Draft Lecture V-VI notes for Les Houches 2014

Joel E. Moore, UC Berkeley and LBNL

(Dated: August 13, 2014)

I. INTRODUCTION

These two lectures are our first exposure to strongly interacting topological phases, defined as those that cannot be understood in terms of free particles. In contrast, the integer quantum Hall effect and topological insulators can be understood in terms of free particles, although these phases are stable in the sense that they survive over a finite region of interaction strength until a phase transition occurs. Our main tool will be quantum field theory, which is a powerful language to describe the long-wavelength physics of interacting systems. After giving some microscopic motivation from the fractional quantum Hall effect (FQHE), we give a first example of field theory applied to spin chains as an example of how an analysis of topological terms in a simple field theory led to a clear experimental prediction (the "Haldane gap") regarding antiferromagnetic integer-spin Heisenberg chains.

We then return to the quantum Hall effect and develop Abelian Chern-Simons theory, an example of a truly topological field theory. Although it is written in terms of one or more $U(1)$ gauge fields, similar to ordinary electromagnetism, its behavior is strikingly different than the conventional field theories with which the reader may already be familiar. In lieu of a microscopic derivation, which has been carried out but is somewhat tedious, we show that it unifies properties such as ground state degeneracy,

II. FQHE BACKGROUND

We give quickly some standard background on the fractional quantum Hall effect in order to motivate the Chern-Simons field theory introduced below. The goal of that field theory is to give a compact universal description of the key features of the topological order in quantum Hall states, similar in spirit to the Ginzburg-Landau field theory of symmetry-breaking phases. Most of this material is standard and can be found in quantum Hall edited volumes and textbooks (Prange and Girvin; Das Sarma and Pinczuk; Jain).

Our discussion centers on the Laughlin wavefunction for two-dimensional electrons ($z_j = x_j + iy_j$ describes the jth electron, $j = 1, \ldots, N$)

$$\Psi_m = \left(\prod_{i<j} (z_i - z_j)^m\right) e^{-\sum_i |z_i|^2/4\ell^2}. \tag{1}$$

The magnetic length is $\ell = \sqrt{\hbar c/eB}$ and the wavefunction is not normalized. This wavefunction clearly can be expanded over the single-electron lowest Landau level wavefunctions in the rotational gauge,

$$\psi_m = z^m e^{-|z|^2/4\ell^2}. \tag{2}$$

where $m = 0, 1, \ldots$ labels angular momentum. For $m = 1$ the Laughlin state is just a Slater determinant for the filled lowest Landau level, but for higher m it is believed not to be a sum of any finite number of Slater determinants in the $N \to \infty$ limit.

We explained the origin of this wavefunction using the pseudopotential approach introduced by Haldane: it is the maximum-density zero-energy state of a repulsive interaction that vanishes for relative angular momentum greater than or equal to m. We checked that its density is $\nu = 1/m$ by looking at the degree of the polynomial factor, which is directly related to $\langle r^2 \rangle$, and argued that it contains "quasihole" excitations of charge $-q/m$, where q is the charge of the electrons. The wavefunction for a quasihole at z_0 is

$$\Psi_{\text{quasihole}} = \left(\prod_i (z_i - z_0)\right) \Psi_m. \tag{3}$$

The fractional charge can be understood by noting that m copies of the extra factor here would lead to the wavefunction with an electron at z_0, but without treating z_0 as an electron coordinate; in other words, a wavefunction with a "hole" added at z_0. It has edge states that at first glance are loosely similar to those in the filled Landau level.

III. TOPOLOGICAL TERMS IN FIELD THEORIES: THE HALDANE GAP AND WZW MODELS

As a warm-up for fully topological field theories, we give an example of how topological terms can have profound consequences in "ordinary" field theories (i.e., theories without gauge fields). By a topological term we mean loosely one whose value in any specific configuration (a path in the path integral) is a topological invariant, so that the set of all paths can be divided into topological sectors by the value of the topological term. A famous example of this phenomenon found by Haldane led to the first understanding of the gapped spin-one Heisenberg antiferromagnet in one spatial dimension, which we will interpret later in section as a symmetry-protected trivial (SPT) phase of interacting particles. We will focus on topological terms that appear in nonlinear σ-models, which despite their unwieldy name are a very basic type of field theory for systems in or near an ordered phase breaking a continuous symmetry.

We first present Haldane's example and then discuss a different kind of topological term that appears in Wess-Zumino-Witten models, again in one spatial dimension. For higher dimensions, only a few general comments are provided at the end, but we return to the subject when we consider nonlinear σ-models for disordered systems in section **??**. The nonlinear σ-model (NLSM) is an example of an effective theory, a simplified description of the low-energy degrees of freedom of a complicated system. Ginzburg-Landau theory is another such effective theory, and one use of the NLSM is as a further simpification of Ginzburg-Landau theory where we have thrown away the "hard" or "massive" fluctuations of the magnitude of the order parameter, keeping only the "soft" or "massless" fluctuations within the order parameter manifold.

For definiteness, consider a d-dimensional XY model, which would be described in Ginzburg-Landau theory by a 2-component real or 1-component complex order parameter. The mean-field physics in the ordered phase as a function of the order parameter is illustrated in Fig. **??**: the order parameter manifold of symmetry-related ground states is a circle, and we can expect that fluctuations along this circle are "soft" in the sense of requiring little energy (as this is a flat direction of the energy) while those perpendicular to the circle are more costly. This order parameter manifold is the same as that considered in the discussion of topological defects in Section **??**, where defects were classified using maps from surfaces enclosing the defect in real space to the order parameter manifold. At low temperature we might expect that a reasonable description of the system is therefore obtained just from fluctuations of the order parameter's direction, leading to a functional integral for the coarse-grained classical partition function:

$$Z_{\mathrm{NLSM}} = \int \mathcal{D}\theta(x) e^{-\beta c \int \frac{(\nabla\theta)^2}{2}\, d^d x}. \tag{4}$$

Here c is a coupling constant with units of energy if $d=2$; one could estimate c simply from the coupling strength in a lattice XY model. The NLSM it is called nonlinear because the circle is defined by a hard constraint on the $\hat{\mathbf{n}}$ field, which in more complicated target manifolds such as the sphere leads to interaction (i.e., nonlinear) terms in the fields obtained in a perturbative expansion; it is called a sigma model because of its first appearance in particle physics.

For a quantum-mechanical model at zero temperature, we might expect on general grounds that imaginary time will become an extra dimension in any Euclidean path-integral representation of the partition function, in the same way as the Dirac-Feynman path integral for quantum mechanics involves integration of the Lagrangian over time (a 0+1-dimensional theory). Now we will obtain a NLSM for a quantum-mechanical problem in one spatial dimension. Heuristically, we might expect an NLSM to be a reasonable description for a quantum model that is "close to" having symmetry-breaking order.

Our approach is to derive a connection between the low-energy, long-wavelength degrees of freedom of the spin path integral of the Heisenberg antiferromagnet. This process is known as Haldane's mapping in the context of spin systems: we will use it to show that there is a topological term present for half-integer spin but not for integer spin, which is believed to explain the different behavior seen numerically and experimentally in these two cases.

