Non-Fermi Liquids as Ersatz Fermi Liquids: General Constraints on Compressible Metals

Dominic V. Else[,1] Ryan Thorngren,[2] and T. Senthil[1]

[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Center for Mathematical Sciences and Applications, Harvard University,*
Cambridge, Massachusetts 02138, USA

A system with charge conservation and lattice translation symmetry has a well-defined filling ν, which is a real number representing the average charge per unit cell. We show that if ν is fractional (i.e., not an integer), this imposes very strong constraints on the low-energy theory of the system and give a framework to understand such constraints in great generality, vastly generalizing the Luttinger and Lieb-Schultz-Mattis theorems. The most powerful constraint comes about if ν is continuously tunable (i.e., the system is charge compressible), in which case, we show that the low-energy theory must have a very large emergent symmetry group—larger than any compact Lie group. An example is the Fermi surface of a Fermi liquid, where the charge at every point on the Fermi surface is conserved. We expect that in many, if not all, cases, even exotic non-Fermi liquids will have the same emergent symmetry group as a Fermi liquid, even though they could have very different dynamics. We call a system with this property an ersatz Fermi liquid. We show that ersatz Fermi liquids share a number of properties in common with Fermi liquids, including Luttinger's theorem (which is thus extended to a large class of non-Fermi liquids) and periodic "quantum oscillations" in the response to an applied magnetic field. We also establish versions of Luttinger's theorem for the composite Fermi liquid in quantum Hall systems and for spinon Fermi surfaces in Mott insulators. Our work makes a connection between filling constraints and the theory of symmetry-protected topological phases, in particular through the concept of "'t Hooft anomalies."

I. INTRODUCTION

In condensed matter physics, systems with prescribed microscopic degrees of freedom (usually electrons) can exhibit varied and exotic emergent behavior at low energies. In general, it is extremely difficult, either analytically or numerically, to predict the nature of the emergent low-energy behavior [as described by an "IR (infrared) theory"] from the properties of the microscopic degrees of freedom. For this reason, it is invaluable to have general results that constrain the nature of the IR theory, given the microscopic properties of the system.

One such result is the Lieb-Schultz-Mattis theorem [1], which states (in the formulation of interest to us here [2]) that in a system in one spatial dimension with a conserved U(1) charge and discrete translation symmetry, if the average charge per unit cell (which we call the "filling") is not an integer, then either the IR theory is gapless or else

it spontaneously breaks one of the symmetries. This result was later generalized to higher dimensions by Oshikawa [3] and Hastings [4]; in higher dimensions, there is also the possibility that a system at fractional filling can be gapped but with nontrivial topological order [5].

A related result is Luttinger's theorem [6], which states that if the IR theory is a Fermi liquid, then the volume enclosed by the Fermi surface (modulo the volume of the Brillouin zone) is determined entirely by the fractional part of the filling, and in particular is independent of the interaction strength. Originally proven perturbatively by Luttinger, the result was later established through a non-perturbative argument (but still assuming that the IR theory is describable by Fermi-liquid theory) by Oshikawa [7]. A generalized version of Luttinger's theorem even holds in a class of phases known as fractionalized Fermi liquids [8,9], which are distinct from conventional Fermi liquids. In such phases, a gapless Fermi liquid coexists with nontrivial topological order. The generalization to Luttinger's theorem can then exactly be determined from the interplay of the symmetry with the topological order [9–11].

These results and others [12–14] raise the question: What is the most general statement that one can make about the relation between the microscopic filling and properties

of the IR theory? In the present work, we answer this question, showing that the microscopic filling is always completely determined by a few properties of the IR theory, namely, (a) its emergent symmetries, (b) the relation between the microscopic translation and U(1) symmetries and the emergent symmetries of the IR theory, and (c) a property of the IR theory called its "'t Hooft anomaly." Some connections between filling and 't Hooft anomalies have previously been explored in Refs. [15–20].

Our work has important implications for the study of "non-Fermi liquids," which are systems that are metallic down to zero temperature but for which the IR physics cannot be described by Fermi-liquid theory. Motivated by the results on filling constraints just mentioned, we introduce the concept of an ersatz Fermi liquid (EFL) as a general framework to understand non-Fermi liquids. An EFL is a system which has the same *kinematic* properties in the deep infrared as a Fermi liquid, though it might have very different *dynamical* properties. By "kinematic" properties, we mean properties that relate to the structure of the Hilbert space that describes the ground state and the low-energy excitations, as opposed to "dynamical" properties which relate to the Hamiltonian that acts in this Hilbert space. More precisely, the kinematic properties comprise properties (a)–(c) described in the previous paragraph and in more detail in the next section. Further, it turns out that the kinematic properties we discuss have a strong topological flavor, so in a suitably vague sense, we can say that an EFL is a system that is "topologically equivalent" to a Fermi liquid. (However, since the dynamical properties of an EFL can be sharply different from that of a Fermi liquid, it will generally *not* be the case that an EFL can be continuously deformed into a Fermi liquid.)

We show that many (though by no means all) well-known aspects of Fermi-liquid phenomenology are, in fact, purely kinematic in nature, and therefore, apply equally well to any EFL. In particular, we show that any EFL has a Fermi surface that hosts long-lived excitations (though these excitations may not be Landau quasiparticles). In fact, our approach leads to a very general perspective on what it means for a system to have a Fermi surface. Furthermore, we find that this Fermi surface must obey Luttinger's theorem (or a generalization thereof, analogous to the "fractionalized Fermi liquids" mentioned above), and that, if the Fermi-surface geometry is such that a Fermi liquid with that geometry would exhibit quantum oscillations in the dependence of physical properties on magnetic field, then any EFL with the same Fermi-surface geometry is also expected to display quantum oscillations with the same periodicity.

Since none of the kinematic properties that define an EFL require that the system be weakly coupled or have a description in terms of quasiparticles, we expect a wide variety of exotic non-Fermi-liquid phenomena to be realizable within the class of EFLs. In fact, we argue based on

the general theory of filling constraints that any IR theory which describes a compressible metal, i.e., the filling can be continuously tuned, [21] must have a very nontrivial emergent symmetry group, larger than any compact Lie group. Such a property is indeed satisfied by EFLs (due to the infinitely many conserved quantities associated with the Fermi surface); whether it could be satisfied in a different way that leads to fundamentally different kinematic properties is an important open problem.

At the very least, however, it is clear that a number of non-Fermi-liquid metals can be fruitfully discussed from the perspective developed in this paper. The simplest are non-Fermi-liquid metals that arise when a Fermi surface is coupled to a critical boson that represents a fluctuating order parameter. Within the standard framework for such quantum critical points, they are seen to be EFLs. A closely related system is a Fermi surface coupled to a gapless U(1) gauge field, which arises in the theory of composite Fermi-liquid metals in the quantum Hall regime, and in some *insulating* quantum spin liquids with Fermi surfaces of emergent electrically neutral quasiparticles. In these states, we find that, unlike a strict EFL, the kinematic properties of the IR theories can differ from those of a Fermi liquid (indeed, unlike a Fermi liquid, these are not states in which the microscopic density can be continuously tuned) though still closely related. The main point still stands though, that the kinematic properties of these IR theories provide a powerful framework to thinking about their universal behavior. In particular, we show that in such systems a version of Luttinger's theorem is still satisfied.

As a final application of our results, we examine the possibility for systems to exhibit disconnected Fermi arcs instead of a closed Fermi surface. We find that, assuming the translational symmetry is unbroken, such a scenario is inconsistent with the IR theory being an EFL, except when the system exists on the boundary of a gapless bulk (such as a Weyl semimetal). This provides strong evidence for the impossibility of Fermi arcs. In fact, we obtain a stronger constraint: In an EFL, the Fermi surface must enclose a volume in the Brillouin zone; this volume is the generalization to EFLs of the "electron sea" in a weakly interacting system.

Before proceeding, let us make a final technical remark. We note that previous work [22–24] generalized the original perturbative proof of Luttinger's theorem to situations where there are a number of fermion and boson fields with U(1) symmetries (either global or gauge). This proof relies on the existence of a Luttinger-Ward functional of the exact Green's functions of the fermions or bosons from which the self-energies can be extracted by functional derivatives. These works lead to the expectation that Luttinger's theorem will be satisfied by some classes of non-Fermi liquids. However, as the quasiparticle is destroyed in the IR in such problems, perturbation theory (even to all orders) should be used with caution. In a *nonperturbative* context, despite the

formal existence [25] of a Luttinger-Ward functional, the conventional proof of Luttinger's theorem fails. This is dramatically illustrated by the fractionalized Fermi-liquid phases. It is thus desirable to have a more general non-perturbative argument for Luttinger's theorem in this context, which we provide in this paper.

II. KINEMATIC PROPERTIES OF THE IR THEORY

Among all the features of the IR theory describing the low-energy properties of a system, there are certain ones that we refer to as kinematic, and these are the subject of this section.

Before describing the first kinematic property, let us note that the microscopic system has a group of symmetries G_{UV}. We are specifically interested in systems where this includes a global U(1) corresponding to charge conservation and translation symmetries (possibly on a lattice).

The first kinematic property of the IR theory is the set of emergent symmetries. We thus introduce the group G_{IR} of emergent symmetries [26] of the IR theory. G_{IR} is, in general, not the same as the microscopic symmetry group G_{UV}. Nevertheless, each element $g \in G_{UV}$ of the microscopic symmetry group gets mapped into an element $\varphi(g) \in G_{IR}$ of the emergent symmetry group, such that $\varphi(g_1)\varphi(g_2) = \varphi(g_1 g_2)$ for all $g_1, g_2 \in G$. There could, of course, be elements of the emergent symmetry group G_{IR} that do not correspond to any microscopic symmetry. Also, a microscopic symmetry $g \in G_{UV}$ could act trivially in the IR theory, in which case $\varphi(g) = 1$ (the identity element in G_{IR}); an example of the latter case would be for systems with a charge gap, in which case the microscopic U(1) symmetry acts trivially in the IR. All these statements can be expressed in a compact mathematical way by saying that φ defines a group homomorphism from G_{UV} to G_{IR} that need not be injective or surjective.

Note that for many of the arguments in this paper, we specifically want G_{IR} to represent the *internal* symmetries of the system (that is, the symmetries which do not move space-time points around). In general, the IR theory also has "trivial" emergent space-time symmetries such as continuous translation symmetry, which we do not include in G_{IR}. It is important to note that the microscopic translation symmetry in general does not map into these trivial translation symmetries, but rather into the internal symmetry G_{IR}. One way to think about this is that since there is a spatial rescaling transformation associated with passing to the IR theory, a microscopic translation symmetry, in fact, has trivial translation action in the IR limit. An alternative perspective is that we imagine that a microscopic translation symmetry acts like a product of an internal symmetry and a trivial translation symmetry, and we worry about only the internal part in defining the map φ.

The other property of the IR theory that is pertinent is the extent to which the full emergent symmetry group G_{IR} can be naturally realized in *some* realization (not necessarily the original microscopic lattice model) of the IR theory. This property is formalized by the concept of the 't Hooft anomaly. Such an anomaly in $d + 1$ space-time dimensions is an obstruction to UV regularizing the theory on a lattice in d spatial dimensions with the full emergent symmetry group G_{IR} realized as an "on-site" microscopic symmetry [28,29]. A powerful alternative but formal characterization of a 't Hooft anomaly [30,31] is that the conservation law corresponding to the G_{IR} symmetry is broken upon coupling to a background gauge field for the symmetry G_{IR}. We consider the 't Hooft anomaly of the emergent symmetry G_{IR} to also be a kinematic property of the IR theory.

't Hooft anomalies are also closely related to the theory of symmetry-protected topological (SPT) phases [32–34]. These are gapped phases of matter in a system with an unbroken global symmetry G with the following property: The ground state cannot be continuously deformed into a trivial product-state ground state while preserving the symmetry without closing the gap if the symmetry G is preserved, but it can be if the symmetry G is lifted. Examples of SPT phases include the celebrated topological insulators [35] and the Haldane phase of the spin-1 antiferromagnetic chain in $d = 1$ [36–38].

The connection between 't Hooft anomalies and SPT phases is that a G SPT phase in $d + 1$ spatial dimensions must have a nontrivial boundary theory, and the boundary theory carries a 't Hooft anomaly. We can say that the 't Hooft anomaly of the boundary theory is "canceled" by inflow from the bulk, in the sense that the conservation laws of the bulk + boundary system are preserved in the presence of background gauge fields. Thus, the classification of 't Hooft anomalies in d spatial dimensions is precisely equivalent to the classification of SPT phases in $d + 1$ spatial dimensions.[39] Such a classification has been explored at great depth, from a variety of perspectives, ranging from physical considerations to very formal ones [28,29,33,40–64]. The aspects of SPT phases that we need to use in this paper, however, are simple enough, at least if we want to understand filling in systems of low spatial dimension $d \leq 2$ that a reader unfamiliar with this literature should still be able to follow our paper.

A. Example: Luttinger liquid in one spatial dimension

Let us illustrate the above general considerations in the case of a system of spinless electrons in a lattice in one spatial dimension. Thus, the microscopic symmetry group G_{UV} is comprised of a U(1) symmetry generated by the total electron number $\hat{Q}$, and a $\mathbb{Z}$ symmetry generated by the lattice translation operator $\mathbb{T}$.

Let us now assume that the IR theory of the electrons is a Luttinger liquid. Thus, the low-energy physics takes place at the two Fermi points at momenta k_L and k_R. Excitations with momentum close to k_L are left movers, and those with

momentum close to k_R are right movers. At low energies, the numbers $\hat{N}_L$ and $\hat{N}_R$ of left and right movers are separately conserved. Therefore, they generate the emergent symmetry group $G_{\text{IR}} = \text{U}(1)_L \times \text{U}(1)_R$.

Next, we need to specify how the microscopic symmetry acts on the IR theory, which we can do by expressing the generators of the microscopic symmetry in terms of the generators of the emergent symmetry. Indeed, we have

$$\hat{Q} \sim \hat{N}_L + \hat{N}_R, \tag{1}$$

$$\mathbb{T} \sim \exp(-i[k_L \hat{N}_L + k_R \hat{N}_R]), \tag{2}$$

where the tilde "$\sim$" refers to an equivalent action on the IR theory.

The emergent symmetry group G_{IR} in a Luttinger liquid has the well-known axial anomaly. This is an example of the 't Hooft anomaly mentioned in the previous subsection. The signature is that, if we turn on an electric field E for the U(1) symmetry generated by $\hat{N}_L + \hat{N}_R$ [which is equivalent to a microscopic electric field, by Eq. (1)], then left- and right-moving charges are no longer separately conserved; instead, if we let $j_L{}^\mu, \mu = 0, 1$ be the current density for the left-moving charge, and similar for $j_R{}^\mu$, we have

$$\partial_\mu j_L{}^\mu = -\frac{E}{2\pi},$$
$$\partial_\mu j_R{}^\mu = \frac{E}{2\pi}. \tag{3}$$

One can easily understand this equation in the case of a noninteracting Fermi gas. In that case, in the absence of electric field, the electrons occupy single-particle states labeled by momentum k. But applying an electric field pointing to the right causes an overall flow of electrons in momentum space according to $\dot{k} = E$. This causes a charge excess to accumulate at k_R and a corresponding charge deficit at k_L.

As we mention above, there is always a bulk-boundary correspondence that relates a 't Hooft anomaly for a symmetry G in d spatial dimensions to a SPT phase in $d + 1$ spatial dimensions whose boundary theory carries the 't Hooft anomaly. In this case, we have $d = 1$ and $G = \text{U}(1) \times \text{U}(1)$. The corresponding SPT phase in $d = 2$ is realized by a "quantum spin Hall" state corresponding to putting spin-up electrons in a quantum Hall state with quantized Hall conductance $\sigma_{xy} = 1$ and spin-down electrons in a quantum Hall state with $\sigma_{xy} = -1$. Here the U(1) × U(1) symmetries correspond to the separate conservation of spin-up and spin-down electrons. If we now consider a system with boundary and apply an electric field parallel to the boundary, due to the Hall conductance this will generate a current of spin-up electrons incident onto the boundary, and a current of spin-down electrons with opposite sign. This precisely accounts for the charge

nonconservation in the boundary theory due to the 't Hooft anomaly.

Finally, let us remark that a useful way to think about SPT phases (and hence, 't Hooft anomalies) is in terms of topological terms describing the response of the SPT phase to background gauge fields. For example, consider the quantum spin Hall state described above. We can theoretically couple to background gauge fields of the two U(1) symmetries, which in the one-dimensional Luttinger liquid are interpreted as the conservation of left and right movers; hence, we denote the gauge fields by A^L and A^R. The response of the quantum spin Hall state is then described by a Chern-Simons action on $(2 + 1)$D space-time:

$$S[A] = \frac{1}{4\pi} \int (A^R \wedge dA^R - A^L \wedge dA^L). \tag{4}$$

III. FILLING CONSTRAINTS

Now we turn to the question of how to understand constraints on the IR theory resulting from the microscopic filling, i.e., the average charge per unit cell, which is a real number ν. Only the fractional part of ν (i.e., $\nu \bmod 1$) should be expected to be detectable in the IR theory, because an atomic insulator (whose IR theory is completely trivial) can have any integer filling. The fundamental observation we make is that the microscopic filling is completely fixed by the kinematic properties of the IR theory, i.e., the emergent symmetry group and the 't Hooft anomaly, along with the mapping from the microscopic symmetry group into the emergent symmetry group.

This is a rare example of a precise relation between a microscopic quantity (the filling) and properties of the IR theory. Such relations are extremely useful, given that it is usually very difficult to determine the IR theory from the microscopic Hamiltonian, either analytically or numerically.

The reason why such a UV-IR correspondence is possible in this case is because the UV and IR are linked through the homomorphism φ that implements the microscopic symmetry inside the emergent IR symmetry. Naively, we can imagine arguing as follows. The lattice filling is defined in systems that have at least a microscopic U(1) symmetry (charge conservation), and a lattice translation symmetry $\mathbb{Z}^d$ (where d is the space dimension). So we set the microscopic symmetry $G_{\text{UV}} = \text{U}(1) \times \mathbb{Z}^d$. This G_{UV} is embedded into the symmetries of the IR theory, so we can talk about the 't Hooft anomaly of the IR theory thought of as a G_{UV}-symmetric theory (which can be computed from the G_{IR} 't Hooft anomaly in light of the homomorphism $\varphi : G_{\text{UV}} \to G_{\text{IR}}$). Then we invoke "UV-IR anomaly matching" to relate the filling ν to this G_{UV} 't Hooft anomaly.

Unfortunately, this argument is not *quite* right. In fact, fractional filling cannot correspond to a nontrivial 't Hooft anomaly for G_{UV} in the usual sense, because there are no

candidate SPT phases with $G_{\rm UV} = {\rm U}(1) \times \mathbb{Z}^d$ symmetry which could constitute the "bulk," for which a system with fractional filling could be the boundary, as we explain in the Appendix A. Indeed, a system with fractional filling seems perfectly well defined on its own without any need for a higher-dimensional bulk.

Instead, we give the correct version of the argument below in various spatial dimensions. The careful reader might notice that they still appear reminiscent of "anomaly matching," albeit for a 't Hooft anomaly that is trivial with respect to the microscopic symmetries. How exactly one could make such a notion precise, we leave as an open question. Nevertheless, the arguments are self-contained and hold without any need for such an interpretation.

A. One spatial dimension

In one spatial dimension, we imagine putting the system on a circle and then very slowly threading 2π flux of the microscopic U(1) symmetry through the circle [3,7,65]. By a standard argument, this transforms the ground state into a low-lying excited state with a different momentum; if we label states by their eigenvalues of the translation operator $\mathbb{T}$, i.e., $\mathbb{T}|\psi\rangle = e^{-ip}|\psi\rangle$, then the momentum gets shifted by $e^{-ip} \rightarrow e^{-i(p+2\pi\nu)}$.

Meanwhile, since the microscopic U(1) symmetry generated by $\hat{Q}$ corresponds to a U(1) symmetry of the IR theory, whose generator we call $\hat{Q}_{\rm IR}$, we can also imagine performing the 2π flux insertion in the IR theory. Now consider the IR symmetry $\tau = \varphi(\mathbb{T}) \in G_{\rm IR}$ corresponding to microscopic translation $\mathbb{T}$. In the IR theory, the ground state can get transformed into a low-lying excited state with a different eigenvalue of τ; that is, the eigenvalue of τ gets shifted according to $\exp(-i\theta) \rightarrow \exp(-i[\theta + \alpha])$ for some α (defined mod 2π). One can argue that α depends *only* on the 't Hooft anomaly of $G_{\rm IR}$ and on the choice of τ and $\hat{Q}_{\rm IR}$; see, for instance, the example below. We write $\alpha = \alpha(\hat{Q}_{\rm IR}|\tau)$ (the dependence on the 't Hooft anomaly is kept implicit).

Now, the key point is that the processes described in the two paragraphs above are in fact the *same* process, just described in two different ways. Therefore, we must equate

$$\nu = \frac{\alpha(\hat{Q}_{\rm IR}|\tau)}{2\pi} \quad ({\rm mod}\ 1). \tag{5}$$

1. Example: Luttinger liquid in one spatial dimension

We can apply the general framework described above to the particular example of a Luttinger liquid in one spatial dimension, as discussed in Sec. II A. In this case, $\hat{Q}_{\rm IR}$ and τ are defined by the right-hand sides of Eqs. (1) and (2).

Threading the flux of $\hat{Q}_{\rm IR}$ generates an electric field of $\hat{Q}_{\rm IR}$ by Faraday's law, and then, from the anomaly equation

Eq. (3), we see that the 2π flux threading creates -1 charge of ${\rm U}(1)_L$ and $+1$ charge of ${\rm U}(1)_R$. Therefore, in light of Eq. (2), the momentum transforms according to

$$e^{-ip} \rightarrow e^{-ip}e^{-i(k_R - k_L)}. \tag{6}$$

Then, from Eq. (5) we find that

$$\nu = \frac{1}{2\pi}(k_R - k_L) \quad ({\rm mod}\ 1), \tag{7}$$

which is nothing other than Luttinger's theorem for a Luttinger liquid in one spatial dimension.

B. Two spatial dimensions

Here the idea is to consider a "2π flux" of the microscopic U(1) symmetry; that is, a very weak background magnetic field spread out over some very large region, such that the total flux is 2π. We can then consider how such a 2π flux transforms under translation symmetry. We claim that, in the presence of fractional filling ν, such 2π fluxes exhibit translational symmetry fractionalization; that is, acting on a 2π flux, the lattice x and y translations $\mathbb{T}_x$ and $\mathbb{T}_y$ obey (in the limit as the 2π flux becomes infinitely spread out spatially) the magnetic algebra

$$\mathbb{T}_x \mathbb{T}_y \mathbb{T}_x^{-1} \mathbb{T}_y^{-1} = e^{2\pi i\nu}. \tag{8}$$

Heuristically, this is clear because a 2π flux sees a background charge density as an effective magnetic field. We give some more careful arguments in Appendix B. This result is also closely connected to the translational symmetry fractionalization of a monopole in an insulator in three spatial dimensions in the presence of polarization, as discussed in Ref. [20].

Meanwhile, in the IR theory, we can consider 2π flux configurations of the IR symmetry generated by $\hat{Q}_{\rm IR}$ that corresponds to the microscopic U(1) symmetry. The homomorphism φ maps $\mathbb{T}_x$ and $\mathbb{T}_y$ into some elements $\tau_x, \tau_y \in G_{\rm IR}$. In the presence of a 't Hooft anomaly, such 2π fluxes indeed can carry a projective representation, i.e.,

$$V(\tau_x)V(\tau_y)V(\tau_x)^{-1}V(\tau_y)^{-1} = e^{i\alpha}, \tag{9}$$

where $V(g)$ denotes the action of a group element $g \in G_{\rm IR}$ on the 2π flux, and $\alpha = \alpha(\hat{Q}_{\rm IR}|\tau_x, \tau_y)$ (defined mod 2π) depends on $\hat{Q}_{\rm IR}$, τ_x, and τ_y, and on the 't Hooft anomaly of $G_{\rm IR}$. Therefore, we must identify

$$\nu = \frac{\alpha(\hat{Q}_{\rm IR}|\tau_x, \tau_y)}{2\pi} \quad ({\rm mod}\ 1). \tag{10}$$

The natural example to consider to illustrate this constraint is a Fermi liquid. However, by contrast to the one-dimensional case, the 't Hooft anomaly of a Fermi liquid

in two spatial dimensions has not previously been discussed. This is the subject of Sec. V.

C. General space dimension

The formulation of filling constraints discussed above can be generalized to arbitrary space dimension d. The idea is to generalize the functions $\alpha(\hat{Q}_{IR}|\tau)$ (in $d=1$) and $\alpha(\hat{Q}_{IR}|\tau_x, \tau_y)$ (in $d=2$) to a function $\alpha(\hat{Q}_{IR}|\tau_1, ..., \tau_d)$ that determines the filling in general spatial dimension d. A convenient way to express this function is in terms of the "topological action" that describes the SPT phase in $d+1$ spatial dimensions which cancels the G_{IR} 't Hooft anomaly in d spatial dimensions by inflow on the boundary. We give the details in Appendix C.

