

Noninvertible Chiral Symmetry and Exponential Hierarchies

Clay Córdova¹ and Kantaro Ohmori²

¹*Enrico Fermi Institute and Kadanoff Center for Theoretical Physics,
University of Chicago, Chicago 60637, Illinois, USA*

²*Department of Physics, Faculty of Science, University of Tokyo, Bunkyo, Tokyo 113-0033, Japan*



(Received 3 August 2022; accepted 2 February 2023; published 8 March 2023)

We elucidate the fate of classical symmetries which suffer from Abelian Adler-Bell-Jackiw anomalies. Instead of being completely destroyed, these symmetries survive as noninvertible topological global symmetry defects with world volume anyon degrees of freedom that couple to the bulk through a magnetic 1-form global symmetry as in the fractional Hall effect. These noninvertible chiral symmetries imply selection rules on correlation functions and arise in familiar models of massless quantum electrodynamics and models of axions (as well as their non-Abelian generalizations). When the associated bulk magnetic 1-form symmetry is broken by the propagation of dynamical magnetic monopoles, the selection rules of the noninvertible chiral symmetry defects are violated nonperturbatively. This leads to technically natural exponential hierarchies in axion potentials and fermion masses.

DOI: [10.1103/PhysRevX.13.011034](https://doi.org/10.1103/PhysRevX.13.011034)

Subject Areas: Condensed Matter Physics,
Particles and Fields

I. INTRODUCTION

Symmetry in its myriad incarnations is a unifying principle of quantum field theory. It organizes universality classes and provides one of the few generally applicable tools to constrain correlation functions. Over the past few years, the concept of symmetry has undergone a profound sequence of generalizations broadening its scope and applicability. In this paper, we continue these developments by exhibiting new symmetries in familiar models of massless quantum electrodynamics and models of axions (as well as non-Abelian generalizations). We further explore mechanisms for weakly breaking these novel symmetries and illustrate that this often leads to technically natural exponential hierarchies in effective field theory.

A. Expanding paradigm of global symmetry

The core idea behind recent progress is the intrinsic formulation of internal symmetry in quantum field theory by topological operators [1]. In this point of view, an ordinary global symmetry gives rise to a topological operator of codimension one (equivalently the dimension of space). These symmetry defects act on local operators and impose selection rules on correlation functions. The fact that they are topological, i.e., invariant under

deformations of their support that do not intersect other operators, encodes in an abstract geometric fashion the fact that these symmetries are conserved. Meanwhile, the fact that ordinary symmetries form a group is captured by the symmetry defect fusion algebra: the operator product of defects associated to group elements g_1 and g_2 leads to the defect associated to g_1g_2 .

From this starting point there are two significant generalizations that appear in an interlocked manner as follows.

Higher-form global symmetry [1].—In this generalization, the topological defects defining the symmetry are allowed to have general codimension $(q + 1)$, corresponding to a so-called q -form global symmetry. Here the charged objects are extended operators of dimension q . Most relevant for our discussion is the example $q = 1$ in four-dimensional quantum field theory. In this case, the associated symmetry defects are topological surface operators. These operators frequently arise in gauge theory and provide a symmetry-based view of confinement.

A particularly simple example of such a symmetry which will feature prominently below is the $U(1)^{(1)}$ magnetic 1-form symmetry of Abelian gauge theory. Denoting the dynamical field strength by f , this symmetry is generated by the current $*f$ which is tautologically conserved. The associated charged line defects are 't Hooft lines which physically model infinitely massive magnetic monopoles.

Noninvertible global symmetry.—In this generalization, the topological defects defining the symmetry are permitted to have a more general fusion product, beyond that captured by a group. For instance, the fusion of two such defects

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

may involve a sum of other defects, representing a variety of possible fusion channels. This algebraic structure is referred to as *noninvertible* to emphasize that, in contrast to grouplike symmetries, a typical topological defect need not admit an inverse under fusion.

A variety of examples of this rich structure have recently been constructed. In two spacetime dimensions, the fusion categories of lines are well known from rational conformal field theory [2–4] and have been investigated in Refs. [5–8] and applied to constrain the dynamics of gauge theories [9]. In spacetime dimension larger than two such defects have also recently been constructed by generalizing Kramers-Wannier duality to higher spacetime dimensions [10–15], as well as by novel notions of gauging [16]. Their algebraic properties have also been developed in Refs. [16,17].

The dynamical implications of these generalized symmetries are yet to be fully understood. This presents an opportunity for investigation that sets the context of this work.

B. Noninvertible chiral symmetry defects

Let us now focus on the main class of examples described below. Our first result, derived in Sec. II, is to clarify the nature of chiral symmetry in Abelian gauge theory. In general, a common understanding is that a classical $U(1)_{\text{chiral}}$ symmetry with an Adler-Bell-Jackiw (ABJ) anomaly [18,19] is destroyed quantum mechanically.

Instead, as we argue below, in a certain cases the chiral symmetry is not destroyed, but its nature is changed from an ordinary invertible symmetry classically to a noninvertible symmetry quantum mechanically. One such case is when the gauge symmetry that causes the ABJ anomaly is Abelian, which we call an Abelian ABJ anomaly. More generally, this can also happen when the gauge group has a nontrivial fundamental group.

In more detail, we construct noninvertible topological defects \mathcal{D}_k which act on local operators as a discrete chiral symmetry rotation by a k th root of unity. Semiclassically, these defects may be understood as a composite of the naive invertible chiral symmetry defect \mathcal{C}_k fused with a three-dimensional topological field theory whose spectrum of lines is precisely chosen to compensate for the anomaly of \mathcal{C}_k . In the terminology of Ref. [20], the relevant topological theory is a minimal Abelian topological quantum field theory (TQFT) $\mathcal{A}^{N,p}$, where the index N specifies that the theory has $\mathbb{Z}_N^{(1)}$ 1-form symmetry, and p controls the spins of the associated Abelian anyons [see Eq. (10) below]. In equations, we thus write

$$\mathcal{D}_k \equiv \mathcal{C}_k \times \mathcal{A}^{N,p} \left(\frac{f}{2\pi} \right), \quad (1)$$

where above the appearance of the $U(1)$ field strength f indicates that the bulk is coupled to the TQFT by gauging its 1-form symmetry using the magnetic 1-form symmetry generated by f . This mimics the familiar coupling of the

electromagnetic gauge field in the fractional quantum Hall effect. [See, e.g., Ref. [21] for further discussion of coupling a TQFT to a quantum field theory (QFT).]

The existence of the symmetry defects \mathcal{D}_k may be viewed as a nonperturbative intrinsic definition of the Abelian ABJ anomalous chiral symmetry valid at the operator level in the associated quantum field theory. As we discuss in detail in Sec. II E, these symmetry defects impose selection rules on correlation functions. For local operator correlation functions in flat space, these are the naive selection rules of the chiral symmetry. However, for more general spacetime manifolds, or correlation functions involving extended operators, the selection rules are modified, but nevertheless encoded by the algebraic properties of \mathcal{D}_k .

Focusing on specific examples, we exhibit the defects \mathcal{D}_k in massless quantum electrodynamics and axion electrodynamics. In particular, the former implies the existence of novel noninvertible symmetries in the vanishing fermion mass limit of the standard model. We also exhibit examples of the symmetry defects \mathcal{D}_k in non-Abelian gauge theory, reviewing the construction of Ref. [12], as well as axion Yang-Mills theory.

Like all symmetries, the noninvertible chiral symmetry defects we construct are scale invariant and preserved under renormalization group flow. We thus anticipate that these defects will find broad application in investigating the dynamics of quantum field theories.

C. Symmetry breaking and hierarchies

Exact global symmetries in quantum field theory provide exact selection rules on correlation functions. However, symmetry can also be useful when it is broken by small effects. In that context, the associated degeneracies and selection rules are weakly violated but are still protected from large corrections in effective field theory. This is known as *technical naturalness* [22].

In Sec. III, we explore the paradigm of weakly broken noninvertible symmetry. We focus on the discrete chiral symmetry defects \mathcal{D}_k and exhibit an interplay between the semiclassical physics of magnetic monopoles and technically natural nonperturbative effects.

The key idea is to utilize the fact that in effective field theory, the magnetic 1-form symmetries that featured in the construction of the defects \mathcal{D}_k cannot be violated by any local operator deformation. In the case of the magnetic 1-form symmetry of Abelian gauge theory this is transparent from the fact that the field strength f is closed by the Bianchi identity. This remains true in the presence of any perturbative interactions or charged matter. Similarly, the discrete magnetic 1-form symmetry of non-Abelian gauge theory cannot be broken by operator deformations or the addition of charged matter.

From a semiclassical point of view, this means that the breaking of these 1-form symmetries can only proceed

through nonperturbative physics or relatedly a change in the topology of field space. In the case at hand, this is achieved by considering models where there are dynamical magnetic monopoles, which can screen the 't Hooft line and so violate the 1-form symmetry nonperturbatively.