First we look for a more general way of writing the Berry phase term for a spin that results from setting up a coherent-state path integral for spin. In order to make a path integral, we should set up an integral over "classical" trajectories; what is the classical trajectory of a spin? One answer is to use the overcomplete basis of coherent states for the spin-S Hilbert space (see the book of Auerbach), which are labeled by a unit vector $\hat{\Omega}$. As S increases, the spin wavefunction becomes more and more concentrated around $\hat{\Omega}$.

$$\omega[\hat{\Omega}] = -\int_0^\beta d\tau \dot{\phi} \cos\theta. \tag{5}$$

For a closed path on the sphere, we showed in Chapter II that this corresponds to the signed spherical area enclosed by the path. An overall ambiguity of $\pm 4\pi$ in this area does not affect the physics, since the area ω appears in the path integral action with a coefficient $-iS$. For a many-spin system, the full action was

$$S[\hat{\Omega}] = -iS \sum_i \omega[\hat{\Omega}_i] + \int_0^\beta d\tau \, \frac{S^2 J}{2} \sum_{ij} \hat{\Omega}_i \cdot \hat{\Omega}_j. \tag{6}$$

For now we return to a single spin to set up an improved way of writing the Berry phase term.

Let the vector potential $\mathbf{A}(\hat{\Omega})$ be assumed to have the following property: its line integral over a closed orbit on the sphere should give the area enclosed by the orbit,

$$\omega = \int_0^\beta d\tau\, \mathbf{A}(\hat{\Omega})\dot{\hat{\Omega}}. \tag{7}$$

Then Stokes's theorem fixes curl $\mathbf{A}$ to be the magnetic field of a magnetic monopole (a vector with uniform outward component):

$$\nabla \times \mathbf{A} \cdot \hat{\Omega} = \epsilon^{\alpha\beta\gamma} \frac{\partial A_\beta}{\partial \hat{\Omega}_\alpha} \hat{\Omega}^\gamma = 1. \tag{8}$$

Two explicit examples to check that this can be done are

$$\mathbf{A}^a = -\frac{\cos\theta}{\sin\theta}\hat{\phi}, \quad \mathbf{A}^b = \frac{1-\cos\theta}{\sin\theta}\hat{\phi}. \tag{9}$$

Clearly $\mathbf{A}^a$ has singularities at the north and south poles, while $\mathbf{A}^b$ has a singularity only at the south pole. (Parenthetical note: actually $\mathbf{A}^b$ is a good representation of the field of a Dirac monopole: a singular flux ("Dirac string") enters through the south pole, and then goes out uniformly over the rest of the sphere. A small circle around the south pole contains flux 4π, which contributes $4\pi S$ to the action, but recall that this winds up giving zero physical contribution to the path integral.)

Now we can use this representation to write concisely the variation of the Berry phase term under a small variation in the path from imaginary time 0 to imaginary time t. Suppose that we want to calculate

$$\delta\omega[\hat{\Omega}] = \int_0^t dt'\, \delta(\mathbf{A} \cdot \dot{\hat{\Omega}}) = \int_0^t dt'\, \left(\frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta} \delta\hat{\Omega}^\beta \dot{\hat{\Omega}}^\alpha + A^\alpha \frac{d}{dt}\delta\hat{\Omega}^\alpha\right) \tag{10}$$

under a small variation of the path $\delta\hat{\Omega}$ that is assumed to keep the endpoints fixed. Now subtract $\frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta}\dot{\hat{\Omega}}^\beta \delta\hat{\Omega}^\alpha$ from the first term and add it to the second, to get

$$\begin{aligned}
\delta\omega[\hat{\Omega}] &= \int_0^t dt'\, \frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta} \epsilon^{\alpha\beta\gamma} (\dot{\hat{\Omega}} \times \delta\hat{\Omega})_\gamma + \int_0^t dt'\, \left(A^\alpha \frac{d}{dt}\delta\hat{\Omega}^\alpha + \frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta}\dot{\hat{\Omega}}^\beta \delta\hat{\Omega}^\alpha\right) \\
&= \int_0^t dt'\, \hat{\Omega} \cdot (\dot{\hat{\Omega}} \times \delta\hat{\Omega}) + \int_0^t dt'\, \frac{d}{dt'}(\mathbf{A} \cdot \delta\hat{\Omega}) = \int_0^t dt'\, \hat{\Omega} \cdot (\dot{\hat{\Omega}} \times \delta\hat{\Omega}).
\end{aligned} \tag{11}$$

Here we used the condition (8) and also, in rewriting the first term, the fact that the quantity in parentheses $(\dot{\hat{\Omega}} \times \delta\hat{\Omega}) \parallel \hat{\Omega}$ because of the constant length of the vector $\hat{\Omega}$.

Now, after this prelude, we are ready to rewrite the full path integral for the many-spin Heisenberg model. The first step is to write the spin $\hat{\Omega}_i$ in terms of two continuous fields of spacetime $\hat{\mathbf{n}}$ and $\mathbf{L}$:

$$\hat{\Omega}_i = \eta_i \hat{\mathbf{n}}(\mathbf{x}_i)\sqrt{1 - \left|\frac{\mathbf{L}(\mathbf{x}_i)}{S}\right|^2} + \frac{\mathbf{L}(\mathbf{x}_i)}{S}. \tag{12}$$

Here η_i alternates between sublattices, $\hat{\mathbf{n}}(\mathbf{x})$ is a unit vector field, sometimes referred to as the Néel field, and $\mathbf{L}$ is constrained to be perpendicular to $\hat{\mathbf{n}}$. Hence a constant value of $\hat{\mathbf{n}}$ corresponds to a classical Néel state. It seems like we have greatly increased the degrees of freedom by this rewriting; what we do now is restrict the allowed Fourier components of the new fields (i.e., the Brillouin zone) to small momenta in such a way that the total number of degrees of freedom is unchanged (cf. Auerbach for details). The spirit of this approximation is that we are interested in long-length-scale physics so details on the scale of the lattice spacing are unimportant. It turns out that we assume slow variations in $\hat{\mathbf{n}}$ but only that $|\mathbf{L}| \ll S$, i.e., that $\mathbf{L}$ is small but not necessarily slowly varying. We now expand the path integral in powers of $|\mathbf{L}|/S$.

A pair of spins gives a contribution

$$\hat{\Omega}_i \cdot \hat{\Omega}_j \approx \eta_i\eta_j - \frac{1}{2}\eta_i\eta_j(\hat{\mathbf{n}}_i - \hat{\mathbf{n}}_j)^2 + \frac{1}{S^2}\left[\mathbf{L}_i\mathbf{L}_j - \frac{1}{2}\eta_i\eta_j(\mathbf{L}_i^2 + \mathbf{L}_j^2)\right] + \frac{1}{S}(\eta_j\mathbf{L}_i\hat{\mathbf{n}}_j + \eta_i\mathbf{L}_j\hat{\mathbf{n}}_i) + \ldots \tag{13}$$

Here the neglected terms are of order $|\mathbf{L}|^2(\hat{\mathbf{n}}_i - \hat{\mathbf{n}}_j)$ or smaller. In the first term, use a Taylor expansion to convert differences of the Néel field into derivatives and keep only the leading contribution. You can show (or it is in Auerbach) that the cross terms (those with both $\mathbf{L}$ and $\hat{\mathbf{n}}$) vanish by the symmetry of the Heisenberg Hamiltonian. The term with two $\mathbf{L}$ factors we rewrite below in Fourier space, where it is much simpler and where we will be able to "integrate it out".

