

Constraints on the symmetric mass generation paradigm for lattice chiral gauge theories

Maarten Golterman¹ and Yigal Shamir²

¹*Department of Physics and Astronomy, San Francisco State University, San Francisco, California 94132, USA*

²*Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, 69978, Tel Aviv, Israel*

 (Received 5 June 2025; accepted 22 December 2025; published 15 January 2026)

Within the symmetric mass generation (SMG) approach to the construction of lattice chiral gauge theories, one attempts to use interactions to render mirror fermions massive without symmetry breaking, thus obtaining the desired chiral massless spectrum. If successful, the gauge field can be turned on, and thus a chiral gauge theory can be constructed in the phase in which SMG takes place. In this paper we argue that the zeros that often replace the mirror poles of fermion two-point functions in an SMG phase should be “kinematical” singularities. We conjecture that the SMG interactions generate opposite-chirality bound states, which combine with the gapped elementary mirror states to form massive Dirac fermions. The propagator zeros can then be avoided by choosing an appropriate set of interpolating fields that contains both elementary and composite fields. This allows us to apply general constraints on the existence of a chiral fermion spectrum which are valid in the presence of (nongauge) interactions of arbitrary strength, including in any SMG phase. Using a suitably constructed one-particle lattice Hamiltonian describing the fermion spectrum, we formulate a generalized no-go theorem which establishes the conditions for the applicability of the Nielsen-Ninomiya theorem to this Hamiltonian. If these conditions are satisfied, the massless fermion spectrum must be vectorlike. We add some general observations on the strong coupling limit of SMG models. We also elaborate on the qualitative differences between four-dimensional and two-dimensional theories that limit the lessons that can be drawn from two-dimensional models. Finally, we compile a list of open questions which must be addressed in any SMG model in order to determine whether or not it is subject to the generalized no-go theorem.

DOI: [10.1103/physrevd.113.014503](https://doi.org/10.1103/physrevd.113.014503)

I. INTRODUCTION

Because of the fermion species-doubling problem, the nonperturbative construction of chiral gauge theories on the lattice is a long-standing challenge.¹ The physical origin of species doubling was first addressed by Karsten and Smit [4], tying the phenomenon to the chiral anomaly, and then generalized by Nielsen and Ninomiya [5]. Since then, the program of putting chiral gauge theories on the lattice has gained partial successes, but no completely worked-out method for doing so exists at present.

Building on earlier work by Ginsparg and Wilson [6], by Kaplan [7], and by Narayanan and Neuberger [8–11], Lüscher successfully constructed anomaly-free Abelian

chiral gauge theories, while requiring one new algebraic constraint on the fermion spectrum beyond the familiar anomaly-cancellation condition [12]. As for non-Abelian chiral gauge theories, he was able to define them to all orders in lattice perturbation theory [13,14].

A second approach is being pursued by Kaplan and collaborators. In this approach, the chiral fermions reside on a four-dimensional boundary of a five-dimensional space, with massive fermion degrees of freedom inside the five-dimensional bulk. The dynamical gauge field lives on the same boundary as the chiral fermions, and is extended into the five-dimensional bulk via a classical differential equation, for which the dynamical four-dimensional gauge field provides the boundary values. The goal is to dampen the gauge field inside the five-dimensional bulk such that, when the fermion spectrum is anomaly free, the long-distance physics would originate exclusively from the coupling of the dynamical gauge field to the chiral fermions at the boundary. Concrete realizations of this approach were proposed in a domain-wall fermion setup [15,16], and more recently in a disk setup [17,18].

¹For reviews, see Refs. [1–3].

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

The scope of this approach is presently under investigation [19,20].

A third approach, which we have pursued, is the gauge-fixing approach. The chiral gauge invariance is explicitly broken on the lattice. Gauge invariance and (conjecturally) unitarity are restored only in the continuum limit, provided that the fermion spectrum is anomaly free. The inclusion of a suitable gauge-fixing lattice action, as well as, in the non-Abelian case, ghost fields, ensures the existence of a novel critical point where the target chiral gauge theory can be defined [21–27]. The current challenges of this approach have to do primarily with the dynamics of the gauge-fixing sector in the non-Abelian case [28,29].

The fourth, and historically the first, approach has a long history. It has evolved into what nowadays is usually called the symmetric mass generation (SMG) approach.² One starts from a lattice gauge theory with massless Dirac fermions. The (anomaly-free) fermion spectrum of the target chiral gauge theory would be obtained by selectively retaining only one of the two chiralities of each Dirac fermion; this is the physical fermion, while the unwanted opposite-chirality component is the “doubler,” or “mirror,” fermion. In order to recover the target chiral gauge theory in the continuum limit one must therefore find a way to decouple all the mirror fermions. The SMG paradigm envisages that this goal can be achieved by adding to the lattice action judiciously chosen strong interactions that do not involve the gauge field. They can be multifermion interactions and/or depend on additional scalar fields introduced especially for this purpose.

Let the gauge symmetry be a compact Lie group G . If we turn off the gauge field, we obtain a so-called *reduced model* in which G is an exact global symmetry. The fermion content of the original gauge theory can be read off by assigning the fermion states of the reduced model to representations of G , if G is non-Abelian, or by specifying the (integer valued) charge of each fermion state, if G is Abelian. The reduced model will typically have a nontrivial phase diagram, and different phases may in principle have a different massless fermion spectrum. The goal of the SMG framework is thus to achieve a novel “SMG phase” with the following features: the global G symmetry is not broken spontaneously; the physical chiral spectrum remains massless; all the unwanted mirror fermions have become massive; and no other undesired massless states have emerged in the process. The continuum limit of the reduced model in the SMG phase should be a theory of free massless chiral fermions, because, if in a *continuum* chiral gauge theory we turn off the gauge field, we are left with a set of free massless chiral fermions. Turning the gauge field back on will then result in a lattice regularization of the desired chiral gauge theory.

The reduced model has a phase diagram spanned by the coupling constants $g_1, g_2, \dots$, of the interactions that we introduce with the aim of decoupling the doublers. In the free theory limit $(g_1, g_2, \dots) = (0, 0, \dots)$, the reduced model must have a spectrum of Dirac fermions. Why? The Hamiltonian (or Lagrangian) of a free lattice fermion has a conserved fermion number symmetry, and associated with it a conserved current.³ The divergence of this current is zero on the lattice, and will remain so in the continuum limit; the current cannot develop an anomalous divergence in any correlation function. The straightforward way for the lattice regularization to guarantee the absence of an anomalous divergence in the continuum limit is to assemble all the fermion states into Dirac fermions with respect to the fermion number symmetry. The same argument applies to the conserved current of any other continuous global symmetry of the theory. These observations were first made by Karsten and Smit, who also derived the simplest no-go theorem in one spatial dimension [4].

A little later, a much more powerful no-go theorem was proved by Nielsen and Ninomiya [5]. One considers a free lattice Hamiltonian with a compact global symmetry. Under this symmetry, every fermion field is endowed with a set of discrete-valued quantum numbers, and the Hamiltonian can be diagonalized in each charge sector separately. The other requirements of the Nielsen-Ninomiya (NN) theorem are the following: lattice translation invariance, which implies that the momentum takes values in a periodic Brillouin zone; a relativistic low-energy spectrum, which implies that every massless fermion is unambiguously either right-handed (RH) or left-handed (LH);⁴ and finally, as a function of the momentum, the Hamiltonian must have a continuous first derivative. Under these assumptions, the NN theorem asserts that there is an equal number of RH and LH massless fermions in every charge sector.⁵

The SMG paradigm envisages that strong-enough interactions have been turned on to achieve an SMG phase with the properties described above. At face value, the presence of these interactions ensures that the NN theorem—which is a theorem about free Hamiltonians—is safely inapplicable. Hence, the massless spectrum in an SMG phase might, in principle, escape the conclusion of the NN theorem and be chiral.

However, a closer look reveals another facet of the reduced model which comes very close to satisfying the assumptions of the NN theorem. The continuum limit of the reduced model is required to be a theory of free

²For a review of the more recent work on SMG, see Ref. [30].

³The exception is Majorana fermions. However, these can belong only to real representations of any compact Lie group G , hence they do not play any role in chiral gauge theories.

⁴See Sec. III for the definition of the handedness.

⁵For no-go theorems in Euclidean space see Refs. [31,32]. For a no-go theorem that applies for a more general notion of conserved charges see Ref. [33].

massless fermions, reproducing the fermion spectrum of the target chiral gauge theory, so that the latter will be recovered when the gauge interactions are turned back on. But this means that close to the continuum limit of the reduced model, its fermionic massless states must be almost free. They can only interact weakly, via irrelevant interactions, that will automatically die out in the continuum limit. This state of affairs takes us almost all the way back to the arena of free bilinear Hamiltonians, to which the NN theorem applies. It is thus quite natural that some generalization of the NN theorem would be applicable in the interacting reduced model as well. Whether or not such a generalized no-go theorem is powerful enough to exclude a chiral fermion spectrum at a particular point in the phase diagram of the reduced model will then depend on the precise conditions of the theorem.

Such a generalization of the NN theorem was derived by one of us in Refs. [34,35]. The key step is to identify a one-particle lattice Hamiltonian $H_{\text{eff}}(\vec{p})$ as the inverse of a suitable Hermitian fermion two-point function $\mathcal{R}(\vec{p})$ at zero frequency,

$$H_{\text{eff}}(\vec{p}) = \mathcal{R}^{-1}(\vec{p}), \quad (1.1)$$

defined for *all* momenta $\vec{p}$ in the Brillouin zone. The precise definition of $\mathcal{R}(\vec{p})$ will be given later. If the Hamiltonian (or Euclidean action) of the underlying reduced model is local, one expects that $\mathcal{R}(\vec{p})$ will be an analytic function of the momentum everywhere in the Brillouin zone, except at *degeneracy points*. A degeneracy point $\vec{p}_c$ is a point in the Brillouin zone where $\mathcal{R}(\vec{p})$ receives a contribution from intermediate states with $E(\vec{p}) \rightarrow 0$ for $\vec{p} \rightarrow \vec{p}_c$. In this paper, we will prove the analyticity of $\mathcal{R}(\vec{p})$ away from the degeneracy points under certain assumptions about the field content and interactions of the reduced model.⁶ We further distinguish between two types of degeneracy points, which we will refer to as *primary singularities* and *secondary singularities*. A primary singularity p_c is defined as a point where $\mathcal{R}(\vec{p}) \rightarrow \infty$ for $\vec{p} \rightarrow \vec{p}_c$, which implies that $H_{\text{eff}}(\vec{p}_c)$ has a zero eigenvalue.⁷ Any other degeneracy point is a secondary singularity. By definition, at a secondary singularity $\mathcal{R}(\vec{p})$ is finite for $\vec{p} \rightarrow \vec{p}_c$, hence $H_{\text{eff}}(\vec{p})$ does not have any zero eigenvalues in the vicinity of $\vec{p}_c$. A primary singularity requires the presence of a massless single-particle intermediate state, whereas for a secondary singularity any single-particle intermediate state must be gapped. The nonanalyticity at a secondary singularity arises from intermediate states containing three or more massless fermions.

⁶The term “degeneracy point” reflects the fact that any state with vanishing energy is degenerate with the second-quantized vacuum.

⁷We define $H_{\text{eff}}(\vec{p})$ at the primary singularities by demanding continuity (see Sec. III).

How does $H_{\text{eff}}(\vec{p})$ fare with the assumptions of the NN theorem? First, obviously, $H_{\text{eff}}(\vec{p})$ is defined over the same Brillouin zone as the reduced model. As in the free case, it is also subject to similar symmetry constraints, since we assume that we are in a phase where the (to be gauged) exact global symmetry G is not broken spontaneously.⁸ Next, as explained above, we require that the massless fermionic asymptotic states are relativistic, and subject to irrelevant interactions only, and that there are no massless bosons. These assumptions will allow us to establish that $H_{\text{eff}}(\vec{p})$ has a continuous first derivative at all degeneracy points, both primary (see Sec. III) and secondary (see Appendix B 4). This, in turn, implies a one-to-one correspondence between the primary singularities and the massless fermion asymptotic states.⁹ Everywhere else in the Brillouin zone, $H_{\text{eff}}(\vec{p})$ will be an analytic function of the momentum, just like $\mathcal{R}(\vec{p})$, with one important exception.

Let us introduce the notion of a *zero* of $\mathcal{R}(\vec{p})$. A point $\vec{p}_0$ in the Brillouin zone is a zero of $\mathcal{R}(\vec{p})$ if the latter has a zero eigenvalue for $p = p_0$. By Eq. (1.1), a zero of $\mathcal{R}(\vec{p})$ turns into a pole of $H_{\text{eff}}(\vec{p})$. Thus, if $\mathcal{R}(\vec{p})$ has a nonempty set of zeros, then $H_{\text{eff}}(\vec{p})$ will not have a continuous first derivative at the same points. As a result, a crucial condition of the NN theorem will be violated, and the theorem will not apply.

The applicability of the NN theorem thus relies on the ability to construct fermion two-point functions that do not have any zeros. This brings us to the following issue: the identification of a suitable set of lattice operators that will serve as interpolating fields for the fermionic asymptotic states of the reduced model at a given point of its phase diagram. It is convenient to define a *complete set* of interpolating fields in a given charge sector by the following two requirements: (1) The massless fermion asymptotic states are in one-to-one correspondence with the primary singularities of $\mathcal{R}(\vec{p})$ for this set of interpolating fields; (2) $\mathcal{R}(\vec{p})$ is free of zeros.

In practice, building a complete set of interpolating fields can be a trial and error process, which could be helped by hints about the dynamics. First, it is quite natural for a lattice fermion field to interpolate more than one massless state, or, equivalently, to have several primary singularities. The simplest example is a naive fermion on a one-dimensional spatial lattice. This is a single-component field with the dispersion relation $E = \frac{1}{a} \sin(ap)$, where a is the lattice spacing. The massless spectrum¹⁰ then consists

⁸For the situation in one spatial dimension see Sec. III.

⁹Physically, a continuous first derivative at a primary singularity means that the velocity of the particle is well-defined for $\vec{p} \rightarrow \vec{p}_c$.

¹⁰In one spatial dimension, with no concept of helicity, massless fermions can be right-moving or left-moving. The abbreviations RH and LH when used in the context of models in one spatial dimension will be taken to refer to right-movers and left-movers, respectively, throughout this paper.

of a RH state at $p_c = 0$ and a LH state at $p_c = \pi/a$. As shown in Appendix B of Ref. [35], attempting to build a one-to-one correspondence between these massless fermion asymptotic states and the interpolating fields results in an “over-complete” set of interpolating fields that suffers from the presence of zeros.

Notice that our definition of a complete set of interpolating fields does not refer explicitly to any gapped fermion asymptotic states. The reason is that gapped states (whether relativistic or not) do not generate singularities at zero frequency. The primary singularities of $\mathcal{R}(\vec{p})$, and thus the zeros of $H_{\text{eff}}(\vec{p})$, correspond to massless fermions only. As a result, both the original NN theorem and its generalization discussed here are “oblivious” to the presence of any gapped fermion states.

This state of affairs is convenient, in that it means that it does not matter if our set of interpolating fields generates any gapped fermion states (in addition to the massless fermion states), so long as it satisfies the requirements of a complete set defined above. However, the requirement that $\mathcal{R}(\vec{p})$ should have no zeros indirectly constrains how Dirac fermions can be interpolated. A massive Dirac fermion consists of a RH and a LH component which are coupled via the mass term. As we will explain in detail below, if both chiralities are interpolated, this should not lead to propagator zeros. However, if only one of the two chiralities is interpolated, while the other is not, this leads to the unwanted appearance of a *kinematical zero* in $\mathcal{R}(\vec{p})$. These kinematical zeros can be avoided by simply adding to the set of interpolating fields new fields that will interpolate the missing chirality component of the relevant Dirac fermions.

A case where a propagator zero is unavoidable is when our starting point is a bilinear Hamiltonian or action that has a built-in pole. Such a (Euclidean) action was proposed long ago by Rebbi [36] in an early effort to put chiral gauge theories on the lattice. However, it was soon realized that a pole in the action acts as a ghost state [32,37]. In particular, in four dimensions it contributes to the one-loop beta function with the same magnitude as a fermion field, but with an opposite sign. We have recently generalized this conclusion using an effective low-energy framework [38].

As the example of Rebbi’s proposal shows, generating a pole in the action requires a highly nonlocal kinetic term. If, on the other hand, the underlying theory is local, it is unlikely that the severe nonlocality needed for generating such ghost states would emerge at the level of the low-energy effective theory [39]. Turned around, this observation suggests that any propagator zeros encountered in the SMG phase of a local reduced model should be kinematical zeros, which are removable by a judicious choice of the interpolating fields.

Our current understanding of the issue of propagator zeros, which has evolved considerably since our earlier paper [38], is the following. We conjecture that the same SMG interactions that gap the mirrors also generate

opposite-chirality bound states. These bound states pair up with the elementary mirror fermions, thereby forming massive Dirac fermions. Hence, in order to obtain a complete set of interpolating field free of propagator zeros, one should add to the set of elementary fields, which interpolate the mirror fermions, a suitable set of composite fields that will interpolate the opposite-chirality bound states.

A popular testbed of the SMG paradigm is the so-called 3450 model [40–44]. The target chiral gauge theory is an Abelian theory in one spatial dimension, with two (say) LH fields with charges 3 and 4, and one RH field with charge 5. Since $3^2 + 4^2 = 5^2$, this fermion content satisfies the anomaly cancellation condition in two spacetime dimensions. When the SMG paradigm is applied to the 3450 model, the starting point is an Abelian lattice theory with massless Dirac fermions with charges 3, 4, and 5, representing a doubled spectrum. One then aims to find an SMG phase where the mirrors have become massive, while the fermion spectrum of the target chiral gauge theory remains massless. The reason for the name “3450 model” is that a neutral “spectator” Dirac fermion is added to the lattice theory. Having a zero charge, this fermion field does not interact with the gauge field. But it is employed in the construction of the interaction terms that, one hopes, would generate the SMG phase in which all the doublers become massive, and one is left with a massless fermion spectrum consisting of LH states with charges 3 and 4, and RH states with charges 5 and 0.

