PHYSICAL REVIEW D 67, 014501 (2003)

Fermion-number violation in regularizations that preserve fermion-number symmetry

Maarten Golterman*
Department of Physics and Astronomy, San Francisco State University, San Francisco, California 94132

Yigal Shamir†
School of Physics and Astronomy, Tel-Aviv University, Ramat Aviv, 69978, Israel

(Received 28 May 2002; published 16 January 2003)

There exist both continuum and lattice regularizations of gauge theories with fermions which preserve chiral U(1) invariance (''fermion number''). Such regularizations necessarily break gauge invariance but, in a covariant gauge, one recovers gauge invariance to all orders in perturbation theory by including suitable counterterms. At the nonperturbative level, an apparent conflict then arises between the chiral U(1) symmetry of the regularized theory and the existence of 't Hooft vertices in the renormalized theory. The only possible resolution of the paradox is that the chiral U(1) symmetry is broken spontaneously in the enlarged Hilbert space of the covariantly gauge-fixed theory. The corresponding Goldstone pole is unphysical. The theory must therefore be defined by introducing a small fermion-mass term that breaks explicitly the chiral U(1) invariance and is sent to zero after the infinite-volume limit has been taken. Using this careful definition (and a lattice regularization) for the calculation of correlation functions in the one-instanton sector, we show that the 't Hooft vertices are recovered as expected.

DOI: 10.1103/PhysRevD.67.014501 PACS number(s): 11.15.Ha

I. INTRODUCTION AND CONCLUSION

Every gauge theory coupled to massless fermions has an anomalous chiral current. Representing all fermions by left-handed Weyl fields, the Noether current associated with a common global U(1) rotation is classically conserved. At the one-loop level, a gauge-invariant definition of the current yields the Adler-Bell-Jackiw anomaly [1]

$$\partial_\mu J_\mu^L = \frac{cg^2}{8\pi^2}\mathrm{tr}F\tilde{F}. \tag{1.1}$$

For our notation see Appendix A. The group-theoretical constant c is additive. (Each Weyl fermion in the fundamental representation contributes $c=\frac{1}{2}$. We will assume the gauge symmetry to be non-anomalous throughout this paper.) One can also define a conserved but gauge non-invariant current

$$\hat{J}_\mu^L = J_\mu^L - g^2 K_\mu, \tag{1.2}$$

where

$$K_\mu = \frac{c}{8\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{tr}\left(A_\nu F_{\rho\sigma} - \frac{1}{3}A_\nu A_\rho A_\sigma\right). \tag{1.3}$$

If a gauge-invariant regularization is used, the gauge-invariant, non-conserved current J_μ^L is defined (up to a Z factor) by the fermion bilinear

$$\sum_i \overline{\psi}_L^i \sigma_\mu \psi_L^i. \tag{1.4}$$

Here i runs over all the left-handed fields. In QCD-like theories, this applies in particular to dimensional regularization as well as to the standard lattice regularization [2,3].

What happens if the regulator is chiral-U(1) invariant? The U(1) current will now be conserved at the one-loop level. Therefore it must, when the cutoff is removed, coincide with the gauge non-invariant current $\hat{J}_\mu^L$ defined in Eq. (1.2). (This is true up to a term $\partial_\nu H_{\mu\nu}$ with $H_{\mu\nu}$ an antisymmetric tensor.) Since, classically, the U(1) Noether current is gauge invariant, this can only happen because the regularization itself is not gauge invariant: a chiral-U(1)-invariant regularization is, necessarily, not gauge invariant.

Does this observation imply that all chiral-U(1)-invariant regularizations must be dismissed? To begin with, in perturbation theory the answer is no, provided the action contains covariant gauge-fixing (and ghost) terms. A covariant gauge will be assumed throughout this paper. In the presence of a longitudinal kinetic term, $(\partial_\mu A_\mu)^2$, the theory is renormalizable by power counting *without* relying on gauge invariance. The renormalization program reduces to an algebraic problem and (provided the gauge symmetry is non-anomalous) one can restore gauge invariance to all orders in perturbation theory by suitable counterterms (see e.g. Ref. [4]).

Beyond perturbation theory, there is an apparent conflict between chiral U(1) invariance of the regularized theory and the fact that instanton-mediated amplitudes violate the conservation of the chiral U(1) charge [5]. It has been pointed out long ago [6–8] that, in a covariant gauge, the breaking of chiral U(1) invariance can be *spontaneous* in a technical sense. The reason is that the enlarged Hilbert space of the gauge-fixed theory can accommodate a new Goldstone pole. The latter is unphysical since it originates from the K_μ part of the current $\hat{J}_\mu^L$. If the regulator is chiral-U(1) invariant, there is, in fact, no other possibility. In this paper, we reexamine this question in the context of a specific lattice-

*Email address: maarten@quark.sfsu.edu
†Email address: shamir@post.tau.ac.il

regularization method. Our analysis reveals that a careful definition of the thermodynamical limit is necessary, just as in the case of conventional spontaneous symmetry breaking. This generalizes to continuum regularizations with chiral U(1) invariance such as, for example, momentum-cutoff schemes (see e.g. Ref. [9]) or the dimensional-reduction scheme of Ref. [10].

The main motivation for our lattice-regularization method is that it may ultimately provide a non-perturbative definition of (anomaly-free) chiral gauge theories. These theories are notoriously difficult to regularize in a gauge-invariant way. In particular, dimensional regularization is not a gauge-invariant regulator in this case. A gauge-invariant *perturbative* regularization for chiral gauge theories has been found only recently in the context of lattice gauge theory [11,12]. Beyond perturbation theory, it is not known if non-Abelian chiral gauge theories can be regularized in a gauge-invariant manner. (For a review of recent work in this direction see Ref. [13].)

According to the gauge-fixing approach, chiral gauge theories are defined as the continuum limit of a lattice theory whose action contains a covariant gauge-fixing term and counterterms [14–17] (see also Ref. [18] for a pedagogical presentation and Ref. [19] for a recent review of lattice chiral gauge theories). Carefully chosen irrelevant terms in the lattice action are essential for the existence and the continuity of the phase transition where the continuum limit is taken. In the Abelian case, we showed that the lattice fermions are indeed chirally coupled to the gauge field and that perturbation theory provides a valid description of the critical point [16].

A fully non-perturbative generalization of the gauge-fixing approach to non-Abelian theories is not a simple task, since one has to confront the issue of Gribov copies. This important problem will not be addressed here (see Ref. [17] for recent progress). Still, the method generates a systematic expansion around the classical vacuum and, provided the fermion spectrum is gauge-anomaly free, it provides a consistent regularization in perturbation theory. By invoking the familiar machinery of collective coordinates [5], it can be used to generate a systematic expansion around other classical solutions. In particular, it is possible to carry out an analytic calculation in an instanton background. This allows us to address the question, first raised in Ref. [20], of how fermion-number violating processes are realized in the gauge-fixing approach.

Let us explain the issue in more detail. The simplest chiral fermion action used in the gauge-fixing approach, the so-called chiral Wilson action, utilizes a set of right-handed *spectator* fields χ_R^i, one per each left-handed field ψ_L^i (see Appendix C for the precise definition). The role of the spectators is to avoid fermion doubling on the lattice [2,3,21]. Thanks to a fermion-shift symmetry they can be proved to decouple in the continuum limit [22].

Each term in the lattice action has one $\overline{\psi}_L$ or $\overline{\chi}_R$ and one ψ_L or χ_R field. The chiral Wilson action is therefore invariant under a common U(1) rotation of *all* fermion fields. The Noether current associated with this symmetry,

$$J_\mu^{\mathrm{L,latt}} \sim \sum_i \left(\overline{\psi}_L^i \sigma_\mu \psi_L^i + \overline{\chi}_R^i \overline{\sigma}_\mu \chi_R^i + \cdots \right), \tag{1.5}$$

is exactly conserved on the lattice (the ellipsis stands for lattice terms with no continuum counterpart; see Appendix C). But since the lattice action is not gauge [nor Becchi-Rouet-Stora-Tyutin (BRST)] invariant in the gauge-fixing approach, the conservation of $J_\mu^{\mathrm{L,latt}}$ is consistent with our earlier general comments. Following Ref. [23], we have verified through an explicit one-loop lattice calculation [24] that the current $J_\mu^{\mathrm{L,latt}}$ indeed reduces in the continuum limit to the current $\hat{J}_\mu^L$ defined in Eq. (1.2).

The perturbative results of Refs. [23,24] are, however, not enough to resolve the following puzzle, which we will refer to as the Banks paradox [20]. In short, the paradox stems from the fact that exact U(1) invariance implies exact conservation of the corresponding U(1) *charge*, namely, of fermion number. To be precise, consider a finite-volume lattice and assume that the boundary conditions respect the U(1) symmetry. The Ward identity that corresponds to a global U(1) rotation reads

$$\langle \delta\mathcal{O} \rangle = q \langle \mathcal{O} \rangle = 0. \tag{1.6}$$

Here q is the fermion number, or the U(1) charge, of $\mathcal{O}$. We stress that, on a finite lattice, this is a rigorous result. (The invariance of the lattice measure follows trivially from the fact that there is an equal number of $d\psi_L$ and $d\overline{\psi}_L$ Grassmann integrals, as well as an equal number of $d\chi_R$ and $d\overline{\chi}_R$ ones.)

The Ward identity (1.6) states that a fermion correlation function $\langle \mathcal{O}(x_1,x_2,\dots) \rangle$ can be non-zero only if $q=0$, namely if the number of ψ_L and χ_R fields is equal to the number of $\overline{\psi}_L$ and $\overline{\chi}_R$ fields. This is true for any finite lattice spacing a and any finite volume and, therefore, also after the continuum limit $a \to 0$ and the infinite-volume limit have been taken. Moreover, in the continuum limit the numbers of χ_R and $\overline{\chi}_R$ fields must by themselves be equal, since the spectator field decouples [22]. Hence the numbers of ψ_L and $\overline{\psi}_L$ fields must be equal too. We have thus reached the paradoxical conclusion that, even though the lattice-fermion spectrum is chiral [16], all fermion-number violating amplitudes vanish. In other words, 't Hooft vertices [5] do not seem to occur. If the gauge-fixing approach would be utilized to define a vector-like theory such as massless QCD, the same reasoning would seem to lead to the erroneous conclusion that the U(1) axial charge is conserved in all physical processes.