These corrections due to monopoles may be quantified by summing over loops of virtual particles. If a monopole arises from a Higgsing process at a scale v , then the monopole mass m_{mon} and cutoff Λ are related as

$$m_{\text{mon}} \sim \frac{v}{e}, \quad \Lambda \sim ev, \quad (2)$$

where e is a coupling constant. A monopole existing for a short proper time on the order of Λ^{-1} then provides a correction to the Lagrangian controlled by its one-particle action S_{mon} :

$$\delta\mathcal{L} \sim \exp(-S_{\text{mon}}) \sim \exp\left(-\frac{m_{\text{mon}}}{\Lambda}\right) \sim \exp\left(-\frac{\#}{e^2}\right). \quad (3)$$

This correction is nonperturbative and small provided the effective field theory is under technical control, with a small coupling constant e .

As is suggested by Eq. (3), there is a close connection between the physics of monopoles described above and the properties of instantons. As we emphasize in Sec. II, noninvertible chiral symmetries arise semiclassically when there are instantons which cannot be realized in \mathbb{R}^4 but can be realized in more complicated spacetime topologies or in the presence of extended operators. From this point of view, the change in field space topology eluded to above is precisely what is needed to allow the instanton to be realized in \mathbb{R}^4 .

Once the 1-form symmetry is violated, the discrete chiral symmetry defects \mathcal{D}_k are no longer topological and hence their selection rules are also broken. This can be seen, for instance, from the fusion rules discussed below and derived in Refs. [11,12,14]. However, since the violation of the 1-form symmetry is nonperturbative, so too is the violation of the \mathcal{D}_k selection rules. Thus, weakly broken 1-form symmetry yields exponentially small violations in the selection rules of the noninvertible symmetry defects \mathcal{D}_k . Moreover, these small corrections are technically natural since they are the leading terms violating \mathcal{D}_k .

In this way we construct a variety of technically natural models with exponential hierarchies in their effective field theory description. This includes axion models (Abelian or non-Abelian) with exponentially suppressed potentials or energy splittings. In this context we make contact with the recent work of Ref. [23], which examined the interplay between axion potentials and magnetic monopoles. We also discuss gauge theories with matter (Abelian or non-Abelian) with exponentially suppressed fermion masses.

We speculate that this new symmetry-based mechanism for generating exponential hierarchies may find applications

in model building or in refining our understanding of naturalness in effective field theory. For recent work utilizing this mechanism in a model of neutrino masses, see Ref. [24].

II. NONINVERTIBLE CHIRAL SYMMETRY FROM ABELIAN ABJ ANOMALIES

In this section, we construct noninvertible symmetry defects in models with Abelian ABJ-like anomalies. Specifically, we find a noninvertible defect for each chiral symmetry rotation by a rational angle. As we discuss below, these noninvertible symmetry defects also make precise the fate of the chiral symmetry selection rules on correlation functions.

Throughout these examples, a key role is played by the bulk 1-form symmetry whose current couples with the internal 1-form symmetry of the topological degrees of freedom on the defect. In the case of Abelian gauge theory this is the continuous magnetic 1-form symmetry $U(1)^{(1)}$ with closed 2-form current $f/2\pi$, while in the non-Abelian gauge theory examples below it is a suitable discrete analog of this magnetic symmetry.

A. Massless electrodynamics

Let us begin with Abelian gauge theory with massless charged fermionic matter (i.e., massless QED) and gauge group $U(1)$. For simplicity we assume that we have N_f species of massless electrons of unit electric charge, though our discussion readily generalizes to models with more complicated matter content; the essential feature is only the ABJ anomaly.

At the classical level this theory admits a chiral symmetry $U(1)_{\text{chiral}}$. Denoting the positively (negatively) charged Weyl fermions as $\chi_{a^\pm}^a$ with a a flavor index, the chiral symmetry acts as

$$\chi_{a,+}^a \rightarrow e^{i\theta} \chi_{a,+}^a, \quad \chi_{a,-}^a \rightarrow \chi_{a,-}^a. \quad (4)$$

When quantum effects are considered, it is well known that this symmetry is violated by the Adler-Bell-Jeckiv anomaly [18,19]. At a technical level, we may view this as arising from an anomaly polynomial related to the anomalous variation in the action by the descent procedure (the five-dimensional inflow action \mathcal{A}_5 obeys $d\mathcal{A}_5 = 2\pi\mathcal{I}_6$):

$$\mathcal{I}_6 = \frac{N_f}{2(2\pi)^3} f \wedge f \wedge F, \quad (5)$$

where f is the field strength of the dynamical $U(1)$ gauge field and F is the background field strength of the putative chiral symmetry. This anomaly polynomial breaks the classical $U(1)_{\text{chiral}}$ symmetry down to a \mathbb{Z}_{N_f} discrete chiral symmetry which is free of ABJ anomalies. In modern language, this \mathbb{Z}_{N_f} symmetry is generated by

codimension-one topological operators with standard grouplike (invertible) fusion rules.

What more can be said about this well-known story? A clue can be seen via a familiar trick in current algebra. Denoting by J_μ the anomalous current and $*J$ the Hodge dual, the anomaly (5) implies:

$$d * J = \frac{N_f}{8\pi^2} f \wedge f. \quad (6)$$

The right-hand side is a globally well-defined closed 4-form. However, if the spacetime manifold is sufficiently simple, in particular if it does not have any closed two cycles, and there are no line operators inserted, then the form is also exact and is the exterior derivative of the Chern-Simons 3-form $\propto a \wedge da$. Thus, in such a simple configuration, one is tempted to redefine the current as

$$*J \xrightarrow{?} *J - \left(\frac{N_f}{8\pi^2} \right) a \wedge da. \quad (7)$$

This modified current is not gauge invariant, but in these simple field configurations certain rational multiples (with denominator N_f) of the total exponentiated charge integrals are well defined and conserved. Thus, one is tempted to declare that the chiral symmetry remains, despite the anomaly Eq. (6). For instance, this manipulation implies that the selection rules of the broken chiral symmetry are valid when considering correlation functions of local operators in a topologically trivial spacetime.

With a view toward later generalizations, we can also rephrase this argument in terms of instantons. Equation (6) implies that the chiral charge is violated by Abelian instantons, but on a topologically trivial spacetime such as \mathbb{R}^4 (or more precisely its infrared (IR) regulated version S^4), with only local operator insertions, no such instantons exist.

The preceding discussion begs the question as to what has become of the chiral symmetry in massless QED (or other similar models with an Abelian ABJ anomaly). As we now show, the correct statement which holds on any spacetime manifold in any configuration of operators is that there is a noninvertible chiral symmetry in the theory. Thus, the true effect of the ABJ anomaly is not to destroy the chiral symmetry, but to change its nature.

It is helpful to consider again the anomaly polynomial Eq. (5). Let us focus on a discrete chiral rotation by a root of unity so that the angle in Eq. (4) is $\theta = 2\pi/k$ for some integer k . This is implemented by a codimension-one domain wall operator \mathcal{C}_k located, say, at $x = 0$, where the chiral background gauge field A (with field strength F) is taken to be

$$A = \theta \delta(x) dx. \quad (8)$$

Applying inflow, we deduce that across this domain wall, the bulk action differs by an effective θ angle:

$$S \rightarrow S + \frac{2\pi i N_f}{k} \int_{x>0} \frac{f \wedge f}{8\pi^2}. \quad (9)$$

This equation implies that in a theory where the dynamical gauge field is a frozen background, the \mathbb{Z}_k chiral domain wall \mathcal{C}_k has a world volume 't Hooft anomaly characterized by the above term by inflow. Once the U(1) is made dynamical, we then see from Eq. (9) that the bulk action jumps across \mathcal{C}_k and hence this defect does not define a symmetry of our theory.

The key idea, following Ref. [12], is now to modify the \mathbb{Z}_k chiral domain wall \mathcal{C}_k by stacking it with a suitable 3D TQFT which also couples to the bulk gauge field a and cancels the apparent anomaly in Eq. (9). After this construction an avatar of the discrete chiral domain wall will remain as a topological defect (symmetry) in the theory, but at the cost of becoming noninvertible.

In more detail, we consider a TQFT with a 1-form symmetry $\mathbb{Z}_N^{(1)}$ with associated background field $B^{(2)}$. This means that among the lines characterizing this theory are Abelian anyons whose fusion algebra forms the group \mathbb{Z}_N . These lines have a spin, with s times the generating line having spin $h(s)$ given by

$$h(s) = \frac{ps^2}{2N} \pmod{1}, \quad (10)$$

where p is coprime to N . Such a topological field theory carries an anomaly defined by inflow as

$$S_{\text{inflow}} = -\frac{2\pi i p}{2N} \int \mathcal{P}(B^{(2)}), \quad (11)$$

where \mathcal{P} is the Pontryagin square operation [25].

As discussed in detail in Ref. [20], any such TQFT can be written as product of a minimal Abelian TQFT $\mathcal{A}^{N,p}$ together with a decoupled sector that plays no role in the following. We denote by $\mathcal{A}^{N,p}(B^{(2)})$ this minimal TQFT coupled to its 1-form symmetry background $B^{(2)}$.

