Anomaly Constraints on Gapped Phases with Discrete Chiral Symmetry

Clay Córdova$^{1,}$ and Kantaro Ohmori$^2$

$^1$Kadanoff Center for Theoretical Physics & Enrico Fermi Institute, University of Chicago
$^2$Simons Center for Geometry and Physics, SUNY

(Dated: January 1, 2020)

We prove that in $(3 + 1)d$ quantum field theories with $\mathbb{Z}_N$ symmetry, certain anomalies forbid a symmetry-preserving vacuum state with a gapped spectrum. In particular, this applies to discrete chiral symmetries which are frequently present in gauge theories as we illustrate in examples. Our results also constrain the long-distance behavior of certain condensed matter systems such as Weyl-semimetals and may have applications to crystallographic phases with symmetry protected topological order. These results may be viewed as analogs of the Lieb-Schultz-Mattis theorem for continuum field theories.

Introduction— Symmetry is a universally applicable tool to constrain the dynamics of strongly-coupled quantum field theories. At the most foundational level, a symmetry organizes energy eigenstates into representations of the symmetry group and provides selection rules constraining the dynamics.

Beyond these elementary considerations there are more subtle aspects of symmetry in quantum field theory such as ‘t Hooft anomalies. These are mild violations of gauge invariance of the partition function of the theory coupled to classical background gauge field sources. Such violations are characterized by local functionals and are intrinsic to the theory when they cannot be eliminated by a choice of local counterterms.

One way to understand anomalies and their properties is via anomaly inflow [1]. In this framework the anomaly of a theory is characterized by a local classical action for background gauge fields in one higher spacetime dimension. These classical actions are sometimes called invertible field theories following [2]. The notion of locality invoked here means that the action obeys certain cutting and gluing rules and applies even for discrete background gauge fields appropriate for discrete global symmetries. The classification of invertible field theories, and in turn anomalies, has a close connection with topology and cobordism theory \[3,10\]. In condensed matter physics, the long-distance limit of symmetry-protected topological order (SPT) is described by such a classical action. The anomalous field theory is then a set of edge modes which may reside on the boundary of the SPT.

A crucial feature of anomalies is that they are invariant under any continuous deformation of a quantum field theory. This includes changing the energy scale of observation so that anomalies are invariant under renormalization group flow \[11\]. Anomalies thus provide powerful input on the possible long-distance behavior of strongly-coupled quantum field theories.

It is useful to organize the possible long-distance physics of a given quantum field theory by the behavior of the vacuum under the global symmetry and whether or not there exists a gap in the energy spectrum. Anomalies computed in a weakly coupled ultraviolet description can then constraint which of these various behaviors is possible at long distances. In particular, while it is believed that any anomaly can be carried by a symmetry breaking, or gapless vacuum, it is sometimes the case that a given anomaly is inconsistent with a gapped and symmetry-preserving vacuum. The fluctuations around such a low energy phase are described by a topological quantum field theory so one can rephrase the above by stating that the possible anomalies of unitary topological field theories with a unique ground state on the sphere are restricted.

Our main result below is to illustrate an example of such a restriction in the context of $(3+1)d$ quantum field theories with $\mathbb{Z}_N$ global symmetry. This symmetry and anomaly occurs in a wide variety of theories, in particular in gauge theories with discrete chiral symmetry where our result implies that a mass gap necessitates discrete chiral symmetry breaking.

The fact that anomalies can prohibit a gap in the energy spectrum is in some cases well-known. For instance in $(3 + 1)d$ field theories, the most familiar cubic anomalies of continuous symmetry groups arising from triangle diagrams in perturbative theories have this feature. One way to argue this is that such a diagram represents a portion of the three-point function of current operators with power law behavior at separated points and hence requires massless degrees of freedom [12].

By contrast, the implications of discrete anomalies, i.e. either discrete anomalies for continuous groups such as parity anomalies in $(2 + 1)d$, or anomalies for finite symmetry groups, is less clear. In particular, these anomalies do not have any obvious imprint on the correlation functions of local operators and so the above logic does not apply. There are general constructions [13,17] showing that large classes of anomalies for discrete symmetry groups can be matched by a suitable symmetry preserving topological field theory.

Converse results have also been obtained. Indeed, in certain cases [13,21] it has been argued that a given discrete anomaly is incompatible with a symmetry preserving gapped phase in $(2 + 1)d$ by investigating the physics of anyons, and analogous results have been found using higher-form symmetry [22]. Recently [23] derived a general obstruction extracted from any anomaly, which must necessarily vanish if the anomaly admits a symmetry preserving gapped phase. We now apply this to anomalies for $\mathbb{Z}_N$ symmetry in $(3 + 1)d$ theories.

