A gauge theory generalization of the fermion-doubling theorem

S. M. Kravec and John McGreevy

Department of Physics, University of California at San Diego, La Jolla, CA 92093

It is possible to characterize certain states of matter by properties of their edge states. This implies a notion of "surface-only models": models which can only be regularized at the edge of a higher-dimensional system. After incorporating the fermion-doubling results of Nielsen and Ninomiya into this framework, we employ this idea to identify new obstructions to symmetry-preserving regulators of quantum field theory. We focus on an example which forbids regulated models of Maxwell theory with manifest electromagnetic duality symmetry.

Contents

Introduction 1

A simple family of topological field theories in 4+1 dimensions 3

Bulk physics

When is this an EFT for an SPT state? 4

This is a model of bosons 5

Surface states

2+1d case with boundary 5

4+1d abelian two-form gauge theory with boundary 5

Symmetries 6

Discussion

7d CS theory and the (2,0) superconformal theory 7

Would-be gauge anomalies and surface-only models 7

Appendix: A simple but non-unitary TFT 8

References 8

INTRODUCTION

This paper is about obstructions to symmetry-preserving regulators of quantum field theories (QFTs), in 3+1 spacetime dimensions. The most famous example of such an obstruction is the theorem of Nielsen and Ninomiya [1–4]. The basic statement of this theorem forbids regularization of free fermions preserving chiral symmetry\(^1\). We will approach the study of such obstructions by thinking about certain states of matter, in one higher dimension, with an energy gap (i.e. the energy of the first excited state is strictly larger than the energy of the groundstate, even in thermodynamic limit). More precisely, we will study the low-energy effective field theories of such states; below the energy gap, these are topological field theories (TFTs) in 4 + 1 dimensions. Such states will be difficult to realize in the laboratory. We will use them to demonstrate an obstruction to any regularization of Maxwell theory which preserves manifest electric-magnetic duality.

To convey the idea of how the study of such extra-dimensional models can be useful for understanding the practical question of symmetry-preserving regulators of 3+1-dimensional QFTs, we must digress on the subject of realizations of symmetries in QFT and in condensed matter. A basic question in condensed matter physics is to enumerate the possible gapped phases of matter. Two gapped phases are equivalent if they are adiabatically connected (varying the parameters in the Hamiltonian whose ground state they are to get from one to the other, without closing the energy gap). An important possible distinguishing feature of different phases is how the symmetries of the Hamiltonian are realized. This leads to Landau’s criterion which characterizes states by what symmetries of the Hamiltonian are broken by the groundstate. Considering this to be understood, it is interesting to refine the question to “How do we distinguish gapped phases that do not break any symmetries?”

A sophisticated answer to this question, vigorously advocated by Wen [5, 6], is topological order. A phase with topological order can be characterized by three related properties:\(^2\)

1. Fractionalization of symmetries, that is, emergent excitations which carry statistics or quantum numbers which are fractions of those of the constituents.

2. Topology-dependent groundstate degeneracy; this is a consequence of property 1, since the groundstates must represent the algebra of flux insertion operators associated with adiabatic braiding of quasiparticle-antiquasiparticle pairs.

3. Long-range entanglement, which may be quantified [7, 8] by the topological entanglement entropy \(\gamma\), a universal piece of the von Neumann entropy of the reduced

---

\(^1\) In odd spacetime dimensions, chiral symmetry is replaced by parity symmetry in this statement.

\(^2\) These are sufficient conditions; a complete characterization is possible in two dimensions in terms of adiabatic modular transformations [6] the generalization to higher dimensions seems not to be known yet.
density of region \( A \) of surface area \( \ell(\mathcal{A}) \) in the ground-state: \( S(A) = \ell(A)\Lambda - \gamma \) (where \( \Lambda \) is nonuniversal). For abelian states, \( \gamma \) is the logarithm of the number of torus groundstates, and vanishes for states with short-ranged entangled states [9].

Topological order is interesting and difficult. Recently, a simpler question has been fruitfully addressed: “What are possible (gapped) phases that don’t break symmetries and don’t have topological order?” (For a nice review, see the second part of [10].) In this paper we will use the spatial-topology-independence of the groundstate degeneracy as our criterion for short-range entanglement (SRE). The \( E_8 \) state in 2+1 dimensions [11–14] is a known example to this characterization.

A way to characterize such nearly-trivial states is to study them on a space with boundary. A gapped state in dimension \( d + 1 \) dimensions with short-ranged entanglement can be (at least partially) characterized (within some symmetry class of Hamiltonians) by (properties of) its edge states (i.e., what happens at an interface with the vacuum, or with another SRE state). The idea is simple: if we cannot adiabatically deform the Hamiltonian in time from one state to another, we must also not be able to deform the Hamiltonian in space from one state to another, without something interesting happening in between. The SRE assumption is playing an important role here: we are assuming that the bulk state has short-ranged correlations, so that changes we might make at the surface cannot have effects deep in the bulk.

A useful refinement of this definition incorporates symmetries of the Hamiltonian: An \textit{symmetry-protected topological (SPT) state}, protected by a symmetry group \( G \), is a SRE, gapped state, which is not adiabatically connected to a product state by local Hamiltonians preserving \( G \). Prominent examples include free fermion topological insulators in 3+1d, protected by \( U(1) \) and \( \mathbb{Z}_2^Q \), which have an odd number of Dirac cones on the surface. Free fermion topological insulators have been classified [15, 16]. Interactions affect the connectivity of the phase diagram in both directions: there are states which are adiabatically connected only through interacting Hamiltonians [17], and there are states which only exist with interactions, including all SPT states of bosons [18–21].