What we are left with, after going from the lattice to integrals using

$$\sum_i F_i \to a^{-d} \int d^d x \sum_i \delta(\mathbf{x} - \mathbf{x}_i) F(x), \tag{14}$$

is the continuum representation

$$H = E_0 + \frac{1}{2} \int d^d x \left[\rho_s \sum_l |\partial_l \hat{\mathbf{n}}|^2 + \int d^d x' \, (\mathbf{L}_x \chi^{-1}_{xx'} \mathbf{L}_{x'}) \right]. \tag{15}$$

Here E_0 is the classical energy

$$E_0 = \frac{S^2}{2} \sum_{ij} J_{ij} \eta_i \eta_j. \tag{16}$$

In the first term, the spin stiffness is

$$\rho_s = -\frac{S^2}{2dNa^d} \sum_{ij} J_{ij} \eta_i \eta_j |\mathbf{x}_i - \mathbf{x}_j|^2. \tag{17}$$

The second or "canting" term in Fourier space is simply

$$\int \frac{d^d q}{(2\pi)^d} \frac{\mathbf{L}_{\mathbf{q}} \mathbf{L}_{-\mathbf{q}}}{J(\mathbf{q}) - J(\pi, \pi, \ldots)}, \quad J(\mathbf{q}) = \sum_j e^{i\mathbf{q}\cdot(\mathbf{x}_i - \mathbf{x}_j)} J_{ij}. \tag{18}$$

Now we just need to rewrite the geometric phase

$$-iS \sum_i \omega_i = -iS \int_0^\beta d\tau \sum_i \mathbf{A}(\hat{\Omega}_i) \dot{\hat{\Omega}}_i. \tag{19}$$

Assume that the vector potential is chosen so that $\mathbf{A}(\hat{\Omega}) = \mathbf{A}(-\hat{\Omega})$, as works for one of the examples above. Now expanding in terms of the new fields,

$$\begin{aligned} -iS \sum_i \omega_i &= -iS \sum_i \eta_i \omega[\hat{\mathbf{n}}_i + \eta_i \mathbf{L}_i / S] \\ &= -iS \sum_i \left[\eta_i \omega[\hat{\mathbf{n}}_i + \frac{\delta \omega}{\delta \hat{\mathbf{n}}_i} \cdot (\mathbf{L}_i / S) \right] \\ &= -i\Upsilon - i \int_0^\beta d\tau \sum_i (\hat{\mathbf{n}}_i \times \partial_\tau \hat{\mathbf{n}}_i \cdot \mathbf{L}_i). \end{aligned} \tag{20}$$

In the last line we used our earlier formula for the variation of ω, and defined

$$\Upsilon = S \sum_i \eta_i \omega[\hat{\mathbf{n}}(\mathbf{x}_i)], \tag{21}$$

switching to the spatial continuum limit.

Now our goal is going to be to combine the classical and geometric terms in order to obtain a simple long-wavelength action. The key step is to note that the second term in (20) couples one power of L to a combination of n fields. So integrating out the L degrees of freedom (a Gaussian integral) will give rise to the following: considering only the terms involving $\mathbf{L}$ and doing the integral in Fourier space, we get (ignoring an unimportant overall constant)

$$Z_L \propto \int \mathcal{D}\hat{\mathbf{n}} \, \exp \left[-\frac{1}{2} \int d\tau \, \frac{d^d q}{(2\pi)^d} (J(\mathbf{q}) - J(\vec{\pi})) (\hat{\mathbf{n}} \times \partial_\tau \hat{\mathbf{n}})_{\mathbf{q}} \cdot (\hat{\mathbf{n}}_\times \partial_\tau \hat{\mathbf{n}})_{-\mathbf{q}} \right]. \tag{22}$$

We can simplify this much further: for long wavelengths we approximate $\chi(\mathbf{q}) \approx \chi(0)$, and use

$$|(\hat{\mathbf{n}} \times \partial_\tau \hat{\mathbf{n}}|^2 = |\partial_\tau \hat{\mathbf{n}}|^2 \tag{23}$$

from the constant length of $\hat{\mathbf{n}}$. to get just (the real-space constant $\chi_0 = a^{-d}\chi(0) = a^{-d}(J(\vec{0}) - J(\vec{\pi})))$

$$Z_L = \int \mathcal{D}\hat{\mathbf{n}} \exp\left(-\frac{1}{2}\int_0^\beta d\tau \int d^d x\, \chi_0 |\partial_\tau \hat{\mathbf{n}}|^2\right). \tag{24}$$

So, putting it all together, we have

$$Z \propto \int \mathcal{D}\hat{\mathbf{n}}\, e^{i\Upsilon} \exp\left[-\frac{1}{2}\int_0^\beta d\tau \int d^d x \left(\chi_0 |\partial_\tau \hat{\mathbf{n}}|^2 + \rho_s |\partial_{x^\alpha} \hat{\mathbf{n}}|^2\right)\right]. \tag{25}$$

This now looks much more symmetric between space and time; if desired, one can just rescale time to make the theory look like it lives in an isotropic $d+1$-dimensional space. This gives

$$Z \propto \int \mathcal{D}\hat{\mathbf{n}}\, e^{i\Upsilon} \exp(-\int d^{d+1}x \mathcal{L}_{NLSM}), \quad \mathcal{L}_{NLSM} = \sum_{\alpha=1}^{d+1} \frac{\partial_{x^\alpha}\hat{\mathbf{n}} \cdot \partial_{x^\alpha}\hat{\mathbf{n}}}{2}. \tag{26}$$

This NLSM is the simplest field theory of maps from the space $\mathcal{R}^{d+1}$ to the unit sphere. We still need to say a bit about the topological term Υ (the capital Greek letter upsilon): in one spatial dimension this term fundamentally modifies the physics, for reasons we shall see. We expand it for slowly varying $\hat{\mathbf{n}}(x)$: recall that Υ is defined to include factors η_i, so

$$\Upsilon^{d=1} = -S\sum_i (\omega[\hat{\mathbf{n}}(x_{2i})] - \omega[\hat{\mathbf{n}}(x_{2i-1})]) = \frac{S}{2}\int \frac{dx}{a}\frac{\delta\omega}{\delta\hat{\mathbf{n}}} \cdot \partial_x \hat{\mathbf{n}} a = 2\pi S \Theta[\hat{\mathbf{n}}(x,\tau)]. \tag{27}$$

Here Θ comes from using our previous variation form for the variation $d\omega$:

$$\Theta = \frac{1}{4\pi}\int d\tau \int dx\, (\hat{\mathbf{n}} \times \partial_\tau \hat{\mathbf{n}} \cdot \partial_x \hat{\mathbf{n}}). \tag{28}$$

This form is known as the Pontryagin index, which is a topological invariant like a winding number. It is an integer and is constant under smooth deformations of $\hat{\mathbf{n}}$. Essentially it measures the number of times the map from $(-L/2, L/2) \times (0, \beta)$ "wraps" the sphere S^2. You can easily construct examples with $\Theta = 0$ (the constant map) and $\Theta = 1$ (spherical projection). If you want a sense for why it is a topological invariant (which is not that hard to show), imagine that someone gives you a sphere wrapped with paper. The paper can't be "contracted to a point" without tearing, unlike a loop drawn on the sphere. So maps $S^1 \to S^2$ are all contractible, while maps $S^2 \to S^2$ are classified by the Pontryagin index.