We now fix N and p such that $p/N = N_f/k$ and $\gcd(p, N) = 1$. We define a modified chiral symmetry defect \mathcal{D}_k , by stacking \mathcal{C}_k with $\mathcal{A}^{N,p}$ and coupling to the bulk by gauging the 1-form symmetry of the TQFT through the dynamical gauge field $f = da$. In equations this means

$$\mathcal{D}_k \equiv \mathcal{C}_k \times \mathcal{A}^{N,p} \left(\frac{f}{2\pi} \right). \quad (12)$$

Thus, the bulk and TQFT are coupled through the fact that the 1-form symmetry on the defect is identified with the magnetic 1-form symmetry of the U(1) gauge theory.

The coupling between the bulk and the defect defined by Eq. (12) takes a more familiar form in the special case $p = 1$ where $\mathcal{A}^{N,p}$ is an Abelian Chern-Simons theory $U(1)_N$ [20,26]. Denoting the defect dynamical gauge field

by c , the bulk-defect coupling arises from a mixed Chern-Simons term:

$$S_{\text{defect}} = \frac{iN}{4\pi} \int c \wedge dc + \frac{i}{2\pi} \int c \wedge f, \quad (13)$$

which mimics the familiar coupling of the electromagnetic gauge field in the fractional quantum Hall effect. This point of view also clarifies the meaning of the gauge noninvariant current appearing in Eq. (7): the fractional Chern-Simons term appearing in the current may be understood as the effective response of the fractional Hall state.

The defect \mathcal{D}_k defines the remnant of the \mathbb{Z}_k chiral symmetry in our model. Repeating the argument for general k or powers thereof, we see that a chiral rotation by any rational angle can be promoted to a general noninvertible topological defect. In particular, all such operators should be viewed as a generalized symmetries and imply selection rules on correlation functions discussed below.

The noninvertible nature of these defects is manifest simply from the fact that the partition function of $\mathcal{A}^{N,p}$ in general does not have unit modulus (and may even vanish). For instance, wrapping \mathcal{D}_k on a three-sphere and using $|Z_{\mathcal{A}^{N,p}}(S^3)| = 1/\sqrt{N}$, we obtain the quantum dimension:

$$\mathcal{D}_k(S^3) = \frac{1}{\sqrt{N}}. \quad (14)$$

Alternatively, one can also exhibit the noninvertibility of these defects through their fusion algebra. In particular, following the derivation in Refs. [11,12,14], one can deduce that the fusion of the defect \mathcal{D}_k with its orientation reversal $\bar{\mathcal{D}}_k$ on a general oriented three-manifold M is

$$\mathcal{D}_k(M) \times \bar{\mathcal{D}}_k(M) = \sum_{S \in H^2(M, \mathbb{Z}_N)} \eta(S) e^{(2\pi i p/N)Q(S)}, \quad (15)$$

where $\eta = \exp[i \int_S (f/2\pi N)]$ is the generator of the \mathbb{Z}_N subgroup of the magnetic 1-form symmetry, the sum is over two cycles $S \subset M$, and $Q(S)$ is the triple self-intersection number. In particular, the right-hand side of Eq. (15) is an example of a condensation defect explored in detail in Refs. [14,27] (see also Refs. [28,29]).

In summary, the existence of the noninvertible topological defects \mathcal{D}_k is an intrinsic statement about the operator content of the quantum field theory. Thus, these defects give a nonperturbative definition of the ABJ anomaly and its physical consequences.

B. Axion electrodynamics

The noninvertible defects \mathcal{D}_k defined above have analogs in any theory with an ABJ-like anomaly. Let us exhibit them in a model of axion electrodynamics. We focus on a minimal model, though our analysis applies to more general field content.

Thus, consider a theory where, in addition to the dynamical U(1) gauge field with field strength f , there is a dynamical periodic scalar field θ (the axion) with

$$\theta \sim \theta + 2\pi\mu, \quad (16)$$

with μ the axion decay constant. The action is

$$S = \frac{1}{2} \int d\theta \wedge *d\theta + \frac{1}{2e^2} \int f \wedge *f + \frac{iL}{8\pi^2\mu} \int \theta f \wedge f, \quad (17)$$

with $L \in \mathbb{Z}$ a possible discrete coupling constant. Ignoring the coupling between the θ and the gauge field, the axion enjoys a continuous shift symmetry with current J where $*J = i\mu *d\theta$. However this current is broken by the coupling between θ and f [the normalization of the current J can be verified by noting that the associated charged operators are $\exp(in\theta/\mu)$ for integer n]:

$$d * J = \frac{L}{8\pi^2} f \wedge f. \quad (18)$$

This equation is directly analogous to the violation of the chiral symmetry in Abelian gauge theory (6) and breaks the axion shift symmetry to \mathbb{Z}_L . However, as in our previous discussion, we note that the violation of the axion shift symmetry is only through Abelian instantons, and hence does not occur in sufficiently simple spacetime topologies with only local operator insertions. We thus anticipate that the axion shift symmetry is not broken, but instead is transformed into a noninvertible topological operator.

We may directly verify this intuition by constructing a suitable symmetry defect \mathcal{D}_k associated to a discrete \mathbb{Z}_k rotation of the axion $\theta \rightarrow \theta + (2\pi\mu/k)$. Letting \mathcal{C}_k denote the naive (broken) discrete \mathbb{Z}_k shift symmetry defect, we see that across \mathcal{C}_k the action shifts as in Eq. (9), effectively generating a rational θ angle $2\pi L/k$ for the dynamical U(1) gauge field. Again choosing N and p coprime with $p/N = L/k$, we define \mathcal{D}_k exactly as in Eq. (12):

$$\mathcal{D}_k \equiv \mathcal{C}_k \times \mathcal{A}^{N,p} \left(\frac{f}{2\pi} \right). \quad (19)$$

Of course the close analogy between the axion shift symmetry and the chiral symmetry in Abelian gauge theory is no accident, as the two can be related by renormalization group flow. Indeed, in a typical model of axions constructed by the Pecci-Quinn mechanism, the UV consists, for instance, of electrodynamics coupled to a neutral complex scalar φ and N_f electrically charged fermions χ_{\pm}^a [30,31]. In addition to the kinetic terms the action contains Yukawa and potential interactions:

$$S \supset \int d^4x \{ \lambda \bar{\varphi} \chi_+^a \chi_-^a + \lambda \varphi \bar{\chi}_+^a \bar{\chi}_-^a - V(\varphi) \}. \quad (20)$$

At the classical level, there is a chiral Peccei-Quinn $U(1)_{PQ}$ symmetry where the different fields have charges

Field	$U(1)_{PQ}$ charge
χ_+^a	+1
χ_-^a	0
φ	+1

At the quantum level, this $U(1)_{PQ}$ has an ABJ anomaly and, hence, following the discussion above, gives rise to a noninvertible discrete chiral symmetry defect \mathcal{D}_k .

We now flow to axion electrodynamics by condensing the scalar with $\langle |\varphi|^2 \rangle = \mu^2$. This freezes the radial mode of φ . However, an axion θ remains as a pseudo-Goldstone mode where

$$\varphi = \mu \exp(i\theta/\mu). \quad (21)$$

The Yukawa couplings give a mass to the fermions, and decoupling them from the axion requires a chiral $U(1)_{PQ}$ rotation. The anomaly then generates an axion coupling between θ and the $U(1)$ gauge fields. Thus, the final action is that of axion electrodynamics where the coupling L in Eq. (17) is identified with N_f the number of UV flavors.

This simple flow illustrates the renormalization group invariance of the existence of the noninvertible symmetry defects. Indeed, the defect \mathcal{D}_k constructed from the discrete chiral symmetry in the UV flows to the noninvertible axion shift symmetry defect in the IR.

In general, the defect \mathcal{D}_k defined in Eq. (19) makes rigorous the sense in which the axion shift symmetry is still present in axion electrodynamics. As discussed in more detail below, this means that at the level of local operator correlation functions on \mathbb{R}^4 , the selection rules implied by the discrete axion shift symmetry hold. One important application of this is to the axion potential. Taking into account only the standard \mathbb{Z}_L invertible symmetry, the axion model admits a possible potential of the form

$$V(\theta) = \sum_{n \in \mathbb{N}} \alpha_n \cos\left(\frac{Ln\theta}{\mu}\right). \quad (22)$$

However, this potential is forbidden by the noninvertible defect \mathcal{D}_k which shifts θ by a fraction of its period. In other words, a vanishing potential in axion electrodynamics is stabilized by the exact noninvertible symmetry generated by the defects \mathcal{D}_k .

C. Non-Abelian gauge theory with matter

A discrete part of the noninvertible symmetry in the previous sections can also arise in a non-Abelian gauge theory. This symmetry was discussed in Ref. [12], and here

we recast the argument so that it aligns with the ABJ anomalous symmetry presented above. We focus on the simplest example of relevance to later models, though as in previous sections the essential feature is the anomaly.

