $\hbar/E_{\rm min}$ where $E_{\rm min}$ is minimal excitation energy). If after some time T λ returns to its original value, then by out previous argument we must have

$$\Psi_{\rm fin}(\mathbf{r},\lambda) = \exp i\alpha\, \Psi_{\rm in}(\mathbf{r},\lambda) \tag{30}$$

Part of the phase increment α is a "dynamical" phase $\left(\hbar^{-1}\int_0^T E(\lambda(t))\,dt\right)$ which depends on the actual time-dependence of $\lambda(t)$ (and on the arbitrary zero of energy). However, in general there may be a second contribution to α which is independent of the detailed dependence of $\lambda(t)$ and is a function *only of the path followed* (recall in the "interesting" case λ is of the nature of a vector or something similar). This is the celebrated *Berry phase*.

It is easy to obtain a formal expression for the Berry phase $\phi_{\rm B}$ in terms of the variation of the groundstate[6] under a small adiabatic change $\delta\lambda$ of the control parameter λ. For such a change we evidently have (since Ψ_0 is normalized)



$$\delta(\arg\Psi_0) = {\rm Im}\,\langle\Psi_0|\delta\Psi_0\rangle = {\rm Im}\,\langle\Psi_0|\partial\Psi_0/\partial\lambda\rangle\delta\lambda \tag{31}$$

and hence

$$\phi_{\rm B} = {\rm Im}\oint d\lambda\,\langle\Psi_0|\partial\Psi_0/\partial\lambda\rangle \tag{32}$$

Note that it is irrelevant to the argument whether the expression on the RHS of (31) has a real part (although the corresponding contribution to (32) must vanish).

A very standard example of a nontrivial Berry phase is that of a spin-1/2 in magnetic field which is oriented in direction specified by θ,ϕ such that θ is constant in time but ϕ is rotated adiabatically (i.e. over a timescale $\ll(\mu B)^{-1}$ from 0 to 2π (see figure). For given θ,ϕ the groundstate satisfies

$$\boldsymbol{n}\cdot\hat{\boldsymbol{\sigma}}|\psi\rangle = |\psi\rangle \tag{33}$$

so if we require that the two component of the spinor wave function be single-valued as a function of ϕ, then the (almost) unique solution is, up to an overall complex but ϕ-independent constant,

$$|\Psi\rangle = \begin{pmatrix} \cos\theta/2 \\ \sin\theta/2\,\exp i\phi \end{pmatrix} \tag{34}$$

If we plug this form of wave function into formula (32) we find

$$\phi_{\rm B} = \int_0^{2\pi} d\phi\,\sin^2\theta/2 = \pi(1-\cos\theta) \tag{35}$$

where the RHS is the solid angle subtraveled by the "orbit" of the field; note that putting $\theta=\pi/2$ (i.e. rotating $\mathbf{B}$ through 360°) gives the standard factor of -1 in the state of

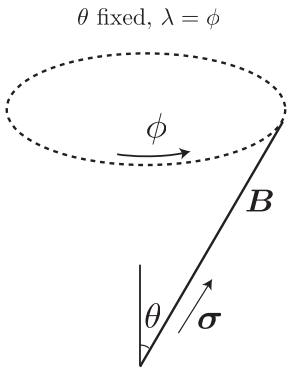
[6] Or any other energy eigenstate.

a spin-1/2 particle. (Note that had we made the other possible "single-valued" choice, namely

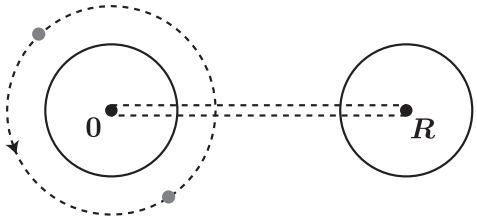
$$|\Psi\rangle = \begin{pmatrix} \cos\theta/2 \, \exp -i\phi \\ \sin\theta/2 \end{pmatrix} \tag{36}$$

we would have got $\phi_{\rm B} = -\pi(1+\cos\theta)$, which is equivalent to (35) modulo 2π).