The important thing to note is that the coefficient in front of this integer is only $2\pi S$, so that there will be a difference between integer and half-integer spins. For integer spin the topological term doesn't do anything, while for half-integer spins, there is interference between terms with odd or even values of the Pontryagin index. So Haldane's mapping explains (with a few approximations along the way!) the profound difference between integer and half-integer spins in one dimension, later confirmed experimentally and numerically. It is actually easier just to solve the spin-half chain using the Bethe ansatz than to explicitly solve its continuum theory with Berry phases, although a proof has been given that the latter is indeed gapless. The Lieb-Schultz-Mattis theorem discussed by Chalker elsewhere in this volume provides a general reason why the half-integer-spin case is gapless. Experimental results by neutron scattering (cf. Nagler et al.) confirm the existence of a gap and also the existence of spin-half edge states at the end of chains, which are discussed elsewhere in this volume by Regnault.

So we have seen how the unusual geometry of spin space, in the path-integral representation, gives rise to a profound difference between integer-spin and half-integer spin chains. Further reading and references for such topics can be found in the book of Auerbach. We can connect the above result to the exact solution by Affleck, Kennedy, Lieb, and Tasaki (AKLT) of a spin-1 chain with additional biquadratic interactions:

$$H = J\sum_i \left[S_i \cdot S_j + \alpha(S_i \cdot S_j)^3\right] \tag{29}$$

with $J > 0$ and $\alpha = 1/3$. The full phase diagram of the bilinear-biquadratic phase diagram from numerical density-matrix renormalization group studies is shown in Fig. 1. We note that the Haldane problem of the purely bilinear chain is in the same gapped phase as the AKLT solution, but that there are other phases as well, and there are also parameter values for which the system is gapless. As the last part of our discussion of topological terms for now, we explain the meaning of the labels $SU(3)_1$ and $SU(2)_2$ on the critical points in Fig. 1, which combine a Lie group with a subscripted integer known as the level. These points are examples of field theories with both conformal invariance and Lie group symmetry known as Wess-Zumino-Witten (WZW) models.

The NLSM for the XY model in equation (30) can be written in a different way if we think about the order parameter manifold (the circle) as the Lie group $U(1)$. Writing $g = e^{i\theta)}$, we note that $\partial_i \theta = g^{-1}\partial_i g$, so

$$Z_{\mathrm{NLSM}} = \int \mathcal{D}\theta(x) e^{-\beta c \int\, d^2x\, \sum_i (g^{-1}\partial_i g)^2/2)}. \tag{30}$$

In taking the trace here, we are looking ahead to a generalization. There is not a Lie group structure on the sphere, but we might be tempted to generalize to other Lie groups, for example by taking $g \in U(N)$ or $SU(N)$. Then $g^{-1}\partial_i g$ is an element of the Lie algebra, which has an inner product known as the Killing form; for $SU(N)$, $\mathcal{K}(X,Y) = 2N\mathrm{Tr}(XY)$. Generalizing the kinetic term that is the only term in the action above to the Lie algebra is straightforward.

However, it turns out that the low-energy physics of this generalization with just the resulting term is quite different than the $U(1)$ case. As for the NLSM into the sphere, the fact that the manifolds of unitary groups are curved once we go beyond the circle, leading to interactions that result in a mass gap. If we want instead to obtain a gapless model with Lie group symmetry, we must add an additional topological term first written down by Wess and Zumino. This term is quite unusual in that it requires extending the manifold on which the theory lives into an extra dimension. Assume $N > 1$ in what follows, and pick $g \in SU(N)$ for definiteness. Let us compactify the two-dimensional space into S^2 as for the Haldane chain above. Given a configuration of the Lie-group field g on the surface of a sphere, we can always find a way to smoothly deform that configuration to the constant configuration since $\pi_2(SU(N))$ is trivial.

We will keep writing the generalized model in Euclidean space although their primary relevance is to quantum models in one spatial dimension. The action of the Wess-Zumino-Witten model in the usual notation is then

$$S = -\frac{k}{8\pi}\int_{S^2} d^2x\, \mathcal{K}(g^{-1}\partial^\mu g, g^{-1}\partial_\mu g) - \frac{k}{24\pi}\int_{B^3} d^3y\, \epsilon^{ijk}\mathcal{K}(g^{-1}\partial_i g, [g^{-1}\partial_j g, g^{-1}\partial_k g]). \tag{31}$$

The meaning of upper and lower indices in the first term is that the metric of spacetime appears. In the second term, in contrast, the ϵ term appears instead of the metric, a sign that the term is topological in the sense of being metric-independent. In the second term, we have chosen a continuation of the field g into the interior B^3 of the sphere S^2. While as mentioned above those continuations certainly exist, we should check to make sure that the physics is independent of precisely which continuation we chose.

This independence is related to another topological fact about $SU(N)$. Consider two different continuations from S^2 into B^3. Actually, as a simpler example, consider two different continuations from S^1 into B^2. We could combine those into a field configuration on S^2, where one continuation gives the northern hemisphere and the other gives the southern hemisphere. In the same way, combining our two continuations from S^2 to B^3 gives a field configuration on S^3. Since $\pi_3(SU(N)) = \mathbb{Z}$, there are integer-valued classes of such configurations, and in fact the Wess-Zumino term is defined so as to compute this topological invariant Z: more precisely, the difference of the integral above for two different continuations into the bulk is k times $2\pi n$, where $n \in \mathbb{Z}$ measures the topological invariant of the map $S^3 \to S^3$ resulting from combining the two continuations as described above.

When we put this action into a quantum path integral, it therefore leads to a quantization of the level k to *integers*. $SU(2)_k$ with $k = 2$ can be viewed as a different representation of the same symmetry as the $SU(2)_1$ realized in the spin-half Heisenberg chain, in the same way as the spins on one site are in different representations of ordinary $SU(2)$. The full demonstration that the model is gapless is beyond our present scope, but at least we have a topological understanding of why the Wess-Zumino term is a natural quantity to consider.

