Models of the fractional quantum Hall effect and of high-temperature superconductivity have followed the recognition that in 2-D systems quantum statistics extends beyond bosons and fermions

Anyons for anyone

FRANK WILCZEK

Quantum mechanics gives a sharp new meaning to the concept of identicalness or indistinguishability, which is not present in classical physics. Its tangible manifestation is in the existence of special "effective interactions" or "statistical forces" between exactly identical particles, that do not exist between particles which are merely very similar.

For example, ^{3}He and ^{4}He atoms are extremely similar in their interactions at low energy (i.e. their chemistry). For almost any particle X, the low-energy scattering of X from ^{3}He and from ^{4}He is essentially the same. The only real exception to this rule arises when X is one of ^{3}He or ^{4}He themselves. The three scattering processes, ^{3}He–^{3}He, ^{3}He–^{4}He, ^{4}He–^{4}He differ not merely by small corrections, but qualitatively. For example, their scattering cross-sections at 90° are in the ratio 0:1:2. This difference is one concrete manifestation of the statistical forces.

Quantum statistics

The special behaviour of exactly identical particles can be understood to arise from the deep structure of quantum theory, in the following way. Let there be two indistinguishable particles A and B. The burden of quantum mechanics is to assign a probability amplitude to each possible process involving A and B. For example, one must assign an amplitude to the process by which A propagates from (x_A, $t = 0$) at time 0 to (x'_A, t_1) at time t_1 while B similarly propagates from (x_B, $t = 0$) to (x'_B, t_1), along a given trajectory (see figure 1). The total amplitude for A and B to propagate between the given points is given as the sum over all possible trajectories with these end points. The principle that one must add the amplitudes for all possible ways of achieving a given physical result is a basic foundation of quantum theory. The idea of summing over all trajectories lies at the heart of Feynman's path integral formulation of quantum theory.

In this formulation, the reason that an essentially new feature appears if A and B are indistinguishable can be made especially transparent. If A and B are indistinguishable, then a process whereby A ends up at x'_B while B ends up at x'_A has a final configuration which cannot be differentiated from the final configuration in the processes we considered before, where A ends up at x'_A and B at x'_B. Thus, according to the general principles of quantum mechanics, the amplitudes for both such possibilities must be added together. One says that one must consider, along with any *direct* process, the *exchange* process wherein the particles have, so to speak, swapped identity.

It is important to recognise that this new feature in the mechanics of identical particles has no analogue in classical physics. In classical physics, one does not sum over "all possible" trajectories in any sense. Given the initial conditions there is just one possible trajectory, namely the one that satisfies the equations of motion.