IV. CONSEQUENCES OF THE FILLING CONSTRAINTS FOR COMPRESSIBLE STATES

The systematic theory of filling constraints described in the previous section has a very important corollary. We want to consider IR theories which can exist at *generic* filling; that is, they are not pinned to a particular filling but instead the filling can be continuously tuned. In other words, the IR theory represents a "compressible" state. What we show is that in this case, for spatial dimensions ≥ 2, the emergent symmetry G_{IR} *cannot* be a compact Lie group. In one dimension, recall that the Luttinger-liquid example discussed in Sec. II A achieves generic filling with only a compact Lie group emergent symmetry $G_{IR} = U(1) \times U(1)$. Note that compact Lie groups include finite groups as special cases, since we do not require the Lie group to be connected; also, in this paper when we refer to Lie groups, we always assume they are finite dimensional.

We wish to emphasize here that as we stated previously, in this paper when we refer to the emergent symmetry group G_{IR}, we are referring specifically to the *internal* symmetries of the IR theory. While translation symmetry is never a compact group, we do not know of any examples of theories that can sensibly arise in condensed matter systems, for which the *internal* symmetry is a noncompact Lie group, and this may, in fact, be impossible. Therefore, our results suggest that the emergent symmetry for IR theories that exist at generic filling must be an infinite-dimensional group [66]. Indeed, this is the case for Fermi liquids, as we describe in the next section.

Our main result is the following:

Theorem 1: Suppose the emergent symmetry group of the IR theory is some compact Lie group G_{IR}. Then, for any spatial dimension $d \geq 2$, the filling ν is constrained to be an integer multiple of $1/N_{G_{IR}}$, for some finite integer $N_{G_{IR}}$ that depends only on the group G_{IR} and the dimension.

The proof is straightforward based on the framework of filling constraints discussed in the previous section. We give the details and reveal what determines $N_{G_{IR}}$ in

Appendix F, and we also give a more formal point of view in Appendix G.

Let us mention a simple way to understand this result for $d=2$. As we mention in the Introduction, the 't Hooft anomaly of a theory in d spatial dimensions implies that the theory can be realized as the boundary of a SPT phase protected by the same symmetry G_{IR} in $(d+1)$-space dimensions. Now suppose that the SPT phase is such that n copies of it are trivial, with n a finite positive integer. For SPT phases protected by a compact Lie group G_{IR} in space dimension $d+1=3$, this is known to be always true. For the d-dimensional boundary theory of interest, this means that for n copies, there is no 't Hooft anomaly. If now we consider n copies of the microscopic lattice system, we see that it has a total filling $n\nu$. Since at this filling the IR theory has no anomaly, it follows that $\nu n = p$ with p an integer, which is essentially the claim of Theorem 1. In $d=3$, this simple argument does not work because there are SPT phases in $d+1=4$ space dimensions such that there is no finite n for which n copies become trivial. (This is also why the argument does not work for $d=1$.) However, Theorem 1 still holds for any $d \geq 2$ as we show in Appendixes F and G.

V. THE KINEMATIC PROPERTIES OF FERMI LIQUIDS

In this section, we return to a familiar IR theory: a Fermi liquid in two spatial dimensions (we briefly discuss higher dimensions as well) and analyze its kinematic properties in the language introduced previously. We see how Fermi liquids, by virtue of having an emergent symmetry group that is "larger" than a compact Lie group, are able to evade the theorem of the previous section and exist at generic filling.

A. Emergent symmetry group

The first step is to identify the emergent symmetry group G_{IR}. We invoke the following well-known property of Fermi liquids: Nonforward scattering terms are irrelevant in the renormalization group sense at low energies, so the quasiparticle number at *each point* on the Fermi surface is separately conserved at low energies. Thus, Fermi liquids have a very large emergent symmetry group [67]. Roughly, we can say that $G_{IR} = "U(1)^\infty$." However, let us be a bit more precise about how one approaches the "∞."

We parametrize the Fermi surface by a continuous parameter θ (we do not require that θ literally represents a geometrical angle), which is periodic; i.e., it lives on a circle. Imagine that we place an IR cutoff on the system (that is, place it in finite volume), which since the Fermi surface exists in momentum space, corresponds to a short-distance cutoff on θ; i.e., θ now takes discrete values. To each such θ value, we associate a U(1) emergent symmetry generated by an integer-valued operator $\hat{N}_\theta$. Hence, a general symmetry operator takes the form

$$\exp\left(-i\sum_\theta f_\theta \hat{N}_\theta\right), \tag{11}$$

where we identify $f_\theta \sim f_\theta + 2\pi$.

Now we send the IR cutoff (the spatial volume) to infinity, which corresponds to sending the spacing between discrete θ values to zero. What we want is to consider symmetry operators that do not depend too sensitively on the precise way in which the short-distance cutoff in θ gets sent to zero. In order to achieve this, we require that, in this continuum limit, the f_θ parameters become *smooth* functions $f(\theta)$. Therefore, in the limit, the emergent symmetries are in one-to-one correspondence with smooth functions from the circle into U(1). The group of all such functions is called the *loop group* [68] of U(1), and we denote it by LU(1). Hence, we conclude that $G_{\text{IR}} = \text{LU}(1)$.

We emphasize that the group structure of LU(1) differs somewhat from naive conceptions of what U(1)$^\infty$ would mean; in particular, LU(1) has only *one* U(1) subgroup, whose elements correspond to taking $f(\theta)$ to be a constant function. Physically, this is because only the *total* charge on the Fermi surface is quantized to be an integer; there is no well-defined concept of the (quantized) charge at a single point on the Fermi surface, only of the linear charge density with respect to θ. Accordingly, we can represent the loop group LU(1) formally by introducing a density operator $\hat{n}(\theta)$ such that the number of quasiparticles between θ and $\theta + d\theta$ is measured by $\hat{n}(\theta)d\theta$. Technically, $\hat{n}(\theta)$ is not really an operator in itself, but an operator-valued distribution which should be integrated against a test function. The elements of the emergent symmetry group can be expressed as

$$\exp\left(-i\int f(\theta)\hat{n}(\theta)\right), \tag{12}$$

where $f(\theta)$ is any smooth function of θ. Note that, because we identify $f(\theta) \sim f(\theta) + 2\pi$, we are allowed to consider functions f with nontrivial winding number around the circle, such that $(1/2\pi)\int \partial_\theta f(\theta)d\theta$ is any integer.

B. The homomorphism $G_{\text{UV}} \to G_{\text{IR}}$

Now it should be clear how to embed the microscopic symmetries into the emergent symmetry group $G_{\text{IR}} = \text{LU}(1)$. Indeed, if $\hat{Q}$ is the generator of the microscopic U(1) symmetry, we have

$$\hat{Q} \sim q \int \hat{n}(\theta)d\theta, \tag{13}$$

where the integer q is the charge of a Landau quasiparticle. Of course, for a Fermi liquid of electrons, $q = 1$, but in principle, one can imagine Fermi-liquid-like states where the quasiparticles carry a different charge. (For example, the quasiparticles could be bound states of an odd number of electrons.)

Meanwhile, if $k_x(\theta)$ and $k_y(\theta)$ represent the components of the lattice momentum of the point on the Fermi surface parametrized by θ, then we have

$$\mathbb{T}_\alpha \sim \exp\left(-i\int k_\alpha(\theta)\hat{n}(\theta)d\theta\right), \tag{14}$$

where $\alpha = x$, y, and $\mathbb{T}_x$, $\mathbb{T}_y$ are the lattice translation operators.

C. The 't Hooft anomaly

Now we are in a position to discuss the 't Hooft anomaly for the emergent LU(1) symmetry. As usual, the 't Hooft anomaly can be understood by coupling to a background gauge field for the symmetry. But first, we must ask, what is a gauge field for an LU(1) symmetry? Since, roughly speaking, an LU(1) symmetry means there is a U(1) symmetry for each point on the circle, we can naively say that a gauge field for an LU(1) symmetry should be a space-time vector field $A_\mu(\theta)$ for each point θ on the circle, with gauge transformations parametrized by a scalar field $\lambda(\theta)$, and gauge transformation

$$A_\mu(\theta) \to A_\mu(\theta) + \partial_\mu\lambda(\theta) \tag{15}$$

(where the derivative is respect to space-time, not θ). Moreover, we require that $A_\mu(\theta)$ and $\lambda(\theta)$ be smooth functions of θ.

However, there is in fact an additional ingredient that is required that is unique to loop groups. To see this, we can appeal to the quasiparticle picture of the Fermi liquid. The spatial components $A_i(\theta)$ describe the Aharanov-Bohm phase (which can be interpreted as a Berry phase) picked up when a localized quasiparticle, localized near position θ on the Fermi surface, is transported in space. But we can also keep a quasiparticle fixed in real space and transport it along the Fermi surface in momentum space. Therefore, the gauge field needs an additional component A_θ to describe the Berry phase associated with such a process. We emphasize though, that even going beyond Fermi liquids, A_θ represents an intrinsic part of what it means to couple to a gauge field for an LU(1) symmetry, regardless of any quasiparticle picture. We return to this point in Sec. VI.

The A_θ component of the gauge field transforms under gauge transformations as $A_\theta \to A_\theta + \partial_\theta\lambda$. Therefore, if we now combine the vector field A_μ with A_θ, we obtain a vector field A on the $(D+1)$-dimensional manifold $M \times S^1$, where M is the space-time manifold, $D = \dim M$ is the space-time dimension, and S^1 is the circle on which the θ variable lives. Moreover, A transforms under gauge transformations precisely as would a U(1) gauge field on $M \times S^1$. Hence, we arrive at our conclusion: An LU(1)

gauge field on M is equivalent to a U(1) gauge field on $M \times S^1$. From this point on, we denote this combined vector field as A_μ, taking the convention that indices such as μ vary both over space-time directions and the θ direction.

Now we can discuss 't Hooft anomalies. One way to characterize a 't Hooft anomaly in two spatial dimensions is in terms of the topological term describing the response of the corresponding SPT phase in three spatial dimensions to background gauge fields. [As usual, the relation is that on a $(3 + 1)$D space-time M_+ whose boundary is a $(2 + 1)$D manifold M, the SPT phase on M_+ gives rise to the 't Hooft anomaly on the space-time M by anomaly inflow.] But from the above discussion, a topological term for LU(1) gauge fields on a $(3 + 1)$D space-time M_+ is equivalent to a topological term for U(1) gauge fields on the $(3 + 1 + 1)$D manifold $M_+ \times S^1$. Hence, we conclude that the appropriate topological term is the 5D Chern-Simons action

$$S[A] = \frac{m}{24\pi^2} \int_{M_+ \times S^1} A \wedge dA \wedge dA, \qquad (16)$$

where the level m is quantized to be an integer [69]. Below, we show that such a topological term, with $m = \pm 1$, indeed reproduces many known properties of spinless Fermi liquids. (Note that the sign of m is defined only relative to a choice of orientation for the Fermi surface, because redefining $\theta \to -\theta$ sends $m \to -m$.)

From the topological term Eq. (16), we can determine the anomaly equation on the boundary by computing the current $j = (\delta S/\delta A)$ and then considering the current incident onto the boundary. We find that, on the boundary, the continuity equation is violated according to

$$\partial_\mu j^\mu = \frac{m}{8\pi^2} \epsilon^{\lambda\sigma\tau\kappa} (\partial_\lambda A_\sigma)(\partial_\tau A_\kappa). \qquad (17)$$

Here the current j depends both on space-time coordinates and on θ. Its space-time components describe the space-time current of the charge at position θ on the Fermi surface, i.e., of the symmetry generated by $\hat{n}(\theta)$. However, recalling that indices are supposed to vary over the θ direction as well as space-time directions, we have to introduce the component $j^\theta = (\delta S/\delta A_\theta)$, which describes flow of charge *along* the Fermi surface. An example of a case where $j^\theta \neq 0$ is a Fermi liquid in a magnetic field described in the next subsection.

Finally, let us mention that there is an alternative picture to understand the inflow of the 't Hooft anomaly that is sometimes helpful. Instead of considering a system with LU(1) symmetry in a $(3 + 1)$D space-time M_+ with boundary, we can consider a system with L^{T^2}U(1) symmetry in a $(2 + 1)$D space-time M without boundary. Here, L^{T^2}U(1) is the group of smooth maps from the Brillouin zone (thought of as a torus $\mathbb{T}^2$) into U(1). In other words,

we imagine that the charge is conserved not just at each point on the Fermi surface, but also at *every* k point in the whole Brillouin zone. This is the case, for example, in a noninteracting Fermi gas. In an interacting Fermi liquid, the physical Hamiltonian presumably has a nonzero scattering rate for quasiparticles in the interior of the Fermi surface, but we can, *theoretically*, imagine an extension of the quasiparticle Hamiltonian from the vicinity of the Fermi surface to its interior in a manner that preserves conservation of quasiparticle number at every k point. This is a familiar construction in Fermi-liquid theory. (The point is that the physics on the Fermi surface will ultimately not depend on the precise form of the Hamiltonian in the interior.)

Then by similar arguments to before, a gauge field for the L^{T^2}U(1) symmetry is equivalent to a U(1) gauge field on $M \times \mathbb{T}^2$. Since this is also a five-dimensional manifold, we can write a similar Chern-Simons term to Eq. (16). Specifically, we write

$$S[A] = \frac{m}{24\pi^2} \int_{M \times \mathcal{D}} A \wedge dA \wedge dA, \qquad (18)$$

where $\mathcal{D} \subseteq \mathbb{T}^2$ is the volume "occupied by electrons," that is, the volume enclosed by the Fermi surface [70]. This gives rise to the same anomaly Eq. (17) on the boundary $\partial(M \times \mathcal{D}) = M \times \partial\mathcal{D}$, where $\partial\mathcal{D}$ is the Fermi surface. Note that in this interpretation, the components A_0, A_x, A_y of A_μ are the usual ones, while the components A_{k_x}, A_{k_y} are k-space gauge fields. Then the five-dimensional manifold has the interpretation of being "phase space" (that is, the space where points are labeled by position and momentum) plus time, and A_μ can be considered a gauge field in phase space. A term similar to Eq. (18) has previously appeared in Ref. [40].

D. Filling constraint and Luttinger's theorem

We consider a theory in two spatial dimensions whose 't Hooft anomaly is canceled by inflow from Eq. (16). According to the discussion of Sec. III B, we first need to determine how the emergent symmetry LU(1) gets represented projectively in such a theory in the presence of a 2π flux [of the microscopic charge U(1) symmetry]. This is something that can be derived from the 't Hooft anomaly that is characterized in the previous subsection. We show in Appendix D that this leads to a projective representation on a 2π flux described by the commutation relations

$$[\hat{n}(\theta), \hat{n}(\theta')] = -i\frac{mq}{2\pi}\delta'(\theta - \theta'), \qquad (19)$$

where $\hat{n}(\theta)$ are the operators introduced in Sec. V A, and δ' is the derivative of the Dirac delta function.

Note that Eq. (19) is precisely the so-called Kac-Moody algebra satisfied by the local density operator of a chiral fermion in one spatial dimension (such as appears, for example, at the boundary of a integer quantum Hall state in two spatial dimensions). Indeed, this relationship is not a coincidence; one can see roughly how it comes about in the case of a Fermi liquid at the level of the semiclassical theory of electron transport. Suppose, for simplicity, that we switch off the interactions so that the Fermi liquid becomes a noninteracting Fermi gas. In the presence of a spatially dependent magnetic field $\mathbf{B}(\mathbf{x})$, the semiclassical equations of motion take the form [71]

$$\frac{d\mathbf{k}}{dt} = -q\mathbf{B}(\mathbf{x}) \times \mathbf{v}(\mathbf{k}), \tag{20}$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{k}) - \Omega(\mathbf{k}) \times \frac{d\mathbf{k}}{dt}, \tag{21}$$

where $\Omega(\mathbf{k})$ is the Berry curvature of the Bloch states, $\mathbf{v}(\mathbf{k}) = [\partial\mathcal{E}(\mathbf{k})/\partial\mathbf{k}]$, $\mathcal{E}(\mathbf{k})$ is the dispersion relation, and we omit terms beyond linear order in the magnetic field strength. In particular, Eq. (20) implies that electrons develop a circulation in momentum space along contours of constant energy, in particular, along the Fermi surface. In two spatial dimensions, the Fermi surface is one dimensional, and the circulation along the Fermi surface is unidirectional (set by the sign of the magnetic field), which indeed resembles a chiral fermion.

This intuitive argument, however, does not fix the coefficient of the right-hand side of Eq. (19). In Appendix E, we give a more careful derivation of Eq. (19) from the semiclassical theory of electron transport, confirming that $m = \pm 1$ for a spinless Fermi liquid. Related expressions have previously been derived in Refs. [72,73].

From Eq. (19), we can compute the projective representation of the translation symmetry in light of the embedding Eq. (14). We find

$$\mathbb{T}_x\mathbb{T}_y\mathbb{T}_x^{-1}\mathbb{T}_y^{-1} = \exp\left(i\frac{mq}{2\pi}\int k_x(\theta)\frac{dk_y(\theta)}{d\theta}d\theta\right) \tag{22}$$

$$= \exp\left(imq\frac{\mathcal{V}_F}{2\pi}\right), \tag{23}$$

where $\mathcal{V}_F$ is the volume in momentum space enclosed by the Fermi surface. (Here we choose a particular convention to define the orientation of the Fermi surface). Hence, from Eq. (10), we conclude that

$$\nu = mq\frac{\mathcal{V}_F}{(2\pi)^2} \quad (\text{mod } 1), \tag{24}$$

which (if we set $m = q = 1$) is precisely Luttinger's theorem for a spinless Fermi liquid in two spatial dimensions.

We wish to emphasize, however, that, in general, Luttinger's theorem Eq. (24) follows directly from Eq. (19), which in turn follows directly from the 't Hooft anomaly. It is not necessary to assume anything about the dynamical properties of the Fermi liquid, e.g., the existence of quasiparticles. Thus, Luttinger's theorem also holds (with a possible integer multiplicative factor mq) for any IR theory that has the same emergent symmetry as a Fermi liquid.

It is interesting to reconsider Eq. (19) and its relation with the 5D Chern-Simons term from the viewpoint that the five-dimensional manifold can be thought of as four-dimensional phase space together with the time direction (see the last two paragraphs of Sec. V C). In the presence of a static 2π-strength magnetic flux in the x, y components of A_μ (with the corresponding components A_x and A_y independent of k_x, k_y, t), the 5D Chern-Simons term Eq. (18) reduces to a 3D Chern-Simons term for the remaining components $A_I = (A_0, A_{k_x}, A_{k_y})$, assuming that they are independent of x and y:

$$S_{3D} = \frac{m}{4\pi}\int d^3x\epsilon_{IJK}A_I\partial_J A_K, \tag{25}$$

where k_x and k_y are integrated over the volume $\mathcal{D}$, i.e., over the interior of the Fermi surface.

Now the claim is that on the Fermi surface, there is a chiral mode carrying the Kac-Moody algebra

$$[\hat{n}(\theta), \hat{n}(\theta')] = -i\frac{m}{2\pi}\delta'(\theta - \theta'). \tag{26}$$

In particular, since a 2π flux of the microscopic U(1) corresponds to a $2\pi q$ flux of A by Eq. (13), we recover Eq. (19).

In the phase-space interpretation, Eq. (25) describes an integer quantum Hall effect in *momentum space* in the interior of the Fermi surface. In other words, we think of the rigid interior of the Fermi surface as hosting an integer quantum Hall state in momentum space when we apply a 2π flux in real space. The Fermi surface is the boundary in momentum space of the interior, and hence hosts a chiral edge state.

E. Extension to higher dimensions

The description of the anomaly extends straightforwardly to higher-dimensional Fermi liquids. Indeed, if M is the $(d+1)$D space-time manifold, and F is a $(d-1)$-dimensional manifold parametrizing the Fermi surface, which is a codimension-one surface in the d-dimensional Brillouin zone, then the emergent symmetry group is $L^F U(1)$, i.e., the group of smooth maps from F to U(1). Including the components of the Berry connection on the Fermi surface promotes the $L^F U(1)$ gauge field to a U(1) gauge field on the $(2d)$-dimensional manifold $M \times F$.

Then we can write down a topological term on a $(d+2)$D space-time M_+ describing the SPT phase whose inflow generates the anomaly of the Fermi liquid given by the $(2d+1)$-dimensional Chern-Simons action:

$$S[A] = \frac{m}{(d+1)!(2\pi)^d} \int_{M_+ \times F} A \wedge \underbrace{(dA \wedge \cdots \wedge dA)}_{d \text{ times}}. \quad (27)$$

VI. ERSATZ FERMI LIQUIDS AND THEIR PHENOMENOLOGY

As we show in Sec. IV, *any* IR theory that can exist at generic filling must have a very large symmetry group, larger than any compact Lie group. In the simplest cases, this emergent symmetry group will be the same as that of a Fermi liquid, i.e., $G_{IR} = LU(1)$. We refer to an IR theory with this emergent symmetry as an EFL. We note that EFLs represent a *class* of theories; there could be many different EFLs with different dynamical properties. Here we examine the properties that all EFLs have in common. In particular, as a Fermi liquid is an example of an EFL, this section provides a fresh perspective on many aspects of Fermi-liquid phenomenology, showing that they arise directly from the emergent symmetry and its 't Hooft anomaly without any need to invoke the detailed dynamical properties of a Fermi liquid.

Going beyond Fermi liquids, an important example of states, which we expect are EFLs, are associated with quantum critical points in metals which are not tied to a particular electron density. As a concrete example, consider a putative quantum critical point associated with the onset of Ising nematic order [74] in a metal in $d = 2$. There is no particular electron density at which this transition will happen in any given system. Indeed, as microscopic parameters are changed continuously, we expect that the electron density at the transition will also change continuously without a change of universality class. The universal critical properties of this transition are described by a theory of electrons near the Fermi surface coupled to the fluctuating order parameter modes. [A model with similar structure also describes *insulating* quantum spin liquid phases with a spinon Fermi surface coupled to a dynamical U(1) gauge field.] In these models, the resulting IR fixed point is not a Fermi liquid [75–86]. Nevertheless, from our point of view, these metallic quantum critical points are expected to be EFLs. (We discuss neutral Fermi surfaces in insulators separately in Sec. VIII C.)

We may understand why such critical points should be EFLs as follows. In previous papers [78,81–83], the ultimate IR fixed point was accessed through a "patch construction," which begins by dividing the Fermi surface into small patches. It was then argued that the important coupling of the fermions within a single patch is to boson fluctuations with momentum tangential to the local Fermi surface. This enables treating the full system by focusing on

a pair of antipodal patches and ignoring their coupling to other such pairs of patches. The patch width is taken to zero at the end. In this scheme, the number of fermions within each patch is conserved. The assumption that this patch description captures the IR fixed point then implies that the linear charge density $\hat{n}(\theta)$ at each point of the Fermi surface is conserved at the fixed point. (There is, potentially, a dangerous interpatch BCS coupling in the pairing channel that could destroy these conservation laws. For the Fermi surface coupled to a gauge field, a weak pairing interaction is irrelevant at the IR fixed point [87], while in the Ising nematic quantum critical point, it is relevant. In the latter case, the non-Fermi-liquid metallic fixed point is pre-empted by the superconducting instability. Our discussion then applies at a scale above this instability.)

We also expect that more complex quantum critical points associated with the death of a Fermi surface can be subsumed under the umbrella of EFLs. Such critical points have been argued [88] to possess a critical Fermi surface even in the absence of Landau quasiparticles. As we see below, the EFL description, if it indeed applies to such quantum critical points, enables us to infer many of their general properties.

A. General properties of EFLs

Let us return to a general EFL. Then, translation symmetry must embed into G_{IR} somehow: That is, we have

$$\mathbb{T}_\alpha \sim \exp\left(-i \int k_\alpha(\theta)\hat{n}(\theta)d\theta\right) \quad (28)$$

for some U(1)-valued functions $k_\alpha(\theta)$. The values of $k_\alpha(\theta)$ can be interpreted as momenta in the Brillouin zone, so this defines a codimension-one surface in the Brillouin zone. We can take this to be the *definition* of a Fermi surface in a general EFL. Of course, this "Fermi surface" may or may not have any signature in, say, angle-resolved photoemission spectroscopy (ARPES) measurements, and if it does, the signature might be different in character from that of a Fermi liquid.

Nevertheless, the fact that $\hat{n}(\theta)$ is a conserved operator for every θ tells us that, at any point θ on the Fermi surface, there are infinitely long-lived excitations which are forbidden from scattering away from that point. To see this, first observe that since $\hat{n}(\theta)$ is conserved, excitations can be labeled by their corresponding eigenvalue $n(\theta)$. An excitation at point θ on the Fermi surface is characterized by

$$n(\theta') = N\delta(\theta' - \theta) \quad (29)$$

for some N. One can show that N is quantized to an integer, because by the definition of LU(1), the operator

$$\hat{N} = \int \hat{n}(\theta)d\theta \quad (30)$$

must have integer eigenvalues, since $\exp(-2\pi i \hat{N})$ is the identity operator. Henceforth, we refer to such excitations as Fermi-surface quanta. From Eqs. (28) and (29), we see that a single Fermi-surface quantum carries momentum $\mathbf{k}(\theta)$. Fermi-surface quanta are the generalization of Landau quasiparticles to a general ersatz Fermi liquid. Their dynamics, however, could be very different from Landau quasiparticles. Moreover, apart from the fact that their number is quantized, the Fermi-surface quanta need not have particularly "quasiparticlelike" properties. Nevertheless, the Berry's phase of a spatially localized Fermi-surface quantum as it is moved in space or over the Fermi surface is still a well-defined quantity, which supplies the general interpretation of the gauge field A_μ (including the A_θ component) discussed in Sec. V C.