IV. WEN-TYPE TOPOLOGICAL PHASES: THE FRACTIONAL QUANTUM HALL EFFECT

A. Chern-Simons theory I: flux attachment and statistics change

We will now start the process of developing a more abstract description of the fractional quantum Hall effect that will help us understand what type of order it has. For example, this will define precisely what it means to say that the physical state is adiabatically connected to the Laughlin wavefunction. Our main tool will be Chern-Simons theory; we briefly encountered the Chern-Simons term of the electromagnetic gauge potential when we discussed quantum Hall layers at the surface of the strong topological insulator, and we will come to that in a moment. However, a more fundamental use of the Chern-Simons theory is to describe the internal degrees of freedom of the quantum Hall

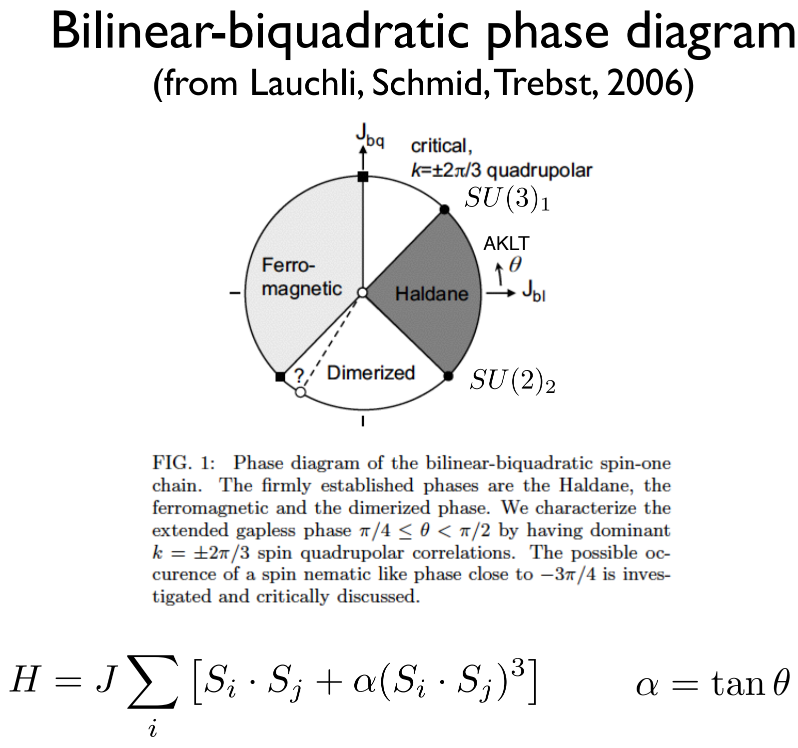

FIG. 1 Phase diagram of the bilinear-biquadratic spin-1 chain. The Hamiltonian is given in equation (29) with $\alpha = \tan\theta$.

liquid. In other words, we will have both an "internal" Chern-Simons theory describing the quantum Hall liquid and a Chern-Simons term induced in the electromagnetic action.

Since that sounds complicated, let's start by understanding why a Chern-Simons theory might be useful. To begin, we come up with a picture for the Laughlin state by noting that, since the filled lowest Landau level has one magnetic flux quantum per electron, the Laughlin state at $m = 3$ (i.e., $\nu = 1/3$) has three flux quanta per electron. To get a picture for how the Laughlin state is connected to the $\nu = 1$ state, we imagine attaching two of these flux quanta to each electron. The resulting "composite fermion" still has fermionic statistics, by the following counting. Interchanging two electrons gives a -1 factor. The Aharonov-Bohm factor from moving an electron all the way around a flux quantum is $+1$, but in this exchange process, each electron moves only half-way around the flux quanta attached to the other electron. So when one of these objects is exchanged with another, the wavefunction picks up three factors of -1 and the statistics is still fermionic.

These composite fermions now can form the integer quantum Hall state in the remaining field of one flux quantum per composite fermion, leading to a $\nu = 1/3$ incompressible state in terms of the original electrons. More generally, the phase picked up by a particle of charge q moving completely around a flux Φ is

$$e^{i\theta} = e^{iq\Phi/(\hbar c)}. \tag{32}$$

We will now see how the Chern-Simons term lets us carry out a "flux attachment" related to the above composite fermion idea: in fact, by attaching three flux quanta rather than two to each electron, we would obtain bosons moving in zero applied field, and the Laughlin state can be viewed as a Bose-Einstein condensate of these "composite bosons" (cf. Zhang, Hansson, and Kivelson, PRL 1988). [1]

The Abelian Chern-Simons theory we will study is described by the Lagrangian density in 2+1 dimensional

[1] One feature of the composite fermion picture that is preferable to the composite boson picture is that the former is naturally described as "topological order", while the latter would lead to a picture of the phase in terms of the symmetry-breaking order of a BEC.

Minkowski spacetime

$$\mathcal{L} = 2\gamma\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda + a_\mu j^\mu \tag{33}$$

where γ is a numerical constant that we will interpret later, a is the Chern-Simons gauge field, and j is a conserved current describing the particles of the theory. Under a gauge transformation $a_\mu \to a_\mu + \partial_\mu\chi$, the Chern-Simons term (the first one) transforms as

$$\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda \to \epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda + \epsilon^{\mu\nu\lambda}\partial_\mu\chi\partial_\nu a_\lambda, \tag{34}$$

where the term with two derivatives of χ drops out by antisymmetry. The new term can be written as

$$\delta S = 2\gamma\int d^2x dt \epsilon^{\mu\nu\lambda}\partial_\mu(\chi\partial_\nu a_\lambda), \tag{35}$$

where again the term with two derivatives acting on a gives zero by antisymmetry. So, *if we can neglect the boundary*, the Abelian Chern-Simons term is gauge-invariant. (As we discussed previously in the discussion of magnetoelectric polarizability, the non-Abelian Chern-Simons term is not gauge-invariant, because "large" (non-null-homotopic) gauge transformations change the integral; this is related to the third homotopy group of $SU(N)$.) Later on we will actually consider a system with a boundary and see how the boundary term leads to physically important effects.

Consider the equation of motion from varying this action. We get

$$4\gamma\epsilon^{\mu\nu\lambda}\partial_n u a_\lambda = -j^\mu. \tag{36}$$

where the 4 rather than 2 appears because the Chern-Simons term has nonzero derivative with respect to both a and ∂a. For a particle sitting at rest, the spatial components of the current vanish, but there must be a flux: writing in components,

$$\int d^2x(\partial_1 a_2 - \partial_2 a_1) = -\frac{1}{4\gamma}\int d^2x\, j^0. \tag{37}$$

Hence a charged particle in the theory gains a flux of the a field (since the left term is just the integral of a magnetic field). If the charge is localized, then the flux is localized as well.

What good is this? Well, we know that when one charged particle with respect to the a field moves around another, it will now pick up an Aharonov-Bohm phase from the attached flux in addition to any statistics factor. The additional statistics factor is

$$\theta = \frac{1}{8\gamma}, \tag{38}$$

where the 1/2 here results because the particles only move halfway around each other in an exchange. In other words, if we started with $\theta = 0$ bosonic particles but added a $\gamma = \frac{1}{8\pi}$ Chern-Simons term, we would obtain fermions, and vice versa. But so far nothing constrains γ, suggesting that in two dimensions, "braiding" statistics is not constrained to be bosonic or fermionic. Particles in two dimensions that are neither bosonic nor fermionic are known as "anyons".