One can avoid the Banks paradox by adding to the lattice action a mass term, $m_{ij}\psi_L^i\psi_L^j + \text{H.c.}$, for the physical fermions. This allows unequal numbers of ψ_L and $\overline{\psi}_L$ fields to be compensated by insertions of the mass term. The broken, global U(1) Ward identity in a finite (lattice) volume is now

$$q\langle \mathcal{O} \rangle = 2 m_{ij} \left\langle \left(\sum_x \psi_L^i(x)\psi_L^j(x) - \text{H.c.} \right) \mathcal{O} \right\rangle. \tag{1.7}$$

The original lattice-fermion action corresponds to the limit $m \to 0$, where m denotes generically the magnitude of the mass terms. We see that fermion-number or axial-charge violating amplitudes can be non-zero provided they behave like m/m in the limit $m \to 0$. Now, we would not expect an "m/m" behavior in a finite volume. [Being valid for $m=0$, Eq. (1.6) in fact implies that this is impossible in the presence of a lattice cutoff.] The conclusion is that, in order to reproduce correctly 't Hooft vertices, the infinite-volume limit and the $m \to 0$ limit *must not commute*. Hence, the U(1) lattice symmetry must be broken *spontaneously*.

Anticipating spontaneous symmetry breaking (SSB) of the U(1) lattice symmetry we define the thermodynamical limit as the infinite-volume limit, followed by the limit $m \to 0$. In this limit the momentum-space U(1) Ward identity reads, for any $p_\mu \neq 0$,

$$i p_\mu \langle \tilde{J}_\mu^{\mathrm{L,latt}}(p) \mathcal{O} \rangle = q \langle \mathcal{O} \rangle. \tag{1.8}$$

Here $\tilde{J}_\mu^{\mathrm{L,latt}}(p)$ is the Fourier transform of $J_\mu^{\mathrm{L,latt}}(x)$. Unlike Eq. (1.7), because here $p_\mu \neq 0$, the explicit m-dependent term now vanishes for $m \to 0$. This is explained in more detail in the Discussion section. The identity (1.8) holds in the regularized theory as well as in the continuum limit. [On the lattice, the p_μ factor on the left-hand side is modified by $O(ap^2)$ terms. Again, remember that there is no room for an anomalous term since the current $J_\mu^{\mathrm{L,latt}}$ is exactly conserved on the lattice.]

Since 't Hooft vertices do exist, this means that there are operators for which the right-hand side of Eq. (1.8) is non-zero. Hence the left-hand side must contain a *Goldstone pole*. As explained earlier, in the continuum limit the Goldstone pole comes from the K_μ part of the conserved U(1) current and is unphysical [6–8].

In this paper we calculate instanton-sector fermion correlation functions on the lattice, in the semi-classical approximation. We start from a lattice-fermion action with an additional, small, mass term that breaks explicitly the unphysical U(1) symmetry. Taking the infinite-volume and continuum limits, followed by the $m \to 0$ limit, we show that the anticipated 't Hooft vertices are recovered. The "m/m" nature of fermion-number and axial-charge violating amplitudes is manifest in our calculation.

The paper is organized as follows. In order to minimize technicalities we begin in Sec. II with one-flavor QCD where, instead of the usual gauge-invariant lattice definition, we define the theory via the gauge-fixing approach in the special case that the (left-handed) fermion spectrum happens to contain one field in the fundamental representation and one field in the complex conjugate one. In Sec. III we work out the anomaly-free SO(10) theory as the prototype of a truly chiral gauge theory. Our conclusions are summarized and discussed in Sec. IV. In particular, we show that the phase of the 't Hooft vertex follows the phase of the applied mass term, thus demonstrating explicitly the existence of the continuously degenerate ground states associated with SSB. The notation is listed in Appendix A, elements of SO(10) group theory are discussed in Appendix B, and lattice definitions are collected in Appendix C. The construction of propagators in the presence of approximate zero modes is discussed in Appendix D.

II. ONE-FLAVOR QCD USING THE GAUGE-FIXING APPROACH

We begin with the simple example of one-flavor massless QCD, an SU(N) gauge theory coupled to one Dirac fermion in the fundamental representation. The anomalous current of Eq. (1.1) is in this case the axial current. Let us recall what the 't Hooft interaction of this theory is. In a fixed instanton background, the massless continuum Dirac operator $\not{D}$ has one left-handed zero mode $u(x) = P_L u(x)$. Therefore the Weyl fields ψ_L and $\overline{\psi_R}$ each have a zero mode. (In an anti-instanton background, the Weyl fields with zero modes are ψ_R and $\overline{\psi_L}$.) The basic axial-symmetry violating correlation function is

$$\langle \psi_L(x) \overline{\psi_R}(y) \rangle = u(x) u^\dagger(y) \mathrm{Det}', \tag{2.1}$$

where the expectation value denotes Grassmann integration only, and Det′ is the (renormalized) fermion determinant with the zero mode removed. Our objective will be to recover this result starting from a lattice action with exact axial U(1) invariance.

The remaining step in a complete semi-classical calculation is the integration over the gauge and ghost fields. This raises no new conceptual issues and therefore we will skip the details. We recall that the integration (or lattice sum) over the instanton position recovers momentum conservation. The Gaussian integration over the non-zero gauge, fermion and ghost fluctuations leads to the replacement of the lattice's bare instanton action, $8\pi^2/g_0^2$, by the renormalized one, $8\pi^2/g_r^2(\rho)$, where ρ is the instanton's size.

One-flavor QCD has a gauge-invariant lattice definition. When using ordinary Wilson fermions, the lattice fermion action is not invariant under axial transformations, and the paradox described in the Introduction does not arise. (In the continuum limit one reproduces the axial anomaly [3], while non-singlet axial symmetries are recovered [25].) In principle, it should be possible to define one-flavor QCD using the gauge-fixing approach, too. While this has many disadvantages compared to the gauge-invariant definition, it has the interesting property that the paradox described in the Introduction *occurs*. By working out one-flavor QCD we are able to address, with minimal technicalities, the main issue of this paper—how a global symmetry of the lattice path integral can be broken in the continuum limit.

The lattice construction of the chiral Wilson action begins with enumerating the left-handed fields of the target theory. For one-flavor QCD we have two Weyl fields ψ_L and ψ_L^c in the fundamental and anti-fundamental representations respectively. As explained in the Introduction, one also needs two right-handed spectator fields that decouple in the continuum limit. These may be denoted χ_R and χ_R^c. In the case at hand, we may take advantage of the Dirac nature of the target theory and trade the left-handed anti-fundamental field with a right-handed fundamental one $\psi_L^c \to \overline{\psi}_R$, $\overline{\psi}_L^c \to \psi_R$. With a similar trade-off for the corresponding spectator field, the lattice fermion action density can be written in the following matrix form:

$$(\overline{\chi}_R \;\; \overline{\psi}_L \;\; \overline{\psi}_R \;\; \overline{\chi}_L) \begin{pmatrix} 0 & 0 & -\frac{a}{2}\hat{\Box} & \overline{\sigma}_\mu \hat{\partial}_\mu \\ 0 & m & \sigma_\mu \hat{D}_\mu & -\frac{a}{2}\hat{\Box} \\ -\frac{a}{2}\hat{\Box} & \overline{\sigma}_\mu \hat{D}_\mu & m & 0 \\ \sigma_\mu \hat{\partial}_\mu & -\frac{a}{2}\hat{\Box} & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \chi_L \\ \psi_R \\ \psi_L \\ \chi_R \end{pmatrix}. \tag{2.2}$$

We use carets to denote lattice derivatives. (For the precise definitions of lattice derivatives and currents see Appendix C.) Observe that the middle two-by-two block in the above matrix operator resembles the massive continuum Dirac operator.

For orientation, we recall that in the conventional definition of Wilson fermions there are of course no spectator fields, and covariant Wilson terms are placed in the same block entries as the mass terms in the above expression. This removes the doublers in a gauge invariant way, while axial symmetry is lost.

In order to later accommodate truly chiral gauge theories, the doublers are removed here by introducing spectator fields and coupling them to the original fermions via the (free) lattice Laplacian $\hat{\Box}$. Since the Wilson terms now couple fields with different gauge-transformation properties, they lead to a breakdown of gauge invariance [26]. In lattice perturbation theory gauge invariance is regained by adding suitable counterterms, and the renormalized diagrams describe one interacting Dirac field, the quark, and one free Dirac field, the spectator. As is usually the case for symmetries broken by the lattice regularization, the above is true provided the external momenta are vanishingly small in lattice units. The choice of a free lattice Laplacian in Eq. (2.2) implies the shift symmetry of the spectator field [22], which reduces considerably the number of counterterms. In particular, there are no counterterms of the form $\overline{\psi}_R \chi_L$, etc., and the spectator field decouples in the continuum limit.

Let us now examine the U(1) symmetries of the action (2.2). Dropping both the Wilson and the mass terms, the action would be invariant under four separate U(1)'s—a fermion-number symmetry for each Weyl field. With the Wilson terms in place, the action is still invariant under two U(1)'s. Finally, for $m \neq 0$, only the invariance under a common U(1) rotation is left. This invariance corresponds to the baryon-number symmetry of QCD.

The additional U(1) symmetry at $m=0$ transforms χ_L and ψ_R with (say) charge $+1$ and χ_R and ψ_L with charge -1. This is the chiral symmetry that leads to the Banks paradox [20]. The existence of this lattice symmetry would seem to lead to the (erroneous) conclusion that the axial charge is conserved in massless (one-flavor) QCD. As mentioned in the Introduction, Refs. [23,24] already showed that the anomaly appears in the triangle diagram as expected. But this still does not explain how the axial charge is *not* conserved in physical processes.

We will now answer this question through an explicit calculation. We calculate the lattice-fermion two point function in the semi-classical approximation for (small) $m>0$. Taking the infinite-volume and continuum limits and finally sending $m \to 0$, we find that the 't Hooft interaction (2.1) is recovered.