An attempt to decouple the mirrors was made in Ref. [40], where a concrete reduced-model realization of the 3450 model was proposed. The goal of Ref. [40] was to obtain an SMG phase by introducing additional scalar fields that interact strongly with the fermion fields via Yukawa couplings. As a probe of the fermion spectrum, the vacuum polarization diagram was calculated in the reduced model, i.e., the two-point function of the conserved current of the global U(1) symmetry to be gauged in the full model. In two space-time dimensions, a massless fermion intermediate state generates a directional discontinuity at zero momentum. In units of the discontinuity created by a single chiral fermion of unit charge, the strength of the total discontinuity is $\sum_i q_i^2$, where the sum runs over the U(1) charges q_i of all the massless chiral fermions, both RH and LH. An undoubled massless spectrum would thus exhibit a discontinuity of total relative strength $3^2 + 4^2 + 5^2 = 50$, coming from the massless LH fermions with charges 3 and 4, and the RH fermion with charges 5.

In Ref. [40] the discontinuity of the vacuum polarization diagram was calculated numerically in the mirror sector at a single point inside the would-be SMG phase. Instead of the desired result, which is a vanishing discontinuity in the mirror sector, the actual result was consistent with a discontinuity of total relative strength of 50. Clearly, the simplest interpretation of this result is that the mirror

fermions remained massless, and thus that the spectrum remained doubled, and not chiral.¹¹ As pointed out in Ref. [40], an alternative interpretation would be the appearance of a new massless Dirac fermion of charge 5, whose LH component is the original mirror state of the fermion field with charge 5, while its RH component is a bound state of the fermion and scalar fields of the model. Such a situation would imply that the total fermion spectrum in each charge sector remains chiral. We comment that while this situation cannot be ruled out based on the result of Ref. [40] alone, it is unlikely in our opinion, even if we set aside the generalized no-go theorem which is the main topic of the present paper. The reason is that the model of Ref. [40] contains no symmetry that would protect the masslessness of such a Dirac fermion, and so it would be a remarkably fine-tuned situation if its mass happened to be zero (within the uncertainties of the calculation) right at the point in the phase diagram where the numerical calculations were carried out.

Recently, building on heuristic arguments presented in Refs. [42,43], a different reduced-model realization of the 3450 model was put forward in Ref. [44], which we will refer to as the ZZWY model. Introducing two judiciously chosen six-fermion interactions, the successful development of an SMG phase with gapped mirror fermions was announced. Since this result is in apparent conflict with the generalization of the NN theorem to interacting (reduced) models, our goal in this paper is to revisit the considerations of the generalized theorem [34,35], using the ZZWY model as a laboratory whenever possible.

This paper is organized as follows. We begin in Sec. II with a fresh examination of the issue of propagator zeros, revisiting our discussion in Ref. [38]. We first explain why propagator zeros tend to arise when mirror fermions are gapped. Because the underlying reduced model is local by assumption, we now realize that the propagator zeros are unlikely to represent ghost states [39], and are thus more likely to be kinematical in nature. We then introduce the (conjectured) bound-state formation mechanism: The same SMG interactions that gap the mirrors also generate opposite-chirality bound states, which, in turn, pair up with the elementary mirror fermions to form massive Dirac fermions free of propagator zeros.¹² We give explicit formulas for the composite operators that would create the bound states in the ZZWY model, and should thus be added to the set of interpolating fields.

In Sec. III we turn to the generalized no-go theorem. In contrast with Refs. [34,35] where analyticity properties were essentially postulated, in this paper we prove the

analyticity of $\mathcal{R}(\vec{p})$ for reduced models defined by a local Hamiltonian that depends on fermion fields only (as is the case for the ZZWY model). We discuss the remaining conditions of the generalized theorem, which include the absence of zeros in $\mathcal{R}(\vec{p})$, and the assumption that the continuum limit of the reduced model is a theory of relativistic free massless fermions, with no additional massless bosonic states. Technical details are mostly relegated to Appendix B. Finally, we briefly explain how the gauge-fixing approach to the construction of four-dimensional lattice chiral gauge theories evades the generalized no-go theorem.

In Sec. IV we prove an independent, but related, simple theorem. We consider a reduced model in which a subset of the fermion degrees of freedom participates in some strong interactions. We then show that when the (uniform) strong-coupling limit is taken, the fermion fields split into two decoupled sectors. In particular, the fermion degrees of freedom that do not participate in the strong interactions then form a decoupled free (or weakly coupled) theory, which is thus subject to the NN theorem. We list several examples, including in particular the application of this theorem to the ZZWY model.

In Sec. V we turn to the implications of the generalized no-go theorem for two-dimensional theories. Two-dimensional theories are, in a sense, more complicated than four-dimensional theories, because many more relevant or marginal operators exist in two dimensions, and their role needs to be understood. We confront the generalized no-go theorem with the properties of two-dimensional SMG reduced models in general, and with the known features of the ZZWY model in particular. While the SMG interactions themselves are always chosen to be irrelevant, they can induce four-fermion interactions without derivatives that respect all the symmetries of the SMG reduced model. In two dimensions, such induced four-fermion interactions are renormalizable, hence they can have a profound impact on the resulting continuum theory.

Our work can only have tentative implications for any specific SMG model; for the generalized no-go theorem to apply, all of its assumptions need to be satisfied, and this must be checked on a case by case basis. In Sec. VI we use our analysis to compile a list of open questions. We focus on the fate of propagator zeros and on the applicability of each assumption of the generalized no-go theorem. These key questions will have to be sorted out in the ZZWY model, as well as in any other SMG model, in order to determine whether or not it succeeds in recovering a chiral gauge theory in the continuum limit.

We end with some concluding remarks in Sec. VII. Appendix A provides details on the ZZWY Hamiltonian, focusing mainly on its bilinear part, including the impact of the strong-coupling limit of Sec. IV, as well as a few details on the interaction Hamiltonian and the symmetries. Proofs of most of the properties of $\mathcal{R}(\vec{p})$ needed for the generalized no-go theorem are relegated to Appendix B.

¹¹See Ref. [40] for the systematic uncertainties involved in the calculation. See also Ref. [41] for an attempt to explain the failure of Ref. [40].

¹²While in Ref. [38] we noted that the presence of bound states could in principle affect our conclusions, we did not appreciate that bound-state formation would turn out to play a key role.

II. SMG, PROPAGATOR ZEROS, AND BOUND-STATE FORMATION

The goal of the SMG paradigm is to gap the mirror chiral component of a Dirac fermion while leaving massless the other chirality, as it is part of the chiral spectrum of the target chiral gauge theory. In order to see why the gapping of the mirror component tends to lead to the appearance of a propagator zero, consider the example of a continuum massless Dirac fermion, with propagator¹³

$$\frac{\not{p}}{p^2}. \quad (2.1)$$

Chiral symmetry follows from the fact that this propagator anticommutes with γ_5 . If, for example, the RH component has been gapped somehow, while the LH component remains massless, then the RH pole must have moved away from zero, so that the same two-point function would now take the new form

$$P_L \frac{\not{p}}{p^2} P_R + P_R \frac{\not{p}}{p^2 - m^2} P_L, \quad (2.2)$$

where $P_{L,R}$ are the chiral projectors. While the last term indicates the presence of a massive state, this propagator now has a RH zero instead of the original massless pole.¹⁴

Assuming that propagator zeros develop in an SMG phase, what is their physical significance? The dynamics of an SMG phase might not be easily tractable because SMG requires strong interactions. Before we turn to this question, let us consider another example taken from a free continuum theory. This time we start from the massive Dirac propagator

$$G = \frac{\not{p} - m}{p^2 - m^2}, \quad (2.3)$$

which has a pole at $p^2 = m^2$, and no zeros. Let us now pick out the RH chirality of this Dirac fermion by applying suitable projectors on the two sides of G . The result, $P_R G P_L$, readily coincides with the rightmost term in Eq. (2.2). This shows that a propagator zero can be a “mundane” kinematical singularity resulting from the application of a projection to an ordinary massive Dirac propagator.

¹³Our conventions are as follows. We use d for the number of spatial dimensions. We denote d -vectors as, e.g., $\vec{x}$, $\vec{p}$, and their inner product as $\vec{x} \cdot \vec{p}$. For quantities that carry spacetime indices we use Minkowski space conventions with $\eta_{\mu\nu} = \text{diag}(1, -1, -1, -1)$. For example, $p^2 = p^\mu p_\mu = \omega^2 - \vec{p}^2$ where $\omega = p_0$.

¹⁴A propagator zero occurs if mixed-chirality regular terms are added to Eq. (2.2) as well.

As we have discussed in the Introduction, the alternative is an “irremovable” propagator zero that arises from a nonlocal action with a pole in its bilinear part, which in turn represents a ghost state. However, the idea is that the SMG phase develops in a theory where, as a rule, the underlying Lagrangian (or Hamiltonian) is local by construction. Hence, it is unlikely that the kind of nonlocality needed for ghost states would develop at the level of the effective theory that controls the long-distance behavior [39].

The propagator zeros found in an SMG phase are thus more likely to be kinematical singularities. This means that it must be possible to reproduce each propagator zero by applying chiral projectors to a pertinent, massive Dirac propagator, as in the simple example we have just considered. But this conclusion immediately raises another question. If the propagator zero can be reproduced by projecting out the RH component of a massive Dirac fermion found in the spectrum of the theory, then, evidently, the corresponding Dirac field must have a LH component as well. However, the LH component of this massive Dirac field must be different from the LH component of the original massless Dirac fermion. The reason is that, first, the SMG interactions are constructed to involve only the RH component of the original massless Dirac fermion (any residual coupling to the LH component is suppressed by design). Second, if the goal of the SMG program is to be achieved, then the LH component of the original massless fermion has to remain massless. The LH component of the massive Dirac fermion should thus be supplied by the SMG dynamics itself, in other words, it must arise as a bound state in the SMG phase! A constraint is that this bound state transforms in the same representation of G as the gapped RH component of the original Dirac fermion, in order to avoid spontaneous breaking of the symmetry group G .

We will now argue that, for each Weyl fermion that has been gapped in an SMG phase, the SMG dynamics can in principle generate an opposite-chirality bound state. Ultimately, the massive spectrum in the SMG phase would thus consist of “hybrid” Dirac fermions, each of which has one chirality component which is elementary, while the other chirality component is a bound state. The full propagator of each massive Dirac fermion is free of zeros.

As we will see, the mechanism of bound-state formation is very general. Nevertheless, to make the discussion more concrete we will consider here the ZZWY lattice construction for the 3450 model [44]. We thus pause to describe the main features of this model. More details may be found in Appendix A.

The ZZWY model is defined by a lattice Hamiltonian. The time coordinate remains continuous while space is discretized. The Hamiltonian consists of a free part H_0 and an interacting part H_{int} . There are four single-component fermion fields ψ_I , where the “species” index $I = 1, 2, 3, 4$ corresponds to charges 3, 4, 5 and 0 under the $U(1)$

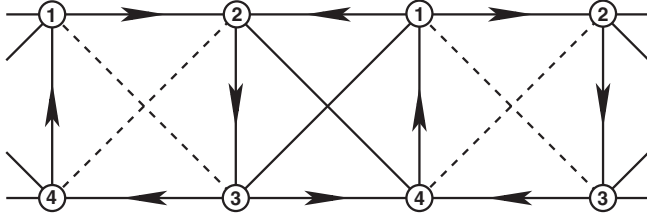


FIG. 1. The lattice of the 3450 model of Ref. [44]. The lattice shown in Fig. 1(a) of that paper is here rotated by 90° clockwise, so that the physical space direction is horizontal. The lattice consists of two interconnected one-dimensional chains. Edge A (on the left in Fig. 1(c) of Ref. [44]) is here the upper chain, while edge B is the lower chain. The directional links have complex coupling $\tau_1 = t_1 e^{i\frac{\pi}{4}}$, while the undirectional links have real coupling t_2 (solid line) or $-t_2$ (dashed line). The unit cell of this lattice is a 2×1 rectangle (in lattice units), and the numbers inside the circles represent different sublattices. To avoid confusion, note that the sublattice index is different from the species index I in Eq. (2.4). For more details, see Appendix A.

symmetry to be gauged.¹⁵ The fermion fields live on the lattice shown in Fig. 1, which consists of two interconnected one-dimensional chains. The free Hamiltonian is given by

$$H_0 = \bar{\psi}_1 \mathcal{H}(\tau_1, t_2) \psi_1 + \bar{\psi}_2 \mathcal{H}(\tau_1, t_2) \psi_2 + \bar{\psi}_3 \mathcal{H}(\tau_1^*, t_2) \psi_3 + \bar{\psi}_4 \mathcal{H}(\tau_1^*, t_2) \psi_4, \quad (2.4)$$

where τ_1 is a complex parameter while t_2 is real [for the precise form of $\mathcal{H}(\tau_1, t_2)$ and the actual values of τ_1 and t_2 , see Appendix A]. $\mathcal{H}(\tau_1, t_2)$ supports one LH chiral mode on edge A, and one RH chiral mode on edge B. For $\mathcal{H}(\tau_1^*, t_2)$ the chiralities are reversed. The spectrum of the target 3450 chiral gauge theory consists of LH fermions with charges 3 and 4, and RH fermions with charges 5 and 0. The physical chiral mode of each lattice species thus always lives on edge A, while the opposite chirality mode on edge B is always the doubler, or mirror fermion. The goal of the SMG program is to gap all the edge-B chiral modes, without breaking spontaneously the U(1) symmetry. Below, it will often be convenient to distinguish between the fermion fields on the two edges using a separate notation: ξ_I for the edge-A fields, and χ_I for the edge-B fields.

We next consider the interaction Hamiltonian, which takes the form

$$H_{\text{int}} = g_1 H_1 + g_2 H_2. \quad (2.5)$$

Here g_1, g_2 are coupling constants, while H_1, H_2 are (lattice sums of) local six-fermion operators. The interaction Hamiltonian couples the edge-B fermions only,

¹⁵The gauge field is turned off in the ZZWY model, i.e., the reduced model is considered in Ref. [44].

$H_{\text{int}} = H_{\text{int}}(\chi_I)$. Notice that the bilinear Hamiltonian H_0 has a separate U(1) fermion number symmetry for every fermion species. Two linear combinations of these U(1) transformations are broken explicitly by H_{int} , while two other linear combinations remain as exact global symmetries of the full Hamiltonian. One of them, of course, corresponds to the gauge symmetry of the 3450 model. For more details on the symmetries, as well as on the interaction terms, see Appendix A 4 or Ref. [44].

Introducing interpolating fields for the (anticipated) bound states,

$$\mathcal{B}_{I,a}(x) = \frac{1}{6} \frac{\delta H_a}{\delta \chi_I^\dagger(x)}, \quad a = 1, 2, \quad (2.6)$$

where the coordinate x labels the edge-B sites, and their sum

$$\mathcal{B}_I(x) = g_1 \mathcal{B}_{I,1}(x) + g_2 \mathcal{B}_{I,2}(x), \quad (2.7)$$

we can reexpress the interaction Hamiltonian as

$$H_{\text{int}} = \sum_x \sum_{I=1}^4 (\chi_I^\dagger(x) \mathcal{B}_I(x) + \text{H.c.}). \quad (2.8)$$

Next, introducing two-component fermion fields

$$\eta_I = \begin{pmatrix} \chi_I \\ \mathcal{B}_I \end{pmatrix}, \quad (2.9)$$

$$\bar{\eta}_I = (\mathcal{B}_I^\dagger \quad \chi_I^\dagger), \quad (2.10)$$

it follows that the interaction Hamiltonian takes the form of a mass term for the (two-dimensional) Dirac fermions $\eta_I(x)$, explicitly,

$$H_{\text{int}} = \sum_x \sum_{I=1}^4 \bar{\eta}_I(x) \eta_I(x). \quad (2.11)$$

The Hamiltonian does not contain explicit kinetic terms for the bound-state fields $\mathcal{B}_I(x)$, but these may be generated dynamically.

Finally let us return to the issue of propagator zeros. Let us assume that near some critical momentum p_0 , the propagator of the η_I field behaves like a massive Dirac propagator,

$$\langle \eta_I \bar{\eta}_I \rangle = \frac{\omega \gamma_0 - (p - p_0) \gamma_1 - m_I}{\omega^2 - (p - p_0)^2 - m_I^2} + \dots, \quad (2.12)$$

where the ellipsis stand for terms suppressed by the lattice spacing, which we will disregard in this section. Here γ_0, γ_1 satisfy the two-dimensional Dirac algebra and, moreover, we assume that the chirality matrix $\gamma_0 \gamma_1$ is equal to the third

Pauli matrix σ_3 in the basis where η_I is given by Eq. (2.9). The two-dimensional chiral projectors are thus $P_{R,L} = \frac{1}{2}(1 \pm \sigma_3)$. If we now consider only the χ_I component, we obtain the projected propagator

$$\langle \chi_I \chi_I^\dagger \rangle = P_R \langle \eta_I \bar{\eta}_I \rangle P_L = P_R \frac{\omega \gamma_0 - (p - p_0) \gamma_1}{\omega^2 - (p - p_0)^2 - m_I^2} P_L. \quad (2.13)$$

Focusing on $\omega = 0$, which will be the relevant case for the discussion in the next section, this projected propagator has a zero eigenvalue for $p - p_0 = 0$. This demonstrates how the mechanism of bound-state formation can turn the propagator zero of Eq. (2.13), anticipated in the SMG phase, into a component of the propagator of a normal, massive Dirac field, Eq. (2.12). As a result, the ghost states associated with an “irremovable” propagator zero are avoided, and the effective theory associated with the long-distance degrees of freedom can be local.

We comment that one can allow for independent renormalization factors, Z_t and Z_s , for the temporal and spatial components in Eq. (2.12). Any deviation of the ratio Z_s/Z_t from unity can be absorbed into a rescaling of the lattice spacing, while a common $Z_s = Z_t$ represents as usual an overall wave-function renormalization factor for the fermion field occurring in the two-point function.