Consider, e.g., a QFT with non-Abelian gauge group $PSU(N_c) \cong SU(N_c)/\mathbb{Z}_{N_c}$, with non-Abelian field strength w . We take the matter content to be N_f Weyl fermions in the adjoint representation λ^a , where a is a flavor index. There is a classical $U(1)_{\text{chiral}}$ symmetry acting on λ^a as $\lambda^a \rightarrow e^{i\theta} \lambda^a$. In this case the anomaly polynomial is

$$\mathcal{I}_6 = \frac{2N_f N_c}{2(2\pi)^3} \text{Tr}(w \wedge w) \wedge F, \quad (23)$$

where Tr is the trace in the fundamental representation and F is again the background $U(1)_{\text{chiral}}$ field strength.

The resulting pattern of discrete chiral symmetry depends on the global form of the gauge group through the properties of the instanton number. Specifically (restricting to spin manifolds for simplicity), we have

$$\mathcal{I} = \frac{1}{8\pi^2} \int \text{Tr}(w \wedge w) \in \begin{cases} \mathbb{Z} & SU(N_c) \\ \frac{1}{N_c} \mathbb{Z} & PSU(N_c). \end{cases} \quad (24)$$

This means the invertible symmetry group is broken down to $\mathbb{Z}_{2N_c N_f}$ for the $SU(N_c)$ theory, and to \mathbb{Z}_{2N_f} for the $PSU(N_c)$ theory.

The $PSU(N_c)$ configuration with fractional \mathcal{I} is called a fractional instanton. Such a configuration is possible because a $PSU(N_c)$ gauge connection a has more freedom than a $SU(N_c)$ gauge connection. Specifically, this is encoded in the discrete magnetic flux, defined cohomologically by the second Stiefel-Whitney class of the gauge bundle $w_2(a) \in H^2(M, \mathbb{Z}_{N_c})$. The fractional part of the instanton number \mathcal{I} is controlled by $w_2(a)$ as [this is an integer multiple of $1/N_c$ because, assuming spacetime is a spin manifold, $\mathcal{P}(w_2(a))$ is even when N_c is even]:

$$\mathcal{I} = \frac{N_c - 1}{2N_c} \mathcal{P}(w_2(a)) \pmod{1}. \quad (25)$$

This fractional instanton breaks the anomaly-free symmetry $\mathbb{Z}_{2N_c N_f}$ of the $SU(N_c)$ theory further down to \mathbb{Z}_{2N_f} .

However, exactly analogous to our discussion of the ABJ anomaly for the chiral symmetry of massless QED, fractional instanton configurations do not exist when the spacetime manifold has a trivial second cohomology group, $H^2(M, \mathbb{Z}_{N_c})$. The main difference is that, in the Abelian case, no instantons exist at all on \mathbb{R}^4 , while for $PSU(N_c)$ only the fractional instantons are absent on \mathbb{R}^4 . Thus, we expect that the generator of $\mathbb{Z}_{2N_c N_f}$ survives as a topological defect, but becomes a noninvertible symmetry.

Proceeding as in the previous sections, we can construct the desired symmetry defect \mathcal{D}_k with $k = 2N_c N_f$ again by

stacking the \mathbb{Z}_k domain wall \mathcal{C}_k with a minimal Abelian TQFT as

$$\mathcal{D}_k = \mathcal{C}_k \times \mathcal{A}^{N_c, 1}(w_2(a)), \quad (26)$$

where the TQFT couples to the Stiefel-Whitney class $w_2(a)$ instead of the field strength. This is the noninvertible defect found in Ref. [12].

D. Axion Yang-Mills theory

Just as the noninvertible chiral symmetry defects in massless QED immediately generalize to noninvertible symmetries of axion electrodynamics, so too do the noninvertible symmetries described in the previous section generalize to noninvertible symmetries of axion Yang-Mills theory. Here we focus on examples with gauge group $\text{PSU}(N_c)$ to parallel the above discussion, though it is straightforward to generalize to other gauge groups.

Thus we consider a QFT, where in addition to the kinetic terms involving the axion θ and the gauge field strength w , the action includes the interaction

$$S \supset \frac{iL}{8\pi^2\mu} \int \theta \text{Tr}(w \wedge w). \quad (27)$$

Here we must carefully keep track of the quantization of the instanton density expressed in Eq. (24). In particular, fractional instantons of $\text{PSU}(N_c)$ imply that the correct periodicity of the axion is

$$\theta \sim \theta + 2N_c\pi\mu. \quad (28)$$

As discussed above, the possibility of fractional instantons also leads to discrete noninvertible symmetry defects. Indeed, the model (27) has a \mathbb{Z}_L standard invertible symmetry which acts by shifting θ by $1/L$ times its full periodicity. However, there is also a noninvertible topological defect \mathcal{D}_k with $k = N_c L$ constructed as in the previous section by stacking a discrete axion shift symmetry \mathcal{C}_k domain wall, with the minimal TQFT $\mathcal{A}^{N_c, 1}(w_2(a))$ as in Eq. (26).

The noninvertible defect \mathcal{D}_k contains the invertible defects by fusion in the sense that $(\mathcal{D}_k)^{N_c}$ is the generator of the invertible \mathbb{Z}_L symmetry. As in the axion-electrodynamics case, \mathcal{D}_k constrains the axion potential by enforcing that it is invariant under shifts of by $1/(N_c L)$ of the full period: $\theta \rightarrow \theta + (2\pi\mu/L)$. In particular, the potential has $N_c L$ minima leading to $N_c L$ degenerate vacua.

E. Selection rules on correlation functions

An ordinary invertible symmetry imposes selection rules; it ensures that some correlation functions vanish, or more generally, relates one correlation function to another. It is thus natural to ask what constraints on

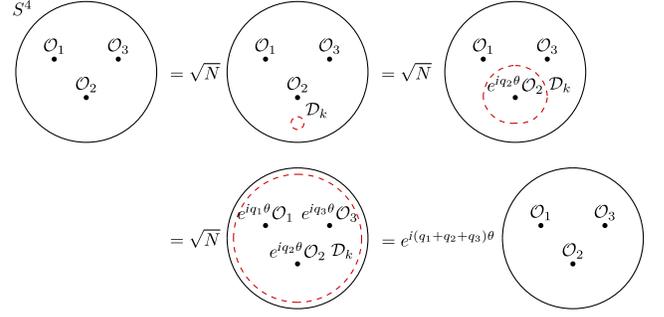


FIG. 1. The action of the noninvertible symmetry \mathcal{D}_k on the correlation function on S^4 with local operator insertions with charge q_1 , q_2 , and q_3 . The rotation angle θ of the symmetry is $2\pi/k$. At the last step the defect is contracted at the other side of the sphere.

correlation functions are implied by the noninvertible topological defects \mathcal{D}_k defined above.

When we consider correlation functions of local operators on S^4 (regarded as an IR regulated \mathbb{R}^4), the selection rule is simple: we just ignore the fact that the defect is noninvertible and treat it as if it were an ordinary (i.e., invertible) chiral symmetry (see also Ref. [32] for related discussion). To formally derive this result, we insert a small S^3 -shaped symmetry defect $\mathcal{D}_k(S^3)$ around the north pole of the ambient S^4 , then let the defect pass through the whole spacetime, contracting at the south pole (see Fig. 1). In this procedure the local operators acted on by the \mathcal{C}_k part of \mathcal{D}_k , but do not interact with the TQFT $\mathcal{A}^{N_c, p}$. [When we insert the defect $\mathcal{D}_k(S^3)$ we have to divide the correlator by a factor of the quantum dimension (14); however, this factor cancels when we contract the defect at the antipodal point.]

The selection rules become more interesting when either the manifold is more complicated or we involve line operators in the correlator. Let us focus on the latter. Consider for example a 't Hooft line operator T of magnetic charge m inserted along a curve γ in spacetime. In an Abelian gauge theory with symmetry \mathcal{D}_k defined in Eq. (12), this operator may be viewed as the worldline of an infinitely heavy Dirac monopole.

How does the noninvertible symmetry act on T ? In this case the TQFT degrees of freedom of the defect \mathcal{D}_k are essential. Examining Eq. (9), we see that the domain wall effectively changes the θ term as Eq. (9). This implies that the line T acquires a fractional electric charge through the Witten effect. Hence the action on T is

$$T(\gamma) \rightarrow T(\gamma) W^{N_f/k}(\gamma, \Sigma), \quad (29)$$

where $W^{N_f/k}(\gamma, \Sigma) = e^{i(N_f/k) \int_{\Sigma} f}$, with $\partial\Sigma = \gamma$, is an open 1-form symmetry generator. When N_f/k is an integer, the operator $W^{N_f/k}(\gamma, \Sigma)$ is a genuine Wilson line operator with the integer charge, but in general, for fractional N_f/k it depends on the surface Σ .