As discussed in the literature [6–8], the continuous degeneracy of ground states associated with SSB of the axial U(1) is parametrized by the vacuum θ angle. The chiral Wilson action (C1) is invariant under a CP transformation, and this remains true in the presence of the mass term introduced in Eq. (2.2) above. Therefore the calculation in this section (as well as in Sec. III) corresponds to a vacuum angle $\theta = 0$. The case of a general θ angle is explained in the Discussion section.

We start with a continuum *regular-gauge* instanton field $A_\mu(x)$ whose size ρ is very large in lattice units, $\rho \gg a$. (Singular-gauge instantons are suppressed in the gauge-fixing approach by the irrelevant terms in the lattice action; see Appendix C.) The lattice gauge field may be defined as $U_\mu(x) = \exp[iaA_\mu(x)]$. This is a smooth configuration. [By this we mean that $U_\mu(x) - I = O(a/\rho)$ and $U_\mu(x) - U_\mu(x+\hat{\nu}) = O((a/\rho)^2)$.] For this lattice gauge field and fixed $m>0$, we denote the matrix operator in Eq. (2.2) by $\gamma_5 \mathcal{M}$. Note that according to this definition $\mathcal{M}$ is Hermitian.

In the formal continuum limit, the lattice operator in Eq. (2.2) goes over to a continuum Dirac operator $D(m) = \gamma_5 H(m)$ which depends on the original instanton field $A_\mu(x)$. One obtains $D(m)$ by dropping the (irrelevant) Wilson terms and replacing the lattice difference operators $\hat{\partial}_\mu$ and $\hat{D}_\mu$ by the corresponding continuum derivatives. For $m>0$, $D(m)$ describes a massive quark (made of $\psi_{R,L}$), whose Dirac operator is $\not{D}+m$, and a decoupled, free massless spectator field (made of $\chi_{R,L}$). Thus, $D(m)$ has no zero modes. We will denote by $G(m)$ the propagator of the Hermitian operator $H(m)$. Both $\mathcal{M}^{-1}$ and $G(m)$ admit a standard spectral decomposition.

Let us now be more precise about how the physical matrix element is obtained. One has to multiply the correlation function in Eq. (2.1) by (normalized) wave functions $f_1^\dagger(x)$ and $f_2(y)$ and integrate (or sum) over x and y. The physical observable is the gauge-field functional average of $\mathrm{Det}(\mathcal{M})\langle f_1|\mathcal{M}^{-1}\gamma_5|f_2\rangle$. We will denote the generic virtuality of the external legs by Q^2. As explained below, the matrix element is dominated by instantons of size $\rho^2 \sim Q^{-2} \gg a^2$, where the last inequality follows because in the continuum limit $a^2 Q^2 \to 0$.

Since the fermion determinant will be $O(m)$, we are interested only in the $O(1/m)$ piece of the propagator(s). We claim that

$$\lim_{a\to 0}\langle f_1|\mathcal{M}^{-1}\gamma_5|f_2\rangle^{\text{singular}}_{\text{lattice}}=\langle f_1|G(m)\gamma_5|f_2\rangle^{\text{singular}}_{\text{continuum}}. \tag{2.3}$$

Here "singular" denotes the $O(1/m)$ piece. The reason why Eq. (2.3) is true is that a non-zero m affects only eigenfunctions with eigenvalues λ in the region $|\lambda|\lesssim m$. Indeed, for all (lattice or continuum) eigenfunctions with, say, $\lambda^2\gg Q^2$, the effect of $m>0$ will be bounded by m^2/Q^2 to some positive power. Therefore they do not contribute to the $O(1/m)$ term. For $\lambda^2\ll Q^2$, the difference between each continuum eigenfunction and the corresponding lattice eigenfunction is bounded by a^2Q^2 to some positive power. Since $m>0$, the inverse eigenvalues λ^{-1} are bounded from above, and the contribution of the entire low-energy lattice spectrum approaches smoothly the continuum one for $a\to 0$. [For the pairing of the lattice and the continuum eigenfunctions we may momentarily assume a very large, but finite, volume, thus making the spectrum discrete; alternatively Eq. (2.3) can also be justified directly in the infinite-volume limit.]

In the infinite-volume limit, the massive continuum propagator satisfies

$$G(x,y;m)\gamma_5=\frac{1}{m}u(x)u^\dagger(y)+O(1). \tag{2.4}$$

Here we show explicitly only the term that diverges for $m\to 0$ [compare Eq. (D15)]. With the understanding that the matrix element is to be taken between smooth wave functions $f_{1,2}$ as described above, we thus have

$$\lim_{a\to 0}\mathcal{M}^{-1}(x,y)\gamma_5=\frac{1}{m}u(x)u^\dagger(y)+O(1). \tag{2.5}$$

For the fermionic determinant similar arguments lead, after renormalization, to

$$\lim_{a\to 0}\mathrm{Det}(\mathcal{M})=m[\mathrm{Det}'+O(m)]. \tag{2.6}$$

The explicit factor of m again comes from the (approximate) zero mode, while the $O(m)$ terms account for the change in the continuous spectrum due to m. Putting this together we thus obtain

$$\langle\psi_L(x)\overline{\psi}_R(y)\rangle=m[\mathrm{Det}'+O(m)]\left(\frac{1}{m}u(x)u^\dagger(y)+O(1)\right). \tag{2.7}$$

Finally, taking the limit $m\to 0$ we recover Eq. (2.1). Equation (2.7) reveals the "m/m" nature of the 't Hooft interaction.

The familiar equation (2.4) above is particularly simple and has been invoked primarily for pedagogical reasons. In Appendix D we show how to handle perturbations that lift the zero modes, but are not proportional to the identity matrix and/or are not spatially constant. This more general formalism will be necessary in the next section. For a few more details on the calculation of the determinant see *Comment 2* in Appendix D.

Our instanton calculation was done in the semi-classical approximation, as is routine in the continuum. Since we have somewhat expanded its scope by using a specific lattice-regularization method, we will briefly review the justification for the semi-classical approximation.

Consider a fermion-number violating amplitude with only a minimal number of fermions, and no other particles, as the external legs. (Here we wish to avoid the controversy about whether the fermion-number violating cross section could become large at very high energies due to multi-boson final states.) As before, denote the generic virtualities of the external legs by Q^2. For instanton size $\rho^2\gg Q^{-2}$, the overlap of the zero modes with the wave functions on the external legs will provide a strong damping factor. Hence the saddle point $\rho_{\rm sp}$ of the integration over the instanton's size is $\rho^2_{\rm sp}\sim Q^{-2}$. If Q^2 is much larger than the confinement scale, the running coupling $g_r=g_r(\rho_{\rm sp}\sim\sqrt{Q^{-2}})$ is small. This justifies the use of the one-instanton approximation.

Ultimately, the most visible consequence of the anomaly in one-flavor QCD is that the lightest pseudo-scalar state (the "η' meson") is *not* light compared to the confinement scale (see e.g. Ref. [8]). The chiral-symmetry breaking effect obtained from the semi-classical instanton calculation is much smaller since it is controlled by the small parameter $\exp[-8\pi^2/g_r^2(\rho_{\rm sp})]$. We resort to this deep Euclidean regime, because only there are we able to apply analytic methods to accurately calculate the consequences of the anomaly.

The above considerations have to do with the asymptotically free nature of the Yang-Mills coupling, and therefore they are completely independent of the regularization method. Moreover, our explicit calculation has demonstrated that no uncontrolled lattice artifacts occur. Finally, we note that the discretization of regular-gauge instantons does yield gauge-field configurations that fail to satisfy the lattice Yang-Mills equation of motion, but only by a small amount $O(a/\rho_{\rm sp})$. Instanton-sector Feynman rules that generate a systematic expansion in $g_r^2(\rho_{\rm sp})$ can be derived in the presence of an approximate classical solution; see e.g. Ref. [27].

III. CHIRAL GAUGE THEORIES

The lesson of the previous section is that a 't Hooft vertex can be interpreted as an order parameter for the spontaneous breaking of the U(1) chiral symmetry in a regularization scheme where chiral (but not gauge) invariance is preserved. The introduction of a small mass term, which is sent to zero after the infinite-volume limit was taken, provides the necessary coupling to an "external magnetic field" and allows the expectation value of a 't Hooft vertex to be non-zero. This reasoning is valid both in the continuum and on the lattice, if one uses the gauge-fixing approach.

The generalization of the previous calculation to 't Hooft vertices that violate the fermion-number symmetry of a chiral gauge theory is relatively straightforward. Starting from the lattice theory, a mass perturbation that lifts the fermionic zero modes will again allow us to keep the (approximate) zero modes under control while taking the infinite-volume and continuum limits. Performing next the limit $m\to 0$, we will recover the 't Hooft vertices as before. The only step which may not be obvious is that a mass perturbation that lifts all zero modes exists in the continuum.