Next, we can consider how the microscopic U(1) symmetry embeds into LU(1). The quantization of charge, i.e., the requirement that $e^{i(2\pi)\hat{Q}} = 1$ (where $\hat{Q}$ is the microscopic charge operator), constrains the embedding to be of the form

$$\hat{Q} \sim q\hat{N} \qquad (31)$$

for some integer q. We can interpret q as the charge of a single Fermi-surface quantum.

Finally, we can consider the 't Hooft anomaly of the LU(1) symmetry. There is not much freedom, since as we see in Sec. V the 't Hooft anomalies are just classified by the level m of the 5D Chern-Simons term Eq. (16), which is quantized to be an integer. In summary, once we fix the shape of the Fermi surface, the kinematic properties of an ersatz Fermi liquid (in the sense of Sec. II) are captured by the integers q and m. [Note that there is a freedom that sends $q \to -q$ and $m \to -m$ simultaneously by redefining the generators of LU(1); moreover, as we mentioned earlier, choosing the reverse orientation of the Fermi surface sends $m \to -m$ while leaving q fixed.]

Next we observe that the arguments of Sec. V D can be applied in any EFL. Therefore, we immediately conclude that the Fermi surface in any EFL satisfies Luttinger's theorem in the form of Eq. (24).

In the remainder of this section, we consider various aspects of Fermi-liquid phenomenology and argue that they hold equally well in any EFL.

B. Response to electric fields

A property of Fermi liquids is that if a uniform electric field is applied, the system responds in essentially the same way as it would in a noninteracting Fermi gas, which is to say that the momenta of quasiparticles get shifted (assuming the electric field is in the x direction) according to $k_x \to k_x + a_x Et$, where a_x is the unit cell size in the x direction. [89] This causes the density operators $\hat{n}(\theta)$ to be no longer conserved in the presence of the electric field. Specializing for simplicity to a Fermi liquid in two spatial

dimensions, the total charge on a segment $[\theta, \theta + d\theta]$ of the Fermi surface gets shifted according to

$$\hat{n}(\theta)d\theta \to \hat{n}(\theta)d\theta + qa_x Et \frac{dk_y(\theta)}{(2\pi)^2/(L_x L_y)}, \qquad (32)$$

where $dk_y(\theta) = k_y(\theta + d\theta) - k_y(\theta)$, and L_x and L_y are the linear dimensions of the system (normalized by a_x and a_y, respectively, the unit cell dimensions) in the x and y directions. We have to divide by the denominator in the second term in Eq. (32) to take into account the density of single-particle states in momentum space in a finite-size system.

We now show that Eq. (32) indeed follows from the 't Hooft anomaly. In order to do that, we want to reformulate Eq. (32) in a way that removes the explicit dependence on L_y. We write Eq. (32) as

$$\frac{d}{dt}\hat{n}(\theta) = \eta(\theta; L_y) := qE\frac{a_x L_x L_y}{(2\pi)^2}\frac{dk_y(\theta)}{d\theta}. \qquad (33)$$

Next, we identify the *difference*

$$\Delta\eta(\theta) := \eta(\theta; L_y + 1) - \eta(\theta; L_y) \qquad (34)$$

$$= qE\frac{a_x L_x}{(2\pi)^2}\frac{dk_y(\theta)}{d\theta} \qquad (35)$$

with the density shift rate associated with applying a uniform electric field in the x direction in the presence of a "flux of y translation" symmetry around the y direction.

To compute this shift from the 't Hooft anomaly, we start with the anomaly Eq. (17) and consider a configuration where

$$A_x = Et, \qquad (36)$$

$$A_y = -k_y(\theta)/L_y, \qquad (37)$$

$$A_t = 0, \qquad (38)$$

$$A_\theta \text{ is independent of } x, y, t. \qquad (39)$$

The motivation for our choice of A_y is that it ensures that the LU(1) flux around the y direction is given by

$$\exp\left(-i\int k_y(\theta)\hat{n}(\theta)d\theta\right), \qquad (40)$$

which coincides with the expression Eq. (14) for the translation symmetry operator $\mathbb{T}_y$. Substituting into Eq. (17) and integrating over x and y gives

$$\partial_t n(\theta) + \partial_\theta J^\theta = \frac{mq}{(2\pi)^2} E \frac{dk_y(\theta)}{d\theta} a_x L_x, \qquad (41)$$

where $J^\theta = \int j^\theta dx dy$, and we can identify $\int j^t dx dy$ with the expectation value $n(\theta) = \langle \hat{n}(\theta) \rangle$. Hence, provided that $\partial_\theta J^\theta = 0$, we recover Eq. (35) if we set $m = 1$.

Our assumption that $\partial_\theta J^\theta = 0$, even in the presence of a background electric field, requires a bit more explanation. J^θ represents a charge circulation along the Fermi surface. As we mention in Sec. V D, in a Fermi liquid a magnetic field induces a chiral flow of quasiparticles along the Fermi surface, in which case $J_\theta \propto B n(\theta)$ and generally $\partial_\theta J_\theta \neq 0$. So we need to address why the same thing could not happen with an electric field. Of course, in a Fermi liquid, one can easily convince oneself that it does not, but we want an argument that holds more generally in any EFL.

In a quantized formulation of the IR theory, the current operator $\hat{J}_\theta$ is defined by

$$\hat{J}_\theta = \int \frac{\delta \hat{H}}{\delta A_\theta} d^2 x, \qquad (42)$$

where $\hat{H}$ is the Hamiltonian of the IR theory. Then we know that, in the absence of an electromagnetic field, $\hat{J}_\theta$ must satisfy (as an operator identity) the conservation law

$$\frac{d}{dt} \hat{n}(\theta, t) + \partial_\theta \hat{J}_\theta(\theta, t) = 0. \qquad (43)$$

On the other hand, in the absence of an external electromagnetic field, by assumption $\hat{n}(\theta)$ should be conserved for each θ, since it is the generator of the emergent symmetry. Therefore, in the absence of electromagnetic field, the current operator identically satisfies

$$\partial_\theta \hat{J}_\theta = 0. \qquad (44)$$

Now observe that because the electric field E enters into the Hamiltonian "temporally," i.e., through the time derivative of A, it is not possible for $\hat{J}_\theta$ defined by Eq. (42) to depend on E because the Hamiltonian is defined on a single time slice of space-time. Therefore, $\hat{J}_\theta$ continues to satisfy Eq. (44) even with an applied electric field. On the other hand, it *is* possible for $\hat{J}_\theta$ to depend on the applied magnetic field, as the Fermi-liquid example demonstrates.

C. Quantum oscillations

Recall that Fermi liquids display "quantum oscillations" when a weak magnetic field B is applied; that is, various physical properties are periodic in $1/B$ [90]. In the case where the physical property under consideration is resistivity, for example, this is known as the Shubnikov–de Haas effect. (What we mean by "weak magnetic field" is that the magnetic flux per unit cell should be much less than 1.

To observe the oscillations at finite temperature T, it is also necessary that the cyclotron energy E_c should satisfy $E_c \gtrsim T$, which for $T > 0$ also places a lower bound on the magnetic fields for which the oscillations are observable.)

It is sometimes stated that observing quantum oscillations is evidence for a quasiparticle description, i.e., of a Fermi liquid. Here, however, we show that there are very general reasons to expect any EFL to display the same periodicity of the quantum oscillations. We do not, however, make any statement about the *amplitude* of the quantum oscillations, and it remains possible that this amplitude and its dependence on parameters such as temperature will still allow for Fermi liquids and non-Fermi liquids to be distinguished.

Our task is complicated by the fact that the quantum oscillations are nonperturbative in the magnetic field, and therefore, it is not clear that one expects them to be directly describable in terms of the deep IR theory. Instead, we find it necessary to appeal to a UV completion.

1. Two spatial dimensions

Let us first consider the case of a system in two spatial dimensions which microscopically has continuous translation symmetry, with a microscopic charge density ρ. In such a case, in the presence of a magnetic field the x and y translation generators P_x and P_y fail to commute, but we can define a "magnetic unit cell" of volume $b_x \times b_y$, such that the flux per magnetic unit cell is 2π and the discrete translations $\mathbb{T}_x = \exp(-ib_x P_x)$ and $\mathbb{T}_y = \exp(-ib_y P_y)$ do commute. Then we can define the magnetic filling $\nu_M = (b_x b_y)\rho = 2\pi\rho/B$, which is the charge per magnetic unit cell. Next, we can apply all the usual results on filling with respect to this discrete translation symmetry; in particular, if ν_M is not an integer, then the Lieb-Schultz-Mattis-Oshikawa-Hastings theorem [1–4] forbids the system from having a trivial (i.e., not topologically ordered) gapped ground state, whereas such a state is permitted for integer ν_M. Moreover, if we assume that the ground state at a given ν_M is itself an EFL with a single Fermi surface, then Luttinger's theorem for EFLs implies that at integer ν_M the Fermi volume must fill all of the Brillouin zone corresponding to the magnetic unit cell; i.e., the Fermi surface becomes degenerate, presumably leading to an instability. Most generally, the point is that since $\nu_M \bmod 1$ always reflects properties of the IR theory according to our general framework of filling constraints, the nature of the IR theory must vary with $\nu_M \bmod 1$.

These considerations motivate our assumption that, in general, observable properties of the ground state vary periodically with ν_M, with period 1. (The periodicity refers to the behavior for $\nu_M \gg 1$, and where ν_M varies over an interval $\Delta\nu_M \ll \nu_M$. On longer scales, there will be some envelope function governing the amplitude of the oscillations.) This behavior is well known in the Fermi-liquid

case, where integer ν_M (noninteger ν_M) corresponds to fully filled (partially filled) Landau levels, respectively. Our periodicity assumption immediately implies that, if we keep ρ fixed and vary B, then the periodicity with respect to $1/B$ is

$$\Delta(1/B) = 1/(2\pi\rho). \qquad (45)$$

Next we can consider a system in two spatial dimensions that microscopically has only discrete translation symmetry. Here we recall that, even for Fermi liquids, there is a condition on the Fermi surface in order to observe quantum oscillations. For a Fermi surface as depicted in Fig. 1, no quantum oscillations will be observed; physically, one can think of this as coming from the fact that the semiclassical orbits of this Fermi surface in a magnetic field are not closed in position space. The general condition for a Fermi liquid to display quantum oscillations is that each connected component of the Fermi surface should have trivial winding number on the Brillouin torus.

Let us therefore examine what we can say about a general EFL that satisfies this condition. To avoid some subtleties in the argument, we assume that $q = 1$. The key point is that when the Fermi surface does not have nontrivial winding number, then there is a consistent way to define the Fermi momentum $\mathbf{k}(\theta)$ without any mod 2π ambiguity. Then we can define the filling that is relevant for quantum oscillations as

$$\nu_o = \frac{m}{(2\pi)^2}\mathcal{V}_F, \qquad \mathcal{V}_F = \int k_x(\theta)\frac{dk_y(\theta)}{d\theta}d\theta, \qquad (46)$$

where $\mathcal{V}_F$ is the volume enclosed by the Fermi surface, and this formula holds without any mod 1 equivalence. Note that ν_o is not necessarily the same as the microscopic filling ν, although by Luttinger's theorem it must differ from it by an integer. Moreover, we can define emergent symmetry generators
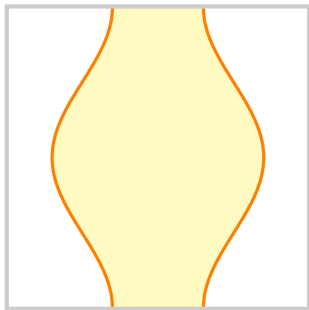

FIG. 1. A Fermi liquid with a Fermi surface that wraps nontrivially around the Brillouin torus does not exhibit quantum oscillations.

$$\hat{P}_\alpha = a_\alpha^{-1}\int k_\alpha(\theta)\hat{n}(\theta) \qquad (47)$$

with a_x, a_y the dimensions of the unit cell, such that $\exp(-ia_\alpha P_\alpha) \sim \mathbb{T}_\alpha$.

This leads us to believe (although we do not give a rigorous proof) that, even though the actual microscopic theory may have had only a discrete translation symmetry, there *exists* a UV completion of the IR theory with continuous translation symmetry whose action on the IR theory is generated by $\hat{P}_\alpha$. Assuming that this is the case, invoking the continuous translations version of Luttinger's theorem shows that the density of U(1) charge in this UV completion must be $\rho = \nu_o/(a_x a_y)$. Then with ρ so defined, we again find that the periodicity of quantum oscillations is given by Eq. (45).

2. Two spatial dimensions: Multiple Fermi surfaces

As a warm-up to going to three spatial dimensions, we extend the results of the previous section to the case where we have several disconnected components of the Fermi surface, instead of just one. For simplicity, we consider the case of two components, although the arguments can easily be generalized. In that case, the emergent symmetry is $L^F U(1)$, i.e., the group of smooth maps from F into U(1), where $F = S^1 \sqcup S^1$, i.e., the disjoint union of two circles. We can represent this by writing the generators of the emergent symmetry as $\hat{n}^{(\lambda)}(\theta)$, where $\theta \in S^1$ and $\lambda \in \{1, 2\}$ labels the two components of the Fermi surface. Similarly, we write the momentum of the Fermi surface as $\mathbf{k}^{(\lambda)}(\theta)$. We assume that each component satisfies the condition of no winding described earlier, which allows us to define the momentum operator

$$\hat{P}_\alpha = \sum_\lambda a_\alpha^{-1}\int k_\alpha^{(\lambda)}(\theta)\hat{n}^{(\lambda)}(\theta)d\theta. \qquad (48)$$

The crucial point is the fact that there are two connected components of the Fermi surface allows us to define two different U(1) emergent symmetries, generated by $\hat{Q}^{(1)}$ and $\hat{Q}^{(2)}$, with

$$\hat{Q}^{(\lambda)} = \int \hat{n}^{(\lambda)}(\theta)d\theta. \qquad (49)$$

Therefore, in the same spirit as the argument of the previous section, we postulate that there is a UV completion with both continuous translation symmetry and $\hat{Q}^{(1)}$ and $\hat{Q}^{(2)}$ realized microscopically. Defining $\rho^{(1)}$ and $\rho^{(2)}$ to be the corresponding charge densities, by similar arguments as earlier we have a generalized Luttinger's theorem

$$\rho^{(\lambda)} = \frac{m}{(2\pi)^2 a_x a_y} \mathcal{V}_F^{(\lambda)}, \qquad \mathcal{V}_F^{(\lambda)} = \int k_x^{(\lambda)}(\theta) k_y^{(\lambda)}(\theta) d\theta.$$

$$(50)$$

Now, in the presence of magnetic field, if we define magnetic fillings $\nu_M^{(\lambda)} = 2\pi\rho^{(\lambda)}/B$, we expect that any observable property should be periodic in each of $\nu_M^{(1)}$ and $\nu_M^{(2)}$ with period 1. In other words, any observable property O can be expressed as

$$O = f(\nu_M^{(1)}, \nu_M^{(2)}),$$

$$(51)$$

where $f(\nu_1, \nu_2)$ is some function such that $f(\nu_1 + 1, \nu_2) = f(\nu_1, \nu_2 + 1) = f(\nu_1, \nu_2)$. This implies that if we vary B while keeping $\rho^{(\lambda)}$ fixed, and assuming that the ratio $\rho^{(1)}/\rho^{(2)}$ is an irrational number, then O varies *quasiperiodically* in $1/B$ with base frequencies $2\pi\rho^{(1)}$ and $2\pi\rho^{(2)}$.

3. Three spatial dimensions

Now we can consider a 3D system. By dimensional reduction, we can consider a 3D system as a 2D system, but the Fermi surface of the 2D system will consist of infinitely many components corresponding to taking slices through the 3D Fermi surface with fixed k_z. By generalizing the previous discussion to N components, and then taking the limit as $N \to \infty$, we find that any observable property should be expressible as

$$O = \mathcal{F}[\nu_\perp],$$

$$(52)$$

where $\mathcal{F}$ is some functional of $\nu_\perp$, and $\nu_\perp$ is a function into $\mathbb{R}/\mathbb{Z}$ (i.e., the real line with $\nu \sim \nu + 1$ identified) defined by

$$\nu_\perp(k_z) = 2\pi\rho_\perp(k_z)\frac{1}{B} \quad \text{mod } 1,$$

$$(53)$$

$$\rho_\perp(k_z) = \frac{m\mathcal{V}_\perp(k_z)}{(2\pi)^2 a_x a_y},$$

$$(54)$$

where $\mathcal{V}_\perp(k_z)$ is the two-dimensional area enclosed by the intersection of the Fermi surface with a plane of fixed k_z.

We show in Appendix H that this implies, so long as the functional $\mathcal{F}$ is sufficiently regular, that the dependence of O on B at small B is dominated by the extremal Fermi-surface cross sections, and we obtain that O varies periodically or quasiperiodically with $1/B$, with base frequencies $2\pi\rho_\perp(k_z^*)$, $k_z^* \in \Sigma$, where Σ is the set of solutions to $(d/dk_z)\mathcal{V}_\perp(k_z) = 0$. This is the same result as for a Fermi liquid.

D. Anomalous Hall effect

The anomalous Hall effect refers to a Hall conductance that exists in a time-reversal broken system in zero external magnetic field. In general, there can be many different effects that contribute to this effect. For a clean free-electron system with a Fermi surface, there is an interesting contribution that arises due to the net Berry curvature Φ_B of the filled Fermi sea. In $d = 2$, we have

$$\sigma_{xy} = \frac{\Phi_B}{2\pi}.$$

$$(55)$$

This contribution can alternately be reexpressed as a Fermi-surface property in terms of the Berry gauge connection:

$$\Phi_B = \int d\theta A_\theta.$$

$$(56)$$

Note that the integrand (as opposed to the integral) on the right-hand side—while involving only the Fermi surface—is not by itself invariant under k-space gauge transformations. Going beyond noninteracting systems, for an interacting Fermi liquid, it has been argued [91] that there will be additional Fermi-surface contributions to the anomalous Hall conductance that, by contrast to the Berry phase contribution above, involve an integral over gauge-invariant quantities defined in local patches of the Fermi surface.

Next we show that the 't Hooft anomaly of an EFL directly implies the Berry phase contribution to the anomalous Hall effect. To that end, we start from Eq. (18) and consider a configuration where the gauge field A has a flux Φ_B through the interior of the Fermi surface in the (k_x, k_y) plane, and take the corresponding gauge field components (A_{k_x}, A_{k_y}) to be independent of (t, x, y). Similar to the discussion surrounding Eq. (25), we now find that the 5D Chern-Simons term reduces (assuming that A_0, A_x, A_y are independent of k_x and k_y) to

$$S_{\text{AH}} = \frac{m\Phi_B}{8\pi^2} \int dt dx dy \epsilon_{IJK} A_I \partial_J A_K$$

$$(57)$$

for the $I, J, K = (0, x, y)$. This then directly corresponds to a contribution to the Hall response $\sigma_{xy} = (\Phi_B/2\pi)$. We can write this directly in terms of the boundary theory at the Fermi surface as $\Phi_B = \int d\theta A_\theta$. In the free-fermion case, this is exactly the Berry phase contribution to the anomalous Hall effect discussed above. In a general EFL, we should regard this as an unavoidable contribution due to the 't Hooft anomaly of the IR theory. Like in the interacting Fermi liquid, the full measured anomalous Hall effect may include other contributions that are "local" on the Fermi surface. Note that Eq. (57) has the structure of a 3D Chern-Simons term but with a coefficient that is not quantized to be an integer in multiples of $1/4\pi$. This is allowed here because the boundary gauge fields are coupled to the gapless modes associated with the Fermi surface. The full boundary action that includes both the Fermi-surface modes and the unquantized Chern-Simons action of

Eq. (57) will be properly gauge invariant (including under large gauge transformations).

E. Chiral magnetic effect

The chiral magnetic effect refers to a phenomenon in a three-dimensional Fermi liquid where if the Chern number of the Berry curvature on the two-dimensional Fermi surface is nonzero (which, in a noninteracting Fermi gas, would occur when the Fermi surface encloses a Weyl point), then the total charge becomes nonconserved in the presence of both an electric field and magnetic field, with $\mathbf{E} \cdot \mathbf{B} \neq 0$ [92]. Recall that the Berry gauge field on the Fermi surface is still defined in a general EFL. Here we show that when this Berry gauge field has a nontrivial Chern number in a 3D EFL, then the system exhibits the chiral magnetic effect.

We start from the anomaly equation that is the analog for an EFL in three spatial dimensions of Eq. (17) (which was stated for an EFL in two spatial dimensions); for convenience, we write it in an index-free form as

$$d(*j) = \frac{m}{48\pi^3} F \wedge F \wedge F, \qquad (58)$$

where d is the exterior derivative on forms, "$*$" is the Hodge star operator, and J is the current expressed as a 1-form, and F is the 2-form gauge curvature (which we can write locally as $F = dA$, although A may not be globally defined). Note that all these forms live on a six-dimensional space $M \times \mathcal{F}$, where M is the four-dimensional space-time manifold, and $\mathcal{F}$ is the two-dimensional manifold that parametrizes the Fermi surface.

Now suppose that we define the total charge current j_{EM}, which is a 1-form on M, by integrating j over the Fermi surface $\mathcal{F}$. Suppose furthermore that we write $F = \pi_{\mathcal{F}}^*(F_{\mathrm{EM}}) + \pi_M^*(F_{\mathrm{Berry}})$, where F_{EM} are 2-forms on M and $\mathcal{F}$, respectively, and $\pi_{\mathcal{F}}^*$ and π_M^* are the pullback operators associated with the projections $\pi_{\mathcal{F}} : \mathcal{F} \times M \to \mathcal{F}$ and $\pi_M : \mathcal{F} \times M$, respectively. In index notation, this would just be saying that F is the sum of two terms, each of which depends only on, and has only components in, M and $\mathcal{F}$, respectively. Then by integrating Eq. (58), we find that

$$d(*j_{\mathrm{EM}}) = \frac{C}{8\pi^2} F_{\mathrm{EM}} \wedge F_{\mathrm{EM}}, \qquad (59)$$

where

$$C = \frac{1}{2\pi} \int_{\mathcal{F}} F_{\mathrm{Berry}} \qquad (60)$$

is the Chern number of the Berry curvature. We can also write Eq. (59) as

$$\partial_\alpha j_{\mathrm{EM}}^\alpha = \frac{C}{8\pi^2} \mathbf{E} \cdot \mathbf{B}, \qquad (61)$$

indicating that the charge is not conserved when the right-hand side is nonzero.

VII. EXTENSION TO SPINFUL SYSTEMS

Throughout this paper, we assume that the only microscopic internal symmetry is U(1), the charge conservation symmetry. Physically, one often wants to consider systems which also have an SU(2) spin-rotation symmetry. In that case, the full internal symmetry group is U(2), which acts by matrix multiplication on the vector $(\psi_\uparrow^\dagger, \psi_\downarrow^\dagger)$ of microscopic spin-up and spin-down electron creation operators. Here we briefly discuss how the general considerations above get extended in that case. (One is still allowed to use the results of previous sections by simply ignoring the spin-rotation symmetry, but taking into account that the additional symmetry will lead to stronger constraints.) For simplicity, we focus on the case of $d = 2$ spatial dimensions.

First, we observe that U(2) has a subgroup $\mathrm{U}(1)_\uparrow \times \mathrm{U}(1)_\downarrow \leq \mathrm{U}(2)$ corresponding to the diagonal unitary matrices. We write the generators of $\mathrm{U}(1)_\uparrow$ and $\mathrm{U}(1)_\downarrow$ as $\hat{Q}_\uparrow$ and $\hat{Q}_\downarrow$; they measure the total number of up-spin and down-spin electrons, respectively. The total charge is the sum $\hat{Q} = \hat{Q}_\downarrow + \hat{Q}_\uparrow$. They both have corresponding fillings $\nu_\uparrow$ and $\nu_\downarrow$ in the ground state; however, U(2) invariance of the ground state immediately implies that $\nu_\uparrow = \nu_\downarrow := \nu$. The total charge density is then $\rho = 2\nu$. Now, if we consider a spinful version of a Fermi liquid, the quasiparticle charge at each Fermi-surface point will be conserved, as will the *total* quasiparticle spin. Note in particular that the quasiparticle spin at each Fermi-surface point is not separately conserved due to the presence of Landau interactions in the spin channel. Thus, for a spinful Fermi liquid, the emergent symmetry group G_{IR} is a quotient of $\mathrm{LU}(1) \times \mathrm{U}(2)$, where the quotient corresponds to identifying two U(1) subgroups: the U(1) subgroup of U(2) corresponding to unitary matrices $e^{i\theta}\mathbb{I}_2$, where $\mathbb{I}_2$ is the (2×2) identify matrix, and the U(1) subgroup of $\mathrm{LU}(1)$ generated by $\hat{N} = \int \hat{n}(\theta) d\theta$, where $\hat{n}(\theta)$ are the generators of $\mathrm{LU}(1)$ as in the previous sections. Therefore, we define a spinful EFL to be a system which has the same emergent symmetry group G_{IR}. We assume that the translation symmetry continues to embed into $\mathrm{LU}(1)$ according to Eq. (14), which defines the Fermi surface of the EFL. Moreover, the mapping of the microscopic U(2) into G_{IR} is induced from a homomorphism $\varphi : \mathrm{U}(2) \to \mathrm{U}(2)_{\mathrm{IR}}$, where the target $\mathrm{U}(2)_{\mathrm{IR}}$ is the corresponding subgroup of G_{IR}. One can show (for example, by considering the induced homomorphism on the Lie algebras), that such maps are characterized by an odd [93] integer q, which again we interpret as the charge of a Fermi-surface quantum (which in a Fermi liquid would be a Landau quasiparticle). Then, we have in particular that the microscopic charge $\hat{Q}$ corresponds in the IR theory to

$$\hat{Q} \sim q(\hat{N}_\uparrow + \hat{N}_\downarrow), \qquad (62)$$

where $\hat{N}_\uparrow$ and $\hat{N}_\downarrow$ are the generators of the $U(1)_{\uparrow,\text{IR}} \times U(2)_{\downarrow,\text{IR}} \leq U(2)_{\text{IR}}$ symmetry corresponding to the diagonal matrices in $U(2)_{\text{IR}}$.

