universe

MDPI

Article

CP Conservation in the Strong Interactions

Wen-Yuan Ai [1], Björn Garbrecht [2,*] and Carlos Tamarit [3]

[1] Theoretical Particle Physics and Cosmology, King's College London, Strand, London WC2R 2LS, UK; wenyuan.ai@kcl.ac.uk
[2] Physik-Department T70, Technische Universität München, James-Franck-Straße, 85748 Garching, Germany
[3] PRISMA+ Cluster of Excellence and MITP, Johannes Gutenberg University, 55099 Mainz, Germany; ctamarit@uni-mainz.de
* Correspondence: garbrecht@tum.de

Abstract: We discuss matters related to the point that topological quantization in the strong interaction is a consequence of an infinite spacetime volume. Because of the ensuing order of limits, i.e., infinite volume prior to summing over topological sectors, *CP* is conserved. Here, we show that this reasoning is consistent with the construction of the path integral from steepest-descent contours. We reply to some objections that aim to support the case for *CP* violation in strong interactions that are based on the role of the *CP*-odd theta-parameter in three-form effective theories, the correct sampling of all configurations in the dilute instanton gas approximation and the volume dependence of the partition function. We also show that the chiral effective field theory derived from taking the volume to infinity first is in no contradiction with analyses based on partially conserved axial currents.

Keywords: quantum chromodynamics; *CP* violation; neutron electric dipole moment

check for updates

Citation: Ai, W.-Y.; Garbrecht, B.; Tamarit, C. *CP* Conservation in the Strong Interactions. *Universe* **2024**, *10*, 189. https://doi.org/10.3390/universe10050189

Academic Editors: Andreas Trautner, Celso C. Nishi, Benjamin Grinstein and Lorenzo Iorio

Received: 3 October 2023
Revised: 1 March 2024
Accepted: 16 April 2024
Published: 23 April 2024

1. Introduction

The strong interactions conserve charge–parity (CP). This has been established through many observations, to the greatest precision in searches for the permanent electric dipole moment (EDM) of the neutron [1,2], and calls for an explanation of certain theoretical aspects.

Only a few years after the discovery of quantum chromodynamics (QCD) being the theory of strong interactions [3], it has been suggested that due to the generic presence of a topological term in the action, the charge–parity symmetry CP should be violated [4–6]. To the present day, there has been no such observation, so arguably, the coefficient θ of the topological term (more precisely $\bar{\theta}$ as we introduce below) would need to be zero, corresponding to unnatural tuning and therefore a problem.

The present paper adds to the discussion following Ref. [7]. There, it has been brought up that the effective decomposition (quantization) of gauge field configurations of finite action into topological sectors of integer winding number without imposing ad hoc boundary conditions can only be derived when taking the volume of Euclidean spacetime to infinity. In contrast, there is no physical motivation for fixed boundary conditions on a finite surface in Euclidean space. As a consequence, the limit of infinite spacetime volume must be taken before summing over topological sectors, and it turns out that CP remains conserved this way.

One of the main points of the present paper is to extend this line of reasoning by demonstrating that this order of limits corresponds to a well-defined integration contour in the path integral constructed from the steepest-descent flows. In contrast, the opposite conventional order of limits results in an integration that is inequivalent in the sense of the Cauchy theorem. Another purpose of this article is to address some criticisms regarding Ref. [7]. This article also partly serves as a review on the related topics, though without including a complete list of references. For some other reviews, see, e.g., Refs. [8–10].

Without going into extensive technical detail, the basic reasons for CP conservation in the strong interaction are given in the summary of Section 2. There, we refer to the sections in the remainder of the paper where the statements of Section 2 are supported at a more technical level.

2. Summary and Outline

Strong interactions are described by a Yang–Mills theory, which generally involves CP-odd parameters through the masses of the quarks as well as through the topological term. The Lagrangian in the Euclidean spacetime is

$$\mathcal{L} = \frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}F_{\mu\nu} + \bar{\psi}\left(\hat{\gamma}_\mu D_\mu + m\mathrm{e}^{\mathrm{i}\alpha\gamma_5}\right)\psi - \frac{\mathrm{i}}{16\pi^2}\theta\,\mathrm{tr}F_{\mu\nu}\widetilde{F}_{\mu\nu}\,, \tag{1}$$

where we use the convention $\mathrm{Tr}(T^aT^b) = \delta^{ab}/2$, $[T^a, T^b] = \mathrm{i}f^{abc}T^c$ for the Lie algebra generators T^a and the structure constants f^{abc}. Above, $F_{\mu\nu} = F^a_{\mu\nu}T^a$ with $F^a_{\mu\nu} \equiv \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu$ being the field strength tensors. $\widetilde{F}^a_{\mu\nu} \equiv \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F^a_{\alpha\beta}$ ($\varepsilon_{1234} = 1$) is the Hodge dual of $F^a_{\mu\nu}$. The covariant derivative takes the form

$$D_\mu\psi_i = \left(\partial_\mu - \mathrm{i}A^a_\mu T^a\right)\psi_i \tag{2}$$

when ψ_i lives in the fundamental representation of the gauge group and

$$D_\mu\psi_i = \partial_\mu\psi_i - \mathrm{i}A^a_\mu[T^a, \psi_i] \tag{3}$$

when ψ_i lives in the adjoint representation.

The Euclidean gamma matrices $\hat{\gamma}_\mu$ are obtained from the Minkowskian counterparts

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \tag{4}$$

(where $\sigma^\mu = (1_2, \vec{\sigma})$ and $\bar{\sigma}^\mu = (1_2, -\vec{\sigma})$ with $\vec{\sigma}^i$ the Pauli matrices) via

$$\hat{\gamma}_4 = \gamma_0 = \gamma^0\,, \quad \hat{\gamma}_i = \mathrm{i}\gamma_i = -\mathrm{i}\gamma^i\,. \tag{5}$$

These matrices satisfy the Clifford algebras

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}1_4\,, \qquad \{\hat{\gamma}_\mu, \hat{\gamma}_\nu\} = 2\delta_{\mu\nu}1_4\,. \tag{6}$$

Following Ref. [7], we use the same γ^5 in Euclidean spacetime as for Minkowski spacetime (In Ref. [7], it is mistakenly stated that $\gamma^5 = \hat{\gamma}_1\hat{\gamma}_2\hat{\gamma}_3\hat{\gamma}_4$.)

$$\gamma^5 \equiv \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3 = -\hat{\gamma}_1\hat{\gamma}_2\hat{\gamma}_3\hat{\gamma}_4\,. \tag{7}$$

Since, in this paper, we mostly work in Euclidean space, from now on, we remove the hat on the Euclidean gamma matrices. Note that the γ^5 in the Euclidean spacetime used here may differ from that used in some other papers, e.g., Ref. [11], by a minus sign. The chirality in Euclidean spacetime in these two notations is thus defined oppositely. The only effect of this appears in applying the Atiyah–Singer index theorem [12] when, e.g., deriving the anomalous axial current by counting the zero modes of the massless Dirac operator.

We note that the terms $\mathrm{tr}F_{\mu\nu}\widetilde{F}_{\mu\nu}$ and $\bar{\psi}\mathrm{i}\gamma^5\psi$ are both parity-odd and charge-conjugation even [13]. The CP-odd parameters are therefore α, the phase pertaining to the mass m of the fermion ψ, and θ, the coefficient of the topological term. The fermion ψ in the fundamental representation is referred to as quark. To focus on the principal aspects, for the most part of the discussion, we take here just one quark flavor and the gauge group to be $\mathrm{SU}(2)$. This is the minimal setup that allows us to study the interplay of the CP-odd parameters α and θ. In particular, we are interested in how α and θ appear in the effective interaction (known as ‘t Hooft operator) that captures nonperturbative effects associated with the chiral anomaly. In the single-flavor model, this operator has the same form as a quark mass term, and therefore, both would be hard to discern. Nonetheless, the main question of whether and how α and θ appear in the effective 't Hooft operator can be answered within this setup. The generalization to the phenomenologically relevant case of several flavors is presented in Ref. [7]. In strong interactions, the group $\mathrm{SU}(2)$ can be viewed as embedded within the $\mathrm{SU}(3)$ color group. The technical details of this construction are reviewed in Ref. [14].

It is well known that the presence of a CP-odd Lagrangian term does not readily imply CP-violating physical effects. A necessary condition is the existence of a CP-odd combination of the Lagrangian terms that is invariant under field redefinitions. In the present case, this condition is met, as the parameter

$$\bar{\theta} = \theta + \alpha \tag{8}$$

is invariant under redefinitions of the quark fields, in particular through anomalous chiral transformations. The chiral anomaly [15,16] also implies that θ is an angular variable, i.e., all observables must be 2π-periodic in θ. Therefore, the integral $1/(16\pi^2)\int \mathrm{d}^4x\,\mathrm{tr}F\widetilde{F}$ that multiplies θ must be an integer in order to contribute to the action and thereby to the partition function.

Here, we ask the question of whether a nonvanishing $\bar{\theta}$ is also a sufficient condition for CP violation in strong interactions. Does $\bar{\theta}$ have physical effects, in particular, is there a neutron EDM depending on its value? Evidently, $\bar{\theta}$ has no impact on the classical equations of motion since the topological term is a total derivative. Nonetheless, under certain assumptions, based on the fact that the third homotopy group of the gauge group or one of its $\mathrm{SU}(2)$ subgroups is $\pi_3(\mathrm{SU}(2)) = \mathbb{Z}$, the energy functional is periodic under so-called large gauge transformations. The situation is therefore reminiscent of a periodic quantum mechanical potential in a crystal, and $\bar{\theta}$ would then correspond to the crystal momentum [4–6,11].

One way to see how far the analogy goes is to study canonical quantization of the gauge theory [4]. Since non-Abelian gauge theories are typically handled through functional quantization, this possibility has not yet been investigated in all aspects pertinent to the present questions. We briefly comment on canonical quantization in Section 4 and in more detail in a separate paper [17].

For the time being, we focus on functional quantization since it has been the principal method to carry out calculations on CP violation in strong interactions ever since the matter was brought up [5,6,11,18]. In the functional approach, we take the partition function

$$Z[\eta,\bar{\eta}] = \int \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\,\mathcal{D}A\,\mathrm{e}^{-\lim\limits_{\Omega\to\infty}\int_\Omega \mathrm{d}^4x\,(\mathcal{L}-\bar{\eta}\psi-\bar{\psi}\eta)} \tag{9}$$

as the defining point of the quantum field theory, where A is the gauge potential. We write this as a functional of external fermionic sources $\eta(x)$, $\bar{\eta}(x)$ as a provision in order to derive quark correlation functions that may or may not exhibit CP invariance. Furthermore, we make it explicit that the integral over Euclidean spacetime is understood as a limiting procedure, taking the spacetime volume Ω to infinity. It is this limit that allows us to state Equation (9) without specifying boundary conditions on the path integral, and, moreover, this way, we obtain the vacuum correlation functions of the theory [19]. We shall review this point in Section 3.

Now, as argued in Ref. [11], the partition function (9) in infinite spacetime volume $\Omega \to \infty$ receives its nonvanishing contributions from saddle points of finite action and fluctuations about these. For these field configurations, the winding number Δn is an integer that labels the topological sector:

$$\Delta n = \lim_{\Omega\to\infty}\int_\Omega \mathrm{d}^4x\,\frac{1}{16\pi^2}\mathrm{tr}F\widetilde{F} \in \mathbb{Z} \quad \text{for nonvanishing contributions to } Z\,. \tag{10}$$

This is the desired outcome because it is consistent with θ being an angular variable, as required by the chiral anomaly. As topological quantization, i.e., integer Δn, is a consequence of $\Omega \to \infty$, we must carry out this limit before summing over topological sectors.

Suppose now that it is valid to organize the calculation of the path integral by adding contributions from the individual topological sectors. Then, Equation (9) implies that the partition function should be evaluated as

$$Z[\eta, \bar{\eta}] = \sum_{\Delta n} \int \mathcal{D}\bar{\psi}\, \mathcal{D}\psi\, \mathcal{D}A_{\Delta n} \mathrm{e}^{-\lim\limits_{\Omega\to\infty} \int_\Omega \mathrm{d}^4x\, (\mathcal{L} - \bar{\eta}\psi - \bar{\psi}\eta)} . \tag{11}$$

The subscript Δn on $\mathcal{D}A_{\Delta n}$ indicates that the path integral is supposed to cover the configurations with given Δn, i.e., to sweep over the given topological sector. The contributions of the topological sectors are evaluated with the limit $\Omega \to \infty$ before they are added together. A rearrangement of the limits will in general lead to different results and is therefore not justified.

In Section 3, we expand on this argument and put it on a more formal footing. The path integrals within individual topological sectors correspond to steepest-descent contours for the exponent of the Euclidean path integral. For different Δn, these contours can only be connected by configurations of infinite action. It thus follows that the arrangement of limits in Equation (11) indeed corresponds to a good contour for the path integral in Equation (9). Further, this formally establishes that the decomposition of the path integral into topological sectors is valid in the first place.

This brings us to the salient point: in Ref. [7], it has been shown that the limit $\Omega \to \infty$ does not commute with the sum over Δn in Equation (11), with the consequence that the quark correlations do not exhibit CP violation. While we review this technical argument in Section 5, the basic reason is that $\lim_{\Delta n \to \pm\infty} \lim_{\Omega\to\infty} \Delta n/\Omega = 0$, see also Section 3. As explained in Section 7, one can thus conclude that there is no EDM for the neutron, no matter what the value of $\bar{\theta}$ is. Since, in the sum over Δn, all integer values are taken, this is where the analogy with the quantum-mechanical crystal breaks down; as for the latter, the number of potential minima may be large but remains finite. Therefore, the order of the path integral and the limit of an infinite spacetime volume are not an issue in that case.

After all, strong interactions are complete without the necessity of tuning the parameter $\bar{\theta}$ to be small or extending the theory by additional scalar fields and nonrenormalizable operators. While this is a gratifying conclusion, a scrutiny of the argument is warranted, not least because the prevalent line of reasoning arrives at the contrary verdict: in order to deduce CP-violation in strong interactions, one would have to impose that the limit $\Omega \to \infty$ is taken last, i.e.,

$$Z[\eta, \bar{\eta}^\dagger] \stackrel{?}{=} \lim_{\Omega\to\infty} \int \mathcal{D}\bar{\psi}\, \mathcal{D}\psi\, \mathcal{D}A\, \mathrm{e}^{-\int_\Omega \mathrm{d}^4x\, (\mathcal{L} - \bar{\eta}\psi - \bar{\psi}\eta)} , \tag{12}$$

and, at the same time, specify boundary conditions on the finite surfaces $\partial\Omega$, such as

$$A_\mu = \mathrm{i}\omega\partial_\mu\omega^{-1} \text{ on some finite } \partial\Omega , \tag{13}$$

where $\omega(x) \in \mathrm{SU}(2)$, which corresponds to a pure gauge and implies topological quantization, i.e., $\Delta n \in \mathbb{Z}$. Equations (12) and (13) together are either directly or indirectly implied in the bulk of the existing literature, including the initial papers on the topic [5,6,11,18]. Although Equation (13) may be motivated by considering fields in the classical ground state, i.e., of vanishing classical energy, on the initial and final spatial hypersurfaces, the quantum ground state also receives contributions from other field configurations. Therefore, the boundary condition (13) does not follow from the partition function (12) (note that Equation (13) moreover assumes that three-dimensional space is finite), unlike the topological quantization (10) that is implied by the partition function (9), see Section 3 for more detail.

To our knowledge, the published papers do not provide a conclusive reason for how the procedure given by Equations (12) and (13) might be deduced from the functional (9) that defines the theory. Given that the limits do not commute, i.e., that Equation (11) is not equivalent to Equations (12) and (13), as shown explicitly in Section 5, there also cannot be such a derivation. (Note that while in Ref. [11], it is shown that the correlators in a large but finite box depend only on the boundary conditions through the Chern–Simons flux, the issue that the limits do not commute is not addressed in that work.) Neither are we aware of an argument why Equation (9), which is the standard textbook expression (up to the fact that we write the infinite-spacetime limit explicitly, which is a purely notational matter), might be incorrect to start with. Unless taking $\Omega \to \infty$, there is also no apparent reason why the boundary condition (13) should be physical.

In the simplest terms, the reason for CP-conservation in strong interactions can thus be stated as follows:

- For θ to be physical, we must have $\Delta n \in \mathbb{Z}$ since θ is an angular variable. Since fixed boundary conditions on a finite surface are not physical, this topological quantization can only follow from $\Omega \to \infty$.
- Given the order of limits that is thus implied, there is no CP violation for $\Omega \to \infty$ since $\lim_{\Delta n \to \pm\infty} \lim_{\Omega \to \infty} \Delta n/\Omega = 0$.

While the main conclusions and technicalities on the absence of CP violation in strong interactions have been presented in Ref. [7], one objective of the present paper is to add a more formal interpretation of the difference between Equations (9) and (11) versus Equations (12) and (13). In Section 3, we recall to that end the reason for taking Euclidean time to infinity in the first place. Then, we show that Equation (11) corresponds to a contour integration that can be derived and assembled from steepest-descent flows, while the prescription of Equations (12) and (13) does not correspond to a connected integration contour. Since the reasoning in the present work is based on infinite Euclidean spacetime as the analytic continuation of Minkowski spacetime, we briefly comment in Section 4 on how calculations in finite Euclidean spacetimes with and without boundaries can be made meaningful, but we leave a detailed discussion to a separate paper. Next, as it allows for an explicit demonstration of CP conservation as a consequence of Equation (11) and for an intuitive interpretation of the matter, we review in Section 5 the dilute instanton gas calculation of Ref. [7]. This lays the ground to address objections concerning the volume dependence of the partition function in Section 6. Since there is no physical interpretation of fixed boundary conditions on finite Euclidean surfaces, we show that the partition function in fact shows the expected behavior when evaluated in finite volumes with open boundary conditions. Further objections are based on effective field theory (EFT) descriptions, which is why we review the role of the θ parameter in the effective 't Hooft vertex as well as in chiral perturbation theory in Section 7. With this preparation, we can reply in Section 8 to an objection using the topological term in hadronic matrix elements and the role of θ in the effective description of the dynamics of the topological current. After some additional comments on the sampling of topological configurations in the different orders of limits and the recent literature, we wrap up and conclude in Section 9. Except for Section 7, where we work in Minkowski spacetime, we work in Euclidean spacetime throughout all other sections.

3. Path Integral and Topological Quantization

3.1. Euclidean Partition Function for Infinite Volume

As a defining point of the quantum field theory, one may take the partition function (9). Such a partition function based on an infinite Euclidean volume is a common starting point for a wide range of calculations. Here, we go for variations with respect to the sources $\eta(x)$ and $\bar{\eta}(x)$ that yield Euclidean correlation functions for the quarks. Eventually, these can be analytically continued to Minkowski spacetime and interpreted as vacuum correlations.

Nonetheless, as some of these matters are contested in the context of CP conservation in strong interactions, we shall briefly revisit here the reasons for taking an infinite

Euclidean spacetime volume, or, more precisely, why we take imaginary time to infinity. For more discussion, see Ref. [19]. In short, it allows one to obtain vacuum correlation functions without specifying the vacuum in terms of a wave functional (which appears to be practically impossible in quantum field theory at the nonlinear level).