Starting from the explicit form of the interaction Hamiltonian of the ZZWY model, we have identified candidate composite fields which can serve as interpolating fields for the opposite-chirality bound states, that, in turn, should pair up with the gapped elementary chiral fermions in order to form massive Dirac fermions in the SMG phase. Clearly, this construction is very general, and could be applied to a wide range of SMG models. In contrast, proving that the said bound states actually exist is a more difficult task that must be addressed on a case-by-case basis. As a rule, we expect that the existence of the bound states can be proved when a strong-coupling expansion is available. Relevant examples include the strong-coupling phase of the Eichten-Preiskill model [45], analyzed in Ref. [46], as well as the qualitatively similar example presented in Ref. [39]. Additional examples of propagator zeros were discussed in Refs. [47,48]. In particular, Ref. [47] employed a strong-coupling expansion similar to that of Refs. [39,46]. The formation of bound states within the strong-coupling expansion was proved in Ref. [46], but was not explicitly considered in Refs. [39,47,48].

As for the ZZWY model itself, no such strong-coupling expansion is available. The technical reason why such an expansion is not feasible is that the multi-fermion interactions contain hopping terms. In order to study whether or not the conjectured bound states exist one must therefore resort to numerical methods.

Let us recap the (conjectural) situation with regard to zeros of $\mathcal{R}(p)$ in the ZZWY model. The mechanism of bound-state formation offers a natural escape route from the ghost states that would otherwise be associated with propagator zeros in the SMG phase. We conjecture that the set $\{\xi_I, \chi_I, \mathcal{B}_I\}$ forms a complete set of interpolating fields. Apart from the original elementary fermion fields of the model: the edge-A fields ξ_I and the edge-B fields χ_I , this set includes composite fields $\mathcal{B}_I$. The fields $\{\chi_I, \mathcal{B}_I\}$ can pair up to form massive Dirac fermions in the SMG phase. The addition of the composite fields $\mathcal{B}_I$ thus eliminates the propagator zeros found if we use only the edge-A and edge-B elementary fields.

Once the two-point functions of the set $\{\xi_I, \chi_I, \mathcal{B}_I\}$ are free of propagator zeros, and provided that the additional conditions stated in the next section are satisfied, the generalized no-go theorem applies. Then the spectrum in each charge channel must be vectorlike, and thus new doublers must appear. Notice that the theorem does not provide any information about which interpolating fields from our complete set generate massless states, nor on the location of any new degeneracy points that might occur in the Brillouin zone. One possibility that is clearly compatible with the mechanism of bound-state formation is that the elementary edge-B fields χ_I together with the composite fields $\mathcal{B}_I$ generate massive states only, and the new doublers are generated by the edge-A fields ξ_I . In this case it may turn out that the smaller set that consists of the edge-A fields ξ_I only already comprises a complete set of interpolating fields free of propagator zeros in the SMG phase. Such a behavior would in fact resemble the spectrum of the edge-A fields in the strong-coupling limit discussed in Sec. IV and Appendix A 3 below.

We stress, however, that we have no concrete knowledge about the actual situation, and that other scenarios might be possible as well. One such alternative scenario is that the new massless doublers of the edge-A massless fermions arise in the SMG phase as bound states interpolated by yet another set of composite fields, denoted $\mathcal{A}_I$. In this case, the complete set of interpolating fields would take the form of $\{\xi_I, \chi_I, \mathcal{B}_I, \mathcal{A}_I\}$. The rest of the argument is unchanged: once we have succeeded in constructing a complete set of interpolating fields with two-point functions free of propagator zeros, the no-go theorem can be applied.

What if it turns out to be impossible to build any complete set of interpolating fields, whose two-point functions are free of propagator zeros? While we cannot rule out this option, we have given in Ref. [38] strong arguments that such irremovable zeros are ghost states, that render the SMG phase of the theory inconsistent. But as we have already stressed in this paper, it is unlikely that the kind of nonlocality needed for ghost states would develop in the SMG phase of any reduced model with a local action [39]. We thus expect that it should always be possible to

find a complete set of interpolating fields whose associated H_{eff} is free of poles.

III. APPLICABILITY OF THE NIELSEN-NINOMIYA THEOREM

The previous section suggests the following conjectural physical picture. In an SMG phase of a local reduced model, propagator zeros are likely to be kinematical singularities. They arise because the gapped mirror components of the elementary fermion fields combine with opposite-chirality bound states to form massive Dirac fermions. As a result, using the elementary fermion fields only as the set of interpolating fields effectively projects onto a single chirality of these massive Dirac fermions, and the projection generates the propagator zeros. The addition of the bound-state composite fields to the set of interpolating fields then provides us with a complete set, with two-point functions that are free of propagator zeros while their primary singularities are in one-to-one correspondence with the massless fermion asymptotic states.

In this section we present a new proof of the generalized no-go theorem [34,35], making certain assumptions about the field content and the Hamiltonian of the reduced model. With these assumptions, the theorem is valid in both $d = 1$ and $d = 3$ spatial dimensions, and the basic framework applies anywhere in the phase diagram of a given reduced model. However, as we discuss later on, some important elements depend on the properties of the SMG phase in question, and, more generally, on whether we are in two or four dimensions.

We assume that a complete set of fermion interpolating fields has been constructed at the point in the phase diagram under consideration. We will use the notation Ψ_a , with a as a generic index, for all the interpolating fields that belong to this complete set, both elementary and composite. For reasons that will become clear in the proof of analyticity, it is advantageous to consider the *retarded* and *advanced* two-point functions (and not the *time-ordered* functions common in the path-integral framework). The retarded anticommutator is defined by

$$R_{ab}(\vec{x}, t) = i\theta(t)\langle 0 | \{ \Psi_a(\vec{x}, t), \Psi_b^\dagger(\vec{0}, 0) \} | 0 \rangle. \quad (3.1)$$

The space and space-time Fourier transforms are (from now on we omit the indices)

$$\hat{R}(\vec{p}, t) = \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} R(\vec{x}, t), \quad (3.2)$$

$$\tilde{R}(\vec{p}, \omega) = \int_0^\infty dt e^{i\omega t} \hat{R}(\vec{p}, t), \quad \text{Im } \omega > 0. \quad (3.3)$$

Excluding the degeneracy points, we also define ($\epsilon > 0$)

$$\mathcal{R}(\vec{p}) = \lim_{\epsilon \rightarrow 0} \tilde{R}(\vec{p}, \omega = i\epsilon). \quad (3.4)$$

Finally, $H_{\text{eff}}(\vec{p})$ is defined by Eq. (1.1).

For the advanced anticommutator $A(\vec{x}, t)$, one replaces $\theta(t)$ by $-\theta(-t)$ in Eq. (3.1). The corresponding Fourier transforms are denoted $\hat{A}(\vec{p}, t)$ and $\tilde{A}(\vec{p}, \omega)$, with now $\text{Im } \omega < 0$. The definition of $\mathcal{A}(\vec{p})$ is similar to Eq. (3.4), with $\omega = -i\epsilon$ replacing $\omega = +i\epsilon$.

The functions $\mathcal{R}(\vec{p})$ and $\mathcal{A}(\vec{p})$ defined above have the following properties:

- (1) *Equality of $\mathcal{R}(\vec{p}) = \mathcal{A}(\vec{p})$.* Except at degeneracy points, where there are intermediate states with vanishing energy, when ω tends to zero from the relevant half-space of the complex plane the common boundary value of the retarded and advanced correlators, $\tilde{R}(\vec{p}, \omega)$ and $\tilde{A}(\vec{p}, \omega)$, is equally given by $\mathcal{R}(\vec{p})$ or $\mathcal{A}(\vec{p})$.

This property follows immediately by introducing a complete set of intermediate states, and noting that the momentum dependence always takes the form of the familiar energy denominators,

$$\frac{1}{\omega \pm E(\vec{p})}, \quad (3.5)$$

where $E(\vec{p})$ is the energy of an intermediate state with momentum $\vec{p}$ (cf. Appendix B 1). Hence, the limit $\omega \rightarrow 0$ exists, except at the degeneracy points.

- (2) *$\mathcal{R}(\vec{p})$ is Hermitian.* The proof is given in Appendix B 1. By Eq. (1.1), $H_{\text{eff}}(\vec{p})$ is Hermitian as well. Strictly speaking, $\mathcal{R}(\vec{p})$ is undefined at the primary singularities, where it diverges. We define $H_{\text{eff}}(\vec{p})$ at the primary singularities by requiring continuity, hence it is Hermitian everywhere in the Brillouin zone. Continuity at a primary singularity is a corollary of Eq. (3.7) below, which, as discussed later on, follows from the requirement that the continuum limit of the reduced model is a theory of free massless fermions.
- (3) *Analyticity.* $\mathcal{R}(\vec{p})$ is an analytic function of $\vec{p}$ except at the degeneracy points. The proof is given in Appendix B 3. It assumes that the reduced model contains fermion fields only, as well as a strong form of locality of the Hamiltonian, and it makes use of the edge-of-the-wedge theorem.

Assuming that $\mathcal{R}(\vec{p})$ has no zeros, it follows that $H_{\text{eff}}(\vec{p})$ is also an analytic function of $\vec{p}$ except at the degeneracy points. What remains to be discussed is the behavior of $H_{\text{eff}}(\vec{p})$ near these degeneracy points. We will be interested mainly in the primary singularities, which is where $H_{\text{eff}}(\vec{p})$ can have zero-energy eigenstates that correspond to the massless fermion asymptotic states. Secondary singularities were discussed in detail in Refs. [34,35], and we will not

repeat this discussion here. Instead, we give in Appendix B 4 an example of a secondary singularity that highlights the role of such points.

We require that the massless fermion excitations are relativistic. This means that the leading behavior near a primary singularity p_c is

$$E = \pm(p - p_c) + \dots, \quad d = 1, \quad (3.6a)$$

$$H_{2 \times 2} = \pm \vec{\sigma} \cdot (\vec{p} - \vec{p}_c) + \dots, \quad d = 3, \quad (3.6b)$$

where the ellipsis stand for terms suppressed by powers of the lattice spacing.¹⁶ The $\pm$ signs define the chirality of the massless state, with a plus (minus) sign for a RH (LH) state. In words, for $d = 1$ the Hamiltonian must have an eigenvalue that behaves like $\sim \pm(p - p_c)$. In $d = 3$ it must be possible to choose a basis for the Hamiltonian such that all the eigenvectors with near-zero energy are assembled into diagonal 2×2 blocks of the form $\sim \pm \vec{\sigma} \cdot (\vec{p} - \vec{p}_c)$.

Analytic corrections to the relativistic behavior (3.6), which come in powers of $a(p - p_c)$, will not violate the conditions of the NN theorem. For example, as already mentioned in the Introduction, a naive fermion in $1 + 1$ dimensions has $E = \frac{1}{a} \sin(ap)$. This dispersion relation is analytic in the whole Brillouin zone, which is topologically a circle; and indeed, in addition to the RH massless state at $p_c = 0$ there is a LH doubler at $p_c = \pi/a$, consistent with the no-go theorems [4,5].

The remaining condition of the NN theorem that needs to be established is that $H_{\text{eff}}(p)$ has a continuous first derivative at each degeneracy point. In order to determine the form of the leading logarithmic corrections we will adopt a low-energy effective field theory (EFT) approach, applicable near $\vec{p}_c$. Here we will discuss the leading logarithmic corrections near the primary singularities. For the secondary singularities, see Appendix B 4 and Refs. [34,35].

Disregarding any massive states, we require that the continuum limit of the reduced model is a theory of relativistic free massless fermions. This means that there are no massless bosonic states, and that the massless fermions interact via irrelevant interactions only, i.e., interactions which vanish in the continuum limit $a\vec{p} \rightarrow 0$. These interactions must moreover be consistent with the symmetries of the underlying reduced model. The leading logarithmic correction near a given primary singularity will arise from a self-energy diagram with two vertices of the least-irrelevant interaction O_{irr} in the EFT that couples to the corresponding massless fermion state. Nonanalytic self-energy corrections can arise only if all the intermediate states are massless. This is why we can ignore all massive asymptotic states in this

discussion.¹⁷ Since we deal with an EFT, the scaling dimension n of O_{irr} is just its canonical mass dimension, which is integer. Being irrelevant means that $n > d + 1$, or equivalently, since n is integer, $n \geq d + 2$. The coupling constant of O_{irr} can be written as Ga^{n-d-1} where G is dimensionless. The self-energy diagram with two O_{irr} vertices thus gives rise to the generic logarithmic corrections

$$E = \pm q(1 + c_1 G^2 (aq)^{2(n-d-1)} \log(q^2)) + \dots, \quad d = 1, \quad (3.7a)$$

$$H_{2 \times 2} = \pm \vec{\sigma} \cdot \vec{q}(1 + c_3 G^2 (aq)^{2(n-d-1)} \log(q^2)) + \dots, \quad d = 3, \quad (3.7b)$$

where $\vec{q} = \vec{p} - \vec{p}_c$ and c_1, c_3 are numerical constants. Since $n - d - 1 \geq 1$, it follows that H_{eff} has at least a continuous second derivative at each primary singularity. In Appendix B 4 we establish that the same is true at the secondary singularities.

At this point we have established that all the conditions of the NN theorem are satisfied by H_{eff} , thereby proving the following generalized no-go theorem:

Consider a reduced model defined on a regular spatial lattice, with a compact global symmetry G that is not broken spontaneously. The G generators are thus discrete-valued conserved charges. Assume also that (i) the Hamiltonian has a finite range, and depends on fermion fields only; (ii) the continuum limit is a theory of relativistic free massless fermions and with no massless bosons; (iii) in any charge sector which supports at least one massless fermion, one can find a complete set of interpolating fields (as defined in the Introduction) so that the corresponding $\mathcal{R}(\vec{p})$ is free of zeros. Then $H_{\text{eff}}(\vec{p})$ satisfies all the assumptions of the Nielsen-Ninomiya theorem, and as a result, the massless fermion spectrum in this charge sector is vectorlike.

Before we continue, we digress to briefly explain how the generalized theorem works in the case of a *free* Hamiltonian. We require translation invariance, so that the Hamiltonian has the general form $\hat{H} = \sum_{\vec{x}, \vec{y}} \sum_{ab} \psi_a^\dagger(\vec{x}) \mathcal{H}_{ab}(\vec{x} - \vec{y}) \psi_b(\vec{y})$. It can then be shown that, as expected, $H_{\text{eff}}(\vec{p})$ is equal to $\mathcal{H}(\vec{p})$, where $\mathcal{H}(\vec{p})$ is the Fourier transform of $\mathcal{H}(\vec{x} - \vec{y})$. A caveat is that the set of interpolating fields used in the construction of the retarded and advanced two-point functions must include all the fermion fields occurring in the Hamiltonian. Omitting some of these fields will result in $H_{\text{eff}}(\vec{p})$ which is different from $\mathcal{H}(\vec{p})$. Moreover, this can lead to the appearance of zeros in $\mathcal{R}(\vec{p})$, and thus to spurious poles in $H_{\text{eff}}(\vec{p})$. For an example of this phenomenon in the context of the ZZWY model, see Appendix A 2.

¹⁶As in Sec. II, in Eq. (3.6) one can allow for a general ratio Z_t/Z_s of renormalization factors, or equivalently, a renormalization of the speed of light. Compare also Eq. (A6a).

¹⁷Gapped states of the lattice Hamiltonian do play a role, however, for the secondary singularities. See Appendix B 4.

In a strict technical sense, many attempts to construct lattice chiral gauge theories will not be subject to the generalized no-go theorem as stated above, if some of its assumptions are not satisfied. At the technical level, the theory could for example be defined in Euclidean space instead of by a lattice Hamiltonian; or the theory may contain massless scalar fields besides the fermion fields. In two dimensions the situation gets further complicated because four-fermion operators without derivatives are marginal, as we discuss in Sec. V below. However, one expects that the generalized theorem is still relevant, because in many cases one can still construct from the fermion two-point functions of the reduced model at $\omega = 0$ an object with properties similar to those of $\mathcal{R}(\vec{p})$, including in particular the analyticity properties if the underlying theory is local. Letting this object play the role of $\mathcal{R}(\vec{p})$ can again give rise to an $H_{\text{eff}}(\vec{p})$ that satisfies all the requirements of the NN theorem.

Generally speaking, if massless bosons emerge in an SMG phase, their role in the physics of the continuum limit would have to be understood. In four dimensions, the presence of a massless scalar usually signals the spontaneous breaking of a global symmetry, for which this massless scalar is a Nambu-Goldstone boson. This is a situation we would like to avoid in the SMG framework. In two dimensions continuous global symmetries cannot be broken spontaneously [49,50], and long-range order is replaced by quasi-long-range order. Again, this is a situation we would normally like to avoid. The reason is simply that reduced models obtained by turning off the gauge field in a *continuum* chiral gauge theory have a massless spectrum that consists of fermions only.

In the rest of this section we discuss the implications of the generalized no-go theorem in four dimensions. The implications for two-dimensional theories, and for the ZZWY model in particular, are deferred to Secs. V and VI.

Our first comment is that, in four dimensions, the effective theory that describes the long-distance behavior of a set of massless fermions is automatically a theory containing irrelevant interactions only. This is because the interaction with the lowest possible mass dimension, namely a four-fermion interaction, is irrelevant.

An important exception is the gauge-fixing approach to the construction of four-dimensional lattice chiral gauge theories, in which one couples, say, the left-handed fermions to the gauge field, while the right-handed fermions are spectators.¹⁸ This construction allows for a Wilson term to remove the doublers at the nonzero corners of the Brillouin zone. The gauge symmetry is broken on the lattice, but with gauge fixing and counterterms restoring Slavnov-Taylor identities, it is recovered in the continuum limit [27]. In the corresponding reduced model one can

construct an object that satisfies most of the properties of $H_{\text{eff}}(\vec{p})$. In particular, it is Hermitian, and analytic away from the (only) degeneracy point at $\vec{p} = 0$, which is a primary singularity.¹⁹ However, this would-be $H_{\text{eff}}(\vec{p})$ does not have a continuous first derivative at the primary singularity in some channels. The origin of this behavior is the presence of a *higher-derivative* scalar field in the reduced model of the gauge fixing approach, which exhibits a coupling-constant dependent critical exponent. In each charge sector, for one handedness the primary singularity is created by the two-point function of a composite operator, which in turn factorizes as the product of decoupled scalar and massless fermion two-point functions. The corresponding sector of $H_{\text{eff}}(\vec{p})$ does not have a continuous first derivative [24,25].

