

Does chiral perturbation theory rule out QCD-based solutions to the strong CP problem?

Thomas D. Cohen

Department of Physics and the Maryland Center for Fundamental Physics University of Maryland, 20742 Maryland, USA



(Received 26 November 2018; published 7 May 2019)

The conventional view is that a solution of the strong CP problem lies beyond QCD. A strong argument supporting this view is that the chiral expansion shows that observables depend on θ (unless a quark mass is zero); this eliminates the possibility that θ is physically irrelevant and appears to necessitate an explanation beyond the standard model. However, scenarios that solve the strong CP problem that are consistent with known chiral behavior exist; in these, QCD becomes nonviable as a theory for nonzero θ . Such scenarios appear to be compatible with lattice studies of the topological susceptibility.

DOI: [10.1103/PhysRevD.99.094007](https://doi.org/10.1103/PhysRevD.99.094007)

I. INTRODUCTION

The resolution [1,2] of the axial $U(1)$ puzzle [3]—the absence of a ninth pseudo-Goldstone boson—gave rise to the strong CP problem. The solution of the $U(1)$ problem requires both the nonconservation of the axial $U(1)$ (due to the anomaly) and a nontrivial topological susceptibility (initially envisioned in terms of instantons [1,2]), which prevents the nonconservation from being rotated away. However, nontrivial topological effects also imply that a possible CP -violating term in the QCD Lagrangian (previously ignored as a total derivative) has physical consequences [4,5]. The coefficient of it is $\bar{\theta}$ (after a chiral rotation to ensure a real mass matrix), which can take any value from $-\pi$ to π . However, empirically, it is exceptionally close to zero. The exceedingly small value of the neutron electric dipole [the current bounds are $-3.2 \times 10^{-26} e\text{-cm} < d_n < 2.8 \times 10^{-26} e\text{-cm}$ (90% C.L.)] [6] led to an estimate [7] that $|\bar{\theta}| \lesssim 10^{-11}$. This is an apparent violation of Gell-Mann's famous dictum that particle physics is totalitarian in the sense that what is not forbidden is mandatory [8]. Why should a parameter not forbidden by symmetry be essentially zero? This is the strong CP problem. It does not appear to be resolvable by anthropic arguments; standard analysis suggests that the Universe would not be substantially different than even if $\bar{\theta}$ were many orders of magnitude bigger than the current bounds [9].

The strong CP problem has motivated scenarios for beyond-the-standard-model (BSM) physics (for a review,

see Ref. [7]). Indeed, new BSM solutions have been proposed in the past few years [10,11]. The most influential class of BSM proposals has been variations on the (approximate) symmetry of Peccei and Quinn [12,13], which implies the existence of a pseudo-Goldstone boson, the axion [14,15]. Such an explanation is attractive in an additional way: the axion could be the dark matter, which has long been known to exist from astrophysical evidence [16]. However, major experimental efforts to detect the axion in a variety of ways (for a review, see Ref. [17]) have been unsuccessful to date. It remains unclear if any of the proposed BSM solutions to the strong CP problem are correct. Perhaps, it is time to reexamine the possibility that the solution of the problem could lie within QCD itself. Scenarios with $m_u = 0$ are not QCD based in this context; they require a BSM explanation of why $m_u = 0$. In any case, $m_u = 0$ scenarios are ruled out by comparisons of current lattice studies to data [18–20], which yield an up quark mass that differs from zero by more than 20 standard deviations.

Of course, there are very good reasons to reject a solution arising from QCD. One of the strongest of these is the calculable and nonzero $\bar{\theta}$ dependence of physical observables in an apparently reliable expansion scheme: chiral perturbation theory (χ PT) [21–25]. Superficially, the results of χ PT by itself appear to exclude the possibility of a QCD-based solution. However, it is worth recalling that systematic expansions can mislead. Consider topological properties through the lens of a strict N_c^{-1} expansion where N_c is the number of colors [26–28]. This systematic expansion suggests that the topological susceptibility χ_T should be largely insensitive to the quark masses since the quark sector is formally down by N_c^{-1} . However, this is clearly wrong for very small quark masses. The interplay between the chiral and N_c^{-1} expansions is required to

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.

understand the correct behavior [26–28]. The question addressed in this paper is whether the χ PT-based analysis that apparently rules out QCD-based solutions to the strong CP problem could similarly yield misleading results.

Before addressing this question in detail, a word about notation is necessary. CP violation can be introduced into QCD via two nominally distinct ways—either through the quark mass terms or via an explicit θ term. However, physically, these are not distinct: due to the anomaly, chiral rotations allows one to convert one type into the other without changing the physics. The combination of parameters $\bar{\theta}$ determines the physical level of CP violation. $\bar{\theta}$ is equal to the parameter specifying the θ term, after chiral rotations have been made that render all quark masses real and non-negative. Throughout this paper, the formalism will be based on a representation of the theory that has all quark mass real and non-negative. Within this representation, there is no distinction between θ and $\bar{\theta}$. Throughout this paper, it will be denoted as $\bar{\theta}$ in order to make clear that it contains all of the CP violation in the theory.

II. CONDITIONS FOR QCD-BASED SOLUTIONS TO THE STRONG CP PROBLEM

The perspective advocated here is that, while χ PT presents severe challenges to QCD-based solutions, it appears possible to reconcile what is known reliably about the chiral expansion with a solution to the strong CP problem within QCD. However, such scenarios require that QCD behaves in a very unexpected way in order to evade the strong constraints imposed by χ PT. Moreover, unlike the breakdown of the $1/N_c$ approximation when quark masses are small, there is no theoretical motivation for why QCD should behave this way—other than the need to solve the strong CP problem. By illustrating how peculiar QCD needs to be to solve the problem by itself, this paper could simply be read as additional evidence for a BSM explanation. However, given the importance of solving the strong CP problem, it is important not to exclude possible solutions even if they require unexpected behavior.

The most natural solution within QCD would be if $\bar{\theta}$ does not affect physical observables, despite the existence of nontrivial topological sectors. Shifman *et al.* (SVZ) [21] considered and rejected this possibility long ago on the basis of χ PT; they showed that the anomaly plus the lack of a massless η' in the chiral limit implies nonzero calculable CP -violating observables at leading nontrivial order in θ and the quark masses. Examples considered by SVZ included the amplitude for $\eta \rightarrow 2\pi$ and the expectation value of the topological charge density, $q(x) \equiv g^2 F\tilde{F}(x)/32\pi^2$ at first order in $\bar{\theta}$.