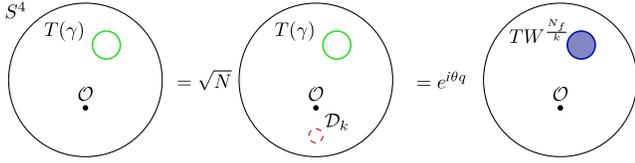


FIG. 2. A noninvertible symmetry transforms a line operator T on a line γ into a line operator attached to a surface operator.

This behavior of the noninvertible symmetry, transforming a genuine line operator to a line attached to a surface, was also found in Refs. [11,12,14]. It is a direct analog of Kramers-Wannier duality, which exchanges the local spin operator and the disorder operator in the two-dimensional Ising model.

Although we have derived the properties of the defect action \mathcal{D}_k in Abelian QED, the result is model independent and intrinsically encoded in the interplay between the bulk 1-form symmetry and the TQFT. Thus, in general, the noninvertible defect \mathcal{D}_k acts on a line charged under the 1-form symmetry and transforms it into a nongenuine line attached to an open 1-form symmetry surface operator. An example of this action on a correlator involving T and local operators on S^4 is shown in Fig. 2. Note that in the presence of T insertions, nonzero correlation functions where the ABJ anomalous charge is classically unbalanced are compatible with the \mathcal{D}_k selection rules.

The noninvertibility of \mathcal{D}_k can also be seen in its action on the Hilbert space of states on a nontrivial spatial manifold. To be concrete, take the spatial manifold to be $M_3 = S^2 \times S^1$. The Hilbert space decomposes into sectors labeled by the 1-form charge: $\mathcal{H}_{S^2 \times S^1} = \bigoplus_m \mathcal{H}_m$, where $m = \int_{S^2} (f/2\pi)$. To study the action of \mathcal{D}_k on this Hilbert space, we need the partition function of the TQFT $\mathcal{A}^{k,1}$ on this spatial manifold. As the TQFT is coupled with f through its 1-form symmetry, the magnetic flux m induces an anyon line along the S^1 direction of M_3 , which is the m th power of the generating line. However, the $S^2 \times S^1$ partition function of $\mathcal{A}^{k,1}$ with an anyon insertion along S^1 vanishes unless the line is trivial. Therefore, we deduce

$$\mathcal{A}^{k,1}[S^2 \times S^1, m] = \begin{cases} 0 & m \neq 0 \pmod{k} \\ 1 & m = 0 \pmod{k}. \end{cases} \quad (30)$$

From the definition (12), we have

$$\mathcal{D}_k[S^2 \times S^1] = C_k P_{m,k}, \quad (31)$$

where $P_{m,k}$ is the projection onto the sectors \mathcal{H}_m with $m = 0 \pmod{k}$. The noninvertibility of \mathcal{D}_k is now manifest from the fact that its action contains a projection operator. Note also that preservation of this operator means that, on \mathcal{H}_m , the chiral central charge is preserved modulo m if $m \neq 0$, and it is exactly preserved on \mathcal{H}_0 .

III. EXPONENTIAL HIERARCHIES FROM MAGNETIC MONOPOLES

The previous section constructed noninvertible analogs of chiral symmetry in gauge theory. Here we consider models where these symmetries are emergent at long distances and show that this scenario naturally leads to exponential hierarchies in the violation of the chiral symmetry selection rules.

As emphasized in Sec. I, the unifying feature of these examples is that the breaking of a magnetic 1-form symmetry proceeds semiclassically through the existence of magnetic monopoles which may screen the 't Hooft lines. This breaking is communicated to the noninvertible chiral symmetry defects \mathcal{D}_k and leads to nonperturbative, but technically natural, violations of the selection rules.

A. Exponentially suppressed axion potentials

As a first pedagogical example, we consider a model of axion electrodynamics where the U(1) gauge field arises from an ultraviolet non-Abelian gauge group by Higgsing. As is familiar, in this model one finds an exponentially small axion potential generated by instantons. Following the analysis of Ref. [23], we recast this nonperturbative potential as the consequence of quantum loops of dynamical magnetic monopoles. These monopoles break the magnetic 1-form symmetry of the infrared theory. In this way we recast the exponentially suppressed axion potential as a technically natural violation of the noninvertible axion shift symmetry defects \mathcal{D}_k .

For concreteness, we focus on the simplest case of SU(2) though the discussion readily generalizes. At high energies, our effective description consists of a non-Abelian gauge field with field strength w , the axion θ , and a real SU(2) triplet Higgs field Φ . In addition to the kinetic terms for θ , w , and Φ , the action contains the interaction terms for the axion and gauge field, as well as the Higgsing potential,

$$S \supset \frac{iL_{UV}}{8\pi^2\mu} \int \theta \text{Tr}(w \wedge w) - \frac{1}{g^2} \int d^4x V(\Phi). \quad (32)$$

We assume the Higgsing potential is chosen such that the triplet condenses, $\langle |\Phi|^2 \rangle = v^2$ leading at long distances to axion electrodynamics and a neutral decoupled scalar with $L_{IR} = 2L_{UV}$. Henceforth for simplicity, we set $L_{UV} \rightarrow 1$.

Let us track the symmetries along this renormalization group flow. In the IR we have the noninvertible axion shift symmetry defects \mathcal{D}_k . However, in this model these symmetries are emergent and explicitly broken by the couplings of the ultraviolet action Eq. (32). A simple way to understand this is again to look at the divergence of the semiclassical current J that acts to shift the axion. In the non-Abelian model (32) we have

$$d * J = \frac{1}{8\pi^2} \text{Tr}(F \wedge F). \quad (33)$$

Unlike the case of axion electrodynamics, the right-hand side of this equation is not in general d exact, even in \mathbb{R}^4 , and hence no modified charge is available. Said differently, there are nontrivial instantons in S^4 which explicitly violate the axion shift symmetry.

One can also ask more directly why the construction of the defect \mathcal{D}_k fails in the non-Abelian gauge theory. In this case, as we cross the shift symmetry wall \mathcal{C}_k , the bulk differs by a rational non-Abelian θ angle $2\pi/k$. However, now we can no longer absorb this anomaly by dressing the defect by a suitable topological field theory. Indeed, our coupling to $\mathcal{A}^{N,p}$ is through the 1-form symmetry of the TQFT and the magnetic 1-form symmetry of the bulk gauge theory. But in the non-Abelian theory this magnetic 1-form symmetry is explicitly broken.

The preceding comments allow us to investigate the breaking of the defect \mathcal{D}_k from the point of view of the IR Abelian gauge theory. In this theory, when the 't Hooft lines are viewed as rigid defect operators, the magnetic 1-form symmetry is preserved. However, when there are dynamical magnetic monopoles, the lines can be screened and the 1-form symmetry is broken. This implies that we can understand the leading symmetry breaking effects for \mathcal{D}_k from the semiclassical physics of magnetic monopoles. In other words, the dynamical monopole or dyon worldlines generated in the vacuum appear as insertions of Wilson-'t Hooft line operators, where the selection rules discussed in Sec. II E do not ensure the vanishing of correlation functions.

Let us focus on the calculation of the effective axion potential which arises from summing over loops of monopoles in the Abelian gauge theory. Our treatment follows that of Ref. [23]. Assuming the interactions between multiple monopoles are small, the effective potential $V(\theta)$ can be usefully computed in the worldline formalism where we sum over closed trajectories of particles with a unit magnetic charge. This can be viewed as a dilute monopole gas approximation. Working at a constant value θ of the axion we then have

$$\int d^4x V(\theta) = Z(\theta), \quad (34)$$

where $Z(\theta)$ is the single particle partition function. In turn, this partition function may be expressed as an integral over worldline proper time τ . This results in a standard expression:

$$V(\theta) \sim \int_0^\infty d\tau \tau^\alpha \exp\left(-\frac{m^2(\theta)}{2}\tau\right), \quad (35)$$

where $m(\theta)$ is the mass of the magnetically charged particle and we have neglected order one coefficients and a power law term in the integrand. Thus, we see that an effective potential will arise for the axion θ when the particles summed in the worldline trajectory have masses that depend on θ . This may be viewed as an analog of the Coleman-Weinberg potential [33] generalized to include loops of solitons. (See also Refs. [34–36] for related discussion.)