This works as follows. The group-valued scalar field $\phi(x) \in G$ arises in the reduced model of the gauge-fixing approach as the longitudinal gauge degree of freedom. Its higher derivative kinetic term originates from a covariant gauge-fixing term which is part of the lattice action in this approach. For $G = \text{U}(1)$, one has $(\partial_\mu A_\mu)^2 \rightarrow (\Box\phi)^2$ when we project A_μ onto its longitudinal part, $A_\mu \rightarrow \partial_\mu \phi$. The higher-derivative kinetic term is present in the non-Abelian case as well. The price to pay is an enlarged Hilbert space, that needs to be projected onto a unitary subspace in the continuum limit. This teaches us that there might exist valid dynamical scenarios in which $H_{\text{eff}}(\vec{p})$ will not have a continuous first derivative at its degeneracy points, but also that such dynamics has got to be rather nontrivial. In the gauge-fixing approach the ultimate result is that, in the continuum limit, the fermion spectrum is free, massless, and chiral with respect to the (unbroken) symmetries of the reduced model. In addition, there is a decoupled unphysical sector associated with the higher-derivative scalar field which, as explained above, represents the longitudinal gauge degree of freedom. In the non-Abelian case, the unphysical sector contains the ghost fields as well.

IV. DECOUPLING IN THE STRONG COUPLING LIMIT

In this section we turn to a different, but related, topic. We consider reduced models in which the fermion fields can be divided into two sets. The first set, denote χ , includes all the fermion degrees of freedom that participate in some strong interaction. The other set, denoted ξ , includes the remaining fermion degrees of freedom, that do not directly participate in any strong interaction. We will then show that in the (uniform) strong-coupling limit the ξ and χ sectors decouple. The relevance of this result is that the decoupled ξ sector is by assumption either free or weakly coupled.

¹⁸The right-handed fermions decouple because of a shift symmetry [51].

¹⁹The gauge-fixing approach is defined in Euclidean space. In this context we define the degeneracy points as the points where the would-be $H_{\text{eff}}(\vec{p})$ is not analytic, and the primary singularities as the points where moreover this $H_{\text{eff}}(\vec{p})$ has a zero eigenvalue.

Therefore, the NN theorem applies to the ξ sector, and its spectrum must be vectorlike in this limit.

To avoid unnecessary technicalities we will present the main result while considering a reduced model that depends on fermion fields only, which are subject to strong interactions only. We will later comment on generalizations that broaden the scope of the result. The total Hamiltonian thus has the form²⁰

$$H = H_0(\xi, \chi) + H_{\text{int}}(\chi; g_1, g_2, g_3, \dots). \quad (4.1)$$

Here $g_1, g_2, g_3, \dots$ are the coupling constants of the various interactions terms, all strong, that couple the χ degrees of freedom. By assumption, the ξ degrees of freedom occur only in the bilinear part H_0 of the Hamiltonian, where they can mix with the χ degrees of freedom.

We now consider the uniform strong-coupling limit defined by writing

$$H = H_0(\psi, \chi) + H_{\text{int}}(\chi; g_1, g_1 \lambda_2, g_1 \lambda_3, \dots), \quad (4.2)$$

where $\lambda_i = g_i/g_1$, $i = 2, 3, \dots$, and then taking the limit $g_1 \rightarrow \infty$ while holding all the λ_i fixed. We first assume that all the interaction terms have the same degree of homogeneity n . We can then rescale the interacting degrees of freedom,

$$\chi \rightarrow g_1^{-1/n} \chi, \quad (4.3)$$

and the Hamiltonian will take the form

$$H = H_0(\psi, g_1^{-1/n} \chi) + H_{\text{int}}(\chi; 1, \lambda_2, \lambda_3, \dots). \quad (4.4)$$

Note that now g_1 occurs explicitly only in H_0 . Finally taking the limit $g_1 \rightarrow \infty$ at fixed λ_i , the Hamiltonian becomes

$$H = H_0(\xi, 0) + H_{\text{int}}(\chi; 1, \lambda_2, \lambda_3, \dots). \quad (4.5)$$

The χ and ξ sets have now decoupled: the χ 's occur only in H_{int} , while the ξ 's occur only in H_0 . Since H_0 is bilinear, it is subject to the NN theorem, and thus the spectrum of the ξ 's must be vectorlike in this limit, if $H_0(\xi, 0)$ satisfies the conditions of the theorem. This is the main result of this section.

If the degree of the interaction term with coupling g_i is n_i , with not all n_i equal, we can consider the following strong-coupling limit. We choose g_1 as the coupling (or, one of the couplings) whose degree of homogeneity n_1 has the smallest value. After rescaling $\chi \rightarrow g_1^{-1/n_1} \chi$, the Hamiltonian becomes

$$H_0(\psi, g_1^{-1/n_1} \chi) + H_{\text{int}}(\chi; 1, g_1^{1-n_2/n_1} \lambda_2, g_1^{1-n_3/n_1} \lambda_3, \dots), \quad (4.6)$$

and after taking the limit $g_1 \rightarrow \infty$ with the λ_i 's fixed as before, we obtain

$$H = H_0(\xi, 0) + H_{\text{int}}(\chi; 1, \lambda_2, \lambda_3, \dots, 0, 0, \dots), \quad (4.7)$$

where $\lambda_2, \lambda_3, \dots$, correspond to any additional interaction terms with the same degree of homogeneity n_1 , whereas the following zeros correspond to all other interaction terms, whose degree of homogeneity is larger than n_1 . Once again, the ξ 's decouple from the χ 's.

These results easily generalize to the case that the reduced model contains also scalar fields, as well as to the case that the ξ 's interact weakly via additional couplings $y_1, y_2, \dots$, that are kept fixed (and small) when the uniform strong-coupling limit is taken. In all of these more general cases, the end result is again that the ξ 's decouple from the χ 's, and, since the ξ sector is either free or weakly coupled, the ξ 's are subject to the NN theorem.

A similar result was previously derived in the so-called waveguide model [52]. That result is now seen to be a special case of the more general phenomenon considered here.

Another model where a similar strong-coupling limit was considered is the Eichten-Preskill (EP) model [45], analyzed in detail in Ref. [46]. The model has a strong-coupling symmetric (PMS) phase. In that phase, however, all the fermion degrees of freedom participate in the strong interaction. In other words, the ξ set is empty. As for the χ set (which includes all the fermion degrees of freedom), a strong-coupling expansion can be used to show that the spectrum in the PMS phase is vectorlike, with the fermion mass tending to infinity in the strong-coupling limit.²¹ The massive Dirac fermions have one chirality component which is elementary in terms of the lattice fields, while the other chirality component is composite. This follows the scenario of bound-state formation we have discussed in Sec. II.

As for the ZZWY model, the interaction Hamiltonian couples only the edge-B fermions, the χ 's, whereas the edge-A fermions, the ξ 's, occur only in the bilinear Hamiltonian. By the theorem derived in this section, in the strong-coupling limit the χ 's and ξ 's decouple. The ξ sector becomes free, and its massless spectrum consists of a single Dirac fermion per fermion species. For more details, see Appendix A 3.

An open question regarding the ZZWY model is whether its phase diagram consists of two phases only (those found in Ref. [44]), or, alternatively, that three (or more) phases

²⁰Of course, one can alternatively work in a Lagrangian formalism.

²¹The PMS phase also has massive composite scalar bound states, and the boundary of the PMS phase is defined by the their mass-squared going negative, indicating the onset of a phase with spontaneous symmetry breaking.

may exist. In the latter case the SMG phase could be an intermediate phase, which borders a weak-coupling phase on one side, and a strong-coupling phase on the other side. More detailed information on the phase diagram should help understand to what extent the strong-coupling limit we studied here is relevant for the properties of the SMG phase.

V. THE GENERALIZED NO-GO THEOREM IN TWO DIMENSIONS

In this section we turn to the implications of the generalized no-go theorem for two-dimensional theories, including in particular the ZZWY model.

The main difference between four dimensions and two dimensions is the following. Assume that the asymptotic states of the reduced model consist of massless fermions only. As we pointed out in Sec. III, in four dimensions this guarantees that all the interactions of these fermions will be irrelevant. The reason is simply that the lowest-dimension interaction is a four-fermion operator, whose mass dimension is six (or higher; depending on whether or not it contains derivatives), which is always irrelevant.

In contrast, in two dimensions a fermion field has mass dimension one half. Derivativeless four-fermion operators thus have mass dimension two, hence they represent renormalizable interactions. In the ZZWY model, the continuous global symmetry of the bilinear part of the Hamiltonian is $U(2) \times U(2)$. Taking into account Fermi statistics, there are three linearly independent, local four-fermion operators which are built from the “physical” edge-A fields only, and are invariant under the full symmetry group. They can be taken to be $j_R j_R$, $j_L j_L$ and $j_R j_L$, where²² $j_L = \sum_{I=1,2} \bar{\xi}_I \xi_I$ and $j_R = \sum_{I=3,4} \bar{\xi}_I \xi_I$. These operators do not require point splitting, and so in the continuum limit they indeed turn into derivativeless, renormalizable four-fermion interactions. When the interactions of the ZZWY model are turned on, the global symmetry reduces to the smaller group $U(1)^2$ (see Appendix A 4). As a result, more four-fermion operators are allowed by the symmetry of the full theory. Now there are altogether six operators made out of the edge-A fields, which can be taken to be $(\bar{\xi}_I \xi_I)(\bar{\xi}_J \xi_J)$ with $I \neq J$.

Induced four-fermion operators in two-dimensional theories have largely been ignored in the SMG literature. In quantum field theory, “everything that can happen will happen,” and thus all these four-fermion operators are expected to be induced. This includes in particular the SMG phase discussed in Ref. [40], and the SMG phase of the ZZWY model.

In the face of this situation, what options are available? The original goal of the SMG paradigm is to construct lattice chiral gauge theories which are not “contaminated”

by any other interactions or massless states not present in the target continuum theory. In two dimensions, this means that the coefficients of all the induced four-fermion interactions without derivatives must be tuned to zero by adding suitable “counterterms” to the lattice theory. Doing so would clearly be a formidable task, given the large number of the expected four-fermion interactions and the technical challenge of determining the induced coupling of every one of them. Nevertheless, let us assume that this has been done in a given two-dimensional reduced model.²³ The continuum limit would then be a theory of free massless fermions, which is therefore subject to the generalized no-go theorem of Sec. III, just like in the four-dimensional case. The massless fermion spectrum in the continuum limit will therefore be vectorlike.

In the actual ZZWY model no effort was made to trace the induced four-fermion interactions, and certainly not to tune them to zero. It must therefore be assumed that all four-fermion interactions consistent with the symmetries of the model are present with $O(1)$ couplings.²⁴ This raises the next question: Is it possible to evade the generalized no-go theorem and build a two-dimensional lattice chiral gauge theory at the price of inducing additional four-fermion interactions? Technically, the presence of marginal (equivalently, renormalizable) interactions in the reduced model invalidates the requirement that $H_{\text{eff}}(p)$ has a continuous first derivative at all the degeneracy points. Indeed, if the self-energy correction in Eq. (3.7a) comes from a marginal operator, we will have $n - d - 1 = 0$. Hence, while $E(p)$ remains a continuous function of p , its derivative is not. A similar conclusion applies if the reduced model has massless bosonic states in addition to the massless fermions, which allows for additional relevant and marginal operators.

Nevertheless, closer scrutiny reveals that doublers may still be present. Our strategy at this point is to appeal to the simplest form of the no-go theorem put forward by Karsten and Smit [4], and apply it to $H_{\text{eff}}(p)$. We consider an eigenvalue $E(p)$ of $H_{\text{eff}}(p)$, assuming only that the dispersion curve $E(p)$ is continuous. Then, as p traverses the periodic Brillouin zone, every crossing of $E(p)$ from negative to positive values must be followed by a crossing in the opposite direction, from positive to negative values.²⁵

²³Notice that it is a nontrivial dynamical question whether an SMG phase would survive such tuning or not.

²⁴The actual point splittings of the six-fermion interaction terms of the ZZWY model were not specified in Ref. [44]. Assuming concrete point splittings for definiteness, we have checked that (derivativeless) four-fermion interactions built from the edge-A fields are induced naturally within the strong-coupling expansion discussed in Sec. IV.

²⁵In general the dispersion curve $E(p)$ can wrap around the Brillouin more than once. In that case, crossings from both negative to positive and from positive to negative values can happen at the same value of p . Notice also that a discontinuous derivative of the form $E(p) \sim |p|$ is not consistent with relativistic invariance.

²²We suppress the Dirac matrices, since for two-dimensional Weyl fields they are purely phase factors.

If, at a zero crossing of $E(p)$ from negative to positive (positive to negative) values, the first derivative is continuous as well, then this zero crossing is a primary singularity associated with a RH (LH) massless fermion. When the first derivative is not continuous at the crossing, the physical nature of the crossing point may not be easy to understand, and it requires a detailed investigation within the model under consideration. Without such a detailed investigation it may not be possible to determine whether or not an independent massless fermion state can be associated with each crossing point. In Appendix B 4 we present a few examples of weakly coupled two-dimensional theories with renormalizable four-fermion interactions, in which the zero eigenvalues of $H_{\text{eff}}(p)$ in the relevant channels can be identified straightforwardly as massless RH or LH fermions, and thus the spectrum is vectorlike.

Nontrivial examples include the strong-coupling symmetric phase of the Smit-Swift model in two dimensions [53], and the gauge-fixing approach to the construction of four-dimensional lattice chiral gauge theories, already discussed in Sec. III. In the latter case the essential feature is the presence of a higher-derivative scalar with a $1/(p^2)^2$ propagator, and thus a similar behavior might be found in two dimensions in the presence of an ordinary massless scalar with a $1/p^2$ propagator. Of course, in both cases some mechanism must be present to tame the infrared divergence associated with the scalar field.²⁶ We stress that in all of these cases, after turning the gauge field back on and taking the continuum limit, additional massless fields and/or additional relevant or marginal interactions will necessarily be present, on top of the physical fields and interactions of the target chiral gauge theory. However, in the case of the gauge-fixing approach all the additional degrees of freedom are unphysical: they include the longitudinal degrees of freedom of the gauge field and ghost fields, that can be shown to decouple from the physical sector in the usual way (provided that the fermion spectrum is anomaly free), at least to all orders in perturbation theory.

VI. ZZWY MODEL—THE OPEN QUESTIONS

We have discussed the conditions for the applicability of the generalized no-go theorem in Sec. III, and the special properties of two-dimensional theories in Sec. V. It should be clear from these discussions that at this point we have not reached a final verdict concerning the ZZWY model. In this section we will summarize the current situation, and offer a “road map” consisting of the open questions that would have to be sorted out before it can be firmly established whether or not the SMG phase of the ZZWY model successfully reproduces the chiral massless spectrum of the 3450 model in the continuum limit. While this

section is tied to the ZZWY model, the questions we formulate will need to be considered for any other SMG model as well.

Our starting point is the following observation. Let us consider the retarded anticommutator $R(\vec{x}, t) = R_{\text{elm}}(\vec{x}, t)$, where the subscript “elm” indicates that the set of lattice fields used in the definition (3.1) consists of the *elementary* fields of the model only. In the notation of Sec. II this set is $\{\xi_I, \chi_I\}$, where ξ_I are the edge-A fields and χ_I the edge-B fields. We similarly attach the same subscript to the associated correlators defined by Eqs. (3.2)–(3.4) as well as to the corresponding one-particle Hamiltonian $H_{\text{eff,elm}}(p) = \mathcal{R}_{\text{elm}}^{-1}(p)$ defined via Eq. (1.1). The properties of these correlators were not studied in Ref. [44] in sufficient detail, yet the general considerations found in the condensed matter literature [30,39,47,48], and reviewed in Sec. II, strongly suggest that $\mathcal{R}_{\text{elm}}(p)$ develops zeros in the SMG phase, and thus $H_{\text{eff,elm}}(p)$ develops poles.

Naively, the presence of poles in $H_{\text{eff,elm}}(p)$ appears to be good news, as it suggests that the NN theorem does not apply and thus the massless fermion spectrum might yet be chiral in the SMG phase. However, as we showed in Ref. [38], if the effective low-energy theory contains poles in (the bilinear part of) its Hamiltonian, this implies the existence of *ghost* states which make the model inconsistent.

The next development came with Ref. [39]. This paper discussed a model where $H_{\text{eff,elm}}(p)$ indeed has poles in the SMG phase. Nevertheless, much like in the EP model [45], a strong coupling expansion can be used to prove that there are no ghost states in the SMG phase [46]. This leads us to the following conjecture, that we now believe is valid very generally, including in the SMG phase of the ZZWY model, even though no strong-coupling expansion exists in this case. The conjecture consists of two parts. The first part is that *if the underlying theory is local then the effective low-energy theory cannot contain ghost states anywhere in the phase diagram*. The second part is that, as a corollary, *it should always be possible to construct a complete set of interpolating fields, and thus $\mathcal{R}(p)$ for this set of interpolating fields has no zeros*. To recall, a complete set of interpolating fields was defined in the Introduction section by the two requirements that $\mathcal{R}(p)$ is free of propagator zeros, while the zeros of $H_{\text{eff}}(p)$ are in one-to-one correspondence with the massless fermion asymptotic states in the given channel.

According to this conjecture any propagator zeros found in an SMG phase must be *kinematical* zeros. What this means is that in the SMG phase there exist additional asymptotic states (apart from those accounted for by the elementary fermion fields), and the propagator zeros arise because of missing interpolating fields for these new asymptotic states. As we proposed in Sec. II, the physical mechanism at work is bound-state formation. That mechanism was established in the EP model using its strong-coupling expansion [46]; we similarly expect that it can be

²⁶For the gauge-fixing approach, see Refs. [23–25].

established in the model of Ref. [39], again using the strong-coupling expansion.