Now we can discuss the form that Luttinger's theorem must take in a spinful EFL. We imagine inserting a 2π flux of $U(1)_{\uparrow,\text{IR}}$ or $U(1)_{\downarrow,\text{IR}}$; this leads to a projective representation of $LU(1)$ described by Eq. (26) with some integer anomaly coefficient $m_\uparrow$ or $m_\downarrow$, respectively. $U(2)$ invariance again implies that $m_\uparrow = m_\downarrow := m$. Therefore, repeating the argument for Luttinger's theorem, and taking into account Eq. (62), we find

$$\frac{mq}{(2\pi)^2}\mathcal{V}_F = \nu \quad (\text{mod } 1), \tag{63}$$

where $\mathcal{V}_F$ is the volume enclosed by the Fermi surface. With respect to the total charge density $\rho = 2\nu$, on the other hand, we have

$$\frac{2mq}{(2\pi)^2}\mathcal{V}_F = \rho \quad (\text{mod } 2). \tag{64}$$

The extra factor of 2 takes into account the two possible spin values, and agrees (setting $m = 1$) with the usual result for spinful Fermi liquids. (Thus, if we define m with respect to the total charge as we do in Sec. VI, we find $m = 2$ for a spinful Fermi liquid or EFL.)

Our discussion goes through with little modification for a Kondo lattice in which itinerant electrons couple to a local spin-1/2 moment in each unit cell. In this case, we can define a $U(1)_\uparrow$ symmetry generated by $\hat{Q}_\uparrow^{\text{tot}} := \hat{S}^z + N/2 + \hat{Q}_\uparrow$, where $\hat{S}^z$ is the total spin component of the local moments, $\hat{Q}_\uparrow$ is defined as before with respect to the itinerant electrons, and N is the total number of unit cells. Assuming no spin ordering of the local moments, we have that $\langle S^z \rangle = 0$, and hence, the total filling of $\hat{Q}_\uparrow^{\text{tot}}$ is $\nu_\uparrow^{\text{tot}} = 1/2 + \nu_\uparrow$, where $\nu_\uparrow$ is the contribution from the itinerant electrons. If we now apply similar arguments to before, we find that Luttinger's theorem for a putative IR Fermi liquid requires a large Fermi surface that counts the local moments as part of the Fermi sea, as previously discussed with the flux threading argument in Ref. [7].

For an alternative point of view, we can think of the Kondo lattice model as being obtained from an Anderson lattice model with a correlated half-filled band of electrons coupled to a separate weakly correlated partially filled band. Then, by definition, the charge ρ per unit cell includes *all* the microscopic electrons, in particular, the contribution from the correlated band. If the IR theory is a spinful Fermi liquid (or spinful EFL), then Luttinger's theorem in the form Eq. (64) shows that the Fermi surface must be large, as in the previous paragraph. Since our discussion (as well as that of Ref. [7]) of Luttinger's theorem does not rely on perturbation theory in the

interaction strength, it holds even in the Kondo limit of the Anderson model.

Note that the arguments of this section assume that all the relevant symmetries—charge conservation, spin rotation, and translation symmetry—are preserved and not spontaneously broken, and that the IR theory is a spinful Fermi liquid or at least a spinful EFL with the same emergent symmetries. If any of these conditions are violated, then the appropriate statement of Luttinger's theorem will be modified, and in such cases, a "small" Fermi surface could be permitted in the Kondo lattice model. The case of fractionalized Fermi liquids, which have a small Fermi surface while preserving all the symmetries, requires separate discussion but can be easily understood within the framework of this paper. We illustrate this for a simple class of such phases in Sec. X.

VIII. FERMI SURFACES COUPLED TO DYNAMICAL GAUGE FIELDS

In this section, we consider cases where the IR theory comprises a Fermi surface coupled to a dynamical gauge field. As we see, these do not constitute EFLs in the strict sense of Sec. VI, but they are still closely related and it still turns out to be helpful to think of the behavior of these systems in terms of their kinematic properties.

We consider three main examples. The first two examples relate to a system of interacting electrons in two spatial dimensions with continuous translational symmetry in a magnetic field B, such that the magnetic filling $\nu = 2\pi\rho/B$ (where ρ is the electron density) is equal to $\nu = 1/2$. Such a system of electrons is believed to form a "composite Fermi liquid" with a Fermi surface of "composite fermions" which are distinct from the microscopic electrons. There are two competing proposals for the IR theory of the system, which we consider separately: the one proposed by Halperin, Lee, and Read (HLR) [77] and the one proposed by Son [94]. In the HLR and Son theories, the system is not quite a conventional Fermi liquid, even when expressed in terms of the composite fermions, because in the IR theory the composite fermions still couple to a dynamical gauge field.

The final example relates to a Mott insulator in two spatial dimensions with discrete translation symmetry and no magnetic field. In certain circumstances, it is believed that the IR theory of such a system can have a "spinon Fermi surface" consisting of emergent fermions coupled to a dynamical gauge field.

A. HLR theory of composite Fermi liquid

We begin by discussing the microscopic symmetries of electrons in $2d$ in a uniform magnetic field. It is well known that in the presence of such a magnetic field, the translation operators $\hat{P}_x$ and $\hat{P}_y$ do not commute. Rather, they satisfy

$$[\hat{P}_x, \hat{P}_y] = iB\hat{Q}, \tag{65}$$

$$[\hat{P}_x, \hat{Q}] = 0, \tag{66}$$

$$[\hat{P}_y, \hat{Q}] = 0, \tag{67}$$

where $\hat{Q}$ is the operator for the total electric charge [generator of the global U(1) symmetry]. This symmetry algebra has to be matched by any putative IR theory. In an infinite system, we think of these operators $\hat{Q}, \hat{P}_x, \hat{P}_y$ as the generators of the symmetry action on local observables; hence, there is an ambiguity in their definition in that one is free to shift any of them by a constant without changing the action on local observables. Thus, one is free to fix them to have zero eigenvalue in the ground state.

Let us now consider the HLR theory of the composite Fermi-liquid ground state at $\nu = \frac{1}{2}$. Our starting point is the Lagrangian for the HLR theory, which can be written as [77,95]

$$L = L[\chi, a] - \frac{2}{4\pi} b \wedge db + \frac{1}{2\pi}(A - a) \wedge db, \tag{68}$$

where a and b are dynamical U(1) gauge fields, and A is the background gauge field for the microscopic U(1) symmetry. Here, χ is a fermion field carrying unit gauge charge of a, and $L[\chi, a]$ is the Lagrangian describing a Fermi surface of these fermions coupled to the gauge field a. This is the corrected version (see, e.g., Ref. [95]) of the Lagrangian initially written by HLR [77], which suffered from an improperly quantized Chern-Simons term.

Varying with respect to the temporal components a_0 and b_0 gives the constraints

$$\hat{n}_\chi - \hat{B}_b/(2\pi) = 0, \tag{69}$$

$$-2\hat{B}_b + (B - \hat{B}_a) = 0, \tag{70}$$

where $\hat{n}_\chi$ is the density of the χ fermions, and $\hat{B}_a$ and $\hat{B}_b$ are the magnetic fields of the a and b gauge fields, respectively. Meanwhile, varying with respect to A_0 allows us to identify the microscopic charge density $\hat{\rho}$ as

$$\hat{\rho} = \hat{B}_b/(2\pi). \tag{71}$$

We define the excess charge according to

$$\hat{N} = \int (\hat{n}_\chi - n_\chi) d^2\mathbf{x}, \tag{72}$$

where n_χ is the charge density in the ground state. We also define the effective magnetic flux that the χ fermions experience (they couple only directly to a) according to

$$\hat{\Phi} = \int \hat{B}_a d^2\mathbf{x}. \tag{73}$$

Using $2n_\chi = B$ and the constraints above, it follows that

$$2\hat{N} = -\frac{\hat{\Phi}}{2\pi}. \tag{74}$$

Now we need to consider what form the IR symmetry takes. Let us first consider what happens if we treat the gauge fields a and b at the mean-field level. In that case, the χ fermions do not experience a magnetic field in the ground state and therefore form a Fermi liquid for which the charge at each point of the Fermi surface is conserved. Thus, the emergent symmetry group is still something resembling LU(1) with generators $\hat{n}(\theta)$. It is not exactly LU(1), however, because if we consider low-energy excitations that carry a net charge $\hat{N} = \int \hat{n}(\theta) d\theta \neq 0$, where $\hat{n}(\theta)$ are the generators of the emergent symmetry group, then from Eq. (74) we see that the χ fermions feel a net magnetic flux. From the discussion of Sec. V D, we expect the magnetic flux to induce a projective representation of LU(1). [96] Hence, the emergent symmetry group is $\tilde{\text{LU}}(1)$, where $\tilde{\text{LU}}(1)$ is obtained from LU(1) by replacing the commutation relation $[\hat{n}(\theta), \hat{n}(\theta')] = 0$ obeyed by the generators $\hat{n}(\theta)$ of LU(1) with

$$[\hat{n}(\theta), \hat{n}(\theta')] = 2i \frac{1}{2\pi} \delta'(\theta - \theta') \int \hat{n}(\theta) d\theta, \tag{75}$$

and with those elements $U_f \in \text{LU}(1)$ corresponding to functions $f: S^1 \to \text{U}(1)$ with nontrivial winding number $W_f = (1/2\pi) \int \partial_\theta f(\theta)$ excluded. [Because otherwise Eq. (75) would imply that $U_f \hat{N} U_f^{-1} = (1 + 2W_f)\hat{N}$, which is mathematically inconsistent unless $W_f = 0$.] From Eqs. (69) and (71), it is clear that the microscopic charge $\hat{Q}$ embeds into the IR symmetry group according to $\hat{Q} \sim \hat{N}$. Meanwhile, the continuous translation symmetry, with generators $\hat{P}_x$ and $\hat{P}_y$ should embed according to

$$\hat{P}_\alpha \sim \int k_\alpha(\theta) \hat{n}(\theta) d\theta, \tag{76}$$

which defines the Fermi surface $\mathbf{k}(\theta)$.

We do not attempt to extend the general discussion of topological terms as in Sec. V C to this more nontrivial symmetry group $\tilde{\text{LU}}(1)$. However, we can at least argue for a version of Luttinger's theorem for the composite Fermi liquid. Unlike in the case of no magnetic field, it turns out to be unnecessary to invoke the 't Hooft anomaly of the IR theory; the result actually follows directly from the structure of the emergent symmetry group.

From the action of translations on the low-energy theory, we find that

$$[\hat{P}_x, \hat{P}_y] \sim \frac{2\mathcal{V}_F}{2\pi} \hat{N}, \tag{77}$$

where

$$V_F = \int k_x(\theta) \frac{dk_y(\theta)}{d\theta} d\theta \qquad (78)$$

is the volume enclosed by the Fermi surface. We emphasize that here $\hat{N}$ is not just a number but an *operator* which acts nontrivially on the states of the IR theory.

On the other hand, microscopically we know that we must have the magnetic translation algebra of Eq. (65) with which we must impose compatibility of Eq. (77). This gives

$$V_F = \pi B, \qquad (79)$$

which is Luttinger's theorem for the composite Fermi liquid.

Now we should discuss to what extent we expect these considerations to survive once we include gauge fluctuations. The only property of the IR theory that we need to derive Luttinger's theorem is that the emergent symmetry group is $\tilde{L}U(1)$. The question is whether this emergent symmetry survives the inclusion of gauge fluctuations. Let us argue that, at any rate, the emergent symmetry group cannot be a compact Lie group, which suggests that the full $\tilde{L}U(1)$ symmetry is preserved. We cannot apply the results of Sec. IV to show this because if we define the magnetic unit cell such that the discrete translations with respect to this unit cell commute, then the filling with respect to this unit cell is supposed to be $\nu = 1/2$ for the composite Fermi liquid, a rational number.

However, we can instead argue as follows. Suppose that the emergent symmetry group G_{IR} is a compact Lie group. Then let $\hat{p}_x$, $\hat{p}_y$, $\hat{q}$ be the elements of the Lie algebra of G_{IR} corresponding to the microscopic translation symmetries $\hat{P}_x$ and $\hat{P}_y$ and the microscopic charge conservation symmetry $\hat{Q}$. Then since the microscopic symmetries obey the algebra of Eqs. (65)–(67), it follows that the same algebra must be satisfied by $\hat{p}_x$, $\hat{p}_y$, $\hat{q}$. Equations (65)–(67) constitute the Heisenberg algebra and cannot be implemented inside of a compact Lie group unless $\hat{q}$ is identically zero. To see this, suppose that G_{IR} is a compact Lie group. Then it admits a finite-dimensional faithful unitary representation. Since $\hat{p}_x$, $\hat{p}_y$, and $\hat{q}$ generate a subgroup of G_{IR}, this induces a finite-dimensional representation of the Heisenberg algebra Eqs. (65)–(67). Now we invoke the famous fact that the Heisenberg algebra does not admit any finite-dimensional representations in which $\hat{q}$ acts nontrivially. [97]

Therefore, we conclude that, if G_{IR} is a compact Lie group, then the charge conservation symmetry acts trivially on the IR theory; in other words, the system has a charge gap and is an electrical insulator. This would be inconsistent with the physics expected of a composite Fermi liquid. Thus, we expect that the emergent symmetry group

remains $\tilde{L}U(1)$ in the presence of gauge fluctuations. Of course, in principle, it could go to some totally different group that is larger than any compact Lie group, but this does not seem very plausible, and moreover, as we discuss in Sec. VI, the charge is still conserved at every point in the Fermi surface in the conventional description of the fixed point for a Fermi surface coupled to dynamical gauge field. Thus, assuming that the emergent symmetry group remains $\tilde{L}U(1)$, we find that the composite Fermi liquid indeed obeys Luttinger's theorem. Note that, though it is suggested based on numerics in Ref. [98] that Luttinger's theorem could be violated in the composite Fermi liquid, more extensive numerics show this to be a finite-size effect [99].

B. Son theory of composite Fermi liquid

We start from the Lagrangian [94,95]

$$\mathcal{L} = i\bar{\chi}\slashed{D}_a\chi - \frac{2}{4\pi}b \wedge db + \frac{1}{2\pi}a \wedge db - \frac{1}{2\pi}A \wedge db + \cdots, \qquad (80)$$

where a and b are dynamical U(1) gauge fields, and the first term represents the Lagrangian for a massless relativistic Dirac fermion field χ coupled to the gauge field a. [100] The "$\cdots$" represents higher-order terms. This is the corrected version [95] of the Lagrangian originally written by Son [94].

Proceeding similarly to Ref. [95], from this Lagrangian one obtains the constraints

$$\hat{N} - \frac{1}{4\pi}\hat{\Phi}_a + \frac{1}{2\pi}\hat{\Phi}_b = 0, \qquad (81)$$

$$-2\hat{\Phi}_b + \hat{\Phi}_a = 0, \qquad (82)$$

$$\hat{Q} = -\frac{1}{2\pi}\hat{\Phi}_b, \qquad (83)$$

where $\hat{N}$ is the charge of the χ fermions, $\hat{\Phi}_a$ and $\hat{\Phi}_b$ are the total magnetic fluxes of the a and b fields, respectively, $\hat{Q}$ is the total microscopic charge, and all of $\hat{N}, \hat{Q}, \hat{\Phi}_a, \hat{\Phi}_b$ represent the excess values compared to the ground state. From these equations, we find that $\hat{N} = 0$ and $\hat{\Phi}_a = 2\hat{\Phi}_b$.

At the mean-field level, the χ fermions form a Fermi liquid, so the charge at each point on the Fermi surface is conserved. However, we have to impose the constraint that $\hat{N} = \int \hat{n}(\theta)d\theta = 0$. Assuming for the moment that the fermions feel no magnetic flux, this would imply that the emergent symmetry group is $\Omega U(1) := LU(1)/U(1)$.

Now if we take into account that the fermions feel an effective magnetic flux given by $\hat{\Phi}_a = 2\hat{\Phi}_b$, we must have the commutation relation

$$[\hat{n}(\theta), \hat{n}(\theta')] = -2i\frac{1}{2\pi}\delta'(\theta - \theta')\frac{\hat{\Phi}_b}{2\pi}. \qquad (84)$$

Since the $\hat{n}(\theta)$'s are conserved, it must also be the case that the magnetic flux $\hat{\Phi}_b$ is conserved and (since it is quantized in units of 2π) generates a $U(1)_{\text{flux}}$ symmetry. We can also observe that, for a general function $f: S^1 \to U(1)$, Eq. (84) implies that the corresponding group element U_f satisfies $U_f\hat{N}U_f^{-1} = \hat{N} + 2W_f\hat{\Phi}_b$, where $W_f = (1/2\pi)\int\partial_\theta f(\theta)d\theta$ is the integer winding number. Since we are imposing that $\hat{N}$ is identically zero (and $\hat{\Phi}_b$ is not), we must exclude U_f with $W_f \neq 0$ from the symmetry group, as in the HLR case.

Hence, the full emergent symmetry group G_{IR} is a central extension of $\Omega_0 U(1)$ by $U(1)_{\text{flux}}$ where $\Omega_0 U(1) = L_0 U(1)$, and $L_0 U(1)/U(1)$ is the subgroup of $LU(1)$ corresponding to functions $f: S^1 \to U(1)$ with trivial winding number. We note that this is a different group compared to the HLR case, which reflects a physical difference between the Son and HLR theories.

Equation (83) tells us how the microscopic $U(1)$ symmetry embeds into G_{IR}, and as before, the continuous translation should embed according to Eq. (76), which defines the Fermi surface. By similar arguments to the HLR case, one again finds Luttinger's theorem Eq. (79) for the composite Fermi liquid. Since again, the only property of the IR theory that we use is the nature of the emergent symmetry group, we expect that this result continues to hold in the presence of gauge fluctuations.

Note that our argument for Luttinger's theorem in the Son theory is very different from the one given by Son [94], who noted that the density of composite fermions is given by $\rho_{\text{CF}} = B/(4\pi)$ and then invoked Luttinger's theorem for a Fermi liquid to relate this to the volume of the Fermi surface. Of course, the composite Fermi liquid is *not* a Fermi liquid beyond mean-field theory, but the arguments of this paper show that Luttinger's theorem is, in fact, far more general. However, from this point of view, it remains unclear whether one ought to expect Luttinger's theorem to hold with respect to ρ_{CF}, since the composite fermions do not carry charge under any global $U(1)$ symmetry. The argument we give above does not suffer from these difficulties.

C. Spinon Fermi surface

Finally, let us consider a system in two spatial dimensions where the IR theory consists of a *spinful* Fermi surface coupled to a dynamical $U(1)$ gauge field. This should apply, for example, to Mott insulators exhibiting a "spinon Fermi surface."

If we treat the dynamical gauge field at the mean-field level, then the fermions form a spinful Fermi liquid described by the considerations of Sec. VII. Thus, the Fermi surface density $\hat{n}(\theta)$ is conserved. In particular, if we use the observation of that section that a spinful Fermi

surface has anomaly coefficient $m = 2$ with respect to the total charge $U(1)$, we conclude that the $\hat{n}(\theta)$'s must satisfy

$$[\hat{n}(\theta), \hat{n}(\theta')] = -2i\frac{1}{2\pi}\delta'(\theta - \theta')\frac{\hat{\Phi}_b}{2\pi}, \qquad (85)$$

where $\hat{\Phi}_b$ is the operator that measures the magnetic flux of the dynamical gauge field. Moreover, we have to impose that the total gauge charge is zero, so $\int \hat{n}(\theta)d\theta = 0$. We see that the emergent symmetry group G_{IR} takes the form $G_{\text{IR}} = G_{\text{IR,charge}} \times SO(3)$, where the charge part $G_{\text{IR,charge}}$ is identical to the emergent symmetry group of the Son theory as we describe above [although for an electrical insulator the microscopic charge, $U(1)$ will act trivially on the IR theory instead of embedding into the emergent $U(1)_{\text{flux}}$], while the $SO(3)$ part accounts for the spin-rotation symmetry of the spinons. [101].

Unlike in the quantum Hall systems discussed above, for a Mott insulator one microscopically has only a discrete (commuting) translation symmetry. The discrete translations must embed into G_{IR} in the usual way:

$$\mathbb{T}_\alpha \sim \exp\left(-i\int k_\alpha(\theta)\hat{n}(\theta)\right), \qquad (86)$$

which defines the Fermi surface. From Eq. (85), the requirement that $\mathbb{T}_x$ and $\mathbb{T}_y$ commute give a nontrivial constraint:

$$\frac{\mathcal{V}_F}{(2\pi)^2} = 0 \quad \text{or} \quad \frac{1}{2} \quad (\text{mod } 1), \qquad (87)$$

where $\mathcal{V}_F$ is the volume enclosed by the Fermi surface.

We can obtain more information by considering the 't Hooft anomaly of the IR theory and invoking the filling constraints of Sec. III B. We assume that the system microscopically has spin-rotation symmetry, so that the full microscopic internal symmetry group is $U(2)$, and in particular, there is a $U(1)_\uparrow \times U(1)_\downarrow$ subgroup as we discuss in Sec. VII, and we can define the corresponding fillings $\nu_\uparrow$ and $\nu_\downarrow$ which satisfy $\nu_\uparrow = \nu_\downarrow := \nu$ and $\rho = 2\nu$, where ρ is the filling of the total charge. For an insulating state without symmetry fractionalization (see Sec. X), the microscopic charge $U(1)$ acts trivially on the IR theory, and hence, from the general theory of filling constraints, we find that ρ must be an integer. On the other hand, for a band insulator, ρ must be an even integer (taking into account the two spin components). Since a Mott insulator by definition is an insulator that exists at fillings where band insulators are impossible, we conclude that ρ for a Mott insulator must be an odd integer, and hence, $\nu = 1/2(\text{mod } 1)$.

From the general discussion of Sec. III B, we now see that if we apply a flux $\Phi_\uparrow$ of $U(1)_\uparrow$ and flux $\Phi_\downarrow$ of $U(1)_\downarrow$, it must be the case that translations act projectively on it,

$$\mathbb{T}_x \mathbb{T}_y \mathbb{T}_x^{-1} \mathbb{T}_y^{-1} = (-1)^{(\Phi_\uparrow + \Phi_\downarrow)/(2\pi)}. \qquad (88)$$

Now we need to see how this comes about in the IR theory. We return to Eq. (85) but take into account that, at the mean-field level where we can treat the dynamical gauge field as a background, the fermions now feel an effective $U(1)_\uparrow$ flux of $\hat{\Phi}_b + \Phi_\uparrow$ and an effective $U(1)_\downarrow$ flux of $\hat{\Phi}_b + \Phi_\downarrow$. Hence, Eq. (85) gets generalized to [102]

$$[\hat{n}(\theta), \hat{n}(\theta')] = -i \frac{1}{2\pi} \delta'(\theta - \theta') \frac{2\hat{\Phi}_b + \Phi_\uparrow + \Phi_\downarrow}{2\pi}, \quad (89)$$

which indeed gives Eq. (88) provided that

$$\frac{\mathcal{V}_F}{(2\pi)^2} = \frac{1}{2} \quad (\text{mod } 1), \qquad (90)$$

which is Luttinger's theorem for a spinon Fermi surface.

As usual, we must consider to what extent we expect these results to hold beyond mean-field theory. The result Eq. (87) depends only on the structure of G_{IR}, so we must ask whether G_{IR} remains the same upon including gauge fluctuations. Since we are assuming the system is a Mott insulator, with integer filling of the microscopic U(1) charge, we cannot use the results of Sec. IV. Still, as we mention in the introduction to Sec. VI, the charge on each patch of the Fermi surface remains conserved in the usual approach to Fermi surfaces coupled to a dynamical gauge field, suggesting that G_{IR} is indeed robust to gauge fluctuations and that Eq. (87) is satisfied generally. To get the stronger result Eq. (90), we also have to assume that the 't Hooft anomaly of the IR theory captured by Eq. (89) remains unchanged by gauge fluctuations. However, the classification of 't Hooft anomalies is discrete—more concretely, there is no way to continuously deform Eq. (89) without leading to inconsistencies—so it seems unlikely that the 't Hooft anomaly would be affected by gauge fluctuations. Moreover, from Eq. (87) the only possibility other than Eq. (90) would be $\mathcal{V}_F/(2\pi)^2 = 0 (\text{mod } 1)$, in which case, the Fermi surface is singular and presumably unstable.