To evaluate Eq. (35) we use the fact that in addition to a fundamental monopole, there is also a tower of dyons whose electric charges q depend on the axion through the Witten effect [37]:

$$q = \left(\ell - \frac{\theta}{2\pi\mu}\right), \quad (36)$$

where $\ell \in \mathbb{Z}$ labels the dyon. The energy spectrum of these dyons and their interactions can usefully be understood by viewing them as the result of quantizing a worldline periodic scalar where the axion θ enters as the worldline θ angle [38]. This results in a familiar energy spectrum for these modes:

$$E_\ell \sim \left(\ell - \frac{\theta}{2\pi\mu}\right)^2, \quad (37)$$

Equations (36) and (37) suggest an ansatz where the θ dependence of the mass is taken to be

$$m_\ell^2(\theta) \approx m_{\text{mon}}^2 + m_{\text{vec}}^2 \left(\ell - \frac{\theta}{2\pi\mu}\right)^2, \quad (38)$$

with the mass of the monopole and W boson respectively given by

$$m_{\text{mon}} = 4\pi v/g, \quad m_{\text{vec}} = gv. \quad (39)$$

The effective potential Eq. (35) is then a sum over dyon species in addition to the integral over proper time.

$$\begin{aligned} V(\theta) &\sim \sum_{\ell \in \mathbb{Z}} \int_0^\infty \frac{d\tau}{\tau^\alpha} \exp\left\{-\frac{\tau}{2} \left[m_{\text{mon}}^2 + m_{\text{vec}}^2 \left(\ell - \frac{\theta}{2\pi\mu}\right)^2\right]\right\} \\ &\sim \sum_{n \in \mathbb{Z}} \int_0^\infty \frac{d\tau}{\tau^\beta} \exp\left\{-\frac{m_{\text{mon}}^2 \tau}{2} - \frac{2\pi^2 n^2}{m_{\text{vec}}^2 \tau} + \frac{i n \theta}{\mu}\right\}, \end{aligned} \quad (40)$$

where above we have used Poisson resummation. Finally, evaluating the proper time integral in a saddle point approximation yields a critical proper time,

$$\tau_* = \frac{2\pi n}{m_{\text{mon}} m_{\text{vec}}}, \quad (41)$$

and a corresponding potential,

$$\begin{aligned} V(\theta) &\approx \sum_{n \in \mathbb{N}} c_n \exp\left(-\frac{2\pi n m_{\text{mon}}}{m_{\text{vec}}}\right) \cos\left(\frac{n\theta}{\mu}\right) \\ &\approx \sum_{n \in \mathbb{N}} c_n \exp\left(-\frac{8\pi^2 n}{g^2}\right) \cos\left(\frac{n\theta}{\mu}\right). \end{aligned} \quad (42)$$

Thus, loops of monopoles generate an effective potential for the axion which is exponentially suppressed.

The result Eq. (42) can also be understood from the ultraviolet point of view as the familiar axion potential generated by non-Abelian instantons. Indeed the weight of each term in the sum (42) is precisely the non-Abelian instanton action. What the presentation in terms of monopoles has accomplished is it has made manifest the link between 1-form symmetry breaking via dynamical monopoles and the exponentially suppressed axion potential. In particular, since this potential is the leading operator deformation violating the axion shift symmetry defects \mathcal{D}_k , it is technically natural.

B. Exponentially small axion energy splittings

The preceding example revisited a familiar small axion potential from the point of view of noninvertible symmetry and presented a link to monopole physics and 1-form symmetry breaking. We can now use the same technique to generate more novel technically natural models with exponential hierarchies.

As an example, we consider an axion Yang-Mills theory with infrared gauge group $\text{SO}(3)$. In addition to the kinetic terms involving the axion θ and the gauge field strength w , the action includes the interaction

$$S \supset \frac{i}{8\pi^2 \mu_{\text{IR}}} \int \theta \text{Tr}(w \wedge w). \quad (43)$$

Noting that $\text{SO}(3) \cong \text{PSU}(2)$; the discussion of Sec. II D applies. Specifically, the full periodicity of the axion is

$$\theta \sim \theta + 4\pi\mu_{\text{IR}}, \quad (44)$$

and there is a noninvertible symmetry defect \mathcal{D}_2 which acts to shift the axion by half of its full period $\theta \rightarrow \theta + 2\pi\mu_{\text{IR}}$. In this case, the topological degrees of freedom on the defect are the semion Chern-Simons theory $\text{U}(1)_2$.

The exact axion shift symmetry defect \mathcal{D}_2 implies exact selection rules on the axion potential. Specifically, a general potential compatible with the full periodicity Eq. (44) takes the form

$$V(\theta) = \sum_{n \in \mathbb{N}} \alpha_n \cos\left(\frac{n\theta}{2\mu_{\text{IR}}}\right), \quad (45)$$

where the coefficients α_n arise, for instance, from instantons which need not be suppressed due to strong non-Abelian

dynamics. However, the symmetry defect \mathcal{D}_2 implies the selection rule:

$$\alpha_n = 0, \quad \text{if } n \text{ odd}. \quad (46)$$

This in turn leads to two ground states where $\langle \theta \rangle = 0$ or $\langle \theta \rangle = 2\pi\mu_{\text{IR}}$, whose exact degeneracy is protected by the topological defect \mathcal{D}_2 .

We now embed this model as the long-distance limit of a flow from a larger gauge group which breaks the 1-form symmetry in the ultraviolet. Specifically, let the parent gauge group be $\text{SU}(3)$ with field strength W and suppose there is in addition a complex Higgs field Φ transforming in the $\mathbf{6}$ of $\text{SU}(3)$ (the symmetric product of two fundamental $\mathbf{3}$'s). In addition to the kinetic terms, the action includes the interactions:

$$S \supset \frac{i}{8\pi^2 \mu_{\text{UV}}} \int \theta \text{Tr}(W \wedge W) - \frac{1}{g^2} \int d^4x V(\Phi). \quad (47)$$

Here μ_{UV} is the ultraviolet axion decay constant, in terms of which the axion has standard periodicity $\theta \sim \theta + 2\pi\mu_{\text{UV}}$.

We now assume that the potential is chosen to condense the Higgs field so that (up to gauge transformations) the expectation value of the scalar $\langle \Phi \rangle = v\delta_{ij}$ (the 3×3 identity matrix.). This expectation value Higgses the gauge group $\text{SU}(3)$ down to the subgroup that preserves δ_{ij} , which is $\text{SO}(3)$. At long distances we thus obtain $\text{SO}(3)$ axion Yang-Mills theory, together with an adjoint scalar field.

We must also keep track of the properties of the instanton density under this pattern of Higgsing. Crucially, the index of embedding of $\text{SO}(3)$ inside $\text{SU}(3)$ is four [39]. This means that a single instanton of $\text{SO}(3)$ has instanton number four when embedded inside $\text{SU}(3)$. In particular, this implies that the relation between the axion decay constants is

$$\mu_{\text{UV}} = 4\mu_{\text{IR}}. \quad (48)$$

Let us next turn to the axion potential generated by the flow. Since the 1-form symmetry of the infrared theory is broken by dynamical monopoles, we anticipate a non-perturbative violation of the \mathcal{D}_2 selection rules. Indeed, assuming that the Higgs scale v is chosen so that at the scale v the $\text{SU}(3)$ theory is weakly coupled, $\text{SU}(3)$ instantons will generate a small change in the axion potential $\delta V(\theta)$ which admits an expansion of the form

$$\delta V(\theta) = \sum_{n \in \mathbb{N}} \beta_n v^4 \exp\left(-\frac{8\pi^2 n}{g^2(v)}\right) \cos\left(\frac{n\theta}{\mu_{\text{UV}}}\right). \quad (49)$$

In particular, using Eq. (48), and comparing to the general IR expansion Eq. (45), we observe that all Fourier modes are now present in $\delta V(\theta)$ [40].

We conclude that the selection rule (46) is now violated nonperturbatively, and hence there is an exponentially suppressed energy splitting between the nearly degenerate states:

$$|\delta V(0) - \delta V(2\pi\mu_{\text{IR}})| \sim v^4 \exp\left(-\frac{8\pi^2}{g^2(v)}\right). \quad (50)$$

Since this is the leading effect violating the defect \mathcal{D}_2 , this exponential potential splitting is technically natural.

C. Exponentially suppressed mass terms in QED

The previous examples used the breaking of noninvertible symmetries to construct axion potentials with interesting features. Here we apply the same mechanism to produce a small breaking of the noninvertible chiral symmetry defect \mathcal{D}_k in massless QED. This results in technically natural models with exponentially suppressed fermion masses.

As the simplest example, let the ultraviolet be SU(2) gauge theory with two Weyl fermions χ_i and ψ_i in the fundamental representation (2) where $i = 1, 2$ is a doublet gauge index. (Note that the number of Weyl doublets must be even for consistency with the Witten anomaly [41].) In addition, we also include a real adjoint scalar Higgs field Φ_{ij} . At short distances, SU(2) instantons completely break the chiral symmetry rotating the fermions χ and ψ , except for the fermion parity $(-1)^F$ which is contained in the SU(2) gauge group. Anticipating the discussion to follow, we also note that this theory does not have any 1-form symmetry.

We now use a potential $V(\Phi)$ to condense the scalar, $\langle \Phi_{ij} \rangle = v\delta_{ij}$. This Higgses the gauge group down to $\text{SO}(2) \cong \text{U}(1)$. The long-distance theory has two fermions χ_+ with charge +1 and two fermions χ_- with charge -1.