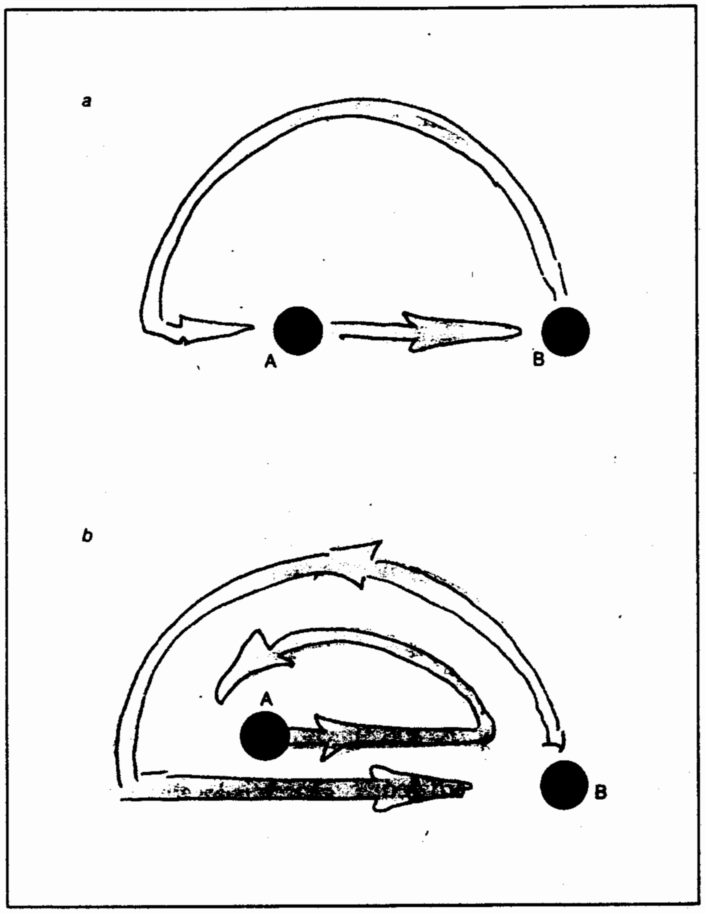
In fact, the contrast between classical and quantum theory for indistinguishable particles is even more drastic than this argument might suggest. Normally, the laws of classical physics provide powerful guidance in the choice of the correct quantum mechanical laws. This is particularly evident in the path integral approach to quantum mechanics, for in that formulation the amplitude assigned to a trajectory is completely determined by classical physics. (The amplitude is equal to the exponential of $2\pi i$ times the classical action divided by Planck's constant, to be precise.) However, by contrast, classical physics provides no guidance whatsoever, even in principle, for the formulation of rules for how we should weight the amplitudes of exchange processes relative to direct processes. This is because the classical equations of motion can be derived by demanding that the action is an extremum (for short trajectories, a minimum) along the actual classical trajectory. The extremum principle, or principle of least action, says that the action for paths which satisfy the classical equations of motion is not changed to first order by *infinitesimal* changes in the path. It has nothing whatsoever to say about paths which cannot be related by a series of infinitesimal changes. In particular, one can change the relative action for paths which cannot be joined continuously in a completely arbitrary way, without altering the classical equations of motion at all.

Now clearly the trajectories involving direct and exchange processes cannot be joined continuously to one another: the particles have either changed their identity or not, and one cannot interpolate continuously between these two discrete logical alternatives. The relative weight which we assign these paths, their relative amplitude, does not affect the classical equations of motion. Thus classical physics gives us no guidance whatsoever as to the choice of weight.

Fortunately, the absence of classical guidance does not leave us totally at sea. The framework of quantum mechanics is so tight that rather few possible choices are consistent with its structure. In particular, one must be consistent with the rule that the amplitude for two successive trajectories is the product of the amplitudes for each trajectory separately. Figure 1 shows the power of this rule in the present context. If one follows an exchange process by a second exchange process, clearly the overall process involves no exchange. Indeed, the double exchange trajectory can be related continuously to the entirely trivial straight-through trajectory, as is shown in the figure. Therefore weighting the trajectories of exchange relative to direct processes by a factor β is only consistent with the product rule for successive trajectories if $\beta^2 = 1$. This means, of course, that $\beta = \pm 1$ are the only two possible cases. They correspond to the familiar possibilities of bosons and fermions.

Beyond bosons and fermions

One might be tempted to stop at this point, satisfied that one has understood the existence of bosons and fermions in a basic and elementary way, and shown that these are the only possibilities. However, a closer examination of the argument reveals a surprising loophole. The unravelling depicted in the figure requires that there are at least *three* spatial dimensions in which to manoeuvre. Leaving aside for a moment the question why one might want to consider such a thing, let us notice that the situation in *two* dimensions is completely different. In two dimensions, it is not the case that repetition of the exchange process leads to a trivial direct process. In two dimensions, one can distinguish unambiguously not only whether particles have swapped identities, but also how many times they have wound around one another. (The number of windings can be visualised as the number of braids formed by their two world-lines.) The requirement that $\beta^2 = 1$ can no longer be derived, since the double exchange is not continuously related to the trivial direct process – the double exchange is not trivial, since it involves a full winding of one particle around the other. In the absence of this consistency requirement, more general choices for β are allowed. In two dimensions, then, one is no longer restricted to bosons and fermions as the only consistent possibilities for quantum statistics. There is a continuous range of possibilities in between.



