The Integer Quantum Hall Effect For Geometers

Edward Witten

Lecture One, TSIMF, December 2015

My lecture today will be on a topic in relatively contemporary physics – the integer quantum Hall effect – which can be understood with the help of some general mathematical knowledge and a very limited knowledge of physics. The integer quantum Hall effect is no longer new – as it was discovered around 1980. But it is still has a very fresh and modern feel. In the course of explaining the integer quantum Hall effect, I will give an elementary explanation of a number of ideas – mostly about effective field theory – that are important in modern physics.

I should say at the outset that the integer quantum Hall effect is a very rich subject that can be studied from many different points of view. I will be explaining only one point of view today.

One bit of physics that I will have to assume is familiar is that Maxwell's equations

$$\mathrm{d}F = \mathrm{d} * F = 0$$

describe propagation of light waves in vacuum. Here F is a two-form on $\mathbb{R}^{1,3}$, that is on $\mathbb{R}^4$ with the pseudo-Riemannian metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}\vec{x}^2$$

where $\vec{x} = (x, y, z) = (x_1, x_2, x_3)$.

Instead of merely viewing F as a two-form, it is better to interpret it as a connection on a $U(1)$ bundle over spacetime. Then one Maxwell equation $\mathrm{d}F = 0$ becomes an identity – the Bianchi identity – and the second one $\mathrm{d} * F = 0$ is the Euler-Lagrange equation derived from the "Maxwell action"

$$I = \frac{1}{4}\int_{\mathbb{R}^{1,3}} F \wedge *F.$$

Let us just check this in case it isn't completely familiar. The condition is supposed to be that I is stationary under a variation of the connection, that is under $F \to F + \mathrm{d}\delta A$. Indeed

$$\delta I = \frac{1}{2}\int_{\mathbb{R}^{1,3}} \mathrm{d}\delta A \wedge *F = -\frac{1}{2}\int_{\mathbb{R}^{1,3}} \delta A \wedge \mathrm{d} * F$$

so the condition for $\delta I = 0$ is indeed the second Maxwell equation $\mathrm{d} * F = 0$.

In this lecture, we are going to consider light (and more general electromagnetic fields) interacting with a material. But to keep things simple, we are going to take the material to be an insulator. A typical insulator for this purpose is a piece of glass. An insulator for us is an inert substance, which has no "relevant" degrees of freedom that we need to be concerned with. So describing how light behaves in an insulator just means describing how Maxwell's equations are modified in the presence of the insulator. By contrast, to describe light interacting with a conducting material, we would need to develop a theory of the charges and currents inside the material; we would have to learn a lot more physics.

Our “matter” system will be at rest and so the problem we will study doesn’t have the relativistic symmetry of the vacuum. Hence we will use a nonrelativistic notation, splitting the two-form F into the “electric field” and the “magnetic field, via

$$F = \mathrm{d}t \, \mathrm{d}\vec{x} \cdot \vec{E} + \frac{1}{2} \mathrm{d}\vec{x} \times \mathrm{d}\vec{x} \cdot \vec{B}.$$

Another point is that it is not a good idea to formulate the problem directly in terms of modifying Maxwell's equations. A random modification of Maxwell's equations would violate physical principles such as conservation of energy. It is better to modify the Maxwell action, not the equations; physically sensible modifications of Maxwell's equations are the ones that come from modifications of the action.

The fundamental reason that corrections to Maxwell's equations come from a modification of the action is that microscopically, nature is governed by a principle of stationary action (or its generalization in quantum mechanics). This has the beautiful property that if it holds microscopically, it also holds macroscopically. The full action including the Maxwell action in vacuum and the contributions due to the material make what we call the "effective action," one of the most important concepts in physics.

What modifications of the action should we consider? This depends on the material. And there are a lot of materials, and people keep fabricating new ones. So more or less anything we can imagine can be realized by some material.

By "anything" I mean you are allowed to imagine adding to the action I any polynomial in the electric and magnetic fields $\vec{E}$ and $\vec{B}$ and their derivatives

$$I_{\text{eff}} = \cdots + \int_M f(\vec{E}, \vec{B}, \mathrm{d}\vec{E}, \mathrm{d}\vec{B}, \mathrm{d}^2\vec{E}, \mathrm{d}^2\vec{B}, \dots).$$

Here $M \subset \mathbb{R}^{1,3}$ is the "world-volume" of the material.