For the present purpose, it is more efficient to apply the logic of Ref. [21] to the topological susceptibility χ_T , and higher-order cumulants of the topological charge defined by

$$\begin{aligned} c^{(2k)} &\equiv \lim_{V \rightarrow \infty} \frac{(-1)^{k+1} \langle Q^{2k} \rangle_c}{V} \\ &= \lim_{V \rightarrow \infty} \frac{(-1)^{k+1} \langle Q^{2k} \rangle - k(2k-1) \langle Q^{2k-2} \rangle \langle Q^2 \rangle + \dots}{V} \end{aligned} \quad (1)$$

[where the brackets indicate the vacuum expectation value evaluated at $\bar{\theta} = 0$, $Q \equiv \int d^4x q(x)$, V is the volume of (Euclidean) space-time, k are positive integers, and the subscript c indicates the connected part.] are nonzero and calculable, and $\chi_T \equiv c^{(2)}$. Odd susceptibilities vanish due to CP . To leading order in χ PT, they are given by

$$\begin{aligned} c^{(2k)} &= \frac{-d^{2k} (m_\pi^2 f_\pi^2 \cos(\frac{\bar{\theta}}{2}) \sqrt{1 + \delta^2 \tan^2(\frac{\bar{\theta}}{2})})}{d\bar{\theta}^{2k}} \Big|_{\bar{\theta}=0} \\ &+ \mathcal{O}(m_\pi^4) \quad \text{with} \quad \delta \equiv \frac{m_d - m_u}{m_d + m_u}, \end{aligned} \quad (2)$$

where f_π is the pion decay constant. As stressed by SVZ, Eq. (2) relies only on the anomaly, a standard chiral expansion which, for simplicity, only takes m_u and m_d as being small—to keep expressions concise, QCD with two light flavors will be discussed here—and the absence of a massless η' in the chiral limit. All expressions are in Euclidean space.

This paper assumes that χ PT accurately reproduces all of the topological cumulants. More precisely, for each $c^{(2k)}$, at any given nontrivial order in χ PT and with any fixed relative error tolerance, ϵ_R , there exists a mass, M , such that whenever $m_\pi^2 < M^2$ the χ PT value is within ϵ_R of the actual value. This assumption is based on more than a natural prejudice that the chiral expansion, which accurately describes much of low energy QCD [29], should also work for the $c^{(2k)}$. It is also based on the fact that lattice studies have provided numerical evidence that χ PT at low order accurately describes χ_T (for state-of-the-art results, see Refs. [30,31]). A key challenge of this paper is to reconcile this assumption with possible loopholes in reasoning based on χ PT that exclude QCD-based solutions to the strong CP problem.

The challenge becomes clear from standard functional arguments which imply [25] that the $c^{(2k)}$ can be written as

$$\begin{aligned} c^{(2k)} &= \frac{d^{2k} \epsilon(\bar{\theta})}{d\bar{\theta}^{2k}} \Big|_{\bar{\theta}=0} \quad \text{with} \\ \epsilon(\bar{\theta}) &\equiv - \lim_{V \rightarrow \infty} \frac{\log(\sum_Q Z_Q(V) e^{iQ\bar{\theta}})}{V}, \end{aligned} \quad (3)$$

where $\epsilon(\bar{\theta})$ is the energy density of the vacuum as a function of $\bar{\theta}$, V is the Euclidean space-time volume, and $Z_Q(V)$ is the partition function for a fixed topological sector. Together with Eq. (2), this implies that the energy density is given by

$$\epsilon(\bar{\theta}) - \epsilon(0) = m_\pi^2 f_\pi^2 \left(1 - \cos\left(\frac{\bar{\theta}}{2}\right) \sqrt{1 + \delta^2 \tan^2\left(\frac{\bar{\theta}}{2}\right)} \right) + \mathcal{O}(m_\pi^4). \quad (4)$$

Clearly, the topological charge density, $q = \epsilon'(\bar{\theta})$, a physically relevant observable, is nonzero and $\bar{\theta}$ dependent. Moreover, it seems inconceivable that higher-order terms in the chiral expansion could act to exactly cancel out the dependence from the lowest-order term. This provides a compelling reason to exclude QCD-based solutions to the strong CP problem that rely on all physical observables being independent of $\bar{\theta}$.

However, there is a loophole in this reasoning: QCD itself might solve that strong CP problem without requiring physical observables to be independent of $\bar{\theta}$. This would happen if some aspect of QCD rendered the theory nonviable unless $\bar{\theta}(\text{mod } 2\pi) = 0$, even though $\bar{\theta}$ *apparently* affects physical observables. Scenarios of this kind are analogous to triviality arguments for ϕ^4 theory in $3 + 1$ dimensions where the value of the interaction term, which clearly would have a physical effect on the dynamics and is allowable by perturbative power counting, is required to be zero since otherwise the theory would not be ultraviolet complete. (The renormalization group using a perturbative kernel is suggestive of triviality; lattice studies are consistent with this expectation [32,33]. For a state-of-the-art calculation, see Ref. [34].) In a solution based on nonviability of QCD away from $\bar{\theta}(\text{mod } 2\pi) = 0$, some currently unrecognized feature of the theory (analogous to the need for ultraviolet completeness) requires $\bar{\theta}$ to be zero. There were proposals along this general line in the past, e.g., Refs. [35,36]. Scenarios such as these are highly speculative and hard to either verify or exclude. In any case, reconciling them with the SVZ-type analysis—which yields physical results at $\bar{\theta} \neq 0$ based on apparently reliable assumptions—appears to be problematic.

Scenarios with QCD nonviable for $\bar{\theta}(\text{mod } 2\pi) \neq 0$ have the function $\epsilon(\bar{\theta})$ ill defined when $\bar{\theta}(\text{mod } 2\pi) \neq 0$. χ PT at any finite order yields an explicit and well-defined function for $\epsilon(\bar{\theta})$, presenting a challenge to such scenarios. Moreover, an assumption underlying this paper, that the $c^{(2k)}$ —even derivatives of $\epsilon(\bar{\theta})$ with respect to $\bar{\theta}$ (evaluated at $\bar{\theta} = 0$)—are nonzero and accurately given in χ PT makes this challenge acute. Superficially, this seems to be an insurmountable challenge to such scenarios; how can a function that only exists at isolated points have well-defined derivatives? Lattice calculations yield a nonzero susceptibility and appear to confirm χ PT. However, despite this challenge, scenarios of this type need not be ruled out due to χ PT.

This can happen, in principle, if $\epsilon(\bar{\theta})$, while undefined for real $\bar{\theta}$ away from $\bar{\theta}(\text{mod } 2\pi) = 0$, is nevertheless well defined for imaginary $\bar{\theta}$. Of course, physically, $\bar{\theta}$ must be

real due to unitarity. However, mathematically additional information about $\epsilon(\bar{\theta})$ can be obtained by extending Eq. (3) to complex $\bar{\theta}$ [37]. Extending the reasoning that gave in Eq. (3) into the complex plane yields $c^{2k} = (-1)^k \frac{d^{2k} \epsilon(i\bar{\theta})}{d(\bar{\theta})^{2k}} \Big|_{\bar{\theta}=0}$. Moreover, there is a caveat to Eq. (3): it holds only if the sum on Q and infinite-volume limit in the definition of $\epsilon(\bar{\theta})$ converge to a finite value for real $\bar{\theta}$ in some finite neighborhood around $\bar{\theta} = 0$. Suppose that it does not so that $\epsilon(\bar{\theta})$ is ill defined for real $\bar{\theta}$ with $\bar{\theta}(\text{mod } 2\pi) \neq 0$ but remains well defined for imaginary $\bar{\theta}$. In such cases, it becomes possible that χ PT could accurately describe $c^{(2k)} = (-1)^k \frac{d^{2k} \epsilon(i\bar{\theta})}{d(\bar{\theta})^{2k}} \Big|_{\bar{\theta}=0}$. The significant point is that the existence of nonzero topological susceptibilities need not require that $\epsilon(\bar{\theta})$ is well defined on the real axis away from $\bar{\theta}(\text{mod } 2\pi) = 0$; it is sufficient for it to be well defined on the imaginary axis. In the remainder of this paper, it will be assumed that this occurs. Such scenarios depend on an obstruction to analytically continuing the functional form of the energy density from the imaginary $\bar{\theta}$ to the real axis, and this implies that the point $\bar{\theta}$ is nonanalytic.