1 The basic exchange process is depicted in (*a*). If we iterate this process, taking care that the trajectories do not cross, we arrive at (*b*). In three dimensions, the big loop may be lifted out of the plane and continuously shrunk to a point, without touching the small loop. Of course, the small loop may also be shrunk, and so the iterated exchange is continuously related to a trivial "non-motion". However, in two dimensions there is no way to disentangle the loops in (*b*)

Particles obeying the more general forms of statistics are called generically *anyons*. The name derives from the fact that they are allowed to have *any* value for the weight β.

Visualising anyons

When faced with a genuine novelty such as the existence of new forms of quantum statistics, it is good intellectual strategy to relate it in any way possible to more familiar things. The peculiar property of the statistical interaction of anyons, that they feel no classical force but do sense one another's presence by acquiring phase as they interweave, is reminiscent of another famous phenomenon of quantum physics. I have in mind the Aharonov–Böhm effect. This, you may recall, is the effect that the interference pattern of a charged particle – which behaves, of course, like a wave – can be altered by the existence of a thin solenoid (i.e. a tube of magnetic flux) encircled by the interfering paths, even if neither passes through the solenoid. This occurs despite the fact that there is no magnetic field outside the flux tube, so that the charged particle feels no classical force whatsoever. The Aharonov–Böhm effect is a purely quantum effect. It comes about because the phases of the amplitudes for trajectories winding around the solenoid are influenced by its presence. The phase change is proportional to the amount of winding, and also to the strength of the magnetic flux through the solenoid.

The interaction popularised by Aharonov and Böhm evidently bears a close mathematical kinship to the statistical interaction that characterises anyons. This kinship is the basis for a way of thinking about anyons that has proved extremely fruitful. The key idea is to imagine that anyons are bosons or fermions that happen also to carry fictitious charge and flux. (The use of the terminology "fictitious fields" is merely to avoid possible confusion between these fields, introduced purely as a mathematical device, and real electromagnetic fields.) If the fictitious charge and flux assignments are appropriately chosen, the Aharonov–Böhm phases they induce will precisely mimic the effect of anyon statistics.

Thus one has the remarkable freedom to trade quantum statistics for concrete dynamical properties, namely fictitious charge and flux. This form of alchemy is known as statistical transmutation. Using it, one can realise bosons as fermions with special additional interactions, or vice versa, or either of them as anyons of any specified type with appropriate fictitious charge and flux attached. Note that it is only in two dimensions where one may take a section perpendicular to the solenoid, and construct a *point* flux source. In three dimensions magnetic field lines cannot terminate, and there is no way of having isolated regions enclosing non-zero net flux. Thus the alchemy for transmuting the statistics of point particles is only available in two space dimensions.

Another necessary requirement for fractional statistics is that the discrete symmetries P and T must be broken. Indeed either spatial reflection or time reversal reverse the sense of winding. The validity of these symmetry operations therefore requires that the phase associated with a clockwise winding should be the same as the phase associated with a counterclockwise winding. However, if β is the phase for the clockwise winding, then the phase associated with the counterclockwise winding must be β^{-1} (since a clockwise winding followed by a counterclockwise winding produces a trajectory that may be continuously deformed to the trivial straight trajectory, the product rule for amplitudes forces this value). Thus, only $\beta = \beta^{-1}$, that is $\beta = \pm 1$ or in other words bosons or fermions, are consistent with the discrete symmetries P and T.

Anyons in nature

You may well be tempted to think, at this point, that anyons are quite fanciful. Particles which can only exist in two dimensions – and even there only when P and T symmetry are violated – might seem to be more appropriate as candidate inmates of some fantasy twisted Flatland, than as possible elements of the real (three dimensional, and very nearly symmetric) world. Upon careful consideration, however, one comes to realise that these apparently daunting obstacles to the existence of anyons in reality are paper tigers.