There are two things that bring some order to this chaos. First, a Taylor series expansion is going to be rapidly convergent since (i) the electric and magnetic fields we can create in practice are small by atomic standards, (ii) derivatives can also be considered small for many purposes since the electric and magnetic fields we make are slowly varying compared to atomic sizes. (For light, this means our discussion today applies to light of visible wavelengths or longer, but not to X-rays or gamma rays.)

The second thing that brings some order to chaos is that the Taylor series expansion is constrained by the microscopic symmetries of a material. We might consider a homogeneous, isotropic material like a liquid, gas, or glass that we can consider to be invariant under the isometries of $\mathbb{R}^3$ – that is rotations and translations – though of course not the isometries of $\mathbb{R}^{1,3}$. The isometries of $\mathbb{R}^3$ are a group $\mathbb{R}^3 \rtimes SO(3)$ – an extension of the group $SO(3)$ of rotations by the group $\mathbb{R}^3$ of translations. This will put a lot of restrictions on the Taylor series expansion.

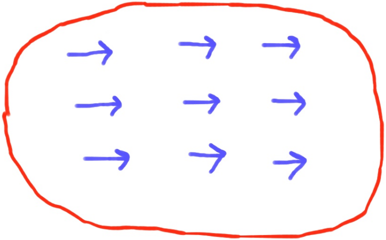
Instead of a homogenous, isotropic material, we can consider a crystal

in which case the symmetry is only a lattice subgroup $\Gamma \subset \mathbb{R}^3$ and a finite subgroup $F \subset SO(3)$ of automorphisms of Γ. For our purposes, the reduction of $\mathbb{R}^3$ to Γ isn't very important, but the reduction of $SO(3)$ to a finite subgroup can be very important; it allows more terms in the Taylor series expansion.

Let us practice by making a Taylor series expansion and discussing the significance of the first few terms. We can start with a term linear in $\vec{E}$:

$$I_{\text{eff}} = I_{\text{Maxwell}} + \int_M \vec{a} \cdot \vec{E}.$$

Here $\vec{a}$ must be a vector in $\mathbb{R}^3$ that is somehow characteristic of the material, so this interaction will not be possible if a material has too much symmetry microscopically. A material with this term in the effective action is "ferroelectric," meaning that in the absence of any externally applied electric field, it spontaneously generates an internal electric field:



To show that such a material is a ferroelectret, one has to solve the Euler-Lagrange equations coming from the combined action

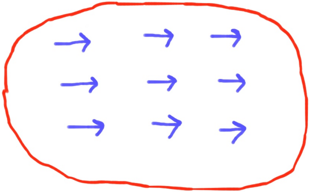
$$I_{\mathrm{eff}} = \frac{1}{2}\int_{\mathbb{R}^{3,1}} (\vec{E}^2 - \vec{B}^2) + \int_M \mathrm{d}^3x\,\mathrm{d}t\;\vec{a}\cdot\vec{E}$$

and we show that a solution for which $\vec{E}$ and $\vec{B}$ vanish outside the material has $\vec{E}$ nonzero inside it.

Similarly, we can contemplate a material in which the effective action has a term linear in $\vec{B}$:

$$I_{\text{eff}} = \cdots + \int_M \mathrm{d}^3x\,\mathrm{d}t\,\vec{a}\cdot\vec{B}.$$

Such a material turns out to be a ferromagnet, that is, it spontaneously generates an internal magnetic field. The picture is the same



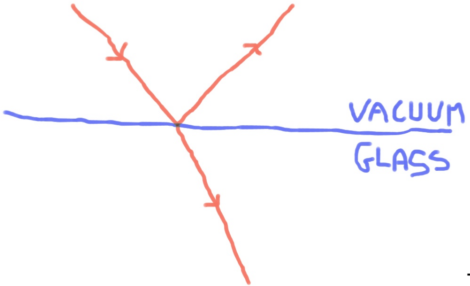
You all know examples of ferromagnets, but the examples you know best probably don't really fit our discussion since they are not insulators.