Why is two spatial dimensions so special? It turns out that an argument about why generalized statistics are possible for point particles in two spatial dimensions but not higher dimensions was given long ago by Leinaas and Myrheim (1976). The key observation is that an exchange path that takes one particle around another and back to its original location is not smoothly contractible in 2D without having the particles pass through each other, while in higher dimensions, such a path is contractible. The consequence of this is that in two dimensions, phase factors are not just defined for permutations of the particles but rather for any "braiding". [2]

B. Chern-Simons theory II: integrating out gauge fields and coupling to electromagnetism

Aside from the composite fermion/composite boson pictures, why might the Chern-Simons theory with Lagrangian density given by (33) describe quantum Hall states? Without working through a detailed derivation starting from

---

[2] Even non-Abelian statistics are possible if there are multiple ground states: the phase factor associated with a particular braid is then a matrix acting on the set of ground states, and two such matrices need not commute.

nonrelativistic quantum mechanics of many interacting electrons in a magnetic field (which is still not all that rigorous; for a discussion, see lecture notes of A. Zee in *Field Theory, Topology, and Condensed Matter Physics*, Springer), we can note the following. A conserved electromagnetic current in 2+1D can always be written as the curl of a gauge field:

$$J^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu a_\lambda. \tag{39}$$

(Note that this electromagnetic current might in general be distinct from the particle current above.) Here a is automatically a gauge field since the $U(1)$ gauge transformation does not modify the current. Gauge invariance forbids the mass term $a^\mu a_\mu$, so the lowest-dimension possible term is the Chern-Simons term, which we write for future use with a different normalization than above:

$$\mathcal{L}_{CS} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda. \tag{40}$$

The point of the new normalization $k = 8\pi\gamma$ compared to (33) is that the boson-fermion statistics transformation above now corresponds just to $k = 1$. We will argue later that k should be an integer for the electron to appear somewhere in the spectrum of excitations of the theory.

Does this term need to appear? No, for example, in a system that has P or T symmetry, it cannot appear. However, if it does appear, then since there is only one spatial derivative, it dominates the Maxwell term at large distances. Effectively we define the quantum Hall phase as one in which $\mathcal{L}_{CS}$ appears in the low-energy Lagrangian; for example, this is true in both the Laughlin state and the physical state with Coulomb interactions, even though the overlap between those two ground-state wavefunctions is presumably zero in the thermodynamic limit.

What if we added the $a_\mu J^\mu$ coupling and integrated out the gauge field? Well, the main reason not to do that is that we obtain a nonlocal current-current coupling. Since the original action is quadratic in the fields, this integration is not too difficult, but an alternate, equivalent way to do it is to solve for a in terms of J. Given a general Lagrangian

$$\mathcal{L} = \phi\mathcal{Q}\phi + \phi J, \tag{41}$$

where $\mathcal{Q}$ denotes some operator, we have the formal equation of motion from varying ϕ

$$2\mathcal{Q}\phi = -J \tag{42}$$

which is solved by

$$\phi = \frac{-1}{2\mathcal{Q}}J. \tag{43}$$

Then substituting this into the Lagrangian (and ignoring some subtleties about ordering of operators), we obtain

$$\mathcal{L} = \frac{1}{4}J\frac{1}{\mathcal{Q}}J - J\frac{1}{2\mathcal{Q}}J = -J\frac{1}{4\mathcal{Q}}J. \tag{44}$$

So for the Chern-Simons term we need to define the inverse of the operator $\epsilon^{\mu\nu\lambda}\partial_\nu$ that appears between the a fields. This is a bit subtle because there is a zero mode of the original operator, related to gauge-invariance: for any smooth function g, $\epsilon^{\mu\nu\lambda}\partial_\nu(\partial_\lambda g) = 0$. To define the inverse, we fix the Lorentz gauge $\partial_\mu a_\mu = 0$. In this gauge, we look for an inverse using

$$(\epsilon^{\mu\nu\lambda}\partial_\nu)(\epsilon^{\lambda\alpha\beta}\partial_\alpha a_\beta) = \epsilon^{\mu\nu\lambda}\epsilon^{\lambda\alpha\beta}(\partial_\nu\partial_\alpha a_\beta). \tag{45}$$

We can combine the ϵ tensors by noting that $\epsilon^{\mu\nu\lambda} = \epsilon^{\lambda\mu\nu}$, so there are two types of nonzero terms in the above: either $\mu = \alpha$ and $\nu = \beta$ or vice versa, with a minus sign in the second case. From the first type of term, we obtain $\partial_\alpha(\partial_\beta a_\beta)$ which is zero by our gauge choice. From the second type, we obtain

$$-\partial_\nu^2 a_\mu. \tag{46}$$

So the inverse of the operator appearing in the Chern-Simons term in this gauge is $-\epsilon^{\mu\nu\lambda}\partial_\nu/\partial^2$, and the Lagrangian (33) with the gauge field integrated out is just

$$\mathcal{L} = \frac{1}{8\gamma}j_\mu\left(\frac{\epsilon^{\mu\nu\lambda}\partial_\nu}{\partial^2}\right)j_\lambda. \tag{47}$$

Aside from showing another interesting difference between the Chern-Simons term and the Maxwell term, we can use this inverse to couple the Chern-Simons theory to an external electromagnetic gauge potential $\mathcal{A}_\mu$. We will set $e = \hbar = 1$ except as noted. We do not include the Maxwell term to give this field dynamics, but rather view it as an imposed field *beyond the magnetic field producing the phase.* For example, we could use this additional field to add an electrical field, and we should find a Hall response. Let's try this:

$$\mathcal{L} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda - \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}A_\mu\partial_\nu a_\lambda = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda - \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu A_\lambda, \tag{48}$$

where in the second step we have dropped a boundary term and used the antisymmetry property of the ϵ tensor. Note that to obtain the second term we have just rewritten $A_\mu J^\mu$ using (39.

Now we can integrate out a_μ using equation (47) above, recalling $\gamma = k/(8\pi)$, and obtain

$$\mathcal{L}_{\mathrm{eff}} = \frac{\pi}{k}J_\mu\epsilon^{\mu\nu\lambda}\partial_\nu\frac{1}{\partial^2}J_\lambda = \frac{1}{4\pi k}\epsilon^{\mu\alpha\beta}\partial_\alpha A_\beta\epsilon^{\mu\nu\lambda}\partial_\nu\frac{1}{\partial^2}\epsilon^{\lambda\gamma\delta}\partial_\gamma A_\delta. \tag{49}$$

where in the second step we have used the rewritten Lagrangian in (48) to identify $J^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu A_\lambda$. As above, the nonzero possibilities are $\alpha = \nu$ and $\beta = \lambda$ (+1) or vice versa (-1), and also $\gamma = \mu$ and $\delta = \nu$ (+1) or vice versa (-1). Working through these, one is left with the $\gamma = \nu$ and $\delta = \mu$ terms,

$$\mathcal{L}_{\mathrm{eff}} = \frac{1}{4\pi k}\epsilon^{\mu\nu\lambda}A_\mu\partial_\nu A_\lambda. \tag{50}$$

This is the *electromagnetic* Chern-Simons term. The electromagnetic current is obtained by varying A:

$$J^\mu = -\frac{\delta\mathcal{L}_{\mathrm{eff}}}{\delta A_\mu} = \frac{1}{2\pi k}\epsilon^{\mu\nu\lambda}\partial_\nu A_\lambda. \tag{51}$$

where the factor of 2 is obtained because the variation can act on either A.

We can see immediately that this predicts a Hall effect: in response to an electrical field along x, we obtain a current along y. What about the factor $1/(2\pi)$? That is here just so that the response, once we restore factors of e and $\hbar$, is

$$\sigma_{xy} = \frac{e^2}{(2\pi)k\hbar} = \frac{1}{k}\frac{e^2}{h}. \tag{52}$$

Here we get a clue about the physical significance of k. Another clue is to consider the electromagnetic charge J^0 induced by a change in the magnetic field δB (i.e., an additional field beyond the one producing the FQHE):

$$J^0 = \delta n = \frac{1}{2\pi k}\delta B. \tag{53}$$

where we have written $J^0 = \delta n$ to indicate that this electromagnetic density describes the change in electron density from the ground state without the additional field. For the IQHE, a change of one flux quantum corresponds to one additional electron, while we can see that the $k = 3$ Chern-Simons theory predicts a change in density $e/3$, consistent with the quasihole and quasiparticle excitations.