Imaginary time arises from the analytic continuation of real time. The corresponding Wick rotation is straightforward for any spacetime that is stationary in real time. Note that taking the spacetime volume Ω to infinity comes as a consequence of the limit of infinite Euclidean time. The spatial geometry is not the decisive reason for this, even though we take here unbounded Cartesian space $\mathbb{R}^3$ for definiteness. The reason for taking time to infinity is as follows: by Wick rotation, the correlations derived from Z for infinite imaginary time correspond to the analytically continued expectation values for the state of the lowest energy that is accessible given the conservation laws. This remains true also for clockwise rotations of the real-time axis by an angle $0 < \vartheta \leq \pi/2$ in the complex plane, provided the infinite time limit is applied prior to taking ϑ to zero (do not confuse ϑ here with the angle θ as the coefficient of the topological term). That is, for given initial and final states $|i, f\rangle$ (which, here, may be taken as the usual linear combinations of field eigenstates of the integer Chern–Simons number to comply with gauge invariance under large gauge transformations, i.e., the so-called θ-vacua), the path integral corresponds to

$$Z_\vartheta = \lim_{t\to\infty} \langle f| \mathrm{e}^{-\mathrm{i}Ht\exp(-\mathrm{i}\vartheta)} |i\rangle \,, \tag{14}$$

where H is the Hamiltonian. It therefore projects on the lowest energy eigenstate that is accessible. For $\vartheta = \pi/2$, one obtains the Euclidean path integral. In this sense, the Minkowskian vacuum correlation functions are analytic continuations of the Euclidean ones with a discontinuity on the real-time axis.

If, instead, we were keeping time purely real or still complex in general but finite, then we would have to weigh each path contributing to the partition function by the ground state wave functional Ψ evaluated at the endpoints of these paths, as we discuss in Section 4. The main reason for taking complex time to infinity is therefore to avoid this complication because then we can evaluate the path integral without explicit use of the vacuum state. In particular, since the ground state wave functional in Yang–Mills theory is not known to an approximation that addresses the present purpose, taking the imaginary part of time to infinity is the main method of making the analytic approximation of vacuum correlators feasible.

3.2. Evaluation of the Partition Function and Topological Quantization

Having specified the problem through the partition function (9), we now turn to its evaluation. In the following, we show that Equation (9) implies an integration contour that joins together the steepest-descent paths for each topological sector Δn. The field configurations on these steepest-descent contours are not bound by any finite spacetime volume, i.e., there is no value R so that $F_{\mu\nu}(x)F_{\mu\nu}(x) = 0$ (at least to some approximation) for $|x| > R$ for all field configurations on a given steepest descent. Therefore, we must evaluate the path integrals for the different topological sectors Δn in the limit $\Omega \to \infty$ before interfering with these [7]. This also implies that organizing the evaluation of the path integral in terms of a sum over contributions from the individual topological sectors is valid in the first place. Further, we argue that the order of limits in Equations (12) and (13) is opposite to what is implied by the form of the correct integration contour.

To start, we state why it is necessary to specify an appropriate integration contour. Since the integrand in Equation (9) for $\bar{\theta} \neq 0 \mod 2\pi$ is not positive definite, explicitly because of the topological term and implicitly because of fermion determinants, we must determine the integration contours to leave the integral well defined. These specify the order in which the integral $\int \mathcal{D}A$ must be carried out so that we can derive how to sum over the topological sectors. In particular, it will allow us to discern whether it is Equation (11) or Equations (12) and (13) that correspond to the correct procedure.

To determine the contours, we first note that, in general, the real part of the Euclidean Yang–Mills action

$$S \supset S_{\rm YM} = \frac{1}{2g^2} \int_\Omega {\rm d}^4 x \, {\rm tr} F_{\mu\nu} F_{\mu\nu} \tag{15}$$

should be bounded from below on its domain and thus have global or local minima that may or may not exhibit degeneracies. If Ω is infinite, this implies that the vector potential at such minima reduces to pure gauge configurations at infinity, i.e.,

$$A_\mu(x) \to {\rm i}\omega(x)\partial_\mu \omega^{-1}(x) \text{ for } |x| \to \infty \quad \text{for local minima of } {\rm Re}[S]\,, \tag{16}$$

where $\omega(x) \in {\rm SU}(2)$. Because the surface at infinity is homeomorphic to S^3, and the third homotopy group of the gauge symmetry is $\pi_3({\rm SU}(2)) = \mathbb{Z}$, this immediately implies topological quantization as in Equation (10) for these minima, i.e., integer winding number Δn. In addition, for each of these minima, there are degeneracies, i.e., flat directions of the action, parameterized by moduli [20]. Note that this reasoning applies without imposing boundary conditions ad hoc (as in Equation (13)) because the limit $\Omega \to \infty$ appears inside the path integral (9). This is in line with standard introductions of path integrals in infinite spacetimes that do not impose particular boundary conditions, see, e.g., Ref. [21], where this point is mentioned explicitly.

Given the minima of the real part of the classical Euclidean action with integer winding number, the contributions to the path integral for the individual topological sectors Δn can then be evaluated on steepest-descent contours (of the negative action $-S$) passing through these minima. The contours are Lefschetz thimbles and are determined through flow equations [22–25]. By this reasoning of steepest-descent contours, upon dealing with the usual ultraviolet divergences and the vacuum contributions in the infinite spacetime volume, the path integrals in the individual topological sectors are convergent. The steepest-descent contours for different Δn do not intersect for finite S because, in the infinite spacetime volume, solutions of different winding number are separated by infinite action barriers. The integration contours over the different sectors Δn can therefore only be connected via configurations of infinite action S that give no contribution to the partition function (9). Note that these infinite action configurations connecting the sectors are allowed precisely because we do not impose boundary conditions when evaluating the partition function. For comparison, in finite volumes with fixed boundary conditions (13), the field configurations in different topological sectors are not continuously connected, not even via a path through configurations of infinite action since paths with a noninteger winding number are forbidden by the boundary conditions (13), no matter whether S is finite or infinite. In Figure 1, we schematically illustrate these integration paths and their crucial differences.

Therefore, we first carry out the integration over the entire infinite spacetime in a given topological sector Δn individually. Then, we can connect this integration with the steepest-descent contour for a different Δn via configurations of infinite action that do not contribute to the path integral. This determines the integration contour that should be used in order to evaluate Equation (9) and that is given by the order of limits specified in Equation (11). Any alternative contour must be a continuous deformation in compliance with the prerequisites of the Cauchy theorem, a criterion that is not met with Equations (12) and (13), which consequently lead to a different result. In practice, this means that, without specifying ad hoc boundary conditions, we must not interfere the different sectors before taking $\Omega \to \infty$. Otherwise, we would partition and rearrange the full integration contour in a noncontinuous way that, in general, leads to an inequivalent result because the integrand, or, more specifically, the sum over the topological sectors, is not positive definite and not absolutely convergent.

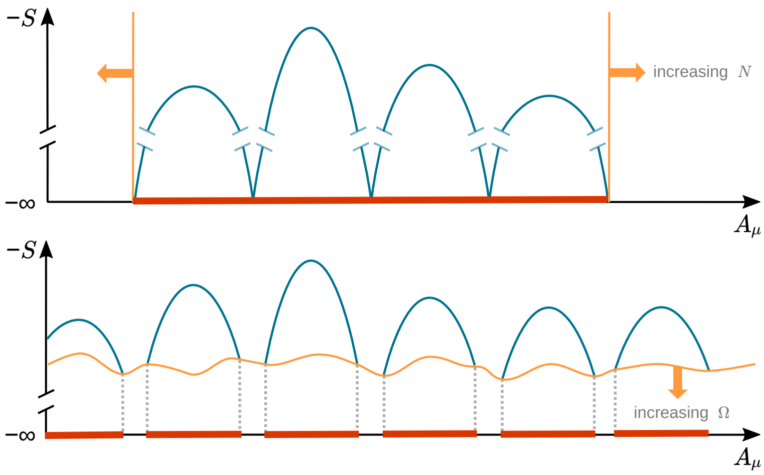


Figure 1. Upper panel: The integration contour given by the steepest-descent trajectories for the partition function (9) may be followed by successive full integrations (i.e., over all field configurations in the infinite spacetime), along the particular thimbles. This corresponds to Equation (11). The thimbles can be thought of as being connected via configurations of infinite action, leading to a connected integrated contour represented by the blue line. Lower panel: Equations (12) and (13) amount to a contour (represented again by the blue lines) that is not connected because there is no continuous deformation from one topological sector into another in finite Ω. The upper and lower contours are therefore not equivalent in the sense of the Cauchy theorem.

We can therefore write the path integral as in Equation (11) and as indicated in Figure 1. We emphasize that the decomposition into topological sectors follows from $\Omega \to \infty$ and is not a consequence of the saddle point approximation (which only comes into play in the dilute instanton gas approximation in Section 5). In turn, finite surfaces $\partial\Omega$ imply that topological charge is not quantized, i.e., it can flow in and out of the volume Ω, unless imposing unphysical constraints.

Since the winding number density $\Delta n/\Omega$ apparently is a measure of CP violation, it is already clear that for each term of the series (11), there is no CP-violating contribution, as $\lim_{\Omega\to\infty} \Delta n/\Omega = 0$. For the quark system, the corresponding calculation is explicitly presented in Section 5.

3.3. Commuting the Order of Limits

The decisive point in the present discussion regarding the CP symmetry of strong interactions is whether Equations (12) and (13) are consistent with Equation (9) and the integration contour that it implies. As we shall review in Section 5, the conclusion that the CP symmetry is violated when $\alpha + \theta \neq 0 \mod \pi$ relies on a partition function as in Equation (12) in conjunction with the ad hoc boundary conditions (13). In Equation (12), the order of limits is therefore opposite to Equation (11) that we have derived from the starting point given by Equation (9).

Now, to further (beyond the apparent contradiction with Equation (9) regarding CP) assess the validity of Equations (12) and (13), we take Ω to be finite and first assume that the boundary conditions from Equation (13) are not imposed. In particular, one can then move topological charge (i.e., instantons in the weak coupling limit) across the boundary $\partial\Omega$ so that there is no topological quantization into sectors with integer winding number Δn and also no conservation of topological charge within Ω. Without these, nontrivial minima of the action do not exist also. To bring topology back into the picture, one might therefore impose the boundary conditions given in Equation (13). These boundary conditions can be derived as a consequence of the infinite spacetime volume $\Omega \to \infty$ in the partition function (9) [11]. However, imposing these boundary conditions and computing the partition function according to Equations (12) and (13) requires commuting the limit of infinite spacetime volume with the sum over infinitely many topological sectors, which is, however, not justified.

Equations (12) and (13) therefore do not follow from Equation (9) so that one may keep looking for alternative arguments. However, while topological quantization also emerges for fixed boundary conditions on a given compact $\partial\Omega$ as in Equation (13), there does not appear to be a valid reason for imposing such configurations. In fact, the vacuum wave functional has nonvanishing support on configurations that do not observe Equation (13) and thus have a nonvanishing classical energy. Since the field operators do not commute with the Hamiltonian, the configurations obeying Equation (13) are as good or as bad as any other field configuration subject to a different boundary condition on $\partial\Omega$ (for the boundary conditions given in Equation (13), Δn are integers, whereas, for more general boundary conditions specified up to gauge transformations, Δn are given by a fixed real number plus any integer). Therefore, there is no preference for choosing pure gauges as a boundary condition on a sequence of finite surfaces, even as these surfaces are taken to infinity. For example, if $\partial\Omega$ is spherical, instantons could be placed at certain angles and close to the radius of $\partial\Omega$, what defines boundary conditions with noninteger Δn. Except that there is not a valid reason to impose Equations (12) and (13), whose consequence for the outcome of the calculation is material, there also is no a priori justification why Ω should be taken to infinity at the same rate for all topological sectors in Equation (12).

Finally, note that for examples that do not involve noncommuting limits, correlations from fixed boundary conditions on finite $\partial\Omega$ converge in imaginary time to the vacuum correlators as $\Omega \to \infty$. However, this does not imply by analogy that in the present case, where we must sum over infinitely many topological sectors, Equations (12) and (13) yield the correct vacuum correlation functions.

4. Finite Euclidean Spacetimes

Above, we have reviewed the reasoning for computing the path integral without specifying boundary conditions in favor of taking complex time to infinity. While yielding the physical correlation functions, taking time to infinity clearly is a mathematical trick. However, as we have discussed in Section 3, simply using Equations (12) and (13) is not a valid procedure. Nonetheless, it should still be possible, at least in principle, to carry out the calculation in a finite spacetime volume.

To this end, we see three ways of doing this. All of these turn out to require the replacement of the fixed boundary conditions (13) with different configurations that again lead to the same conclusion of CP conservation. These particular possibilities are:

- We can take a finite time interval at the price of having to project on a vacuum wave functional (see the present section).
- In order to avoid the projection on the wave functional, we can stay within functional quantization and consider a finite subvolume of infinite Euclidean space. Then, we have to integrate over all possible boundary configurations on the subvolume (see Section 6).
- We can take compact spacetimes without boundaries. For definiteness, consider here a four-torus with a finite Euclidean time interval of length $\beta = 1/T$, where T is the temperature. The relation with Minkowski spacetime is given by its correspondence with the canonical thermodynamic partition function. This once again requires canonical quantization that restricts the form of the wave functionals (see the present section).

4.1. Projecting on the Wave Functional

Regarding the first option, unless taking complex time to infinity or assuming finite temperature, we need to specify the ground state in order to obtain the correct boundary conditions for the path integral. That is, when restricting to a real time interval from t' to t, we must weigh each path contributing to the partition function

$$Z_{\mathrm{M}}(t_2,t_1) = \int \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\,\mathcal{D}A\,\Psi^*[A_\mu(t_2,\mathbf{x}),\psi(t_2,\mathbf{x})]\,\Psi[A_\mu(t_1,\mathbf{x}),\bar{\psi}(t_1,\mathbf{x})] \times \mathrm{e}^{\mathrm{i}\int_{\Omega_{12}} \mathrm{d}^4x\,(\mathcal{L}_{\mathrm{M}}-\bar{\eta}\psi-\bar{\psi}\eta)}, \tag{17}$$

by the unknown ground state wave functional

$$|\Psi,t\rangle = \Psi[A_\mu(\mathbf{x},t),\bar{\psi}(\mathbf{x},t))]\,|A_\mu(\mathbf{x},t)\rangle \otimes |\psi(\mathbf{x},t)\rangle \tag{18}$$

evaluated at the endpoints of these paths. Here, $\mathcal{L}_{\mathrm{M}}$ is the Minkowskian Lagrangian, $|A_\mu(\mathbf{x},t)\rangle$ and $|\psi(\mathbf{x},t)\rangle$ are field eigenstates and Ω_{12} is the four-volume bounded by the three-volumes at the times t_1 and t_2.

In turn, for physical boundary conditions that do account for fluctuations about the classical minimal-energy states, topological quantization cannot be assumed. Note while this means that the action generally receives contributions from field configurations of noninteger Δn, this does not preclude the wave functional from transforming with a phase factor under large gauge transformations. However, one may wonder whether one can still assume topological quantization in an approximate sense, so that one might still obtain a good result from the sum over path integrals with boundary conditions corresponding to pure gauges imposed on some finite volume. To settle this question, we would have to find the ground state wave functional in canonical quantization of Yang–Mills theory [26], which does not appear to be practically possible. Finite temperature field theory also relies on canonical quantization, but tangible conclusions may be drawn in that context, as we discuss next.

4.2. Compact Spacetimes without Boundary

As for introducing finite spacetime volume through temperature, a notable example is de Sitter space, where the Euclidean counterpart is a sphere, and hence Ω is finite for this spacetime. The latter can be interpreted as the representation of a canonical ensemble of states over a static patch of de Sitter space with a Lorentzian signature. Another important situation where Ω is finite is when space is a three torus and we consider a canonical ensemble as well. Then, the Euclidean representation is a four torus and corresponds to the continuum limit of computations in lattice QCD. Both of these very relevant examples of finite Euclidean volume therefore require first the canonical quantization in the respective background geometries in order to connect these with physical observables in Lorentzian spacetimes. This is mandatory in order to evaluate the trace of the canonical density matrix from which the path integral representation can then be derived.

In canonical quantization of gauge theory, large gauge transformations on the spatial section can play a special role. These are gauge transformations that are not continuously connected with the identity but give rise to equivalence classes that are a representation of the homotopy group. This also happens for infinite volumes (provided that the gauge field configurations approach a unique value at spatial infinity) and leads to the well-known θ-vacua [4–6]. However, the infinite volume limit taken inside the path integral readily implies CP conservation so that it is interesting to further focus on finite spatial volumes.

On finite spatial volumes, large gauge transformations are only singled out when fixing the gauge corresponding to periodic (i.e., single-valued) gauge potentials on the torus or single-valued gauge potentials on the sphere. Imposing single-valuedness, we find that the canonical quantization on these finite spatial volumes only admits states that are invariant under large gauge transformations, i.e., without any phase incurred, and that there hence can be no CP violation. This also resolves the matter of renormalizability of the states raised in Refs. [27,28]. On the other hand, without imposing single-valuedness, all gauge transformations on the spatial sections are continuously connected with the identity transformation, and, again, no CP-violating effects can be deduced. The details of this argument shall be published elsewhere [17].

5. Dilute Instanton Gas Approximation

The most sensitive probe of possible CP violation associated with strong interactions is the EDM of the neutron. At the relevant energy scale, QCD is deeply in the nonperturbative regime. This is a well-known and obvious drawback for any analytical approximation. Yet, one can observe from semiclassical calculations that instantons play a central role in the spontaneous breaking of chiral symmetry as well as in mediating the effects from the anomalous axial $\mathrm{U}(1)_\mathrm{A}$ symmetry that notably explains the large mass of the η' meson [29,30]. It is therefore also strongly indicated that the role of the topological term can be understood from a semiclassical evaluation of the effective fermion interaction mediated by instantons, i.e., the 't Hooft operator [29,30]. This corresponds to the expectation that the presence or absence of CP violation should prevail when crossing between the strongly and weakly coupled regimes at low and high energies, respectively. Moreover, the generic arguments in Section 3 as well as in Refs. [7,19], where cluster decomposition and the index theorem are used, do not refer to the semiclassical approximation.

The semiclassical approximation therefore remains of substantial interest, being the only analytic procedure to make quantitative statements about CP violation in strong interactions. Also, it offers a very useful perspective on the central issues with this topic.