In this section we demonstrate the existence of the necessary mass perturbation by working out the example of an SO(10) chiral gauge theory. [Attempting to construct the necessary mass perturbation for the most general anomaly-free chiral gauge theory may be tedious, and the SO(10) example is general enough to encompass the standard model as well as the most popular grand unification schemes.] In a one-generation SO(10) theory the Weyl fermions reside in the complex **16** representation. We introduce covariant derivatives $(M,N=1,\ldots,10)$

$$D_\mu=\partial_\mu+iA_\mu^{MN}\Sigma_{MN},$$

$$\overline{D}_\mu=\partial_\mu+iA_\mu^{MN}\overline{\Sigma}_{MN}, \tag{3.1}$$

in the **16** and the $\overline{\mathbf{16}}$ representations respectively. The SO(10) generators are defined via

$$\frac{i}{2}[\Gamma_M,\Gamma_N]=\frac{1}{2}(1+\Gamma_{11})\Sigma_{MN}+\frac{1}{2}(1-\Gamma_{11})\overline{\Sigma}_{MN}. \tag{3.2}$$

We use the 32 by 32 representation of the ten-dimensional gamma matrices given in Appendix B. The (continuum) Lagrangian is

$$\mathcal{L}=\overline{\psi_L}\sigma_\mu D_\mu\psi_L. \tag{3.3}$$

In an instanton background there are four left-handed zero modes, one for each quark or lepton. We will show that a suitable mass term lifts all four zero modes. To prepare for the introduction of the mass term we first rewrite the Lagrangian in terms of Majorana-like fermions

$$\Psi=\begin{pmatrix}\epsilon\mathcal{C}\overline{\psi_L}^T\\ \psi_L\end{pmatrix},\qquad \overline{\Psi}=(\overline{\psi_L}\quad \psi_L^T\epsilon\mathcal{C}), \tag{3.4}$$

satisfying

$$\overline{\Psi}\equiv\Psi^TC_4\mathcal{C}, \tag{3.5}$$

where ϵ is the anti-symmetric two-dimensional tensor and C_4 is the four-dimensional charge conjugation matrix (see Appendix A). The 16 by 16 matrix $\mathcal{C}$, which is related to the ten-dimensional charge-conjugation matrix, is defined in Eq. (B3). It satisfies $\mathcal{C}^*=\mathcal{C}^T=\mathcal{C}^{-1}=\mathcal{C}$. In terms of the Majorana-like fields the Lagrangian is rewritten as

$$\mathcal{L}=\frac{1}{2}\overline{\Psi}D_0\Psi, \tag{3.6}$$

where

$$D_0=\gamma_\mu(D_\mu P_L+\overline{D}_\mu P_R)=\begin{pmatrix}0 & \sigma_\mu D_\mu\\ \overline{\sigma}_\mu\overline{D}_\mu & 0\end{pmatrix}. \tag{3.7}$$

Note that $D_0^\dagger\neq -D_0$: unlike the QCD case, D_0 is not anti-Hermitian. One can show that

$$D_0^\dagger C_4\mathcal{C}=C_4\mathcal{C}D_0^*. \tag{3.8}$$

Appendixes A and B contain a number of useful relations which have been used above.

Equipped with the Majorana formulation we introduce a mass term

$$\frac{m}{2}\overline{\Psi}\Psi=\frac{m}{2}(\psi_L^T\epsilon\mathcal{C}\psi_L+\overline{\psi_L}\epsilon\mathcal{C}\overline{\psi_L}^T). \tag{3.9}$$

The mass term breaks explicitly the fermion-number symmetry and, in the limit $m\to 0$, provides the "seed" for spontaneous symmetry breaking. (The mass term also breaks the chiral gauge invariance; see below.) The fermion operator becomes

$$D(m)=D_0+m. \tag{3.10}$$

Equation (3.8) holds for $D(m)$ too.

We will soon prove that $D(m)$ has no zero modes, for $m\neq 0$. But first, we give a simple physical explanation why this should be expected. Observe that

$$\mathcal{C}\otimes I=-iC_{10}\Gamma_{10},$$

where I is the two-by-two identity matrix. Introducing a 32-component spinor Ψ' whose first 16 components are equal to Ψ we may write

$$\overline{\Psi}\Psi\equiv\Psi^TC_4\mathcal{C}\Psi=-\frac{i}{2}(\Psi')^TC_4C_{10}\Gamma_{10}(1+\Gamma_{11})\Psi'. \tag{3.11}$$

Because of Γ_{10}, the mass term can be thought of as coming from the vacuum expectation value of a Higgs field in the **10** representation. This vacuum expectation value breaks SO(10) down to SO(9). Since all spinor representations of SO(9) are real, the fermions can acquire Majorana masses consistently with SO(9) invariance. Moreover, the 16-dimensional representation of SO(9) is irreducible, and therefore *all* 16 fermions acquire a Majorana mass.

We will now show in more detail that there are no exact zero modes for $m\neq 0$ and that the fermion-number-violating 't Hooft interaction is recovered in the limit $m\to 0$. We describe the main steps here, relegating further technical details to Appendix B. In order to obtain information on the fermion propagator for $m\neq 0$ we will need the general formalism of Appendix D, which applies to Hermitian operators. We will thus consider the following Hermitian operator (and corresponding propagator):

$$\mathcal{D}(m)=\begin{pmatrix}0 & D(m)\\ D^\dagger(m) & 0\end{pmatrix},\qquad \mathcal{G}(m)=\begin{pmatrix}0 & G^\dagger(m)\\ G(m) & 0\end{pmatrix}. \tag{3.12}$$

Note that $D(m)$ carries a four-component spinor index, and $\mathcal{D}(m)$ carries an eight-component spinor index. Let us first enumerate the zero modes for $m=0$. There are the four original left-handed zero modes u_i that belong to the **16** and satisfy $\sigma_\mu D_\mu u_i=0$. In addition, define

$$v_i=-\epsilon\mathcal{C}u_i^* . \tag{3.13}$$

Using the left handedness of u_i and Eq. (3.8) one has

$$D_0^\dagger\begin{pmatrix}0\\ v_i\end{pmatrix}=-C_4\mathcal{C}\left(D_0\begin{pmatrix}0\\ u_i\end{pmatrix}\right)^*=0. \tag{3.14}$$

Therefore the v_i are left-handed zero modes of $D_0^\dagger$ that belong to $\overline{\mathbf{16}}$. The propagator $\mathcal{G}_0(x,y)$ is orthogonal to all eight zero modes and satisfies

$$\mathcal{D}_0\mathcal{G}_0(x,y)=\delta^4(x-y)-\mathcal{P}(x,y), \tag{3.15}$$

where the zero-mode projector is

$$\mathcal{P}(x,y)=\begin{pmatrix}0&0&0&0\\ 0&v_i(x)v_i^\dagger(y)&0&0\\ 0&0&0&0\\ 0&0&0&u_i(x)u_i^\dagger(y)\end{pmatrix}. \tag{3.16}$$

From these equations one can read off the relations satisfied by the chiral propagator:

$$D_0G_0(x,y)=\delta^4(x-y)-P_Lv_i(x)v_i^\dagger(y)P_L ,$$

$$G_0(x,y)\overleftarrow{D}_0=\delta^4(x-y)-P_Lu_i(x)u_i^\dagger(y)P_L .$$

The derivative acting to the left has a minus sign.

We now turn to $m\neq 0$. As explained earlier, the fermion-number symmetry is broken explicitly. This is reflected by the fact that $G(m)$ does not anti-commute with γ_5 [compare Eq. (3.21) below]. An inspection of Eqs. (3.10) and (3.12) reveals that to first order, the mass perturbation can have non-zero matrix elements only between a **16** and a $\overline{\mathbf{16}}$ zero mode. Let

$$\lambda_{ij}=\langle v_i|m|u_j\rangle=m\langle v_i|u_j\rangle. \tag{3.17}$$

We find, using Eq. (D15),

$$G(m)=|u_i\rangle\lambda_{ij}^{-1}\langle v_j|+O(1). \tag{3.18}$$

The first term on the right-hand side is $O(1/m)$. Furthermore, using Eq. (3.13),

$$\lambda_{ij}=m\int u_i^T\epsilon\mathcal{C}u_j , \tag{3.19}$$

which implies that λ_{ij} is antisymmetric. In the zero-mode sector, the Majorana-fermion determinant is the analytic square root (Pfaffian)

$$\det{}^{1/2}(\lambda)=\frac{1}{8}\epsilon_{ijkl}\lambda_{ij}\lambda_{kl} . \tag{3.20}$$

In Appendix B we prove that $\det(\lambda)\neq 0$ for (almost) every embedding of the instanton in SO(10).

Next consider correlation functions in the instanton sector. In terms of the original Weyl fields,

$$G(m)=\langle\Psi\bar\Psi\rangle=\begin{pmatrix}\langle\epsilon\mathcal{C}\overline{\psi}_L^T\overline{\psi}_L\rangle & \langle\epsilon\mathcal{C}\overline{\psi}_L^T\psi_L^T\epsilon\mathcal{C}\rangle\\ \langle\psi_L\overline{\psi}_L\rangle & \langle\psi_L\psi_L^T\epsilon\mathcal{C}\rangle\end{pmatrix}. \tag{3.21}$$

Let us first see what happens if we saturate two fermion fields $\psi_L(x)\psi_L(y)$ by the $O(1/m)$ part of the propagator. Using Eqs. (3.13), (3.18), and noting the lower-right entry of Eq. (3.21), we obtain a factor

$$u_i(x)u_j(y)\lambda_{ij}^{-1} . \tag{3.22}$$

By itself, this will give a vanishing result in the limit $m\to 0$ because $\det^{1/2}(\lambda)$ is $O(m^2)$. Next consider saturating the product of four fields $\psi_L(x)\psi_L(y)\psi_L(z)\psi_L(w)$. Summing over all possible contractions and paying attention to Fermi statistics we get

$$u_i(x)u_j(y)u_k(z)u_l(w)(\lambda_{ij}^{-1}\lambda_{kl}^{-1}+\lambda_{ik}^{-1}\lambda_{lj}^{-1}+\lambda_{il}^{-1}\lambda_{jk}^{-1}). \tag{3.23}$$

The expression in parentheses is completely anti-symmetric in the four indices i,j,k,l. [If we would try to saturate $2n$ fermion fields for some $n>2$ with the $O(1/m)$ part of the propagator, the result would be identically zero due to the anti-symmetrization.] To evaluate the sum we write

$$(\lambda_{ij}^{-1}\lambda_{kl}^{-1}+\lambda_{ik}^{-1}\lambda_{lj}^{-1}+\lambda_{il}^{-1}\lambda_{jk}^{-1})=c\,\epsilon_{ijkl} , \tag{3.24}$$

and contracting with another ϵ_{ijkl} we find

$$c=\frac{1}{8}\epsilon_{ijkl}\lambda_{ij}^{-1}\lambda_{kl}^{-1}=\det{}^{1/2}(\lambda^{-1}). \tag{3.25}$$

This cancels against Eq. (3.20), and in the limit $m\to 0$, one is left with

$$\langle\psi_L(x)\psi_L(y)\psi_L(z)\psi_L(w)\rangle=\epsilon_{ijkl}u_i(x)u_j(y)\times u_k(z)u_l(w)\mathrm{Det}' , \tag{3.26}$$

which is the expected 't Hooft vertex.