If a material has too much symmetry to allow a term linear in $\vec{E}$ or $\vec{B}$, it can still have a quadratic term

$$\int_M \mathrm{d}^3x \, \mathrm{d}t \, \left(\epsilon \vec{E}^2 + \mu \vec{B}^2\right).$$

These terms are possible in a homogeneous, isotropic material which has all the symmetries of $\mathbb{R}^3$. A material in which these are the dominant effects is called a dieletric. As long as we just look at what happens inside the dielectric, it is fairly obvious that what happens is that light waves still propagate, just not at the velocity that they would have in vacuum. That is because the action is the same as it would be in vacuum, but with different coefficients of the $\vec{E}^2$ and $\vec{B}^2$ terms. We say that the material is characterized by a dielectric constant that depends on ϵ and μ.

We miss a lot if we just describe what happens inside a dielectric. Interesting things happen at an interface between vacuum and a dielectric or between two different dielectrics. For instance, light entering the dielectric is bent or refracted, and some is reflected

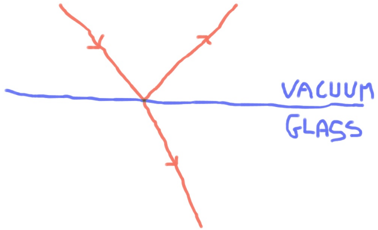


The reflected light is "polarized." Also, if an external electric or magnetic field is applied, charges or currents appear on the surface of the material, and the field inside is different from the field outside. (To understand the charges and currents, one needs something I will explain later.) All this has both interesting and useful applications.

Now let us consider a dielectric with a crystal structure that doesn't have much symmetry. The effective action can still have a term quadratic in E, but in the absence of $SO(3)$ symmetry, this quadratic term doesn't have to be $\vec{E}^2$; we can consider in the effective action some general quadratic function $V(\vec{E})$ or similarly $W(\vec{B})$:

$$I_{\text{eff}} = \cdots + \int_M \mathrm{d}^3x \, \mathrm{d}t \, \left(V(\vec{E}) + W(\vec{B}) \right) .$$

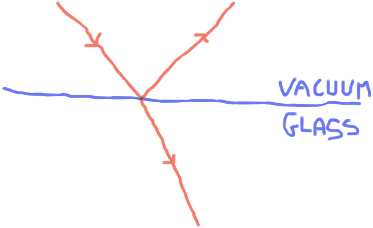
The speed of light in the material now depends on the direction of propagation, as well as the polarization; likewise the angle of refraction:



Since the effective action is allowed to depend on derivatives of $\vec{E}$ and $\vec{B}$, let us try it:

$$I_{\text{eff}} = \cdots + \int_M \mathrm{d}^3x\,\mathrm{d}t\,\sum_{i=1}^{3} \frac{\partial}{\partial x^i}\vec{E} \cdot \frac{\partial}{\partial x^i}\vec{E}.$$

Now the velocity at which light propagates – and its reflection and refraction properties

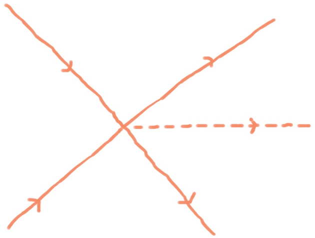


depend on color. (We say that the index of refraction is frequency-dependent.) This is a better description of a real dielectric.

I will just consider one more example. What happens if we include one more term in the Taylor series expansion? – a a cubic term $H(\vec{E})$ or $H(\vec{E}, \vec{B})$. (A cubic term is not $SO(3)$-invariant, so we are discussing a material with not too much symmetry.) Thus we consider

$$I_{\mathrm{eff}} = \cdots + \int_M \mathrm{d}^3x \, \mathrm{d}t \, H(\vec{E}, \vec{B}).$$

What happens now is that Maxwell's equations become nonlinear – since a cubic term in the effective action leads to a quadratic term in the Euler-Lagrange equations. This leads to what is called nonlinear optics, provided one has intense laser beams at one's disposal:



What we have hopefully learned from this is that: (1) a material is characterized by an effective action; (2) the effective action for a suitably chosen material can be more or less anything subject to general principles; (3) a lot of physics follows from simple statements about the effective action. By considering only insulators, we kept things simple: we could consider effective actions for a $U(1)$ gauge field only, without having to discuss what are the relevant degrees of freedom characterizing the material.