To summarize what we have learned so far, we now see that Chern-Simons theory predicts a connection between the Hall quantum, the statistics of quasiparticles in the theory (from the previous section), and the effective density induced by a local change in the magnetic field. Here "quasiparticles", which we will discuss later, means whatever particle couples to the Chern-Simons theory as in the preceding section, which need not be an electron.

C. Chern-Simons theory III: topological aspects and gapless edge excitations

One obvious respect in which the Chern-Simons theory is topological is that, because ϵ rather than the metric tensor g was used to raise the indices, there is no dependence on the metric. In Zee's language, it describes a world without rulers or clocks. Since the stress-energy tensor in a relativistic theory is determined by varying the Lagrangian with respect to the metric, the stress-energy tensor is identically zero.

How can a theory be interesting if all its states have zero energy, as in the pure Chern-Simons theory? Well, one interesting fact is that the number of zero-energy states is dependent on the manifold where the theory is defined. We will not try to compute this in general but will solve the theory for the case of the torus. It is quite surprising that we can solve this 2+1-dimensional field theory exactly; the key will be that there are very few physical degrees of freedom once the $U(1)$ gauge invariance is taken into account.

We wish to solve the pure Chern-Simons theory with action

$$\mathcal{L}_{CS} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda \tag{54}$$

on the manifold $\mathbb{R}(\text{time}) \times T^2(\text{space})$. The gauge invariance is under $a_\mu \to a_\mu + \partial_\mu\chi$, χ an arbitrary scalar function. Given an arbitrary configuration of the gauge field a_μ, we first fix $a_0 = 0$ by the gauge transformation $a_\mu \to a_\mu + \partial_\mu\chi$ with $\chi = -\int a_0\,dt$. The Lagrangian is then

$$\mathcal{L} = -\frac{k}{4\pi}\epsilon^{ij}a_i\dot{a}_j, \tag{55}$$

where $i, j = 1, 2$. The equation of motion from varying the original Lagrangian with respect to a_0 now gives a constraint

$$\epsilon_{ij}\partial_i a_j = 0. \tag{56}$$

There is still some gauge invariance remaining in a_1, a_2: we can add a purely spatially dependent χ, so that a_0 remains 0, to make $\partial_i a_i = 0$ (exercise). Then $(a_i(t), a_j(t))$ have zero spatial derivatives and hence are purely functions of time. The Lagrangian (55) is now just the minimal coupling of a particle moving in a position-dependent vector potential; thinking of (a_1, a_2) as the coordinates of a particle moving in the plane, and noting that a constant magnetic field can be described by the vector potential $(By/2, -Bx/2) = (Ba_2/2, -Ba_1/2)$, we see that this is the interaction term of a particle in a constant magnetic field.

So far, using gauge invariance we can reduce the degrees of freedom from a 2+1-dimensional field theory to the path integral for the quantum mechanics of a particle moving in two dimensions. There is one last bit of gauge invariance we need to use. This will reduce the space on which our particle moves, which so far is $\mathbb{R}^2$ because the gauge fields are noncompact, to the torus T^2 on which the theory is defined. We consider a gauge transformation of the form $a_j \to a_j - iu^{-1}\partial_j u$, where u is purely a function of space. Note that if we can write $u = \exp(i\theta)$, this becomes a conventional gauge transformation $a_j \to a_j + \partial_j\theta$. This gauge transformation will not break the previous two gauge constraints if $\nabla^2\theta = 0$.

However, the periodicity of the torus means that we might not be able to define θ periodically, even if u is defined globally and the gauge transformation is indeed periodic. Taking the torus to be $L_1 \times L_2$, the following θ has zero Laplacian everywhere and gives rise to a periodic u and hence a periodic gauge transformation, even if θ is not itself periodic:

$$\theta = \frac{2\pi n_1 a_1}{L_1} + \frac{2\pi n_2 a_2}{L_2}. \tag{57}$$

The effect of this gauge transformation is that we can shift the particle's trajectory by an arbitrary constant integer multiple of L_1 in the x direction and L_2 in the y direction. To make the torus equivalent to the unit torus, we can rescale $a_i(t) = (2\pi/L_i)q_i(t)$. So finally we have shown

$$S = \int d^2x\,dt\,\frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda = -\frac{kL_1L_2}{4\pi}\int dt\,\frac{(2\pi)^2}{L_1L_2}\epsilon^{ij}q_i\dot{q}_j. \tag{58}$$

Here one L_1L_2 factor is from the spatial integrals and one is from the change of variable from a_i to q_i. We still haven't done anything quantum-mechanical to solve the path integral. However, we can temporarily add a term $m\dot{q}_i^2/2$ to the Lagrangian and recognize it as the path integral for a particle moving on the torus in a constant magnetic field. The gauge potential is $A_i = k\pi\epsilon_{ij}q_j$, which corresponds to a magnetic field $B = 2\pi k$ (this factor of 2 always appears in the rotational gauge). This is in our theorist's units with $\hbar = e = 1$; it means that there are a total of k flux quanta through the torus.

The limit we care about for pure CS theory is $m \to 0$, which takes all states not in the lowest Landau level to infinite energy. This makes sense because in a topological theory there can be no energy scale; the states either have some constant energy (the lowest Landau level here), which can be taken to zero, or infinite energy (the other Landau levels here). A quick calculation shows that there are exactly k states in the lowest Landau level on the torus pierced by k flux quanta; note that the "shift" of 1 extra level on the sphere is absent. For example, the lowest Landau level with one flux quantum through the *sphere* corresponds to the coherent-state path integral for a $s = 1/2$ particle (see problem sets), with 2 degenerate states.

The conclusion is that the parameter k also controls the ground-state degeneracy on the torus. An argument (X.G. Wen and Q. Niu, Phys. Rev. B41, 9377 (1990)) (regrettably direct calculation seems to be more difficult) shows that the general degeneracy of the pure Abelian CS theory on a 2-manifold of genus g is k^g. So for a topological theory, the physical content of the model is determined not just by explicit parameters in the action, such as k, but also by the topology of the manifold where the theory is defined. In this sense topological theories are sensitive to global or "long-ranged" properties, even though the theory is massive/gapped. (Of course, in the pure CS theory there is no notion of length so the distinction between local and global doesn't mean much, but adding a Maxwell term or something like that would not modify the long-distance properties; it would just mean that the other Landau levels are no longer at infinite energy.)

Bulk-edge correspondence

We noted above that the Chern-Simons term has different gauge-invariance properties from the Maxwell term: in particular, in a system with a boundary, it is not gauge-invariant by itself because the boundary term we found above need not vanish. Our last goal in this section is to see that this gauge invariance leads to the free massless chiral boson theory at the edge,

$$S_{\text{edge}} = \frac{k}{4\pi}\int dt\, dx\, (\partial_t + v\partial_x)\phi\partial_x\phi. \tag{59}$$

Here k is exactly the same integer coefficient as in the bulk CS theory, while v is a nonuniversal velocity that depends on the confining potential and other details. Note that the kinetic term here is "topological" in the sense that it does not contribute to the Hamiltonian, because it is first-order in time. The second term is not topological and hence shouldn't be directly obtainable from the bulk theory.