The $\mathbb{Z}_N$ SPT from $(4 + 1)d$ Fermions— Let us describe the $\mathbb{Z}_N$ anomaly of interest. An intuitive way to proceed is to first think of the $\mathbb{Z}_N$ as part of a larger $U(1)$ symmetry. In that case, there are two distinct types of anomalies, a cubic $U(1)$ anomaly and a mixed anomaly between $U(1)$ and Poincaré symmetry (sometimes called a mixed $U(1)$-gravity anomaly). When $U(1)$ is reduced to $\mathbb{Z}_N$ both types of anomalies survive (though
the classification becomes somewhat complicated [24, 25]. Our focus is on the discrete analog of the mixed anomaly. As we will see below, often \( \mathbb{Z}_N \) symmetries with such an anomaly arise from chiral symmetries that act on fermions.

More directly, we can summarize the \( \mathbb{Z}_N \) anomaly by inflow from a \((4 + 1)d\) SPT protected by the same symmetry group. The SPT, which we call \( T_N^k \), can be associated with the long-distance limit of a \( k \) copies of a massive \((4 + 1)d\) fermion \( \Psi \) with real mass term \( m\bar{\Psi}\). The \( \mathbb{Z}_N \) symmetry acts on these fields as \( \Psi \rightarrow e^{2\pi i m} \Psi \). In the limit \( |m| \rightarrow \infty \) the theory becomes trivially gapped, but maintains subtle information about the symmetry which constrains the boundary physics.

We can see the imprint of the \( \mathbb{Z}_N \) symmetry by taking space to be a compact smooth spin 4-manifold \( M_4 \) with minimal non-zero value, 48, of the Pontryagin number \( \int_{M_4} p_1(TM_4) \). On such a manifold, the Atiyah-Singer index theorem [26] implies that the fermions have \( 2k \) complex zero modes \( \lambda_i \), \( i = 1, \cdots 2k \). Therefore, when the mass of \( M_4 \) is small compared to the mass, the system is described by the \( 0 + 1d \) complex fermions \( \lambda_i \) which are acted on by the \( \mathbb{Z}_N \) symmetry.

Quantization proceeds in a standard fashion. Let \( \{ \} \) be the state annihilated by the operators \( \lambda_i \); \( \lambda_i \{ \} = 0 \). By convention we can take \( \{ \} \) to be invariant under the \( \mathbb{Z}_N \) symmetry. Now set \( |+\rangle = \prod_{i=1}^{2k} \lambda_i^0 |\rangle \). On this state the generator, \( U \), of the \( \mathbb{Z}_N \) symmetry acts as

\[
U |+\rangle = e^{-\frac{4\pi i m}{N}} |+\rangle .
\]

(1)

The effect of the \((4 + 1)d\) fermion mass \( m \) is to break the degeneracy between these states and leave a unique vacuum. We set conventions so that the energy of \(|+\rangle\) is positively proportional to \( m \). Then, when \( m \) is positive the system flows to the trivial phase, with a unique \( \mathbb{Z}_N \) invariant ground state. On the other hand, when \( m \) is negative and \( N \nmid 2k \), the ground state \(|+\rangle\) is not invariant under the symmetry and thus defines a nontrivial SPT \( T_N^k \).

From the construction above it is clear that this SPT always admits a simple boundary defined by \( k \) massless \((3 + 1)d\) Weyl fermions. Such a boundary can be constructed by allowing the mass of the fermions \( \Psi \) to modulate along the fifth dimension \( x^5 \); we set \( m > 0 \) where \( x^5 < 0 \), \( m < 0 \) where \( x^5 > 0 \), and \( m = 0 \) at \( x^5 = 0 \). In the region \( x^5 < 0 \) the system flows to the trivial phase, and in the region \( x^5 > 0 \) the system flows to \( T_N^k \). On the interface, \( x^5 = 0 \), there are localized massless \((3 + 1)d\) Weyl fermions.

The possible fermionic \( \mathbb{Z}_N \) SPTs in \((4 + 1)d\) were classified in [24,25]. Their result takes the form of \( \mathbb{Z}_{a_N} \oplus \mathbb{Z}_{b_N} \), where \( a_N, b_N \) are integers depending on \( N \) whose explicit form is given in the references. The SPT \( T_N^k \) corresponds to a generator of this classification. Our analysis above shows that if \( N \nmid 2k \) the phase \( T_N^k \) is trivial on \( M_4 \times S^2 \) when the \( \mathbb{Z}_N \) symmetry is not activated along \( M_4 \). This does not in general mean that \( T_N^k \) is trivial as an SPT. However, it is known [24,25] that if \( N \mid 2k \), the phase \( T_N^k \) admits a gapped symmetry-preserving boundary. As we now argue, if \( N \nmid 2k \) no such boundary exists [27].