In the present context, the semiclassical approach is given by the dilute instanton gas approximation. Stationary and quasi-stationary points of the action are described in terms of instantons and their individual collective coordinates [10,11,29,30]. Stationary points are the classical solutions. These are the minima of the action for each topological sector characterized by winding number Δn. For $\Delta n = \pm 1$, they are given by Belavin–Polyakov–Schwarz–Tyupkin (BPST) (anti-)instanton solutions [31], whose classical Yang–Mills action is

$$S_{\mathrm{BPST}} = \frac{8\pi^2}{g^2}\,. \tag{19}$$

Explicitly, the BPST instanton reads in the regular gauge

$$A^a_\mu = 2\eta_{a\mu\nu}\frac{(x-x_0)_\nu}{(x-x_0)^2+\rho^2}, \tag{20}$$

where x_0 and ρ are free parameters corresponding to the center location of the instanton and its size, respectively. Here, $\eta_{a\mu\nu}$ are the 't Hooft symbols [30]

$$\eta_{a\mu\nu} = \begin{cases} \varepsilon_{a\mu\nu}, & \mu,\nu = 1,2,3 \\ -\delta_{a\nu}, & \mu = 4 \\ \delta_{a\mu}, & \nu = 4 \\ 0, & \mu = \nu = 4 \end{cases}\,. \tag{21}$$

Similarly, one can define $\bar{\eta}_{a\mu\nu}$ by a change in the sign of δ in the above equation. For the anti-instanton, we should replace $\eta_{a\mu\nu}$ with $\bar{\eta}_{a\mu\nu}$ in Equation (20).

To visualize the BPST instanton (in analogy to Ref. [32] for one-dimensional instantons), we consider some explicit expressions with $x_0 = 0$, i.e., with the center set at the origin. For example,

$$A^3_1(x) = \frac{2x_2}{x^2+\rho^2}\,, \tag{22}$$

$$A^3_2(x) = -\frac{2x_1}{x^2+\rho^2}\,. \tag{23}$$

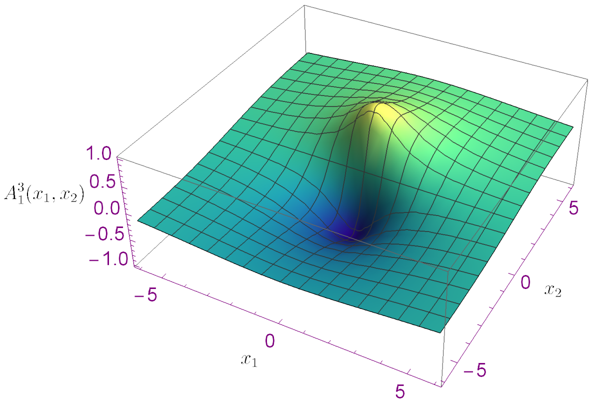
$A_1^3(x)$ is symmetric in the hyperplane $\{x_1, x_3, x_4\}$ and A_2^3 symmetric in the hyperplane $\{x_2, x_3, x_4\}$. Without loss of generality, in Figure 2, we show these as a function of x_1, x_2 by taking $x_3 = x_4 = 0$ in arbitrary units. The field strength components read

$$F^a_{\mu\nu} = -4\eta_{a\mu\nu}\frac{\rho^2}{x^2+\rho^2}\,. \tag{24}$$

As an example, we plot F^3_{21} as a function of x_1 and x_2 in Figure 3. These quantities are gauge-dependent. The gauge-independent quantities are

$$\mathrm{tr}F_{\mu\nu}F^{\mu\nu} = \mathrm{tr}F_{\mu\nu}\widetilde{F}^{\mu\nu} = -\frac{96\rho^4}{[x^2+\rho^2]^4}\,. \tag{25}$$

For the anti-instanton, one would have $\mathrm{tr}F_{\mu\nu}F^{\mu\nu} = -\mathrm{tr}F_{\mu\nu}\widetilde{F}^{\mu\nu}$. We plot these in Figure 4. From the graph, one can see that the instanton indeed has a radius characterized by the value of ρ ($\rho = 1$ in the plot).



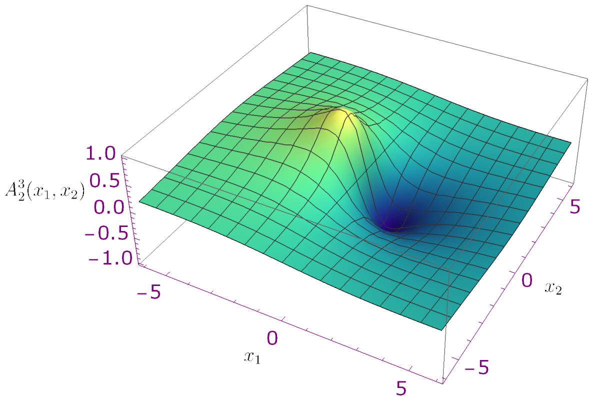


Figure 2. Visualization of the gauge potential components A_1^3 and A_2^3 in the BPST instanton solution on the $\{x_1, x_2\}$ plane ($\rho = 1$) .

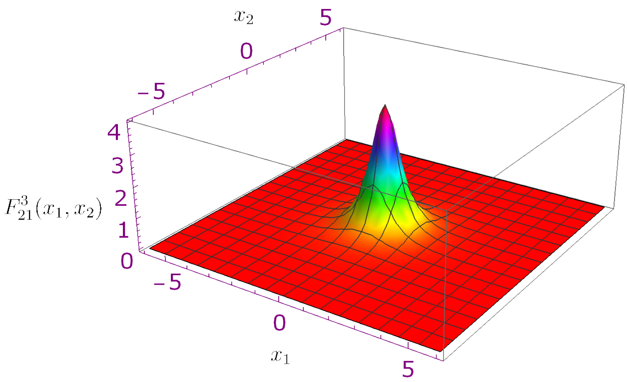


Figure 3. The field strength component F^3_{21} for the BPST instanton solution as a function of x_1 and x_2 with $x_3 = x_4 = 0$, $\rho = 1$.

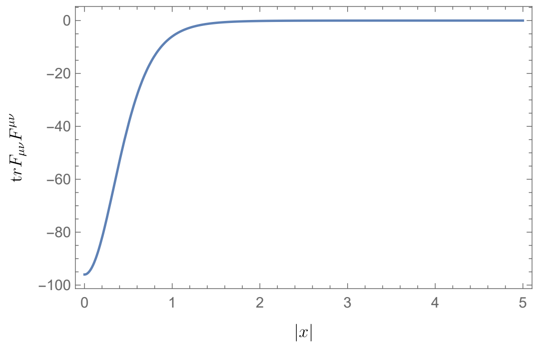


Figure 4. $\mathrm{tr}F_{\mu\nu}F^{\mu\nu} = \mathrm{tr}F_{\mu\nu}\widetilde{F}^{\mu\nu}$ as a function of $|x|$ with $\rho = 1$ for the BPST instanton.

For $|\Delta n| > 1$, they should be obtained from the Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction [33]. In the dilute gas picture, they correspond to Δn instantons, no anti-instantons for $\Delta n > 0$ and $-\Delta n$ anti-instantons, no instantons for $\Delta n < 0$. The collective coordinates for the individual instantons describe their size, their gauge orientation as well as their position in Euclidean spacetime. As we follow the steepest-descent contours, the action S evolves towards larger values, and we can encounter quasi-stationary points. These can be described in terms of the number n of instantons and $\bar{n}$ of anti-instantons, where both of these objects can coexist within such configurations. In the sector (i.e., on the thimble) characterized by Δn, it must hold that $n - \bar{n} = \Delta n$. Each of these individual instantons and anti-instantons is again parametrized in terms of the aforementioned collective coordinates. In Figure 5, we illustrate the typical structure of the thimbles in the semiclassical approximation.

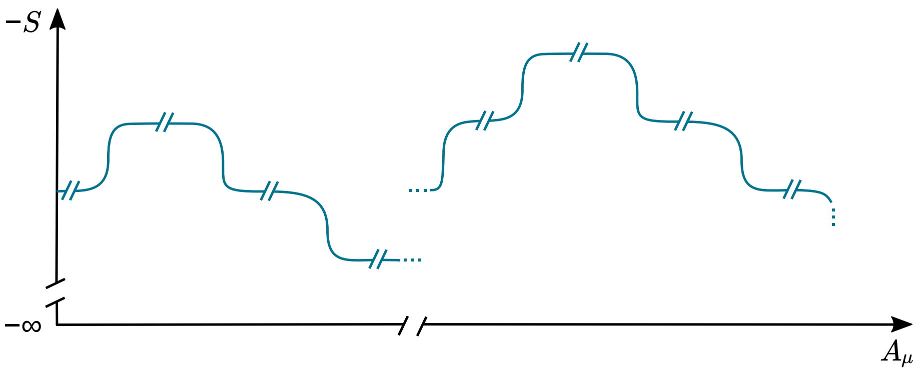


Figure 5. Some projections of steepest-descent contours (thimbles) in the semiclassical approximation. Within each topological sector, there is a set of points with minimal action. This is continuously connected via steps of height $2S_{\mathrm{BPST}}$, corresponding to the addition of pairs of instantons and anti-instantons, to additional plateaus. Double lines indicate that these plateaus extend over infinite sets in field space, in the direction of the (approximate) collective coordinates.

Now, we aim to integrate out the gluon fields in order to see explicitly what quark correlations breaking the anomalous chiral symmetry $\mathrm{U}(1)_A$ they leave behind. We follow Ref. [7], but here, we work with Euclidean time for simplicity.

The relevant quark correlation function is given by

$$\langle \psi(x)\bar{\psi}(x')\rangle = -\frac{\delta^2 \log Z[\eta,\bar{\eta}]}{\delta\bar{\eta}(x)\delta\eta(x')}\bigg|_{\eta=\bar{\eta}=0} = \frac{\int \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\,\mathcal{D}A\,\psi(x)\bar{\psi}(x')\mathrm{e}^{-\lim\limits_{\Omega\to\infty}\int_\Omega \mathrm{d}^4x\,\mathcal{L}}}{\int \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\,\mathcal{D}A\,\mathrm{e}^{-\lim\limits_{\Omega\to\infty}\int_\Omega \mathrm{d}^4x\,\mathcal{L}}}\,. \quad (26)$$

While this is a standard expression, one should note that the numerator and denominator in this equation are not well defined in the thermodynamic limit $\Omega \to \infty$, even when ultraviolet divergences have been renormalized. However, this does not force us to keep Ω finite. Rather, divergent extensive contributions in the numerator and denominator from spacetime regions far away from x and x' cancel. In standard perturbation theory, these contributions are represented by vacuum diagrams where the divergence results from their overall invariance under spacetime translations. We go into more detail regarding this point in Section 6. In the present semiclassical evaluation, we shall see how to deal with these extensive contributions a bit further down the line of argument.

To proceed with the evaluation of Equation (26), we approximate the Green's function of the quarks in the background of one anti-instanton ($\Delta n = -1$) as

$$S(x,x') = \sum\!\!\!\!\!\!\!\int \frac{\hat{\psi}_{\bar{\lambda}}(x)\hat{\psi}^\dagger_{\bar{\lambda}}(x')}{\bar{\lambda}} \approx \frac{\hat{\psi}_{0\mathrm{L}}(x)\hat{\psi}^\dagger_{0\mathrm{L}}(x')}{m\,\mathrm{e}^{-\mathrm{i}\alpha}} + \sum\!\!\!\!\!\!\!\int_{\lambda\neq 0} \frac{\hat{\psi}_{\lambda}(x)\hat{\psi}^\dagger_{\lambda}(x')}{\lambda}\,. \tag{27}$$

The middle expression is the exact spectral sum representation in terms of the eigenvalues $\bar{\lambda}$ of the Dirac operator of massive quarks in the anti-instanton background, and $\hat{\psi}_{\bar{\lambda}}$ are the corresponding eigenfunctions. As for the approximation on the right, by $\hat{\psi}_{0\mathrm{L,R}}$, we denote the 't Hooft zero modes of the massless Dirac operator in the corresponding one anti-instanton or instanton background [29,30], which are purely chiral and where their handedness is indicated by L, R. The nonzero eigenvalues of the massless Dirac operator are given by λ and the pertaining eigenmodes by ψ_λ. Note that the contribution breaking chiral symmetry, i.e., the first term in the approximate expression, aligns with the CP phase α pertaining to the quark mass and not with the angle θ. For real masses, this approximation has been used in Refs. [34,35].

In the semiclassical approximation, we carry out the path integral by taking the quasistationary configurations of the action, i.e., with n instantons and $\bar{n}$ anti-instantons, and evaluate the leading fluctuations, i.e., the functional determinants corresponding to one-loop order. For such a quasistationary background, the Green's function for the quarks should be well approximated by [35]

$$\begin{aligned} S_{n,\bar{n}}(x,x') \approx S_0(x,x') + &\sum_{\bar{\nu}=1}^{\bar{n}} \frac{\hat{\psi}_{0\mathrm{L}}(x-x_{0,\bar{\nu}})\hat{\psi}^\dagger_{0\mathrm{L}}(x'-x_{0,\bar{\nu}})}{m\,\mathrm{e}^{-\mathrm{i}\alpha}} \\ + &\sum_{\nu=1}^{n} \frac{\hat{\psi}_{0\mathrm{R}}(x-x_{0,\nu})\hat{\psi}^\dagger_{0\mathrm{R}}(x'-x_{0,\nu})}{m\,\mathrm{e}^{\mathrm{i}\alpha}}\,. \end{aligned} \tag{28}$$

Here, $x_{0,\nu}$ and $x_{0,\bar{\nu}}$ are the locations of instantons and anti-instantons, respectively, and S_0 is the Green's function of a Dirac fermion with mass $m\exp(\mathrm{i}\alpha\gamma^5)$ in a translation-invariant (i.e., void of instantons) background. This approximation neglects contributions from overlapping instantons, which are more suppressed as the instanton gas becomes more dilute. While the Green's function close to the individual instantons and anti-instantons is dominated and therefore approximated by the 't Hooft zero modes, sufficiently far away from the points $x_{0,\nu}$ and $x_{0,\bar{\nu}}$, the Green's function is given by the form in the background without instantons, i.e.,

$$S_0(x,x') = (-\gamma_\mu\partial_\mu + m\,\mathrm{e}^{-\mathrm{i}\alpha\gamma^5}) \int \frac{\mathrm{d}^4 p}{(2\pi)^4}\,\mathrm{e}^{-\mathrm{i}p(x-x')}\,\frac{1}{p^2+m^2}\,. \tag{29}$$

In Equation (28), we note the alignment of the instanton-induced breaking of chiral symmetry with the quark masses so that there is no indication of CP violation at this level but also note that θ has not yet entered into the calculation.

Given the Green's functions (28), we can proceed with evaluating the fermion correlation on the thimble (or equivalently in a fixed topological sector) characterized by the winding number Δn:

$$
\begin{aligned}
&\langle\psi(x)\bar{\psi}(x')\rangle_{\Delta n} \\
&= \sum_{\substack{\bar{n},n\geq 0\\ n-\bar{n}=\Delta n}} \int \mathcal{D}A_{\bar{n},n}\mathcal{D}\bar{\psi}\mathcal{D}\psi\,\psi(x)\bar{\psi}(x')\mathrm{e}^{-S[A,\bar{\psi},\psi]} \\
&= \sum_{\substack{\bar{n},n\geq 0\\ n-\bar{n}=\Delta n}} \frac{1}{\bar{n}!n!}\left(\prod_{\bar{\nu}=1}^{\bar{n}}\int_\Omega \mathrm{d}^4x_{0,\bar{\nu}}\mathrm{d}\Sigma_{\bar{\nu}}J_{\bar{\nu}}\right)\left(\prod_{\nu=1}^{n}\int_\Omega \mathrm{d}^4x_{0,\nu}\mathrm{d}\Sigma_{\nu}J_{\nu}\right)S_{\bar{n},n}(x,x') \\
&\qquad\qquad\times\,\mathrm{e}^{-S_{\mathrm{BPST}}\,(\bar{n}+n)}\mathrm{e}^{-\mathrm{i}(\bar{n}-n)(\alpha+\theta)}(-\Theta\varpi)^{(\bar{n}+n)}\,.
\end{aligned}
\tag{30}
$$

The symbol $\mathcal{D}A_{\bar{n},n}$ implies that the path integral is evaluated in terms of fluctuations and moduli about the classical background, i.e., (quasi-)stationary point, made up from n instantons and $\bar{n}$ anti-instantons. The integration over collective coordinates other than the locations of the instantons and anti-instantons are denoted by $\mathrm{d}\Sigma_{\nu,\bar{\nu}}$, and the Jacobians from the transformation of the zero modes in the path integral in favor of the collective coordinates are denoted by $J_{\nu,\bar{\nu}}$. The one-loop determinant of the gauge field about a single instanton or anti-instanton (denoted by $\bar{A}$ below) with the zero modes omitted and divided by the gauge field determinant in the background $A=0$ is given by

$$
\varpi \equiv \frac{1}{\sqrt{\det'_{\bar{A}}/\det_{A=0}}}\,,
\tag{31}
$$

where the prime on the determinant indicates the omission of zero eigenvalues. In an analogous manner, Θ represents the modulus of the ratio of the fermionic determinants in the one-(anti)instanton and $A=0$ backgrounds,

$$
\Theta \equiv \left|\frac{\det(-\not{D}-m\mathrm{e}^{\mathrm{i}\alpha\gamma^5})}{\det(-\not{\partial}-m\mathrm{e}^{\mathrm{i}\alpha\gamma^5})}\right|\,.
\tag{32}
$$

As usual, the partition function diverges in the thermodynamic limit so that we keep the spacetime volume Ω finite for now. Nonetheless, we need to take $\Omega\to\infty$ before eventually summing over the topological sectors Δn, as the latter are only a consequence of infinite spacetime volume and to remain true to the integration contour implied by Equation (9), cf. the discussion in Section 3.

In order to normalize, i.e., to divide out vacuum contributions, we also need the partition function in a fixed topological sector. Proceeding as for the fermion correlation, we obtain

$$
\begin{aligned}
Z_{\Delta n} &= \sum_{\substack{\bar{n},n\geq 0\\ n-\bar{n}=\Delta n}} \int \mathcal{D}A_{\bar{n},n}\mathcal{D}\bar{\psi}\mathcal{D}\psi\,\mathrm{e}^{-S[A,\bar{\psi},\psi]} \\
&= \sum_{\substack{\bar{n},n\geq 0\\ n-\bar{n}=\Delta n}} \frac{1}{\bar{n}!n!}\left(-\Omega J\,\Theta\,\varpi\,\mathrm{e}^{-S_{\mathrm{BPST}}}\int\mathrm{d}\Sigma\right)^{(\bar{n}+n)}\mathrm{e}^{-\mathrm{i}(\bar{n}-n)(\alpha+\theta)}\,.
\end{aligned}
\tag{33}
$$

Next, we turn to the collective coordinates and integrate out the location of a single anti-instanton as

$$
\begin{aligned}
&\int_\Omega \mathrm{d}^4x_{0,\bar{\nu}}\;S(x,x') \\
\approx&\int_\Omega \mathrm{d}^4x_{0,\bar{\nu}}\left[S_0(x,x')+\frac{\hat{\psi}_{0\mathrm{L}}(x-x_{0,\bar{\nu}})\hat{\psi}^\dagger_{0\mathrm{L}}(x'-x_{0,\bar{\nu}})}{m\mathrm{e}^{-\mathrm{i}\alpha}}+\cdots\right] \\
=&\,\Omega\,(S_0(x,x')+\cdots)+m^{-1}\mathrm{e}^{\mathrm{i}\alpha}h(x,x')P_{\mathrm{L}}\,.
\end{aligned}
\tag{34}
$$

The dots above represent contributions from the zero modes of the (anti)-instantons whose centers were not integrated over. This expression defines the overlap function $h(x,x')$—a rank-two tensor in spinor space:

$$h(x,x')P_{\mathrm{L}} = \int_{\Omega} \mathrm{d}^4 x_{0,\bar{\nu}}\, \hat{\psi}_{0\mathrm{L}}(x - x_{0,\bar{\nu}})\hat{\psi}^{\dagger}_{0\mathrm{L}}(x' - x_{0,\bar{\nu}})\,, \tag{35}$$

$$h(x,x')P_{\mathrm{R}} = \int_{\Omega} \mathrm{d}^4 x_{0,\nu}\, \hat{\psi}_{0\mathrm{R}}(x - x_{0,\nu})\hat{\psi}^{\dagger}_{0\mathrm{R}}(x' - x_{0,\nu})\,. \tag{36}$$

Further, we integrate over the remaining collective coordinates as

$$\bar{h}(x,x') \equiv \frac{\int \mathrm{d}\Sigma\, h(x,x')}{\int \mathrm{d}\Sigma}\,. \tag{37}$$

Notice that we ignore here the fact that for the classical instanton, the integral over the dilatational mode is divergent. The running coupling will however render the correlations finite in a more complete calculation.