A simplifying property is that the set of SPT states (protected by a given symmetry group \( G \)) forms a group. The addition law is tensor product of Hilbert spaces, and addition of all interaction terms allowed by \( G \). The inverse operation is simply addition of the mirror image. We emphasize that with topological order, destroying the edge states does not modify the nontrivial physics (e.g., fractional charges) in the bulk: states with topological order do not form a group. A conjecture for the identity of this group (in \( d \) space dimensions, for given \( G \)) is the group cohomology group \( H^{d+1}(G, U(1)) \) [22]. Exceptions to this conjecture have been found [19, 20, 23, 24] and are understood by [12]. For our present purposes, we do not need to know how to classify these states. The existence of this group structure has the following implication [19, 20, 23–25], which we may pursue by examples.

Suppose that the edge theory (e.g., with vacuum) of a \( (D + 1) \)-dimensional SPT\(_G \) state were realized otherwise—that is, intrinsically in \( D \) dimensions, with a local Hamiltonian respecting \( G \). Then we could paint the conjugate local theory on the surface of a \( (D + 1) \)-dimensional space hosting the SPT state without changing anything about the bulk. By allowed (\( G \)-preserving, local) deformations of the surface Hamiltonian, we could then completely destroy the edge states. But this contradicts the claim that we could characterize the \( (D + 1) \)-dimensional SPT state by its edge theory. We conclude that the edge theories of SPT\(_G \) states cannot be regularized intrinsically in \( D \) dims, while preserving \( G \). We will refer to them as “surface-only models”. More generally, we can consider the interface between pairs of SRE states; the edge theory for interface between any pair of SPT states is also a surface-only theory.

The Nielsen-Ninomiya result is implied by this logic. Consider free massive (relativistic, for convenience) fermions in 4+1 dimensions: \( S = \int d^{4+1}\Psi\big(\partial^\mu + m_\Psi\big)\Psi \). The two signs of \( m \) label distinct phases. One proof of this arises by coupling to an external gauge field via the fermion number current \( \Delta S = \int d^4\chi A^\mu\gamma_\mu\Psi \). The effective action involves a quantized topological term:

\[
\log \int[D\Psi] e^{iS_{4+1}[\Psi,A]} = \frac{m}{|m|} \int \frac{d^5x}{24\pi^2\epsilon_{abcde}A_aF_{bc}F_{de}}.
\]

The domain wall between the two phases hosts an exponentially localized chiral fermion field [26, 27], a fact which has been useful for lattice simulations [27, 28].

A more famous \( D = 3 + 1 \) analog begins with a massive fermion in 3+1 dimensions:

\[
S = \int d^{3+1}\Psi \big(\partial^\mu + m_\Psi + i\hat{n}\gamma^5\big)\Psi
\]

If we demand time-reversal (\( \mathcal{T} \)) invariance, \( \hat{n} = 0 \), and \( \pm m \) label distinct SPT states protected by \( \mathbb{Z}_2^Q \). Coupling to an external gauge field

\[
\log \int[D\Psi] e^{iS_{3+1}[\Psi,A]} = \frac{m}{|m|} \int \frac{d^4x}{32\pi^2\epsilon_{abcd}F_{ab}F_{cd}}.
\]

produces a quantized magnetoelectric effect with \( \theta = 0, \pi \) [29]. The domain wall hosts a single Dirac cone in 2+1d. (We must emphasize that in both the above examples the protecting symmetry which distinguishes these states from the trivial insulator includes charge conservation.)

This suggests the following strategy for generalizing these results away from free fermions. Study a simple (unitary) gapped or topological field theory in 4+1 dimensions without
topological order, with symmetry $G$. Consider the model on the disk with some boundary conditions. The resulting edge theory is a “surface-only theory with respect to $G$” — it cannot be regulated by a local $3 + 1$-dim’l model while preserving $G$.

What does it mean to be a surface-only state? Such a model is perfectly consistent and unitary; it can be realized at the edge of some gapped bulk theory. However, it cannot be regularized in a local way consistent with the symmetries in absence of the bulk.

It (probably) means these QFTs will not be found as low-energy EFTs of solids or in cold atom lattice simulations. Why ‘probably’? This perspective does not rule out emergent topological order, with symmetry $G$. Consider the model on the disk with some boundary conditions. The resulting edge theory is a “surface-only theory with respect to $G$” — it cannot be regulated by a local $3 + 1$-dim’l model while preserving $G$.

A SIMPLE FAMILY OF TOPOLOGICAL FIELD THEORIES IN 4+1 DIMENSIONS

Here we will study simple field theories in 4+1 dimensions, whose path integral is gaussian. We follow an analog of the K-matrix approach used in [13] to study 2+1 dimensional SPT states.

Specifically, consider $2N_B$ 2-form potentials $B^I_{MN}$, in 4+1 dimensions, with action

$$S_{CS}[B] = \frac{K_{IJ}}{2\pi} \int_{\mathbb{R} \times \Sigma_4} B^I \wedge dB^J.$$  

(1)

$M, N = 0, \ldots, 4$ is a 4+1 dimensional spacetime index, $I, J = 1, \ldots, 2N_B$, and $\Sigma$ is a smooth 4-manifold (sometimes with boundary).

In $4\ell + 1$ dims, $K$ is a skew-symmetric integer $2N_B \times 2N_B$ matrix. This basic difference from 2+1 dimensional CS theory (where the K-matrix is symmetric) arises from the fact that the wedge product of two-forms is symmetric, and hence $B \wedge dB = \frac{1}{2} d(B \wedge B)$, and thus the symmetric part of $K$ produces a total derivative. The action is independent of a choice of metric on $\mathbb{R} \times \Sigma_{2p}$. Just as one may add an irrelevant Maxwell term to CS gauge theory, these models may be considered as the $g \to \infty$ limit of the non-topological models with ‘topologically massive’ [30] propagating two-forms with action

$$S = \int \frac{\tau_{IJ}}{M} dB^I \wedge *dB^J + S_{CS}[B];$$  

(2)

the mass scale $M$ determines the energy gap.

These models have a long history, including [31–41]. The canonical quantization of closely-related models was discussed in [35, 36]. (The difference from the models we study comes from issues related to the compactness of the gauge group, as in (4) below. This includes the quantization of the level and the resulting conclusion about the finiteness of the dimension of the Hilbert space.) The linking number of (homologically-trivial) pairs of surfaces was constructed as an observable of this theory in [36, 37].