The mass term (3.9) breaks not only the unwanted fermion-number symmetry; it also breaks chiral gauge invariance. This, however, does not lead to any disaster; in fact, we know that in the UV-regulated theory the gauge symmetry is already broken by the regulator. The crucial point is that, in a covariantly gauge-fixed theory, the ultraviolet behavior of the vector-boson propagator is $1/p^2$ for all polarizations. Consequently, the theory remains renormalizable even if terms that break gauge invariance are added.

Moreover, the addition of a mass term does not change the nature of the coupling of the theory. The one-loop beta function is unaffected by the mass term, and so the theory is still asymptotically free. Also, provided we are careful to employ a mass-independent renormalization prescription, universality of the renormalized coupling should be preserved. The same considerations imply that the continuum limit of the lattice theory should exist for $m\neq 0$, too. In the renormalized theory, a fermion-mass term is expected to induce a vector-boson mass term, and so unitarity will be violated by $O(m)$ effects. In the limit $m\rightarrow 0$ (keeping physical scales fixed) unitarity should be recovered.

IV. DISCUSSION

The classically conserved chiral U(1) symmetry is not preserved by any gauge-invariant regularization, and in the quantized theory physical observables exist that violate this symmetry. In this paper we have considered an important aspect of regularization methods which are not gauge invariant but, instead, respect chiral U(1) invariance. Our concrete motivation to do so is the gauge-fixing approach to (chiral) lattice gauge theories. We showed that even with a chiral-U(1) invariant regulator, a careful treatment of the *infrared* limit reproduces correctly the gauge invariant, chiral-symmetry-violating 't Hooft vertices. In essence, our conclusions are as follows.

(I) Since the chiral U(1) symmetry is preserved by the regularization but is not respected by physical amplitudes, it must be broken spontaneously. (II) Therefore, in order to obtain the physical amplitudes, one should introduce a mass perturbation that breaks the chiral symmetry explicitly, and take the limit $m\rightarrow 0$ *after the infinite-volume limit. (III) The Hilbert space of the gauge-fixed theory will contain a corresponding Goldstone pole, but it is unphysical, because it originates from an unphysical (gauge non-invariant) conserved current.*

The second and third statements are closely tied to the first one. When there is SSB the thermodynamical limit is defined by introducing an "external magnetic field" (here the mass term) which is switched off after the infinite-volume limit has been taken. As for the existence of a massless pole, it is a consequence of locality and the Goldstone theorem which, in the present context, has been noted before in Refs. [6,7].

The mass perturbation allows us to avoid the Banks paradox, namely the apparent conflict between the symmetries of the regularized theory and of the physical amplitudes. Starting from a gauge-fixed lattice theory that has the unwanted chiral symmetry, we have demonstrated through explicit examples how this mechanism works for the anomalous axial symmetry of QCD-like theories and for fermion-number-violating processes in chiral gauge theories.

Zero modes belong to the low end of the fermion spectrum which, in general, is sensitive to infrared details such as having $m\neq 0$, finite versus infinite volume, and choices of boundary conditions. Keeping $m>0$ provides an infrared regularization for all fermionic correlation functions and allows us to take the infinite-volume limit without difficulty. Once in infinite volume, we obtain the correct 't Hooft vertices because we evidently have the correct number of (approximate) zero modes.

As explained earlier, because of CP invariance, the vacuum angle θ was equal to zero in the previous sections. Considering the one-flavor QCD example for simplicity, let us examine a mass term with a general axial-U(1) phase. (The generalization to chiral gauge theories is straightforward; see Ref. [28] for a discussion of θ vacua in the context of the standard Wilson action.) The lattice action is now defined by replacing the mass term in Eq. (2.2) with a new mass term pointing in an arbitrary axial-U(1) direction (the parameter m is real):

$$m\overline{\psi}e^{i\theta\gamma_5}\psi=m(e^{i\theta}\overline{\psi_L}\psi_R+e^{-i\theta}\overline{\psi_R}\psi_L). \tag{4.1}$$

By applying an axial-U(1) transformation ($\psi\rightarrow e^{-i\theta\gamma_5/2}\psi$, $\chi\rightarrow e^{i\theta\gamma_5/2}\chi$, etc.) and using the invariance of the lattice theory for $m=0$, we can relate the value of any $\theta\neq 0$ correlation function to its value at $\theta=0$. For the correlation function of Eq. (2.1) the result is

$$\langle\psi_L\overline{\psi_R}\rangle_\theta=e^{i\theta}\langle\psi_L\overline{\psi_R}\rangle_{\theta=0}. \tag{4.2}$$

The subscript θ refers to the angle of the mass term (4.1). [Note that, in Eq. (4.2), $\exp(i\theta)$ may be re-expressed as $\exp(-iq\theta/2)$ where $q=-2$ is the axial charge of $\psi_L\overline{\psi_R}$.]

Equation (4.2) is a rigorous result in the lattice theory. A similar relation holds in the thermodynamical limit. Hence the order parameter for axial-U(1) symmetry breaking—the 't Hooft vertex—acquires a phase equal to the phase of the applied mass term. This proves that the SSB ground states are indeed parametrized by the (vacuum) θ angle.

The relevance of θ vacua to the U(1) problem was discussed in detail in the literature. In a continuum treatment the path integral can only be expanded perturbatively around selected classical fields, and θ vacua must be incorporated "by hand." Here, we showed that the θ angle of the 't Hooft vertex arises as an unavoidable consequence of the lattice regularization.

While our explicit lattice calculation does not depend on this, let us recall some related observations from Refs. [6–8]. Any sector with a fixed topological charge, such as the one-instanton sector, is a superposition of all the θ vacua. But for any $m\neq 0$, however small, the vacuum energy density is a non-trivial function of θ. This explains why a unique θ vacuum is selected in the thermodynamical limit.

Had the infinite-volume limit been taken while keeping $m=0$, the θ vacua would have remained exactly degenerate. This prescription would yield a vanishing result for chiral-symmetry violating amplitudes. But since *clustering* would be violated, this prescription is inconsistent. Returning to the lattice regularization, this may be explained as follows. If the lattice volume is finite, $V<\infty$, the limit $m\rightarrow 0$ must be smooth and, hence, independent of θ: namely,

$$\langle\psi_L\overline{\psi_R}\rangle_{m\rightarrow 0,V<\infty,\theta}=\langle\psi_L\overline{\psi_R}\rangle_{m\rightarrow 0,V<\infty,\theta=0}. \tag{4.3}$$

Since this must be true for all θ simultaneously with Eq. (4.2), the conclusion is that the 't Hooft vertex vanishes on a finite lattice, if we set $m=0$. [This argument is really an alternative explanation of the Ward identity (1.6), but makes the role of the axial phase more explicit.] An implication is that, clearly, one would have to keep $m\neq 0$ in a numerical simulation in order to recover 't Hooft vertices.

The 't Hooft vertices are characterized by an "m/m" behavior in the thermodynamical limit. Given a chiral U(1) Ward identity, should an "m/m" behavior be expected from any other term except the symmetry-breaking expectation value? The answer is no. As a concrete example consider the following momentum-space Ward identity in one-flavor QCD with $m\neq 0$, as defined on the lattice via the gauge-fixing approach (Sec. II). For $|pa|\ll 1$ it reads [compare Eq. (C7)]

$$ip_\mu\langle \tilde{J}^{\rm latt}_{5\mu}(p)J_5\rangle = -2m\langle \tilde{J}_5(p)J_5\rangle + 2\langle \bar{\psi}\psi\rangle. \tag{4.4}$$

In this equation, $J^{\rm latt}_{5\mu}$ is the conserved U(1) axial current in the limit $m\to 0$, and $\bar{\psi}\psi$ and J_5 are the local scalar and pseudo-scalar lattice densities. As in Eq. (1.8) the tilde denotes Fourier transform. The expectation value of $\bar{\psi}\psi$ corresponds to the limit $x=y$ in Eq. (2.1). While in this limit the semi-classical calculation ceases to be reliable, we take the non-zero result for the 't Hooft vertex as evidence that $\langle\bar{\psi}\psi\rangle\neq 0$.

The contribution of an approximate Goldstone pole to $\langle\tilde{J}_5(p)J_5\rangle$ should be proportional to $(p^2+vm)^{-1}$ where v is a dimensionful constant. The corresponding contribution to the Ward identity goes like $m/(p^2+vm)$ and vanishes in the limit $m\to 0$. The contribution of all other excitations to $\langle\tilde{J}_5(p)J_5\rangle$ should be less infrared singular. Therefore nothing that behaves like "m/m" should arise from $\langle\tilde{J}_5(p)J_5\rangle$, as long as we are careful to keep the momentum *not* strictly zero. Indeed, sending $p\to 0$ as a *limit* is an inherent part of the Goldstone theorem (see for example Ref. [29]). For $m\to 0$ we thus obtain

$$ip_\mu\langle J^{\rm latt}_{5\mu}(p)J_5\rangle = 2\langle\bar{\psi}\psi\rangle. \tag{4.5}$$

This equation is a special case of the Ward identity (1.8).

The problem of fermion-number violation in lattice chiral gauge theories was previously also addressed in Ref. [30]. In the (axial) Schwinger model [31], these authors examined a lattice-fermion Hamiltonian that has a "superfluous" U(1) global symmetry. They monitored the response of the fermion ground state to an adiabatic evolution of the (Abelian) gauge field that changes the topological charge of the gauge vacuum. They found that a U(1) charge of the anticipated amount is produced in the process.

The clash between chiral-U(1) and gauge symmetries is at the heart of their argument. Because of the lack of exact gauge invariance at the lattice level, the initial and final bare vacua are not gauge transforms of each other and their bare U(1) charges are different. During the evolution the bare charge is necessarily conserved. But since the bare charge of the ground state changes in the process, there is a corresponding change in the *normal-ordered* charge defined with respect to the ground state.

The introduction of $m\neq 0$ in our work was necessary to control the infrared behavior of a dynamical gauge-fermion system that undergoes spontaneous symmetry breaking of a peculiar nature. In contrast, in Ref. [30] only the response of the spectrum of the axial Dirac operator to an external gauge field was considered, and so it was not necessary to introduce the mass perturbation.

In conclusion, in this paper we have demonstrated convincingly that, in spite of the exact chiral U(1) invariance of the lattice action in the gauge-fixing approach, fermion-number violating processes do occur, thus resolving the questions raised in Ref. [20].