Strictly speaking, the dilute instanton gas approximation is only applicable when the integral over the dilatational mode converges, i.e., when contributions from both large and small instantons are cut off. This is naturally the case in the ultraviolet, where small instantons are suppressed for asymptotically free theories as $g \to 0$. In the infrared, one may attempt to keep g perturbatively small by a bespoke particle content that controls the running coupling. As a matter of principle, an infrared cutoff can also be enforced when the gauge symmetry is spontaneously broken so that the size of the instantons is limited by the inverse gauge boson mass [29,30]. While none of this applies to strong interactions, such considerations show that the dilute instanton gas is a meaningful concept. Note that an infrared cutoff for the instanton size has no implications for the integration over the locations $x_{0,\nu}, x_{0,\bar{\nu}}$ of the instanton centers. The preservation of Poincaré symmetry, and, in particular, Lorentz invariance, demands that the values of $x_{0,\nu}, x_{0,\bar{\nu}}$ should remain unconstrained. Hence, the dependence of the results on the spacetime volume is unchanged in the presence of a cutoff for the instanton size. As a consequence, the former has no consequence for the order of limits of infinite spacetime volume and infinite maximal absolute value of the topological charge. Note that even with the aforementioned size cutoff, it is clear that either expression (11) or (12) can be technically evaluated. Further, the presence or absence of divergences from infrared instantons does not decide which order of limits must be taken because the presence of sectors of integer Δn is a topological argument that does not depend on the validity of the semiclassical expansion. We eventually note here that the validity of the dilute instanton gas and its generalization toward the inclusion of interactions between instantons has been addressed in Refs. [9,10].

The present point of view is that the saddle point approximation in the dilute instanton gas approach, while not quantitatively applicable to strong interactions, yields information about the symmetries that are respected by the theory. This does not only apply to the present work that argues in favor of the evaluation of the partition function according to Equation (11) but also to Refs. [5,18] that assume Equations (12) and (13). To our knowledge, Ref. [18] is the only paper that explicitly evaluates the ’t Hooft operator for nonzero $\bar{\theta}$. While the saddle point approximation is an important cross check, we note that the conclusion about the absence of CP violation does not rely on it, cf. the boxed argument in Section 2 and in the part of Section 3 on the evaluation of the partition function and topological quantization as well as Refs. [7,19]. Furthermore, as will be reviewed at the end of Section 6, the results obtained with the dilute instanton gas can be recovered from general arguments based on cluster decomposition and the index theorem, without making use of the dilute gas approximation.

Integrating now over all locations of instantons and anti-instantons, we obtain the correlation function for fixed Δn:

$$
\begin{aligned}
&\langle\psi(x)\bar{\psi}(x')\rangle_{\Delta n}\\
&=\sum_{\substack{\bar{n},n\geq 0\\ n-\bar{n}=\Delta n}}\frac{1}{\bar{n}!n!}\left[\bar{h}(x,x')\left(\frac{\bar{n}}{m\mathrm{e}^{-\mathrm{i}\alpha}}P_{\mathrm{L}}+\frac{n}{m\mathrm{e}^{\mathrm{i}\alpha}}P_{\mathrm{R}}\right)\Omega^{\bar{n}+n-1}+S_0(x,x')\Omega^{\bar{n}+n}\right]\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\kappa^{\bar{n}+n}(-1)^{n+\bar{n}}\mathrm{e}^{\mathrm{i}\Delta n(\alpha+\theta)}\\
&=\left[\left(\mathrm{e}^{\mathrm{i}\alpha}I_{\Delta n+1}(2\kappa\Omega)P_{\mathrm{L}}+\mathrm{e}^{-\mathrm{i}\alpha}I_{\Delta n-1}(2\kappa\Omega)P_{\mathrm{R}}\right)\frac{\kappa}{m}\bar{h}(x,x')+I_{\Delta n}(2\kappa\Omega)S_0(x,x')\right]\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times(-1)^{\Delta n}\mathrm{e}^{\mathrm{i}\Delta n(\alpha+\theta)}\,,
\end{aligned}
\tag{38}
$$

where $\kappa=\int\mathrm{d}\Sigma\, J\,\Theta\,\varpi\,\mathrm{e}^{-S_{\mathrm{BPST}}}$ is the instanton density per spacetime volume and $I_\nu(x)$ is the modified Bessel function. (Note that in Minkowskian spacetime, we define $\mathrm{i}\kappa=\int\mathrm{d}\Sigma\, J\Theta\varpi\mathrm{e}^{-S_{\mathrm{BPST}}}$ in Ref. [7]. The κ in both cases is the same and is real due to the fact that the Jacobian J in Minkowski spacetime contains an additional factor of i compared to its Euclidean counterpart.)

The terms involving the overlap function $\bar{h}$ are due to the instanton effects on the quarks and break chiral symmetry. While we should expect that these scale in the same way with the spacetime volume Ω as the term with S_0, i.e., the contribution from regions between instantons, the explicit dependence on Ω in Equation (38) is different. However, we see that the scaling after all is the same. Relax for the moment the constraint of fixed Δn and use that κ may be interpreted as the likelihood for finding an instanton in a unit four-volume. Then, for large Ω, the sum is dominated by a particular value of $n\approx\bar{n}$:

$$
\langle n\rangle=\frac{\sum_{n=0}^{\infty}n\frac{(\kappa\Omega)^n}{n!}}{\sum_{n=0}^{\infty}\frac{(\kappa\Omega)^n}{n!}}=\kappa\Omega\,.
\tag{39}
$$

Moreover, the relative fluctuation vanishes in the infinite-volume limit [7]:

$$
\frac{\sqrt{\langle(n-\langle n\rangle)^2\rangle}}{\langle n\rangle}=\frac{1}{\sqrt{\kappa\Omega}}\,.
\tag{40}
$$

This means that in the coefficients in front of the chiral projection operators within the middle expression in Equation (38), we can replace $n,\bar{n}\to\kappa\Omega$. This basic behavior, i.e., that the central value for the number of instantons is given by $\kappa\Omega$, is also reflected by the fact that for large arguments, the modified Bessel functions become independent of their index, i.e., $\lim_{x\to\infty}I_{\Delta n}(x)/I_{\Delta n'}(x)=1$. Since all the modified Bessel functions in Equation (38) tend to the same value, we see directly from this expression that there is no relative CP phase between the terms from the quark masses and instanton-induced breaking of chiral symmetry in the infinite-volume limit. Correspondingly, the partition function for fixed Δn turns out as [36]

$$
Z_{\Delta n}=I_{\Delta n}(2\kappa\Omega)\,(-1)^{\Delta n}\mathrm{e}^{\mathrm{i}\Delta n(\alpha+\theta)}\,.
\tag{41}
$$

Now, when calculating the correlation function as the sum over the topological sectors, we have to take the limit $\Omega\to\infty$ first for the reasons explained in Section 3. Because of the divergence in the thermodynamic limit, the numerator and denominator have to be treated together, and we obtain

$$
\begin{aligned}
\langle\psi(x)\bar{\psi}(x')\rangle=&\lim_{\substack{N\to\infty\\ N\in\mathbb{N}}}\lim_{\Omega\to\infty}\frac{\sum_{\Delta n=-N}^{N}\langle\psi(x)\bar{\psi}(x')\rangle_{\Delta n}}{\sum_{\Delta n=-N}^{N}Z_{\Delta n}}\\
=&S_{0\mathrm{inst}}(x,x')+\kappa\bar{h}(x,x')m^{-1}\mathrm{e}^{-\mathrm{i}\alpha\gamma^5}\,.
\end{aligned}
\tag{42}
$$

In Section 6, we show that this procedure amounts to dropping the divergent extensive contributions that correspond to the vacuum diagrams in standard perturbation theory. In this final result, the phase from the quark mass in $S_{0\text{inst}}$, cf. Equation (29), is aligned with the phase from the instanton-induced effects in the term with the overlap function $\bar{h}$, so that there are no CP-violating effects.

One may wonder about Equation (42) why we take the limit $N \to \infty$ in front of the fraction, whereas, by Equation (9), it appears that it should hold for numerator and denominator separately in the first place. As we have noted though, without normalization by vacuum contributions, the partition function is not well defined in the thermodynamic limit. The present procedure is necessary to divide out the extensive contributions causing the divergence. It is unique in the sense that we carry out the integrals over each steepest-descent contour before interfering them. Doing otherwise would correspond to a partitioning and reordering of the full integration contour that consists of the steepest-descent contours connected via configurations of infinite action, see Figure 1 for illustration. This amounts to an incorrect manipulation of a path integral that is not absolutely convergent.

Now, we consider what happens when the limits are ordered the other way around, i.e., sum over the topological sectors before taking $\Omega \to \infty$, according to Equations (12) and (13). We reiterate though that this procedure is not valid because topological quantization can only be deduced in infinite spacetime volume. As for the fermion correlation, one obtains

$$\begin{aligned}
&\sum_{\bar{n},n\geq 0} \frac{1}{\bar{n}!n!}\Big[\bar{h}(x,x')(\bar{n}\, m^{-1}\mathrm{e}^{\mathrm{i}\alpha}P_{\mathrm{L}} + n\, m^{-1}\mathrm{e}^{-\mathrm{i}\alpha}P_{\mathrm{R}})(\Omega)^{\bar{n}+n-1} + S_0(x,x')(\Omega)^{\bar{n}+n}\Big] \\
&\qquad\qquad\qquad\qquad \times (-\kappa)^{\bar{n}+n}\mathrm{e}^{\mathrm{i}\Delta n(\alpha+\theta)} \\
=&\Big[-\Big(\mathrm{e}^{-i\theta}P_{\mathrm{L}} + \mathrm{e}^{\mathrm{i}\theta}P_{\mathrm{R}}\Big)\frac{\kappa}{m}\bar{h}(x,x') + S_0(x,x')\Big]\mathrm{e}^{-2\kappa\Omega\cos(\alpha+\theta)}\,,
\end{aligned} \tag{43}$$

and for the partition function

$$\sum_{n,\bar{n}} \frac{1}{n!\bar{n}!}(-\kappa\Omega)^{\bar{n}+n}\mathrm{e}^{-\mathrm{i}(\bar{n}-n)(\alpha+\theta)} = \mathrm{e}^{-2\kappa\Omega\cos(\alpha+\theta)}\,. \tag{44}$$

Taking the ratio, the overall exponential factors cancel, but now there is a misalignment between the phases in $S_{0\text{inst}}$ and in the instanton-induced term. This means that as $\Omega \to \infty$, there is an infinite amount of destructive interference that suppresses the statistically more likely contributions with approximately equal numbers of n and $\bar{n}$ (see Equation (39)) in favor of outliers for which $\Delta n/\Omega$ does not go to zero. Equations (43) and (44), if they were correct, would signal CP-violating effects. Note that in either result, terms that break the $\mathrm{U}(1)_A$ symmetry from both instanton-mediated effects and the quark mass m are present. When we turn to the phenomenology of strong interactions and generalization to several flavors in Section 7, we shall recall that both the breaking of chiral symmetry through the quark masses and from instantons are necessary in order to explain the spectrum of mesons, in particular why the η'-meson is much heavier than the pions [18,29,30]. In either order of limits, this phenomenology is explained. Therefore, the meson spectrum alone cannot be used in order to conclude the correct order of limits.

6. Thermodynamic Limit and Cluster Decomposition

We establish here that with the limiting procedure in Equation (42), contributions that are divergent due to the infinite spacetime volume cancel between numerator and denominator. This corresponds to the usual cancellation of vacuum diagrams when evaluating connected correlation functions in standard perturbation theory (i.e., without expanding around nontrivial classical solutions).

The present argument is also interesting for what concerns Equation (9). The partition function is defined in the limit $\Omega \to \infty$ in the first place. This appears as an obstacle to using $\log Z$ as an extensive, volume-dependent quantity in line with what is familiar from thermodynamics. We shall see here that an expression with such a property can nonetheless be defined when restricting Z to some subvolume of Ω. Note that as such a restriction is arbitrary, no boundary conditions on the subvolume can be placed.

While we are working here at zero temperature, we note that one can use the Polyakov line at finite temperature in order to control and study the deconfinement phase transition, including contributions from the gradient expansion of the quark determinant [37].

We use a well-known line of reasoning [38] and consider the expectation value of an operator $\mathcal{O}$ in an infinite spacetime volume Ω and interfere different topological sectors Δn as

$$\langle \mathcal{O} \rangle = \lim_{\substack{N\to\infty \\ N\in\mathbb{N}}} \lim_{\Omega\to\infty} \frac{\sum\limits_{\Delta n=-N}^{N} f(\Delta n) \int\limits_{\Delta n} \mathcal{D}\phi\, \mathcal{O}\, \mathrm{e}^{-S_\Omega[\phi]}}{\sum\limits_{\Delta n=-N}^{N} f(\Delta n) \int\limits_{\Delta n} \mathcal{D}\phi\, \mathrm{e}^{-S_\Omega[\phi]}}\,, \tag{45}$$

where $\mathcal{D}\phi$ is the path integral measure over all fields involved. Now, let $\mathcal{O}$ be an operator corresponding to a correlation function evaluated for some spacetime points. For example, in Equation (42), $\mathcal{O} = \psi(x)\bar{\psi}(x')$. For the action, we write S_Ω to indicate that it is obtained from integrating the Lagrangian over the spacetime volume Ω. As for the Lagrangian, we take it not to include the topological term $-\mathrm{i}\theta \mathrm{tr} F\widetilde{F}/(16\pi^2)$. Rather, we have the function $f(\Delta n)$ taking care of the dependence on the topological sector.

Now, consider partitioning the spacetime volume as $\Omega = \Omega_1 \cup \Omega_2$ so that $\Delta n(\Omega) = \Delta n_1(\Omega_1) + \Delta n_2(\Omega_2)$. We further assume that the spacetime arguments of the operator fall within Ω_1, and we write $\mathcal{O}_1$ in favor of $\mathcal{O}$ to indicate this. We can thus write

$$\langle \mathcal{O}_1 \rangle = \lim_{\substack{N_1,N_2\to\infty \\ N_1,N_2\in\mathbb{N}}} \lim_{\Omega\to\infty} \frac{\sum\limits_{\Delta n_1=-N_1}^{N_1} \sum\limits_{\Delta n_2=-N_2}^{N_2} f(\Delta n_1+\Delta n_2) \int\limits_{\Delta n_1} \mathcal{D}\phi\, \mathcal{O}_1\, \mathrm{e}^{-S_{\Omega_1}[\phi]} \int\limits_{\Delta n_2} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_2}[\phi]}}{\sum\limits_{\Delta n_1=-N_1}^{N_1} \sum\limits_{\Delta n_2=-N_2}^{N_2} f(\Delta n_1+\Delta n_2) \int\limits_{\Delta n_1} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_1}[\phi]} \int\limits_{\Delta n_2} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_2}[\phi]}}\,. \tag{46}$$

Since there may be instantons sitting right at the boundaries of the two subvolumes, $\Delta n_{1,2}$ will not be strictly integer. However, if the instanton gas is sufficiently dilute, integer winding numbers may still correspond to an adequate approximation.

Now, as required by the cluster decomposition principle, provided Ω_1 is chosen large enough, $\langle \mathcal{O}_1 \rangle$ must not depend on contributions from Ω_2 to the path integral. This is generally the case when the numerator and the denominator decompose into factors that only depend on Δn_1, Ω_1 or Δn_2, Ω_2, respectively. Then, the contributions from the volume Ω_2 can be reduced from the fractions. This generally happens when

$$f(\Delta n_1 + \Delta n_2) = f(\Delta n_1) f(\Delta n_2) \Rightarrow f(\Delta n) = \mathrm{e}^{\mathrm{i}\Delta n\theta}\,. \tag{47}$$

Therefore, the contributions from the topological term, that we have left aside thus far, can indeed be accounted for through $f(\Delta n)$. Note that the argument holds for either order of limits, i.e., the one from Equation (9) that implies Equation (11), which is imposed here, as well as for the commuted version from Equation (12).

To carry out the limits, we write Equation (46) as

$$\langle \mathcal{O}_1 \rangle = \lim_{\substack{N_1,N\to\infty \\ N_1,N\in\mathbb{N}}} \lim_{\Omega\to\infty} \frac{\sum\limits_{\Delta n=-N}^{N} \sum\limits_{\Delta n_1=-N_1}^{N_1} f(\Delta n) \int\limits_{\Delta n_1} \mathcal{D}\phi\, \mathcal{O}_1\, \mathrm{e}^{-S_{\Omega_1}[\phi]} \int\limits_{\Delta n_2=\Delta n-\Delta n_1} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_2}[\phi]}}{\sum\limits_{\Delta n=-N}^{N} \sum\limits_{\Delta n_1=-N_1}^{N_1} f(\Delta n) \int\limits_{\Delta n_1} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_1}[\phi]} \int\limits_{\Delta n_2=\Delta n-\Delta n_1} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_2}[\phi]}}\,. \tag{48}$$

Corresponding to Equation (41), the integrations over the volume Ω_2 lead to

$$\langle \mathcal{O}_1 \rangle = \lim_{\substack{N_1, N \to \infty \\ N_1, N \in \mathbb{N}}} \lim_{\Omega_2 \to \infty} \frac{\sum_{\Delta n=-N}^{N} \sum_{\Delta n_1=-N_1}^{N_1} f(\Delta n)\, I_{\Delta n - \Delta n_1}(2\kappa\Omega_2)(-1)^{N_f(\Delta n - \Delta n_1)} \mathrm{e}^{\mathrm{i}\,\alpha(\Delta n - \Delta n_1)} \int_{\Delta n_1} \mathcal{D}\phi\, \mathcal{O}_1\, \mathrm{e}^{-S_{\Omega_1}[\phi]}}{\sum_{\Delta n=-N}^{N} \sum_{\Delta n_1=-N_1}^{N_1} f(\Delta n)\, I_{\Delta n - \Delta n_1}(2\kappa\Omega_2)(-1)^{N_f(\Delta n - \Delta n_1)} \mathrm{e}^{\mathrm{i}\,\alpha(\Delta n - \Delta n_1)} \int_{\Delta n_1} \mathcal{D}\phi\, \mathrm{e}^{-S_{\Omega_1}[\phi]}}\,. \tag{49}$$

The explicit exponential factors here are phases from the fermion determinants. Note that these have not been absorbed in κ, which we have defined to be real.