Moore [42] refers to these models as both trivial and difficult. They are trivial in that the path integral is gaussian, and difficult in the sense that there are lots of subtleties in defining what’s being integrated over. We will not highlight all these subtleties, but we will endeavor to not be wrong. Moore and collaborators [42] have studied such generalizations of CS theory and their edge states with great care. In particular, the paper [40] constructs the partition function for the theory we study below and relates it to a piece of the $\mathcal{N} = 4$ super Yang-Mills partition function.

This relation arises because these models are realized as a ‘topological sector’ of type IIB strings on $AdS_5 \times X_5$ [38]; the simplest case with two 2-forms arises for $X_5 = S^3$, where the forms may be identified as $B^1 \equiv B_{NSNS}, B^2 \equiv C_{RR}$ which couple to fundamental strings and D1-branes respectively; the CS term (1) arises from

$$S_{IIB} \equiv \frac{1}{2\pi} \int_{AdS_5 \times S^5} F^{(5)}_{RR} \wedge B \wedge dC = \frac{N}{2\pi} \int_{\mathbb{R} \times \Sigma} B \wedge dC$$

and allows for the ending of $N$ F-strings holographically dual to the baryon vertex of $\mathcal{N} = 4$ SYM [44, 45]. S-duality of IIB string theory acts by interchanging the two-forms. This symmetry of (1) will play a crucial role below.

We focus on the Abelian case. The action $S$ is invariant under the gauge redundancies

$$B^I \simeq B^I + d\lambda^I$$  

(3)

where $\lambda^I$ are one-forms. We will impose large gauge equivalences:

$$B^I \simeq B^I + n^\alpha \omega_\alpha, \quad [\omega^\alpha] \in H^2(\Sigma, \mathbb{Z}), \quad n^\alpha \in \mathbb{Z}^{l_\Sigma(\Sigma)}$$  

(4)

where $b^2 \equiv \text{dim}H^2(\Sigma, \mathbb{Z})$ is the 2nd Betti number of $\Sigma$. In the case of 2+1 CS gauge theory of $\mathbb{R} \times \Sigma_2$, the analogous identification arises naturally from large gauge transformations [33, 34]

$$A \equiv A + ig^{-1}dg, \quad \text{with} \quad g(x) = e^{i \sum_{\alpha=1}^{l_\Sigma(\Sigma)} \int_\alpha n^\alpha \omega_\alpha}$$

3 This paper [43] gives a readable description of the dangers here. We can go a long way toward avoiding running afoul of the subtleties which are resolved by formulating the model in terms of differential cohomology if we only consider manifolds whose homology has no torsion. Even with this restriction we can say something interesting. More subtle distinctions between 4+1d phases that only arise on manifolds with torsion in their homology will have to wait for future work.
where $\omega_0$ form a basis of $H^1(\Sigma_2, \mathbb{Z})$, and $x_0$ is a base point. In the 4+1 dimensional case, we don’t know what a group-valued one-form is, but retain the natural identification (4). A mathematical formalism which produces this identification in a ‘natural’ way is described in [43]. This identification was not imposed in the otherwise-identical models studied in [35].

This class of models has been used [39, 40, 42, 46–50] to ‘holonomically’ define the partition function of the edge theory. (These papers focus mainly on the case of bulk spacetime dimension $D = 4\ell + 3$: 1+1d chiral CFTs, conformal blocks of the 5+1d $(2,0)$ theory.) In this paper, we are using this same relation to a different end. Finally, we note that the simplest model (1) with one pair of two-forms is equivalent to a $\mathbb{Z}_k$ 2-form gauge theory [48, 51]. The case we will be most interested in, with $k = 1$, can therefore be regarded as a kind of ‘$\mathbb{Z}_1$ gauge theory’, which nevertheless has something to teach us.

**BULK PHYSICS**

*When is this an EFT for an SPT state?*

For this subsection we suppose that $\partial \Sigma$ is empty. We wish to ascertain the size of the Hilbert space, and its dependence on the topology of $\Sigma$. Thinking of this as the EFT for some 4+1 dimensional gapped state of matter, at energies below the gap, this number is the groundstate degeneracy (up to $e^{-\text{(system size)-gap}}$ finite-size effects). If this degeneracy is dependent on the topology of $\Sigma$, then this state has topological order in the bulk [5]. A closely related calculation appears in section 6 of [43], in the case of a theory of $p$-forms in 2$p + 1$ dimensions, for the case of odd $p$.

Many aspects of the problem are analogous to the case of 2+1 dimensional CS 1-form gauge theory. The kinetic term in (2) is analogous to the Maxwell term, and $g$ has units of energy. With $g < \infty$ in (2), there are excited states of energy $E \propto g$, analogous to higher Landau levels. As in 2+1d, this is an example of an inclusion of metric dependence which does not change the nature of the phase. We will focus on the limit $g \to \infty$.

And as in the 2+1d case, the identification (4) means that the gauge-invariant operators are labelled by cohomology classes, here $[\omega_\alpha] \in H^2(\Sigma, \mathbb{Z})$: $$\mathcal{F}_\omega(m) \equiv e^{2\pi i mt} \exp{B^f}$$ where the identification on $B$ implies $m^f \in \mathbb{Z}$. These operators satisfy a Heisenberg algebra: $$\mathcal{F}_{\omega_\alpha}(m)\mathcal{F}_{\omega_\beta}(m') = \mathcal{F}_{\omega_\beta}(m')\mathcal{F}_{\omega_\alpha}(m)e^{2\pi im_\alpha m_\beta (K^{-1})^{ij} I_{\alpha\beta}}.$$ This kind of algebra is familiar from the quantum Hall effect, and from 2+1 dimensional CS theory. Its only irreducible representation is the Hilbert space of a particle on a ‘fuzzy torus’ of dimension $2N_B b^2(\Sigma)$. The number of states depends on the intersection form on the second homology of $\Sigma$: $$\int_\Sigma \omega_\alpha \wedge \omega_\beta = I_{\alpha\beta}.$$ This is a $b^2(\Sigma) \times b^2(\Sigma)$ symmetric matrix, which has various properties guaranteed by the theory of 4-manifold topology [52]. For a smooth, compact, oriented, simply-connected 4-manifold, it is unimodular (i.e. has $| \det I | = 1$). It is even if $\Sigma$ is a spin manifold (i.e. it admits spinor fields, i.e. its first Steiffel-Whitney class vanishes). Its signature is not definite.