The challenges posed by χ PT to a solution to the strong CP problem within QCD can be met, provided $\epsilon(\bar{\theta})$ and the $c^{(2k)}$ satisfy the following conditions:

- (i) $\epsilon(\bar{\theta})$ is finite and real for $\bar{\theta}$ real with $\bar{\theta}(\text{mod } 2\pi) = 0$ and for $\bar{\theta}$ purely imaginary.
- (ii) $\epsilon(\bar{\theta})$ is ill defined, corresponding to a nonviable physical theory when $\bar{\theta}(\text{mod } 2\pi) \neq 0$ with $\bar{\theta}$ real.
- (iii) $\epsilon(\bar{\theta})$ is an analytic function for $\bar{\theta} \neq 0$ with $\bar{\theta}$ purely imaginary.
- (iv) $\epsilon(i\bar{\theta})$ is not analytic at $\bar{\theta} = 0$.
- (v) All derivatives of $\epsilon(i\bar{\theta})$ with respect $\bar{\theta}$ are finite and well defined at $\bar{\theta} = 0$, despite the nonanalyticity at zero of condition (iv).
- (vi) The topological cumulants are well defined, finite, and given by $c^{(2k)} = (-1)^k \frac{d^{2k} \epsilon(i\bar{\theta})}{d(\bar{\theta})^{2k}} \Big|_{\bar{\theta}=0}$.
- (vii) For any $c^{(2k)}$ and any given nontrivial order in χ PT with any fixed relative error tolerance, ϵ_R , there exists a mass, M , such that whenever $m_\pi^2 < M^2$, χ PT accurately reproduces $c^{(2k)}$ to within ϵ_R .
- (viii) The chiral expansion for $\epsilon(\bar{\theta})$ when $\bar{\theta}$ is purely imaginary is asymptotic; for any fixed order, any imaginary $\bar{\theta}$ and any fixed relative error tolerance, ϵ_R , there exists an M such that whenever $m_\pi^2 < M^2$, χ PT accurately reproduces $\epsilon(\bar{\theta})$ to within ϵ_R .
- (ix) The nonanalyticity of $\epsilon(i\bar{\theta})$ at $\bar{\theta} = 0$ of condition (iv) is due to contributions that are subleading to all orders in χ PT.

Conditions (i) and (iii) encode the expected behavior when $\bar{\theta}$ is imaginary that leads to well-defined topological susceptibilities, while condition (ii) is the heart of scenarios that make QCD nonviable for real $\bar{\theta}$ away from $\bar{\theta}(\text{mod } 2\pi) = 0$.

Conditions (vi) and (vii) place severe constraints on such scenarios due to χ PT. Conditions (iv) and (viii) play essential roles in evading such constraints: condition (iv) reconciles conditions (ii) and (iii), while condition (ix) reconciles conditions (ii), (iv), (iii), and (viii).

III. ILLUSTRATIVE EXAMPLE

Consider the following illustrative functional form, which illustrates the type of function that satisfies all nine conditions,

$$\begin{aligned}
 e^{\text{illus}}(\bar{\theta}) - \epsilon(0) & \quad (5a) \\
 &= m_\pi^2 f_\pi^2 g(\bar{\theta}, \delta) \int_0^\infty dx e^{-x+x^2\left(\frac{m_\pi^2 f_\pi^2}{\Lambda^4}\right)g(\bar{\theta}, \delta)} \\
 &= \sum_{j=0}^{\infty} \frac{(2j)!}{j!} \Lambda^4 \left(\frac{m_\pi^2 f_\pi^2}{\Lambda^4}\right)^{j+1} g(\bar{\theta}, \delta)^{j+1} \\
 &= \Lambda^4 \left(\left(\frac{m_\pi^2 f_\pi^2 g(\bar{\theta}, \delta)}{\Lambda^4}\right) + 2 \left(\frac{m_\pi^2 f_\pi^2 g(\bar{\theta}, \delta)}{\Lambda^4}\right)^2 + \right. \\
 &\quad \left. 12 \left(\frac{m_\pi^2 f_\pi^2 g(\bar{\theta}, \delta)}{\Lambda^4}\right)^3 + 120 \left(\frac{m_\pi^2 f_\pi^2 g(\bar{\theta}, \delta)}{\Lambda^4}\right)^4 + \dots \right) \quad (5b)
 \end{aligned}$$

$$\begin{aligned}
 \text{with } g(\bar{\theta}, \delta) &= 1 - \cos\left(\frac{\bar{\theta}}{2}\right) \sqrt{1 + \delta^2 \tan^2\left(\frac{\bar{\theta}}{2}\right)} \\
 &= \frac{(1 - \delta^2)\bar{\theta}^2}{8} - \frac{(1 + 2\delta^2 - 3\delta^4)\bar{\theta}^4}{384} + \dots \quad (5c)
 \end{aligned}$$

where Λ is a parameter with dimensions of mass.

Equation (5b), the formal chiral expansion of e^{illus} , is asymptotic and valid when $\bar{\theta}$ is imaginary; at lowest order, it yields Eq. (4). The form of $g(\bar{\theta}, \delta)$ in Eq. (5c) accounts for isospin violation and ensures that all $\bar{\theta}$ dependence vanishes when m_u or m_d is zero. Note that $g(\bar{\theta}, \delta)$ has an important property: it is greater than zero when $\bar{\theta}$ is real and away from $\bar{\theta}(\text{mod } 2\pi) = 0$, equal zero when $\bar{\theta}(\text{mod } 2\pi) = 0$, and less than zero when $\bar{\theta}$ is imaginary. From the integral form of Eq. (5), this implies that energy density is divergent for real $\bar{\theta}(\text{mod } 2\pi) \neq 0$ but convergent for imaginary $\bar{\theta}$; thus, conditions (i) and (ii) hold. One might hope that, despite the nonanalytic behavior at $\bar{\theta} = 0$, one could analytically continue the e^{illus} from the imaginary axis to real $\bar{\theta}$. One can; but the function so obtained is multibranched, and all branches have an imaginary part—indicating that they are not physically viable.

It is straightforward to verify that conditions (i)–(ix) hold for e^{illus} . The numerical coefficients in the chiral expansion of Eq. (5b) explain why: the coefficient of the j th term is $\frac{(2j)!}{j!}$, which grows faster than any power law in j . Thus, the

radius of convergence for the chiral expansion is strictly zero. Moreover, the series is actually in $m_\pi^2 g(\bar{\theta}, \delta)$ where $g(\bar{\theta}, \delta)$ in a series in $\bar{\theta}$ begins at order $\bar{\theta}^2$. Thus, the radius of convergence of an expansion in $\bar{\theta}$ is also zero, and the point $\bar{\theta} = 0$ is nonanalytic. The rapid growth in these coefficients implies that, while χ PT at low order can accurately describe the $c^{(2k)}$ (as required by condition), the value of m_π for which χ PT at fixed order is accurate rapidly drops with k .