IX. IMPOSSIBILITY OF FERMI ARCS AND OTHER CONSTRAINTS ON THE FERMI SURFACE

A. Fermi arcs

In the celebrated pseudogap normal phase of the cuprate high-temperature superconducting materials, ARPES measurements observe "Fermi arcs" rather than a closed Fermi surface (see, e.g., Ref. [103] for a review). One possible explanation is that these materials still have a closed Fermi surface, but for some reason parts of the Fermi surface are not easily visible to the ARPES probe. On the other hand, three-dimensional Weyl semimetals exhibit Fermi arcs at their two-dimensional boundaries [104], but these can exist only at the boundary of a gapless three-dimensional bulk.

It is an important fundamental question to ask whether a system with true Fermi arcs (not related to the above mechanisms) can ever exist.

Specifically, in this section we ask whether it is possible to extend the framework of EFLs to substitute a closed Fermi surface with a Fermi arc. We find that there is an obstruction to doing so. Therefore, in light of the results of Sec. IV, the only remaining route to Fermi arcs, assuming that they can exist in systems at generic filling, would be to find some other way to have an emergent symmetry group larger than any compact Lie group, different from what occurs in EFLs.

Let us consider a theory in d spatial dimensions described by a suitable generalization of the EFL class discussed in Sec. VI. Specifically, we imagine that the Fermi surface is a parametrized by a $(d-1)$-dimensional manifold F, but, instead of requiring that F be a closed manifold as before, we allow it to be a manifold with boundary. Then we assume that the emergent symmetry group is $L^F U(1)$, the space of smooth maps from F to $U(1)$. We can discuss 't Hooft anomalies as before, and again we find that the 't Hooft anomaly should be described by inflow from a Chern-Simons term on a higher-dimensional manifold, for example, for $d = 2$:

$$S[A] = \frac{m}{24\pi^2} \int_{M_+ \times F} A \wedge dA \wedge dA, \qquad (91)$$

where M_+ is an extension of the $(d+1)$-dimensional space-time manifold to a $(d+2)$-dimensional manifold. Now, however, we immediately encounter a problem if F has a boundary ∂F, because the Chern-Simons term is not gauge invariant—that is, the system has a 't Hooft anomaly—in the presence of boundaries (unless $m = 0$). At the boundary of M_+, the gauge variation is canceled by the 't Hooft anomaly of the EFL, but since the interior of M_+ is supposed to correspond to a gapped SPT bulk, there is no way to cancel the anomaly at $M_+ \times \partial F$.

Should we be worried about this anomaly? In general, an uncanceled 't Hooft anomaly signals a violation of charge conservation. We can see this concretely in spatial dimension $d = 2$ if we assume a particular model of a Fermi arc which has Fermi-liquid-like quasiparticles. Then in the presence of a magnetic field, the quasiparticles have a chiral flow along the Fermi surface as discussed in Sec. V D, but when they reach the end of the Fermi arc they have no choice but to simply disappear. This violates conservation of microscopic charge (if the microscopic charge q of a quasiparticle is nonzero) and conservation of momentum (even when $q = 0$). Beyond this simple model, the 't Hooft anomaly indicates the same issue with lack of charge or momentum conservation, and therefore, we must impose that, if F has a boundary, then $m = 0$. However, by the discussion of Sec. V D, this would imply the filling

$\nu = 0$ mod 1, so it cannot correspond to a state that exists at generic filling.

Of course, this is only a proof of the impossibility of Fermi arcs if we assume that the system is described by the EFL framework. We cannot make more general statements, except to recall from Sec. IV that any system that can exist at generic filling, yet is not an EFL, needs to find some other way of having an emergent symmetry group that is larger than any compact Lie group.

Finally, we should address one physical mechanism that one might think of to generate Fermi arcs, where the system couples to a critical boson at some wave vector **q**. Then, in the quasiparticle picture described above, one could imagine that a quasiparticle, upon reaching the end of a Fermi arc at momentum **k**, jumps to the beginning of another arc at momentum **k** − **q** while emitting a boson of momentum **q**. In the language of emergent symmetries, this would correspond to introducing another operator $\hat{N}_{\text{boson}}$ that generates a U(1) group, and then writing the translation symmetry as

$$\mathbb{T}_\alpha = \exp\left(-i \int k_\alpha(\theta)\hat{n}(\theta)d\theta\right) \exp\left(-iq_\alpha \hat{N}_{\text{boson}}\right). \quad (92)$$

This would amount to postulating a slightly different emergent symmetry group, namely, a modification of LU(1) where one allows the corresponding functions $f : S^1 \to \mathrm{U}(1)$ to have at least a finite number of discontinuities. The question is whether there is a meaningful nontrivial anomaly that one could write for such a group, as is necessary if it is to describe a system with noninteger microscopic filling. It seems likely that there is not, for reasons that we explain at the end of the next subsection.

B. More general constraints

Returning to EFLs, we can also obtain some more general constraints than just the impossibility of Fermi arcs. Indeed, the most general statement is that the Fermi surface must "enclose" a volume in a suitable generalized sense. This obviously fails for Fermi arcs, but it also rules out Fermi surfaces which are closed but nevertheless fail to enclose a volume in the Brillouin zone; e.g., see Fig. 2(c).

Let us imagine generally that we construct the Fermi surface out of a collection of "patches" which might be glued together along their boundaries and might not all have the same values of the anomaly coefficient m. Then the requirement that the Chern-Simons term be gauge invariant places a constraint on how patches are allowed to be glued together. For example, in two spatial dimensions, the rule is that the sum of the anomaly coefficients for all the segments of Fermi surface (taking into account the orientation of the segment) intersecting at a given point must be zero. This is the most general sense in which the Fermi surface must be "closed." Thus, the configuration
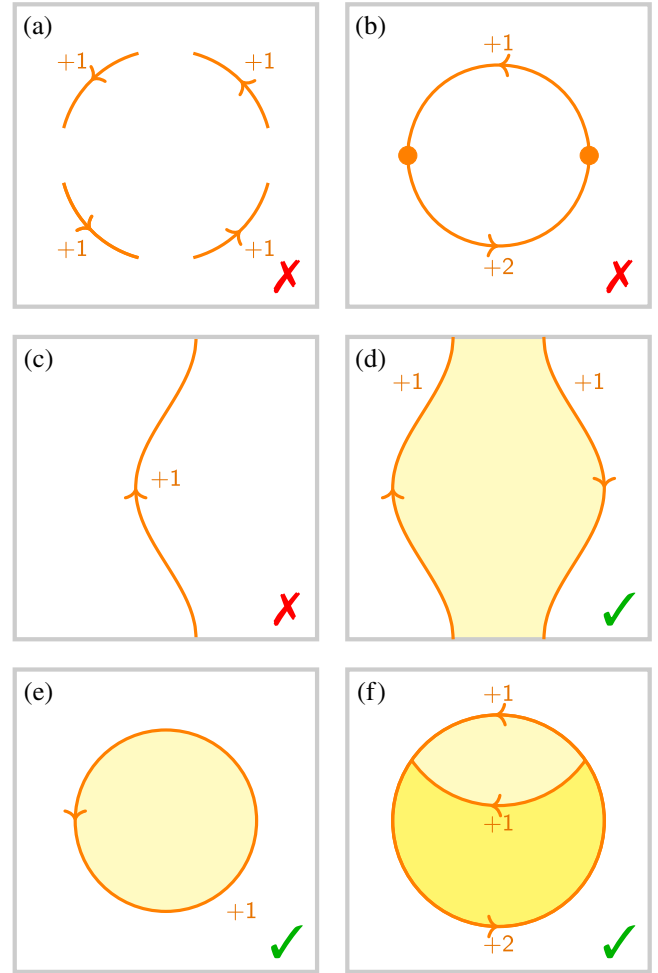

FIG. 2. Allowed (✓) and disallowed (✗) Fermi-surface configurations in a two-dimensional Brillouin zone. Each Fermi-surface segment has an associated anomaly coefficient m shown by the number next to the segment. An arbitrary choice of orientation of the segment (denoted with arrows) determines the convention for the sign of m. For allowed configurations, it is always possible to define an integer function $n(\mathbf{k})$—depicted here by the different shading colors of different regions—such that the boundary between two regions carries the anomaly coefficient determined by the difference of $n(\mathbf{k})$ between the two regions. In a Fermi gas, $n(\mathbf{k})$ represents the number of occupied bands at the point $\mathbf{k}$.

shown in Fig. 2(f) is allowed, but the one shown in Fig. 2(b) is not.

In fact, we can make an even stronger statement: Configurations such as the one shown in Fig. 2(c), even though the Fermi surface is closed, are also disallowed (except possibly in the case where the Fermi-surface quanta are uncharged, i.e., $q = 0$). In this case, the Chern-Simons term is gauge invariant, but there is still an issue with the implementation of translation symmetry. Recall that for $d = 2$, the translation symmetry $\mathbb{T}_x$ and microscopic charge operator $\hat{Q}$ can be expressed in terms of the LU(1) generators $\hat{n}(\theta)$ as

$$\mathbb{T}_x \sim \exp\left(-i\int k_x(\theta)\hat{n}(\theta)d\theta\right), \qquad (93)$$

$$\hat{Q} \sim q\int \hat{n}(\theta)d\theta. \qquad (94)$$

In light of the commutation relations Eq. (19), when acting on a 2π flux we have

$$\mathbb{T}_x\hat{Q}\mathbb{T}_x^{-1} \sim \hat{Q} + mqW_x, \qquad (95)$$

where

$$W_x = \frac{1}{2\pi}\int \frac{dk_x(\theta)}{d\theta}d\theta \qquad (96)$$

is a winding number of the Fermi surface on the Brillouin torus. However, microscopically the translation operator should commute with the charge, so we must impose that $mqW_x = 0$. Another way to see this (which generalizes more easily to higher dimensions) would be impose conservation of microscopic charge on the electric field response discussed in Sec. VI B.

From such considerations, we can obtain the most general constraint on the Fermi surface. We find that it must be the case that each point $\mathbf{k}$ in the Brillouin zone can be assigned an integer $n(\mathbf{k})$, such that the boundary between two regions with different $n(\mathbf{k})$ carries a Fermi surface with anomaly coefficient m given by the difference Δn. In a Fermi gas, $n(\mathbf{k})$ has a natural interpretation: It describes the number of bands that are occupied at the point $\mathbf{k}$. What we find is that such an $n(\mathbf{k})$ can always be defined in any consistent EFL. Note that, in general, the Fermi surface determines only $n(\mathbf{k})$ up to addition of a $\mathbf{k}$-independent integer. Repeating the analysis of Sec. V D in this more general setting, we can also now formulate the most general version of Luttinger's theorem:

$$\frac{q}{(2\pi)^d}\int d^2\mathbf{k}\,n(\mathbf{k}) = \nu \quad (\text{mod } 1), \qquad (97)$$

where the integral is over the whole Brillouin zone.

We remark that a compact mathematical way to state the above results is if we think of the Fermi surface, together with associated anomaly coefficients m of the different patches, as defining a *chain* (in the homology theory sense) $\omega \in C_{d-1}(\mathbb{T}^d, \mathbb{Z})$, where $\mathbb{T}^d$ represents the Brillouin zone. Then the requirement that the Chern-Simons term be gauge invariant implies that this chain is closed, i.e., it has trivial boundary $\partial\omega = 0$, whereas from the winding number arguments, we find that this chain must be *exact*, $\omega = \partial\kappa$ for some κ.

Finally, we note that in retrospect the result should have been clear, because however the microscopic translations act in the IR theory (which is specified by specifying the

Fermi surface), it must be possible to compute a filling ν by considering the anomaly of the emergent symmetry group. For Fermi surfaces that do not enclose a volume, there is evidently no meaningful way to define an associated ν. It is for this reason that we do not expect there to be any nontrivial anomalies associated with the modified symmetry group described in the last paragraph of Sec. IX A, again because it would lead to Fermi surfaces that do not enclose a volume and therefore have no meaningful ν associated with them.

X. THE ROLE OF FRACTIONALIZATION

The analyses in Secs. III and IV are not sufficient to describe systems which admit fractionalized excitations, i.e., localized excitations which cannot be created locally. To see this, consider a specific and familiar example, namely, a system of bosons on a lattice in two spatial dimensions with microscopic filling $\nu = 1/2$. A symmetry-preserving ground state that is allowed is one where the IR theory is equivalent to a $\mathbb{Z}_2$ gauge theory [105–107]. Let us first recall the physical properties of this state and how it is allowed at filling $\nu = 1/2$. The key property is that the ground state is topologically ordered and admits fractionalized excitations, which we can label as $1, e, m, f$, where e is the $\mathbb{Z}_2$ gauge charge, m is the $\mathbb{Z}_2$ gauge flux, and f is the composite of e and m. Moreover, these excitations can exhibit "symmetry fractionalization" of the microscopic symmetries, which roughly means that they carry fractional quantum numbers. Specifically, in the $\nu = 1/2$ Mott insulator, one of the excitation types (say, e) carries *half-quantized* charge under the microscopic U(1) symmetry. This is allowed because these excitations are fractionalized and can be created only in pairs. Moreover, the m particle also experiences translational symmetry fractionalization [107–109], which is to say that $\mathbb{T}_x\mathbb{T}_y\mathbb{T}_x^{-1}\mathbb{T}_y^{-1} = -1$ when acting on a single m excitation. Again, this does not contradict the fact that translations commute microscopically, because the global number of m particles is always even. Physically, one can think of the translational symmetry fractionalization as saying that the ground state has an e particle in each unit cell; the -1 phase factors then come from the -1 braiding phase for m going around e. But then since the e particle has $1/2$ charge under the microscopic U(1), these e particles contribute a filling $\nu = 1/2$. It is also instructive to ask how this state is compatible with the momentum balance constraints from flux threading introduced by Oshikawa [7]. The point is that adiabatic threading of a 2π flux through one cycle of a torus shifts the ground state from one topological sector to another which is at a different total crystal momentum [5]. This change of the ground-state momentum is then able to match what is expected from flux threading.

How should we think about the emergent IR symmetry of this state? Deep in the IR, at scales below the gap to all quasiparticle excitations, we can describe it as a purely

topological theory. Naively, the only emergent symmetry of the resulting theory might seem to be $G_{IR} = \mathbb{Z}_2$ generated by the "electromagnetic duality" which exchanges a $\mathbb{Z}_2$ gauge charge with a gauge flux; however, for the theory under consideration, neither translations nor the microscopic charge U(1) map into this duality symmetry. Therefore, following the analysis of Sec. III we immediately conclude that $\nu = 0 \mod 1$. How then, are we to make sense, in the context of our general framework, of the fact that such a IR theory actually can exist at $\nu = 1/2$?

We have to think about what went wrong in the argument of Sec. III B. The assumption we make there is that a 2π flux of the microscopic symmetry generated by the microscopic charge $\hat{Q}$ will correspond in the IR theory to a 2π flux of $\hat{Q}_{IR}$, the generator of the corresponding IR symmetry. In any system with a charge gap, we have that $\hat{Q}_{IR}$ is identically zero, so we conclude that such an object is completely trivial, leading to a trivial filling. However, in the $\mathbb{Z}_2$ gauge theory example, a 2π flux of the microscopic symmetry is evidently a nontrivial object since acting on it we have $\mathbb{T}_x \mathbb{T}_y \mathbb{T}_x^{-1} \mathbb{T}_y^{-1} = -1$. In fact, as a defect of the IR theory, it is equivalent to an m particle. In general, an interpretation of nontrivial symmetry fractionalization is precisely the statement that certain flux configurations of the microscopic symmetry look like topologically nontrivial excitations of the IR theory. Therefore, it is always necessary to take into account symmetry fractionalization when computing the filling.

Let us first recall a few cases from previous works where symmetry fractionalization of the microscopic symmetry affects the filling. The first case generalizes the deconfined $\mathbb{Z}_2$ gauge theory for bosons at $\nu = 1/2$ (as we discuss above) to situations where the IR theory is an arbitrary gapped $(2+1)$D gapped topological phase. The computation of ν has been discussed in Refs. [10,11]. The basic idea is that the translational symmetry fractionalization is described by a class in $H^2(\mathbb{Z} \times \mathbb{Z}, A) \cong A$, where A is the group of Abelian anyons. Roughly, we can imagine that there is a background a particle sitting in each unit cell. Meanwhile, the symmetry fractionalization of U(1) is described by an Abelian anyon b, which we can think of as the anyonic charge carried by a 2π flux. Because of the background of a particles, if b particles have nontrivial braiding phase S_{ab} with a particles, then they carry a projective representation of translation symmetry. Thus, we find that the filling is given by

$$\nu = \frac{1}{2\pi i} \log S_{ab} \quad (\text{mod } 1). \tag{98}$$

A nice way to interpret this equation is that since a 2π flux has nontrivial braiding S_{ab} with an a particle, then the a particle must carry fractional charge $q_a = (1/2\pi i) \log S_{ab} (\mod 1)$ under the U(1) symmetry. Therefore, since we have an a particle in each unit cell, the average charge per unit cell is $\nu = q_a$.

Building on the previous case, we can imagine a scenario where the IR theory is equivalent to stacking a gapless theory $\mathcal{C}_{gapless}$ without any topologically nontrivial excitations and a gapped topological theory $\mathcal{C}_{gapped}$. In that case, the microscopic filling ν is simply the sum of the contributions from each component:

$$\nu = \nu_{gapless} + \nu_{gapped}, \tag{99}$$

where $\nu_{gapless}$ is computed according to Sec. III, and ν_{gapped} is computed according to the previous paragraph. An example of such a case is the FL* Fermi liquid discussed in Refs. [8,9].

If we want to construct a more general theory beyond these examples, we might start to become a bit more uncertain about what exactly we mean in general by a "topologically nontrivial excitation." Instead, we introduce an alternative formalism that is more readily generalizable [110–113]. To motivate this formalism, note that in the $\mathbb{Z}_2$ gauge theory example, $G_{IR} = \mathbb{Z}_2$ is not in fact the only emergent symmetry of the IR theory. Indeed, gapped topological phases in two spatial dimensions, such as $\mathbb{Z}_2$ gauge theory, have emergent *1-form* symmetries [27]. This is a general feature of gapped topological phases. In fact, one can think about the 1-form symmetry operators, which act on closed strings when $d = 2$, as becoming the open string operators that create topologically nontrivial excitations at the end points when the string is broken.

Once one has introduced the emergent higher-form symmetries, one can phrase symmetry fractionalization entirely in terms of these symmetries without needing to talk about excitations at all. In general, a d-dimensional system can have emergent 0-form, 1-form,..., up to $(d-1)$-form symmetries. These symmetries combine into a structure called a "d-group" [112,114–116], which we denote $\underline{\underline{G}}_{IR}$. Meanwhile, the microscopic symmetries can also be thought of as forming a d-group $\underline{\underline{G}}_{UV}$, although normally we consider the case where the microscopic symmetries are all 0-form, and there are no nontrivial higher-form symmetries. Then the correct way to describe the action of the microscopic symmetries on the IR theory is through an n-group homomorphism

$$\underline{\underline{\rho}} : \underline{\underline{G}}_{UV} \to \underline{\underline{G}}_{IR}. \tag{100}$$

One might think if the microscopic symmetries are all 0-form, then we do not need to worry about the higher-form part of $\underline{\underline{G}}_{IR}$, but it turns out that even in this case the homomorphism $\underline{\rho}$ contains not just the data of how the microscopic symmetries map into emergent 0-form symmetries, but also the data of the symmetry fractionalization of the microscopic symmetries, which can be interpreted as relating to the interplay between the microscopic symmetries and the emergent higher-form symmetries.

It is instructive to consider another case, in addition to gapped topological phases, where higher-form symmetries modify the relation between microscopic filling and the properties of the IR theory, namely, superfluids, which spontaneously break the charge U(1) symmetry. A superfluid in d spatial dimensions has an emergent U(1) $(d-1)$-form symmetry, since the winding number of the order parameter along any closed curve is conserved in the IR theory. Moreover, a superfluid is a compressible phase which can exist at any filling. One can show that the filling is encoded in the symmetry fractionalization of the translation symmetry by the $(d-1)$-form symmetry. Physically, this is reflected in the fact that a vortex sees an "effective magnetic field" given by the charge density [117,118].

Finally, it is necessary for us to address the question of whether symmetry fractionalization can allow for the spirit of Theorem 1 to be bypassed; that is, is it possible for a system whose emergent symmetry is described by a compact Lie d-group to exist at *generic* filling? The superfluid example above shows that we have to restrict to systems in which the microscopic U(1) and translation symmetries are not spontaneously broken. In that case, we conjecture that a compact Lie d-group symmetry is incompatible with generic filling. To motivate this, first we expect that in order to have generic filling, at the least, the system needs to have an emergent continuous $(d-1)$-form symmetry (e.g., a 1-form symmetry in $d=2$ or a 2-form symmetry in $d=3$). The usual way to get emergent continuous higher-form symmetries is to have an emergent deconfined gauge field. However, in $d=3$, for example (we return to the more subtle $d=2$ case below), we know that a deconfined U(1) gauge theory has two 1-form symmetries (electric and magnetic) but no 2-form symmetry. We can contrast this with gapped topological phases in $d=3$, which typically have both a 1-form symmetry and 2-form symmetry.

We also recall that a deconfined U(1) gauge theory in $d=2$ spatial dimensions is dual to a superfluid, and that the only way for such a theory to be stable (other than by coupling to charged gapless fermions, which would explicitly break the 1-form symmetry) is if the monopole events (in the superfluid language, charge-creation events) are forbidden by one of the microscopic symmetries, in which case, this symmetry will be spontaneously broken.

Therefore, it appears, at least in these well-known mechanisms for obtaining emergent higher-form symmetries, that there is no way to obtain generic filling without spontaneously breaking either the microscopic U(1) or translation symmetry. Whether there is some less familiar theory in d spatial dimensions with an emergent $(d-1)$-form symmetry that can achieve this, we leave as an open question.

XI. DISCUSSION: RELATION TO "VIOLATIONS OF LUTTINGER'S THEOREM"

In this work, we argue that Luttinger's theorem, or a suitable generalization, in fact should hold in great generality, both for Fermi liquids and at least a certain class of non-Fermi liquids. Therefore, we must contend with past works that have reported a violation of Luttinger's theorem in various settings.

In certain cases [119–121], attempts have to been made to formulate Luttinger's theorem in a way that could apply beyond perturbation theory in the interaction strength by expressing it in terms of the volume enclosed both by traditional Landau Fermi surfaces on which the electron Green's function $G(\mathbf{k}, \omega)$ has a pole at $\omega = 0$ and by "Luttinger surfaces" in which $\mathrm{Re}G(\mathbf{k}, \omega)$ changes sign via a zero of $G(\mathbf{k}, 0)$ instead of a pole. However, it has been shown that no such result holds in general [122,123]. In our work, we define a Fermi surface in a general system by its emergent symmetries, rather than by any particular property of the electron Green's function. Our version of Luttinger's theorem is therefore very different from the one disproven by Refs. [122,123].

A potential violation of Luttinger's theorem has been discussed in the context of holographic models [124–127]. For a system that is holographically dual to a gravitational theory that includes a black hole, the charge hidden behind the event horizon does not contribute to the volume enclosed by the Fermi surface. It is not clear, however, whether such a scenario can ever actually occur at zero temperature, since an event horizon would be associated with a nonzero entropy density, seemingly in tension with the third law of thermodynamics. If somehow the event horizon could be stabilized at zero temperature, however, one could envision it being associated with some exotic emergent symmetry group that is larger than any compact Lie group but different from that of a Fermi liquid, in which case, a violation of Luttinger's theorem would be compatible with the considerations of this paper.

Next we discuss the numerical study of the $t - J$ model in Ref. [128]. By an analytic continuation of a high-temperature series expansion, Ref. [128] computed the properties of the $t - J$ model down to around temperature $T \sim 0.2J$ and found that, for several different definitions of the Fermi surface at finite temperature, the volume inside the Fermi surface appears to be plateauing at a value less than the value predicted by Luttinger's theorem. However, our arguments in favor of Luttinger's theorem apply only at zero temperature. Since the ground state of the $t - J$ model is unknown and could well exhibit exotic physics such as high-temperature superconductivity, the significance of these finite-temperature observations remains unclear, especially given the ambiguity of the definition of the Fermi surface at finite temperature and the fact that the results of Ref. [128] also contain hints of Fermi arcs (see Sec. IX A).

Finally, we discuss Ref. [129], which studied an $\mathrm{SU}(N)$ generalization of the $t - J$ model in a $1/N$ expansion where a Fermi-liquid mean-field solution becomes exact in the limit as $N \to \infty$. Reference [129] found that the

fluctuations appear to cause a violation of Luttinger's theorem to leading order in $1/N$. According to our general arguments, a violation of Luttinger's theorem must correspond to a change in the emergent symmetry group (such that the system is no longer an EFL). However, the mean-field solution is an EFL, in particular, a Fermi liquid, and it seems unlikely that fluctuations could give rise to a *new* emergent symmetry not already present while remaining perturbative. Therefore, assuming there is no error in the calculations of Ref. [129], it seems that we must interpret their results as signaling a breakdown of perturbation theory, such that the radius of convergence in $1/N$ is zero. In that case, there is no reason to believe that Luttinger's theorem is violated at all.