Consider the simplest nontrivial case where $b^2(\Sigma) = 1$, which case $I$ is a $1 \times 1$ matrix. The simplest example is $\Sigma = \mathbb{P}^2$ where $I = 1$ [52].

We may arrive at the same conclusion by expanding the action in a basis of cohomology representatives: $$B^f = \sum_{\alpha=1}^{b^2(\Sigma_4, \mathbb{Z})} \omega_\alpha b^{f\alpha}(t), \quad \text{span}\{\omega_\alpha\} = H^2(\Sigma_4, \mathbb{Z}),$$ so $$S = \frac{K_{IJ}}{2\pi} \int dt \int_\Sigma \omega_\alpha \wedge \omega_\beta b^{f\alpha} b^{f\beta}$$ which describes a particle in $b^2(\Sigma)$ dimensions with a magnetic field in each pair of dimensions of strength $k$.

First we further assume that there is just one pair of $B$-fields, $N_B = 1$ and (WLOG by a $GL(N_B, \mathbb{Z})$ rotation) we take $K = k \alpha \sigma^\gamma$. Then the hilbert space is that of a particle on a periodic 1d lattice with $k$ sites: $$\mathcal{F}_1[\Omega] = [\Omega], \quad |w\rangle = \mathcal{F}_2^w[\Omega], \quad w = 1...k.$$ Here $\mathcal{F}_1 \equiv \mathcal{F}_1(1)$. (Note that $\mathcal{F}_1^1 = I_1$.)

Now keep this simplest $\Sigma$ but take more pairs of $B$-fields. We can skew-diagonalize the $K$-matrix, so that we find $N_B$ copies of the previous discussion, with various $k$, which are the skew-eigenvalues of $K$:

$$K = \begin{pmatrix}
0 & k_1 & 0 & 0 & \ldots \\
-k_1 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & k_2 & \ldots \\
0 & 0 & -k_2 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots 
\end{pmatrix}.$$ 

So the dimension of the Hilbert space of the bulk in this case is

$$\dim \mathcal{H}(b_2(\Sigma) = 1, I = 1) = \prod_{a=1}^{N_B} k_a = \text{Pf } K,$$

the pfaffian.

If $I > 1$, it multiples $k_a$ in the commutation relation. This gives

$$\dim \mathcal{H}(b_2(\Sigma) = 1, I) = \prod_{a=1}^{N_B} k_a = \text{Pf } (IK).$$
Now consider a more general 4-manifold where \(b_2 > 1\). Then there are \(b_2 N_B\) pairs of canonically-conjugate variables. By \(GL(2 b_2 N_B, \mathbb{Z})\) rotations, we can choose a basis to diagonalize \(\mathcal{I} \otimes K\). Since \(\mathcal{I}\) is symmetric, this matrix is still skew-symmetric. If we call its skew-eigenvalues \(\lambda_A\), and the number of groundstates is

\[
\dim \mathcal{H} = \prod_{A=1}^{b_2(\Sigma) N_B} \lambda_A = \text{Pf} (\mathcal{I} \otimes K) .
\]

The conclusion from this discussion is that if \(\mathcal{I} \otimes K\) has any skew-eigenvalues which are larger than one, the system has bulk topological order: a topology-dependent groundstate degeneracy.

This is a model of bosons

We observe that if \(\Sigma\) is a spin manifold, then we may take the entries in \(K\) to be half-integer, while preserving the consistency of the theory (in particular, the fact that the number of groundstates is an integer). A choice of spin structure was not required for the calculation we did here, but would be for other observables. Such models would provide low-energy effective theories for fermionic models, analogous to odd other observables. Such models would provide low-energy effective theories for fermionic models, analogous to odd other observables.

The preceding statements provide low-energy evidence that the models (1) for integer \(K\) can arise as the low-energy EFT for models of bosons. High-energy evidence for this claim comes from [53], which constructs a local bosonic lattice model which produces this EFT at low energies.

SURFACE STATES

2+1d case with boundary

Here we briefly review some relevant properties of 2+1d CS gauge theory on a spacetime of the form \(\mathbb{R} \times \Sigma\) with a spatial boundary, \(\partial \Sigma\) [5, 33, 34, 48, 49]. We will further assume for convenience that \(H^p(\Sigma, \partial \Sigma) = \delta^{0,p}\) – the topology (specifically, the relative cohomology) of the bulk is trivial.

The gauge group consists only of gauge transformations which go to the identity at the boundary. Transformations which act nontrivially do not leave the CS action invariant; if we wished to mod out by them we would need to add degrees of freedom to cancel this variation, and would arrive at the same description. The CS equation of motion for \(A_0\) is a constraint which imposes \(0 = F\) which means that

\[
A = g^{-1} dg
\]

which we can choose to vanish in the bulk. In the abelian case we can take \(A = d\phi\), with \(\phi \simeq \phi + 2\pi\) compact. We arrive at a theory of compact free bosons.