ACKNOWLEDGMENTS

We would like to thank Pierre van Baal, Aharon Casher and Mike Creutz for useful discussions. Both of us would like to thank the Institute for Nuclear Theory at the University of Washington for hospitality and support provided for part of this work. This research is supported in part by a grant from the United States–Israel Binational Science Foundation, and by the U.S. Department of Energy.

APPENDIX A: NOTATION

The group generators are normalized according to

$$\mathrm{tr}T_aT_b=\frac{1}{2}\delta_{ab}. \tag{A1}$$

The dual tensor is

$$\tilde{F}_{\mu\nu}=\frac{1}{2}\epsilon_{\mu\nu\lambda\rho}F_{\lambda\rho}. \tag{A2}$$

The topological charge is

$$\nu=\frac{g^2}{16\pi^2}\mathrm{tr}\int d^4xF\tilde{F}. \tag{A3}$$

The Hermitian gamma matrices obey the Dirac algebra

$$\{\gamma_\mu,\gamma_\nu\}=\delta_{\mu\nu}, \quad \gamma_\mu^\dagger=\gamma_\mu. \tag{A4}$$

In four dimensions we use the representation

$$\gamma_\mu=\begin{pmatrix}0 & \sigma_\mu\\ \bar{\sigma}_\mu & 0\end{pmatrix}, \quad \gamma_5=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{A5}$$

$$\sigma_\mu=(\vec{\sigma},i), \quad \bar{\sigma}_\mu=\sigma_\mu^\dagger=(\vec{\sigma},-i). \tag{A6}$$

The chiral projectors are

$$P_R=(1+\gamma_5)/2, \quad P_L=(1-\gamma_5)/2. \tag{A7}$$

Charge conjugation matrices play a key role in the Majorana formulation of Sec. III. In any even dimension the charge conjugation matrix is defined by (see e.g. Ref. [32])

$$C\gamma_\mu = -\gamma_\mu^T C, \tag{A8}$$

and satisfies $C^{-1}=C^\dagger=C^T$. In $8n+2$ and $8n+4$ dimensions, $C^T=-C$, while in $8n+6$ and $8n$ dimensions, $C^T=C$. For the above four-dimensional gamma matrices the charge-conjugation matrix can be chosen as

$$C_4=\gamma_3\gamma_1=\begin{pmatrix}\epsilon & 0\\ 0 & \epsilon\end{pmatrix}. \tag{A9}$$

It is unique up to a sign. The two-dimensional antisymmetric tensor (with $\epsilon_{12}=1$) is

$$\epsilon=i\sigma_2=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}. \tag{A10}$$

For any even dimension and $\mu\neq\nu$ one has

$$C\gamma_\mu\gamma_\nu=-\gamma_\mu^T C\gamma_\nu=\gamma_\mu^T\gamma_\nu^T C=(\gamma_\nu\gamma_\mu)^T C=-(\gamma_\mu\gamma_\nu)^T C. \tag{A11}$$

In four dimensions one has

$$C\gamma_\mu(1\pm\gamma_5)=[C\gamma_\mu(1\mp\gamma_5)]^T, \tag{A12}$$

a relation which generalizes to $4n$ dimensions.

APPENDIX B: SO(10)-OLOGY

We define the ten-dimensional gamma matrices by the following tensor products:

$$\begin{aligned}
\Gamma_1&=\sigma_1\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_2&=\sigma_2\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_3&=\sigma_3\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_4&=I\otimes\sigma_2\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_5&=I\otimes\sigma_3\otimes\sigma_1\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_6&=I\otimes I\otimes\sigma_2\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_7&=I\otimes I\otimes\sigma_3\otimes\sigma_1\otimes\sigma_1,\\
\Gamma_8&=I\otimes I\otimes I\otimes\sigma_2\otimes\sigma_1,\\
\Gamma_9&=I\otimes I\otimes I\otimes\sigma_3\otimes\sigma_1,\\
\Gamma_{10}&=I\otimes I\otimes I\otimes I\otimes\sigma_2,\\
\Gamma_{11}&=I\otimes I\otimes I\otimes I\otimes\sigma_3.
\end{aligned} \tag{B1}$$

The ten-dimensional charge conjugation matrix is

$$C_{10}=i\mathcal{C}\otimes\sigma_2, \tag{B2}$$

where

$$\mathcal{C}=\sigma_2\otimes\sigma_3\otimes\sigma_2\otimes\sigma_3. \tag{B3}$$

Notice that C_{10} anti-commutes with Γ_{11}. With the above definitions, Eq. (A11) reads

$$\begin{pmatrix}0 & \mathcal{C}\\ -\mathcal{C} & 0\end{pmatrix}\begin{pmatrix}\Sigma_{MN} & 0\\ 0 & \overline{\Sigma}_{MN}\end{pmatrix}=-\begin{pmatrix}\Sigma_{MN}^T & 0\\ 0 & \overline{\Sigma}_{MN}^T\end{pmatrix}\begin{pmatrix}0 & \mathcal{C}\\ -\mathcal{C} & 0\end{pmatrix}. \tag{B4}$$

Equivalently

$$\mathcal{C}\overline{\Sigma}_{MN}=-\Sigma_{MN}^T\mathcal{C}=-\Sigma_{MN}^*\mathcal{C}, \tag{B5}$$

a relation which is needed for the derivation of Eq. (3.8).

If D satisfies Eq. (3.8) and has no zero modes, its inverse satisfies

$$G(x,y)\equiv\langle\Psi(x)\overline{\Psi}(y)\rangle=-C_4\mathcal{C}G^T(y,x)C_4\mathcal{C}. \tag{B6}$$

By taking a suitable limit, this generalizes to the case where D has exact zero modes and $G(x,y)$ is constructed from the non-zero modes only.

We now show that, for almost every embedding of an instanton in SO(10), $\det(\lambda)\neq 0$ [cf. Eq. (3.19)]; i.e., the mass term of Sec. III lifts the four zero modes. We will first show that $\det(\lambda)\neq 0$ for a particular embedding. We introduce 16 by 16 matrices $\mathcal{S}_i(l)$, $i=1,\ldots,3$, $l=1,\ldots,4$, defined to be tensor products of four two-by-two matrices with σ_i as the lth factor and the identity for the rest. The SO(10) generators Σ_{MN}, with $M,N=1,\ldots,4$, generate two SU(2) groups. For $\mathcal{S}_3(2)=\pm 1$ we label them $\mathrm{SU}(2)_{R,L}$.

Each zero mode is written explicitly as $u_i=u_{\alpha\beta_1\beta_2\beta_3\beta_4,i}(x)$, where $\alpha,\beta_1,\beta_2,\beta_3,\beta_4=1,2$. Here α is the spin index. The other four indices correspond to the tensor product that defines the SO(10) generators in the **16** representation. Assuming that the instanton resides in e.g. the $\mathrm{SU}(2)_L$ subgroup defined above, the zero modes have $\mathcal{S}_3(2)=-1$ (equivalently $\beta_2=2$). Their $\mathrm{SU}(2)_L$ index is β_1. Using the explicit solution of the isospin one-half zero mode (for a regular-gauge instanton) we have

$$u_{\alpha\beta_1\beta_2\beta_3\beta_4,i}(x)=\mathcal{N}\delta_{\alpha,\beta_1}\delta_{\beta_2,2}\mathcal{O}_{\beta_3\beta_4,i}f(x^2), \tag{B7}$$

where $f(x^2)=(x^2+\rho^2)^{-3/2}$ and $\mathcal{N}$ is a normalization factor. The constants $\mathcal{O}_{\beta_3\beta_4,i}$ define the four independent zero modes. We will label them by the eigenvalues of $\mathcal{S}_2(3)$ and $\mathcal{S}_3(4)$. Replacing the index $i=1,\ldots,4$, by a pair of indices $\tau_1,\tau_2=1,2$, we take $\mathcal{O}_{\beta_3\beta_4,\tau_1\tau_2}=i(\sigma_2)_{\beta_3\tau_1}(\sigma_3)_{\beta_4\tau_2}$. In matrix notation, $\mathcal{O}=i\sigma_2\otimes\sigma_3$, and $\mathcal{O}$ has similar properties to the four-dimensional charge-conjugation matrix. Putting together Eqs. (3.19), (B3) and (B7) we get

$$\begin{aligned}\lambda_{\tau_1\tau_2,\tau_1'\tau_2'}&=-m\mathcal{N}^2\int d^4x f^2(x^2)\mathrm{tr}(\epsilon^2)(\mathcal{O}^T(i\sigma_2\\ &\quad\otimes\sigma_3)\mathcal{O})_{\tau_1\tau_2,\tau_1'\tau_2'}=m(i\sigma_2)_{\tau_1\tau_1'}(\sigma_3)_{\tau_2\tau_2'},\end{aligned} \tag{B8}$$

which proves $\det(\lambda)\neq 0$ for this special case. In the first row, the explicit $i\sigma_2\otimes\sigma_3$ comes from the last two factors in the tensor product (B3), while $\mathrm{tr}\,\epsilon^2$ comes from the first factor in this tensor product, and the explicit ϵ in Eq. (3.19). The transition from the first to the second row implicitly defines the normalization constant.

Suppose now that a global rotation $R\in SO(10)$ is applied to the above special embedding of the instanton. The new zero modes are $u_i'=Ru_i$. We claim that $\det(\lambda(R))\neq 0$ for almost every R. The proof is simple. Suppose on the contrary that $\det[\lambda(R)]=0$ for every R in some open subset of SO(10). Since the embedding and, hence, $\det(\lambda)$ are analytic functions of R, this would imply that $\det(\lambda)=0$ for all R. This, however, contradicts Eq. (B8) in the special case $R=I$. Therefore $\det(\lambda)=0$ may be true, at most, on a measure zero subset of SO(10).