Addressing the ZZWY model, which lacks a strong-coupling expansion in the SMG phase, we proposed in Sec. II a general procedure for identifying the bound states that combine with the gapped elementary mirror fermions to form massive Dirac fermions. We stress that while these massive Dirac fermions are expected to decouple in the continuum limit, in order to avoid propagator zeros one should nevertheless add to the set of interpolating fields suitable composite fields to generate the bound-state component of each massive Dirac fermion.²⁷ It goes without saying that further (numerical) investigations of the ZZWY model are required in order to establish whether the scenario of bound-state formation is indeed at work in the SMG phase. If true, a complete set of interpolating fields can be constructed in the SMG phase of the ZZWY model in every charge sector which supports (single) massless fermion asymptotic states (as discussed in detail in Sec. II), and one could proceed to examine the conditions for the applicability of the generalized no-go theorem.

The conditions of the generalized no-go theorem were listed in Sec. III. Condition (i) is satisfied by construction for the ZZWY model: the model depends on fermion fields only, and has a finite-range Hamiltonian. Next, following the preceding discussion, we expect condition (iii) to be satisfied because of our conjecture: there exists a complete set of interpolating fields at every point in the phase diagram and in every charge sector which admits single-particle massless fermion states.

The remaining condition of the theorem is condition (ii), which asserts that the continuum limit is a theory of free massless fermions. However, as we have discussed in Sec. V it is practically certain that this condition is *not* satisfied by the ZZWY model, because all the renormalizable four-fermion interactions consistent with the symmetries of the interacting theory must have been induced in the SMG phase. This particular concern is, of course, limited to models in two dimensions.

As we then explained in Sec. V, at this point there are two basic alternatives. The first is to stick to the goal of obtaining in the continuum limit a chiral gauge theory (once the gauge field has been turned back on), without any additional four-fermion interactions. This would require the modification of the ZZWY model by new four-fermion counterterms, whose couplings would be tuned so as to eliminate the renormalizable four-fermion interactions that will be induced in the ZZWY model in its present form. Once this—technically challenging—task would have been completed, it is an open question whether or not the SMG phase will survive. In any case, once all the renormalizable

four-fermion interactions have been tuned away by the counterterms, condition (ii) of the generalized no-go theorem will also be satisfied. As a result, the spectrum will be vectorlike.

The only alternative would be to opt for a different goal, by allowing the target continuum chiral gauge theory to depend on the additional, induced four-fermion interactions. We believe this is the best one can hope for in the ZZWY model in its present form, given that no attempt was made to track and subtract the induced four-fermion interactions in this model. As we have further discussed in Sec. V, while the generalized no-go theorem is technically no longer applicable, also in this case it is far from obvious that the massless spectrum will be chiral. We also remark that allowing for four-fermion interactions carries with it the risk of drastically altering the dynamics of the (chiral) gauge theory. Indeed it is well known that in two dimensions, four-fermion interactions can by themselves give rise to fermion mass generation without symmetry breaking [54].

Finally, we mention the direct test of the massless fermion spectrum carried out in Ref. [40] within their attempt to put on the lattice the same target theory, the two-dimensional 3–4–5 Abelian chiral gauge theory. In the reduced model, one considers the two-point function of the conserved current of the global U(1) symmetry to be gauged, and calculates (numerically, if necessary) its zero-momentum discontinuity. As we explained in the Introduction, the magnitude of this discontinuity provides a direct test whether or not the massless fermion spectrum can coincide with the chiral spectrum of the target 3–4–5 Abelian chiral gauge theory.

VII. CONCLUSION

The symmetric mass generation, or SMG, paradigm aims to construct lattice chiral gauge theories by finding an SMG phase in the reduced model (where the gauge field has been turned off) in which the to-be-gauged exact global symmetry is not broken spontaneously, while the massless fermion spectrum is chiral with respect to that symmetry, and yields the (anomaly-free) spectrum of the target chiral gauge theory. To accomplish this, interactions are introduced in the reduced model whose sole purpose is to gap the unwanted mirror fermions predicted by the Nielsen-Ninomiya theorem.

The SMG literature stresses that a necessary condition for the success of the SMG approach is that all symmetries (continuous or discrete) which are anomalous in the target chiral gauge theory must be broken explicitly in the lattice theory. But since the fermion spectrum is determined at the level of the reduced model, this condition must apply already in the reduced model, even though the gauge field has been turned off. In other words, this condition translates into the requirement that all the exact (global) symmetries of the reduced model in the relevant SMG phase must not

²⁷See Sec. II for further discussion of scenarios for the construction of a complete set of interpolating fields in the SMG phase of ZZWY model.

have an anomaly in the target chiral gauge theory.²⁸ We will refer to this requirement as the SMG anomaly paradigm.

While it seems easy to motivate the SMG anomaly paradigm at an intuitive level, it has remained at odds with numerous lattice studies. The paradigm goes back to the Eichten-Preskill model [45]. However, it was shown long ago that the SMG phase of the Eichten-Preskill model (known in the relevant literature as a PMS phase) fails to support a chiral massless spectrum [46] for essentially the same dynamical reasons as the corresponding phase of the Smit-Swift model [55–58], even though the Eichten-Preskill model adhered to the SMG anomaly paradigm, whereas the Smit-Swift model did not. The Nielsen-Ninomiya theorem, and its generalization discussed in this paper, are also oblivious to the SMG anomaly paradigm. It is thus a logical possibility that the underlying reason for the potential failure of the SMG approach is the remarkably wide scope of the Nielsen-Ninomiya theorem, as reflected by the generalized no-go theorem, while the SMG anomaly paradigm does not play an important role.

The original Nielsen-Ninomiya theorem does not apply in an SMG phase, because it is a theorem about free lattice theories. However, the generalization of this theorem to the case that interactions of arbitrary strength are present applies in principle everywhere in the phase diagram of any reduced model, including in an SMG phase. Physically, what allows for the generalized theorem is the requirement that the continuum limit of the reduced model will be a theory of free massless fermions. Of course, for the SMG paradigm to succeed, the spectrum of these massless fermions must be chiral. These requirements, if satisfied, ensure that after the gauge field is turned on one recovers a chiral gauge theory without any additional “contaminating” interactions. But, basically because the continuum limit of the reduced model has to be a free theory, starting from the fermion two-point functions of the reduced model it is often possible to construct a one-particle lattice Hamiltonian H_{eff} that will satisfy all the conditions of the Nielsen-Ninomiya theorem. Once the Nielsen-Ninomiya theorem applies, the spectrum must be vectorlike.

In this paper, the required properties of H_{eff} were established in a fairly specific setting: a lattice Hamiltonian defined on a spatial lattice, which depends on fermion fields only, and has a finite range. However, experience teaches us that objects with essentially the same properties as this paper’s H_{eff} can be constructed in a much more general setting. This includes lattice models formulated within the Euclidean path-integral framework, where two-point functions with similar features as $\mathcal{R}(\vec{p})$ can be constructed for $\omega = 0$, including similar analyticity properties if the underlying theory is local. Whenever such a construction is possible, the Nielsen-Ninomiya theorem eventually applies.

The starting point of this paper was the role of propagator zeros in an SMG phase, previously discussed in Refs. [38,39]. Our renewed examination of this issue strongly suggests that a complete set of interpolating fields whose two-point functions are free of propagator zeros should exist everywhere in the phase diagram of any local reduced model built along the lines of the SMG paradigm (see Sec. VI for a summary). While finding the desired complete set in any particular model might require a trial-and-error process, we stress that the inability to build such a complete set would likely imply that the propagator zeros represent “irremovable” ghost states. This situation is not only highly unlikely given that the underlying theory is local [39], but moreover, if realized, would render the theory inconsistent [38].

As we have extensively discussed in this paper, a key assumption of the generalized no-go theorem is that the continuum limit of the reduced model is a theory of free massless fermions. This requirement leads to a major difference between four-dimensional and two-dimensional theories. In four dimensions, once the massless spectrum consists of fermions only, the continuum limit is automatically a free theory, because all multifermion interactions are irrelevant in four dimensions. This makes it easier to satisfy the conditions of the generalized no-go theorem. In contrast, four-fermion interactions without derivatives are renormalizable in two dimensions. This not only makes the situation in two dimensions much more complicated than in four dimensions, it also significantly limits any lessons that might be drawn from two-dimensional models for the physically interesting case of four dimensions.

ACKNOWLEDGMENTS

We thank Yi-Zhuang You and Cenke Xu for fruitful discussions. M. G. and Y. S. both benefited from the workshop on Symmetric Mass Generation held in May 2024 at the Simons Center for Geometry and Physics at SUNY Stony Brook. M. G. also thanks the Kavli Institute of Theoretical Physics at UC Santa Barbara, where part of this work was carried out during the program “What is Particle Theory,” for hospitality. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award No. DE-SC0013682 (M. G.). Y. S. is supported by the Israel Science Foundation under Grant No. 1429/21.

DATA AVAILABILITY

The data supporting this study’s findings are available within the article.

APPENDIX A: THE ZZWY HAMILTONIAN

In this appendix we provide some technical details about the Hamiltonian of the ZZWY model [44]. In Appendix A 1

²⁸See, for example, Ref. [30] and references therein.

we discuss the bilinear part of the Hamiltonian H_0 and its spectrum. In Appendix A 2 we discuss the inverse of the free Hamiltonian, and give another example of a propagator zero. In Appendix A 3 we discuss the effect of the decoupling of the edge-A and edge-B fermions in the strong-coupling limit of Sec. IV. In Appendix A 4 we briefly discuss the interaction Hamiltonian H_{int} and its symmetries.

1. The bilinear Hamiltonian

Here we discuss the bilinear part of the Hamiltonian, defined in Eq. (2.4). For $I = 1, 2$, namely for the fermion species with charges 3 and 4 under the U(1) symmetry to be gauged, the Hamiltonian matrix is $\mathcal{H}(\tau_1, t_2)$; whereas for $I = 3, 4$, or the fermion species with charges 5 and 0, it is $\mathcal{H}(\tau_1^*, t_2)$. The actual values of the parameters in the ZZWY model are

$$\begin{aligned} \tau_1 &= t_1 e^{i\frac{\pi}{4}} = e^{i\frac{\pi}{4}}, \\ t_2 &= 1/2. \end{aligned} \quad (\text{A1})$$

Here we will focus on the bilinear Hamiltonian for one of the first two species, $\hat{H} = \bar{\psi} \mathcal{H}(\tau_1, t_2) \psi$, omitting the index I .

The fermion fields reside on the lattice shown in Fig. 1, which constitutes two interconnected one-dimensional chains. The unit cell of this lattice is a 2×1 rectangle in units of the lattice spacing a . The (one-dimensional) momentum p thus lives in a Brillouin zone $0 \leq p \leq 2\pi/(2a) = \pi/a$. Henceforth we will set $a = 1$.

The unit cell has four independent degrees of freedom, which are associated with the four sublattices indicated by the numbers inside the circles in Fig. 1, which start at the upper-left corner of the unit cell and go clockwise. Consistent with Fig. 1(c) of Ref. [44], we will also refer to the upper chain in our figure as edge A, and to the lower chain as edge B. Hence edge A consists of sublattices 1 and 2, while edge B consists of sublattices 3 and 4.

Labeling the degrees of freedom of the four sublattices as ψ_k , $k = 1, 2, 3, 4$, one has²⁹

$$\begin{aligned} \hat{H} &= \sum_x t_1 (e^{i\pi/4} \psi_1^\dagger(2x) (\psi_2(2x+1) + \psi_2(2x-1)) + e^{-i\pi/4} \psi_4^\dagger(x) (\psi_3(2x+1) + \psi_3(2x-1))) \\ &+ e^{-i\pi/4} \psi_1^\dagger(2x) \psi_4(2x) + e^{i\pi/4} \psi_2^\dagger(2x+1) \psi_3(2x+1) + \text{H.c.}) + \sum_x t_2 (\psi_1^\dagger(2x) (\psi_3(2x-1) - \psi_3(2x+1)) \\ &+ \psi_4^\dagger(2x) (\psi_2(2x-1) - \psi_2(2x+1)) + \text{H.c.}). \end{aligned} \quad (\text{A2})$$

In this expression we have kept the parameters t_1, t_2 free, to make it easier to trace the origin of the various terms. In terms of the degrees of freedom associated with the four sublattices, the Hamiltonian matrix in momentum space is

$$\mathcal{H} = \begin{pmatrix} 0 & 2t_1 e^{i\frac{\pi}{4}} \cos(p) & -2it_2 \sin(p) & t_1 e^{-i\frac{\pi}{4}} \\ 2t_1 e^{-i\frac{\pi}{4}} \cos(p) & 0 & t_1 e^{i\frac{\pi}{4}} & 2it_2 \sin(p) \\ 2it_2 \sin(p) & t_1 e^{-i\frac{\pi}{4}} & 0 & 2t_1 e^{i\frac{\pi}{4}} \cos(p) \\ t_1 e^{i\frac{\pi}{4}} & -2it_2 \sin(p) & 2t_1 e^{-i\frac{\pi}{4}} \cos(p) & 0 \end{pmatrix}. \quad (\text{A3})$$

The spectrum of this Hamiltonian is shown in Fig. 2, using the values of the parameters from Eq. (A1). This reproduces Fig. 1(b) of Ref. [44].

As can be seen in Fig. 2, the spectrum consists of two massless and two massive branches,³⁰ and the critical momentum where the energy of the two chiral modes vanishes is $p = \frac{\pi}{2}$. Expanding to quadratic order in the relative momentum $k = p - p_c = p - \frac{\pi}{2}$, and solving the eigenvalue equation, we find that the eigenvectors of the two chiral modes are (in an arbitrary normalization)

$$v_+ = \begin{pmatrix} e^{-i\frac{\pi}{4}} \frac{k}{8} \\ -\frac{ik}{8} \\ e^{-i\frac{3\pi}{4}} \\ 1 \end{pmatrix}, \quad v_- = \begin{pmatrix} e^{-i\frac{3\pi}{4}} \\ -1 \\ e^{-i\frac{\pi}{4}} \frac{k}{8} \\ \frac{ik}{8} \end{pmatrix}, \quad (\text{A4})$$

²⁹We arbitrarily place sublattices 1 and 4 on the even sites, and thus sublattices 2 and 3 on the odd sites. We thank YiZhuang You for some clarifications about the structure of the lattice.

³⁰Notice that the spectrum consists of a single continuous curve that wraps around the Brillouin zone four times.

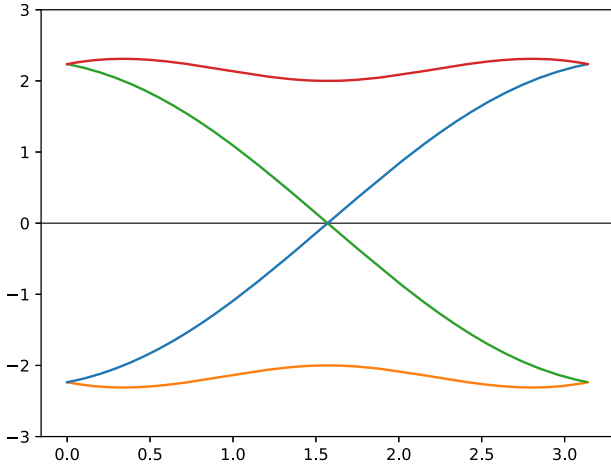


FIG. 2. Spectrum of the bilinear Hamiltonian $\hat{H}$, reproducing Fig. 1(b) of Ref. [44]. The horizontal axis runs from 0 to π . The green (blue) branch corresponds to the LH (RH) chiral mode supported mainly on edge A (edge B). The plot shows that the spectrum has the correct mod π periodicity. Also seen is that the eigenvalues spectrum forms a single smooth curve that winds around the Brillouin zone four times.

while the eigenvectors of the two massive modes are

$$u_+ = \begin{pmatrix} e^{-i\frac{\pi}{4}} \left(1 + k + \frac{k^2}{2}\right) \\ i \left(1 + k + \frac{k^2}{2}\right) \\ e^{i\frac{\pi}{4}} \\ 1 \end{pmatrix}, \quad u_- = \begin{pmatrix} e^{i\frac{\pi}{4}} \\ -1 \\ e^{-i\frac{\pi}{4}} \left(1 + k + \frac{k^2}{2}\right) \\ -i \left(1 + k + \frac{k^2}{2}\right) \end{pmatrix}. \quad (\text{A5})$$

The energies are

$$E(v_{\pm}) = \pm 2k, \quad (\text{A6a})$$

$$E(u_{\pm}) = \pm \left(2 + \frac{k^2}{2}\right). \quad (\text{A6b})$$

For the chiral mode v_+ , the relative momentum k and energy E have the same sign, making it the RH mode by definition. For v_- they have opposite signs, and thus it is the LH mode. As can be seen in Eq. (A4), the RH mode v_+ is supported mainly on sublattices 3 and 4, that is, on edge B, while the LH mode v_- is supported mainly on sublattices 1 and 2, hence on edge A.³¹

Appealing to the tensor-product matrix notation introduced in Appendix A 2 below, the matrix $[\sigma_1 \otimes \sigma_3]$ anticommutes with $\mathcal{H}(p)$. Hence, the application of $[\sigma_1 \otimes \sigma_3]$ to an eigenvector with eigenvalue E gives rise to an

³¹For any $k \neq 0$, the support of the chiral mode has a small component on the opposite edge.

eigenvector with eigenvalue $-E$, as can be verified using the explicit form of the eigenvectors in Eqs. (A4) and (A5). Notice that the factor of σ_1 in this tensor product interchanges the two edges, consistent with the fact that the RH and LH modes live on opposite edges.

Turning to the last two species in Eq. (2.4), it is straightforward to check that if $v_+(p)$ is an eigenmode of $\mathcal{H}(\tau_1, t_2)$ with relative momentum $k = p - p_c$, then $v_+^*(p)$ is an eigenmode of $\mathcal{H}(\tau_1^*, t_2)$ with the same energy $E(p)$ but with an opposite relative momentum. The behavior of $v_-(p)$ is similar. It follows that the roles of the RH and LH modes are interchanged for $\mathcal{H}(\tau_1^*, t_2)$.