XII. CONCLUSION AND OUTLOOK

In this work, we make significant progress toward an understanding of strongly correlated metals, going beyond both Fermi-liquid theory and perturbation theory in the interaction strength. We highlight the essential role of the emergent symmetry of the IR theory and its 't Hooft anomaly.

Going forward, we believe that our work leads to a new perspective on studies of non-Fermi liquids: From our point of view, a crucial question one should ask about any potential non-Fermi-liquid state is what its emergent symmetry group is, and if it is supposed to represent a compressible state, in what way does the emergent symmetry group and its 't Hooft anomaly enable the system to satisfy the filling constraints developed in this paper for a continuously tunable filling ν? (In particular, in this case, the emergent symmetry group must be larger than any compact Lie group.)

What one might expect to find is that, in fact, *every* non-Fermi liquid is simply an EFL as defined in Sec. VI or a slight variant of it (for example, including fractionalization as discussed in Sec. X). It is a very important open question to determine whether there might be a fundamentally different way to realize compressible metallic states, with a totally different emergent symmetry group (which must still be larger than any compact Lie group). If such a possibility were to be realized, we would still expect that consideration of the emergent symmetry group and its 't Hooft anomaly would still be a powerful way to determine properties of the system, as we find for EFLs.

In particular, it is quite energizing to imagine how our results apply to the strange metal phase in cuprates [103], assuming that it is possible for it to extend all the way down to a critical point at zero temperature and assuming that it can exist in a translationally invariant system without disorder. The filling at which the critical point occurs does not appear to be near any particular rational filling. Therefore, it is likely that the filling is generic and can be continuously tuned depending on the microscopic parameters, which as we note, leads to very strong

constraints on the IR theory. Although the strange metal has extremely exotic properties, there does not appear to be any obvious reason why it could not still be an EFL. (In particular, we emphasize again that the "Fermi-surface quantum" excitations of an EFL do *not* need to have the nature of quasiparticles, which are certainly expected to be absent in the strange metal.) Identification of the strange metal as being in the EFL class, or even a significant generalization of it, would have further profound implications, which we explore in a subsequent work [130].

APPENDIX A: FILLING DOES NOT CORRESPOND TO THE BOUNDARY OF A SPT PHASE

Here we explain why, contrary to what one might initially assume, filling in d spatial dimensions does not correspond to a boundary of a bulk SPT phase in $d + 1$ spatial dimensions, when viewed in terms of the microscopic symmetries. The relevant symmetry group is $G_{IR} = \mathbb{Z}^d \times U(1)$. One can show [131–133] that the SPT classification for such a symmetry group in D spatial dimensions is given by

$$\mathcal{C}_D \times \prod_{r=1}^{d} \mathcal{C}_{D-r}^{\times \binom{d}{r}}, \qquad (A1)$$

where $\mathcal{C}_m$ is the classification of U(1) SPT phases in m spatial dimensions, $\mathcal{C}^{\times k}$ denotes the product of k copies of $\mathcal{C}$, and $\binom{d}{r}$ is the binomial coefficient. Physically, $\mathcal{C}_D$ correspond to "strong" SPT phases, i.e., those protected by U(1) alone, while the remaining terms correspond to "weak" SPT phases that can be understood by stacking layers of lower-dimensional strong SPT phases.

If filling in d spatial dimensions could somehow correspond to a SPT phase in $D = d + 1$ spatial dimensions, then it would have to be a SPT phase that inherently relies on all d translation symmetries. Therefore, the only term in Eq. (A1) that could be relevant is the $r = d$ term, $\mathcal{C}_1$. The problem is that there *are* no one-dimensional SPT

phases protected by U(1) alone, in either bosonic or fermionic systems. Therefore, $\mathcal{C}_1 = 0$. So we conclude that there are no SPT phases in $d + 1$ spatial dimensions that could correspond to filling in d dimensions.

APPENDIX B: TRANSLATION SYMMETRY FRACTIONALIZATION ON A 2π FLUX IN 2D SYSTEMS AT FRACTIONAL FILLING

In this Appendix, we give a more detailed argument for why a 2π flux in a 2D system with fractional filling should transform projectively under the translation symmetry, as discussed in Sec. III B.

First of all, let us recall that in the presence of a magnetic field, one should introduce gauge-invariant translation operators in the following way. For simplicity of notation, let us assume that the system is defined in continuous space (although with only a discrete translation symmetry); similar considerations can be made in a tight-binding model. We introduce the generator of gauge-invariant continuous translations according to

$$\tilde{\mathbf{P}} = \mathbf{P} - \int \mathbf{A}(\mathbf{x})\hat{\rho}(\mathbf{x})d^2\mathbf{x}, \qquad (B1)$$

where $\mathbf{A}(\mathbf{x})$ is the magnetic vector potential, $\hat{\rho}(\mathbf{x})$ is the local density operator, and $\mathbf{P}$ is the generator of translations when $\mathbf{A} = 0$. These translation operators satisfy the commutation algebra

$$[\tilde{P}_\alpha, \tilde{P}_\beta] = i\epsilon_{\alpha\beta} \int B(\mathbf{x})\hat{\rho}(\mathbf{x})d^2\mathbf{x}, \qquad (B2)$$

where $B(\mathbf{x}) = \epsilon^{\alpha\beta}\partial_\alpha A_\beta(\mathbf{x})$ is the magnetic field. We then define the discrete translation operators $\mathbb{T}_x$ and $\mathbb{T}_y$ by

$$\mathbb{T}_\alpha = \exp(-i\mathbf{a}^{(\alpha)} \cdot \tilde{\mathbf{P}}), \qquad (B3)$$

where $\mathbf{a}^{(\alpha)}$ is the lattice translation vector in the α direction.

Now consider any state $|\psi\rangle$ that obeys the "cluster property"

$$\langle\hat{\rho}(\mathbf{x})\hat{\rho}(\mathbf{x}')\rangle - \langle\hat{\rho}(\mathbf{x})\rangle\langle\hat{\rho}(\mathbf{x}')\rangle \le C|\mathbf{x} - \mathbf{x}'|^{-\eta} \quad (B4)$$

for some constants C and $\eta > 0$, and $\langle\cdot\rangle = \langle\psi| \cdot |\psi\rangle$ denotes expectation values with respect to $|\psi\rangle$. Then it follows from Eq. (B2) that, if we fix that $\int B(\mathbf{x})d^2\mathbf{x} = 2\pi$ but make the magnetic field more and more spread out, then the variance

$$\langle[\tilde{P}_\alpha, \tilde{P}_\beta]^2\rangle - \langle[\tilde{P}_\alpha, \tilde{P}_\beta]\rangle^2 \qquad (B5)$$

goes to zero in the limit where the magnetic field is infinitely spread out. Therefore, when acting on such a state $|\psi\rangle$, and in this limit, we can replace the commutator by its expectation value:

$$[\tilde{P}_\alpha, \tilde{P}_\beta] \sim \langle[\tilde{P}_\alpha, \tilde{P}_\beta]\rangle = i\epsilon_{\alpha\beta} \int B(\mathbf{x})\rho(\mathbf{x})d^2\mathbf{x} \sim \frac{2\pi\nu}{v} i\epsilon_{\alpha\beta}, \qquad (B6)$$

where v is the volume of one unit cell and $\rho(\mathbf{x}) = \langle\hat{\rho}(\mathbf{x})\rangle$. To get to the last identification, we use the fact that $B(\mathbf{x})$ is very slowly varying, and that the integral of $\rho(\mathbf{x})$ over one unit cell is ν. Finally, from Eq. (B6) and the definition Eq. (B3), we find that

$$\mathbb{T}_x\mathbb{T}_y\mathbb{T}_x^{-1}\mathbb{T}_y^{-1} \sim e^{2\pi i\nu}, \qquad (B7)$$

which is Eq. (8).

APPENDIX C: FILLING CONSTRAINTS IN GENERAL DIMENSION FROM THE TOPOLOGICAL ACTION OF SPT PHASES

Here we discuss the sense in which the microscopic filling can be computed from the IR theory in general spatial dimension d, generalizing the discussion of Secs. III A and III B. First of all, we know that there is a one-to-one correspondence between the 't Hooft anomaly of the IR theory and a SPT phase in $d + 1$ spatial dimensions. Since the SPT phase is topological, at long wavelengths a G_{IR} SPT phase in $d + 1$ spatial dimensions is described by a (Euclidean) topological action $S[M, A]$ that depends on the $(d + 2)$-dimensional space-time manifold M and a gauge field A for the G_{IR} symmetry. This action can be evaluated on any oriented space-time manifold M. (In some cases, there may be some additional structures required on M, for example, fermionic systems require a spin structure or spin$_c$ structure.) This action is purely imaginary (i.e., has no real part), does not depend on a metric on M, and provided that G_{IR} is a compact Lie group, depends only on the SPT phase and is independent of the microscopic details of the SPT ground state. In general, the action is defined only modulo $2\pi i$ since this does not affect the amplitude e^{-S}.

Let $\tau_1, \ldots, \tau_d \in G_{\mathrm{IR}}$ be the elements of the IR symmetry group corresponding to microscopic translations. We can define a generalization of the function α from the previous subsections in the following way. We set the space-time manifold M to be the $(d + 2)$-dimensional torus T^{d+2}. We require that over the first d of the $d + 2$ directions of the torus, the gauge field A is flat, with the fluxes through the associated 1-cycles given by $\tau_1, \ldots, \tau_d$. Then, over the final two directions on the torus, we require that there be a 2π flux of the U(1) subgroup of G_{IR} (generated by $\hat{Q}_{\mathrm{IR}}$) corresponding to the microscopic U(1) symmetry. Finally, we define $\alpha(\hat{Q}_{\mathrm{IR}}|\tau_1, \ldots, \tau_d) = -iS[M, A]$. One can show that this agrees with the previous definitions of α in the cases $d = 1, 2$.

Thus, our hypothesis for the form of the filling constraint is

$$\nu = \frac{1}{2\pi}\alpha(\hat{Q}_{\text{IR}}|\tau_1, \dots, \tau_d) \quad (\text{mod } 1). \qquad \text{(C1)}$$

We do not devise a physical derivation analogous to the $d = 1$ and $d = 2$ cases; however, we expect that this should be possible, since α should relate to the anomalous action of translations on a 2π flux, which in d spatial dimensions is a $(d-2)$-dimensional object.

We can extend this definition to include higher-form symmetries as follows. As we discuss in Sec. X, in this case we can include the possibility of symmetry fractionalization by considering the map $\rho: G_{\text{UV}} \to G_{\text{IR}}$ to be a map of higher groups, where G_{IR} contains all higher-form symmetries. We show that there is a G_{IR} gauge field A on T^{d+2} associated with ρ which generalizes the A above, so that $\nu = -iS[T^{d+2}, A]/2\pi$. First, we equip the $d+2$ torus T^{d+2} with a $\mathbb{Z}^d \times \text{U}(1)$ gauge field A' defined as follows. For $j = 1, \dots, d$, the jth 1-cycle C_j of T^{d+2} has $\int_{C_j} A'$ equal to the generator of the jth $\mathbb{Z}$ factor. On the last two coordinates, which form a T^2, we place a 2π flux in the U(1) factor. We then take $A = \rho(A')$.

Note that, as we mention above, in certain cases the topological action could depend on other data (such as a spin structure for fermionic systems). However, we expect that if the action still depends nontrivially on these additional data for the particular space-time manifold M and gauge field A considered here, this corresponds to cases where the microscopic symmetry $G_{\text{UV}} = \mathbb{Z}^d \times \text{U}(1)$ still has a nontrivial 't Hooft anomaly in d spatial dimensions. In this case, the system cannot be realized as a lattice system in d spatial dimensions with translation symmetry and on-site U(1) symmetry, only as a boundary of a SPT phase in $d + 1$ spatial dimensions. In such systems, the filling of the d-dimensional system does not need to be well defined.

APPENDIX D: PROJECTIVE REPRESENTATION OF LOOP GROUP ON MONOPOLE FROM 'T HOOFT ANOMALY

In this Appendix, we show that Eq. (19) follows from the anomaly Eq. (17).

Let us work on a four-dimensional space parametrized by coordinates (t, x, y, θ). We start by considering a configuration where the A_x and A_y components of the gauge field depend only on x and y, and the A_t, A_θ components depend only on z,t. Then, we integrate Eq. (17) over x and y. We obtain

$$\partial_\mu \tilde{j}^\mu = \frac{m\phi}{2\pi}(\partial_t A_\theta - \partial_\theta A_t), \qquad \text{(D1)}$$

where the index μ now varies only over t and θ, and we define $\phi = (1/2\pi)\int(\partial_x A_y - \partial_y A_x)dxdy$, which is the number of flux quanta passing through the system, and $\tilde{j}^\mu = \int j^\mu dxdy$. Effectively, what we do is compactify our

original system in two spatial dimensions to one in zero spatial dimensions. This 0D system has a 't Hooft anomaly that depends on ϕ, and it is canceled by inflow from a 2D SPT phase. Recall that such 't Hooft anomalies correspond to projective representations of the symmetry group. If we set $\phi = 1$, this corresponds to the projective representation of the symmetry group acting on a 2π flux.

In our case, where the symmetry group is $G = \text{LU}(1)$, we can express the projective representation through this central extension of the commutation algebra of the symmetry generators $\hat{n}(\theta)$, i.e.,

$$[\hat{n}(\theta), \hat{n}(\theta')] = c(\theta, \theta'), \qquad \text{(D2)}$$

where $c(\theta, \theta')$ is a c number (i.e., not an operator) to be determined.

The projective representation should be independent of the choice of Hamiltonian, so we are free to choose $H = \int f(\theta)\hat{n}(\theta)d\theta$ for some arbitrary test function $f(\theta)$, which amounts to coupling the zero Hamiltonian to the gauge field $A_t = f(\theta)$. Then we find that Heisenberg evolution $\hat{n}(\theta, t) = e^{itH}\hat{n}(\theta)e^{-itH}$ satisfies

$$\frac{d}{dt}\hat{n}(\theta, t) = i\int c(\theta', \theta)f(\theta')d\theta'. \qquad \text{(D3)}$$

We should compare this with the anomaly Eq. (D1), in which we choose A_θ to be independent of time, we have $\tilde{j}^\theta = (\delta H/\delta A_\theta) = 0$, and we identify the time component $\tilde{j}^t$ with the expectation value of $\hat{n}(\theta, t)$. We find that, to be consistent with Eq. (D3), it must be that

$$i\int c(\theta', \theta)f(\theta')d\theta' = -\frac{m\phi}{2\pi}\frac{d}{d\theta}f(\theta), \qquad \text{(D4)}$$

which is true for an arbitrary test function f only if

$$c(\theta, \theta') = -i\frac{m\phi}{2\pi}\delta'(\theta - \theta'). \qquad \text{(D5)}$$

Finally, to obtain the extra factor of q on the right-hand side of Eq. (19), we recall that Eq. (13) implies that ϕ, which we can interpret as the number of flux quanta of the IR U(1) symmetry generated by $\int \hat{n}(\theta)d\theta$, is q times the number of flux quanta of the microscopic U(1) symmetry.

APPENDIX E: PROJECTIVE REPRESENTATION OF LOOP GROUP IN A FERMI LIQUID

In this Appendix, we derive Eq. (19) in a Fermi liquid by working in terms of the semiclassical theory of band electrons. Strictly speaking, we treat only noninteracting electrons, but since our arguments are purely geometric and the result ultimately can be reduced to a contribution coming solely from the Fermi surface, we expect that

the same results hold in a Fermi liquid in light of the quasiparticle picture.

For clarity, in this Appendix we temporarily work in units where $\hbar \neq 1$. Then we can write Eq. (19) as

$$[\hat{n}(\theta), \hat{n}(\theta')] = -i \frac{\Phi m q}{(2\pi)^2 \hbar} \delta'(\theta - \theta'), \qquad \text{(E1)}$$

where $\Phi = \int B(\mathbf{x}) d^2\mathbf{x}$ is the total magnetic flux. Therefore, in the semiclassical theory, what we want to show is the Poisson bracket

$$\{n(\theta), n(\theta')\} = -\frac{\Phi m q}{(2\pi\hbar)^2} \delta'(\theta - \theta'). \qquad \text{(E2)}$$

Note that for each θ, $n(\theta)$ is a semiclassical observable, i.e., a real-valued function on the semiclassical phase space.

What we actually prove is the following equivalent statement:

$$\{n_f, n_g\} = -\frac{\Phi q m}{(2\pi\hbar)^2} \int f(\theta) \frac{dg(\theta)}{d\theta} d\theta \qquad \text{(E3)}$$

for any smooth functions f and g, where we define

$$n_f = \int f(\theta) n(\theta) d\theta. \qquad \text{(E4)}$$

We can relate n_f and n_g to the density $n(\mathbf{k}, \mathbf{x})$ of semiclassical particles according to

$$n_f = \int f_+(\mathbf{k}) n(\mathbf{k}, \mathbf{x}) d^2\mathbf{x} d^2\mathbf{k}, \qquad \text{(E5)}$$

where $f_+(\mathbf{k})$ is any smooth extension of f away from the Fermi surface such that $f_+(\mathbf{k}_F(\theta)) = f(\theta)$, where $\mathbf{k}_F(\theta)$ is the momentum of the Fermi surface as a function of θ. Note that by $n(\mathbf{k}, \mathbf{x})$ we mean that for *each* $\mathbf{k}$, $\mathbf{x}$, $n(\mathbf{k}, \mathbf{x})$ is a function of phase-space variables. If there are N particles in total, then the phase-space variables are labeled by $(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)}, \mathbf{k}^{(1)}, ..., \mathbf{k}^{(N)})$, and we can write

$$n(\mathbf{k}, \mathbf{x})(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)}, \mathbf{k}^{(1)}, ..., \mathbf{k}^{(N)})$$
$$= \sum_{i=1}^{N} \delta^2(\mathbf{k} - \mathbf{k}^{(i)}) \delta^2(\mathbf{x} - \mathbf{x}^{(i)}). \qquad \text{(E6)}$$

Here, $\mathbf{k}^{(j)}$ denotes the gauge-invariant momentum. That means that the Poisson brackets of the $\mathbf{x}^{(j)}$'s and $\mathbf{k}^{(j)}$'s get modified due to the magnetic field, as we discuss below.

Now we can compute $\{n_f, n_g\}$. We first do it for a single-particle phase space $(\mathbf{k}^{(1)}, \mathbf{x}^{(1)})$. Then we have

$$n_f(\mathbf{k}^{(1)}, \mathbf{x}^{(1)}) = f_+(\mathbf{k}^{(1)}), \qquad \text{(E7)}$$

and hence,

$$\{n_f, n_g\} = \frac{\partial f_+(\mathbf{k}^{(1)})}{\partial k_\alpha^{(1)}} \frac{\partial g_+(\mathbf{k}^{(1)})}{\partial k_\beta^{(1)}} \{k_\alpha^{(1)}, k_\beta^{(1)}\} \qquad \text{(E8)}$$

$$= -B(\mathbf{x}^{(1)})[\nabla_{\mathbf{k}^{(1)}} f_+(\mathbf{k}^{(1)})] \times [\nabla_{\mathbf{k}^{(1)}} g_+(\mathbf{k}^{(1)})], \qquad \text{(E9)}$$

where we use the fact that the momenta $\mathbf{k}^{(1)}$ obey the magnetic algebra [134]

$$\{k_\alpha^{(1)}, k_\beta^{(1)}\} = -qB(\mathbf{x}^{(1)})\varepsilon_{\alpha\beta} \qquad \text{(E10)}$$

(here we choose the convention that $\mathbf{k}$ carries units of momentum, not inverse length). For general N, we have

$$\{n_f, n_g\} = -q \int B(\mathbf{x}) n(\mathbf{x}, \mathbf{k})[\nabla_{\mathbf{k}} f_+(\mathbf{k})]$$
$$\times [\nabla_{\mathbf{k}} g_+(\mathbf{k})] d^2\mathbf{x} d^2\mathbf{k}, \qquad \text{(E11)}$$

where in this equation both the left- and right-hand sides have an implicit dependence on the phase-space variables $(\mathbf{k}^{(1)}, ..., \mathbf{k}^{(N)}, \mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)})$ on the right-hand side through $n(\mathbf{x}, \mathbf{k})$.

Now let us write $n(\mathbf{x}, \mathbf{k}) = n_0(\mathbf{k}) + \delta n(\mathbf{x}, \mathbf{k})$, where $n_0(\mathbf{k})$ has no implicit dependence on the phase-space variables and represents the equilibrium distribution of particles in the ground state (keeping in mind the Pauli exclusion principle). For low-energy configurations, the contribution of δn to Eq. (E11) is negligible compared to n_0. Hence, we can replace n with n_0 in Eq. (E11) and we find

$$\{n_f, n_g\} = -q\Phi \int n_0(\mathbf{k})[\nabla_{\mathbf{k}} f_+(\mathbf{k})] \times [\nabla_{\mathbf{k}} g_+(\mathbf{k})] d^2\mathbf{k}, \qquad \text{(E12)}$$

$$= -q\Phi \int n_0(\mathbf{k}) \nabla_{\mathbf{k}} \times [f_+(\mathbf{k}) \nabla_{\mathbf{k}} g_+(\mathbf{k})] d^2\mathbf{k}. \qquad \text{(E13)}$$

Now we observe that $n_0(\mathbf{k}) = 1/(2\pi\hbar)^2$ in the occupied region of the Brillouin zone, and 0 outside. So by Stokes theorem, we obtain

$$\{n_f, n_g\} = -\frac{qm\Phi}{(2\pi\hbar)^2} \oint_{\text{Fermi surface}} [f_+(\mathbf{k}) \nabla_{\mathbf{k}} g_+(\mathbf{k})] \cdot d\mathbf{k} \qquad \text{(E14)}$$

$$= -\frac{qm\Phi}{(2\pi\hbar)^2} \int f(\theta) \frac{dg(\theta)}{d\theta} d\theta, \qquad \text{(E15)}$$

where $m = \pm 1$ (relative to an arbitrary choice of orientation of the Fermi surface, $m = +1$ or -1 corresponds to which side of the Fermi surface the occupied region of the Brillouin zone is on). This completes the derivation of Eq. (E3).

APPENDIX F: PROOF OF THEOREM 1

In this Appendix, we give a proof of Theorem 1 from Sec. IV. The result, in fact, follows directly from the following mathematical property of the function α defined in Sec. III A for spatial dimension $d = 1$, in Sec. III B for $d = 2$, and in Appendix C for general d.

Lemma 1: Let $\mathcal{A}$ be an Abelian subgroup of G_{IR}. The function $\alpha(\mathcal{Q}|\tau_1, \ldots, \tau_d)$ defined in Sec. III is linear in each argument when restricted to $\mathcal{A}$. That is, for $a_1, \ldots, a_d, a_1' \in \mathcal{A}$, and where $\mathcal{Q}$ is the infinitesimal generator of a U(1) subgroup of $\mathcal{A}$, then we have

$$\alpha(\mathcal{Q}|a_1 a_1', a_2, \ldots, a_d) = \alpha(\mathcal{Q}|a_1, a_2, \ldots, a_d) \\ + \alpha(\mathcal{Q}|a_1', a_2, \ldots, a_d) \quad (\mathrm{mod}\ 2\pi), \qquad (\mathrm{F1})$$

and similar for each of the arguments after the "|".

Proof.—It is easy to verify this given the concrete definitions of α in the cases $d = 1$ and $d = 2$. In general d, where α is defined as discussed in Appendix C, the result follows from the hypothesis that the topological action is a cobordism invariant [53,55,59]. ∎

Lemma 1 is all we need to give the following:

Proof of Theorem 1.—Let $\tau_1, \ldots, \tau_d \in G_{\mathrm{IR}}$ be the IR symmetries corresponding to the microscopic translation symmetries and let $\hat{Q}_{\mathrm{IR}}$ be the infinitesimal generator of the IR symmetry corresponding to the microscopic U(1) rotation symmetry. Since the microscopic translations commute with each other and with charge, $\tau_1, \ldots, \tau_d$ and $\hat{Q}_{\mathrm{IR}}$ must all commute. Let $\mathcal{A}$ be a maximal Abelian subgroup of G_{IR} (that is, a subgroup that cannot be enlarged as a subgroup of G_{IR} while remaining Abelian) that contains $\tau_1, \ldots, \tau_d$ and the subgroup generated by $\hat{Q}_{\mathrm{IR}}$. One can show [136] that $\mathcal{A}$ is a compact Abelian Lie subgroup of G_{IR}. Let $C_{\mathcal{A}}$ be the number of connected components of $\mathcal{A}$, and let $\mathcal{A}_0 \leq \mathcal{A}$ be the connected component of the identity element. Then, since $[a]^{C_{\mathcal{A}}} = 1$ for any element $[a] \in \mathcal{A}/\mathcal{A}_0$, we conclude that $a^{C_{\mathcal{A}}} \in \mathcal{A}_0$ for any $a \in \mathcal{A}$.