The Hilbert space of edge states is specified universally by the bulk – the CS action evaluated on (5) gives

\[
S_0 = \int_{\mathbb{R} \times \partial \Sigma} \frac{k_{IJ}}{4\pi} \partial_I \phi^I \partial_J \phi^J
\]

which determines the canonical equal-time commutators

\[
[\phi^I(x), \phi^J(x')] = 2\pi i k_{IJ} \theta(x-x') .
\]

However, the Hamiltonian governing these bosons comes entirely from the choice of boundary conditions, since the bulk action is linear in time derivatives. The non-universal velocity of the chiral edge modes is not encoded in the Chern-Simons action; this is natural from the quantum Hall point of view, where this velocity is determined by the slope of the potential confining the Hall droplet. The non-universal velocity term arises if we choose the ‘boosted’ boundary condition \(A_0 + v A_1|_{\partial \Sigma} = 0\), where \(x\) is the coordinate along the boundary. (Note that this is a boundary condition, not a gauge choice – \(v\) affects the physics.) This gives

\[
S[\phi] = \frac{k}{2\pi} \int (\partial_i + v \partial_{\Sigma}) \phi \partial_j \phi ,
\]

which shows that the edge boson is chiral; the sign of \(v\) is correlated with the sign of \(k\) in order that the energy be bounded below. Alternatively, and more covariantly, we may use units with \(v=1\) and impose \(A - * A|_{\partial \Sigma} = 0\) [42]. We may neglect the possibility of a matrix of velocities if we include a coupling to disorder [54, 55], which we should unless we are interested in SPT states protected by translation invariance.

4+1d abelian two-form gauge theory with boundary

As in the previous discussion, we consider the bulk spacetime to be \(\mathbb{R} \times \Sigma_4\) with nontrivial boundary \(\partial \Sigma_4\), and trivial topology in the bulk. The gauge transformations must act trivially at the boundary.

We will focus on the simplest case where \(K = k i \sigma^2\). Consider

\[
S[B, C] = \frac{k}{2\pi} \int_{\mathbb{R} \times \Sigma_4} (B \wedge dC - C \wedge dB) + \int_{\mathbb{R} \times \partial \Sigma_4} \left( \frac{1}{4g^2} C \wedge \ast_4 C + \frac{k}{4\pi} B \wedge C \right)
\]

A convenient boundary condition (imposed by free variation of (6) with the boundary terms indicated) is:

\[
\left( \frac{k}{2\pi} B - \frac{1}{2g^2} \ast_4 C \right) |_{\partial \Sigma_4} = 0 .
\]

The path integral over \(B\) produces a delta function forcing \(C\) to be flat [48]:

\[
\int |DB| e^{iS} = \delta[dC] \implies C = da .
\]

\[
S[C = da] = \frac{1}{4g^2} \int_{\mathbb{R} \times \partial \Sigma_4} da \wedge \ast_4 da
\]
This is ordinary Maxwell theory. We know how to regularize this theory with a local lattice boson model. We infer from the logic described above that we are forced to break some symmetry of (1) in order to realize its edge states in a local way. What symmetry must we break in writing such a local model?

We answer this burning question in the next section, after which we comment on more general boundary conditions. The general abelian case produces a collection of copies of Maxwell theory. This can be seen by block-diagonalizing $K$ by a GL$(2N_B, \mathbb{Z})$ rotation.

**SYMMETRIES**

CS gauge theories in 2+1d have bulk ‘topological’ conserved currents. When these theories arise from quantum Hall states, some linear combination of these currents can be interpreted as electron number:

$$0 = \partial^\mu J_\mu \implies J_\mu = \frac{1}{2\pi} \epsilon_{\mu\nu\rho} \partial_\nu A_\rho.$$  

$J$ is conserved when $A$ is single-valued. In 4+1d, the analog is pairs of string currents

$$J^I_{\mu
u} = \frac{1}{2\pi} \epsilon_{\mu
u\rho\sigma} \partial_\rho B^I_{\sigma\lambda}.$$  

We can use these symmetries to demonstrate that different $K$ label nontrivial, different states.

In $D = 2+1$ CS theory, we can couple the particle currents $J^I = * dA^I$ to external 1-form potentials, $A_I$:

$$\log \int [DA^I] e^{\int \frac{1}{2\pi} A^I dA^I} J_1 A^I = \int_{2+1} (4\pi K^{-1})_{IJ} A^I dA^J$$

This term in the effective action demonstrates a quantized Hall response; in the absence of topological order ($\det k = 1$), $\sigma_{xy} = \frac{1}{2\pi}$ is an integer. This is one way to distinguish the integer quantum Hall state from a trivial insulator. In the 2+1d case, the thermal Hall response $\propto c_L - c_R$ makes this distinction even in the absence of charge conservation.

The analogous stratagem in $D = 4+1$ is to couple the string currents $J^I = * dB^I$ to external 2-form potentials, $B_I$:

$$\log \int [DB^I] e^{\int \frac{k}{2\pi} B dB^I + \int J_1 B^I} = \int (4\pi K^{-1})_{IJ} B^I dB^J.$$  

This term exhibits a quantized ‘string Hall’ response to external 3-form field strengths, which again is an integer in the absence of topological order, $\text{Paff}(K) = 1$. This response distinguishes this state from the completely trivial state of bosons in 4+1 dimensions.

- Translation invariance is a red herring. In fact, breaking translation invariance with static disorder helps to produce a uniform speed of light in the edge theory. Indeed, the lattice model [53] should have the same edge states.

- Stringy symmetries: $J^B_{(0)b} = E_t, J^C_{(0)b} = -B_t$. $E_t = \partial_t a_t - \partial_t a_t B_t \equiv \epsilon_{tij}(\partial_t a_j - \partial_j a_i)$ are ordinary E&M fields

$$J^B_{y0} = \epsilon_{ijk}\partial_t C_{jk} = \epsilon_{ijk}\partial_t a_j a_k = \vec{\nabla} \cdot \vec{B}.$$  

This is ordinary charge, of course it has to be conserved.

- $C$: $(B, C) \rightarrow -(B, C)$ is $(\vec{E}, \vec{B}) \rightarrow -(\vec{E}, \vec{B})$. This is preserved in pure U(1) lattice gauge theory.