Recall that in Ref. [44] the chiral modes on edge A of the four fermion species are identified with the fields of the target 3450 model, and the goal is to gap the mirror modes on edge B. It follows that the chiral modes with charges 3 and 4 of the target 3450 model are LH, while the chiral modes with charges 5 and 0 are RH.

2. Inverse of the bilinear Hamiltonian and propagator zeros

Here we construct the inverse of the momentum-space Hamiltonian matrix $\mathcal{H}(p)$, Eq. (A3), and examine its behavior near the primary singularity $p_c = \pi/2$. This allows us to give an example of a propagator zero, obtained by projecting $\mathcal{H}^{-1}(p)$ on the fermion fields of one edge only.

We begin by introducing a convenient representation of $\mathcal{H}(p)$ in terms of tensor products of Pauli matrices,

$$\mathcal{H}(p) = -2 \cos(p) [I \otimes \tilde{\sigma}_2] + [\sigma_1 \otimes \tilde{\sigma}_1] + \sin(p) [\sigma_2 \otimes \sigma_3], \quad (\text{A7})$$

where we have used the values of τ_1 and t_2 in Eq. (A1). The tensor products correspond to the splitting of $\mathcal{H}(p)$ into 2×2 blocks. In each tensor product, the first factor corresponds to the block structure, and as a basis for it we take the usual Pauli matrices together with the identity matrix I . The second factor corresponds to the internal structure of the blocks, and as a basis we use a rotated set of Pauli matrices, again together with the identity matrix. The rotated matrices are

$$\begin{aligned} \tilde{\sigma}_1 &= e^{-i\frac{\pi}{8}\sigma_3} \sigma_1 e^{+i\frac{\pi}{8}\sigma_3} = 2^{-1/2}(\sigma_1 + \sigma_2), \\ \tilde{\sigma}_2 &= e^{-i\frac{\pi}{8}\sigma_3} \sigma_2 e^{+i\frac{\pi}{8}\sigma_3} = 2^{-1/2}(\sigma_2 - \sigma_1), \\ \tilde{\sigma}_3 &= \sigma_3. \end{aligned} \quad (\text{A8})$$

We start by noting that

$$\mathcal{H}^2(p) = 2 + 3 \cos^2(p) + 2 \sin(p) [\sigma_3 \otimes \tilde{\sigma}_2]. \quad (\text{A9})$$

The tensor-product matrix $[\sigma_3 \otimes \tilde{\sigma}_2]$ commutes with $\mathcal{H}(p)$. Using Eq. (A9) it is straightforward to check that

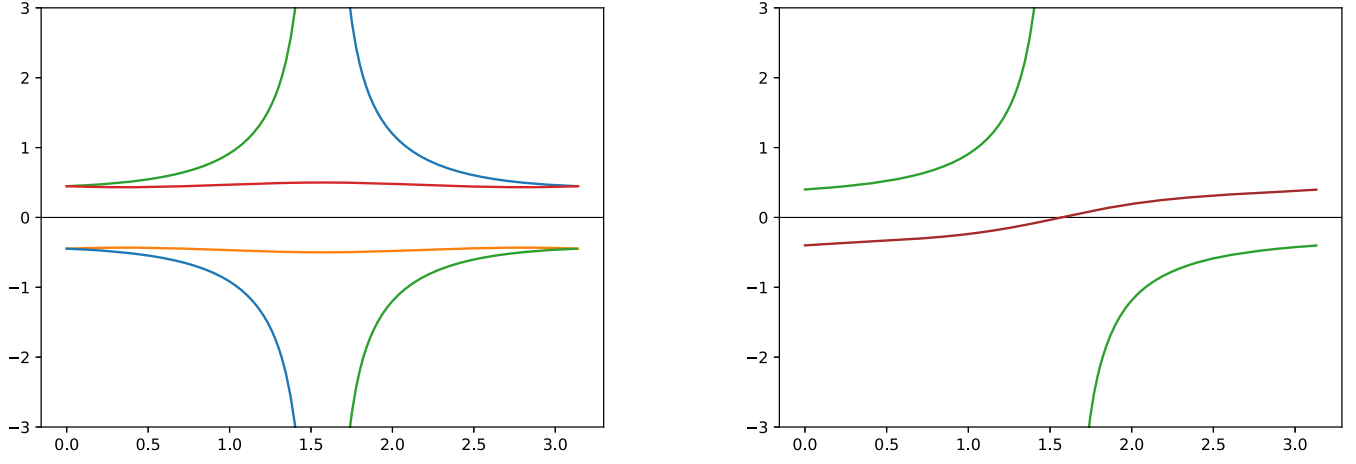


FIG. 3. Eigenvalues of $\mathcal{H}^{-1}(p)$. Left: eigenvalues of the full $\mathcal{H}^{-1}(p)$. Colors match those in Fig. 2. In particular, the poles of the LH (RH) chiral modes are shown in green (blue). Right: eigenvalues of the upper-left 2-by-2 block of $\mathcal{H}^{-1}(p)$. While the LH pole (green) supported near edge A is reproduced, the RH pole got replaced by a propagator zero.

$$\mathcal{H}^{-1}(p) = \mathcal{H}(p)(2 + 3\cos^2(p) - 2\sin(p)[\sigma_3 \otimes \tilde{\sigma}_2]) \times (16\cos^2(p) + 9\cos^4(p))^{-1}. \quad (\text{A10})$$

Again considering small $k = p - \pi/2$ near the primary singularity $p_c = \pi/2$, and introducing projectors $P_{\pm} = \frac{1}{2}(1 \pm [\sigma_3 \otimes \tilde{\sigma}_2])$, we find

$$\mathcal{H}^{-1}(k) = P_-[I \otimes \tilde{\sigma}_2] \frac{1}{2k} + P_+[\sigma_1 \otimes \tilde{\sigma}_1] \frac{1}{2} + \dots, \quad (\text{A11})$$

which is consistent with the eigenvalue spectrum (A6). Notice that the first term on the right-hand side captures the two massless chiral modes of opposite chiralities, Eq. (A6a).

We next turn to another example of a propagator zero. We recall that in the free case $\mathcal{R}(p) = \mathcal{H}^{-1}(p)$ (see Sec. III). $\mathcal{H}^{-1}(p)$ has a LH pole on edge A, and a RH pole on edge B. We thus expect that a propagator zero will appear if we project $\mathcal{H}^{-1}(p)$ onto the fields of one edge only, since in this case the pole originating from the other edge will be missing.

The projectors on the edge-A and edge-B fields are $P_{A,B} = \frac{1}{2}(1 \pm [\sigma_3 \otimes I])$. Let us check what happens if we project the propagator on (say) the edge-A fields. First, if we sandwich $\mathcal{H}(p)$ itself between P_A projectors the result is

$$P_A \mathcal{H}(p) P_A = -2\cos(p) P_A [I \otimes \tilde{\sigma}_2] P_A. \quad (\text{A12})$$

We next use that $P_{A,B}$ commute with $P_{\pm}$, as well as with $[I \otimes \tilde{\sigma}_2]$. Again considering small $k = p - \pi/2$ we find

$$P_A \mathcal{H}^{-1}(k) P_A = P_A [I \otimes \tilde{\sigma}_2] \left(P_- \frac{1}{2k} + P_+ \frac{k}{2} \right) + \dots \quad (\text{A13})$$

In order to understand the physical content of this result, notice that within the matrix block form we are using, the two P_A projectors on the left-hand side of Eq. (A13) pick out the upper-left 2-by-2 block of $\mathcal{H}^{-1}$, which is the diagonal block with $\sigma_3 = +1$. Making the substitution $\sigma_3 = +1$, the 4×4 projectors $P_{\pm}$ reduce to 2×2 projectors $\tilde{P}_{2\pm} = \frac{1}{2}(1 \pm \tilde{\sigma}_2)$. Hence

$$(P_A \mathcal{H}^{-1}(k) P_A)_{2 \times 2} = \tilde{\sigma}_2 \left(\tilde{P}_{2-} \frac{1}{2k} + \tilde{P}_{2+} \frac{k}{2} \right) + \dots = \left(-\tilde{P}_{2-} \frac{1}{2k} + \tilde{P}_{2+} \frac{k}{2} \right) + \dots \quad (\text{A14})$$

While the first term on the right-hand side is the familiar LH pole of edge A, the missing RH pole of edge B got replaced by a propagator zero. (This is a RH zero in the sense that the corresponding eigenvalue of the propagator has the same sign as k .) The eigenvalues of the 2×2 matrix (A14) form a single curve with the following properties: the curve covers the Brillouin zone twice, it is continuous everywhere except at the LH pole, and it has no additional zeros except for the RH zero we have found at $p = \pi/2$.

We show the eigenvalues of the complete $\mathcal{H}^{-1}(p)$ in the left panel of Fig. 3. The poles of the LH and RH chiral modes, which are supported on edge A and edge B respectively, are clearly seen. In the right panel, we show the eigenvalues of the upper-left 2-by-2 block of $\mathcal{H}^{-1}(p)$. This corresponds to a two-point function in which we retain only the edge-A degrees of freedom. The LH pole, which is supported mainly on edge A, is affected only in a minor way. But the RH pole, which is supported mainly on the missing edge-B degrees of freedom, is absent. In its place, there is now a—kinematical—propagator zero.

Finally, to avoid confusion we would like to stress the difference between what we have discussed here, and the

situation described in Appendix A 3 below. Here we are interested in the effect of applying a projection to the inverse, $P_A \mathcal{H}^{-1}(p) P_A$, and, in particular, the propagator zero that this projection generates. By contrast, in Appendix A 3 the question is what is the physical effect of applying a projection to the Hamiltonian itself, $P_A \mathcal{H}(p) P_A$ [already given in Eq. (A12)], as we will be considering a strong-coupling limit that turns the edge-A fermions into a free theory, decoupled from the edge-B fermions.

The appearance of a zero in the projected inverse $P_A \mathcal{H}^{-1}(p) P_A$ is yet another example of a general phenomenon we have discussed in this paper: if our set of interpolating fields is under-complete, and the poles of some components of the fermion asymptotic states are missing from its two-point function, this is bound to lead to zeros in the same two-point function. Specifically, the case we have considered here corresponds to constructing³² the (retarded) two-point function of the edge-A fields only, while leaving out the edge-B fields, even though the fields of both edges are coupled in the bilinear Hamiltonian.

Of course, when we deal with a bilinear Hamiltonian there is no difficulty to identify an appropriate set of interpolating fields. As already pointed out in Sec. III, in this case $H_{\text{eff}}(\vec{p})$ becomes equal to the momentum-space Hamiltonian matrix $\mathcal{H}(\vec{p})$, provided that our set of interpolating fields includes all the fields appearing in the Hamiltonian. Assuming that the original free Hamiltonian is short ranged, $\mathcal{H}(\vec{p})$ is an analytic function of $\vec{p}$, and thus so is $H_{\text{eff}}(\vec{p})$. This is true regardless of the precise relation between the lattice fermion fields and the massless fermion states of the Hamiltonian. In particular, it is possible that a particular lattice field generates more than one massless fermion state, as in the case of naive fermions; or that certain linear combinations of the lattice fields do not generate any massless states near a particular primary singularity, as in the case of the massive modes of the ZZWY Hamiltonian [Eq. (A6b)].

By contrast, in the presence of potentially strong interactions, such as in an SMG phase, the presence of propagator zeros in the two-point function of all the elementary fermion fields may signal that in this phase the fermion asymptotic states contain new degrees of freedom which can be generated by composite fields only.

3. Massless spectrum in the strong-coupling limit

Following the theorem of Sec. IV, when the coupling constants of the interaction Hamiltonian tend to infinity simultaneously, the edge-B fermions decouple from the edge-A fermions. For each of the four species, the edge-A fermions then form a decoupled free theory. Since edge A consists of sublattices 1 and 2, it follows that their (free)

³²In the free limit where all the interactions are turned off, of course.

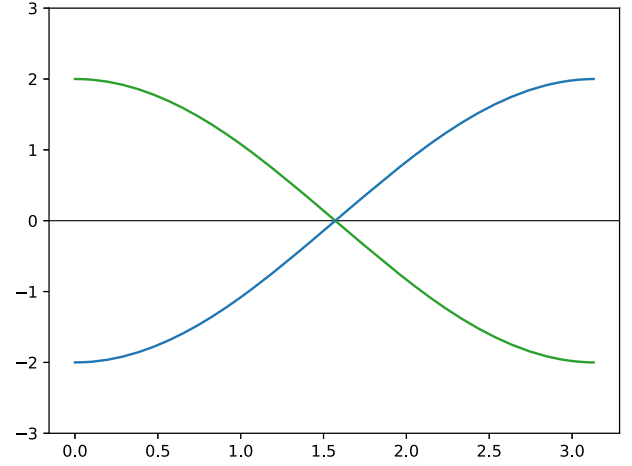


FIG. 4. Spectrum of the “left-over” Hamiltonian—the upper-left 2×2 block of the Hamiltonian matrix (A3).

Hamiltonian is the upper-left 2×2 block of the full Hamiltonian matrix (A3), given in Eq. (A12). The spectrum of this “left-over” Hamiltonian is shown in Fig. 4. As expected, it shows doubling.

A comparison of Fig. 4 with Fig. 2 reveals some interesting facts. First, the critical momentum remains at $p_c = \pi/2$. Also, at a qualitative level, the effect of the strong coupling limit is to eliminate the two massive branches from the spectrum of the bilinear Hamiltonian, while the two massless branches remain qualitatively the same as before.

As for the edge-B degrees of freedom of the four fermion species, they form a single strongly coupled system, for which no strong-coupling expansion is available. As mentioned in Sec. II, this is because of the presence of hopping terms in the interactions (see also Appendix A 4).

4. H_{int} and symmetries

Here we briefly review the algebraic structure underlying the construction of the interaction Hamiltonian H_{int} of the ZZWY model and its symmetries. Both the operator content and the symmetries of H_{int} are encoded in the following set of orthogonal vectors,

$$\tilde{q}_1 = (3, 4; 5, 0), \quad (\text{A15a})$$

$$\tilde{q}_2 = (4, -3; 0, -5), \quad (\text{A15b})$$

$$\tilde{\ell}_1 = (1, -2; 1, 2), \quad (\text{A15c})$$

$$\tilde{\ell}_2 = (2, 1; -2, 1). \quad (\text{A15d})$$

This set plays a key role in the formal arguments for the decoupling of the mirrors in the SMG phase, which we will not repeat here. For the full details, see Refs. [42–44].

Consider first the two $\tilde{q}$ vectors. The first vector, $\tilde{q}_1$, is recognized as the set of charge assignments of the four fermion species under the U(1) symmetry to be gauged. Its first two entries provide the charges of the LH fields of the target theory, and its last two entries provide the charges of the RH fields. The second vector, $\tilde{q}_2$, defines a linearly independent set of charge assignments, which is also anomaly free. By construction, the two U(1) transformations defined by both $\tilde{q}_1$ and $\tilde{q}_2$ will be respected by H_{int} .

The $\tilde{\ell}$ vectors encode the two six-fermion operators introduced in H_{int} , which are consistent with the U(1) symmetries associated with the two $\tilde{q}$ vectors. Recalling that the interaction Hamiltonian involves only the edge-B fermions, $\tilde{\ell}_1$ corresponds to a six-fermion operator with the schematic structure $\chi_1(\chi_2^\dagger)^2\chi_3\chi_4^2$, while $\tilde{\ell}_2$ corresponds to $\chi_1^2\chi_2(\chi_3^\dagger)^2\chi_4$. The orthogonality of the four vectors in Eq. (A15) implies that these six-fermion operators are invariant under the two U(1) symmetries associated with the $\tilde{q}$ vectors. Since the elementary fermions have a single degree of freedom per lattice site, same-field products (for example, χ_1^2) must be point-split. The corresponding hopping terms turn into derivatives in the (classical) continuum limit. We note that Refs. [42–44] did not discuss the four-fermion operators consistent with the same U(1) symmetries (see Sec. V).

APPENDIX B: PROPERTIES OF TWO-POINT FUNCTIONS

In this appendix we establish some properties of the two-point functions used in the proof of the generalized no-go theorem in Sec. III. In Appendix B 1 we establish the Hermiticity of the relevant two-point functions for $\omega \rightarrow 0$. In Appendix B 2 we prove under certain assumptions that $\mathcal{R}(\vec{p})$ is infinitely differentiable except at the degeneracy points. In Appendix B 3 we prove the stronger result that $\mathcal{R}(\vec{p})$ is analytic except at the degeneracy points. We recall that $\mathcal{A}(\vec{p}) = \mathcal{R}(\vec{p})$ away from degeneracy points (see Sec. III). Finally, in Appendix B 4 we explain the role of secondary singularities.

1. Hermiticity

Here we establish the Hermiticity properties of the retarded anticommutator. Specifically, we will consider the Fourier transform of one of the two terms that make it up [see Eqs. (3.1)–(3.4)],

$$S_{ab}(\vec{p}, \omega) = i \int_0^\infty dt e^{i\omega t} \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} \langle 0 | \Psi_a(\vec{x}, t) \Psi_b^\dagger(\vec{0}, 0) | 0 \rangle, \quad (\text{B1})$$

and prove that it is Hermitian for $\text{Im } \omega \rightarrow 0$. The proof for the other term works in the same way, as it does for the

advanced function. As we will see below the field $\Psi_a(\vec{x}, t)$ can be both elementary or composite.