If $\epsilon(\bar{\theta})$ in QCD were given by $e^{\text{illus}}(\bar{\theta})$, the strong CP problem would be solved. Of course, in QCD, $\epsilon(\bar{\theta}) \neq e^{\text{illus}}(\bar{\theta})$, which was given simply to illustrate that functions satisfying all the conditions exist. $e^{\text{illus}}(\bar{\theta})$ is consistent with a chiral Lagrangian given to all orders in m_π^2 but treated at tree level. In practice, higher-order terms in χ PT will develop chiral logarithms due to infrared behavior in loops along with the powers of m_π^2 associated with tree-level terms [22–25]. However, these chiral logs are entirely fixed by lower-order terms in the theory and should not affect whether or not the conditions are satisfied. The critical issue for this is whether the coefficients of higher-order terms in the chiral expansion of $\epsilon(\bar{\theta})$ grow sufficiently rapidly; at this stage, we do not know whether or not they do in QCD.

This paper has focused on $\epsilon(\bar{\theta})$. However, it should be clear that, if scenarios of this sort were valid, other observables such as the $\eta \rightarrow 2\pi$ amplitude considered by SVZ [21] would be expected to have analogous behavior. The matrix elements for these observables would be finite and describable in χ PT for imaginary $\bar{\theta}$ but would become ill defined for real $\bar{\theta}(\text{mod } 2\pi) \neq 0$, and the function would be nonanalytic at $\bar{\theta} = 0$ with the nonanalyticity subleading to all orders in χ PT.

IV. DISCUSSION

The central argument of this paper is that there is a loophole in the logic that χ PT excludes QCD-based solutions to the strong CP problem. Of course, there remain very strong reasons to doubt that QCD itself can be the solution of the strong CP problem. One of these is simply that the loophole requires QCD to behave in a very surprising manner that differs from the familiar ways we expect quantum field theories to behave. Such behavior is radically differently from any quantum field theory that the community has seen over the decades. A second reason is that the motivation for considering the conditions proposed here is entirely phenomenological; they were proposed solely to ensure that $\bar{\theta}(\text{mod } 2\pi) \neq 0$ without violating established properties of QCD. Accordingly, there is no underlying theoretical basis for expecting QCD to satisfy them. The illustrative model of Sec. III illuminates this. While the illustrative model satisfies conditions (i)–(ix) and thereby demonstrates that forms that do so are not excluded mathematically, it is also contrived. The model did not emerge as a natural outgrowth of any theoretically motivated

mechanism; rather, it was essentially reverse engineered for the sole purpose of satisfying the conditions. The upshot of this is that it is hard to see why a functional form satisfying the conditions should emerge from QCD.

On the other hand, there are no easy solutions to the strong CP problem. Solutions require one to conjecture entire new sectors of BSM physics—a major intellectual leap—or to conjecture that the standard model behaves in a very unfamiliar way—which is also a major intellectual leap. Thus, it seems sensible to explore the possibility of all possible solutions.

It is important to stress the way in which this class of scenarios solves the strong CP problem. In these scenarios, the theory is only viable as a physical theory for values of $\bar{\theta}$ satisfying $\cos(\bar{\theta}) = 1$. At first glance, this might appear to be inconsistent with $\epsilon(\bar{\theta})$ being well defined and analytic when $\bar{\theta}$ is purely imaginary. Indeed, one might worry that the existence of a well-defined function on the imaginary axis means that the theory allows CP -violating physics, provided that one takes $\bar{\theta}$ to be imaginary. However, such concerns are misplaced. It is important to distinguish between the theory as a mathematical object and one describing physics. Physically, the theory is only sensible when $\bar{\theta}$ is real. If it is not, the theory is not unitary and is not an acceptable description of nature. On the other hand, one can define the path integral for the theory mathematically, regardless of the phase of $\bar{\theta}$ (assuming that it converges appropriately) even when this renders the theory unphysical. Thus, if this class of scenario is correct, then the strong CP problem is solved: the only physically viable value of $\bar{\theta}$ in the theory has $\cos(\bar{\theta}) = 1$ and no CP violation. This remains true despite the fact that theory is mathematically sensible for unphysical imaginary values of $\bar{\theta}$.

Clearly solutions to the strong CP problem that exploit the loophole noted in this work are very different from axion solutions. In axion models, $\bar{\theta}$ effectively becomes dynamical; its value is free to respond to the environment. The strong CP problem is solved since the effective potential for $\bar{\theta}$ has a minimum at $\cos(\bar{\theta}) = 1$; if a region of space had $\cos(\bar{\theta}) \neq 1$, it would not be in a stable equilibrium and would slide toward the minimum. In contrast, in solutions of the sort considered here, $\bar{\theta}$ does not adjust itself to a value where CP violation vanishes. Rather, $\bar{\theta}$ is not dynamical but a fixed value. The solution requires that for reasons currently not understood—but also not completely ruled out, given our current state of knowledge—QCD does not exist as a viable physical theory unless $\cos(\bar{\theta}) = 1$; thus, if the standard model contains QCD, it must have $\bar{\theta}(\text{mod } 2\pi) = 0$.

Ideally, one should be able to verify or exclude this class of solution—or at least develop evidence in favor of or opposed to it. A key problem with trying to rule out scenarios of the sort considered here is their phenomenological (as opposed to theoretical) motivation. This lack of an underlying theoretical picture means that one cannot rule

out this class of scenario by undermining the underlying theoretical assumptions.

Nevertheless, one obvious future research direction is to see whether there are reliable theoretical arguments that can close the loophole considered in this paper. In essence, that would mean a “no-go” theorem that would rule out this class of solutions to the strong CP problem. In this context, the interplay between perturbative and nonperturbative aspects of a chiral expansion suggest analysis along using ideas of resurgence (for a review of the state of the art, see Refs. [38,39]) might shed light on the issue.

If the loophole cannot be closed using reliable theoretical arguments, other ways of obtaining evidence in support of or against solutions of this type should be considered. Clearly, the most straightforward way to exclude these scenarios would be via a direct and reliable calculation of $\epsilon(\bar{\theta})$. If one could show that the $\epsilon(\bar{\theta})$ is well defined and calculable in QCD away from $\cos(\bar{\theta}) = 1$, the scenario is ruled out. Unfortunately, the only known systematic reliable numerical method for computing nonperturbative observables directly from QCD is via Euclidean-space lattice calculations, and direct lattice calculations of $\epsilon(\bar{\theta})$ are well beyond the state of the art due to a sign problem [37]. Accordingly, it is necessary to consider what kind of indirect evidence one can obtain.

As it happens, if such scenarios were correct, Monte Carlo calculations in Euclidean space without exponentially bad sign problems could, in principle, provide compelling, if indirect, evidence for them. However, there are severe practical limitations to such calculations for QCD. These practical limitations will almost certainly make any such evidence obtainable in the foreseeable future quite indirect. Still, even indirect evidence for such scenarios would be very significant.