The theory of the bulk and boundary is certainly invariant under "restricted" gauge transformations that vanish at the boundary: $a_\mu \to a_\mu + \partial_\mu\chi$ with $\chi = 0$ on the boundary. From (35) above, the boundary term vanishes if $\chi = 0$ there. This constraint means that degrees of freedom that were previously gauge degrees of freedom now become dynamical degrees of freedom. We will revisit this idea later.

To start, choose the gauge condition $a_0 = 0$ as in the previous section and again use the equation of motion for a_0 as a constraint. [3] Then $\epsilon^{ij}a_j = 0$ and we can write $a_i = \partial_i\phi$. Substituting this into the bulk Chern-Simons Lagrangian

$$\begin{aligned} S &= -\frac{k}{4\pi}\int \epsilon^{ij}a_i\partial_0 a_j\, d^2x\, dt = -\frac{k}{4\pi}\int (\partial_x\phi\partial_0\partial_y\phi - \partial_y\phi\partial_0\partial_x\phi)\, d^2x\, dt \\ &= -\frac{k}{4\pi}\int (\partial_x(\phi\partial_0\partial_y\phi) - \partial_y(\phi\partial_0\partial_x\phi))\, d^2x\, dt \\ &= -\frac{k}{4\pi}\int (\nabla\times\mathbf{v})_z\, d^2x\, dt = -\frac{k}{4\pi}\int \mathbf{v}\cdot d\mathbf{l}\, dt, \end{aligned} \tag{60}$$

where $\mathbf{v}$ is the vector field

$$\mathbf{v} = (\phi\partial_0\partial_x\phi, \phi\partial_0\partial_y\phi). \tag{61}$$

(You might wonder why this doesn't let us transform the action simply to zero in the case of the torus studied in the previous section. The reason is that using Stokes's theorem in the second line, we have assumed the disk topology–since the torus has nontrivial topology, we are not allowed to use Stokes's theorem to obtain zero, cf. "Preliminaries" lecture notes.) So at the boundary, which we will assume to run along x for compactness, the resulting action is, after an integration by parts,

$$S_{\text{edge}} = \frac{k}{4\pi}\int \partial_t\phi\,\partial_x\phi\, dx\, dt. \tag{62}$$

We're almost done–this predicts a "topological" edge theory determined by the bulk physics; this edge theory is topological in that the Hamiltonian is zero. However, in order to obtain an accurate physical description we need to include non-universal, non-topological physics arising from the details of how the Hall droplet is confined. One approach to this is to start from a hydrodynamical theory of the edge and then recognize one term in that theory as S_{edge} above. The other term in that theory is a nonuniversal velocity term, and the combined action is

$$S_{\text{edge}} = \frac{k}{4\pi}\int (\partial_t\phi - v\partial_x\phi)\,\partial_x\phi\, dx\, dt. \tag{63}$$

---

[3] Here and before we are assuming that the Jacobians from our gauge-fixings and changes of variables are trivial. That this is the case is argued in S. Elitzur et al., Nuclear Physics B **326**, 108 (1989). Another nice discussion in this paper is how, for the non-Abelian case, the bulk can be understood as providing the Wess-Zumino term that keeps the edge theory gapless.

Here the nonuniversal parameter v clearly has units of a velocity, and in the correlation functions of the theory discussed below indeed appears as a velocity. The Hamiltonian density is

$$\mathcal{H} = \frac{kv}{4\pi}(\partial_x \phi)^2 \tag{64}$$

Note that for the Hamiltonian to be positive definite, the product kv needs to be positive: in other words, the sign of the velocity is determined by the bulk parameter k even thought the magnitude is not, and the edge is indeed chiral. (The density at the edge is found from the hydrodynamical argument to be proportional to $\partial_x \phi/(2\pi)$, so the above interaction term corresponds to a short-ranged density-density interaction; as usual, we will neglect the differences that arise if the long-ranged Coulomb interaction is retained instead.)

D. Chern-Simons theory IV: connecting edge theory to observables

We give a quick overview of how the above theory leads to detailed predictions of several edge properties. The general approach to treating one-dimensional electronic systems via free boson theories is known as "bosonization", and is the subject of several books. [4]. While we will not calculate the main results in detail, it turns out that there is a close similarity between the 1-dimensional free (chiral or nonchiral) boson Lagrangian and the theory of the algebraic phase of the XY model studied previously.

The reason such a connection exists is simple: the Euclidean version of the nonchiral version of the above free boson theory is just the 2D Gaussian theory. However, we know from the study of the XY model that subtleties such as the Berezinskii-Kosterlitz-Thouless transition arise when the variable appearing in the Gaussian theory is taken to be periodic, as when it describes an angular variable in that model. One of the surprising results we found was a power-law phase with continuously variable exponents: the correlations of spin operators $S_x + iS_y = \exp(i\theta)$ go as a power-law with the coefficient depending on the prefactor of the Gaussian.

The connection between the edge theory above and physical quantities is that the electron correlation function is represented in the bosonized theory as $e^{ik\phi}$: effectively ϕ describes a single quasiparticle and k quasiparticles make up the electron. The electron propagator in momentum space is likewise here found to have an exponent that depends on k: there is a factor of k^2 from the k's in the electron operator, and a factor of k^{-1} from the quasiparticle propagator since k appears as a coefficient in the Lagrangian. The result is

$$G(q,\omega) \propto \frac{(vq+\omega)^{k-1}}{vq-\omega}. \tag{65}$$

This describes an electron density of states $N(\omega) \propto |\omega|^{k-1}$, and this exponent can be measured in tunneling exponents: $dI/dV \propto V^{k-1}$. As a sanity check, the $k=1$ case describes a constant density of states and the predicted conduction is Ohmic: $I \propto V$.

Experimental agreement is reasonable but hardly perfect; at $\nu = 1/3$ the observed tunneling exponent $I \propto V^\alpha$ observes $\alpha \approx 2.7$, which is far from the Ohmic value ($\alpha = 1$) but reasonably close to the predicted value $\alpha = 3$. The tunneling exponent also does not appear to be perfectly constant when one is on a Hall plateau, as the theory would predict. Other measurements include "noise" measurements that attempt to see the quasiparticle charge directly, and in recent years interferometry measurements that try to check more subtle aspects of the theory.

In closing we comment briefly on the generalization of the above Chern-Simons and edge theories to more complicated (but still Abelian) quantum Hall states. These states, as suggested by the hierarchy picture, have multiple types of "particles", and two particles can have nontrivial statistics whether or not they belong to the same species. These statistics are defined by a universal integer "K matrix" that can be taken as a fundamental aspect of the topological order in the state. (Information must also be provided about the allowed quasiparticle types.) The resulting CS theory is

$$\mathcal{L} = \frac{1}{4\pi} K^{IJ} a^I_\mu \partial_\nu a^J_\lambda \tag{66}$$

This effective theory works for all but a few proposed quantum Hall states. I believe Prof. Stern will discuss these exotic "non-Abelian" quantum Hall states elsewhere in this volume.

[4] For example, M. Stone, *Bosonization*, World Scientific