No Symmetry-Preserving Boundary Topological Order—We constrain the possible \((3 + 1)d\) boundary theories \( B \) following the general logic of [23]. Consider the SPT \( T_N^k \) on \( M_4 \times \mathbb{R}_+ \), where \( M_4 \) is a compact smooth spin-4-manifold with Pontryagin number 48 as above, and \( \mathbb{R}_+ = \{ x^5 | x^5 \geq 0 \} \) is the half-line. The theory \( B \) lives on the boundary \( x^5 = 0 \).

The path-integral on the region \( 0 \leq x^5 \leq x \) with \( x > 0 \) defines a boundary state \(|\psi\rangle \) in the Hilbert space of \( T_N^k \) on \( M_4 \times \{ x \} \). From our discussion above, we know that this Hilbert space has dimension one and is spanned by \(|+\rangle\). Therefore, \(|\psi\rangle = Z_B[M_4] |+\rangle \) where the coefficient \( Z_B[M_4] \) should be regarded as the partition function of the theory \( B \) on \( M_4 \). If the boundary \( B \) does not explicitly break the \( \mathbb{Z}_N \) symmetry, the boundary state satisfies \( U \langle B | \rangle = \langle B | \rangle \). However, from \([1]\) we know that \( U |+\rangle \neq |+\rangle \). Therefore we conclude

\[
Z_B[M_4] = 0 .
\]

(2)

When \( B \) is the \((3 + 1)d\) massless Weyl fermion, this equation is satisfied because of the fermion zero modes.

Now we assume that the boundary is a unitary spin-topological field theory \( B_{\text{top}} \). Since we are only interested in symmetry-preserving boundaries, we further assume that the Hilbert space of states, \( \mathcal{H}_{S^3} \), of \( B_{\text{top}} \) on a sphere \( S^3 \) is one-dimensional. We argue that these assumptions are incompatible with [24,25].

Define the state \(|0\rangle \in \mathcal{H}_{S^3} \) by the path-integral over the disk \( D^4 \). The \( S^4 \) partition function is the norm: \( Z_{B_{\text{top}}}[S^4] = \langle 0|0 \rangle \) which is positive by unitarity. More generally, given a connected compact 4-manifold \( X_4 \), we define a state \(|X_4⟩ \) in \( \mathcal{H}_{S^3} \) by the path-integral over \( X_4 \setminus D^4 \), where \( D^4 \) is a small ball around a point in \( X_4 \). The inner product \( \langle X_4|Y_4 \rangle \) of two of such states is identified with the partition function \( Z_{B_{\text{top}}}[X_4 \# Y_4] \) on the connected sum \( X_4 \# Y_4 \). Noting also that \(|0,0⟩ \in \mathcal{H}_{S^3} \) is the unit operator on \( \mathcal{H}_{S^3} \), we have

\[
Z_{B_{\text{top}}}[X_4 \# Y_4] = Z_{B_{\text{top}}}[X_4]Z_{B_{\text{top}}}[Y_4]Z_{B_{\text{top}}}[S^4]^{-1} .
\]

(3)

Next we make use of [28], which implies that: given a simply-connected smooth spin 4-manifold \( X_4 \), there exists an integer \( \ell \) such that

\[
X_4 \# (-X_4) \# (S^2 \times S^2) \# \chi \equiv (S^2 \times S^2) \# \chi (X_4 - 2) ,
\]

(4)

where \( -X_4 \) is the orientation reverse of \( X_4 \), \( (S^2 \times S^2) \# \chi \) is the connected sum of \( \ell \) copies of \( S^2 \times S^2 \), and \( \chi (X_4) \) is the Euler number of \( X_4 \). In general, the partition function on \( X_4 \) and \( -X_4 \) are complex conjugates. Hence \([3]\) and \([4]\) imply that for a simply connected spin manifold \( X_4 \), the absolute value of the partition function depends only on its Euler number:

\[
|Z_{B_{\text{top}}}[X_4]|^2 = Z_{B_{\text{top}}}[S^2 \times S^2]^{\chi (X_4) - 2} Z_{B_{\text{top}}}[S^4]^{4 - \chi (X_4)} .
\]

(5)

Finally, following [23] we prove that

\[
Z_{B_{\text{top}}}[S^2 \times S^2] \neq 0 .
\]