We begin by noting that

$$\Psi_a(\vec{x}, t) = e^{i(Ht - \vec{p}\cdot\vec{x})} \Psi_a(\vec{0}, 0) e^{-i(Ht - \vec{p}\cdot\vec{x})}, \quad (\text{B2})$$

hence

$$\begin{aligned} \langle 0 | \Psi_a(\vec{x}, t) | \vec{p}, n \rangle &= e^{-i(E_n(\vec{p})t - \vec{p}\cdot\vec{x})} \langle 0 | \Psi_a(\vec{0}, 0) | \vec{p}, n \rangle \\ &\equiv e^{-i(E_n(\vec{p})t - \vec{p}\cdot\vec{x})} v_a(\vec{p}, n). \end{aligned} \quad (\text{B3})$$

Here and below, the index n labels all the states with momentum $\vec{p}$. We next introduce a complete set of intermediate states and perform the $\vec{x}$ summation in Eq. (B1), which projects out the intermediate states with momentum $\vec{p}$,

$$\begin{aligned} \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} \langle 0 | \Psi_a(\vec{x}, t) \Psi_b^\dagger(\vec{0}, 0) | 0 \rangle \\ = \sum_n e^{-iE_n(\vec{p})t} v_a(\vec{p}, n) v_b^*(\vec{p}, n). \end{aligned} \quad (\text{B4})$$

The matrix elements product $v_a(\vec{p}, n) v_b^*(\vec{p}, n)$ is manifestly Hermitian. It remains to perform the time integral. Assuming that ω has a positive imaginary part, for a particular intermediate state this integral is

$$i \int_0^\infty dt e^{i(\omega - E_n(\vec{p}))t} = -\frac{1}{\omega - E_n(\vec{p})}. \quad (\text{B5})$$

Finally sending $\text{Im } \omega \rightarrow 0$ this factor becomes real. This establishes the Hermiticity of $S_{ab}(\vec{p}, \omega)$ provided that $\text{Re } \omega$ is not equal to $E_n(\vec{p})$ for any intermediate state, so that $S_{ab}(\vec{p}, \omega)$ is well defined. In particular, it follows that $\mathcal{R}(\vec{p})$ is Hermitian except at the degeneracy points. As explained in Sec. III, the inverse $H_{\text{eff}}(\vec{p})$ is well defined and Hermitian everywhere in the Brillouin zone.

2. Smoothness

We will consider a Hamiltonian H satisfying the following conditions:

- (i) H depends on fermion fields only.
- (ii) It is possible to express H as a sum $H = \sum_{\vec{x}} \mathcal{H}(\vec{x})$, where the Hamiltonian density $\mathcal{H}(\vec{x})$ is a *local composite operator*.

We define a local composite operator $\mathcal{B}(\vec{x})$ by the following requirements. First, $\mathcal{B}(\vec{x})$ is translationally covariant. Second, $\mathcal{B}(\vec{x})$ is the sum of a finite number of terms, where each term is the product of a finite number of elementary (fermion) fields with its own (finite) coupling constant. Finally, the support of $\mathcal{B}(\vec{x})$ is limited to lattice sites $\vec{y}$ whose distance from $\vec{x}$ is bounded. The smallest $0 \leq R_{\mathcal{B}} < \infty$ such that $\|\vec{y} - \vec{x}\| \leq R_{\mathcal{B}}$ for all $\vec{y}$ in the support

of $\mathcal{B}(\vec{x})$ is the *range* of $\mathcal{B}(\vec{x})$. We will denote the range $R_{\mathcal{H}}$ of the Hamiltonian density as R_0 for short.³³

For simplicity we will consider here two-point functions of the elementary fermion fields, which in turn satisfy the canonical anticommutation relations. The generalization to the case that some fermion fields are local composite operators is straightforward, and is left for the reader.

We start with some preliminaries. The canonical anticommutation relations imply that, as an operator acting on the Fock space, the norm of an elementary fermion field is bounded by $\|\psi_a(\vec{x})\| = 1$ (in lattice units). It follows from our assumptions that the norm of the Hamiltonian density $\mathcal{N} = \|\mathcal{H}(\vec{x})\|$ is finite, $0 < \mathcal{N} < \infty$.

We next turn to the time-dependent field,

$$\psi_a(\vec{x}, t) = e^{iHt} \psi_a(\vec{x}) e^{-iHt}. \quad (\text{B6})$$

The norm of the time-dependent anticommutator satisfies the trivial bound

$$\|\{\psi_a(\vec{x}, t), \psi_b^\dagger(\vec{0}, 0)\}\| \leq 2. \quad (\text{B7})$$

Introducing multicommutators of the Hamiltonian H with an elementary fermion field

$$\begin{aligned} \llbracket H, \psi_a \rrbracket_0 &= \psi_a, \\ \llbracket H, \psi_a \rrbracket_1 &= [H, \psi_a], \\ \llbracket H, \psi_a \rrbracket_2 &= [H, \llbracket H, \psi_a \rrbracket_1] = [H, [H, \psi_a]], \\ \llbracket H, \psi_a \rrbracket_n &= [H, \llbracket H, \psi_a \rrbracket_{n-1}], \end{aligned} \quad (\text{B8})$$

the time-dependent field can be expressed in terms of the canonical field at $t = 0$ as

$$\psi_a(\vec{x}, t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \llbracket H, \psi_a(\vec{x}) \rrbracket_n. \quad (\text{B9})$$

This Taylor series controls the locality properties of correlation functions, as we will see.

We will now prove that the retarded function $\hat{R}(\vec{p}, t)$ is infinitely differentiable for all $\vec{p}$ in the Brillouin zone. Provided $\text{Im } \omega > 0$, the same is true for $\tilde{R}(\vec{p}, \omega)$, which is also an analytic function of ω . The advanced functions $\hat{A}(\vec{p}, t)$ and $\tilde{A}(\vec{p}, \omega)$ have similar properties, except the support of $\tilde{A}(\vec{p}, \omega)$ is the lower half-plane $\text{Im } \omega < 0$.

We begin by introducing the function $n(\vec{x})$, defined as the smallest integer such that $n(\vec{x}) \geq \|\vec{x}\|/R_0$. This means that the term with $n = n(\vec{x})$ in the Taylor series (B9) is the first term where the support of $\llbracket H, \psi_a(\vec{x}) \rrbracket_n$ may include the origin. It follows that the first term in Eq. (B9) that

contributes to the anticommutator $\{\psi_a(\vec{x}, t), \psi_b^\dagger(\vec{0}, 0)\}$ must have $n \geq n(\vec{x})$. Estimating the norm of the anticommutator using the norm of the $n = n(\vec{x})$ term we arrive at

$$\|\{\psi_a(\vec{x}, t), \psi_b^\dagger(\vec{0}, 0)\}\| \sim \frac{(2\mathcal{N}t)^{n(\vec{x})}}{n(\vec{x})!} \sim \left(\frac{2eR_0\mathcal{N}t}{\|\vec{x}\|} \right)^{\frac{\|\vec{x}\|}{R_0}}. \quad (\text{B10})$$

The rightmost expression becomes smaller than one for

$$\|\vec{x}\| > 2eR_0\mathcal{N}t. \quad (\text{B11})$$

For fixed t , the estimate (B10) vanishes faster than exponentially for $\vec{x} \rightarrow \infty$.

We now use the bounds (B7) and (B10) to establish the analyticity properties of the Fourier transform $\hat{R}(\vec{p}, t)$. In this subsection we consider the case that $\vec{p}$ is real. In Appendix B 3 we will extend the discussion to complex $\vec{p}$. In the “cone”

$$\|\vec{x}\| \leq 2eR_0\mathcal{N}t, \quad (\text{B12})$$

the norm $\|\{\psi(\vec{x}, t), \psi^\dagger(\vec{0}, 0)\}\|$ can only be bounded by 2, a bound which is always valid. For $\|\vec{x}\| \gg 2eR_0\mathcal{N}t$ we may use Eq. (B10), and the norm of the anticommutator vanishes faster than exponentially. Since $\hat{R}(\vec{p}, t)$ involves a spatial sum, it follows that this sum is dominated by the region defined by inequality (B12). This implies that the norm of $\hat{R}(\vec{p}, t)$ behaves like

$$\sim C(2eR_0\mathcal{N}t)^d, \quad (\text{B13})$$

where again d is the number of spatial dimensions, and C is a geometrical factor.³⁴ Similarly, the k th derivative of $\hat{R}(\vec{p}, t)$ with respect to $\vec{p}$ behaves like $\sim C(2eR_0\mathcal{N}t)^{d+k}$, since each $\vec{p}$ -derivative adds one power of $\vec{x}$. It follows that $\hat{R}(\vec{p}, t)$ and all its $\vec{p}$ derivatives exist. Thus, $\hat{R}(\vec{p}, t)$ is infinitely differentiable for any $\vec{p}$ in the Brillouin zone, for any fixed $t \geq 0$.

We next perform the time Fourier transform. Allowing for k derivatives of $\hat{R}(\vec{p}, t)$ with respect to $\vec{p}$, together with m additional ω derivatives, the integrand in Eq. (3.3) behaves like

$$\sim C(2eR_0\mathcal{N}t)^{d+k} t^m e^{-t \text{Im } \omega}. \quad (\text{B14})$$

Hence, the t -integral converges for any $\text{Im } \omega > 0$.³⁵ It thus follows that $\tilde{R}(\vec{p}, \omega)$ is an analytic function of ω in the

³³The norm $\|\vec{z}\|$ can be for instance the usual L_2 norm $(\sum_{i=1}^d z_i^2)^{1/2}$ or the taxi-driver's distance $\sum_{i=1}^d |z_i|$. We note that the Hamiltonian must contain hopping terms, and thus $R_0 \geq 1$.

³⁴In this appendix we use the notation C for all the geometrical factors we encounter, but it should be noted that they are in general different from each other.

³⁵This is where the use of the retarded (or advanced) two-point function is crucial.

upper half plane, $\text{Im } \omega > 0$, as well as an infinitely differentiable function of $\vec{p}$ everywhere in the Brillouin zone. The advanced function $\tilde{A}(\vec{p}, \omega)$ has similar properties, except that it is analytic in the lower half plane, $\text{Im } \omega < 0$. Finally, by the arguments of Sec. III it follows that the common $\omega \rightarrow 0$ boundary value $\mathcal{R}(\vec{p}) = \mathcal{A}(\vec{p})$ is infinitely differentiable with respect to $\vec{p}$, except at the degeneracy points.

3. Analyticity

The results of Appendix B 2 are sufficient for the no-go theorem, which only requires a continuous first derivative. Nevertheless, for completeness we will use here the edge-of-the-wedge theorem to obtain the stronger result that $\mathcal{R}(\vec{p})$ is an analytic function of $\vec{p}$ except at the degeneracy points.³⁶

In order to apply the edge-of-the-wedge theorem we need to specify the “edge” and the two “wedges.” We start with the edge $\mathcal{E}$, which will be an open subset of the real $(\vec{p}, \omega)$ space, where $\vec{p}$ belongs to the Brillouin zone. For each $\vec{p}$, let $E_{\min}(\vec{p}) \geq 0$ be the infimum (greatest lower bound) of the energy of the states of the second-quantized Hamiltonian with this total momentum. We then define an interval $\Omega(\vec{p})$ as the open subset of the (real) ω axis with $|\omega| < E_{\min}(\vec{p})$. The edge $\mathcal{E}$ is defined as the union of $\Omega(\vec{p})$ for all $\vec{p}$ in the Brillouin zone. Note that the degeneracy points are *not* included in $\mathcal{E}$, because if $\vec{p}$ is a degeneracy point then $E_{\min}(\vec{p}) = 0$ and thus $\Omega(\vec{p}_c)$ is an empty set. For any other $\vec{p}$, the edge $\mathcal{E}$ contains the open interval $\Omega(\vec{p})$ which, in particular, includes the point $\omega = 0$. It follows that for real $\omega \in \Omega(\vec{p})$ the (common) limit

$$\lim_{\epsilon \rightarrow 0} \tilde{R}(\vec{p}, \omega + i\epsilon) = \lim_{\epsilon \rightarrow 0} \tilde{A}(\vec{p}, \omega - i\epsilon) \quad (\text{B15})$$

exists, and is continuous in $\mathcal{E}$.

The two wedges are defined as $W^\pm = \mathcal{E} \pm i\mathcal{U}$, where $\mathcal{U}$ is the intersection of an open cone in $\mathbf{R}^{d+1} = (\text{Im } \vec{p}, \text{Im } \omega)$ with a ball of radius $r > 0$. (An open cone $\mathcal{U}$ is an open set such that if $u \in \mathcal{U}$ then $su \in \mathcal{U}$ for any $s > 0$.) We next discuss the choice of $\mathcal{U}$. In Appendix B 2 we already allowed ω to have an imaginary part, and now we seek a generalization to the case that $\vec{p}$ has an imaginary part as well. First, using Eq. (B10), for $\|\vec{x}\| \gg 2eR_0\mathcal{N}t$ the summand in Eq. (3.2) is bounded by

$$\sim e^{\|\text{Im } \vec{p}\| \|\vec{x}\|} \left(\frac{2eR_0\mathcal{N}t}{\|\vec{x}\|} \right)^{\frac{\|\vec{x}\|}{R_0}} = \left(\frac{2R_0\mathcal{N}te^{R_0\|\text{Im } \vec{p}\|+1}}{\|\vec{x}\|} \right)^{\frac{\|\vec{x}\|}{R_0}}. \quad (\text{B16})$$

This still vanishes faster than exponentially for asymptotically large $\|\vec{x}\|$. Hence $\hat{R}(\vec{p}, t)$ exists and is continuous as

³⁶For the precise statement and a proof of the edge-of-the-wedge theorem, see for example Ref. [59].

before, and the same is true for all of its derivatives with respect to $\vec{p}$, where now $\vec{p}$ can take complex values. It follows that $\hat{R}(\vec{p}, t)$ is an analytic function of $\vec{p}$.

Estimating the norm of $\hat{R}(\vec{p}, t)$ requires more care. We begin by noting that Eq. (B16) becomes smaller than one when $\|\vec{x}\|$ becomes larger than $s_0(\text{Im } \vec{p})t$, where [compare Eq. (B11)]

$$s_0(\text{Im } \vec{p}) = 2R_0\mathcal{N}e^{R_0\|\text{Im } \vec{p}\|+1}. \quad (\text{B17})$$

It follows that we can neglect the region $\|\vec{x}\| > s_0(\text{Im } \vec{p})t$. We need to perform the $\vec{x}$ -summation over the complement region $\|\vec{x}\| \leq s_0(\text{Im } \vec{p})t$ and, as before, in that region we will use the trivial bound (B7) on the anticommutator, which is applicable everywhere. The summand is then bounded by

$$2e^{\|\text{Im } \vec{p}\| \|\vec{x}\|} \leq 2e^{\|\text{Im } \vec{p}\| s_0(\text{Im } \vec{p})t} \leq 2e^{\|\text{Im } \vec{p}\| s_0^{\max} t}, \quad (\text{B18})$$

where $s_0^{\max}$ is the maximum of $s_0(\text{Im } \vec{p})$ for $\|\text{Im } \vec{p}\| \leq r$. Using the right-hand side as a uniform bound for $\|\vec{x}\| \leq s_0(\text{Im } \vec{p})t$ leads to the following (over-)estimate³⁷ of the norm of $\hat{R}(\vec{p}, t)$

$$\sim C(s_0(\text{Im } \vec{p})t)^d e^{\|\text{Im } \vec{p}\| s_0^{\max} t}. \quad (\text{B19})$$

In order for the t -integral to converge, we thus impose

$$\|\text{Im } \vec{p}\| s_0^{\max} < \text{Im } \omega. \quad (\text{B20})$$

This condition defines the open cone, and thus the open set $\mathcal{U}$ as its intersection with the ball of radius r , for all $\text{Im } \omega > 0$. It then follows that $\tilde{R}(\vec{p}, \omega)$ is an analytic function of both $\vec{p}$ and ω in the wedge $\mathcal{E} + i\mathcal{U}$. The same is true for $\tilde{A}(\vec{p}, \omega)$ in the wedge $\mathcal{E} - i\mathcal{U}$.

Having defined an edge and two wedges that satisfy the assumptions of the edge-of-the-wedge theorem, the theorem asserts that there exists an open set $\mathcal{D} \subset \mathbf{C}^{d+1}$ such that (a) $\mathcal{D}$ contains the union of W^+ , W^- and $\mathcal{E}$; (b) there is an analytic function $F(\vec{p}, \omega)$ defined on $\mathcal{D}$ whose restriction to W^+ (W^-) is $\tilde{R}(\vec{p}, \omega)$ [$\tilde{A}(\vec{p}, \omega)$], and whose restriction to the edge $\mathcal{E}$ is given by Eq. (B15). In particular, for $\omega = 0$ we have $F(\vec{p}, 0) = \mathcal{R}(\vec{p})$, provided that $\vec{p}$ is not a degeneracy point.³⁸ This completes the proof that $\mathcal{R}(\vec{p})$ is an analytic function of $\vec{p}$ everywhere in the Brillouin zone except at the degeneracy points.

³⁷The generalization to the case that derivatives with respect to $\vec{p}$ and ω are taken works as before [compare Eqs. (B13) and (B14)].

³⁸On physical grounds we expect that for general real $\vec{p}$, the analytic function $F(\vec{p}, \omega)$ will have two cuts in the complex ω plane along the real ω axis: one starts at $\omega = E_{\min}(\vec{p})$ and goes to $+\infty$, and the other starts at $\omega = -E_{\min}(\vec{p})$ and goes to $-\infty$. At the degeneracy points, the end points of the two cuts meet at $\omega = 0$.

4. Secondary singularities

Here we explain the role of secondary singularities via several examples involving four-fermion interactions. We first consider an example of a self-energy correction near a primary singularity, of the kind already discussed in Sec. III. We then consider an example of a self-energy correction near a secondary singularity, and discuss the similarities and differences between the two cases. As both examples are somewhat abstract, we give a further example how they can be realized in the context of weakly coupled theories.

To keep things as simple as possible we will consider a Brillouin zone extending over the standard interval $[0, 2\pi/a]$ for every momentum component. In $d = 1$, we will assume that primary singularities exist at $p = 0$ and $p = \pi/(2a)$. In $d = 3$ we will assume the same situation for p_x , while the p_y and p_z components of the primary singularities under consideration are always zero. Below, we will focus on the massless states associated with the primary singularity for which p (or p_x) equals $\pi/(2a)$, which we will eventually use in our example of a secondary singularity.