Now, we are aiming for an expression for $\langle \mathcal{O}_1 \rangle$ with finite Ω_1, without making reference to Ω_2. This proves to be possible because the contributions from Ω_2 can be interpreted as vacuum factors that reduce out from the normalized expectation value.

Since we must take $\Omega \to \infty$ to have well-defined integer Δn, we also have to take here $\Omega_2 \to \infty$. The Bessel functions with a factor Ω_2 in their argument then go to a common limit so that we can factorize out the sum over Δn. We are left with

$$\langle \mathcal{O}_1 \rangle = \frac{\sum_{\Delta n_1=-\infty}^{\infty} \int_{\Delta n_1} \mathcal{D}\phi\, (-1)^{-N_f \Delta n_1} \mathrm{e}^{-\mathrm{i}\,\alpha \Delta n_1} \mathcal{O}_1\, \mathrm{e}^{-S_{\Omega_1}[\phi]}}{\sum_{\Delta n_1=-\infty}^{\infty} \int_{\Delta n_1} \mathcal{D}\phi\, (-1)^{-N_f \Delta n_1} \mathrm{e}^{-\mathrm{i}\,\alpha \Delta n_1} \mathrm{e}^{-S_{\Omega_1}[\phi]}}\,. \tag{50}$$

We therefore see that taking the limits as in Equation (42) leads to the correct cancellation of "disconnected" terms, in particular those that originate from regions that are far separated from the spacetime arguments of the observable $\mathcal{O}_1$.

Moreover, in Equation (50), the θ-angle from the function $f(\Delta n)$ does not occur anymore. We can see this as a consequence of phases incurred in Ω_1 being canceled against complementary phases from $\Omega \setminus \Omega_1$. The remaining explicit dependence on the unphysical phase α cancels when the fermionic part of the path integral $\mathcal{D}\phi$ is carried out. Since the path integral here is restricted to Ω_1, which is finite, we can compute the expectation values in finite volumes after all from a partition function in the form of Equation (12) but with the parameter $\bar{\theta}$ set to zero. This way, the logarithm of the partition function can be taken as an extensive quantity.

In the previous derivation, when going from Equation (48) to (49) we made use of the result for the partition function in the dilute instanton gas approximation, Equation (41). However, it is worth pointing out that the latter result can be derived from the cluster decomposition principle alone, without making use of the dilute instanton gas approximation. One can start by noting that the factorization of the path integrals in the denominator in Equation (46) can be written in terms of the following relations between the partition functions $Z_{\Delta n}(\Omega)$ in the full volume and their counterparts $Z_{\Delta n_1}(\Omega_1)$, $Z_{\Delta n_2}(\Omega_2)$ for the subvolumes Ω_1, Ω_2,

$$Z_{\Delta n}(\Omega) = \sum_{\Delta n_1} Z_{\Delta n_1}(\Omega_1) Z_{\Delta n - \Delta n_1}(\Omega_2)\,. \tag{51}$$

Equation (51) is an infinite set of identities that can be used to solve for $Z_{\Delta n}(\Omega)$ from a set of minimal assumptions. First, we note that $Z_{\Delta n}(\Omega)$ are complex. For starters, they receive a phase $\mathrm{e}^{\mathrm{i}\theta\Delta n}$ due to the θ term. Further complex phases in $Z_{\Delta n}(\Omega)$ can only come from the phases α_i of the fermion masses. At least at the leading order, the fermionic path integration yields determinants of the massive Dirac operator in a background of topological charge Δn, which can be fully general and is not assumed to be precisely captured by the dilute instanton gas approximation. The phase of the total fermionic determinant is then fixed by the Atiyah–Singer index theorem [12] and for a single fermion is given by $e^{i\alpha\Delta n}$. As a consequence, one can write

$$Z_{\Delta n}(\Omega) = e^{i(\theta+\alpha)\Delta n}\tilde{g}_{\Delta n}(\Omega)\,, \quad \tilde{g}_{\Delta n}(\Omega) \in \mathbb{R}\,. \tag{52}$$

Parity considerations and appropriate limits of the cluster-decomposition relation (51) can be used to motivate the simple ansatz [7]

$$\tilde{g}_{\Delta n}(\Omega) = \Omega^{|\Delta n|} f_{\Delta n}(\Omega^2), \quad f_{\Delta n}(0) \neq 0\,. \tag{53}$$

Notably, the previous ansatz together with the assumption of analyticity in Ω give rise to a unique solution for the infinite tower of identities in Equation (51), which can be written as

$$Z_{\Delta n}(\Omega) = e^{i(\theta+\alpha)} I_{\Delta n}(2\beta\Omega)\,, \quad \beta \equiv f_{\Delta n=1}(0) \in \mathbb{R}\,. \tag{54}$$

As advertised, this recovers the result of Equation (41) without making use of the dilute instanton gas approximation.

Equation (54) can be taken even further, as it allows one to rederive the phases of fermionic correlators and confirm the conclusions of this section without using the dilute instanton gas approximation. Defining a complex mass parameter $\mathfrak{m}$ as

$$\mathfrak{m} \equiv m e^{i\alpha}\,, \tag{55}$$

the mass terms in the Lagrangian of Equation (1) can be written as

$$\mathcal{L} \supset \mathfrak{m}\bar{\psi}P_R\psi + \mathfrak{m}^*\bar{\psi}P_L\psi\,. \tag{56}$$

Then, one can view the complex mass parameters as sources for integrated correlators,

$$\frac{\partial}{\partial\mathfrak{m}} Z_{\Delta n} = -\int d^4x\, \langle\bar{\psi}P_R\psi\rangle_{\Delta n}\,, \quad \frac{\partial}{\partial\mathfrak{m}^*} Z_{\Delta n} = -\int d^4x\, \langle\bar{\psi}P_L\psi\rangle_{\Delta n}\,. \tag{57}$$

As the partition functions of Equation (54) have been derived on general grounds, the previous correlators are meant to include nonperturbative effects. Noting that the reality condition in the β parameter of Equation (54) implies $\beta = \beta(\mathfrak{m}\mathfrak{m}^*)$ and writing $\alpha = -(i/2)\log(\mathfrak{m}/\mathfrak{m}^*)$ yields the following spacetime-averaged correlators [7]

$$\begin{aligned}
\frac{1}{\Omega}\int d^4x\, \langle\bar{\psi}P_R\psi\rangle &= \lim_{N\to\infty}\lim_{\Omega\to\infty} \frac{\sum_{|\Delta n|<N}\int d^4x\, \langle\bar{\psi}P_R\psi\rangle_{\Delta n}}{\Omega\sum_{|\Delta m|<N} Z_{\Delta m}} = -2\mathfrak{m}^*\,\partial_{\mathfrak{m}\mathfrak{m}^*}\beta(\mathfrak{m}\mathfrak{m}^*)\,,\\
\frac{1}{\Omega}\int d^4x\, \langle\bar{\psi}P_L\psi\rangle &= \lim_{N\to\infty}\lim_{\Omega\to\infty} \frac{\sum_{|\Delta n|<N}\int d^4x\, \langle\bar{\psi}P_L\psi\rangle_{\Delta n}}{\Omega\sum_{|\Delta m|<N} Z_{\Delta m}} = -2\mathfrak{m}\,\partial_{\mathfrak{m}\mathfrak{m}^*}\beta(\mathfrak{m}\mathfrak{m}^*)\,.
\end{aligned} \tag{58}$$

It is readily seen that the total phase of the fermionic correlators, including nonperturbative effects, is aligned with the phases of the tree-level masses in the Lagrangian. This generalizes the result of Equation (42) and leads again to the conclusion of no CP violation. Again, the order of limits plays a crucial role in Equation (58).

7. Effective Theories and Effective Operators

We shall now draw the connection from the results of the semiclassical approximation corresponding to integrating out gluons from Section 5 with observables probing CP conservation or violation in strong interactions. The main object of interest in that context is the 't Hooft vertex, which can be inferred from the correlation function (42) as the Lagrangian term

$$-\bar{\psi}(x)\Gamma e^{i\alpha\gamma^5}\psi(x)\,. \tag{59}$$

This vertex generates the same correlation functions as in Equation (42) for the EFT where gluons have been integrated out. Figure 6 illustrates how such a model fits into the picture of the different EFTs discussed in the present context. In addition, there will also be in general nonlocal operators from the long-range interactions of the gluons, because with quark degrees of freedom still in the theory, there is no cutoff parameter that allows for a local expansion. The new operators appear in favor of the gluon kinetic term $F_{\mu\nu}F_{\mu\nu}$ as well as the topological term $F_{\mu\nu}\widetilde{F}_{\mu\nu}$, which disappear together with the gluons.

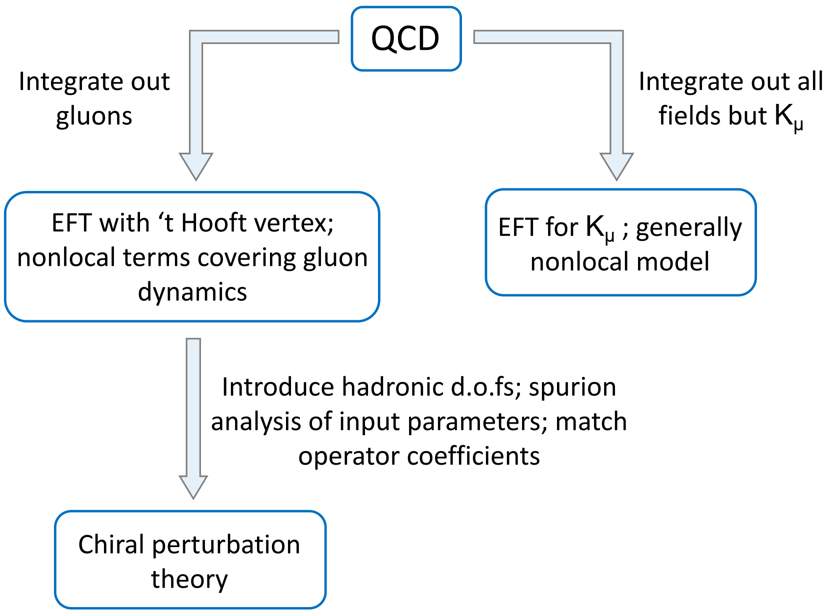


Figure 6. Schematic representation of the different effective field theories (EFTs) discussed in the present article and their relation with QCD. Note that the nonlocal operators due to gluon dynamics are all but under calculational control.

In the dilute instanton gas approximation in Section 5, one has integrated out gluons in the semiclassical approximation. For this to be valid, the theory should be perturbative throughout, which can be achieved in principle by adding a bespoke matter content that controls the renormalization group evolution in this peculiar way. Certainly, however, this is neither the case for the theory specified in Equation (1) with the gauge group SU(2) and one quark flavor nor for QCD with SU(3) and three flavors of light quarks. As a consequence, one should expect substantial deviations from the correlation function given in Equation (42), in particular for large distances between x and x'. Nonetheless, there should still be a small distance, high energy contribution of this form. When drawing conclusions about CP conservation or violation, one therefore must make the assumption that the CP-odd coefficient of the 't Hooft vertex appears in the same way within the extra operators that have to be added in principle to account for the low-energy behavior. Note, however, that this shortcoming applies to the conclusions based on either order of limits when the calculation is carried out semiclassically.

To some extent, the above matter is addressed by the concluding argument from Section 5, where the leading fermion correlations are constrained without the dilute instanton gas approximation but using instead cluster decomposition and the index theorem. There, no assumption about the fermion correlation is made but for its $U(1)_A$-violating form. While the resulting fermion correlation then can only be stated in the coincident limit, the conclusions about CP conservation based on the order of the infinite volume limit and the sum over topological sectors should therefore extend to the nonperturbative low-energy regime as well.

The underlying theory that we are concerned with after all is QCD, which is specified (now with the gauge group $\mathrm{SU}(3)$, N_f flavors of light quarks and in *Minkowski* spacetime) as (we choose $\varepsilon_{1230} = +1$ in Minkowski spacetime)

$$\mathcal{L}_{\mathrm{M}} \supset -\frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}F^{\mu\nu} + \sum_{i,j=1}^{N_f} \bar{\psi}_i\left(\mathrm{i}\gamma^\mu D_\mu\delta_{ij} - M_{ij}P_{\mathrm{R}} - M_{ij}^\dagger P_{\mathrm{L}}\right)\psi_j + \frac{1}{16\pi^2}\theta\,\mathrm{tr}F_{\mu\nu}\widetilde{F}^{\mu\nu}\,, \quad (60)$$

where in the mass-diagonal basis

$$M_{ij} = \delta_{ij}m_j \quad \text{(no sum over } j)\,. \quad (61)$$

For the model as in Equation (60), the vertex corresponding to Equation (59) is

$$-\Gamma_{N_f}\mathrm{e}^{-\mathrm{i}\bar{\alpha}}\prod_{j=1}^{N_f}(\bar{\psi}_jP_{\mathrm{L}}\psi_j) - \Gamma_{N_f}\mathrm{e}^{\mathrm{i}\bar{\alpha}}\prod_{j=1}^{N_f}(\bar{\psi}_jP_{\mathrm{R}}\psi_j)\,, \quad (62)$$

where

$$\bar{\alpha} = \sum_{i}^{N_f}\alpha_i = \arg\det M\,. \quad (63)$$

We need to sort in what way (cf. Figure 6) this is connected to the EFT of hadrons that is valid at low energies and should describe those possible CP-violating effects that are accessible by current precision experiments. A principal obstacle to systematically deriving quantitative predictions lies of course within the circumstance that perturbation theory is not valid anymore at low energies.

Yet, the symmetries, even when realized approximately only, offer a standard method of constraining the EFT. In the fundamental theory, as well as on the EFT side, one can introduce operators of the physical fields coupled to external sources (sometimes called spurions) so that these operators are invariant under local symmetry transformations. On the side of the EFT, the coefficient of these operators has to be obtained through computational or experimental matching. Variation with respect to these sources then allows one to express matrix elements of the fundamental theory in terms of parameters of the EFT.

In the present case, we can apply this method by perceiving the quark masses that break the chiral flavor symmetries $\mathrm{SU}(N_f)_A$ as well as the operators breaking $\mathrm{U}(1)_A$ as external sources that transform according to these explicitly broken symmetries. In the following discussion, we occasionally let $N_f = 2$, meaning that only up and down quarks are considered, for simplicity. But for expressions explicitly depending on N_f, we keep N_f general. First, we parametrize a chiral transformation as

$$\psi \to LP_{\mathrm{L}}\psi + RP_{\mathrm{R}}\psi\,, \quad (64)$$

where $P_{\mathrm{L,R}} = (1 \mp \gamma^5)/2$ and L, R are independent unitary matrices. For an axial transformation, $R = L^{-1}$ so that the $\mathrm{SU}(2)_A$ transformations are given by

$$\psi_i \to [\mathrm{e}^{\mathrm{i}\vec{\gamma}\cdot\vec{\sigma}\gamma^5}]_{ij}\psi_j\,. \quad (65)$$

The Lagrangian (60) would remain invariant if the mass matrix transformed as

$$M \to LMR^\dagger = \mathrm{e}^{-\mathrm{i}\vec{\gamma}\cdot\vec{\sigma}}M\mathrm{e}^{-\mathrm{i}\vec{\gamma}\cdot\vec{\sigma}}\,. \quad (66)$$

In this transformation, M corresponds to a spurion field.

The corresponding EFT Lagrangian (cf. Figure 6) with the lowest-order terms is (see, e.g., Refs. [8,39,40] where the effective theory is derived from integrating out quark fields)

$$\mathcal{L}_{\mathrm{M}}^{\mathrm{EFT}} = \frac{f_\pi^2}{4}\mathrm{Tr}\,\partial_\mu U \partial^\mu U^\dagger + \frac{f_\pi^2 B_0}{2}\,\mathrm{Tr}(MU + U^\dagger M^\dagger) + |\lambda| \mathrm{e}^{-\mathrm{i}\xi} f_\pi^4 \det U + |\lambda| \mathrm{e}^{\mathrm{i}\xi} f_\pi^4 \det U^\dagger\,, \tag{67}$$

where

$$U = U_0 \mathrm{e}^{\frac{\mathrm{i}}{f_\pi}\Phi} = U_0 \tilde{U}\,, \quad U_0 = \langle U \rangle = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\varphi_u} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\varphi_d} \end{pmatrix}, \quad \Phi = \begin{bmatrix} \pi^0 + \eta' & \sqrt{2}\,\pi^+ \\ \sqrt{2}\,\pi^- & -\pi^0 + \eta' \end{bmatrix}. \tag{68}$$

In the equations above, f_π is the pion decay constant and λ, B_0 are EFT coefficients to be determined experimentally or computationally and B_0 is directly related to the magnitude of the chiral quark condensate. The phases of the latter correspond to φ_u, φ_d, and we have assumed a diagonal mass matrix M. The squared pion and η' masses are then given by

$$\begin{aligned} m_\pi^2 &= B_0(m_u \cos(\alpha_u + \varphi_u) + m_d \cos(\alpha_d + \varphi_d))\,, \\ m_{\eta'}^2 &\approx 16|\lambda| f_\pi^2 \cos(\xi - \varphi_u - \varphi_d)\,, \end{aligned} \tag{69}$$

where we have made the phenomenologically valid approximation that $m_\pi^2 \ll m_{\eta'}^2$. We leave the terms with the parameter λ aside just yet as $\det U$ is invariant under $\mathrm{SU}(2)_A$ but transforms with $\mathrm{U}(1)_A$ in a way that we shall approach shortly. In correspondence with Equation (65), the meson fields behave under axial transformations as

$$U \to RUL^\dagger = \mathrm{e}^{\mathrm{i}\vec{\gamma}\cdot\vec{\sigma}} U \mathrm{e}^{\mathrm{i}\vec{\gamma}\cdot\vec{\sigma}} \tag{70}$$

The term with the parameter B_0 should be matched so that the correct correlation functions are produced. Corresponding to the invariance of the underlying theory (60), the Lagrangian (67) is invariant under the simultaneous $\mathrm{SU}(2)_A$ transformations (66) and (70).

Now, after all, the (up and down) quark masses do not transform under $\mathrm{SU}(2)_A$, they rather break this symmetry explicitly. We can still perceive these as local sources though, that perturb the correlators of the theory about the case with full $\mathrm{SU}(2)_A$ symmetry. For the local source, we can then take the fixed physical values of M so that Equation (67) accounts for the perturbation through the quark masses to linear order. In the EFT, one can continue this to higher orders pending on the precision that is aimed for.