APPENDIX C: LATTICE FORMULAS

The free symmetric lattice derivative is defined for any function f_x by

$$\hat{\partial}_{x,\mu}f=\frac{1}{2a}(f_{x+\hat{\mu}}-f_{x-\hat{\mu}}),$$

where $\hat{\mu}$ is a unit vector in the μ direction. The corresponding covariant derivative is

$$\hat{D}_{x,\mu}f=\frac{1}{2a}(U_{x,\mu}f_{x+\hat{\mu}}-U^{\dagger}_{x-\hat{\mu},\mu}f_{x-\hat{\mu}}),$$

where $U_{x,\mu}$ is the link variable. The free lattice Laplacian is

$$\hat{\Box}_{x,\mu}f=\frac{1}{a^2}(f_{x+\hat{\mu}}+f_{x-\hat{\mu}}-2f_x).$$

Given a set of left-handed fields ψ_L^i and corresponding spectator fields χ_R^i, the chiral Wilson Lagrangian is [compare the upper-right block in Eq. (2.2)]

$$\mathcal{L}=\sum_i\left(\overline{\psi}_L^i\sigma_\mu\hat{D}_\mu\psi_L^i+\overline{\chi}_R^i\overline{\sigma}_\mu\hat{\partial}_\mu\chi_R^i-\frac{ar}{2}(\overline{\chi}_R^i\hat{\Box}\psi_L^i+\overline{\psi}_L^i\hat{\Box}\chi_R^i)\right),\tag{C1}$$

where r is Wilson parameter. This action is invariant under a U(1) rotation of all fermion fields

$$\psi_L^i\to e^{i\alpha}\psi_L^i,\quad \chi_R^i\to e^{i\alpha}\chi_R^i,\quad \overline{\psi}_L^i\to e^{-i\alpha}\overline{\psi}_L^i,\quad \overline{\chi}_R^i\to e^{-i\alpha}\overline{\chi}_R^i.\tag{C2}$$

The conserved U(1) current is (see Ref. [24])

$$\begin{aligned}J^{\mathrm{L,latt}}_{x,\mu}=\frac{1}{2}\sum_i\{&\overline{\psi}^i_{L,x}\sigma_\mu U_{x,\mu}\psi^i_{L,x+\hat{\mu}}+\overline{\psi}^i_{L,x+\hat{\mu}}\sigma_\mu U^{\dagger}_{x,\mu}\psi^i_{L,x}\\&+\overline{\chi}^i_{R,x}\overline{\sigma}_\mu\chi^i_{R,x+\hat{\mu}}+\overline{\chi}^i_{R,x+\hat{\mu}}\overline{\sigma}_\mu\chi^i_{R,x}-r(\overline{\psi}^i_{L,x}\chi^i_{R,x+\hat{\mu}}\\&+\overline{\chi}^i_{R,x}\psi^i_{L,x+\hat{\mu}}-\overline{\psi}^i_{L,x+\hat{\mu}}\chi^i_{R,x}-\overline{\chi}^i_{R,x+\hat{\mu}}\psi^i_{L,x})\}.\end{aligned}\tag{C3}$$

It satisfies the conservation equation

$$\sum_\mu(J^{\mathrm{L,latt}}_{x,\mu}-J^{\mathrm{L,latt}}_{x-\hat{\mu},\mu})=0.\tag{C4}$$

In the special case of one-flavor QCD let us introduce Dirac fermions $\psi=(\psi_R,\psi_L)$, $\overline{\psi}=(\overline{\psi}_L,\overline{\psi}_R)$, $\chi=(\chi_R,\chi_L)$, $\overline{\chi}=(\overline{\chi}_L,\overline{\chi}_R)$. The axial transformation is

$$\psi\to e^{-i\alpha\gamma_5}\psi,\quad \chi\to e^{i\alpha\gamma_5}\chi,\quad \overline{\psi}\to\overline{\psi}e^{-i\alpha\gamma_5},\quad \overline{\chi}\to\overline{\chi}e^{i\alpha\gamma_5},\tag{C5}$$

and the axial current is [note that in Eq. (2.2) we set $r=1$]

$$\begin{aligned}J^{\mathrm{latt}}_{x,5\mu}=\frac{1}{2}\{&\overline{\psi}_x\gamma_5\gamma_\mu U_{x,\mu}\psi_{x+\hat{\mu}}+\overline{\psi}_{x+\hat{\mu}}\gamma_5\gamma_\mu U^{\dagger}_{x,\mu}\psi_x\\&-\overline{\chi}_x\gamma_5\gamma_\mu\chi_{x+\hat{\mu}}-\overline{\chi}_{x+\hat{\mu}}\gamma_5\gamma_\mu\chi_x-r(\overline{\psi}_x\gamma_5\chi_{x+\hat{\mu}}\\&-\overline{\chi}_x\gamma_5\psi_{x+\hat{\mu}}-\overline{\psi}_{x+\hat{\mu}}\gamma_5\chi_x+\overline{\chi}_{x+\hat{\mu}}\gamma_5\psi_x)\}.\end{aligned}\tag{C6}$$

For $m\neq 0$ the axial current satisfies the partial-conservation equation

$$\frac{1}{a}\sum_\mu(J^{\mathrm{latt}}_{x,5\mu}-J^{\mathrm{latt}}_{x-\hat{\mu},5\mu})=-2mJ_{x5}.\tag{C7}$$

The difference operator on the left-hand side (the free backward derivative) becomes $[1-\exp(-iap_\mu)]/a=ip_\mu+\cdots$ in momentum space. The local scalar and pseudo-scalar lattice densities are $\overline{\psi}_x\psi_x$ and $J_{x5}=\overline{\psi}_x\gamma_5\psi_x$. As usual they are related by an axial rotation. They do not mix with the corresponding spectator-field densities thanks to the shift symmetry [22].

We now explain why singular-gauge instantons are suppressed on the lattice by the gauge-fixing action of Ref. [33]. Recall that, in a singular gauge, the instanton's vector potential near the gauge singularity (located at $x=0$) is $A_\mu\sim\Phi(x)\partial_\mu\Phi^\dagger(x)/g$ where $\Phi(x)=\sigma_\mu x_\mu/|x|$. The magnitude of this vector potential grows like $1/(g|x|)$.

On the lattice let us make the (bare-field) rescaling $A_{x,\mu}\to A'_{x,\mu}=g_0A_{x,\mu}$. Dropping the prime, we expand the link variable as $U_{x,\mu}=\exp(iaA_{x,\mu})$. The gauge-fixing action contains the expected longitudinal kinetic term

$$\frac{1}{\xi_0g_0^2}\mathrm{tr}\left(\sum_\mu\partial_\mu A_\mu\right)^2=\frac{1}{2\xi_0g_0^2}\sum_a\left(\sum_\mu\partial_\mu A^a_\mu\right)^2,\tag{C8}$$

plus irrelevant terms. Here ξ_0 is the bare gauge-fixing parameter and index summations have been shown explicitly. The leading irrelevant term that contributes to the classical potential is

$$\frac{a^2}{2\xi_0g_0^2}\mathrm{tr}\left(\sum_\mu A^2_\mu\sum_\nu A^4_\nu\right).\tag{C9}$$

The irrelevant terms break BRST invariance, and so there is no reason that regular-gauge and singular-gauge instantons will have the same lattice action.

Consider now some lattice discretization of the singular-gauge vector potential. Inevitably, the (rescaled) vector potential will be $O(1/a)$ in the hypercube(s) containing the point $x=0$ and the vicinity. For such a vector potential $U_{x,\mu}-I=O(1)$. The positivity of expression (C9) and (since there are infinitely many other irrelevant terms) of the gauge-fixing action as a whole [33] guarantees that the lattice action will be an $O(1)$ quantity *times* $1/g_0^2$. In the continuum limit $g_0\to 0$ any such lattice configuration is suppressed.

APPENDIX D: CONTINUUM PROPAGATORS IN THE PRESENCE OF APPROXIMATE ZERO MODES

Let $H=H_0+\alpha V$ be a Schrödinger-like (i.e. elliptic and self-adjoint) operator in a d-dimensional open infinite space. We assume that H_0 has a finite number of (normalized) zero modes $u_i(x)$ and that H has no zero modes. The inverse of H, denoted G, is defined by

$$HG=1, \tag{D1}$$

where both sides are considered as operators acting on a suitable Hilbert space. In this appendix we explain how to construct a systematic approximation for G. [Equation (D1) may be rewritten in the familiar form $HG(x,y)=\delta(x-y)$ by taking the matrix element of Eq. (D1) between "position eigenstates" $\langle x|$ and $|y\rangle$.]

Let us introduce the notation $|i\rangle$ for the zero modes of H_0, and $|p\rangle$ for the rest of its spectrum. (In instanton problems the zero modes are the only bound states, and $|p\rangle$ denotes the continuum of scattering states.) One has $H_0|p\rangle=E(p)|p\rangle$. The propagator G_0 is defined by

$$H_0G_0=1-|i\rangle\langle i|=|p\rangle\langle p| \tag{D2}$$

and has the spectral representation

$$G_0=|p\rangle(G_0)_{pq}\langle q|, \quad (G_0)_{pq}=E^{-1}(p)\,\delta(p-q). \tag{D3}$$

It is convenient to expand $G=G(\alpha)$ too using the eigenmodes of H_0 as a complete orthonormal basis:

$$G=|i\rangle\beta_{ij}^{-1}\langle j|+|p\rangle G_{pq}\langle q|+|p\rangle f_{pj}\langle j|+|j\rangle f_{pj}^{*}\langle p|. \tag{D4}$$

In this expansion, the basis vectors are fixed (α independent), while the α dependence is carried by the spectral functions β_{ij}, G_{pq} and f_{pj}. Below, we show that the spectral functions can be expanded as power series in α where β_{ij} is $O(\alpha)$ and G_{pq} and f_{pj} are $O(1)$. They will be used to construct approximations for G.