(6)

The path-integral over \( S^2 \times S^2 \) defines a state \(|\phi⟩ \) in the Hilbert space on \( S^2 \times S^1 \), and \( Z_{B_{\text{top}}}[S^2 \times S^2] = ⟨\phi|\phi⟩ \). As
TABLE I. Applications of the obstruction to gauge theories with various gauge groups $G$ and matter representations $R$. The final column lists the minimum number of vacua implied by our obstruction if the theory is gapped. In the case with symmetric or antisymmetric tensor representations, the number of enforced vacua is $\frac{N_c}{2}$ or $N_c$ depending on if $N_c$ is 0 modulo 4 or not.

\[
\begin{array}{cccc}
G & R & I(R) & \text{dim}(R) & \text{Obstruction?} & \text{Enforced vacua} \\
SU(N_c) & \text{fund} & 1 & N_c & \text{No} & 1 \\
SO(N_c) & \text{vec} & 2 & N_c & \text{No} & 1 \\
SU(N_c) & \text{adj} & 2N_c & N_c^2 - 1 & \text{Yes} & N_c \\
SU(N_c) & \text{sym} & N_c + 2 & N_c(N_c+1) & \text{Yes} & \frac{N_c}{2} \text{ or } N_c \\
SU(N_c) & \text{asym} & N_c - 2 & N_c(N_c-1) & \text{Yes} & \frac{N_c}{2} \text{ or } N_c \\
SU(2) & 4 & 4 & \text{Yes} & 5 \\
\end{array}
\]

$S^4$ can be obtained by gluing $D^2 \times S^2$ and $S^1 \times D^3$ along their boundary $S^1 \times S^2$, we have $Z_{B_{\text{top}}} [S^3] = (\phi \bar{\phi})$ where the state $\phi$ is defined by path-integral over $S^1 \times D^3$. Since $Z_{B_{\text{top}}} [S^3] > 0$ from unitarity we conclude $|\phi| \neq 0$ and thus $Z_{B_{\text{top}}} [S^2 \times S^2] = (\phi | \phi) \neq 0$.

Therefore, (5) imply that the partition function on a simply-connected $X_4$, of any unitary spin topological field theory with a unique state on $S^3$ obeys

\[
Z_{B_{\text{top}}} [X_4] \neq 0. \tag{7}
\]

In particular, taking $X_4$ to have Pontryagin number 48, e.g., the K3 manifold, this contradicts (2).

To summarize, we have shown that the $(4 + 1)d$ SPT $T_N^k$ associated with $k(4 + 1)d$ massive fermions charged under the $Z_N$ symmetry cannot have a symmetry-preserving gapped boundary when $N \nmid 2k$. If the boundary is gapped, the symmetry must be broken down to $Z_{\text{gcd}(N,2k)}$ at least. On the way, we have also proven that the partition function of a $(3+1)d$ spin-topological field theory on a simply-connected manifold obeys the simple relation (5).

Applications to Gauge Theories—Here we apply our results to the chiral $Z_N$ symmetries in a gauge theory with simple gauge group $G$ and $N_f$ copies of massless Weyl fermions $\psi$ in a representation $R$. If necessary for gauge anomaly cancellation, we also add $N_f$ Weyl fermions in the complex conjugate representation $\bar{R}$.

We consider the chiral symmetry acting on $\psi$ (and not on $\bar{\psi}$) simultaneously. This classical $U(1)$ symmetry is explicitly broken because of the ABJ anomaly. However, depending on $R$, a discrete chiral symmetry group remains:

\[
U(1) \to Z_{N_f I(R)}, \tag{8}
\]

where $I(R)$ is the index of the representation $R$ [29].

This discrete chiral $Z_{N_f I(R)}$ symmetry often has the type of anomaly discussed in the previous sections. (The appropriate $k$ above is simply the total number of fermions.) We thus deduce that such theories cannot be gapped without discrete chiral symmetry breaking when $I(R) \nmid 2 \text{dim}(R)$. When this obstruction occurs in examples is summarized in Table I.

In particular, adjoint QCD has the obstruction. Indeed, when $N_f = 1$ this is $\mathcal{N} = 1$ super-Yang-Mills theory, and it is well known [30] that at long distances the theory is confining and gapped with $N_c$ vacua arising from discrete chiral symmetry breaking due to gaugino condensation.