Now, consider $\mathrm{U}(1)_A$ transformations

$$\psi_i \to \mathrm{e}^{\mathrm{i}\beta\gamma^5}\psi_i \tag{71}$$

and recall the expression for $\bar{\alpha}$ from Equation (63). The fundamental theory (60) would remain invariant if the quark mass transformed as

$$M \to \mathrm{e}^{-2\mathrm{i}\beta} M\,, \quad \text{so that} \quad \bar{\alpha} \to \bar{\alpha} - 2\beta N_f\,. \tag{72}$$

The chiral anomaly requires that the coefficient θ of the topological term goes as

$$\theta \to \theta + 2N_f\beta \tag{73}$$

in order to keep the Lagrangian invariant. Note that this implies that the combination

$$\bar{\theta} = \theta + \bar{\alpha} \tag{74}$$

is invariant under chiral rephasings and in general is nonzero. The presence of such an invariant does however not yet guarantee that it leads to physical effects.

We thus see that there are two local sources that transform under the symmetry $\mathrm{U}(1)_A$: θ and $-\alpha$. Noting that under this symmetry

$$\det U \to \mathrm{e}^{2\mathrm{i}N_f\beta} \det U\,, \tag{75}$$

the EFT Lagrangian (67) remains invariant if either

$$\xi = \begin{cases} -\bar{\alpha} \\ \theta \end{cases} . \tag{76}$$

In principle, one may also allow linear combinations of the parameters $-\bar{\alpha}$ and θ. As this does not follow from either order of limits for the sum over topological sectors and spacetime volume that we discuss here, we do not consider this combination option further.

We also note that the operator with the coefficient B_0 breaks $\mathrm{U}(1)_A$. Therefore, instead of the quark mass phase in M, one could also use θ to write this as an invariant operator with the help of chiral-variant source fields. However, the symmetric theory should respond to $\mathrm{U}(1)_A$-breaking perturbations through a quark mass term in the same way as it does for $\mathrm{SU}(2)_A$-breaking. In this sense, the term with B_0 is unique to linear order in M. The explicit breaking of $\mathrm{U}(1)_A$ through instantons is independent of the quark masses, cf. Equation (42) together with the fact that $\kappa = \mathcal{O}(|M|)$, and, therefore, M does not appear in the terms with $\det U$.

Now, recall that Equation (42) leads to the effective vertex (62) in the theory where gluons have been integrated out. At this level, θ has disappeared so that the only option for the EFT Lagrangian (67) is

$$\xi = -\bar{\alpha} . \tag{77}$$

The CP-odd coefficients can then be removed by an overall field redefinition. On the other hand, if it were $\xi = \theta$, there would be a residual CP-odd term.

Further, note that Equation (69) shows that the mass of the η' in general does not vanish in the limit of $m_{u,d} \to 0$, no matter which of the values ξ takes in Equation (76). In turn, the fact that the η' is heavy compared to the pions as such does not lead to a conclusion about which is the correct order of limits.

Finally, the parameter ξ in the coefficient of the 't Hooft operator enters the calculation of the nucleon EDM as follows: Given the EFT Lagrangian (67) and choosing a basis in which M is diagonal, the minimum of the field U is given by U_0 as in Equation (68), where, in the limit of $|\lambda| \gg B_0 m_d / f_\pi^4$, and for $\xi + \alpha_u + \alpha_d$ in the first quadrant, one has [41,42]

$$m_u \sin(\varphi_u + \alpha_u) = m_d \sin(\varphi_d + \alpha_d) = \frac{\sin(\xi + \alpha_u + \alpha_d)}{\sqrt{\frac{1}{m_u^2} + \frac{1}{m_d^2} + \frac{2\cos(\xi+\alpha_u+\alpha_d)}{m_u m_d}}} . \tag{78}$$

Going beyond the assumption $|\lambda| \gg B_0 m_d / f_\pi^4$ leads to a mixing of the flavor eigenstates π^0 and η' within the mass eigenstates.

In order to expand in terms of the meson fields, following Equations (68) and (78) leads to

$$\begin{aligned} \mathrm{Tr}[MU + U^\dagger M^\dagger] =& \frac{1}{2}[m_u \cos(\varphi_u + \alpha_u) + m_d \cos(\varphi_d + \alpha_d)]\mathrm{Tr}[\tilde{U} + \tilde{U}^\dagger] \\ &+ \frac{\mathrm{i}}{2}[m_u \sin(\varphi_u + \alpha_u) + m_d \sin(\varphi_d + \alpha_d)]\mathrm{Tr}[\tilde{U} - \tilde{U}^\dagger] \\ &+ \text{terms mixing } \pi^0 \text{ with } \eta' . \end{aligned} \tag{79}$$

The operators in the second line are CP-odd, and we note that

$$\mathrm{Tr}\left[\tilde{U} - \tilde{U}^\dagger\right] = \frac{2\mathrm{i}}{f_\pi}\mathrm{Tr}[\Phi] - \frac{\mathrm{i}}{3f_\pi^3}\mathrm{Tr}\left[\Phi^3\right] = \frac{4\mathrm{i}}{f_\pi}\eta' - \frac{\mathrm{i}}{f_\pi^3}\left(\frac{2}{3}\eta'^3 + 2\eta'\left[(\pi^0)^2 + 2\pi^+\pi^-\right]\right) . \tag{80}$$

Substituting this into the term with B_0 in Equation (67) and expanding in the meson fields, one generally would obtain CP-violating effects if $\xi \neq -\bar{\alpha}$, the most immediate consequence of which would be $\eta' \to 2\pi$ (recall that the meson fields are CP-odd) through the interaction term

$$\begin{aligned}\mathcal{L}_{\mathrm{M}}^{\mathrm{EFT}} \supset & \frac{B_0 \sin(\xi + \alpha_u + \alpha_d)}{f_\pi \sqrt{\frac{1}{m_u^2} + \frac{1}{m_d^2} + \frac{2\cos(\xi+\alpha_u+\alpha_d)}{m_u m_d}}} \Big[(\pi^0)^2 + 2\pi^+\pi^- \Big] \eta' \\ \approx & \frac{m_u m_d (\xi + \alpha_u + \alpha_d)}{(m_u + m_d)^2} \frac{m_\pi^2}{6 f_\pi} \mathrm{Tr}[\Phi^3] \,. \end{aligned} \tag{81}$$

The latter expression (which follows when assuming $\xi + \alpha_u + \alpha_d \ll 1$) is shown here for comparison with Equation (8) of Ref. [43]. In the latter, the quark masses are taken as real (see Equation (5) in [43]), which means that the parameter θ of that reference should correspond to $\pm\bar{\theta}$ in Equation (74). Matching the resulting signs for the phases of the quark masses in Equation (8) of Ref. [43] leads to an identification with $-\bar{\theta}$. It then follows that up to the central issue that which value in Equation (76) is taken by ξ, the results from the present EFT description and the partially conserved axial currents in Ref. [43] are therefore in agreement, as they should be. We further compare the two approaches given different values of ξ in Section 8. Finally, let us note also that the coefficient in Equation (81) is different in the approximation of three light flavors where an extra factor of $\sqrt{2/3}$ occurs.

To see what the above CP-odd interactions of the pions and η' would imply for the nucleons, one can add their interactions to the EFT Lagrangian as

$$\mathcal{L}_{\mathrm{M}}^{\pi,\mathrm{p,n}} \supset \mathrm{i}\bar{N}\partial\!\!\!/ N - \Big(m_N \bar{N} U P_{\mathrm{L}} N + \mathrm{i}c \bar{N} U^\dagger \partial\!\!\!/ U P_{\mathrm{L}} N + d \bar{N} M^\dagger P_{\mathrm{L}} N + e \bar{N} U M U P_L N + \mathrm{h.c.} \Big) \,, \tag{82}$$

where the nucleon doublet transforms as

$$N \to L\, P_{\mathrm{L}} N + R\, P_{\mathrm{R}} N \,. \tag{83}$$

Again, promoting M to a source that transforms under the axial symmetries rather than breaking these, this Lagrangian is invariant. Substituting the expectation value of the chiral condensate (68), (78) for small $\xi + \bar{\alpha}$, expanding in the meson field and applying field redefinitions $N \to \mathcal{N}$ so as to obtain the canonically normalized flavor eigenstates of the nucleons, one finds the interaction terms [41]

$$\mathcal{L}_{\mathrm{M}}^{\mathrm{neutron}} \supset c_1 \partial_\mu \pi^a \bar{\mathcal{N}} \frac{\tau^a}{2} \gamma^\mu \gamma^5 \mathcal{N} + c_2 (\xi + \bar{\alpha}) \bar{\mathcal{N}} \pi^a \frac{\tau^a}{2} \mathcal{N} \,. \tag{84}$$

The first of these is CP-even, as it couples two axial currents, π^a being a pseudoscalar field. The second term is CP-odd, as it couples a scalar density with a pseudoscalar field. At one loop level, if it were $\xi \neq -\bar{\alpha}$, this would induce an EDM through the famous diagrams shown in Figure 7. Note that the weak interactions make an additional contribution to the neutron EDM [44,45], which, however, is too small and usually neglected in the discussion of CP in strong interactions.

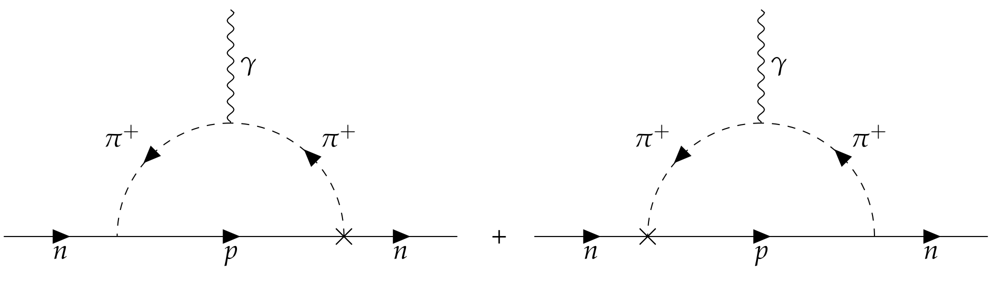


Figure 7. Leading order contribution to the neutron EDM from the strong interactions for $\xi \neq -\bar{\alpha}$. The CP-violating vertex is indicated by a cross. A similar diagram involves π^-.

8. Some Objections and Answers to These

We review here and reply to some objections that we have been made aware of, mostly to the extent that these can be related to (partly earlier) articles or to conference talks that have been published online.

8.1. Instanton Configurations in the Different Limits

As argued in Section 3, the different orders of limits correspond to a partitioning and rearrangement of the integration contour that leads to inequivalent results for the path integration. Still, upon taking the limits, the path integral covers the same field configurations. Yet, arguments have been put forward that partitioning the contour should be harmless or that the order of limits as in Equations (9) and (11) does not include all relevant contributions as opposed to Equations (12) and (13) [46].

Regarding the partitioning of the contour, one may state that for a given configuration of finite action, one can find a radius R so that [11]

$$A_\mu(x) = \mathrm{i}\omega(x)\partial_\mu\omega^{-1}(x) + \mathcal{O}(1/R^2) \tag{85}$$

and, consequently,

$$\Delta n = \frac{1}{16\pi^2}\int\limits_{|x|<R} \mathrm{d}^4x\,\mathrm{tr}F\widetilde{F} = \text{integer} + \mathcal{O}(1/R)\,. \tag{86}$$

Therefore, in this sense, even for finite volumes, one can at least approximately categorize certain configurations by an integer Δn. However, R is not universal and not even a function of Δn since the individual units of winding number can be separated arbitrarily far for given Δn. Therefore, we cannot take this as an argument for the calculation in fixed volumes Ω to be equivalent to the full result.

Concerning Equation (11) perhaps not accounting for all relevant configurations, one may attempt to argue as follows: Consider the dilute instanton gas picture and let ρ be the radius of an instanton. Then, demanding that the instantons and anti-instantons do not overlap, the maximum number of instantons and anti-instantons $n+\bar{n}$ satisfies $|\Delta n| \leq N \leq (n+\bar{n})_{\max} \approx \Omega/\rho^4$. Therefore, there is room to take $N \to \infty$ as $\Omega \to \infty$, and it seems not to be appropriate to cap Δn by some finite value in an infinite volume as Equation (11) suggests.

However, such a cap is not forced by Equation (11) in the following sense: Decompose the full spacetime into subvolumes, e.g., $\Omega = \Omega_1 \cap \Omega_2$. For each topological sector Δn, there is a constraint $\Delta n_1 + \Delta n_2 = \Delta n$, where $\Delta n_1, \Delta n_2$ are the winding numbers (which may not be integers precisely) in the subvolumes, but no constraint on Δn_1 and Δn_2 separately. Therefore, Δn_1 can be arbitrarily large. For an observer in a subvolume, say Ω_1, the path integral (11) therefore includes configurations of arbitrarily large winding number density $\Delta n_1/\Omega_1$ within Ω_1. Hence, there is no cap on Δn_1 in the finite subvolume Ω_1. Indeed, this is already clear from Equation (50), which is derived from the cluster decomposition principle.

8.2. Chiral Limit

In a theory with at least one massless quark, the parameter $\bar{\alpha}$ can be chosen arbitrarily with no consequence to the Lagrangian. Without further ado, this implies that $\bar{\theta}$ in Equation (8) cannot be physical and that there is no CP violation in such a model. In the dilute instanton gas picture, this behavior results from the suppression of single instantons through the zero-mode from the fermion determinant that makes the factor ϖ in Section 5 and consequently κ vanishes proportionally to the absolute value of the quark mass determinant.

In this sense, as $\kappa \sim m$, we can take in Equation (42) $m \to 0$ and obtain a well-defined limit. Since the gluons and therefore the topological term have been integrated out, the parameter α in this expression is still arbitrary but unphysical, as it can be removed by a chiral rotation of the fermion field. The same reasoning applies to the effective operator (62).

8.3. Effective Theory for the Topological Current

We respond here to comments concerning Ref. [7] that have been made in Ref. [47], see also Ref. [48]. It is argued there that finite spacetime volumes Ω lead to an unphysical breaking of the conservation of the winding number Δn (or the density $\Delta n/\Omega$) so that the absence of CP violation would be an artifact of such regularization. But, apparently, in Equation (9) and consequently Equation (11) that lead us here as well as in Ref. [7] to the conclusion of no CP violation, $\Omega \to \infty$ is the first limit that is taken. This is in contrast to Equations (12) and (13), which lead to CP-violating observables but where the topological sectors are fixed prior to taking $\Omega \to \infty$. Therefore, the criticism of producing finite volume artifacts would rather be an issue for the latter prescription, and, in fact, it is, as we have discussed in the previous sections.

While this comes as a rather immediate conclusion, it is of interest to see how the EFT for the topological current, which is introduced in Refs. [47,49,50], fits into the present considerations of imposing boundary conditions and orders of limits. (See Figure 6 for where it stands in relation to the other EFTs that are discussed here.) The topological current can be defined as

$$K_\mu = \frac{1}{4\pi^2}\epsilon_{\mu\nu\alpha\beta}\mathrm{tr}\left[\frac{1}{2}A_\nu\partial_\alpha A_\beta - \frac{\mathrm{i}}{3}A_\nu A_\alpha A_\beta\right], \tag{87}$$

where we recall $\mathrm{tr}(T^aT^bT^c) = \mathrm{i}f^{abc}/4$ for $\mathrm{SU}(2)$. The topological charge density and hence the topological term can be explicitly written as a total divergence

$$q = \frac{1}{16\pi^2}\mathrm{tr}F_{\mu\nu}\widetilde{F}_{\mu\nu} = \partial_\mu K_\mu\,. \tag{88}$$

One of the interesting points concerning the current K_μ is the form of its two-point correlation. Some information can be extracted from the chiral susceptibility

$$\begin{aligned}\chi_\Omega =&\frac{1}{\Omega}\,\langle\Delta n^2\rangle\Big|_{\bar\theta=1/2(1-(-1)^{N_f})\pi}\\ =&\frac{1}{\Omega}\left\langle\left(\int_\Omega \mathrm{d}^4x\,q(x)\right)^2\right\rangle\Bigg|_{\bar\theta=1/2(1-(-1)^{N_f})\pi} = \int_\Omega \mathrm{d}^4x\,\langle q(x)q(0)\rangle|_{\bar\theta=1/2(1-(-1)^{N_f})\pi}\,.\end{aligned} \tag{89}$$

We evaluate this for the CP-even values of $\bar\theta = 1/2(1-(-1)^{N_f})\pi$. When summing over topological sectors in a finite spacetime volume with fixed boundary conditions according to Equations (12) and (13), the vacuum energy is minimized at the chosen value of $\bar\theta$ and χ remains positive. Choosing instead the starting point Equation (9) and consequently Equation (11), the value of $\bar\theta$ is irrelevant by the arguments of Sections 5 and 6. We attach a subscript on χ in order to indicate that it is important which volume is referred to. Significant differences can arise when Ω corresponds to a subvolume of the spacetime as opposed to the full spacetime, in particular when boundary conditions control the overall topological fluctuations. For expressions that apply to all volumes, we omit the subscript.

One should also pay attention to the fact that χ will in general have connected and disconnected parts. If there is CP violation in a certain setup, e.g., imposed by unphysical boundary conditions, then $\langle\Delta n\rangle/\Omega \neq 0$. When one aims to characterize the volume-scaled variance of the topological charge, one should then subtract the disconnected contributions, i.e., consider

$$\chi_\Omega - \frac{\langle\Delta n\rangle^2}{\Omega}\,. \tag{90}$$

Of course, one may also define χ without the disconnected pieces to start with, which we do not do here for the sake of simpler expressions.

Applying the result (50) to Equation (89) yields

$$\begin{aligned}\chi_{\Omega_1} \equiv& \frac{1}{\Omega_1}\left\langle \left(\int_{\Omega_1} \mathrm{d}^4x\, q(x)\right)^2 \right\rangle \\ =& \frac{1}{\Omega_1} \frac{\sum\limits_{\Delta n_1=-\infty}^{\infty} \int \mathcal{D}A_{\Delta n_1}\mathcal{D}\bar{\psi}\mathcal{D}\psi \int_{\Omega_1} \mathrm{d}^4z\, q(z) \int_{\Omega_1} \mathrm{d}^4z'\, q(z')\mathrm{e}^{-S_{\Omega_1}[A_\mu]}(-1)^{-N_f\Delta n_1}\mathrm{e}^{-\mathrm{i}\,\bar{\alpha}\Delta n_1}}{\sum\limits_{\Delta n_1=-\infty}^{\infty} \int \mathcal{D}A_{\Delta n_1}\mathcal{D}\bar{\psi}\mathcal{D}\psi \mathrm{e}^{-S_{\Omega_1}[A_\mu]}(-1)^{-N_f\Delta n_1}\mathrm{e}^{-\mathrm{i}\,\bar{\alpha}\Delta n_1}} \\ =& \frac{1}{\Omega_1} \frac{\sum\limits_{n_1=0}^{\infty}\sum\limits_{\bar{n}_1=0}^{\infty} \frac{1}{n_1!\bar{n}_1!}(n_1-\bar{n}_1)^2(\kappa\Omega_1)^{n_1+\bar{n}_1}}{\sum\limits_{n_1=0}^{\infty}\sum\limits_{\bar{n}_1=0}^{\infty} \frac{1}{n_1!\bar{n}_1!}(\kappa\Omega_1)^{n_1+\bar{n}_1}} = 2\kappa\,. \end{aligned} \tag{91}$$

That is, when calculating the susceptibility from the partition function (9) but evaluating Equation (89) in a finite subvolume Ω_1 of infinite Euclidean spacetime, χ_{Ω_1} is nonzero. On the other hand, as $\lim_{\Omega\to\infty} \Delta n/\Omega = 0$, in infinite volume we have

$$\chi_{\Omega\to\infty} = 0\,. \tag{92}$$

One next introduces the Fourier transforms of the correlation functions, which we indicate here by a tilde:

$$\widetilde{\langle qq\rangle}(p) = \int \mathrm{d}^4(x-y)\,\mathrm{e}^{\mathrm{i}p(x-y)}\langle q(x)q(y)\rangle = p_\mu p_\nu \widetilde{\langle K_\mu K_\nu\rangle}(p)\,. \tag{93}$$

From Equation (89), one can conclude the following infrared behavior [51]:

$$\lim_{p\to 0} p_\mu p_\nu \widetilde{\langle K_\mu K_\nu\rangle}(p) = \int \mathrm{d}^4x\,\langle \partial_\mu K_\mu(x)\partial_\nu K_\nu(y)\rangle = \chi\,. \tag{94}$$

The former equation implies, e.g., in transverse gauge [51], where $\partial_\mu\epsilon_{\mu\nu\rho\sigma}A_\sigma = 0$,

$$\widetilde{\langle K_\mu K_\nu\rangle}(p) = \frac{\chi\, p_\mu p_\nu}{p^4} + \mathcal{O}(p^2) \tag{95}$$

so that one would expect for $\chi \neq 0$ a simple massless pole in $\widetilde{\langle K_\mu K_\nu\rangle}(p)$ for $p_\mu \to 0$. Here, we have further dropped the spacetime volume subscript on χ.