Effectively two-dimensional systems arise when one prepares atomically thin layers of material, or for strongly anisotropic materials (such as graphite or the copper oxide high-temperature superconductors) which allow electron motion much more easily within planes than out of them. In thinking about the theory of such materials, it is important to recognise the fact that in quantum mechanics motion in the third direction may be *rigorously forbidden* at low energies. This can occur because the energy levels are quantised, and a definite non-zero minimum of energy must be injected to excite motion in the third direction.

The discrete symmetries P and T are very nearly respected in the fundamental laws of physics, but they are often badly violated by external conditions or by the spontaneous development of order in materials. For example an external magnetic field violates P and T, and therefore of course so does any ferromagnet. Several other less familiar but still common types of magnetically ordered states also spontaneously violate these symmetries, as does superfluid ^{3}He in its A phase.

Thus the necessity for effective two-dimensionality and for violation of the symmetries P and T is far from providing an insuperable barrier to the existence of anyons. It does focus our search, however: we must search among the quasiparticles of effectively two-dimensional systems either subject to explicit P and T violating conditions (e.g. a magnetic field) or exhibiting spontaneous order which violates these symmetries. Within this domain, theory gives us much encouragement. It turns out that some of the simplest and most natural P- and T-violating quantum field theories in two space dimensions have anyons in their spectrum. Since one expects that the low-energy excitations of materials will be described by the simplest quantum field theories of the appropriate symmetry, one is therefore encouraged to anticipate many real-world realisations of anyons.

This anticipation is rewarded in the new states of matter underlying the fractional quantised Hall effect. The effect is observed in the behaviour of electrons trapped in a layer at the boundary between different semiconductors and subjected to strong ($\gtrsim$ 1 T) magnetic fields at millikelvin temperatures.

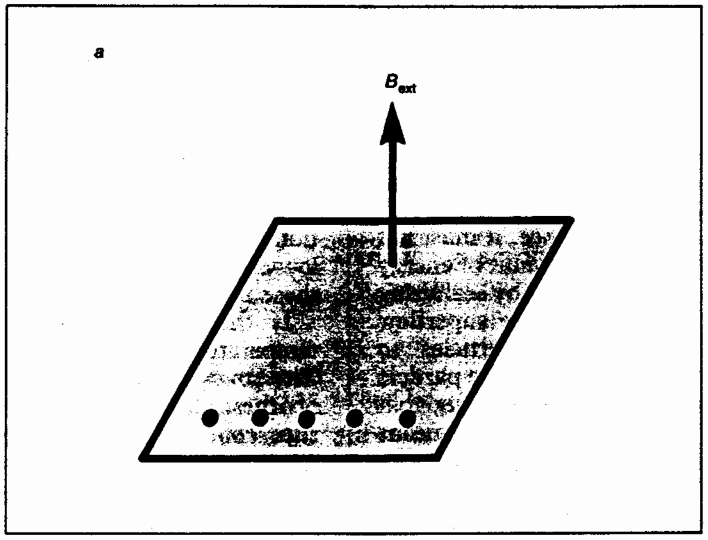
Unfortunately a full discussion of the FQHE is beyond the scope of this article, but a sample of the flavour of the subject can be garnered from the following small taste. By ruthlessly suppressing many fascinating aspects and technical ramifications of this effect, one may state its essence as follows. Just as certain atoms containing certain special numbers of electrons are particularly energetically favourable and stable (leading to the existence of inert elements or noble gases at these atomic numbers), and nuclei containing certain special numbers of neutrons or protons are particularly energetically favourable and stable (leading to "magic" nuclei), so too there are certain special densities at which the two-dimensional electron gas in a magnetic field is especially stable. More precisely – and this is very important – it is the ratio of density to applied magnetic field that is the crucial factor. This ratio is called the filling factor. At the preferred filling factors the electron gas becomes an incompressible fluid, since to compress it (by definition) one would need to alter the density – and that is precisely what the electrons don't want you to do.