- $\mathcal{T}$: $t \rightarrow t, x^M \rightarrow -x^M, i \rightarrow -i, B \rightarrow -B, C \rightarrow C$ as two-forms. Acts in the usual way on the EM field as $(E, B) \rightarrow (E, -B)$.

**Electric-magnetic Duality**: $(B, C) \rightarrow (C, -B)$

The global symmetry which interchanges $B$ and $C$ is a manifest symmetry of the bulk theory. It acts on the boundary gauge field as electromagnetic duality $(\vec{E}, \vec{B}) \rightarrow (\vec{B}, -\vec{E})$.

- Just as in string theory and gauge theory [56], the $\mathbb{Z}_2$ EM transformation just described is a subgroup of a classical SL$(2, \mathbb{R})$ symmetry acting on $(B, C)^T$ as a doublet $\begin{pmatrix} B \\ C \end{pmatrix} \rightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} B \\ C \end{pmatrix}$ with $ad - bc = 1$. The continuous parts of this group, where $a, b, c, d \notin \mathbb{Z}$ do not preserve the integrality condition (4), and are not symmetries of the quantum theory. The SL$(2, \mathbb{Z})$ subgroup is generated by

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. $$  

$S$ is the EM duality transformation above, and $T$ adds a boundary term $\int_{\partial B \times \partial \Sigma} \frac{k}{4\pi} C \wedge C$; this shifts the $\theta$ angle of the surface gauge theory by $2\pi$.

This EM duality symmetry is effectively unbreakable in the 4+1d CS theory. Even with arbitrary boundary conditions, breaking Lorentz invariance, the scaling freedom in the relationship between $B, C$ and $a, \bar{a}$, and in the duality map itself always allow a symmetry operation exchanging $a$ and $\bar{a}$. The general boundary condition is of the form

$$0 = (B + c_1 C + c_2 \ast B + c_3 \ast C) |_{\partial \Sigma}. $$
Of the three real coefficients $c_i$, one may be absorbed in the speed of light, a second may be absorbed in the relationship between $C$ and the Maxwell potential $a$, and the third may be absorbed in the duality map, $(B, C) \rightarrow (AC, \lambda^{-1} B)$.

The only way to break the manifest EM duality symmetry of the bulk is to add charged matter, in the form of strings which end at the boundary. Any matter we add has a definite charge vector $(q_e, q_m)$. Even if the matter comes in dual pairs, condensing matter with a definite charge vector will gap out the boundary photon (and the matter with the dual charge vector).

Charged matter in the surface Maxwell theory arises from strings which are minimally coupled to the bulk 2-forms and terminate at the boundary. The precise spectrum of matter is described; it is specified in a lattice UV completion [53].

We conclude that it is not possible to regularize Maxwell theory in a way which retains manifest electromagnetic duality symmetry. This is consistent with the existence of $U(1)$ lattice gauge theory, and its UV completions in terms of lattice models with local Hilbert spaces [11, 57], in which EM duality is not manifest on the lattice. Of course, we still do expect this symmetry to emerge in the IR limit of these models.

**DISCUSSION**

7d CS theory and the (2,0) superconformal theory

There are many generalizations of the kind of nearly-trivial bulk theory we have studied above. Consider the following 6+1d Chern-Simons theory

$$S_7[C^{(3)}] = \frac{k}{4\pi} \int_{\mathbb{R} \times \Sigma_6} C^{(3)} \wedge dC^{(3)}$$

For $k = 1$, there is no topological order, and the model is completely trivial in the bulk. To study the edge states, we examine solutions of the bulk equations of motion $C^{(3)} = dc^{(2)}$, with the boundary condition: $C_{0ij} = v(\star_6 C)_{1ij}$. The resulting action is

$$S_7[C^{(3)} = dc^{(2)}] = \frac{k}{4\pi} \int_{\mathbb{R} \times \Sigma_6} \epsilon \partial c^{(2)} \cdot \left( \partial c^{(2)} + v \epsilon \partial c^{(2)} \right).$$

c is a self-dual 2-form potential in 5+1d. In this case (as in the 2+1d CS theory and the integer QHE), no symmetry is required to protect the edge states. We conclude that this model cannot be regularized intrinsically in 5+1 dimensions.

This model describes the ‘topological sector’ of the $A_0 (2,0)$ superconformal theory in 6d [38, 39, 50] – the worldvolume theory of an M5-brane in M-theory. The conjecture that these exotic QFTs can be consistently decoupled from gravity underlies much recent progress in string theory (e.g. [58] and references thereto). In particular, by compactification, its existence makes various deep 4d QFT dualities manifest. This model reduces to the simple model without fermions or scalars when the partition function preserves enough supersymmetry [39]. We conclude that the (2,0) theory cannot be regularized preserving a sufficient amount of supersymmetry to allow this. It is possible that further inquiry in this direction can provide guidance for better definitions of string theory.

Would-be gauge anomalies and surface-only models

Many surface-only obstructions, including the Nielsen-Ninomiya examples, are directly related to anomalies in the protecting symmetry: They would be gauge anomalies if we tried to gauge the protecting symmetry. This property has been used to identify nontrivial SPT states in [59].

We described an obstruction to regularizing a self-dual 2-form theory in $D = 5 + 1$; this, too, can be understood in terms of a known anomaly. Just as for chiral CFTs (the chiral boson theory is a theory of a self-dual 1-form field strength), a gravitational anomaly was relevant here. In 1+1 dimensions, the chiral central charge $c_{L} - c_{R}$, which determines the magnitude of the thermal Hall effect, measures an anomaly that obstructs coupling the theory to gravity. A similar gravitational anomaly arises for the 6d self-dual tensor theory [60].

Are all of surface-only obstructions to realizing a symmetric regulator merely anomalies we would find should we try to gauge the symmetry? Thinking of anomalies as ‘symmetries broken by any regulator’ the answer would seem to be ‘yes’ [25]. It is true that the NN theorem can be interpreted as the statement that no chirally-symmetric regulator can produce the correct chiral anomaly in a background gauge field coupled to the vector current.