The basic reason why lattice calculations could, in principle, provide indirect support for such scenarios comes from conditions (i)–(vi), which do not depend in detail on chiral properties. These provide a distinctive signature. Consider the function $\epsilon(\bar{\theta})$ along the imaginary $\bar{\theta}$ axis where it is well defined in the scenarios considered here. The conditions imply that at $\bar{\theta} = 0$ all derivatives of $\epsilon(i\bar{\theta})$ are finite and well defined but the function is nonanalytic. Thus, if one were to write $\epsilon(i\bar{\theta})$ as a Taylor expansion around some point in the complex plane, its radius of convergence would approach zero as that point approaches $\bar{\theta} = 0$.

Given that $\epsilon(\bar{\theta})$ is an even function and that up to a k -dependent sign the topological cumulants, $c^{(2k)}$, are given by the $(2k)$ th derivative of $\epsilon(i\bar{\theta})$ with respect to $\bar{\theta} = 0$, it follows from the ratio test that R , the radius of convergence of a Taylor expansion for $\epsilon(\bar{\theta})$, in general is given by

$$R = \lim_{n \rightarrow \infty} \sqrt{\frac{|c^{(2n)}|}{(2n)!}} \bigg/ \sqrt{\frac{|c^{(2n+2)}|}{(2n+2)!}}, \quad (6)$$

provided the limit exists. This means that if

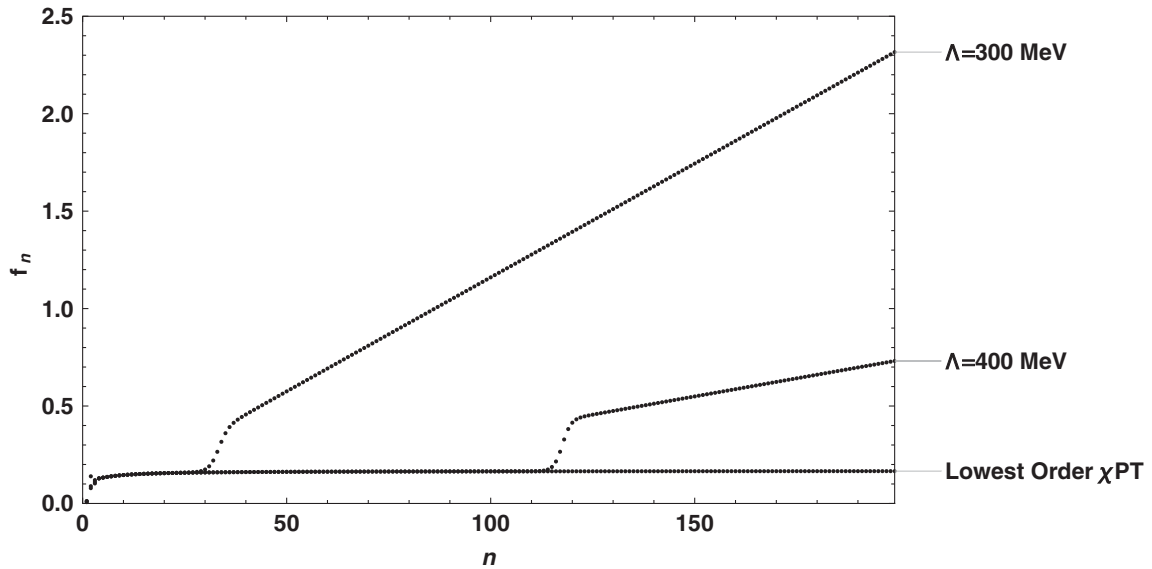


FIG. 1. $f_n \equiv \frac{|c^{(2n+2)}|}{(n+1)(n+2)|c^{(2n)}|}$ for $\epsilon^{\text{illus}}(\bar{\theta})$, the illustrative model of Sec. III. The $c^{(2n)}$ are topological cumulants. Two different values of the parameter Λ are shown. For comparison, the same ratio is also given for lowest-order chiral perturbation theory. If this ratio diverges as $n \rightarrow \infty$, then $\epsilon(i\bar{\theta})$ is nonanalytic at $\bar{\theta} = 0$.

$$\lim_{n \rightarrow \infty} f_n \rightarrow \infty \quad \text{where} \quad f_n \equiv \frac{|c^{(2n+2)}|}{(n+1)(n+2)|c^{(2n)}|} \quad (7)$$

the radius of convergence is zero, a necessary condition for this class of scenario to be valid. Provided that such behavior could be documented for QCD, one would have strong evidence in favor of such scenarios.

Since topological cumulants are calculable without a sign problem (either by directly computing fluctuations in the topological charge or by computing numerical derivatives of $\epsilon(i\bar{\theta})$), one could look for indications that in QCD f_n grows without bound as $n \rightarrow \infty$ (as is required by scenarios of the sort considered here). Such studies are necessarily indirect: lattice calculations can only provide information about a finite number of topological cumulants. Thus, one cannot determine whether f_n is actually diverging or not; the best one can do is see whether it *appears* to be growing in a manner consistent with divergence. Still, a convincing indication of this sort might be regarded as compelling evidence for such scenarios.

To see how in principle this could work, suppose, for the sake of illustration, that $\epsilon(i\bar{\theta})$ in QCD is given exactly by $\epsilon^{\text{illus}}(i\bar{\theta})$, the illustrative model of Sec. III, and that lattice calculations determining numerous topological cumulants with negligible error were tractable. In Fig. 1, the ratio f_n defined in Eq. (7) is plotted as a function of n for this model with two different values of the parameter Λ , 300 and 400 MeV; for comparison, this ratio is also given for lowest-order chiral perturbation theory. The ratio in lowest-order chiral perturbation theory saturates at a finite value, as one expects when $\epsilon(i\bar{\theta})$ has a nonzero radius of convergence. In contrast, for the toy model with either $\Lambda = 300$ MeV or $\Lambda = 400$ MeV, the ratio quite clearly appears

to be growing linearly with n asymptotically. To the extent that the behavior is indicative of the true asymptotic behavior (as is the case for the toy model), one can conclude that the Taylor series for $\epsilon(i\bar{\theta})$ around zero has a zero radius of convergence as required by the scenarios considered here. The behavior seen in Fig. 1 is quite dramatic. If one could obtain lattice data of this sort for QCD and find that it had this behavior, one would have compelling, if somewhat indirect, evidence for a scenario of the sort considered here.

Unfortunately, Fig. 1 also indicates practical problems in implementing such an approach that render it intractable for QCD. Note that for low n the data for f_n do not even hint that it will ultimately grow linearly with n for asymptotically large n . Rather, up to a fairly large value of n , the ratio appears to be saturating to a finite value, as one would expect with an analytic function around $\bar{\theta} = 0$ —falsely suggesting that this functional form does not satisfy the conditions. For the model with $\Lambda = 300$ MeV, the first noticeable hint that the ratio is not saturating is around $n = 30$ (requiring a calculation of topological cumulants up to $c^{(62)}$); for the model with $\Lambda = 400$ MeV, the first noticeable hint that it is not saturating is around $n = 110$ (requiring a calculation of topological cumulants up to $c^{(222)}$).