Now since $\mathcal{A}$ is a compact Abelian Lie group, the connected component $\mathcal{A}_0$ must be isomorphic to U(1)^r for some r. Hence, we can write elements of $\mathcal{A}_0$ as $(\theta_1, \ldots, \theta_r)$ for angular variables $\theta_1, \ldots, \theta_r$. We now return to the function $\alpha(\hat{Q}_{\mathrm{IR}}|\tau_1, \ldots, \tau_d)$ that computes the filling. Henceforth, we pick a given $\hat{Q}_{\mathrm{IR}}$ and hold it fixed; therefore, we write simply $\alpha(\tau_1, \ldots, \tau_d)$. The linearity of α established by Lemma 1 implies that α is continuous in each argument, and, moreover, that if $a_1 = (\theta_1, \ldots, \theta_r) \in \mathcal{A}_0$, then

$$\alpha(a_1, \ldots, a_d) = \sum_{j=1}^{r} \theta_j m_j(a_2, \ldots, a_d) \quad (\mathrm{mod}\ 2\pi) \quad (\mathrm{F2})$$

for some continuous integer function $m_j : \mathcal{A}^{\times(d-1)} \to \mathbb{Z}$, where we suppress the dependences on $\hat{Q}_{\mathrm{IR}}$, which we hold fixed. But now continuity implies that m_j is zero if any of its arguments are in $\mathcal{A}_0$. We therefore conclude that $\alpha(a_1, \ldots, a_d) = 0 \ (\mathrm{mod}\ 2\pi)$ if any *two* of its arguments are in $\mathcal{A}_0$. Therefore,

$$(C_{\mathcal{A}})^2 \alpha(\tau_1, \ldots, \tau_d) = \alpha(\tau_1^{C_{\mathcal{A}}}, \tau_2^{C_{\mathcal{A}}}, \tau_3, \ldots, \tau_d) = 0 \\ (\mathrm{mod}\ 2\pi), \qquad (\mathrm{F3})$$

and hence, the filling ν is an integer multiple of $1/(C_{\mathcal{A}})^2$.

Finally, to see that we can take $N_{G_{\mathrm{IR}}}$ finite for any given compact Lie group G_{IR}, we note [138] that for any such G_{IR}, there are only finitely many conjugacy classes of maximal Abelian Lie subgroups $\mathcal{A} \leq G_{\mathrm{IR}}$. ∎

For an alternative perspective that gives more precise bounds on $N_{G_{\mathrm{IR}}}$, see Appendix G.

APPENDIX G: BOUNDS ON FILLING DENOMINATORS FOR COMPACT G_{IR}

Theorem 1 says that a compact G_{IR} is inconsistent with an irrational filling fraction. In this Appendix, we discuss how, given an assumed G_{IR}, in a fixed space dimension d, one can compute a denominator $N_{G_{\mathrm{IR}}}$ such that G_{IR} is consistent only with rational filling fractions of the form $m/N_{G_{\mathrm{IR}}}$, for $m \in \mathbb{Z}$. We show the following:

Theorem 2: In $d > 1$ space dimensions, for any compact emergent symmetry G_{IR} (which could be a subgroup of the full group of emergent symmetries) containing the translation generators $\tau_1, \ldots, \tau_d$ and the subgroup U(1) corresponding to microscopic particle-number conservation, the maximum denominator $N_{G_{\mathrm{IR}}}$ divides the largest order of elements in the torsion part of the (co)bordism group $\Omega_{d+2}(G_{\mathrm{IR}})^{\mathrm{tors}} = \Omega^{d+2}(G_{\mathrm{IR}})^{\mathrm{tors}}$, i.e., the group of $(d+1)$-space-dimensional SPT phases such that for some n, a stack of n copies of the SPT phase is trivial.

Theorem 3: In $d = 2$ space dimensions, assuming a compact emergent symmetry G_{IR} satisfying the conditions of Theorem 2 and furthermore of the form $G_{\mathrm{IR}} = H \times \mathrm{U}(1)$, the maximum denominator $N_{G_{\mathrm{IR}}}$ divides $|\pi_0(G_{\mathrm{IR}})| \cdot |\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}|$, where $\pi_0(G_{\mathrm{IR}})$ is the set of connected components of G_{IR} and $\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}$ is the torsion part of its fundamental group. If there is also a finite 1-form symmetry K, then $N_{G_{\mathrm{IR}}}$ divides $|\pi_0(G_{\mathrm{IR}})| \cdot |\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}| \cdot |K|$.

Theorem 4: In $d = 3$ space dimensions, assuming a compact emergent symmetry G_{IR} satisfying the conditions of Theorem 2 and furthermore of the form $G_{\mathrm{IR}} = H \times \mathrm{U}(1)$, the maximum denominator $N_{G_{\mathrm{IR}}}$ divides $|\pi_0(G_{\mathrm{IR}})| \cdot \mathrm{GCD}[|\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}|, |\pi_0(G_{\mathrm{IR}})|]$, where $\pi_0(G_{\mathrm{IR}})$ is the set of connected components of G_{IR} and $\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}$ is

the torsion part of its fundamental group and GCD denotes the greatest common divisor. In particular, for connected G_{IR}, $N_{G_{\text{IR}}} = 1$, so a compact connected emergent symmetry is incompatible with any fractional filling. With a finite 1-form symmetry K, then $N_{G_{\text{IR}}}$ divides $|\pi_0(G_{\text{IR}})| \cdot \text{GCD}[|\pi_1(G_{\text{IR}})^{\text{tors}}|, |\pi_0(G_{\text{IR}})|] \cdot \text{GCD}[|K|, \pi_0(G_{\text{IR}})]$. Furthermore, if there is a finite 2-form symmetry J, then $N_{G_{\text{IR}}}$ divides $|\pi_0(G_{\text{IR}})| \cdot \text{GCD}[|\pi_1(G_{\text{IR}})^{\text{tors}}|, |\pi_0(G_{\text{IR}})|] \cdot \text{GCD}[|K|, |\pi_0(G_{\text{IR}})|] \cdot |J|$.

We note that since $\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}$ is finite, Theorem 1 follows from Theorem 2.

Proof of Theorem 2.—As we argue in Appendix C, the filling fraction associated with G_{IR} may be computed from the emergent G_{IR} 't Hooft anomaly as well as the data of how U(1) and translation symmetries are realized in G_{IR}. In particular, the translation generators $\tau_1, \ldots, \tau_d$ and the U(1) generator $\hat{Q}$ define a G_{IR} gauge field on $T^{d+2} = T^d \times T^2$, where the τ_j define the fluxes through the d coordinate 1-cycles of T^d and $\hat{Q}$ defines a 2π flux of T^2. The emergent 't Hooft anomaly is associated with a cobordism invariant of G_{IR} gauge fields on $d + 2$ manifolds [the partition function of the $[(d + 1) + 1]$D SPT phase associated with the 't Hooft anomaly], and that cobordism invariant evaluated on this particular background equals $e^{2\pi i \nu}$.

Let us denote the group of cobordism invariants by $\Omega^{d+2}(G_{\text{IR}})$. In the notation of Ref. [55], it is $\Omega^{d+2}_{\text{spin}}(B(G_{\text{IR}}/\mathbb{Z}^F_2), \xi)$, where ξ is an orientable bundle whose second Stiefel-Whitney class classifies the extension of $G_{\text{IR}}/\mathbb{Z}^F_2$ by fermion parity $\mathbb{Z}^F_2$. The group $\Omega^{d+2}(G_{\text{IR}})$ sits in a (split) short exact sequence [59]

$$\text{Hom}\left[\Omega_{d+3}(G_{\text{IR}}), \mathbb{Z}\right] \longrightarrow \Omega^{d+2}(G_{\text{IR}})$$
$$\downarrow$$
$$\text{Hom}\left[\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}, U(1)\right]$$

where $\Omega_{d+2}(G_{\text{IR}})$ and $\Omega_{d+3}(G_{\text{IR}})$ are the associated bordism groups of manifolds, and $\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}$ indicates the torsion subgroup of $\Omega_{d+2}(G_{\text{IR}})$, elements x for which there exists an n such that $nx = 0$.

The kernel of this extension $\text{Hom}[\Omega_{d+3}(G_{\text{IR}}), \mathbb{Z}]$ is characterized by generalized Chern-Simons invariants described by Chern-Weil forms evaluated on a bounding $d + 3$ manifold M with $\partial M = T^{d+2}$. We show that for $d > 1$ these invariants do not contribute to the filling fraction. More precisely, the map $\Omega^{d+2}(G_{\text{IR}}) \to U(1)$ given by evaluating the cobordism invariant on such a T^{d+2} background as we describe above sends the kernel to zero. Therefore, the maximum denominator $N_{G_{\text{IR}}}$ divides $|\text{Hom}[\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}, U(1)]| = |\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}|$, which is finite. This also gives an alternative proof of Theorem 1. In fact, $N_{G_{\text{IR}}}$ divides the largest order of elements in $\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}$, which we use below. This group is

isomorphic to its cobordism partner $\Omega^{d+2}(G_{\text{IR}})^{\text{tors}}$ because the short exact sequence above splits.

Now we show that for $d > 1$ the generalized Chern-Simons invariants do not contribute to the filling fraction. For even d, there are no generalized Chern-Simons invariants and we are done, so let us consider odd d instead. We use a flat M so that there are no metric contributions to the Chern-Simons invariant. We also use the fact that the remaining Chern-Simons invariants are defined for G_{IR} bundles by taking a trace in a unitary representation $R: G_{\text{IR}} \to U(n)$ and is therefore equal to a Chern-Simons invariant $U(n)$ bundle associated with it by the map R. Then, we use the fact that $\tau_1, \ldots, \tau_d$ and the U(1) generator $\hat{Q}$ commute in G_{IR}; hence, their images in $U(n)$ also commute and can be simultaneously diagonalized. This reduces the calculation to an Abelian Chern-Simons invariant for the subgroup $U(1)^n$ of diagonal matrices in $U(n)$. Taking the trace reduces this to a calculation for a U(1) bundle.

We can extend our U(1) bundle from $T^{d+2} = T^d \times T^2$ to $D^2 \times T^{d-1} \times T^2$, where D^2 is a disk whose boundary is the first coordinate cycle of T^d. The extended gauge field A is defined so that it has curvature only along D^2 [integrating to $\text{Tr}_R(\tau_1)$ as required by the extension] and along T^2 (where it has some Chern number). The Chern-Simons invariant is proportional to

$$\int_{D^2 \times T^{d-1} \times T^2} (dA)^{\frac{d+3}{2}}.$$

However, since the 2-form dA has only nonzero components in four directions, D^2 and T^2, its power in the integrand above vanishes if $d + 3 > 4$, i.e., if $d > 1$ [141]. This proves the claim. ∎

Proof of Theorem 3.—We can get a handle on the maximum denominator $N_{G_{\text{IR}}}$, which is proved above to divide $|\Omega_{d+2}(G_{\text{IR}})^{\text{tors}}|$, using the Atiyah-Hirzebruch spectral sequence for computing the latter. This gives us an upper bound

$$N_{G_{\text{IR}}} \text{ divides } \prod_{j=0}^{d+2} |H_j(BG^b_{\text{IR}}, \Omega^{\text{spin}}_{d+2-j})^{\text{tors}}|,$$

where $G^b_{\text{IR}} = G_{\text{IR}}/\mathbb{Z}^F_2$, and Ω^{spin}_k is the usual spin bordism group, which in the low degrees of physical interest is

$$\Omega^{\text{spin}}_0 = \mathbb{Z},$$
$$\Omega^{\text{spin}}_1 = \mathbb{Z}_2,$$
$$\Omega^{\text{spin}}_2 = \mathbb{Z}_2,$$
$$\Omega^{\text{spin}}_3 = 0,$$
$$\Omega^{\text{spin}}_4 = \mathbb{Z},$$
$$\Omega^{\text{spin}}_5 = 0.$$

This upper bound can be further improved by taking into account the known differentials of the dual spectral sequence for cobordism $\Omega^{d+2}(G_{\mathrm{IR}})$, which has isomorphic torsion, $\Omega^{d+2}(G_{\mathrm{IR}})^{\mathrm{tors}} = \Omega_{d+2}(G_{\mathrm{IR}})^{\mathrm{tors}}$. See Ref. [142] for a review.

Another thing we can do is consider the special case $G_{\mathrm{IR}} = H \times \mathrm{U}(1)$, where $\mathrm{U}(1)$ and H is compact. Because $\tau_1, \ldots, \tau_d$ and $\hat{Q}$ must commute, they are always contained in such a group. In this case, by taking considering $(d+2)$-manifolds of the form $M \times T^2$ with an arbitrary G_{IR} bundle on M and a $\mathrm{U}(1)$ bundle with Chern number 1 on T^2, our $(d+2)$-dimensional cobordism invariant defines a d-dimensional cobordism invariant for M. By similar arguments as above, this gives us the further constraint that $N_{G_{\mathrm{IR}}}$ divides the element of highest order in $\Omega_d(G_{\mathrm{IR}})^{\mathrm{tors}} = \Omega^d(G_{\mathrm{IR}})^{\mathrm{tors}}$.

For $d = 2$, using the Atiyah-Hirzebruch spectral sequence for $\Omega^2(G_{\mathrm{IR}})$, we find in the case $G_{\mathrm{IR}} = H \times \mathrm{U}(1)$, $N_{G_{\mathrm{IR}}}$ divides

$$|\Omega_2^{\mathrm{spin}}| \cdot |H^1(BG_{\mathrm{IR}}^b, \Omega_1^{\mathrm{spin}})| \cdot |H^3(BG_{\mathrm{IR}}^b, \Omega_0^{\mathrm{spin}})^{\mathrm{tors}}|.$$

One can show that since the fermion parity is a subgroup of the $\mathrm{U}(1)$ factor, that the first two factors do not contribute to the spin bordism because the spectral sequence has a nonzero differential there. The third part is $|\pi_0(G_{\mathrm{IR}})| \cdot |\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}|$ torsion, so

$$N_{G_{\mathrm{IR}}} \text{ divides } |\pi_0(G_{\mathrm{IR}})| \cdot |\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}|.$$

We note that we can often get even better bounds if we better understand the cobordism invariants for G_{IR}. For example, if $G_{\mathrm{IR}} = \mathbb{Z}_2^r \times \mathrm{U}(1)$, one can show the maximum denominator is $N_{G_{\mathrm{IR}}} = 2$, even though this group has 2^r components.

To include a finite 1-form symmetry K, there is one more term which contributes to the spectral sequence, $H^3(B^2 K, \mathbb{Z}) = K$, which is $|K|$ torsion. ∎

Proof of Theorem 4.—For $d = 3$, $G_{\mathrm{IR}} = H \times \mathrm{U}(1)$, the contributions to the Atiyah-Hirzebruch spectral sequence for $\Omega^3(G_{\mathrm{IR}})$ are

$$|H^1(BG_{\mathrm{IR}}^b, \Omega_2^{\mathrm{spin}})| \cdot |H^2(BG_{\mathrm{IR}}^b, \Omega_1^{\mathrm{spin}})| \cdot |H^4(BG_{\mathrm{IR}}^b, \Omega_0^{\mathrm{spin}})^{\mathrm{tors}}|.$$

As in $d = 2$, the first factor does not contribute. The second factor has one piece that can contribute, coming from the map $H^2(BU(1), \mathbb{Z}) \to H^2(BG_{\mathrm{IR}}^b, \mathbb{Z}_2)$, but it turns out to give a spin Chern-Simons term for the $\mathrm{U}(1)$ factor and we can disregard it; i.e., it does not contribute to the torsion in $\Omega^3(G_{\mathrm{IR}})$. The third piece can be bounded using the Serre spectral sequence for the extension

$$G_{\mathrm{IR}}^{b,0} \to G_{\mathrm{IR}}^b \to \pi_0(G_{\mathrm{IR}}^b) = \pi_0(G_{\mathrm{IR}}),$$

from which we find it is $|\pi_0(G_{\mathrm{IR}})| \cdot \mathrm{GCD}[|\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}|, |\pi_0(G_{\mathrm{IR}})|]$ torsion, so

$N_{G_{\mathrm{IR}}}$ divides $|\pi_0(G_{\mathrm{IR}})| \cdot \mathrm{GCD}[|\pi_1(G_{\mathrm{IR}})^{\mathrm{tors}}|, |\pi_0(G_{\mathrm{IR}})|]$.

Note that when G_{IR} is connected, we reestablish the result $N_{G_{\mathrm{IR}}} = 1$: connected, compact G_{IR} are incompatible with fractional filling in $d = 3$.

When there is also a finite 1-form symmetry K, it contributes a term to $H^4(BG_{\mathrm{IR}}^b, \Omega_0^{\mathrm{spin}})$ via $H^1(B\pi_0(G_{\mathrm{IR}}^b), H^3(B^2 K, \mathbb{Z}))$, which is $\mathrm{GCD}[|\pi_0(G_{\mathrm{IR}}^b)|, |K|]$ torsion.

When there is also a finite 2-form symmetry J, it contributes a term to $H^4(BG_{\mathrm{IR}}^b, \Omega_0^{\mathrm{spin}})$ via $H^4(B^2 J, \mathbb{Z}) = J$. ∎

Finally, we note that the bounds so derived hold for *every subgroup* G_{IR} of the full emergent symmetry, so long as it contains the translation generators $\tau_1, \ldots, \tau_d$ and the $\mathrm{U}(1)$ particle-number symmetry. Thus, one does not need to know the entire emergent symmetry to obtain a bound. Further, by varying G_{IR}, one can sometimes find a better bound for the problem at hand.

APPENDIX H: ASYMPTOTIC ANALYSIS FOR QUANTUM OSCILLATIONS

In this Appendix, we derive the asymptotic form for quantum oscillations stated in Sec. VI C 3. Abstracting out from the specific details, the situation is that we have some functional $\mathcal{F}[f]$, where f is a function of a single variable into $\mathbb{R}/\mathbb{Z}$ that can be expressed as

$$f(x) = tg(x), \tag{H1}$$

and we wish to extract the asymptotic dependence of $\mathcal{F}[f]$ as $t \to \infty$ while keeping the function g fixed.

First, we imagine approximating f by its values on the discrete points $x_1, \ldots, x_k$. Then we can expand $\mathcal{F}$ as a Fourier series

$$\mathcal{F}[f] = \sum_{\mathbf{n} \in \mathbb{Z}^k} a_{\mathbf{n}} e^{2\pi i t (n_1 g(x_1) + \cdots + n_k g(x_k))}, \tag{H2}$$

where the sum is over integer vectors $\mathbf{n} \in \mathbb{Z}^k$. We can rewrite this as

$$\mathcal{F}[f] = \mathcal{S}^{(0)} + \sum_{n \neq 0} \mathcal{S}_n^{(1)} + \sum_{n_1 \neq 0, n_2 \neq 0}^{\infty} \mathcal{S}_{n_1, n_2}^{(2)} + \cdots, \tag{H3}$$

where $\mathcal{S}^{(0)} = a_{\mathbf{0}}$ and

$$\mathcal{S}_n^{(1)} = \sum_{j=1}^{k} a_{(j:n)} e^{2\pi i t m g(x_j)}, \tag{H4}$$

$$\mathcal{S}_{n_1, n_2}^{(2)} = \sum_{j_1 \neq j_2} a_{(j_1 : n_1, j_2 : n_2)} e^{2\pi i t (n_1 g(x_1) + i t n_2 g(x_2))}, \tag{H5}$$

$$\mathcal{S}_{n_1, n_2, n_3}^{(3)} = \cdots, \tag{H6}$$

where $(j_1 : m_1, j_2 : m_2)$, for example, is the vector $\mathbf{n}$ obtained by setting $n_{j_1} = m_1$ and $n_{j_2} = m_2$ and $n_j = 0$ for $j \notin \{m_1, m_2\}$. By taking the limit as $k \to \infty$ and the x_j's become dense, we find that $\mathcal{S}_\ell \to \mathcal{I}_\ell$, where

$$\mathcal{I}_n^{(1)} = \int dx a_n(x) e^{2\pi i t n g(x)}, \tag{H7}$$

$$\mathcal{I}_{n_1, n_2}^{(2)} = \int dx_1 dx_2 a_{n_1, n_2}(x_1, x_2) e^{2\pi i t (n_1 g(x_1) + n_2 g(x_2))}, \tag{H8}$$

$$\mathcal{I}_{n_1, n_2, n_3}^{(3)} = \cdots. \tag{H9}$$

Now we invoke the theory of stationary phase integrals which tells us that, for a function $h(\mathbf{x})$ of a q-dimensional variable $\mathbf{x}$, as $t \to \infty$,

$$\int d^q \mathbf{x} \varphi(\mathbf{x}) e^{2\pi i t h(\mathbf{x})} \sim \sum_{\mathbf{x}_* \in \mathbf{\Sigma}} c_{\mathbf{x}_*} t^{-q/2} e^{2\pi i t f(\mathbf{x}_*)} \tag{H10}$$

for some constants $c_{\mathbf{x}_*}$ and where the sum is over the set $\mathbf{\Sigma}$ of solutions to $\nabla f(\mathbf{x}) = 0$. Hence, we find that

$$\mathcal{I}_n^{(1)} \sim c_n^{(1)} t^{-1/2} \sum_{x_* \in \Sigma} e^{2\pi i t n g(x_*)}, \tag{H11}$$

$$\mathcal{I}_{n_1, n_2}^{(2)} \sim c_{n_1, n_2}^{(2)} t^{-1} \sum_{x_*, x_*' \in \Sigma} e^{2\pi i t (n_1 g(x_*) + n_2 g(x_*'))}, \tag{H12}$$

$$\mathcal{I}_{n_1, n_2, n_3}^{(3)} \sim \cdots, \tag{H13}$$

where Σ is the set of solutions to $g'(x) = 0$. The oscillatory part of each of these expressions is consistent with a quasiperiodic function of t with base frequencies $\omega(x_*) = 2\pi g(x_*)$, $x_* \in \Sigma$.

---

[1] E. Lieb, Theodore Schultz, and Daniel Mattis, *Two Soluble Models of an Antiferromagnetic Chain*, Ann. Phys. (N.Y.) **16**, 407 (1961).

[2] M. Oshikawa, M. Yamanaka, and I. Affleck, *Magnetization Plateaus in Spin Chains: "Haldane Gap" for Half-Integer Spins*, Phys. Rev. Lett. **78**, 1984 (1997).

[3] M. Oshikawa, *Commensurability, Excitation Gap, and Topology in Quantum Many-Particle Systems on a Periodic Lattice*, Phys. Rev. Lett. **84**, 1535 (2000).

[4] M. B. Hastings, *Sufficient Conditions for Topological Order in Insulators*, Europhys. Lett. **70**, 824 (2005).

[5] G. Misguich, C. Lhuillier, M. Mambrini, and P. Sindzingre, *Degeneracy of the Ground-State of Antiferromagnetic Spin-1/2 Hamiltonians*, Eur. Phys. J. B **26**, 167 (2002).

[6] J. M. Luttinger, *Fermi Surface and Some Simple Equilibrium Properties of a System of Interacting Fermions*, Phys. Rev. **119**, 1153 (1960).

[7] M. Oshikawa, *Topological Approach to Luttinger's Theorem and the Fermi Surface of a Kondo Lattice*, Phys. Rev. Lett. **84**, 3370 (2000).

[8] T. Senthil, S. Sachdev, and M. Vojta, *Fractionalized Fermi Liquids*, Phys. Rev. Lett. **90**, 216403 (2003).

[9] T. Senthil, M. Vojta, and S. Sachdev, *Weak Magnetism and Non-Fermi Liquids near Heavy-Fermion Critical Points*, Phys. Rev. B **69**, 035111 (2004).

[10] A. Paramekanti and A. Vishwanath, *Extending Luttinger's Theorem to Z_2 Fractionalized Phases of Matter*, Phys. Rev. B **70**, 245118 (2004).

[11] P. Bonderson, M. Cheng, K. Patel, and E. Plamadeala, *Topological Enrichment of Luttinger's Theorem*, arXiv:1601.07902.

[12] H. Watanabe, H. C.F Po, A. Vishwanath, and M. Zaletel, *Filling Constraints for Spin-Orbit Coupled Insulators in Symmorphic and Nonsymmorphic Crystals*, Proc. Natl. Acad. Sci. U.S.A. **112**, 14551 (2015).

[13] Y.-M. Lu, Y. Ran, and M. Oshikawa, *Filling-Enforced Constraint on the Quantized Hall Conductivity on a Periodic Lattice*, Ann. Phys. (Amsterdam) **413**, 168060 (2020).

[14] N. Bultinck and M. Cheng, *Filling Constraints on Fermionic Topological Order in Zero Magnetic Field*, Phys. Rev. B **98**, 161119(R) (2018).

[15] M. Cheng, M. Zaletel, M. Barkeshli, A. Vishwanath, and P. Bonderson, *Translational Symmetry and Microscopic Constraints on Symmetry-Enriched Topological Phases: A View from the Surface*, Phys. Rev. X **6**, 041068 (2016).

[16] G. Y. Cho, C.-T. Hsieh, and S. Ryu, *Anomaly Manifestation of Lieb-Schultz-Mattis Theorem and Topological Phases*, Phys. Rev. B **96**, 195105 (2017).

[17] C.-M. Jian, Z. Bi, and C. Xu, *Lieb-Schultz-Mattis Theorem and Its Generalizations from the Perspective of the Symmetry-Protected Topological Phase*, Phys. Rev. B **97**, 054412 (2018).