Before we go on, we need one more lesson about the effective action. A full description of nature would involve the $U(1)$ gauge field A and other variables such as electrons and atomic nuclei. Let us just write Φ for those variables. So the microscopic action would be something like

$$I_{\rm micro} = I_{\rm Maxwell} + \tilde{I}(\Phi, A) = \frac{1}{4}\int_{\mathbb{R}^{3,1}} F \wedge *F + \tilde{I}(\Phi, A).$$

What is called the electromagnetic current $\mathcal{J}$ is the extra term in the Euler-Lagrange equations:

$$\mathrm{d} * F + \frac{\delta I(\Phi, A)}{\delta A} = 0,$$

so

$$\mathcal{J} = \frac{\delta I(\Phi, A)}{\delta A}.$$

One explanation of this definition is that since $\mathrm{d}^2 = 0$, the equation $\mathrm{d} * F + \mathcal{J} = 0$ implies current conservation

$$\mathrm{d}\mathcal{J} = 0.$$

When we talk about an effective action, this means, roughly, that I_{micro} has been extremized with respect to Φ to get an effective action for A only

$$I_{\mathrm{eff}} = \frac{1}{4} \int_{\mathbb{R}^{3,1}} F \wedge *F + \tilde{I}(A).$$

This does not change the relation of the electromagnetic current to the extra term in the action

$$\mathcal{J} = \frac{\delta \tilde{I}(A)}{\delta A}.$$

This gives us a framework to understand the currents and charges in an insulating material.

Now we are going to consider $2+1$-dimensional materials – thin films. A thin film could be a sheet of paper, a monoatomic layer floating in empty space, a very thin layer at the surface of another material, or in general anything that has 2 spatial directions that are sufficiently large compared to the third.

We again assume that our material is an insulator, so as before, we can describe it by an effective action. This effective action can have the sort of terms that we are by now familiar with

$$I_{\mathrm{eff}} = \cdots + \int_M \left(\vec{a} \cdot \vec{E} + \epsilon \vec{E} \cdot \vec{E} + \dots \right)$$

with the sole difference that now M has dimension $2+1$ rather than $3+1$. But in addition, one more term is possible that does not have a close analog in 3+1 dimensions.

This is the Chern-Simons function, which we can define roughly as

$$I_{\rm CS} = \frac{1}{4\pi}\int_M A \wedge {\rm d}A.$$

I will assume that it is known that, for A a connection on a $U(1)$ bundle $\mathcal{L} \to M$, where M is an oriented three-manifold without boundary, $I_{\rm CS}$ can be defined as a gauge-invariant map to $\mathbb{R}/2\pi\mathbb{Z}$. (The naive formula that I wrote assumes that $\mathcal{L}$ is trivial.) We will discuss later what happens when M has a boundary. Now we need the following fundamental fact about quantum mechanics, essentially first exploited by Dirac: the action does not have to be well-defined as a real number, but it does have to be well-defined as a map to $\mathbb{R}/2\pi\mathbb{Z}$. The reason for that last statement is that the starting point in quantum mechanics is the Feynman path integral, in which the integrand is e^{iI}.

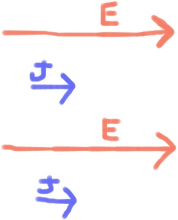
Accordingly, $2+1$-dimensional materials are characterized by an integer k. The effective action is

$$I_{\rm eff} = I_{\rm Maxwell} + \cdots + k\, I_{\rm CS}$$

where k must be an integer. The integer k is 0 for a sheet of paper, but since 1980 (when a serendipitous discovery was made by K. von Klitzing, leading to the 1985 Nobel Prize) materials with nonzero k have been known. Integrality of k for many materials is measured to 7 or 8 digits, making it one of the most precisely known facts in physics.