The sharp onset at some large value of n of behavior inconsistent with saturation makes verifying such a scenario for QCD highly problematic, even assuming QCD behaved according to this class of scenario. Suppose, hypothetically, this class of scenario were correct for QCD and moreover algorithms for the determination of topological cumulants of QCD advanced to the point where they could be calculated with high accuracy up to $c^{(40)}$

(which corresponds to $n = 19$). Even with this large number of topological cumulants, it could easily still be the case that one would have no numerical evidence for the scenario. Given the toy model, it is not implausible that one might need accurate cumulants up through $c^{(100)}$ (i.e., $n = 49$) or some larger value before there was compelling evidence for the scenario.

In any case, a reliable and accurate calculation of topological cumulants through $c^{(40)}$ will almost certainly be well beyond our reach for QCD for the foreseeable future—unless some radically improved algorithm is found. Although topological cumulants for fixed n are calculable on a Euclidean lattice studies without exponential sign problems, with all current approaches, the difficulty in computing them grows with the order of the cumulant. If one computes the $c^{(2n)}$ via fluctuations in the topological charge, the signal to noise worsens as n increases. If one attempts to extract them as numerical derivatives of $\epsilon(i\bar{\theta})$, one is faced with the need to compute $\epsilon(i\bar{\theta})$ with increasing accuracy to accurately get higher derivatives; this is particularly difficult, given that the function is being evaluated numerically via Monte Carlo methods. Even in, even for the computationally much simpler case of Yang-Mills theory, only a few topological cumulants have been computed or even bounded (see, e.g., Ref. [40]).

In Fig. 1, f_n appears to be saturating with n until at some comparatively large there is a sharp onset of behavior inconsistent with saturation. This qualitative behavior is easy to understand. Condition (ix) implies that the non-analyticity of $\epsilon(i\bar{\theta})$ at $\bar{\theta} = 0$ is due to contributions that are subleading to all orders in χ PT. This comes about because there are contributions to $\frac{c^{2n}}{(2n)!}$ that simultaneously diminish with n due to a chiral suppression that scales like $(\frac{m_\pi}{\Lambda'})^{2n}$ (where Λ' is a typical hadronic scale) and, due to a numerical factor, grows with n faster than exponentially. This behavior can easily be seen to occur in e^{illus} , the toy model of Sec. III. For small values of n , the chiral suppression overwhelms the numerical coefficient, and these contributions are swamped by contributions that are leading order in a chiral expansion. However, the numerical coefficients grow faster with n than the diminution due to chiral suppression. Thus, at some value of n , this contribution ceases to be negligible. Moreover, the rapid growth of the numerical contribution implies that, once the term ceases to be negligible, it rapidly becomes dominant. Thus, one expects the sort of rapid onset of behavior incompatible with saturation seen in Fig. 1.

The value of n where this change of behavior sets in depends on the scale of the chiral suppression. In the toy model, this is fixed by the dimensionless combination $\frac{m_\pi^2 f_\pi^2}{\Lambda^4}$. Thus, larger values of Λ in the model will lead to more chiral suppression and accordingly the onset of growth incompatible with saturation occurring at larger n . This is

seen in Fig. 1, where for $\Lambda = 400$ MeV this onset is pushed out to the neighborhood of $n = 110$, whereas it occurs around $n = 30$ for $\Lambda = 300$ MeV.

One might worry that the prospects for directly detecting this sort of behavior in QCD may be even more dire than suggested by Fig. 1. The two models used to illustrate the issue had $\Lambda = 300$ MeV and $\Lambda = 400$ MeV, which might be regarded as quite low scales. The parameter Λ controls the scale of the chiral suppression, and one might assume that it should be taken to be naturally at a typical hadronic scale of order 1 GeV; values Λ at that scale would push the onset of behavior incompatible with saturation in this model to extremely large values of n .

On the other hand, it is hard to estimate the natural scale for Λ . The size of the chiral suppression in the model is fixed by the combination $(\frac{m_\pi^2 f_\pi^2}{\Lambda^4})$ of which the form was picked in order to have the $m_\pi^2 f_\pi^2$ structure match with the leading-order χ PT result. However, f_π^2 is numerically quite small on the scale of hadronic physics. One could just as well have written this as $\frac{m^2}{\Lambda'^2}$ with $\Lambda' = \frac{\Lambda^2}{f_\pi}$. The models with $\Lambda = 300$ MeV correspond to $\Lambda' = 958$ MeV, which may be regarded as naturally sized, while $\Lambda = 400$ MeV corresponds to $\Lambda' = 1720$ MeV, which may be regarded as large. The numerical value of the parameter one uses to parametrize the chiral suppression is a matter of bookkeeping as well as physics. Regardless of how natural the model parameters are, the model strongly suggests that, even if the scenarios considered in this paper were correct, it seems extraordinarily unlikely that this would be revealed by lattice studies of the topological cumulants in the foreseeable future.

Fortunately, one need not completely rule out the possibility of indirect evidence in support of such scenarios. For example, it is conceivable that volume dependence could be used to obtain evidence for or against the type of QCD-based solution to the strong CP problem considered in this paper. It has long been known [41,42] that interplay between topology, the chiral limit, and the infinite-volume limit is subtle. However, by exploiting known behaviors near the chiral limit, one can make concrete predictions for topological behavior, taking into account finite-volume effects, e.g., in the so-called ϵ -regime. Unfortunately, one cannot just borrow these results here since the essence of the current scenarios is that the leading-order chiral effects yield misleading results for the θ dependence. Nevertheless, one might imagine that if the full interplay of topological, finite volume, and chiral effects were understood in the context of this class of scenario, then finite-volume studies might have distinctive behavior which might act as a signature. If things were particularly fortuitous, it is possible that such a volume dependent behavior signature might turn out to be far more practical than directly computing numerous topological cumulants.

However, since this class of scenario was proposed for essentially phenomenological reasons and lacks a

theoretically motivated detailed mechanism, it is not currently possible to deduce from first principles how finite-volume effects will alter the delicate interplay of chiral effects with the analyticity of behavior of $\epsilon(\bar{\theta})$ near $\theta = 0$. Thus, one cannot easily anticipate whether finite-volume effects might prove to be a useful tool in discerning whether QCD obeys such a scenario.

If the solution to the strong CP problem is due to a scenario of this type, there is another way to find evidence in its favor. Recall that a principal reason that the scenarios considered here may seem implausible *a priori* is that, after decades of experience with quantum field theory, the community has never encountered any theories that behave in such a manner. Thus, a more general strategy would be to concentrate on quantum field theories that, while distinct from QCD, are more or less related and more tractable. If one could see compelling evidence for behavior analogous to conditions (i)–(ix) in such a theory, that would greatly increase the plausibility of the scenario. Indeed, evidence for behavior analogous conditions (i)–(vi) would be enough to substantially increase the plausibility; one can view conditions (i)–(ix) as the critical underlying ones, while conditions (vii)–(ix) could emerge naturally in QCD if the first six held, in order to reconcile chiral physics with the behavior of implied by (i)–(vi).

The behavior exhibited in Fig. 1 suggests one obvious line of attack. Evidence that the ratio f_n is not saturating to a finite value (as expected from models satisfying the first six conditions) requires far smaller values of n for the model with $\Lambda = 300$ MeV than the model with $\Lambda = 400$ MeV. The difference between the two models is that the chiral suppression is more pronounced for the $\Lambda = 400$ MeV case. Indeed, conditions (vi)–(ix) imply that as the chiral suppression grows the value of n needed before the onset of nonsaturating behavior does as well. Similarly, as the chiral suppression shrinks, the value of n needed before the onset of nonsaturating behavior does as well. In QCD, one can make the size of chiral suppression small, simply by increasing the value of the quark masses.