[18] M. A. Metlitski and R. Thorngren, *Intrinsic and Emergent Anomalies at Deconfined Critical Points*, Phys. Rev. B **98**, 085140 (2018).

[19] Y. Yao and M. Oshikawa, *Generalized Boundary Condition Applied to Lieb-Schultz-Mattis-Type Ingappabilities and Many-Body Chern Numbers*, Phys. Rev. X **10**, 031008 (2020).

[20] X.-Y. Song, Y.-C. He, A. Vishwanath, and C. Wang, *Electric Polarization as a Nonquantized Topological Response and Boundary Luttinger Theorem*, arXiv:1909.08637.

[21] A perhaps more standard definition of "compressible" would be that the partial derivative of the particle number with respect to the chemical potential is nonzero; however, we do not use this definition in this paper.

[22] P. Coleman, I. Paul, and J. Rech, *Sum Rules and Ward Identities in the Kondo Lattice*, Phys. Rev. B **72**, 094430 (2005).

[23] S. Powell, S. Sachdev, and H. P. Büchler, *Depletion of the Bose-Einstein Condensate in Bose-Fermi Mixtures*, Phys. Rev. B **72**, 024534 (2005).

[24] L. Huijse and S. Sachdev, *Fermi Surfaces and Gauge-Gravity Duality*, Phys. Rev. D **84**, 026001 (2011).

[25] M. Potthoff, *Non-Perturbative Construction of the Lut-tinger-Ward Functional*, Condens. Matter Phys. **9,** 557 (2006).

[26] For most of the paper, we focus on the most familiar kind of emergent symmetry—what has retroactively been renamed "0-form symmetry," i.e., symmetries that act everywhere in space-time at once. Recently, the importance of so-called "higher-form symmetries" has started to be understood [27]. Loosely speaking, the restriction to the ordinary 0-form symmetries means that there are no emergent "fractionalized excitations." We say a pointlike excitation is "fractionalized" if it cannot be created by a pointlike creation operator, but rather is created at the end points of an open string by an operator supported on the string. There are similar definitions for higher-dimensional excitations. Fractionalized excitations emerge, for instance, in phases where there is topological order and/or deconfined emergent gauge fields. For such systems, there are emergent higher-form symmetries that play an important role, but in the interest of simplifying the exposition, we defer discussion of this point to later, and for the moment, we assume that all the emergent symmetries are 0-form symmetries. We eventually return to this point in Sec. X.

[27] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, *Generalized Global Symmetries*, J. High Energy Phys. 02 (2015) 172.

[28] X. Chen, Z.-X. Liu, and X.-G. Wen, *Two-Dimensional Symmetry-Protected Topological Orders and Their Protected Gapless Edge Excitations*, Phys. Rev. B **84,** 235141 (2011).

[29] D. V. Else and C. Nayak, *Classifying Symmetry-Protected Topological Phases through the Anomalous Action of the Symmetry on the Edge*, Phys. Rev. B **90,** 235137 (2014).

[30] G. 't Hooft, in *NATO Advanced Study Institute Seminar Series* (Springer, Boston, 1980), Vol. 59, p. 135.

[31] A. Kapustin and R. Thorngren, *Anomalous Discrete Symmetries in Three Dimensions and Group Cohomology*, Phys. Rev. Lett. **112,** 231602 (2014).

[32] Z.-C. Gu and X.-G. Wen, *Tensor-Entanglement-Filtering Renormalization Approach and Symmetry-Protected Topological Order*, Phys. Rev. B **80,** 155131 (2009).

[33] T. Senthil, *Symmetry-Protected Topological Phases of Quantum Matter*, Annu. Rev. Condens. Matter Phys. **6,** 299 (2015).

[34] X.-G. Wen, *Colloquium: Zoo of Quantum-Topological Phases of Matter*, Rev. Mod. Phys. **89,** 041004 (2017).

[35] M. Z. Hasan and C. L. Kane, *Colloquium: Topological Insulators*, Rev. Mod. Phys. **82,** 3045 (2010).

[36] F. D. M. Haldane, *Continuum Dynamics of the 1-D Heisenberg Antiferromagnet: Identification with the O(3) Nonlinear Sigma Model*, Phys. Lett. **93A,** 464 (1983).

[37] F. D. M. Haldane, *Nonlinear Field Theory of Large-Spin Heisenberg Antiferromagnets: Semiclassically Quantized Solitons of the One-Dimensional Easy-Axis Néel State*, Phys. Rev. Lett. **50,** 1153 (1983).

[38] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, *Entanglement Spectrum of a Topological Phase in One Dimension*, Phys. Rev. B **81,** 064439 (2010).

[39] There can also be 't Hooft anomalies associated with "invertible" topological phases which do not require any symmetry to protect them.

[40] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, *Topological Field Theory of Time-Reversal Invariant Insulators*, Phys. Rev. B **78,** 195424 (2008).

[41] F. Pollmann, E. Berg, A. M. Turner, and M. Oshikawa, *Symmetry Protection of Topological Phases in One-Dimensional Quantum Spin Systems*, Phys. Rev. B **85,** 075125 (2012).

[42] X. Chen, Z.-C. Gu, and X.-G. Wen, *Classification of Gapped Symmetric Phases in One-Dimensional Spin Systems*, Phys. Rev. B **83,** 035107 (2011).

[43] N. Schuch, D. Pérez-García, and I. Cirac, *Classifying Quantum Phases Using Matrix Product States and Projected Entangled Pair States*, Phys. Rev. B **84,** 165139 (2011).

[44] X. Chen, Z.-C. Gu, and X.-G. Wen, *Complete Classification of One-Dimensional Gapped Quantum Phases in Interacting Spin Systems*, Phys. Rev. B **84,** 235128 (2011).

[45] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, *Symmetry Protected Topological Orders and the Group Cohomology of Their Symmetry Group*, Phys. Rev. B **87,** 155114 (2013).

[46] Z.-C. Gu and X.-G. Wen, *Symmetry-Protected Topological Orders for Interacting Fermions: Fermionic Topological Nonlinear σ Models and a Special Group Supercohomology Theory*, Phys. Rev. B **90,** 115141 (2014).

[47] M. Levin and Z.-C. Gu, *Braiding Statistics Approach to Symmetry-Protected Topological Phases*, Phys. Rev. B **86,** 115109 (2012).

[48] Y.-M. Lu and A. Vishwanath, *Theory and Classification of Interacting Integer Topological Phases in Two Dimensions: A Chern-Simons Approach*, Phys. Rev. B **86,** 125119 (2012).

[49] A. Vishwanath and T. Senthil, *Physics of Three-Dimensional Bosonic Topological Insulators: Surface-Deconfined Criticality and Quantized Magnetoelectric Effect*, Phys. Rev. X **3,** 011016 (2013).

[50] L. Fidkowski, X. Chen, and A. Vishwanath, *Non-Abelian Topological Order on the Surface of a 3D Topological Superconductor from an Exactly Solved Model*, Phys. Rev. X **3,** 041016 (2013).

[51] C. Wang, A. C. Potter, and T. Senthil, *Classification of Interacting Electronic Topological Insulators in Three Dimensions*, Science **343,** 629 (2014).

[52] C. Wang and T. Senthil, *Interacting Fermionic Topological Insulators/Superconductors in Three Dimensions*, Phys. Rev. B **89,** 195124 (2014).

[53] A. Kapustin, *Symmetry Protected Topological Phases, Anomalies, and Cobordisms: Beyond Group Cohomology*, arXiv:1403.1467.

[54] M. A. Metlitski, L. Fidkowski, X. Chen, and A. Vishwanath, *Interaction Effects on 3D Topological Superconductors: Surface Topological Order from Vortex Condensation, the 16 Fold Way and Fermionic Kramers Doublets*, arXiv:1406.3032.

[55] A. Kapustin, R. Thorngren, A. Turzillo, and Z. Wang, *Fermionic Symmetry Protected Topological Phases and Cobordisms*, J. High Energy Phys. 12 (2015) 052.

[56] M. Cheng, Z. Bi, Y.-Z. You, and Z.-C. Gu, *Classification of Symmetry-Protected Phases for Interacting Fermions in Two Dimensions*, Phys. Rev. B **97**, 205109 (2018).

[57] A. Kitaev, *Homotopy-Theoretic Approach to SPT Phases in Action: Z 16 Classification of Three-Dimensional Superconductors, Institute for Pure Applied Mathematics, UCLA*, http://www.ipam.ucla.edu/abstract/?tid=12389&pcode=STQ2015.

[58] E. Witten, *Fermion Path Integrals and Topological Phases*, Rev. Mod. Phys. **88**, 035001 (2016).

[59] D. S. Freed and M. J. Hopkins, *Reflection Positivity and Invertible Topological Phases*, arXiv:1604.06527.

[60] C. Z. Xiong, *Minimalist Approach to the Classification of Symmetry Protected Topological Phases*, J. Phys. A **51**, 445001 (2018).

[61] A. Kapustin and R. Thorngren, *Fermionic SPT Phases in Higher Dimensions and Bosonization*, J. High Energy Phys. 10 (2017) 080.

[62] Q.-R. Wang and Z.-C. Gu, *Towards a Complete Classification of Symmetry-Protected Topological Phases for Interacting Fermions in Three Dimensions and a General Group Supercohomology Theory*, Phys. Rev. X **8**, 011055 (2018).

[63] D. Gaiotto and T. Johnson-Freyd, *Symmetry Protected Topological Phases and Generalized Cohomology*, J. High Energy Phys. 05 (2019) 007.

[64] Q.-R. Wang and Z.-C. Gu, *Construction and Classification of Symmetry-Protected Topological Phases in Interacting Fermion Systems*, Phys. Rev. X **10**, 031055 (2020).

[65] R. B. Laughlin, *Quantized Hall Conductivity in Two Dimensions*, Phys. Rev. B **23**, 5632 (1981).

[66] We remind the reader again of the restriction that for the moment we assume that all the emergent symmetries are 0-form symmetries, which we reexamine in Sec. X. In fact, Theorem 1 continues to hold in the presence of *finite* higher-form symmetries.

[67] F. D. M. Haldane, in *Proceedings of the International School of Physics*, "Enrico Fermi," Course No. CXXI, edited by R. A. Broglia and J. R. Schrieffer (North-Holland, Amsterdam, 1994), pp. 5–29.

[68] A. Pressley and G. Segal, *Loop Groups* (Oxford University Press, Oxford, 1986).

[69] To see the quantization, observe that the total charge carried by the SPT phase under the U(1) subgroup of LU(1) corresponding to setting the function $f(\theta)$ in Eq. (12) to be independent of θ is given by the integral (over a fixed time slice)

$$\int j^0 = \int \frac{\delta S}{\delta A_0} = \frac{m}{8\pi^2} \int dA \wedge dA = mC[A],$$

where $C[A]$ is the second Chern number and is quantized to be an integer. Hence, m must be an integer in order for the total charge of the SPT phase to always be an integer.

[70] By a particle-hole transformation, which cannot modify the physics, the "occupied volume" could equivalently be taken to be the complement $\mathcal{D}^c$ of $\mathcal{D}$ in the Brillouin zone.

Observe, however, that if we replace $\mathcal{D}$ with $\mathcal{D}^c$ in Eq. (18), the only change in the anomaly Eq. (17) on the Fermi surface would be an additional minus sign, which as we mentioned earlier just corresponds to a choice of orientation of the Fermi surface.

[71] M.-C. Chang and Q. Niu, *Berry Phase, Hyperorbits, and the Hofstadter Spectrum: Semiclassical Dynamics in Magnetic Bloch Bands*, Phys. Rev. B **53**, 7010 (1996).

[72] S. Golkar, D. X. Nguyen, M. M. Roberts, and D. T. Son, *Higher-Spin Theory of the Magnetorotons*, Phys. Rev. Lett. **117**, 216403 (2016).

[73] D. G. Barci, E. Fradkin, and L. Ribeiro, *Bosonization of Fermi Liquids in a Weak Magnetic Field*, Phys. Rev. B **98**, 155146 (2018).

[74] This is associated with spontaneously breaking C_4 lattice rotation symmetry to C_2.

[75] T. Holstein, R. E. Norton, and P. Pincus, *de Haas–van Alphen Effect and the Specific Heat of an Electron Gas*, Phys. Rev. B **8**, 2649 (1973).

[76] P. A. Lee, N. Nagaosa, and X.-G. Wen, *Doping a Mott Insulator: Physics of High-Temperature Superconductivity*, Rev. Mod. Phys. **78**, 17 (2006).

[77] B. I. Halperin, P. A. Lee, and N. Read, *Theory of the Half-Filled Landau Level*, Phys. Rev. B **47**, 7312 (1993).

[78] J. Polchinski, *Low-Energy Dynamics of the Spinon-Gauge System*, Nucl. Phys. **B422**, 617 (1994).

[79] C. Nayak and F. Wilczek, *Non-Fermi Liquid Fixed Point in 2 + 1 Dimensions*, Nucl. Phys. **B417**, 359 (1994).

[80] B. L. Altshuler, L. B. Ioffe, and A. J. Millis, *Low-Energy Properties of Fermions with Singular Interactions*, Phys. Rev. B **50**, 14048 (1994).

[81] S.-S. Lee, *Low-Energy Effective Theory of Fermi Surface Coupled with U(1) Gauge Field in 2 + 1 Dimensions*, Phys. Rev. B **80**, 165102 (2009).

[82] M. A. Metlitski and S. Sachdev, *Quantum Phase Transitions of Metals in Two Spatial Dimensions. I. Ising-Nematic Order*, Phys. Rev. B **82**, 075127 (2010).

[83] D. F. Mross, J. McGreevy, H. Liu, and T. Senthil, *Controlled Expansion for Certain Non-Fermi-Liquid Metals*, Phys. Rev. B **82**, 045121 (2010).

[84] Y. Schattner, S. Lederer, S. A. Kivelson, and E. Berg, *Ising Nematic Quantum Critical Point in a Metal: A Monte Carlo Study*, Phys. Rev. X **6**, 031028 (2016).

[85] D. Dalidovich and S.-S. Lee, *Perturbative Non-Fermi Liquids from Dimensional Regularization*, Phys. Rev. B **88**, 245106 (2013).

[86] S.-S. Lee, *Recent Developments in Non-Fermi Liquid Theory*, Annu. Rev. Condens. Matter Phys. **9**, 227 (2018).

[87] M. A. Metlitski, D. F. Mross, S. Sachdev, and T. Senthil, *Cooper Pairing in Non-Fermi Liquids*, Phys. Rev. B **91**, 115111 (2015).

[88] T. Senthil, *Critical Fermi Surfaces and Non-Fermi Liquid Metals*, Phys. Rev. B **78**, 035103 (2008).

[89] This appears because we normalize our momenta to be dimensionless.

[90] D. Shoenberg, *Magnetic Oscillations in Metals* (Cambridge University Press, Cambridge, England, 1984).

[91] J.-Y. Chen and D. T. Son, *Berry Fermi Liquid Theory*, Ann. Phys. (Amsterdam) **377**, 345 (2017).

[92] D. T. Son and N. Yamamoto, *Berry Curvature, Triangle Anomalies, and the Chiral Magnetic Effect in Fermi Liquids*, Phys. Rev. Lett. **109**, 181602 (2012).

[93] The restriction to odd integers does not come up in Sec. VI, but we would have found the same restriction if we assumed that microscopically the fermion parity is $(-1)^{\hat{Q}}$ where $\hat{Q}$ is the total charge (as is the case in an electronic system), and moreover required (as is the case in all known examples) that an elementary Fermi-surface quantum be fermionic. In the spinful case, the restriction to odd q comes directly from the assumed form of G_{IR}, although in principle one could get even q if we replaced U(2) with SO(3) × U(1) in the definition of G_{IR}.

[94] D. T. Son, *Is the Composite Fermion a Dirac Particle?*, Phys. Rev. X **5**, 031027 (2015).

[95] N. Seiberg, T. Senthil, C. Wang, and E. Witten, *A Duality Web in 2 + 1 Dimensions and Condensed Matter Physics*, Ann. Phys. (Amsterdam) **374**, 395 (2016).

[96] One might ask why the magnetic flux does not also induce a chiral flow of quasiparticles along the Fermi surface, violating conservation of $\hat{n}(\theta)$. The point is that for a low-energy excitation, the variation of the local charge density, and therefore, of the flux of a, should be spread over an arbitrarily large volume, and hence, the effective local magnetic field, which determines the velocity of the chiral flow, actually goes to zero.

[97] To see this, just decompose into irreducible representations. In any such irreducible representation, by Schur's lemma $\hat{q}$ acts proportionally to the identity $\hat{q} = Q\mathbb{I}$ for some scalar Q. Then, taking the trace of $[\hat{p}_x, \hat{p}_y] = iB\hat{q}$ shows that $Q = 0$.

[98] A. C. Balram, C. Tőke, and J. K. Jain, *Luttinger Theorem for the Strongly Correlated Fermi Liquid of Composite Fermions*, Phys. Rev. Lett. **115**, 186805 (2015).

[99] A. C. Balram and J. K. Jain, *Fermi Wave Vector for the Partially Spin-Polarized Composite-Fermion Fermi Sea*, Phys. Rev. B **96**, 235102 (2017).

[100] We follow the regularization convention adopted in Ref. [95], where the action for the massless Dirac fermion implicitly includes the so-called "η invariant." Physically, this amounts to assuming that the system also contains an additional massive Dirac fermion in addition to the massless one explicitly represented in the Lagrangian.

[101] It is SO(3) rather than SU(2) or U(2) because of the constraint that the total gauge charge, i.e., the net spinon number, is zero for any state in the IR theory.

[102] We remark that if we set $\Phi_\uparrow = \Phi_\downarrow = 2\pi$, then the effect on Eq. (89) can be absorbed by redefining $\hat{\Phi}_b \to \hat{\Phi}_b - 2\pi$. Since $\hat{\Phi}_b$ always has such an ambiguity in terms of its action on a background flux (such ambiguities are the reason why fluxes are able to carry projective representations in the first place), we conclude that $\Phi_\uparrow = \Phi_\downarrow = 2\pi$ does not carry a nontrivial projective representation. This corresponds to the statement that the projective representations of $G_{\mathrm{IR,charge}}$ are $\mathbb{Z}_2$ classified, whereas the projective representations of LU(1) (see Sec. V D) were $\mathbb{Z}$ classified. These results are consistent with the fact that the total microscopic charge U(1) acts trivially on the IR theory since it is an electrical insulator.

[103] B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, and J. Zaanen, *From Quantum Matter to High-Temperature Superconductivity in Copper Oxides*, Nature (London) **518**, 179 (2015).

[104] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov, *Topological Semimetal and Fermi-Arc Surface States in the Electronic Structure of Pyrochlore Iridates*, Phys. Rev. B **83**, 205101 (2011).

[105] N. Read and S. Sachdev, *Large-N Expansion for Frustrated Quantum Antiferromagnets*, Phys. Rev. Lett. **66**, 1773 (1991).

[106] X. G. Wen, *Mean-Field Theory of Spin-Liquid States with Finite Energy Gap and Topological Orders*, Phys. Rev. B **44**, 2664 (1991).

[107] T. Senthil and M. P. A. Fisher, *Z_2 Gauge Theory of Electron Fractionalization in Strongly Correlated Systems*, Phys. Rev. B **62**, 7850 (2000).

[108] R. A. Jalabert and S. Sachdev, *Spontaneous Alignment of Frustrated Bonds in an Anisotropic, Three-Dimensional Ising Model*, Phys. Rev. B **44**, 686 (1991).

[109] S. Sachdev and M. Vojta, in *Proceedings of the International Workshop on Magnetic Excitations in Strongly Correlated Electrons, Hammatsu, Japan, 1999*, Vol. 69 (Physical Society of Japan, 2000), 1.

[110] P. Etingof, *Dmitri Nikshych, Victor Ostrik, and with an Appendix by Ehud Meir, Fusion Categories and Homotopy Theory*, arXiv:0909.3140.

[111] M. Barkeshli, P. Bonderson, M. Cheng, and Z. Wang, *Symmetry Fractionalization, Defects, and Gauging of Topological Phases*, Phys. Rev. B **100**, 115147 (2019).

[112] F. Benini, C. Córdova, and P.-S. Hsin, *On 2-Group Global Symmetries and Their Anomalies*, J. High Energy Phys. 03 (2019) 118.

[113] P.-S. Hsin and A. Turzillo, *Symmetry-Enriched Quantum Spin Liquids in* $(3 + 1)$D, J. High Energy Phys. 09 (2020) 022.

[114] J. C. Baez and A. D. Lauda, Higher-Dimensional Algebra V: 2-Groups, arXiv:math.QA/0307200.

[115] S. Gukov and A. Kapustin, *Topological Quantum Field Theory, Nonlocal Operators, and Gapped Phases of Gauge Theories*, arXiv:1307.4793.

[116] A. Kapustin and R. Thorngren, *Higher Symmetry and Gapped Phases of Gauge Theories*, arXiv:1309.4721.

[117] C. Lannert, M. P. A. Fisher, and T. Senthil, *Quantum Confinement Transition in a d-Wave Superconductor*, Phys. Rev. B **63**, 134510 (2001).

[118] L. Balents, L. Bartosch, A. Burkov, S. Sachdev, and K. Sengupta, *Putting Competing Orders in Their Place near the Mott Transition*, Phys. Rev. B **71**, 144508 (2005).

[119] I. Dzyaloshinskii, *Some Consequences of the Luttinger Theorem: The Luttinger Surfaces in Non-Fermi Liquids and Mott Insulators*, Phys. Rev. B **68**, 085113 (2003).

[120] T. D. Stanescu and G. Kotliar, *Fermi Arcs and Hidden Zeros of the Green Function in the Pseudogap State*, Phys. Rev. B **74**, 125110 (2006).

[121] K.-Yu. Yang, T. M. Rice, and F.-C. Zhang, *Phenomenological Theory of the Pseudogap State*, Phys. Rev. B **73**, 174501 (2006).

[122] A. Rosch, *Breakdown of Luttinger's Theorem in Two-Orbital Mott Insulators*, Eur. Phys. J. B **59,** 495 (2007).

[123] K. B. Dave, P. W. Phillips, and C. L. Kane, *Absence of Luttinger's Theorem due to Zeros in the Single-Particle Green Function*, Phys. Rev. Lett. **110,** 090403 (2013).

[124] S. A. Hartnoll, D. M. Hofman, and A. Tavanfar, *Holographically Smeared Fermi Surface: Quantum Oscillations and Luttinger Count in Electron Stars*, Europhys. Lett. **95,** 31002 (2011).

[125] S. A. Hartnoll, in *Black Holes in Higher Dimensions*, edited by Gary T. Horowitz (Cambridge University Press, New York, 2012), pp. 387–419.

[126] N. Iqbal and H. Liu, *Luttinger's Theorem, Superfluid Vortices and Holography*, Classical Quantum Gravity **29,** 194004 (2012).

[127] K. Hashimoto and N. Iizuka, *A Comment on Holographic Luttinger Theorem*, J. High Energy Phys. 07 (2012) 064.

[128] W. O. Putikka, M. U. Luchini, and R. R. P. Singh, *Violation of Luttinger's Theorem in the Two-Dimensional $t - J$ Model*, Phys. Rev. Lett. **81,** 2966 (1998).

[129] E. Cappelluti and R. Zeyher, *Violation of Luttinger's Theorem in Strongly Correlated Electronic Systems within a $1/N$ Expansion*, Int. J. Mod. Phys. B **13,** 2607 (1999).

[130] D. V. Else and T. Senthil, *Strange Metals as Ersatz Fermi Liquids*, arXiv:2010.10523.

[131] A. Kitaev, *Periodic Table for Topological Insulators*, AIP Conf. Proc. **1134,** 22 (2009).

[132] R. Thorngren and D. V. Else, *Gauging Spatial Symmetries and the Classification of Topological Crystalline Phases*, Phys. Rev. X **8,** 011040 (2018).

[133] D. V. Else, W. W. Ho, and P. T. Dumitrescu, *Long-Lived Interacting Phases of Matter Protected by Multiple Time-Translation Symmetries in Quasiperiodically Driven Systems*, Phys. Rev. X **10,** 021032 (2020).

[134] In general, there there can be a Berry curvature correction to this equation [135], but it enters only at next-to-leading order in the magnetic field strength, so we disregard it here.

[135] C. Duval, Z. Horváth, P. A. Horváthy, L. Martini, and P. C. Stichel, *Berry Phase Correction to Electron Density and Exotic Dynamics*, Mod. Phys. Lett. B **20,** 373 (2006).

[136] By using the closed subgroup theorem for Lie groups [137].

[137] J. M. Lee, *Introduction to Smooth Manifolds* (Springer, New York, 2013).

[138] This follows [139] from the "neighboring subgroup theorem" of Ref. [140].

[139] https://mathoverflow.net/questions/361149/does-a-compact-lie-group-have-finitely-many-conjugacy-classes-of-maximal-abelian.

[140] D. Montgomery and L. Zippin, *A Theorem of Lie Groups*, Bull. Am. Math. Soc. **48,** 448 (1942).

[141] We note that for $d = 1$, the Chern-Simons invariant is an integer multiple of $\mathrm{Tr}_R(\tau_1)$.

[142] R. Thorngren, *Anomalies and Bosonization*, arXiv:1810.04414.