We begin with the self-energy correction for a massless fermion at the primary singularity $p_c = \pi/(2a)$ for $d = 1$, or $\vec{p}_c = (\pi/(2a), 0, 0)$ for $d = 3$. To avoid cumbersome notation we will mostly omit the vector symbol below, but the discussion applies to both $d = 1$ and $d = 3$. We assume that the collection of massless states associated with primary singularities at p_c is described by a set of lattice interpolating fields ψ_i , $i = 1, \dots, N$, with conserved charges $Q_i(\psi_j) = \delta_{ij}$ under the corresponding U(1) symmetries.³⁹ As explained in Sec. III, self-energy corrections can be calculated using an EFT approach, and the leading (momentum dependent) self-energy correction is a two-loop diagram with two four-fermion vertices. We will also make contact with the discussion in Sec. V by presenting examples in which the four-fermion interactions are renormalizable in $d = 1$. For this to be the case, the four-fermion interactions have to exist already at the lattice level without the need of point splitting (which turns into additional derivatives in the continuum limit). In the case at hand we assume $N = 2$, and then take the four-fermion interaction to be the local operator $(\psi_1^\dagger \psi_1)(\psi_2^\dagger \psi_2)$. The resulting self-energy is given in Eq. (3.7a) for $d = 1$ and Eq. (3.7b) for $d = 3$. This self-energy diagram corresponds to a process in which an initial state of a single ψ particle with small $q = p - p_c$ splits into a virtual state consisting of two particles and one antiparticle, which then recombine. To produce the singularities in Eq. (3.7), the lattice momenta of the two particles have to be close to p_c , whereas the lattice momentum of the antiparticle is close to $-p_c$. This

allows the process to occur while preserving the lattice momentum.⁴⁰

We next turn to an example of a self-energy diagram near a secondary singularity. In our example there is a secondary singularity at $p_s = 3\pi/(2a)$, associated with an intermediate state of three ψ fermions. Now the single-particle state on the external legs will in general be a different fermion species, which we will denote as χ . We again avoid the need of point splitting (or equivalently, derivatives) by assuming $N = 3$, and then taking the local four-fermion interaction to be $(\chi^\dagger \psi_1 \psi_2 \psi_3 + \text{H.c.})$. This interaction preserves in particular the U(1) symmetry associated with a common rotation of all the ψ fermions, with charge $Q = Q_1 + Q_2 + Q_3 = 3$, under which $Q(\chi) = 3$. The four-fermion interaction now enables the process of a single χ particle splitting (or decaying) into three ψ particles. Since a single χ intermediate state also contributes in the same channel, by the definition of a secondary singularity it should be gapped near p_s , in other words, its energy for $p = p_s$ should be $|E_0| > 0$. Let us introduce $q_i = p_i - p_c$ for the three ψ virtual particles, as well as $q = q_1 + q_2 + q_3$ and $p = p_1 + p_2 + p_3$, so that $q = p - p_s$. A straightforward diagrammatic calculation of the self-energy correction for the χ particle, coming from the three massless ψ intermediate states, gives rise for small q to the schematic form [compare Eq. (3.7)],

$$E = E_0 \pm c_1 G^2 q (aq)^{2(n-d-1)} \log(q^2) + \dots, \quad d = 1, \quad (\text{B21a})$$

$$H_{2 \times 2} = E_0 \pm c_3 G^2 \vec{\sigma} \cdot \vec{q} (aq)^{2(n-d-1)} \log(q^2) + \dots, \quad d = 3, \quad (\text{B21b})$$

where again n is the mass dimension of the interaction, and the ellipsis indicate subleading and/or analytic corrections. As in Eq. (3.7), G is a dimensionless coupling constant. Our discussion implies that H_{eff} has a finite value at the secondary singularity, but it is not analytic there, in agreement with the general considerations of Sec. III. In the example we have considered here, for $d = 3$ the four-fermion interaction is irrelevant: one has $n - d - 1 = 2$, hence expression (B21b) has four continuous derivatives. In contrast, for $d = 1$ the four-fermion interaction is marginal: $n - d - 1 = 0$. As a result, expression (B21a) is continuous, but does not have a continuous derivative at p_s , as discussed in Sec. V.

In summary, we see that the nonanalytic behavior near secondary singularities is similar to that near the primary ones. The main difference is that at the primary singularities H_{eff} has zero eigenvalues, which are required to be

³⁹Handedness plays little role in the argument, and is therefore suppressed.

⁴⁰Since an antiparticle is associated with a hole in the Dirac sea, in the charge $Q = -1$ sector the primary singularity of $\mathcal{R}(\vec{p})$ is at $-\pi/(2a)$.

relativistic, whereas at the secondary singularities it does not. In fact, at a generic point in the phase diagram any secondary singularity is inherently nonrelativistic, because the gapped single-particle state in the relevant channel has energy E_0 which is $O(1)$ in lattice units. Nevertheless, the behavior of H_{eff} near the secondary singularities must be understood, because the NN theorem requires a continuous first derivative throughout the entire Brillouin zone.

Primary and secondary singularities occur in the two-point functions of both weakly interacting and strongly interacting theories. The latter case includes SMG models, which are difficult to study. We can make the discussion more concrete by demonstrating how the primary and secondary singularities arise in the appropriate two-point functions of weakly coupled lattice theories, which can be studied systematically using perturbation theory. Of course, in our weakly coupled examples, the massless fermions can always be combined into Dirac fermions, and the theory is vectorlike; the goal is mainly to illustrate the nature of primary and secondary singularities.

For the example of the primary singularity, all we need to do is specify a bilinear Hamiltonian for the ψ_i fields, which, specializing to $d = 1$, can be taken to be

$$H_i = \frac{\sqrt{2}}{a} \int_0^{2\pi/a} \frac{dp}{2\pi} \psi_i^\dagger(-p) [\cos(ap - \pi/4) - \cos(\pi/4)] \psi_i(p), \quad (\text{B22})$$

in momentum space. It is easy to check that this Hamiltonian supports one RH and one LH massless states, at $p = 0$ and $p = p_c = \pi/(2a)$ respectively, each leading to a primary singularity in the two-point function $\langle \psi_i \psi_i^\dagger \rangle$. Once the (by assumption, weak) four-fermion interaction $(\psi_1^\dagger \psi_1)(\psi_2^\dagger \psi_2)$ is added, we recover the analytic structure discussed above near both of these primary singularities.

For the example of the secondary singularity, we need a bilinear Hamiltonian for the additional χ field, which we simply take to be

$$H_\chi = \frac{1}{a} \int_0^{2\pi/a} \frac{dp}{2\pi} \chi^\dagger(-p) \sin(ap) \chi(p). \quad (\text{B23})$$

When the interaction $(\chi^\dagger \psi_1 \psi_2 \psi_3 + \text{H.c.})$ is added, the χ field couples to three-particle states of the ψ fields, and some of these states will have vanishing energy near $p_s = 3\pi/(2a)$. The single χ state with momentum $p \sim 3\pi/(2a)$ is gapped with energy $|E_0| = a^{-1} + \mathcal{O}(G^2)$ for $p = p_s$, hence we will indeed recover the secondary singularity discussed above in the two-point function $\langle \chi \chi^\dagger \rangle$.

The presence of the renormalizable four-fermion interaction in $d = 1$ implies that, in certain channels, the resulting H_{eff} will not have a continuous first derivative [for primary singularities, see Eq. (3.7a), for secondary singularities, see Eq. (B21a) above]. As a result, one of the main assumptions of the generalized no-go theorem is not satisfied. As we have explained in Sec. V, under these circumstances one can tentatively still identify RH (LH) massless states with a branch $E(p)$ of H_{eff} that crosses zero from negative to positive (positive to negative) energy. The weakly coupled lattice Hamiltonians presented here are examples where this identification is valid. For more subtle situations where this is not (or not necessarily) the case, see Sec. V.

Notice that we have assumed that the primary singularity is at $\pi/(2a)$, which is equal to $2\pi/a$ times $1/4$, a rational number. It was shown in Refs. [34,35] that this is a completely general feature: every momentum component of a primary singularity must be equal to $2\pi/a$ times a rational number. The same is then also true for all the secondary singularities. This implies that one can define a reduced d -dimensional Brillouin zone in which all the (finitely many) primary and secondary singularities collapse to $\vec{k} = 0$. The momentum $\vec{k}$ in this reduced Brillouin zone can then be identified with the physical momentum, which transforms homogeneously under Lorentz transformations in the continuum limit. For the full discussion, see Refs. [34,35].

-
- [1] Y. Shamir, Lattice chiral fermions, *Nucl. Phys. B, Proc. Suppl.* **47**, 212 (1996).
 - [2] M. Golterman, Lattice chiral gauge theories, *Nucl. Phys. B, Proc. Suppl.* **94**, 189 (2001).
 - [3] E. Poppitz and Y. Shang, Chiral lattice gauge theories via mirror-fermion decoupling: A mission (im)possible?, *Int. J. Mod. Phys. A* **25**, 2761 (2010).
 - [4] L. H. Karsten and J. Smit, Lattice fermions: Species doubling, chiral invariance, and the triangle anomaly, *Nucl. Phys.* **B183**, 103 (1981).
 - [5] H. B. Nielsen and M. Ninomiya, Absence of neutrinos on a lattice. 1. Proof by homotopy theory, *Nucl. Phys.* **B185**, 20 (1981); **B195**, 541(E) (1982); Absence of neutrinos on a lattice. 2. Intuitive topological proof, *Nucl. Phys.* **B193**, 173 (1981).
 - [6] P. H. Ginsparg and K. G. Wilson, A remnant of chiral symmetry on the lattice, *Phys. Rev. D* **25**, 2649 (1982).
 - [7] D. B. Kaplan, A method for simulating chiral fermions on the lattice, *Phys. Lett. B* **288**, 342 (1992).

- [8] R. Narayanan and H. Neuberger, Chiral fermions on the lattice, *Phys. Rev. Lett.* **71**, 3251 (1993).
- [9] R. Narayanan and H. Neuberger, A construction of lattice chiral gauge theories, *Nucl. Phys.* **B443**, 305 (1995).
- [10] H. Neuberger, Exactly massless quarks on the lattice, *Phys. Lett. B* **417**, 141 (1998).
- [11] H. Neuberger, More about exactly massless quarks on the lattice, *Phys. Lett. B* **427**, 353 (1998).
- [12] M. Lüscher, Abelian chiral gauge theories on the lattice with exact gauge invariance, *Nucl. Phys.* **B549**, 295 (1999).
- [13] M. Lüscher, Weyl fermions on the lattice and the nonabelian gauge anomaly, *Nucl. Phys.* **B568**, 162 (2000).
- [14] M. Lüscher, Lattice regularization of chiral gauge theories to all orders of perturbation theory, *J. High Energy Phys.* **06** (2000) 028.
- [15] D. M. Grabowska and D. B. Kaplan, Nonperturbative regulator for chiral gauge theories?, *Phys. Rev. Lett.* **116**, 211602 (2016).
- [16] D. M. Grabowska and D. B. Kaplan, Chiral solution to the Ginsparg-Wilson equation, *Phys. Rev. D* **94**, 114504 (2016).
- [17] D. B. Kaplan, Chiral gauge theory at the boundary between topological phases, *Phys. Rev. Lett.* **132**, 141603 (2024).
- [18] D. B. Kaplan and S. Sen, Weyl fermions on a finite lattice, *Phys. Rev. Lett.* **132**, 141604 (2024).
- [19] M. Golterman and Y. Shamir, Conserved currents in five-dimensional proposals for lattice chiral gauge theories, *Phys. Rev. D* **109**, 114519 (2024).
- [20] D. B. Kaplan and S. Sen, Regulating chiral gauge theory at $\theta = 0$, [arXiv:2412.02024](#).
- [21] Y. Shamir, The standard model from a new phase transition on the lattice, *Phys. Rev. D* **57**, 132 (1998).
- [22] M. F. L. Golterman and Y. Shamir, A gauge fixing action for lattice gauge theories, *Phys. Lett. B* **399**, 148 (1997).
- [23] W. Bock, M. F. L. Golterman, and Y. Shamir, On the phase diagram of a lattice U(1) gauge theory with gauge fixing, *Phys. Rev. D* **58**, 054506 (1998).
- [24] W. Bock, M. F. L. Golterman, and Y. Shamir, Lattice chiral fermions through gauge fixing, *Phys. Rev. Lett.* **80**, 3444 (1998).
- [25] W. Bock, M. Golterman, and Y. Shamir, Gauge fixing approach to lattice chiral gauge theories: Part 2, *Nucl. Phys. B, Proc. Suppl.* **63**, 581 (1998).
- [26] M. Golterman and Y. Shamir, Fermion number violation in regularizations that preserve fermion number symmetry, *Phys. Rev. D* **67**, 014501 (2003).
- [27] M. Golterman and Y. Shamir, SU(N) chiral gauge theories on the lattice, *Phys. Rev. D* **70**, 094506 (2004).
- [28] M. Golterman and Y. Shamir, Running couplings in equivariantly gauge-fixed SU(N) Yang-Mills theories, *Phys. Rev. D* **73**, 014510 (2006).
- [29] M. Golterman and Y. Shamir, Dimensional transmutation in the longitudinal sector of equivariantly gauge-fixed Yang-Mills theory, *Phys. Rev. D* **90**, 034504 (2014).
- [30] J. Wang and Y. Z. You, Symmetric mass generation, *Symmetry* **14**, 1475 (2022).
- [31] L. H. Karsten, Lattice fermions in Euclidean space-time, *Phys. Lett. B* **104**, 315 (1981).
- [32] A. Pelissetto, Lattice nonlocal chiral fermions, *Ann. Phys. (N.Y.)* **182**, 177 (1988).
- [33] D. Friedan, A proof of the Nielsen-Ninomiya theorem, *Commun. Math. Phys.* **85**, 481 (1982).
- [34] Y. Shamir, Constraints on the existence of chiral fermions in interacting lattice theories, *Phys. Rev. Lett.* **71**, 2691 (1993).
- [35] Y. Shamir, On the Absence of chiral fermions in interacting lattice theories, [arXiv:hep-lat/9307002](#).
- [36] C. Rebbi, Chiral invariant regularization of fermions on the lattice, *Phys. Lett. B* **186**, 200 (1987).
- [37] M. Campostrini, G. Curci, and A. Pelissetto, Can a nonlocal lattice fermion formulation avoid the doubling problem?, *Phys. Lett. B* **193**, 279 (1987).
- [38] M. Golterman and Y. Shamir, Propagator zeros and lattice chiral gauge theories, *Phys. Rev. Lett.* **132**, 081903 (2024).
- [39] M. Zeng, F. Xu, D. C. Lu, and Y. Z. You, Optical conductivity in symmetric mass generation insulators, [arXiv:2405.05339](#).
- [40] C. Chen, J. Giedt, and E. Poppitz, On the decoupling of mirror fermions, *J. High Energy Phys.* **04** (2013) 131.
- [41] Y. Kikukawa, Why is the mission impossible? Decoupling the mirror Ginsparg-Wilson fermions in the lattice models for two-dimensional Abelian chiral gauge theories, *Prog. Theor. Exp. Phys.* **2019**, 073B02 (2019).
- [42] J. Wang and X. G. Wen, Nonperturbative regularization of (1 + 1)-dimensional anomaly-free chiral fermions and bosons: On the equivalence of anomaly matching conditions and boundary gapping rules, *Phys. Rev. B* **107**, 014311 (2023).
- [43] M. DeMarco and X. G. Wen, A novel non-perturbative lattice regularization of an anomaly-free 1 + 1d Chiral SU(2) gauge theory, [arXiv:1706.04648](#).
- [44] M. Zeng, Z. Zhu, J. Wang, and Y. Z. You, Symmetric mass generation in the 1 + 1 dimensional chiral fermion 3-4-5-0 model, *Phys. Rev. Lett.* **128**, 185301 (2022).
- [45] E. Eichten and J. Preskill, Chiral gauge theories on the lattice, *Nucl. Phys.* **B268**, 179 (1986).
- [46] M. F. L. Golterman, D. N. Petcher, and E. Rivas, Absence of chiral fermions in the Eichten-Preskill model, *Nucl. Phys.* **B395**, 596 (1993).
- [47] Y. Z. You, Z. Wang, J. Oon, and C. Xu, Topological number and fermion Green's function for strongly interacting topological superconductors, *Phys. Rev. B* **90**, 060502 (2014).
- [48] K. Slagle, Y. Z. You, and C. Xu, Exotic quantum phase transitions of strongly interacting topological insulators, *Phys. Rev. B* **91**, 115121 (2015).
- [49] N. D. Mermin and H. Wagner, Absence of ferromagnetism or antiferromagnetism in one-dimensional or two-dimensional isotropic heisenberg models, *Phys. Rev. Lett.* **17**, 1133 (1966).
- [50] S. R. Coleman, There are no Goldstone bosons in two-dimensions, *Commun. Math. Phys.* **31**, 259 (1973).
- [51] M. F. L. Golterman and D. N. Petcher, The decoupling of right-handed neutrinos in chiral lattice gauge theories, *Phys. Lett. B* **225**, 159 (1989).
- [52] M. F. L. Golterman and Y. Shamir, Domain wall fermions in a waveguide: The phase diagram at large Yukawa coupling, *Phys. Rev. D* **51**, 3026 (1995).
- [53] W. Bock, A. K. De, E. Focht, and J. Smit, Fermion Higgs model with strong Wilson-Yukawa coupling in two-dimensions, *Nucl. Phys.* **B401**, 481 (1993).

- [54] E. Witten, Chiral symmetry, the $1/n$ expansion, and the $SU(N)$ thirring model, *Nucl. Phys.* **B145**, 110 (1978).
- [55] W. Bock and A. K. De, Unquenched investigation of fermion masses in a chiral fermion theory on the lattice, *Phys. Lett. B* **245**, 207 (1990).
- [56] W. Bock, A. K. De, K. Jansen, J. Jersak, T. Neuhaus, and J. Smit, Phase diagram of a lattice $SU(2) \otimes SU(2)$ scalar fermion model with naive and Wilson fermions, *Nucl. Phys.* **B344**, 207 (1990).
- [57] M. F. L. Golterman, D. N. Petcher, and J. Smit, Fermion interactions in models with strong Wilson-Yukawa couplings, *Nucl. Phys.* **B370**, 51 (1992).
- [58] M. F. L. Golterman, D. N. Petcher, and E. Rivas, The large Wilson-Yukawa coupling expansion, *Nucl. Phys.* **B377**, 405 (1992).
- [59] S. G. Krantz, *Function Theory of Several Complex Variables* (John Wiley, New York, 1982), p. 133 theorem 3.5.3.