How is k measured? Conceptually, what may be the most obvious thing to do is to scatter light from our thin film. The Chern-Simons interaction will produce a distinctive effect on the polarization (and intensity) of the scattered light. Surprisingly, this measurement apparently hasn't been made, though in the last couple of years, people are thinking about how to make it in practice.

In practice, one instead measures "Hall currents." When one applies an electric field to a conductor, there is an induced current that is usually in the direction of the applied field:

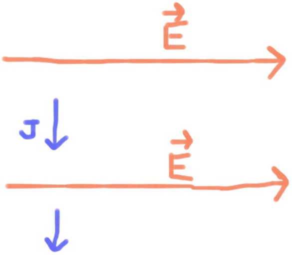


The constant of proportionality between the electric field $\vec{E}$ and the current $\vec{J}$ is called the conductivity:

$$\vec{J} = \sigma \vec{E}$$

It vanishes for insulators, by definition. (To explain the notation here, I earlier defined the current as a 3-form $\mathcal{J} = \delta I(\Phi, A)/\delta A$, but physicists usually define $J = \star\mathcal{J}$ and – using the pseudo-Riemmanian metric on spacetime – view J not as a 1-form but as a vector field representing flow of charge. Let us use that language.)

Though we usually expect the flow of current to be proportional to the applied electric field $\vec{E}$, there are circumstances in which an applied electric field generates a current that is *perpendicular* to the field:



Perpendicular currents were first observed in the 19th century in the presence of a magnetic field and are known as Hall currents.

A magnetic field was also important in von Klitzing's work. But subsequently it has been discovered that the quantum Hall effect – which I am describing – can occur in the absence of a magnetic field and so I am not emphasizing the magnetic field today. I should however mention that the example with a magnetic field does give an intuitive understanding of how Hall currents are possible.

Let us see how the Chern-Simons function leads to a Hall current. With

$$I^* = k\, I_{\mathrm{CS}} = \frac{k}{4\pi}\int_M A \wedge \mathrm{d}A = \frac{k}{4\pi}\int_{\mathbb{R}^{1,3}} \delta_M \wedge A \wedge \mathrm{d}A$$

we find that the contribution of I_{CS} to the current 3-form is

$$\mathcal{J}_{\mathrm{CS}} = \frac{k}{2\pi}\delta_M \wedge F.$$

Here δ_M is the Poincaré dual to $M \subset \mathbb{R}^{1,3}$, which I used to convert an integral over M to one over $\mathbb{R}^{1,3}$.

Let us suppose that that $\mathbb{R}^3$ is parametrized by Cartesian coordinates x, y, z and that M is supported at $z = 0$, so $\delta_M = \delta(z)\mathrm{d}z$. Let us take an electric field of strength E_x in the x direction, meaning $F = E_x \mathrm{d}x \wedge \mathrm{d}t$. So

$$\mathcal{J}_{\mathrm{CS}} = \frac{kE_x}{2\pi}\delta(z)\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}t.$$

So we set $J_{\mathrm{CS}} = *\mathcal{J}_{\mathrm{CS}}$ and find

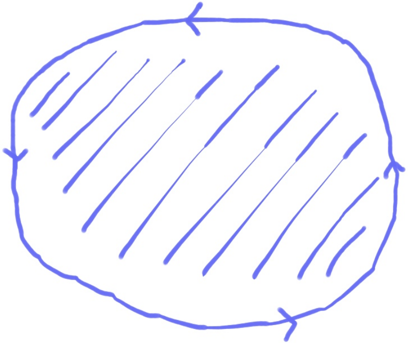
$$J_{\mathrm{CS}} = \frac{k}{2\pi}E_x \mathrm{d}y\delta(z)$$

showing that the current has a magnitude $k/2\pi$ times the magnitude of the electric field and is perpendicular to the electric field. The delta function just means that the current only flows inside the material. Integrality of the parameter k in this formula is measured to great precision.