Now let us put together the results of the two considerations above. As an example, imagine a quantum particle with a coordinate $\mathbf{r} - \boldsymbol{a}$ which can be localized either at point $\mathbf{0}$ or at point $\boldsymbol{R}$ (or be in a quantum superposition of those two states). Further, imagine a second particle (with coordinate $\mathbf{r}_b$) which is constrained by energy considerations to be in a p-state relative to $\boldsymbol{a}$, i.e. the relative wave function must have the form (up to irrelevant dependence on $|\mathbf{r}_b - \mathbf{r}_a|$) $\exp i\phi$, where $\phi \equiv \arg|\mathbf{r}_b - \mathbf{r}_a|$. Then the most general wave function of the two-particle system will be of the form



$$\Psi(\mathbf{r}_a, \mathbf{r}_b) = a\Psi_{\mathbf{0}} \exp i\phi_{\mathbf{0}}(\mathbf{r}_b) + b\Psi_{\boldsymbol{R}} \exp i\phi_{\boldsymbol{R}}(\mathbf{r}_b) \tag{37}$$



where $\phi_{\mathbf{0}} \equiv \arg \mathbf{r}_b$, $\phi_{\boldsymbol{R}} \equiv \arg(\mathbf{r}_b - \boldsymbol{R})$. This wave function is of course perfectly single-valued with respect to $\mathbf{r}_b$. Now suppose we "pin" $\mathbf{r}_b$, e.g. with a stray external potential, and move it adiabatically around $\mathbf{0}$ as shown in the figure. ("Adiabatically" in this example means "on a timescale long compared to the inverse of the matrix element for the transition of $\mathbf{r}_a$ from $\mathbf{0}$ to $\boldsymbol{R}$). What is the resultant phase change (Berry phase) $\phi_{\rm B}$? We may parametrize the dashed path in the figure by $\lambda \equiv \phi_{\mathbf{0}}$; then we have

$$\frac{\partial \Psi_0}{\partial \lambda} = i\left\{a\Psi_{\mathbf{0}} \exp i\phi_{\mathbf{0}} + b\Psi_R \frac{\partial \phi_{\boldsymbol{R}}}{\partial \phi_{\mathbf{0}}} \exp i\phi_{\boldsymbol{R}}\right\} \tag{38}$$

Since $\Psi_{\mathbf{0}}$ and $\Psi_{\boldsymbol{R}}$ are each assumed orthogonal and normalized, this means

$$\mathrm{Im}\,\langle\Psi|\partial\Psi/\partial\phi_{\mathbf{0}}\rangle = |a|^2 + |b|^2 \partial\phi_{\boldsymbol{R}}/\partial\phi_{\mathbf{0}} \tag{39}$$

and the Berry phase $\phi_{\rm B}$ is given by

$$\phi_{\rm B} = 2\pi|a|^2 + |b|^2 \int_0^{2\pi} (\partial\phi_{\boldsymbol{R}}/\partial\phi_{\mathbf{0}})\, d\phi_{\mathbf{0}} \tag{40}$$

However, since the path does not encircle $\boldsymbol{R}$, it is clear that the integral in (40), which is simply the total change in $\phi_{\boldsymbol{R}}$ due to the encirclement of $\mathbf{0}$, is zero. Hence

$$\phi_{\mathrm{B}} = 2\pi|a|^2 \neq 2\pi \tag{41}$$

or in words, since the mean charge enclosed by the path has the fractional value $|a|^2$,

$$\boxed{\text{statistical phase} = 2\pi \times \text{(fractional) average charge enclosed by path}} \tag{42}$$

Thus, it is natural that the phase acquired by one Laughlin quasiparticle encircling another is $2\pi\nu$, and since this is *two* exchange processes, the exchange phase is $\pi\nu$.