It is not clear, however, whether our present notion of ‘anomaly’ is general enough to make this the correct answer. Here is some evidence that there can be obstructions more general than obstructions to gauging symmetries:

1. A known example [59] is SPT states protected by time-reversal $\mathbb{Z}_T^+$ . It is not clear what it would mean to gauge an antunitary symmetry (all numbers are real?), so it is not clear that obstructions to regularization arising from demanding $\mathbb{Z}_T^+$ symmetry can be thought of as would-be discrete gauge anomalies.

2. From our discussion above, we can conclude that it is impossible to gauge electromagnetic duality in Maxwell theory. Previous literature suggesting that it is impossible to gauge EM duality includes [61–63]. In other models in other dimensions, it is possible to gauge analogs of EM duality [64]. We do not know how to describe this phenomenon in terms of a known $\mathbb{Z}_2$ anomaly. We note, however, that a model which resulted from identifying configurations related by the $S$
transformation could not be CPT invariant, since CPT in \(4k\) spacetime dimensions relates helicity \(\pm 1\) states. We observe (following [48]) that the model we get on the surface is exactly in the form given by [56]. We are forced to conclude that the manifestly EM-duality-invariant (but not manifestly Lorentz-invariant) model described in [56] cannot be regularized. (Note that Lorentz invariance is not the protecting symmetry: breaking Lorentz-invariance in the bulk of the 5d CS theory has no effect on the surface states.)

3. Finally, it would be very interesting to find obstructions to supersymmetry-preserving regulators. Gauging supersymmetry leads to supergravity. Consistency of quantum supergravity is much harder to verify than the absence of quantum violations of the current.

Acknowledgements: We would like to thank Maissam Barkeshli, Xiaoliang Qi, Congjun Wu, Cenke Xu and especially T. Senthil and Brian Swingle for discussions. We also thank Brian Swingle for collaboration on the related work [53], and for comments on the manuscript. JM is grateful for the Perimeter Institute for hosting the workshop Entanglement and Emergence, II, where part of this work was done. This research was supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under cooperative research agreements DE-FG0205ER41360 and DE-SC0009919, in part by the Alfred P. Sloan Foundation, and in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development & Innovation.

Appendix: A simple but non-unitary TFT

Here we consider analogous gaussian models where the fields are grassman-valued \(p\)-forms in \(2p+1\) dimensions:

\[
S = \frac{K_{IJ}}{4\pi} \int \gamma^I \wedge d\gamma^J.
\]

Our motivation for thinking about this was that this reverses the behavior of the K matrices as a function of the parity of \(p\), relative to the case of bosonic forms. That is: in \(2+1\) dimensions (odd \(p\)), the K-matrix is antisymmetric, while in \(4+1\) dimensions (even \(p\)) the K-matrix is symmetric.

There is an analog in \(1+1\) dimensions which is called the \(bc\) CFT, an example of which arises in string theory. There \(b,c\) are grassmann fields with action \(\int b\partial c\). There one can consider the spin of \(b,c\) to be \((\lambda, 1-\lambda)\) respectively; it is a CFT for any \(\lambda\) with (chiral) central charge \(c_\lambda = -3(2\lambda - 1)^2 + 1. \lambda = \frac{1}{2}\) is free complex fermions, \(c_\frac{1}{2} = 1\). The spin doesn't make any difference in terms of counting of degrees of freedom in \(1+1d\), it just affects the form of the stress tensor. If we take the spins to be 0, 1 (\(\lambda = 0\)), so that we have a one-form and a zero-form we get \(c = -2\). It is a non-unitary theory (the central charge is positive in a unitary CFT).

Such a system violates the usual connection between spin and statistics, which is general for unitary relativistic quantum field theories (indeed we will find that it is not unitary). If we add a Maxwell-like term \(\frac{1}{2} \partial^2 (\partial\gamma)^2\) then surely this action will propagate ghosts. But in the topological limit \(m \to 0\), nothing propagates, and we might hope it is unitary, but has no relativistic UV completion. As we will see by finding the space of states on a closed manifold, it is not.

If we study the quantization of this model on a closed manifold \(\Sigma_{2p}\) we can expand

\[
\gamma^I = \sum_{\alpha=1}^{b(p+1,\mathbb{Z})} \omega_\alpha \theta^{\alpha I}
\]

and find

\[
S = \int dt \frac{K_{IJ}}{4\pi} \mathcal{I}_{\alpha\beta} \theta^{\alpha I} \dot{\theta}^{\beta J},
\]

with \(\mathcal{I}_{\alpha\beta} = \int_{2p} \omega_\alpha \wedge \omega_\beta\) the intersection form on \(p\)-cycles. Quantizing this system leads to the canonical anticommutation relations

\[
\{\theta^{\alpha I}, \theta^{\beta J}\} = 4\pi \hbar (K^{-1})^{IJ} (\mathcal{I}^{-1})^{\alpha\beta}.
\]

In any dimension \(2p\) there are manifolds with indefinite signature of the intersection form on the middle homology. If the RHS is of indefinite signature, the Hilbert space representing this algebra has states of negative norm.

\[1\] H. B. Nielsen and M. Ninomiya, “Absence of Neutrinos on a Lattice. 1. Proof by Homotopy Theory,” Nucl. Phys. B185 (1981) 20.

\[2\] H. B. Nielsen and M. Ninomiya, “Absence of Neutrinos on a Lattice. 2. Intuitive Topological Proof,” Nucl. Phys. B193 (1981) 173.

\[3\] H. B. Nielsen and M. Ninomiya, “No Go Theorem for Regularizing Chiral Fermions,” Phys. Lett. B105 (1981) 219.

\[4\] D. Friedan, “A PROOF OF THE NIELSEN-NINOMIYA THEOREM,” Commun.Math.Phys. 85 (1982) 481–490.

\[5\] X.-G. Wen, Quantum Field Theory of Many-Body Systems. Oxford Univ. Press, Oxford, 2004.