The case $N_f = 2$ has also been recently studied [31–33] (see also [34, 35] for foundational work). The most plausible IR phase consists of $N_c$ vacua each supporting a $\mathbb{C}P^1$ sigma model. Although this phase is gapless, the chiral $Z_N$, symmetry acts trivially on the $\mathbb{C}P^1$. Therefore, without additional massless degrees of freedom, the anomaly of the discrete chiral symmetry must be matched by spontaneous symmetry breaking and the discrete vacuum degeneracy is inevitable.

In some examples, one can realize spontaneous symmetry breaking of the discrete chiral symmetry semi-classically. We add to the theory to a gauge-neutral complex scalar $\phi$ and include a Yukawa coupling $\phi \bar{\psi} \psi$, where we assume that a gauge-neutral bilinear of $\psi$ exists [30]. To preserve the $Z_{N_f I(R)}$ symmetry, we let $\phi$ transform as $\phi \to e^{\frac{i\pi m}{N_f I(R)}} \phi$.

We now spontaneously break the discrete symmetry via a potential condensing the scalar $\phi$. Classically the phase $\theta$ of $\phi$ is a periodic massless scalar. However, gauge instantons generate a potential for $\theta$ through the Peccci-Quinn mechanism [32]. This potential is $\frac{2\pi}{N_f I(R)}$ periodic, leaving $N_f I(R)$ vacua breaking $Z_{N_f I(R)}$ down to $Z_{2}$.

Weyl Semimetal—The anomaly obstruction derived above also has applications to systems of interest in condensed matter physics. Here we highlight one example.

In materials called Weyl semimetals [38–41], the low energy physics is effectively described by two Weyl fermions. A simple model of a Weyl semimetal [38,39] is realized by the Bloch Hamiltonian acting on two-component functions of $\bar{\psi}$ [38]:

\[
\mathcal{H}_k = 2t_y \sin k_y \sigma_y + 2t_z \sin k_z \sigma_z + (2t_x (\cos k_x - \xi) + m (2 - \cos k_y - \cos k_z)) \sigma_x, \tag{9}
\]

where $t_y, t_z, \xi \text{ and } m$ are parameters and $\sigma_i$ are the Pauli matrices. When $-1 \leq \xi \leq 1$, there are band crossings at $k_x = \pm k_0 = \pm \cos^{-1}(\xi)$ with energy $E = 0$ (see Fig. 1). The dispersion around the crossings is approximately linear and thus, when half-filled, the low energy physics can be described by relativistic left-handed and right-handed Weyl fermions $\psi_1$ and $\psi_2$ corresponding to the crossing points $k_x = \pm k_0$. The
electromagnetic $U(1)_{\text{EM}}$ is a global symmetry of this theory and acts on the Bloch wave function as a uniform phase rotation. This in turn acts on the effective relativistic fermion fields as $\psi_1 \rightarrow e^{i\alpha} \psi_1$ and $\bar{\psi}_2 \rightarrow e^{-i\alpha} \bar{\psi}_2$ with phase $\alpha$.

On the other hand, the discrete translation $T_x$ along the $x$ direction acts on the effective fermions as the axial rotation $\psi_1 \rightarrow e^{i\xi_0} \psi_1$ and $\bar{\psi}_2 \rightarrow e^{-i\xi_0} \bar{\psi}_2$, because at a general point on the Brillouin zone it acts on Bloch wave function as multiplication of $e^{ik_x}$. When the parameters in (9) are tuned to special values, the location $\xi_0$ of the crossing occurs at $\frac{2\pi}{N}$ for some integer $N$. In this case, the $T_x$ translations give rise to a $\mathbb{Z}_N$ axial symmetry.

When the truncation occurs, the effective theory has a $\mathbb{Z}_N$-$U(1)_{\text{EM}}$ mixed anomaly discussed in [42], as well as the $\mathbb{Z}_N$-gravity mixed anomaly extensively discussed above. As remarked in [42], these anomalies exist only in the IR effective theory and not in the UV lattice system, and therefore the gapless modes are not protected under arbitrary continuous deformations of the Hamiltonian. Indeed, the system can easily be gapped out by setting $|\xi| > 1$, see Fig [1](b).

The anomalies, however, can constrain the dynamics when the Lorentz symmetry and truncation of the translation symmetry to $\mathbb{Z}_N$ in the IR are preserved. In particular, our main results says that, under the assumption above, the $\mathbb{Z}_N$-gravity anomaly protects the gapless modes even if general interactions are included [43].

It would be interesting to find applications of our obstruction to lattice theories with crystallographic symmetry. For instance, in a model with point group containing $\mathbb{Z}_N$, if the long-distance physics is relativistic and realizes the anomaly the only way to gap the system while preserving the point group would be to violate the effective Lorentz symmetry.

Acknowledgements— We thank K. Sackel for discussions.