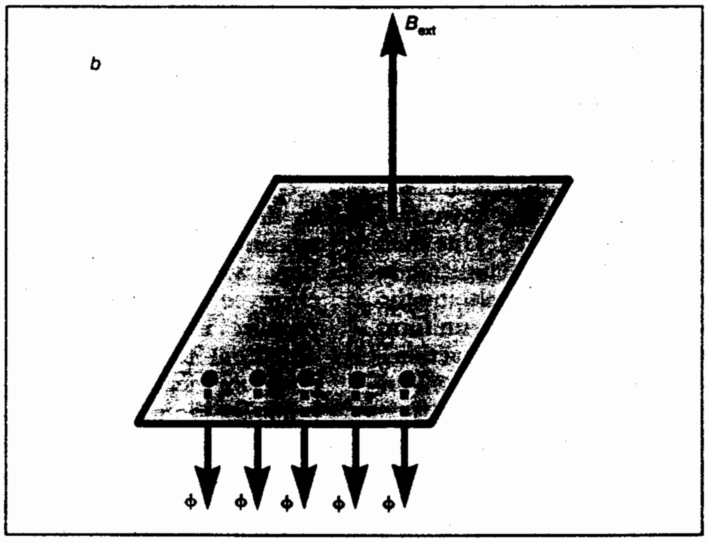
The existence of one of the preferred filling factors can be understood in a fairly straightforward way, similar to the way we understand closed shells in atoms or nuclei. That is, one ignores the interaction between electrons and solves the Schrödinger equation to find the possible states of single electrons in a magnetic field. It turns out that the energies of these states form bands, called Landau levels, separated by energy gaps. An especially stable configuration arises when there is one electron occupying each possible state within the lowest Landau level, that is, when one has a filled Landau level. (One may not assign *more* than one electron to each state, because of the Pauli exclusion principle.) The density of electrons relative to the density for a filled Landau level is called the filling fraction. For the especially stable configuration mentioned above the filling fraction is 1. Its stability reflects itself in the incompressibility of the electron gas at this filling fraction, which leads to the characteristic phenomenon of the quantised Hall effect. Indeed the central aspect of that effect is that even if the filling fraction is not quite 1 the electron gas organises itself almost everywhere into the desirable density that does correspond to filling fraction 1, accommodating the difference by nucleating small immovable pockets where surplus charge is localised. Thus the character of the mobile fluid, including its electrical resistance, is completely unchanged even if the electron density or the strength of the magnetic field to which it is subjected changes slightly. Such resistance "plateaus" are the defining experimental signature of the quantum Hall effect.

Thus the existence of the quantum Hall effect for filling fraction 1, or more generally for any integer filling fraction (i.e. for any whole number of exactly filled Landau levels), though as an historical fact it came as a surprise, is not terribly subtle to understand theoretically. It is essentially a single-particle effect, and does not involve any correlations among the particles, other than those implicit in the Pauli exclusion principle.

Much more subtle and interesting from a theoretical point of view is that other discrete fractional values of the filling factor – for example, exactly one-third the obviously preferred value – are also observed to be especially preferred. There are plateaus in resistance at these densities, too. That is the fractional quantised Hall effect (FQHE).

One way to understand this is shown in figure 2. In our minds, we can imagine starting from the preferred filling fraction and gradually localising magnetic flux onto the electrons. As we do this, we change the electrons from fermions into anyons of various types, as we have discussed. If we go far enough in the process, though, we eventually come right back to fermions. This happens when altogether we have gathered up the appropriate flux quantum. At that point our fantasy impinges on reality, since we have produced an acceptable state for physical (fermionic) electrons. In the process, the value of the background magnetic field has been altered, and with it the filling factor. It is not difficult to show that this procedure generates the favourable states at fractional filling factors continuously out of the well understood integer filling factor state. By continuously interpolating through a series of anyon states, one has connected two special electron states of very different character, and understood in a simple, satisfying way why the favourable filling factors are what they are.