We start by substituting Eq. (D4) into Eq. (D1). Using $H_0|i\rangle=0$ we get

$$1=\alpha V|i\rangle\beta_{ij}^{-1}\langle j|+HG', \tag{D5}$$

where G' consists of the last three terms on the right-hand side of Eq. (D4). By taking the matrix element between zero-mode states $\langle i|$ and $|j\rangle$ and using $\langle p|i\rangle=0$ we get

$$\delta_{ij}=\alpha\langle i|V|k\rangle\beta_{kj}^{-1}+\alpha\langle i|V|q\rangle f_{qj}. \tag{D6}$$

By taking the matrix element between $\langle p|$ and $|j\rangle$ we get

$$0=\alpha\langle p|V|k\rangle\beta_{kj}^{-1}+H_{pq}f_{qj}, \tag{D7}$$

where

$$H_{pq}=\langle p|H|q\rangle. \tag{D8}$$

In order to solve for β_{ij} and f_{qj} we have to invert H_{pq}. Note that H_{pq} is the continuous-index kernel of the operator $H^{\perp}\equiv|p\rangle H_{pq}\langle q|$ which is, by definition, the projection of H onto the subspace orthogonal to the zero modes of H_0. The inverse of the projected operator is defined by $H^{\perp}(H^{\perp})^{-1}=|p\rangle\langle p|$. The corresponding kernel satisfies $(H^{\perp})^{-1}\equiv|p\rangle H_{pq}^{-1}\langle q|$. One has

$$\langle p|H|q\rangle=\langle p|H_0|q\rangle+O(\alpha)=E(p)\,\delta(p-q)+O(\alpha). \tag{D9}$$

Therefore $(H^{\perp})^{-1}$ exists, and [compare Eq. (D3)]

$$H_{pq}^{-1}=E^{-1}(p)\,\delta(p-q)+O(\alpha). \tag{D10}$$

We are now ready to solve for β_{ij} and f_{qj}. Multiplying Eq. (D7) by H_{pq}^{-1} we get

$$f_{qj}=-\alpha H_{qp}^{-1}\langle p|V|k\rangle\beta_{kj}^{-1}, \tag{D11}$$

and substituting this in Eq. (D6) we get

$$\beta_{ij}=\langle i|(\alpha V-\alpha^2V|q\rangle H_{qp}^{-1}\langle p|V)|j\rangle. \tag{D12}$$

Equation (D10) implies that $H_{qp}^{-1}=O(1)$, and hence $\beta_{ij}=O(\alpha)$ and $f_{qj}=O(1)$.

The parametric α dependence of the last spectral function is determined to be $G_{pq}=O(1)$. To show this, take the matrix element of Eq. (D5) between scattering states $\langle p|$ and $|q\rangle$, which gives

$$\delta(p-q)=H_{pp'}G_{p'q}+\alpha\langle p|V|k\rangle f_{qk}^{*}. \tag{D13}$$

Since we already know that $f_{qj}=O(1)$, the last term on the right-hand side is sub-leading. It follows that $G_{pq}=H_{pq}^{-1}+O(\alpha)=(G_0)_{pq}+O(\alpha)$. Moreover, by combining these results it follows that the spectral functions can be expanded as a power series in α, starting at the above-specified order for each spectral function. (See, however, *Comment 1* below.)

The propagator G involves β_{ij}^{-1}, and so it has a Laurent series starting at order $1/\alpha$. To find the singular, $O(1/\alpha)$ piece of the propagator, we keep only the $O(\alpha)$ term on the right-hand side of Eq. (D12). We get $\beta_{ij}=\lambda_{ij,1}+O(\alpha^2)$, where

$$\lambda_{ij,1}=\alpha\langle i|V|j\rangle. \tag{D14}$$

This expression is recognized as the first-order energy shifts of the zero modes, as calculated using degenerate perturbation theory. Substituting in Eq. (D4) we obtain

$$G=|i\rangle\lambda_{ij,1}^{-1}\langle j|+O(1). \tag{D15}$$

While Eq. (D15) is all we will be using in the body of the paper, it is instructive to go one step further and construct also the $O(1)$ part of the propagator. To this end we approximate H^{-1}_{pq} by $(G_0)_{pq}$. Equation (D12) then yields $\beta_{ij}=\lambda_{ij,2}+O(\alpha^3)$ where

$$\lambda_{ij,2}=\langle i|\alpha V-\alpha^2 V G_0 V|j\rangle, \tag{D16}$$

which, as expected, includes also the second-order energy shifts. Next, making a similar approximation in Eq. (D11) we have

$$|p\rangle f_{pj}\langle j|=-\alpha G_0 V|i\rangle\lambda^{-1}_{ij,1}\langle j|+O(\alpha). \tag{D17}$$

This involves the first-order correction to the zero mode wave function. Putting everything together we find that, at $O(1)$, the propagator may be compactly expressed as

$$G=|i,1\rangle\lambda^{-1}_{ij,2}\langle j,1|+G_0+O(\alpha), \tag{D18}$$

where

$$|i,1\rangle=(1-\alpha G_0 V)|i\rangle. \tag{D19}$$

We conclude this appendix with two technical comments.

Comment 1. Perturbation theory is *a priori* not valid, if the perturbation αV results in the disappearance of any bound states from the spectrum. What we have done amounts to showing that perturbation theory can still be used to construct an $O(\alpha^n)$ approximation for the propagator, provided that the integrals occurring at $(n+1)$th order in perturbation theory converge. As already mentioned, in this paper we actually use only Eq. (D15). In four dimensions fermionic zero modes fall like $|x|^{-3}$ or faster for $|x|\to\infty$. Even if we perturb by a spatially constant mass term m (that does not vanish at infinity) Eq. (D15) gives the correct expression for the $O(1/m)$ part of the propagator.

Comment 2. In a semi-classical calculation we also have to separate out the approximate zero mode contribution to the determinant. As discussed in detail in Ref. [27], this can be done by splitting the functional integration into separate integrations over the amplitudes of the (approximate) zero modes and over the orthogonal subspace. To leading order, the integration over the zero mode subspace gives rise to $\det(\lambda_1)$ [see Eq. (D14) above]. The integration over the orthogonal subspace has a non-vanishing finite limit (after subtracting UV divergences).

[1] S. Adler, Phys. Rev. **177**, 2426 (1964); J. Bell and R. Jackiw, Nuovo Cimento A **60**, 47 (1969).

[2] K. Wilson, in *New Phenomena in Sub-Nuclear Physics*, Erice, 1975, edited by A. Zichichi (Plenum, New York, 1977).

[3] L.H. Karsten and J. Smit, Nucl. Phys. **B183**, 103 (1981).

[4] O. Piguet and S.P. Sorella, *Algebraic Renormalization: Perturbative Renormalization, Symmetries and Anomalies* (Springer-Verlag, Berlin, 1995).

[5] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976); Phys. Rev. D **14**, 3432 (1976).

[6] J.B. Kogut and L. Susskind, Phys. Rev. D **11**, 3594 (1975).

[7] S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, England, 1985).

[8] G. 't Hooft, Phys. Rep. **142**, 357 (1986).

[9] J. Zinn-Justin, hep-th/0201220.

[10] I. Jack, D.R.T. Jones, and K.L. Roberts, Z. Phys. C **62**, 161 (1994); **63**, 151 (1994).

[11] M. Lüscher, J. High Energy Phys. **06**, 028 (2000).

[12] H. Suzuki, Nucl. Phys. **B585**, 471 (2000).

[13] M. Lüscher, "Chiral gauge theories revisited," hep-th/0102028.

[14] A. Borelli, L. Maiani, G.-C. Rossi, R. Sisto, and M. Testa, Phys. Lett. B **221**, 360 (1989); Nucl. Phys. **B333**, 335 (1990); L. Maiani, G.-C. Rossi, and M. Testa, Phys. Lett. B **292**, 397 (1992).

[15] Y. Shamir, Phys. Rev. D **57**, 132 (1998); M. Golterman and Y. Shamir, Phys. Lett. B **399**, 148 (1997).

[16] W. Bock, M. Golterman, and Y. Shamir, Phys. Rev. D **58**, 054506 (1998); Phys. Rev. Lett. **80**, 3444 (1998); Phys. Rev. D **58**, 034501 (1998); **58**, 097504 (1998); W. Bock, M. Golterman, K.C. Leung, and Y. Shamir, *ibid.* **62**, 034507 (2000).

[17] W. Bock, M. Golterman, M. Ogilvie, and Y. Shamir, Phys. Rev. D **63**, 034504 (2001); Y. Shamir and M. Golterman, hep-lat/0209003.

[18] W. Bock, M. Golterman, and Y. Shamir, Nucl. Phys. B (Proc. Suppl.) **63**, 147 (1998); **63**, 581 (1998).

[19] M. Golterman, Nucl. Phys. B (Proc. Suppl.) **94**, 189 (2001).

[20] T. Banks, Phys. Lett. B **272**, 75 (1991); T. Banks and A. Dabholkar, Phys. Rev. D **46**, 4016 (1992).

[21] H.B. Nielsen and M. Ninomiya, Nucl. Phys. **B185**, 20 (1981); **B193**, 173 (1981); **B195**, 541(E) (1982).

[22] M.F.L. Golterman and D.N. Petcher, Phys. Lett. B **225**, 159 (1989).

[23] M.J. Dugan and A.V. Manohar, Phys. Lett. B **265**, 137 (1991).

[24] See the 3rd paper of Ref. [16].

[25] M. Bochicchio, L. Maiani, G. Martinelli, G.C. Rossi, and M. Testa, Nucl. Phys. **B262**, 331 (1985); C. Curci, Phys. Lett. **167B**, 425 (1986).

[26] P.D.V. Swift, Phys. Lett. **145B**, 256 (1984); J. Smit, Acta Phys. Pol. B **17**, 531 (1986).

[27] A. Casher and Y. Shamir, hep-th/9908074.

[28] E. Seiler and I.O. Stamatescu, Phys. Rev. D **25**, 2177 (1982); **26**, 534(E) (1982).

[29] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble, in *Advances in Particle Physics*, edited by R.L. Cool and R.E. Marshak (Wiley, New York, 1968), Vol. 2.

[30] W. Bock, J.E. Hetrick, and J. Smit, Nucl. Phys. **B437**, 585 (1995).

[31] J. Schwinger, Phys. Rev. **82**, 664 (1951).

[32] L. Alvarez-Gaumé, in *Fundamental Problems of Gauge Field Theory*, edited by G. Velo and A.S. Wightman, Proceedings of the 6th International School of Mathematical Physics, Erice, Italy, 1985, NATO Advanced Study Institute, Series B: Vol. 141 (Plenum, New York, 1986).

[33] See the 2nd paper of Ref. [15].