Optimistically, one might hope that if QCD behaves according to a scenario of this type, tractable calculations using sufficiently large quark masses might be possible in the foreseeable future that suggest a growth of f_n that is inconsistent with saturation.

There are a number of caveats to this. Clearly, if such behavior is seen with large quark masses, it will not fully establish the scenario. If chiral suppression plays no role in the observation, the best one can test are conditions (i)–(vi) and not conditions (vi)–(ix), which depend on chiral symmetry. Moreover, it is at least theoretically conceivable that scenarios of this sort only apply for a certain domain of quark masses and that this domain includes large quark masses but not realistic ones. This is a minor concern. If, contrary to all previous experience with quantum field theory, QCD with large quark mass satisfies conditions

(i)–(vi) (providing a natural explanation for the strong CP problem), it becomes extremely plausible, given the empirical fact that CP violation in strong interactions is nonexistent or extremely small, that the same behavior would extend down to light quark masses and thereby require conditions (vi)–(ix).

A more significant concern given the limited number of topological cumulants that are likely to be accessible is the possibility of a “false positive” in which the ratio f_n is seen to be growing with n for a few calculable small values of n in a manner suggesting that the ratio may not saturate, when in fact it does but does so at slightly large values of n . There is also the possibility of “false negatives.” One situation that could arise is that the behavior near the chiral limit plays a central role in realizing such scenarios and that they only are realized in QCD for a domain in which the quark masses are sufficiently small. If this were the case, then calculations at large quark masses would miss the effect.

There is also a practical concern. One is likely to be able to compute only a small number of topological cumulants. It is quite possible that for those n which are calculable the f_n appears to be saturating with n but ultimately there will be the onset of behavior where it increases without bound, but this onset is in a regime beyond where the $c^{(2n)}$ are calculable—even if the quark masses are large.

If one is attempting to render a study tractable by increasing the quark masses, it is reasonable to consider the extreme case where they go to infinity, leaving a pure gauge theory. Doing this not only eliminates chiral suppression altogether, leading to the prospect of seeing signatures of the scenario at smaller values of n , but it also has the obvious practical advantage that calculations lacking a quark functional determinant are far more straightforward numerically and hence one can compute $c^{(2n)}$ coefficients to larger values of n . Of course, removing quarks entirely from the problem makes the theory that much further from QCD and thus that much harder to draw definitive conclusions. Still, it would go a long way towards establishing a QCD-based solution to the strong CP problem if nontrivial evidence were found suggesting that Yang-Mills theory satisfied with conditions (i)–(vi); the fact that such conditions were satisfied for a theory related to QCD combined with the phenomenological fact that CP violation in QCD is either nonexistent or very small would make such a scenario plausible.

While Yang-Mills theory has no chiral suppression, numerical challenges may emerge due to suppression in $1/N_c$ (where N_c is the number of colors in the theory)—but fortunately these are much less severe than those due to chiral suppression. Witten showed long ago [26] that standard N_c , counting rules imply that as the large N_c limit is approached, the topological cumulants scale as

$$c^{(2n)} \sim N_c^{2-2n}. \quad (8)$$

Thus, as n increases, the $c^{(2n)}$ decrease parametrically quite quickly with N_c . For example, for $n = 5$, this suppression factor is N_c^{-8} , which for $N_c = 3$ is $3^{-8} \approx .0001524$. The principal difficulty this poses is that small numbers are often hard to compute accurately via numerical means and are particularly difficult to extract accurately via Monte Carlo methods. Thus, it may be hard to compute c_n 's up to even moderately large n , and those that are computed are apt to be noisy. This is likely to limit range in n that one can explore. Moreover, the noise in the c_n will translate to noise in the f_n . Note that the scaling in Eq. (8) implies that all of the $f_n \sim N_c^{-2}$ and thus are parametrically small. This means that the growth of f_n with n will also be parametrically small. Detecting clear signs of growth of f_n with n might be difficult to see without going to larger n , given both the parametrically small growth and the substantial noise in the extraction of the f_n .

Fortunately, the numerical difficulties due to this $1/N_c$ suppression are likely to be much less severe than the ones associated with chiral suppression for models satisfying conditions (i)–(ix). In particular, there is no reason to believe that N_c^{-2} suppression in the growing value of f_n should be masked at small n by a leading-order contribution as happens with chiral suppression. Thus, there is no reason to suspect that $1/N_c$ suppression would induce the type of the behavior seen in Fig. 1, where the evidence of asymptotic growth is undetectable up to some comparatively large value of n where it sets in suddenly. This greatly improves the prospects of seeing the effect in pure gauge theory.

One strategy for reducing numerical difficulties associated with N_c^{-2} suppression is to reduce N_c . Rather than studying SU(3) pure gauge theory, one could study SU(2). There are two virtues to doing this: the calculations intrinsically require fewer computational resources, possibly allowing computations of f_n with more accuracy or to larger n , and the N_c^{-2} suppression effect for each f_n is reduced, simplifying the task of identifying the growth of f_n . Of course, there is also a downside—SU(2) Yang-Mills theory is obviously further from QCD than SU(3) Yang-Mills theory. Nevertheless, if one found evidence that SU(2) pure gauge theory appeared to be consistent with conditions (i)–(vi), it would certainly make a QCD-based solution to the strong CP problem far more plausible.

There is a hierarchy of theories starting with QCD itself, going through QCD with artificially large quark masses and SU(3) pure gauge theory to SU(2) pure gauge theory. As one descends this hierarchy, the theories become progressively less like QCD, but if the f_n coefficients grow with n asymptotically, then evidence for it becomes progressively more accessible. It is noteworthy, however, that at present, even toward the bottom of the hierarchy, too few c_n coefficients have been computed to even begin exploring the issue—the state of the art for Yang-Mills theory [40] only goes up to calculations of $c^{(6)}$, allowing

only two f_n 's, and the $c^{(6)}$ computations are essentially bounds rather than well-determined values. Thus, at present, for Yang-Mills theories one can only compute f_1 and set a bound for f_2 ; this is insufficient to see even the hint of a trend valid at large n . However, it is at least conceivable that in the not-too-distant future, it may be possible to extract a couple more f_n 's. Even if these showed an increase with n , it would be insufficient to establish compelling evidence for conditions (i)–(vi). However, it could be suggestive of it. Thus, in the short term, evidence of this sort will necessarily be quite indirect.

There is another way to proceed, namely the study quantum field theories that may be quite far removed from QCD (including theories in space-time dimensions smaller than $3 + 1$). As noted previously, a principal reason to be skeptical that QCD could satisfy conditions (i)–(ix) is that, over decades of experience with quantum field theory, the community has never encountered a theory that satisfies anything resembling these conditions. The converse, however, is also true. Were one to find *any* quantum field theory where there is a reliable way to show that an analog of conditions (i)–(ix) [or even conditions (i)–(vi)] can be shown to hold and to emerge naturally from the structure of the theory, it would greatly increase the plausibility of this class of solution to the strong CP problem. Moreover, if one were to find a theory that did satisfy the analog of these conditions, there is the prospect that one might well gain insight into the underlying mechanism causing the conditions to be satisfied and to develop a physical picture of what is behind it. This, in turn, might hint at what could be happening in QCD.