2 It can be a good approximation to replace a uniform background magnetic field (*a*) by an equivalent flux localised on particles (*b*). Thus one approximates many-body systems in an external magnetic field by systems with slightly different quantum statistics in a slightly different field

Recently, Martin Greiter, X G Wen and I have introduced a new wrinkle on this analysis. The quantised Hall states described so far descend one full Landau level (filling fraction 1). These states are characterised by filling fractions with an odd denominator. Another interesting class of states follows upon applying the same sort of construction starting with a gas of electrons *not* subjected to any magnetic field – a free Fermi gas (equivalent to filling fraction ∞). From this starting point, one may simultaneously add flux localised on the particles and delocalised flux – i.e. a uniform magnetic field – in the other direction, in such a way that the magnetic field vanishes when averaged over a large area. It may be shown that the effect of adding both these perturbations simultaneously is equivalent to adding certain precisely defined short-ranged interactions among the original fermions. Interactions of this type can be analysed using the traditional, well developed tools of many-body theory. Following the same strategy as before, then, we can build up from small changes of statistics and magnetic field to such a large change that we come all the way around back to fermions (in a strong magnetic field). In this way, we relate the state of fermions with a specific local interaction to the state of fermions at filling fraction 1/2. Remarkably, the local interaction induces pairing, much as in the BCS theory of superconductivity. Following this logic, we expect that a sort of superconductivity occurs in the Hall effect at filling fraction 1/2. Its consequences are being vigorously explored at present. Recent experiments have revealed striking, and so far unexplained, peculiarities in the behaviour of the Hall effect at filling fraction 1/2. Although it would be premature to claim, it is not unreasonable to hope that the fascinating new "quasi-superconducting" state of matter suggested here by anyon ideas is present in reality.

Anyons also appear directly in the FQHE. According to the construction I have just outlined, electrons in the FQHE are actually in a way "super-fermions". If one electron is interchanged with another, one gets not only the phase π always associated with fermions but also (in the simplest, filling-factor 1/3 case) the extra 2π phase associated with the gathered-up flux, for a total of 3π. Since an interchange is half a winding, this means that one electron circling around another has its phase altered by 6π. That is three times the minimal phase it is possible to associate with a point in the two-dimensional electron gas, without disturbing the motion of the electrons circling around the point. This strongly suggests that it ought to be possible, in the FQHE state with filling factor 1/3, to define another kind of point-like excitation which is in some sense equivalent to one third of an electron. In fact, this turns out to be true in a very strong sense. The quasiparticles of filling factor 1/3 FQHE carry one-third of the charge of the electron – and 1/3 of the electron's statistics! Mathematically, when one of these quasiparticles winds around another the amplitude is multiplied by the cube root of −1. In the FQHE states at other filling factors, quasiparticles carrying other types of exotic anyon statistics occur.

Anyon superconductivity

Perhaps the most spectacular result in the theory of anyons to date is the proof that the existence of anyon statistics in and of itself can lead to a new form of superfluidity. More precisely, it has been demonstrated that gases of certain types of anyons will condense into superfluids at low temperatures, even if the interactions among them are otherwise strongly repulsive. (Actually, the temperature below which superconductivity appears is typically much higher than that predicted by other mechanisms.) As a corollary, since superfluids of charged particles are superconductors, one has a new and extremely robust mechanism of superconductivity.

The simplest way to appreciate the reason for the existence of anyon superfluidity is by reference to the classic criteria formulated by Landau in the context of ^4He. He argued that one will have superfluidity when two requirements are met simultaneously. First, the system must be compressible, so that there are extremely low-energy states of density flow (i.e. sound waves). Second, there must not be any other low-energy excitations, into which these sound waves could potentially dissipate. For instance, it must not be possible for a sound wave to give up a little bit of energy by letting some particle in the wave slow down by scattering off an ambient particle. Roughly speaking, the superflow must be a correlated motion of all the particles so constructed that the energetic cost of letting any one particle get "out of line" is greater than the gain in letting that particle slow down. In the absence of any reasonable means of dissipation, the flow will last forever – this is the essence of superfluidity.