If the bare mass term for the fermions vanishes in the ultraviolet, then at long distances this model will have emergent noninvertible chiral symmetries acting on the fermions of the type discussed in Sec. II A. Crucially, these chiral symmetries forbid a fermion mass term. However, since the 1-form symmetry is broken at short distances, we expect that this noninvertible symmetry is broken nonperturbatively by monopole bubbles, generating an exponentially suppressed mass term. This expectation can be directly verified using SU(2) instanton calculus. The 't Hooft vertex coming from the sector of instanton number one generates a bilinear $\chi\psi$ term. At long distances this gives a mass correction suppressed by the instanton factor $\sim e^{-8\pi^2/g^2(v)}$.

Thus, this simple example generates a nonperturbative violation of the noninvertible chiral symmetry selection rules. However, it is unsatisfactory, and not technically natural, since there is no mechanism prohibiting a bare fermion mass term in the ultraviolet.

To remedy this, and construct a technically natural model with exponentially light fermion masses, we consider a slight generalization of the model above. We now take the ultraviolet to be SU(2) gauge theory with $2N_f$ Weyl fermions and also make the adjoint scalar Φ_{ij} complex. We denote the first of the fermions doublets as χ , and the rest of the doublets as ψ^S , with flavor index $S = 1, \dots, 2N_f - 1$. We aim to flow at long distances to QED with one light electron χ_+ and positron χ_- .

To achieve this in a technically natural way, we impose on the ultraviolet model the following (invertible, ABJ-anomaly-free) symmetry:

$$\chi \rightarrow \zeta\chi, \quad \psi^S \rightarrow \tilde{\zeta}\zeta\psi^S, \quad \Phi \rightarrow (\tilde{\zeta}\zeta)^{-2}\Phi, \quad (51)$$

where $\zeta^{2N_f} = 1$ and $\tilde{\zeta}^{2N_f-1} = 1$ are roots of unity. This transformation generates a $\mathbb{Z}_{2N_f(2N_f-1)}$ invertible symmetry. This symmetry prohibits the bare mass for the fermions and the Yukawa coupling between ϕ and χ . However, it allows a Yukawa interaction of the form:

$$y_{ST}\phi_{ij}\psi^{S,i}\psi^{T,j}, \quad (52)$$

where y_{ST} is a symmetric matrix of couplings.

We again Higgs to U(1) by giving an expectation value $\langle \Phi_{ij} \rangle = v\delta_{ij}$. The Yukawa term (52) implies that the $2N_f - 1$ SU(2) fundamental fermions ψ^S get a mass $v y_{ST}$ and only one electron-positron pair $\chi = (\chi_+, \chi_-)$ remains light. Again, the mass term for χ is protected at long distances only by the noninvertible chiral symmetry which is broken by dynamical monopoles.

Following our previous examples, we thus expect that SU(2) instantons give an exponentially suppressed mass to χ . To see this directly consider the 't Hooft vertex of the SU(2) theory:

$$\frac{1}{v^{3N_f-4}} \exp\left(-\frac{8\pi^2}{g^2(v)}\right) \left(\chi \prod_S \psi^S\right). \quad (53)$$

Contracting two of these 't Hooft vertices with the Yukawa interactions, we obtain an effective two-instanton interaction:

$$\frac{y^{2N_f-1}}{v^{2N_f-2}} \exp\left(-\frac{16\pi^2}{g^2(v)}\right) \chi^2 (\Phi^\dagger)^{2N_f-1}, \quad (54)$$

which is neutral under the $\mathbb{Z}_{2N_f(N_f-1)}$ symmetry. Here, the expression (54) is to be understood as a sum of all possible gauge contractions, and y represents a typical value of elements in the symmetric Yukawa matrix y_{ST} defined in Eq. (52). Note also that when $N_f \geq 2$, this interaction term is irrelevant so that its bare value is naturally suppressed by a further UV (e.g., Planck) scale.

We now determine the effect of the interaction (54) at long distances by evaluating Φ at its expectation value $v\delta_{ij}$.

This produces an effective mass term generated by two SU(2) instantons:

$$vy^{2N_f-1} \exp\left(-\frac{16\pi^2}{g^2(v)}\right)\chi^2. \quad (55)$$

An alternative way to see this two-instanton contribution is to retain the massive fermions ψ^S in the IR theory. Then the one-instanton 't Hooft vertex, when contracted with the Yukawa, gives a tiny fermion mixing term:

$$vy^{2N_f-2} \exp\left(-\frac{8\pi^2}{g^2(v)}\right)\chi\psi^S, \quad (56)$$

which in turn gives a tiny nonzero value to the lighter mass eigenvalue via the seesaw mechanism.

As with our models of axions, this is a technically natural suppression of the mass term, protected by the weakly broken noninvertible chiral symmetry.

D. Exponentially suppressed mass terms in QCD

In the previous example the long-distance theory was Abelian gauge theory with a noninvertible chiral symmetry which is defined for all rational angles. Here we present a variant of the above model where the infrared theory is a non-Abelian gauge theory whose noninvertible chiral symmetry protecting the fermion masses is broken by an ultraviolet instanton.

The short-distance model is a SU(N_c) gauge theory, which we Higgs down to SO(N_c). As matter content we take N_f fundamental and antifundamental Weyl fermions and two complex scalars Φ_1, Φ_2 in the representation with two symmetric fundamental gauge indices. As before, we separate one of the fundamental fermions χ from the rest of the fundamental fermions ψ^S , $S = 1, \dots, N_f - 1$, and we indicate all of the antifundamental fermions as $\bar{\psi}^{S'}$, $S' = 1, \dots, N_f$.

On this ultraviolet model, we impose an invertible $\mathbb{Z}_{2N_f(N_f-1)}$ symmetry acting as

$$\begin{aligned} \chi &\rightarrow \zeta\chi, & \psi^S &\rightarrow \tilde{\zeta}\zeta\psi^S, & \bar{\psi}^{S'} &\rightarrow (\tilde{\zeta}\zeta)\bar{\psi}^{S'}, \\ \Phi_1 &\rightarrow (\tilde{\zeta}\zeta)^2\Phi_1, & \Phi_2 &\rightarrow (\tilde{\zeta}\zeta)^{-2}\Phi_2, \end{aligned} \quad (57)$$

where $\zeta^{2N_f} = 1$ and $\tilde{\zeta}^{2(N_f-1)} = 1$ are roots of unity generating the symmetry action. This symmetry allows the Yukawa interactions $y_{ST}(\Phi_1)^\dagger\psi^S\psi^S$ and $y'_{S'T'}\Phi_2\bar{\psi}^{S'}\bar{\psi}^{T'}$.

By giving an expectation value to the Higgs fields $\langle\Phi_{1,2}^{ij}\rangle = v_{1,2}\delta^{ij}$, the theory is Higgsed down to a SO(N_c) gauge theory. There is a single light Weyl fermion χ in the vector representation of SO(N_c), while the other fermions acquire a mass directly from the Yukawa interactions. (For simplicity, below we do not distinguish the

separate expectation values v_1 and v_2 , indicating both of their scales as v .)

Note that when $N_c = 3$, this pattern of Higgsing is the same as that investigated in Sec. III B and also discussed in Ref. [12]. In particular, in infrared SO(N_c) gauge theory the discrete chiral symmetry $\chi \rightarrow i\chi$ is noninvertible and forbids a mass term for χ . This holds for a general N_c .

In the parent SU(N_c) theory, dynamical monopoles break the 1-form symmetry leading to a nonperturbative violation of the noninvertible chiral symmetry selection rules. This results in a technically natural mass term for χ suppressed by an ultraviolet instanton factor.

To see this directly, we again examine the SU(N_c) 't Hooft vertex:

$$\frac{1}{v^{3N_f-4}} \exp\left(-\frac{8\pi^2}{g^2(v)}\right)\chi \prod_{S=1}^{N_f-1} \psi^S \prod_{S'=1}^{N_f} \bar{\psi}^{S'}. \quad (58)$$

As before, we can contract two such vertices and the Yukawa couplings to arrive at the effective interaction:

$$\frac{y^{2N_f-1}}{v^{2N_f-2}} \exp\left(-\frac{16\pi^2}{g^2(v)}\right)\chi^2\Phi_1^{N_f-1}(\Phi_2^\dagger)^{N_f}. \quad (59)$$

Evaluating at the expectation value of the Higgs fields, this turns into an exponentially suppressed mass term for the SO(N_c) vector fermion χ :

$$vy^{2N_f-1} \exp\left(-\frac{16\pi^2}{g^2(v)}\right)\chi^2. \quad (60)$$

This tiny mass is protected by the noninvertible chiral symmetry.

ACKNOWLEDGMENTS

We thank L. Delacretaz, H. Fukuda, and J. Harvey for useful discussions. We also thank the authors of Ref. [42] for sharing a draft of their work. C. C. is supported by the U.S. Department of Energy DE-SC0009924 and the Simons Collaboration on Global Categorical Symmetries. K. O. is supported in part by JSPS KAKENHI Grant-in-Aid, No. 22K13969 and the Simons Collaboration on Global Categorical Symmetries.