This prospect suggests the possible utility of a review of known quantum field theories (in any number of dimensions) that are tractable in some manner and have some analog to the θ -term. Such theories could be tractable for a number of reasons. They could be one of the rare examples of an exactly solvable model. Alternatively, they could be tractable numerically on a lattice in a manner that allowed testing the conditions. Clearly, numerical treatments on a lattice of theories with nontrivial topology are far more likely to be viable for theories in lower space-time dimensions than in $3 + 1$ dimensions. Consider, for example, Ref. [43], a calculation of a theory in $1 + 1$ dimensions that was viable more than a decade and half ago; an analogous $3 + 1$ dimensional calculation remains impractical today. A third class of tractable examples would be a theory that is known to be solvable in some limit and that admits systematic corrections away from that limit. Of course, with this third class, one must bear in mind that any conclusions based on such a study may give misleading results (after all, the entire premise of this paper is that the chiral expansion could give misleading results for θ dependence). It is conceivable that one could identify one or more such theories that could satisfy some type of

analog of conditions (i)–(vi) that may not have been recognized as such in the past.

In searching for such a theory, the analogy to QCD need not be perfect. Of course, the further the theory is from QCD, the less evidence for the analog of conditions (i)–(vi) suggests that QCD will also satisfy these. Similarly, the weaker the analogy of the quantities in the tractable theory to those in QCD, the weaker the suggestion that QCD will satisfy them. Nevertheless, if scenarios that solve the strong CP without invoking BSM physics in the manner considered in this paper are correct, then in the short term, the

best prospect for finding evidence for them—albeit rather indirect evidence—is via the study of tractable analog theories.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the U.S. Department of Energy under Contract No. DE-FG02-93ER-40762. Comments by A. Cherman were extremely helpful.

-
- [1] G. 't Hooft, *Phys. Rev. D* **14**, 3432 (1976).
 - [2] G. 't Hooft, *Phys. Rep.* **142**, 357 (1986).
 - [3] S. Weinberg, *Phys. Rev. D* **11**, 3583 (1975).
 - [4] C. G. Callan, Jr., R. F. Dashen, and D. J. Gross, *Phys. Lett.* **63B**, 334 (1976).
 - [5] R. Jackiw and C. Rebbi, *Phys. Rev. Lett.* **37**, 172 (1976).
 - [6] J. M. Pendlebury *et al.*, *Phys. Rev. D* **92**, 092003 (2015).
 - [7] J. E. Kim and G. Carosi, *Rev. Mod. Phys.* **82**, 557 (2010).
 - [8] M. Gell-Mann, *Nuovo Cimento* **4**, 848 (1956).
 - [9] M. Dine, *Annu. Rev. Nucl. Part. Sci.* **65**, 43 (2015).
 - [10] A. Hook, *Phys. Rev. Lett.* **114**, 141801 (2015).
 - [11] R. T. D'Agnolo and A. Hook, *Phys. Lett. B* **762**, 421 (2016).
 - [12] R. D. Peccei and H. R. Quinn, *Phys. Rev. Lett.* **38**, 1440 (1977).
 - [13] R. D. Peccei and H. R. Quinn, *Phys. Rev. D* **16**, 1791 (1977).
 - [14] S. Weinberg, *Phys. Rev. Lett.* **40**, 223 (1978).
 - [15] F. Wilczek, *Phys. Rev. Lett.* **40**, 279 (1978).
 - [16] G. Bertone and D. Hooper, *Rev. Mod. Phys.* **90**, 045002 (2018).
 - [17] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, *Annu. Rev. Nucl. Part. Sci.* **65**, 485 (2015).
 - [18] S. Aoki *et al.*, *Eur. Phys. J. C* **77**, 112 (2017).
 - [19] S. Aoki *et al.*, *Eur. Phys. J. C* **74**, 2890 (2014).
 - [20] Z. Fodor, C. Hoelbling, S. Krieg, L. Lellouch, T. Lippert, A. Portelli, A. Sastre, K. K. Szabo, and L. Varnhorst, *Phys. Rev. Lett.* **117**, 082001 (2016).
 - [21] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, *Nucl. Phys.* **B166**, 493 (1980).
 - [22] R. Brower, S. Chandrasekharan, J. W. Negele, and U. J. Wiese, *Phys. Lett. B* **560**, 64 (2003).
 - [23] Y.-Y. Mao and T.-W. Chiu (TWQCD Collaboration), *Phys. Rev. D* **80**, 034502 (2009).
 - [24] S. Aoki and H. Fukaya, *Phys. Rev. D* **81**, 034022 (2010).
 - [25] F.-K. Guo and U.-G. Meiner, *Phys. Lett. B* **749**, 278 (2015).
 - [26] E. Witten, *Nucl. Phys.* **B156**, 269 (1979).
 - [27] G. Veneziano, *Nucl. Phys.* **B159**, 213 (1979).
 - [28] P. Di Vecchia and G. Veneziano, *Nucl. Phys.* **B171**, 253 (1980).
 - [29] S. Scherer, *Adv. Nucl. Phys.* **27**, 277 (2003).
 - [30] C. Bonati, M. D'Elia, M. Mariti, G. Martinelli, M. Mesiti, F. Negro, F. Sanfilippo, and G. Villadoro, *J. High Energy Phys.* **03** (2016) 155.
 - [31] S. Aoki, G. Cossu, H. Fukaya, S. Hashimoto, and T. Kaneko (JLQCD Collaboration), *Prog. Theor. Exp. Phys.* (2018), 043B07.
 - [32] M. Luscher and P. Weisz, *Nucl. Phys.* **B290**, 25 (1987).
 - [33] M. Luscher and P. Weisz, *Nucl. Phys.* **B295**, 65 (1988).
 - [34] T. Korzec and U. Wolff, *Proc. Sci., LATTICE2014* (2015) 305.
 - [35] G. Schierholz, *Nucl. Phys. B, Proc. Suppl.* **37**, 203 (1994).
 - [36] Y.-S. Wu and A. Zee, *Nucl. Phys.* **B258**, 157 (1985).
 - [37] Y. Cai, T. Cohen, A. Goldbloom-Helzner, and B. McPeak, *Phys. Rev. D* **93**, 114510 (2016).
 - [38] I. Aniceto, G. Basar, and R. Schiappa, *arXiv:1802.10441*.
 - [39] G. V. Dunne and M. Unsal, *Annu. Rev. Nucl. Part. Sci.* **66**, 245 (2016).
 - [40] C. Bonati, M. D'Elia, and A. Scapellato, *Phys. Rev. D* **93**, 025028 (2016).
 - [41] H. Leutwyler and A. V. Smilga, *Phys. Rev. D* **46**, 5607 (1992).
 - [42] F. C. Hansen and H. Leutwyler, *Nucl. Phys.* **B350**, 201 (1991).
 - [43] H. Fukaya and T. Onogi, *Phys. Rev. D* **68**, 074503 (2003).