Now, per Equation (87), K_μ can be understood as the Hodge dual of a three-form field. Therefore, it is of interest to compare it with the action for a massless three-form L_μ,

$$S = \int_\Omega \mathrm{d}^4x\,(\partial_\mu L_\mu)^2\,. \tag{96}$$

The expression "massless" refers here to the absence of a mass term in the action and is not supposed to indicate that there is a massless propagating degree of freedom, which in fact there is not. The variation of this action is given by

$$\delta S = -2\int_\Omega \mathrm{d}^4x\,(\partial_\nu\partial_\mu L_\mu)\delta L_\nu + 2\int_{\partial\Omega} \mathrm{d}a_\nu\,(\partial_\mu L_\mu)\delta L_\nu\,, \tag{97}$$

where a_ν is a normal surface element on Ω. Discarding the boundary term for the moment, one would conclude that $\partial_\mu L_\mu = \text{const}$. But then again, dropping the boundary term is in general not justified because it in general contributes to the equations of motion. The situation here is different from the model specified by the Lagrangian (1) in an infinite spacetime volume Ω: There, solutions of finite action exist while the topological term does not need to vanish and can be written per Equations (10) and (88) as a boundary term. However, within a given topological sector the boundary term gives a fixed contribution $-\mathrm{i}\theta\Delta n$ to the action. Therefore, it has no impact on the equations of motion.

Back to the action (96), we must therefore make further assumptions about the boundary term. While there are many different options, we follow Ref. [50] that works with solutions

$$\partial_\mu L_\mu = \theta_L \,, \tag{98}$$

where θ_L is an integration constant. While not stated explicitly in Ref. [50], this appears to require boundary conditions $\delta L_\mu = 0$ on $\partial\Omega$ and

$$\theta_L = \frac{1}{\Omega}\int_{\partial\Omega} \mathrm{d}a_\nu\, L_\nu \,. \tag{99}$$

Interpreting for the moment L_μ as a four-dimensional electric field in the background of a constant charge density θ_L, it should be clear that Equation (98) has solutions that satisfy the condition (99). Note that for a given θ_L, $L_\mu \sim \theta_L \Omega^{1/4}$. To keep the boundary conditions well defined in terms of finite L_ν, we therefore have to impose these on finite Ω or we have to take $\theta_L = 0$.

Given the boundary condition $\delta L_\mu = 0$, we can also derive the equation of motion

$$\partial^x_\mu \partial^x_\nu \langle L_\nu(x) L_\rho(y)\rangle = -\delta_{\mu\rho}\delta^4(x-y) \tag{100}$$

for the correlation function. It further implies

$$\langle(\partial_\mu L_\mu(x))(\partial_\nu L_\nu(y))\rangle = 4\delta^4(x-y) + \text{const.} \tag{101}$$

To arrive at the above equation, one uses the translation invariance to deduce that the correlator is a function of $(x-y)$. Then, one can take $\partial^x_\mu = -\partial^y_\mu$ and contract μ and ρ in Equation (100).

Having stated all this, it is interesting to follow Ref. [50] and consider the action (96) as an EFT of the full model (1), of which all degrees of freedom but K_μ have been integrated out. The effective action should generally feature higher-order terms in $\partial_\mu L_\mu$ beyond Equation (96) as well as nonlocal expressions that account for the rich phenomenology of QCD compared to the present simplistic model [50]. Yet, it appears that the following characteristic features of QCD are recovered.

Equation (101) indicates that the topological fluctuations correspond to white noise. In that sense, the model (96) without higher-order and nonlocal terms captures the infrared behavior on scales where the fluctuations are indeed expected to be uncorrelated. We may use the δ-function term in relation (101) to identify

$$L_\mu = \frac{2K_\mu}{\sqrt{\chi}} \,. \tag{102}$$

Indeed, substituting the above into Equation (101) (ignoring the constant term for the moment) and taking a Fourier transform, one recovers Equation (94).

On the basis of these correspondences, the argument of Ref. [50] then goes that the integration constant θ_L selects a vacuum state with a nonzero value of the CP-odd observable $\partial_\mu L_\mu$ that it is therefore proportional, at least at the linearized level, to the parameter θ in the microscopic theory, which would thus turn out to be physical. However, as we shall see next, we disagree with this interpretation. Because of the different boundary conditions, the EFT for L_μ cannot correspond to either of the QCD partition functions (12) or (9), due to reasons that will also make clear why θ_L is not proportional to an angle.

Translating the boundary condition (99) back to the fundamental theory implies that one fixes

$$\Delta n = \int_{\partial\Omega} \mathrm{d}a_\nu\, K_\nu = \frac{1}{16\pi^2} \int_\Omega \mathrm{d}^4x\, F\widetilde{F}\,. \tag{103}$$

Moreover, these boundary conditions do not imply that the physical fields vanish on $\partial\Omega$, and, in fact, per Equation (98), they do not for $\theta_L \neq 0$, because $\theta_L \propto \partial_\mu L_\mu \propto \partial_\mu K_\mu$, which is gauge invariant. The present setup is therefore not an EFT for the partition function specified through Equations (12) and (13). For the latter, one samples over all integer values of Δn and the physical fields vanish on $\partial\Omega$.

A consequence of this discrepancy is the fact that from Equations (12) and (13), one concludes that $\langle \Delta n \rangle/\Omega = 2\mathrm{i}(-1)^{\mathrm{N_f}}\kappa \sin\bar{\theta}$ is purely imaginary [7], whereas in the EFT and for the boundary conditions that are assumed here, Δn is given by the fixed real value (103). The two setups therefore are not only different, but they also predict distinctly inequivalent results for the observables. There should also be no relation between θ from the model of Equations (12) and (13) and θ_L, which explains why the EFT in terms of the latter does not show any periodic behavior, as would be required for an angular variable.

One should notice that while Δn is fixed over the full volume Ω, it will nonetheless fluctuate in any subvolume, cf. Equation (101). Still, the boundary conditions on $\partial\Omega$ are required, and these are very different from those in Equation (13). Further, if we interpret L_μ as an effective infrared field, nonvanishing $\partial_\mu L_\mu$ on $\partial\Omega$ does not conflict with vanishing microscopic fields on that surface. But following the above remark concerning real versus imaginary $\langle \Delta n \rangle$, the expectation values for Δn in Ω still do not agree between the EFT and the setup specified through Equations (12) and (13).

It is also clear now that the present EFT (96) with the boundary condition (99) is also not equivalent to the partition function (9) that we take as the object to theoretically define strong interactions. The EFT corresponds to imposing a fixed flux (99)—not necessarily an integer—on a finite surface. This is in contrast to the case of Equation (9), for which we see from Equation (11) that all integer values of the topological flux at infinity are accounted for in the action. Imposing a fixed flux on a finite surface cannot yield the correct vacuum correlations, not least since a general finite Euclidean $\partial\Omega$ does not have a geometrically meaningful continuation to some domain in Minkowski space.

Though, as just has been argued, the EFT is not compatible with the usual QCD partition functions, one may nevertheless point out that the topological susceptibility in the EFT can have a similar infrared behavior as in the theory (9). To see this, consider the case $\theta_L = 0$ for simplicity. The generalization can be carried out by accounting for disconnected contributions to the correlations in Equation (89). Since the topological flux Δn vanishes, this means by Equation (89) that $\chi_\Omega = 0$. In fact, this requirement fixes the constant term in Equation (101). This clearly resembles the behavior exhibited in Equation (92). For either setup, there is no problem with $\chi_\Omega = 0$ on the full volume Ω because there is no inhibition of topological fluctuations in subvolumes. Note that in the EFT for the topological current, for $\theta_L \neq 0$, it follows that $\chi_\Omega = \langle \Delta n \rangle^2/\Omega$ is purely disconnected (because the winding number in Ω is fixed through the flux on $\partial\Omega$) and still does not correspond to the value that is locally observed in a subvolume and includes connected contributions.

Now, in Ref. [47], it is suggested that the calculations in Ref. [7] (i.e., using the partition function (11)) are equivalent to choosing boundary conditions on a finite surface that lift the mass of the three-form, supposedly preventing to capture the true massless dynamics. However, as discussed above, the vanishing of the chiral susceptibility from the partition function (11) for the full volume in Equation (92) also occurs for the EFT for the topological current, up to disconnected contributions. In either case, this has the implication that the massless simple pole in $\langle \widetilde{K_\mu K_\nu} \rangle(p)$ disappears when evaluating Equation (94) for the full spacetime. In this sense, both models behave very similarly, and no conclusion can be drawn that the EFT captures dynamics that Equation (11) does not. In particular, one cannot support the statement from Ref. [47] that with Equation (11), one loses crucial information about CP-violating vacuum states based on some ill-behaved infrared behavior of, e.g., χ compared to the EFT.

Though also for the three-form EFT, Equation (94) does not yield a massless simple pole, it is of interest to give a more direct argument for why the presence of such a pole is not necessary for consistency with observed topological fluctuations to start with. First observe that Equation (94) is only true when the spacetime integral is taken over the full spacetime; otherwise, the Fourier transform is not simply projected into its zero momentum limit. To see this, consider a topological susceptibility defined by integrating over a finite subvolume $\Omega_1 \subset \Omega$

$$\chi_{\Omega_1} = \int_{\Omega_1} \mathrm{d}^4x \, \langle \partial_\mu K_\mu(x) \partial_\nu K_\nu(y) \rangle = \int_{\Omega_1} \mathrm{d}^4x \int \frac{\mathrm{d}^4p}{(2\pi)^4} \, \mathrm{e}^{-\mathrm{i}p(x-y)} p_\mu p_\nu \langle \widetilde{K_\mu K_\nu} \rangle(p) \,. \tag{104}$$

In an infinite spacetime Ω, when $\Omega_1 = \Omega$ is also infinite, the spacetime integral yields a delta-function and one recovers the result of Equation (94). For a finite $\Omega_1 \subset \Omega$, however, the topological susceptibility will receive contributions from the correlator $\langle \widetilde{K_\mu K_\nu} \rangle(p)$ evaluated at nonzero momentum.

Lattice calculations in frozen topologies (that have $\chi_\Omega = 0$ up to disconnected contributions) still observe nonzero topological susceptibility when sampling over a subvolume of the full lattice [52]. Hence, there is the logical possibility of a topological susceptibility that vanishes for the full spacetime yet is nonzero when defined over subvolumes. This is exactly what happens with QCD defined as in Equation (9), as can be seen in Equations (91) and (92). Now, with χ vanishing in the full spacetime volume, Equation (94) implies that the correlator $\langle \widetilde{K_\mu K_\nu} \rangle(p)$ cannot have a massless pole. In Ref. [47], it is stated without proof that this implies a massive pole instead. Let us adopt this assumption or more loosely introduce m_K as an infrared regulator that may or may not correspond exactly to the effect from the finite-volume cutoff on the correlation function. We may also have m_K stipulate the absence of the simple pole in Equation (95). Either way,

$$\langle \widetilde{K_\mu K_\nu} \rangle(p) = \frac{\chi_0 \, g_{\mu\nu}}{p^2 + m_K^2} \,. \tag{105}$$

Provided such a massive pole is consistent with nonzero χ_{Ω_1}, there is then no motivation for a massless pole for the three-form EFT of Ref. [47]. Hence, any conclusion derived from insisting on a massless pole in the EFT does not apply.

One should therefore demonstrate how one can reconcile the topological susceptibility being zero at infinite volume while remaining nonzero at finite subvolumes. It turns out that such behavior can actually be derived from the massive correlator of Equation (105). For estimating the susceptibility defined over a finite region Ω_R of radius R, one can consider an integration with a cutoff function e^{-R^2/r^2}, which suppresses the fluctuations for $r > R$,

$$\chi_{\Omega_\Lambda} \equiv \int \frac{\mathrm{d}^4p}{(2\pi)^4} \int_{\mathbb{R}^4} \mathrm{d}^4x \, \mathrm{e}^{-R^2/r^2} \mathrm{e}^{-\mathrm{i}p(x-y)} p_\mu p_\nu \langle \widetilde{K_\mu K_\nu} \rangle(p) \,. \tag{106}$$

Substituting Equation (105) and considering $y = 0$, the result is

$$\chi_{\Omega_R} = \chi_0 \left(1 - \frac{R^4 m^4}{16} \, \mathrm{e}^{\frac{R^2 m_K^2}{4}} \, \mathrm{Ei}\left(-\frac{1}{4} m_K^2 R^2 \right) - \frac{1}{4} \, R^2 m_K^2 \right) \sim \begin{cases} \chi_0, & R \ll \frac{1}{m_K}, \\ 0, & R \gg \frac{1}{m_K}. \end{cases} \tag{107}$$

Above, $\mathrm{Ei}(z) = -\int_{-z}^{\infty} \mathrm{e}^{-t}/t \, \mathrm{d}t$ is the exponential integral function.

As advertised, one obtains a zero topological susceptibility in the full volume but also constant susceptibility for subvolumes with associated length scales $L \ll 1/m_K$. This behavior is compatible with lattice results [52] and matches the result of Equations (91) and (92), which are based on the dilute instanton gas approximation.

8.4. Taking θ as a Perturbation and PCAC

Sometimes it is argued that θ can be seen to be physical without making topological considerations by referring to calculations of possible CP-violating effects using identities for partially conserved axial currents (PCAC), e.g., in the discussion of Section 3 of Ref. [53] or in Ref. [43]. (Note though that these papers do not state explicitly that their calculations would allow them to come to this result without making topological arguments.) However, we show here that such conclusions ultimately rely on the assumption $\xi = \theta$ for Equation (76). The answer to the question of whether $\xi = \bar{\alpha}$ or $\xi = \theta$ relies however on topology so that PCAC methods or related arguments cannot be used in order to bypass this step.

To review such points, we express the anomalous divergence of the axial U(1) current as

$$\partial_\mu \sum_{j=1}^{N_f} \langle \psi_j^\dagger \gamma^5 \gamma_\mu \psi_j \rangle = \frac{2N_f}{16\pi^2} \mathrm{tr} F\widetilde{F} + 2 \sum_{j=1}^{N_f} \langle \psi_j^\dagger \gamma^5 m_j \mathrm{e}^{\mathrm{i}\alpha_i \gamma^5} \psi_j \rangle \,. \tag{108}$$

Using this relation, we can therefore see that

$$\langle A | \frac{\theta}{16\pi^2} \mathrm{tr} F\widetilde{F} | B \rangle = -\frac{\theta}{N_f} \langle A | \sum_{j=1}^{N_f} \psi_j^\dagger \gamma^5 m_j \mathrm{e}^{\mathrm{i}\alpha_i \gamma^5} \psi_j | B \rangle \,, \tag{109}$$

where we have dropped the total divergence of the axial current. We have inserted a factor θ on each side in view of the subsequent discussion, which is however arbitrary at this point.

Now, we first follow Ref. [53] and carry out chiral field redefinitions so that in Equation (60) the mass matrices are purely real (and diagonal) and in particular $\bar{\alpha} = 0$ and $\theta = \bar{\theta}$. Moreover, one can take the freedom of non-anomalous $\mathrm{SU}(2)_A$ chiral field redefinitions to set $\alpha_j = 0$ individually. Assuming θ to be small, one may view the left-hand side of Equation (109) as a perturbative insertion of the CP-odd topological term into a hadronic transition matrix element. When $|A\rangle$ and $|B\rangle$ are eigenstates of CP with opposite eigenvalues, a nonzero result would then signal CP violation.

Before proceeding, we note that we could just as well have set $\theta = 0$, $\bar{\alpha} = +\bar{\theta}$ and treated the axial $\mathrm{U}(1)_A$ phase of the quark masses as the perturbation. This is an equivalent point of view that is taken, e.g., in Ref. [43] and is perhaps more true to the circumstance that the gluons have been integrated out in the chiral EFT so that the operator $\theta F\widetilde{F}$ has been removed as well. Either way, the right-hand side of Equation (109) turns out as

$$-\frac{\bar{\theta}}{N_f} \langle A | \sum_{j=1}^{N_f} \psi_j^\dagger \gamma^5 m_j \psi_j | B \rangle \,. \tag{110}$$

This way, all explicit phases have been pulled out in terms of a single factor in front of the matrix element (110). If, in addition, the chiral condensate has no phase, i.e., in Equation (68) $U_0 = \mathbb{1}_{N_f}$ or, in other words, the condensate points into the real direction, the matrix element (110) can be evaluated using standard reduction formulae as presented, e.g., in Ref. [54] leading to a nonzero result.

Therefore, taking $\theta = 0$ and $\bar{\alpha}$ as a perturbation and under the additional assumption that the chiral condensate points into the real direction, $U_0 = \mathbb{1}_{N_f}$, one sees here CP violation. This extra assumption, however, is exactly what characterizes the choice $\xi = \theta$ in Equation (76)—as opposed to the correct choice in Equation (67)—as the ground state implied by $\xi = \theta = 0$ can always be chosen as $U_0 = \mathbb{1}_{N_f}$. This can be seen to follow from Equations (68) and (78), provided that one uses the freedom to perform field redefinitions without changing $\theta = 0$ (or, equivalently, without changing $\bar{\alpha} = \alpha_u + \alpha_d$) to set α_u, α_d as follows,

$$\begin{aligned} \alpha_u &= \frac{m_d}{m_u + m_d}\,\bar{\alpha}, \\ \alpha_d &= \frac{m_u}{m_u + m_d}\,\bar{\alpha}. \end{aligned} \tag{111}$$

If instead $\xi = -\bar{\alpha}$, U_0 can be calculated according to Equations (68) and (78). Then, U_0 will be complex and aligned with $-\bar{\alpha}$, leading to additional perturbative corrections that will compensate for Equation (110). That this cancellation has to happen is just a consequence of the fact that we can dial $\bar{\alpha}$ to zero in the chiral effective Lagrangian (67) by a redefinition of the field U.