There are all sorts of pressing things to explain about this situation. Perhaps one of the most pressing is to discuss what happens if our thin film has a boundary. In practice, this is almost always the case! Mathematically, if A is a connection on a $U(1)$ bundle $\mathcal{L} \to M$, where M is an oriented three-manifold with boundary N, then it is not possible to define the Chern-Simons function $I_{\rm CS}(A)$ as a gauge-invariant function valued in $\mathbb{R}/2\pi\mathbb{Z}$. However, we can do the following. Let $\mathcal{L}_N \to N$ be the restriction of $\mathcal{L}$ to N and let $\mathcal{A}$ be the infinite-dimensional space of connections on $\mathcal{L}_N$. Then one can define a hermitian line bundle $\mathcal{M} \to \mathcal{A}$ and one can define $e^{iI_{\rm CS}(A)}$ as a section of $\mathcal{M}$ and hence $e^{ikI_{\rm CS}(A)}$ as a section of $\mathcal{M}^k$. (One version of this is due to Ramadas, Singer, Weitsman (1989).)

But we can't do physics with an exponentiated action that is a section of a line bundle. The exponential $\exp(iI)$ has to be a well-defined complex number. So something is missing. In fact, the assumption that our thin film is an insulator with no relevant degrees of freedom of its own cannot be true near the boundary of the film. There must be degrees of freedom on the boundary whose effective action produces a section of $\mathcal{M}^{-k}$. There is not a unique way for this to happen, but the simplest way relies on the fact that if we use the metric of spacetime to define a complex structure on N, and on the line bundle $\mathcal{L}_N \to N$, then the determinant of the $\bar{\partial}$ operator acting on $\mathcal{L}_N$ is a line bundle over $\mathcal{A}$ that is isomorphic to $\mathcal{M}^{-1}$ or $\mathcal{M}$ (depending on orientation).

Physically this means that charged fermions living on the boundary of our thin film and propagating in only one direction around the 1-dimensional boundary



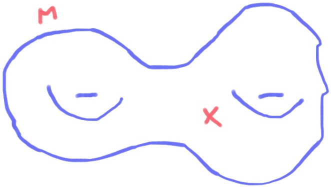
have an effective action whose exponential is a section of $\mathcal{M}^{-1}$. These fermions contribute the famous "edge currents" of quantum Hall systems.

Another point that may be obvious to a topologist is that we might consider the case that our thin film M is the boundary of a three-manifold X and we are given an extension of the line bundle $\mathcal{L} \to M$ and the connection A to a line bundle with connection $\mathcal{L} \to X$. In this case the substitution

$$I_{\mathrm{CS}}(A) = \frac{1}{4\pi} \int_M A \wedge \mathrm{d}A \to \frac{1}{4\pi} \int_X F \wedge F$$

gives a completely gauge-invariant definition of $I_{\mathrm{CS}}(A)$ as a real-valued function. However, this definition does depend on the choice of X. If there is a natural X in the problem, this means that there is no reason for k to be an integer.

There is a natural X if our thin film lives on the surface of a 3-dimensional material:



In recent years, a fascinating theory of "topological insulators" has been developed. A tiny ingredient in this theory is the fact that the surface of a 3-dimensional material can behave under certain conditions somewhat like an abstract thin film, but with a crucial difference that there is no reason that k has to be an integer.

My work on Khovanov homology, which I am talking about in the other two lectures, can be viewed as a sort of topological counterpart of the theory of topological insulators. The thin film that supports the action I_{CS} is where the knot lives, in the context of my other lectures.

There is one more thing that I really should tell you, even though I will just be telling you the topic of another lecture. There is something called the fractional quantum Hall effect, which is even more interesting than the integer quantum Hall effect, which is what I have described. The fractional quantum Hall effect arises for a thin film for which our assumption “no relevant degrees of freedom” is not quite true, at least not in a strong sense. The fractional quantum Hall effect leads to a fascinating role of topological quantum field theory in condensed matter physics, with links to the Jones polynomial and other quantum invariants of knots, and potential applications to quantum computing, and more.

What makes the fractional quantum Hall effect fundamentally different from the integer one is that to understand the fractional quantum Hall effect, we have to become quantum field theorists. By contrast, we've been able to understand the integer Hall effect almost as classical physicists – or geometers – with just a pinch of quantum mechanics now and then.