\[6\] X.-G. Wen, “Topological order: from long-range entangled quantum matter to an unification of light and electrons,” 1210.1281.

\[7\] M. Levin and X.-G. Wen, “Detecting Topological Order in a Ground State Wave Function,” Phys. Rev. Lett. 96 (Mar, 2006) 110405.

\[8\] A. Kitaev and J. Preskill, “Topological Entanglement Entropy,” Phys. Rev. Lett. 96 (Mar, 2006) 110404.

\[9\] T. Grover, A. M. Turner, and A. Vishwanath, “Entanglement entropy of gapped phases and topological order in three dimensions,” Phys. Rev. B 84 (Nov., 2011) 195120, 1108.4038.
[10] A. M. Turner and A. Vishwanath, “Beyond Band Insulators: Topology of Semi-metals and Interacting Phases,” Commun.Math.Phys. 121 (1989) 351.
[34] S. Elitzur, G. W. Moore, A. Schwimmer, and N. Seiberg, “Remarks on the Canonical Quantization of the Chern-Simons-Witten Theory,” Nucl.Phys. B326 (1989) 108.
[35] G. T. Horowitz, “Exactly Soluble Diffeomorphism Invariant Theories,” Commun.Math.Phys. 125 (1989) 417.
[36] M. Blau and G. Thompson, “Topological Gauge Theories of Antisymmetric Tensor Fields,” Annals Phys. 205 (1991) 130–172.
[37] G. T. Horowitz and M. Srednicki, “A QUANTUM FIELD THEORETIC DESCRIPTION OF LINKING NUMBERS AND THEIR GENERALIZATION,” Commun.Math.Phys. 130 (1990) 83.
[38] E. Witten, “AdS / CFT correspondence and topological field theory,” JHEP 9812 (1998) 012, hep-th/9812012.
[39] G. W. Moore, “Anomalies, Gauss laws, and Page charges in M-theory,” Comptes Rendus Phys. 6 (2005) 251–259, hep-th/0409158.
[40] D. Belov and G. W. Moore, “Conformal blocks for AdS(5) singletons,” hep-th/0412167.
[41] S. A. Hartnoll, “Anyonic strings and membranes in AdS space and dual Aharonov-Bohm effects,” Phys.Rev.Lett. 98 (2007) 111601, hep-th/0612159.
[42] G. Moore, “Minicourse of three lectures on Generalized Abelian Gauge Theories, Self-Duality, and Differential Cohomology, at the Simons Center Workshop on Differential Cohomology, Simons Center for Geometry and Physics, Stonybrook.”
[43] D. S. Freed, G. W. Moore, and G. Segal, “Heisenberg Groups and Noncommutative Fluxes,” Annals Phys. 322 (2007) 236–285, hep-th/0605200.
[44] D. J. Gross and H. Ooguri, “Aspects of large N gauge theory dynamics as seen by string theory,” Phys.Rev. D58 (1998) 106002, hep-th/9805129.
[45] E. Witten, “Baryons and branes in anti-de Sitter space,” JHEP 9807 (1998) 006, hep-th/9805112.
[46] S. Shatashvili, “unpublished.”
[47] E. Witten, “On flux quantization in M theory and the effective action,” J.Geom.Phys. 22 (1997) 1–13, hep-th/9609122.
[48] J. M. Maldacena, G. W. Moore, and N. Seiberg, “D-brane charges in five-brane backgrounds, Appendix A,” JHEP 0110 (2001) 005, hep-th/0108152.
[49] D. Belov and G. W. Moore, “Classification of Abelian spin Chern-Simons theories,” hep-th/0505235.
[50] D. Belov and G. W. Moore, “Holographic Action for the Self-Dual Field,” hep-th/0605038.
[51] T. H. Hansson, V. Ogievetsky, and S. L. Sondhi, “Superconductors are topologically ordered,” Annals of Physics 313 (Oct., 2004) 497–538, arXiv:cond-mat/0404327.
[52] S. K. Donaldson and P. B. Kronheimer, The Geometry of Four-Manifolds. Oxford Univ. Press, Oxford, 1990.
[53] S. Kravcev, J. McGreevy, and B. Swingle, “work in progress.”
[54] C. Kane, M. P. Fisher, and J. Polchinski, “Randomness at the edge: Theory of quantum Hall transport at filling nu = 2/3,” Phys.Rev.Lett. 72 (1994) 4129–4132.
[55] C. Kane and M. P. Fisher, “Edge state transport.”
[56] J. H. Schwarz and A. Sen, “Duality symmetric actions,” Nucl.Phys. B411 (1994) 35–63, hep-th/9304154.
[57] M. A. Levin and X.-G. Wen, “String-net condensation: A physical mechanism for topological phases,” Phys. Rev. B 71 (Jan, 2005) 045110.
[58] D. Gaiotto, “N=2 dualities,” JHEP 1208 (2012) 034, 0904.2715.
[59] M. Levin and Z.-C. Gu, “Braiding statistics approach to symmetry-protected topological phases,” Phys. Rev. B 86 (Sept., 2012) 115109, 1202.3120.
[60] L. Alvarez-Gaume and E. Witten, “Gravitational Anomalies,” Nucl. Phys. B234 (1984) 269.
[61] S. Deser, “No local Maxwell duality invariance,” Class. Quant. Grav. 28 (2011) 085009, 1012.5109.
[62] C. Bunster and M. Henneaux, “Sp(2n,R) electric-magnetic duality as off-shell symmetry of interacting electromagnetic and scalar fields,” PoS HRMS2010 (2010) 028, 1101.6064.
[63] A. Saa, “Local electromagnetic duality and gauge invariance,” Class. Quant. Grav. 28 (2011) 127002, 1101.3927.
[64] M. Barkeshli and X.-G. Wen, “Effective field theory and projective construction for Z_k parafermion fractional quantum Hall states,” Phys. Rev. B 81 (Apr., 2010) 155302, 0910.2483.