Either of these requirements may fail, and each does in well known examples. For electrons in an insulator, or for a full Landau level of electrons in a magnetic field, the requirement of compressibility fails. In both these cases there is a definite preferred value of the density, and there are no density oscillations (i.e. sound waves) of low energy. There are no low-energy density flows, super or not. On the other hand electrons in a metal do not have a preferred density, and they support low-energy density oscillations, known as zero sound. However, zero sound waves can decay by losing energy to ambient electrons near the Fermi surface. They are certainly not superflows, whose main characteristic is that they persist indefinitely.

As we discussed above, one can begin to understand the properties of a gas of anyons by modelling the anyons as fermions with fictitious charge and flux attached. One then relates the behaviour of the anyons to the more familiar case of interacting fermions in a magnetic field, by replacing the fictitious flux localised on the particles with its uniform average, together with a residual short-ranged interaction.

Now if this procedure leads us to a system of fermions which *exactly fills* an integral number of Landau levels, then there are none of those low-energy, single-particle excitations whose existence would yield a means of dissipation, and in essence provide abundant leaks fatal to potential superflows. The absence of such excitations is easy to understand. Indeed, the fact that the fermions have completely filled a Landau level means that it takes a relatively large amount of energy – the energy gap to the next, unoccupied level – to change the motion of any single electron. (On the other hand if one has a partially filled level, the electrons in this level can change their state of motion by sliding to unoccupied states *within* the same level, at no great cost in energy.)

Ordinary electrons in external magnetic fields, however, do not satisfy the second condition for a superfluid. Since they have a preferred density, determined by and pinned to the magnitude of the external magnetic field, they do not support density oscillations or sound waves. They have very limited ability to flow. However, on this point anyons differ decisively from the system of fermions in a uniform magnetic field that we have used to approximate them. The crucial fact leading to this difference is that for anyons, unlike for fermions in a fixed magnetic field, when the particle density varies, the local value of the average magnetic field varies along with it. Thus there is no need to promote particles into the next Landau level in response to a gentle density gradient – rather, there is always an excellent fit between the local density and the local magnetic field, leading to exact integral filling. Because the density can be modulated smoothly with no energetic penalty, there are sound waves of very low energy. Following Landau's logic, the existence of these sound waves, together with the absence of low-energy, single-particle excitations, leads to superfluidity.

At a more formal level, one can show by direct but rather arduous calculations that the residual interactions, which distinguish the anyon system from its approximate realisation by fermions in a uniform magnetic field, restore the low-energy sound waves that do not exist in the absence of these interactions.

Is the superconductivity of the copper oxide high-temperature superconductors driven by this mechanism? The strong two dimensionality and peculiar spin ordering of these materials, as well of course as their robust superconductivity, might lead one to suspect so. How will we tell? It would certainly be encouraging to find that these materials violated the discrete symmetries P and T, since as we have seen violation of these symmetries is closely tied to the existence of anyons. Do they? At present the situation is unclear. Experimentalists have attempted to detect such symmetry violated by looking for specific asymmetric effects, such as the existence of an internal magnetic field or a tendency of the material to rotate the polarisation of light passing through it. Two of these experiments (performed by groups at Bell Labs, USA, and at Dortmund, Germany) appear to demonstrate such symmetry violation, but another of comparable sensitivity (by a group at Stanford) has found no trace of it.

Conclusion

In the span of a few years anyons have evolved from an amusing curiosity into a rich area of theoretical investigation, which has already made meaningful contact with observed phenomena in condensed matter physics and promises much more. In the future, it will come to seem as unnatural to restrict oneself to bosons and fermions in quantum mechanics as it would be to restrict oneself to real numbers in mathematical analysis.

Further reading

A much fuller discussion of all aspects of anyon physics, with references and including reprints of many of the significant original papers, is available in *Fractional Statistics and Anyon Superconductivity* F Wilczek 1990 World Scientific, Singapore

[redacted]