We can illustrate this point explicitly for the example of $\eta' \to 2\pi$ discussed in Refs. [43,53] as well as in Section 7 of the present paper. Suppose we start with the theory with $\theta = 0$ and $\bar{\alpha} = 0$. Then, $\xi = 0$ for either order of limits and there is no CP violation. Next, we reintroduce $\bar{\alpha} \neq 0$ with $|\bar{\alpha}| \ll 1$ through the Lagrangian term

$$\delta\mathcal{L}_{\bar{\theta}} = -\mathrm{i}\bar{\theta}\frac{m_u m_d}{m_u + m_d}\sum_{i=1}^{N_f}\bar{\psi}_i\gamma^5\psi_i\,, \tag{112}$$

where, here, $N_f = 2$ for simplicity and we have written $\bar{\alpha}$ as $\bar{\theta}$ since $\theta = 0$ per the present assumptions. Equation (112) follows from expanding the mass terms in the Lagrangian (60) to first order in α_i, restricting to two flavors and substituting Equation (111). We could have used the approximate $\mathrm{SU}(2)_A$ symmetry to attribute the phases differently among the flavors u and d, but the present form allows most straightforwardly to connect with the relations from Refs. [43,53] and Section 7. In particular, inserting the Lagrangian of Equation (112) between the vacuum and $|\eta'\pi^0\pi^0\rangle$, using PCAC relations, leads to

$$\langle 0|\delta\mathcal{L}|\eta'\pi^0\pi^0\rangle = \frac{m_u m_d}{(m_u + m_d)^2}\frac{m_\pi^2}{f_\pi}\bar{\theta}\,. \tag{113}$$

The previous result would give the total contribution to the CP-violating matrix element under the crucial assumption (which is not made explicit in references such as [43,53], neither do these articles contain a derivation of the expectation value $\langle U\rangle$) that there are no additional CP-odd phases from the quark condensates when taking matrix elements on the physical states. Equation (113) matches Equation (5) of Ref. [53] up to a factor of $\sqrt{2/3}$, which can be understood from the fact that the former reference included a third light quark which was not accounted for here. Of course, this result from PCAC relations also follows directly from the EFT Lagrangian term (81) when assuming $\xi = \theta = 0$, so that $\xi + \alpha_u + \alpha_d = \alpha_u + \alpha_d = \bar{\theta}$.

Rather than relying on assumptions, the values of the phases in the quark condensate are calculable, as was shown in our effective theory analysis in Section 7, in which the condensate phases were included in Equation (68) from the beginning and determined by solving the tadpole conditions, cf. Equation (78). Now, when choosing the evaluation of the partition function according to Equations (12) and (13), the chiral condensate U_0 aligns with θ. This follows from the fact that in this order of limits, one should identify $\xi = \theta$ in Equation (67). Then, Equations (78) and (111) relate φ_u, φ_d to θ. As $\theta = 0$ in the present setup, one obtains $\varphi_u = \varphi_d = 0$, so that the chiral condensate induces no additional phases and Equation (113) then is the only contribution to the total matrix element, as mentioned before. If instead the evaluation is performed as in Equation (11), which corresponds to an evaluation of the partition function (9) on a connected integration contour, $\xi = -\bar{\alpha}$ in Equation (67), so that Equations (68) and (78) give $\det U_0 = \mathrm{e}^{-\mathrm{i}\bar{\alpha}}$. Therefore, this deviation from an orientation of the chiral condensate in the real direction must be accounted for, which leads to additional complex phases in matrix elements, and thus extra contributions to Equation (113). The existence of additional phases to those leading to the PCAC result of Equation (113) is captured by our EFT result (81) in the following manner: The angle $\bar{\alpha} = \alpha_u + \alpha$ (which gives the part that can alternatively be evaluated through the PCAC relation (113)) is compensated by $\xi = -\bar{\alpha}$, leading to a vanishing CP-violating amplitude.

In summary, the arguments from Refs. [43,53], while technically correct if it were $U_0 = \mathbb{1}_{N_f}$ for $\theta = 0$, do not demonstrate that this assumption actually holds. To assess this, one, after all, has to consider topology in order to derive the effective 't Hooft vertex and to sort out the correct limiting procedure, i.e., either Equation (11) or Equations (12) and (13). As Equation (11) turns out to be correct, U_0 aligns with $-\bar{\alpha}$ instead of θ so that there is no CP violation.

8.5. Relation of the Present Arguments to Some Recent Literature

There are other recent papers that argue for CP conservation in strong interactions. We do not discuss these here comprehensively but state why they substantially differ from the line of reasoning in the present work, or, respectively, if there is some prospect to establish connections.

In Ref. [55], the starting assumption is a finite toroidal four-dimensional geometry, where the different topological sectors are weighted as $\exp(\mathrm{i}\Delta n\theta)$. Because of the finite volume, this decisively differs from Equation (11). It is further reported in that work that for $\theta \neq 0$, no confinement occurs so that the experimentally observed confinement must result from $\theta = 0$. We can neither confirm this latter statement nor is it central to the discussion in the present work.

In Ref. [56], it is stated that topological charge, i.e., winding number, is not observable and that no 't Hooft operator can be derived. We do not reproduce this, and the 't Hooft operator (62) here is of physical consequence and is central to the present discussion.

The paper [57] makes the point that coherence between topological sectors must not lead to observable consequences in the presence of causal horizons, which arguably only happens when $\theta = 0$. In the present work, it turns out that θ has no material consequence in infinite Euclidean spacetime and hence neither in Minkowski spacetime. However, the present result (50) is suggestive in that the partition function in a finite subvolume with free boundary conditions of Euclidean space, as derived from Equation (9), does not exhibit the parameter θ any more. It would therefore be interesting to understand the possible relation between the Euclidean subvolumes and the domains within a causal horizon of Ref. [57].

9. Conclusions

Topological quantization appears central to the correct assessment of the CP-odd θ parameter in QCD. It is therefore important to deduce its origin and to account for the implications of the setup of the problem. When we take Euclidean spacetime as the analytic continuation of Minkowski spacetime and do not impose ad hoc boundary conditions, time must be taken to infinity. Then, integration contours can be deduced that imply that we are first to evaluate the path integral in the individual topological sectors in infinite volume and then to sum over these sectors. As a consequence, there is no CP violation present in QCD, particularly not in the effective 't Hooft vertex. While this corresponds to a stringent reasoning within zero-temperature QCD, beyond the present work, we shall next address finite-temperature QCD as a setup with a clear physical meaning in a finite spacetime volume.

Author Contributions: All authors have substantially contributed to the conceptualization and investigations leading to the present results as well as to writing and review of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: W.Y.A. is supported by the UK Engineering and Physical Sciences Research Council (EPSRC), under Research Grant No. EP/V002821/1. The research of C.T. was supported by the Cluster of Excellence Precision Physics, Fundamental Interactions, and Structure of Matter (PRISMA+, EXC 2118/1) within the German Excellence Strategy (Project-ID 390831469).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Baker, C.A.; Doyle, D.D.; Geltenbort, P.; Green, K.; Van der Grinten, M.G.D.; Harris, P.G.; Iaydjiev, P.; Ivanov, S.N.; May, D.J.R.; Pendlebury, J.M.; et al. An Improved experimental limit on the electric dipole moment of the neutron. *Phys. Rev. Lett.* **2006**, *97*, 131801. [CrossRef] [PubMed]
2. Abel, C.; Afach, S.; Ayres, N.J.; Baker, C.A.; Ban, G.; Bison, G.; Bodek, K.; Bondar, V.; Burghoff, M.; Chanel, E.; et al. Measurement of the Permanent Electric Dipole Moment of the Neutron. *Phys. Rev. Lett.* **2020**, *124*, 081803. [CrossRef] [PubMed]
3. Fritzsch, H.; Gell-Mann, M.; Leutwyler, H. Advantages of the Color Octet Gluon Picture. *Phys. Lett. B* **1973**, *47*, 365–368. [CrossRef]
4. Jackiw, R.; Rebbi, C. Vacuum Periodicity in a Yang-Mills Quantum Theory. *Phys. Rev. Lett.* **1976**, *37*, 172–175. [CrossRef]
5. Callan, C.G., Jr.; Dashen, R.F.; Gross, D.J. The Structure of the Gauge Theory Vacuum. *Phys. Lett. B* **1976**, *63*, 334–340. [CrossRef]
6. Callan, C.G., Jr.; Dashen, R.F.; Gross, D.J. Toward a Theory of the Strong Interactions. *Phys. Rev. D* **1978**, *17*, 2717. [CrossRef]
7. Ai, W.Y.; Cruz, J.S.; Garbrecht, B.; Tamarit, C. Consequences of the order of the limit of infinite spacetime volume and the sum over topological sectors for CP violation in the strong interactions. *Phys. Lett. B* **2021**, *822*, 136616. [CrossRef]
8. Diakonov, D.; Petrov, V.Y.; Pobylitsa, P.V. A Chiral Theory of Nucleons. *Nucl. Phys. B* **1988**, *306*, 809. [CrossRef]
9. Schäfer, T.; Shuryak, E.V. Instantons in QCD. *Rev. Mod. Phys.* **1998**, *70*, 323–426. [CrossRef]
10. Diakonov, D. Instantons at work. *Prog. Part. Nucl. Phys.* **2003**, *51*, 173–222. [CrossRef]
11. Coleman, S. *Aspects of Symmetry: Selected Erice Lectures*; Cambridge University Press: Cambridge, UK, 1985. [CrossRef]
12. Atiyah, M.F.; Singer, I.M. The index of elliptic operators on compact manifolds. *Bull. Am. Math. Soc.* **1969**, *69*, 422–433. [CrossRef]
13. Branco, G.; Lavoura, L.; Silva, J. *CP Violation*; International Series of Monographs on Physics; Oxford University Press: Oxford, England, 1999.
14. Vainshtein, A.I.; Zakharov, V.I.; Novikov, V.A.; Shifman, M.A. ABC of Instantons. *Sov. Phys. Usp.* **1982**, *25*, 195. [CrossRef]
15. Bell, J.S.; Jackiw, R. A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ model. *Nuovo Cim. A* **1969**, *60*, 47–61. [CrossRef]
16. Adler, S.L. Axial vector vertex in spinor electrodynamics. *Phys. Rev.* **1969**, *177*, 2426–2438. [CrossRef]
17. Ai, W.Y.; Garbrecht, B.; Tamarit, C. The QCD θ-parameter in canonical quantization. *arXiv* **2024**, arXiv:2403.00747. Available online: https://arxiv.org/abs/2403.00747 (accessed on 17 April 2024).
18. 't Hooft, G. How Instantons Solve the U(1) Problem. *Phys. Rept.* **1986**, *142*, 357–387. [CrossRef]
19. Ai, W.Y.; Cruz, J.S.; Garbrecht, B.; Tamarit, C. The limits of the strong CP problem. *arXiv* **2022**, arXiv:2205.15093.
20. Dorey, N.; Hollowood, T.J.; Khoze, V.V.; Mattis, M.P. The Calculus of many instantons. *Phys. Rept.* **2002**, *371*, 231–459. [CrossRef]
21. Pokorski, S. *Gauge Field Theories*, 2nd ed.; Cambridge Monographs on Mathematical Physics, Cambridge University Press: Cambridge, UK, 2000. [CrossRef]
22. Witten, E. Analytic Continuation Of Chern-Simons Theory. *AMS/IP Stud. Adv. Math.* **2011**, *50*, 347–446.
23. Witten, E. A New Look at the Path Integral Of Quantum Mechanics. In *Surveys in Differential Geometry*; Mrowka, T., Yau, S.T., Eds.; International Press of Boston: Boston, MA, USA, 2010; Volume 15, pp. 345–420. Available online: https://www.intlpress.com/site/pub/files/_fulltext/journals/sdg/2010/0015/0001/SDG-2010-0015-0001-a011.pdf (accessed on 17 April 2024).
24. Tanizaki, Y.; Koike, T. Real-time Feynman path integral with Picard–Lefschetz theory and its applications to quantum tunneling. *Annals Phys.* **2014**, *351*, 250–274. [CrossRef]
25. Ai, W.Y.; Garbrecht, B.; Tamarit, C. Functional methods for false vacuum decay in real time. *J. High Energy Phys.* **2019**, *12*, 095. [CrossRef]
26. Jackiw, R. Introduction to the Yang-Mills Quantum Theory. *Rev. Mod. Phys.* **1980**, *52*, 661–673. [CrossRef]
27. Okubo, S.; Marshak, R.E. Argument for the nonexistence of the 'strong CP problem' in QCD. *Prog. Theor. Phys.* **1992**, *87*, 1059–1062. [CrossRef]
28. Gracia-Bondía, J.M.; Várilly, J.C. Ensuring locality in QFT via string-local fields. *arXiv* **2023**, arXiv:2207.06522. Available online: https://arxiv.org/abs/2207.06522 (accessed on 17 April 2024).
29. 't Hooft, G. Symmetry Breaking Through Bell-Jackiw Anomalies. *Phys. Rev. Lett.* **1976**, *37*, 8–11. [CrossRef]
30. 't Hooft, G. Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle. *Phys. Rev. D* **1976**, *14*, 3432–3450; Erratum in *Phys. Rev. D* **1978**, *18*, 2199. [CrossRef]
31. Belavin, A.A.; Polyakov, A.M.; Schwartz, A.S.; Tyupkin, Y.S. Pseudoparticle Solutions of the Yang-Mills Equations. *Phys. Lett. B* **1975**, *59*, 85–87. [CrossRef]
32. Jentschura, U.D.; Surzhykov, A.; Zinn-Justin, J. Generalized Nonanalytic Expansions, PT-Symmetry and Large-Order Formulas for Odd Anharmonic Oscillators. *SIGMA* **2009**, *5*, 005. [CrossRef]
33. Atiyah, M.F.; Hitchin, N.J.; Drinfeld, V.G.; Manin, Y.I. Construction of Instantons. *Phys. Lett. A* **1978**, *65*, 185–187. [CrossRef]
34. Shifman, M.A.; Vainshtein, A.I.; Zakharov, V.I. Instanton Density in a Theory with Massless Quarks. *Nucl. Phys. B* **1980**, *163*, 46–56. [CrossRef]
35. Diakonov, D.; Petrov, V.Y. A Theory of Light Quarks in the Instanton Vacuum. *Nucl. Phys. B* **1986**, *272*, 457–489. [CrossRef]

36. Leutwyler, H.; Smilga, A.V. Spectrum of Dirac operator and role of winding number in QCD. *Phys. Rev. D* **1992**, *46*, 5607–5632. [CrossRef] [PubMed]
37. Diakonov, D.; Oswald, M. Covariant derivative expansion of fermionic effective action at high temperatures. *Phys. Rev. D* **2004**, *70*, 016006. [CrossRef]
38. Weinberg, S. *The Quantum Theory of Fields. Vol. 1: Foundations*; Cambridge University Press: Cambridge, UK, 2005. [CrossRef]
39. Pich, A. Chiral perturbation theory. *Rept. Prog. Phys.* **1995**, *58*, 563–610. [CrossRef]
40. Scherer, S. Introduction to chiral perturbation theory. *Adv. Nucl. Phys.* **2003**, *27*, 277.
41. Srednicki, M. *Quantum Field Theory*; Cambridge University Press: Cambridge, UK, 2007.
42. Cheng, H.Y. The Strong CP Problem Revisited. *Phys. Rept.* **1988**, *158*, 1. [CrossRef]
43. Crewther, R.J.; Di Vecchia, P.; Veneziano, G.; Witten, E. Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics. *Phys. Lett. B* **1979**, *88*, 123; Erratum in *Phys. Lett. B* **1980**, *91*, 487. [CrossRef]
44. Ellis, J.R.; Gaillard, M.K.; Nanopoulos, D.V. Lefthanded Currents and CP Violation. *Nucl. Phys. B* **1976**, *109*, 213–243. [CrossRef]
45. Ellis, J.R.; Gaillard, M.K. Strong and Weak CP Violation. *Nucl. Phys. B* **1979**, *150*, 141–162. [CrossRef]
46. Choi, K. Discussion on hep-th/2001.07152 "Consequences of the order of the limit of infinite spacetime volume and the sum over topological sectors". Presented at the Workshop Axions across Boundaries between Particle Physics, Astrophysics, Cosmology and forefront Detection Technologies, The Galileo Galilei Institute, Florence, Italy, 26 April 2023–9 June 2023. Available online: https://www.youtube.com/watch?v=pJ7zwaFoDPE (accessed on 17 April 2024).
47. Dvali, G. Strong-CP with and without Gravity. *arXiv* **2022**, arXiv:2209.14219. Available online: https://arxiv.org/abs/2209.14219 (accessed on 17 April 2024).
48. Dvali, G. One more discussion on hep-th/2001.07152 "Consequences of the order of the limit of infinite spacetime volume and the sum over topological sectors". Presented at the Workshop Axions across Boundaries between Particle Physics, Astrophysics, Cosmology and forefront Detection Technologies, The Galileo Galilei Institute, Florence, Italy, 26 April 2023–9 June 2023. Available online: https://www.youtube.com/watch?v=7TW6bD4Nig4 (accessed on 17 April 2024).
49. Dvali, G. Three-Form Gauging of Axion Symmetries and Gravity. *arXiv* **2005**, arXiv:hep-th/0507215. Available online: https://arxiv.org/abs/hep-th/0507215 (accessed on 17 April 2024).
50. Dvali, G. Topological Origin of Chiral Symmetry Breaking in QCD and in Gravity. *arXiv* **2017**, arXiv:1705.06317. Available online: https://arxiv.org/abs/1705.06317 (accessed on 17 April 2024).
51. Luscher, M. The Secret Long Range Force in Quantum Field Theories With Instantons. *Phys. Lett. B* **1978**, *78*, 465–467. [CrossRef]
52. Aoki, S.; Fukaya, H.; Hashimoto, S.; Onogi, T. Finite volume QCD at fixed topological charge. *Phys. Rev. D* **2007**, *76*, 054508. [CrossRef]
53. Shifman, M.A.; Vainshtein, A.I.; Zakharov, V.I. Can Confinement Ensure Natural CP Invariance of Strong Interactions? *Nucl. Phys. B* **1980**, *166*, 493–506. [CrossRef]
54. Donoghue, J.F.; Golowich, E.; Holstein, B.R. *Dynamics of the Standard Model*, 2nd ed.; Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press: Cambridge, UK, 2014. [CrossRef]
55. Nakamura, Y.; Schierholz, G. The strong CP problem solved by itself due to long-distance vacuum effects. *Nucl. Phys. B* **2023**, *986*, 116063. [CrossRef]
56. Yamanaka, N. Unobservability of the topological charge in nonabelian gauge theory: Ward-Takahashi identity and phenomenological aspects. *arXiv* **2022**, arXiv:2212.11820.
57. Torrieri, G.; Truran, H.D. The strong CP problem, general covariance, and horizons. *Class. Quant. Grav.* **2021**, *38*, 215002. [CrossRef]