Note added.—Recently, we were notified of Ref. [42], which independently constructs and studies the noninvertible chiral symmetry defects in massless QED.

-
- [1] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, *Generalized Global Symmetries*, *J. High Energy Phys.* **02** (2018) 172.

- [2] J. Fuchs, I. Runkel, and C. Schweigert, *TFT Construction of RCFT Correlators I. Partition Functions*, *Nucl. Phys.* **B646**, 353 (2002).
- [3] J. Frohlich, J. Fuchs, I. Runkel, and C. Schweigert, *Kramers-Wannier Duality from Conformal Defects*, *Phys. Rev. Lett.* **93**, 070601 (2004).
- [4] J. Frohlich, J. Fuchs, I. Runkel, and C. Schweigert, *Duality and Defects in Rational Conformal Field Theory*, *Nucl. Phys.* **B763**, 354 (2007).
- [5] L. Bhardwaj and Y. Tachikawa, *On Finite Symmetries and Their Gauging in Two Dimensions*, *J. High Energy Phys.* **03** (2018) 189.
- [6] C.-M. Chang, Y.-H. Lin, S.-H. Shao, Y. Wang, and X. Yin, *Topological Defect Lines and Renormalization Group Flows in Two Dimensions*, *J. High Energy Phys.* **01** (2019) 026.
- [7] R. Thomgren and Y. Wang, *Fusion Category Symmetry I: Anomaly In-Flow and Gapped Phases*, [arXiv:1912.02817](https://arxiv.org/abs/1912.02817).
- [8] R. Thomgren and Y. Wang, *Fusion Category Symmetry II: Categoriosities at $c = 1$ and Beyond*, [arXiv:2106.12577](https://arxiv.org/abs/2106.12577).
- [9] Z. Komargodski, K. Ohmori, K. Roumpedakis, and S. Seifnashri, *Symmetries and Strings of Adjoint QCD₂*, *J. High Energy Phys.* **03** (2021) 103.
- [10] M. Koide, Y. Nagoya, and S. Yamaguchi, *Non-Invertible Topological Defects in 4-Dimensional \mathbb{Z}_2 Pure Lattice Gauge Theory*, *Prog. Theor. Exp. Phys.* **2022**, 013B03 (2022).
- [11] Y. Choi, C. Cordova, P.-S. Hsin, H. T. Lam, and S.-H. Shao, *Noninvertible Duality Defects in 3 + 1 Dimensions*, *Phys. Rev. D* **105**, 125016 (2022).
- [12] J. Kaidi, K. Ohmori, and Y. Zheng, *Kramers-Wannier-like Duality Defects in (3 + 1)D Gauge Theories*, *Phys. Rev. Lett.* **128**, 111601 (2022).
- [13] Y. Hayashi and Y. Tanizaki, *Non-Invertible Self-Duality Defects of Cardy-Rabinovici Model and Mixed Gravitational Anomaly*, *J. High Energy Phys.* **08** (2022) 036.
- [14] Y. Choi, C. Cordova, P.-S. Hsin, H. T. Lam, and S.-H. Shao, *Non-Invertible Condensation, Duality, and Triality Defects in 3 + 1 Dimensions*, [arXiv:2204.09025](https://arxiv.org/abs/2204.09025).
- [15] J. Kaidi, G. Zafrir, and Y. Zheng, *Non-Invertible Symmetries of $\mathcal{N} = 4$ SYM and Twisted Compactification*, *J. High Energy Phys.* **08** (2022) 053.
- [16] K. Roumpedakis, S. Seifnashri, and S.-H. Shao, *Higher Gauging and Non-Invertible Condensation Defects*, [arXiv:2204.02407](https://arxiv.org/abs/2204.02407).
- [17] L. Bhardwaj, L. Bottini, S. Schafer-Nameki, and A. Tiwari, *Non-Invertible Higher-Categorical Symmetries*, *SciPost Phys.* **14**, 007 (2022).
- [18] S. L. Adler, *Axial Vector Vertex in Spinor Electrodynamics*, *Phys. Rev.* **177**, 2426 (1969).
- [19] J. S. Bell and R. Jackiw, *A PCAC Puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ Model*, *Nuovo Cimento A* **60**, 47 (1969).
- [20] P.-S. Hsin, H. T. Lam, and N. Seiberg, *Comments on One-Form Global Symmetries and Their Gauging in 3d and 4d*, *SciPost Phys.* **6**, 039 (2019).
- [21] A. Kapustin and N. Seiberg, *Coupling a QFT to a TQFT and Duality*, *J. High Energy Phys.* **04** (2014) 001.
- [22] G. 't Hooft, *Naturalness, Chiral Symmetry, and Spontaneous Chiral Symmetry Breaking*, *NATO Sci. Ser. B* **59**, 135 (1980).
- [23] J. J. Fan, K. Fraser, M. Reece, and J. Stout, *Axion Mass from Magnetic Monopole Loops*, *Phys. Rev. Lett.* **127**, 131602 (2021).
- [24] C. Córdova, S. Hong, S. Koren, and K. Ohmori, *Neutrino Masses from Generalized Symmetry Breaking*, [arXiv:2211.07639](https://arxiv.org/abs/2211.07639).
- [25] The Pontryagin square is a cohomology operation:
- $$\mathcal{P}: H^2(M, \mathbb{Z}_N) \rightarrow H^4(M, \mathbb{Z}_{\gcd(N,2N)}).$$
- In the case where $B^{(2)} \in H^2(M, \mathbb{Z}_N)$ admits an integral uplift $\tilde{B}^{(2)} \in H^2(M, \mathbb{Z})$ (the main case of interest in the present paper), we have $\mathcal{P}(B^{(2)}) = \tilde{B}^{(2)} \wedge \tilde{B}^{(2)} \bmod 2N$.
- [26] M. Barkeshli, P. Bonderson, M. Cheng, and Z. Wang, *Symmetry Fractionalization, Defects, and Gauging of Topological Phases*, *Phys. Rev. B* **100**, 115147 (2019).
- [27] D. Gaiotto and T. Johnson-Freyd, *Condensations in Higher Categories*, [arXiv:1905.09566](https://arxiv.org/abs/1905.09566).
- [28] D. V. Else and C. Nayak, *Cheshire Charge in (3 + 1)-Dimensional Topological Phases*, *Phys. Rev. B* **96**, 045136 (2017).
- [29] T. Johnson-Freyd, *(3 + 1)D Topological Orders with Only a \mathbb{Z}_2 -Charged Particle*, [arXiv:2011.11165](https://arxiv.org/abs/2011.11165).
- [30] J. E. Kim, *Weak Interaction Singlet and Strong CP Invariance*, *Phys. Rev. Lett.* **43**, 103 (1979).
- [31] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *Can Confinement Ensure Natural CP Invariance of Strong Interactions?*, *Nucl. Phys.* **B166**, 493 (1980).
- [32] D. Harlow and H. Ooguri, *Symmetries in Quantum Field Theory and Quantum Gravity*, *Commun. Math. Phys.* **383**, 1669 (2021).
- [33] S. R. Coleman and E. J. Weinberg, *Radiative Corrections as the Origin of Spontaneous Symmetry Breaking*, *Phys. Rev. D* **7**, 1888 (1973).
- [34] M. Kawasaki, F. Takahashi, and M. Yamada, *Suppressing the QCD Axion Abundance by Hidden Monopoles*, *Phys. Lett. B* **753**, 677 (2016).
- [35] Y. Nomura, S. Rajendran, and F. Sanches, *Axion Isocurvature and Magnetic Monopoles*, *Phys. Rev. Lett.* **116**, 141803 (2016).
- [36] M. Kawasaki, F. Takahashi, and M. Yamada, *Adiabatic Suppression of the Axion Abundance and Isocurvature due to Coupling to Hidden Monopoles*, *J. High Energy Phys.* **01** (2018) 053.
- [37] E. Witten, *Dyons of Charge $e\theta/2\pi$* , *Phys. Lett.* **86B**, 283 (1979).
- [38] W. Fischler and J. Preskill, *Dyon-Axion Dynamics*, *Phys. Lett. B* **125**, 165 (1983).
- [39] C. Csaki and H. Murayama, *Instantons in Partially Broken Gauge Groups*, *Nucl. Phys.* **B532**, 498 (1998).
- [40] There are also naively modes violating the periodicity Eq. (44). Consistency of the above analysis therefore requires that odd numbered instantons of SU(3) do not contribute to the axion potential. A related vanishing of the contribution to the superpotential from odd numbered SU(3) instantons in the same Higgsing pattern in a supersymmetric context was observed in Ref. [39]. Both phenomena likely follow from the fact that in these models the dyon is a fermion.
- [41] E. Witten, *An SU(2) Anomaly*, *Phys. Lett. B* **117**, 324 (1982).
- [42] Y. Choi, H. T. Lam, and S.-H. Shao, *Noninvertible Global Symmetries in the Standard Model*, *Phys. Rev. Lett.* **129**, 161601 (2022).