Enhanced EDMs from small instantons

Ravneet S. Bedi[*] and Tony Gherghetta[†]
School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA

Maxim Pospelov[‡]
William I. Fine Theoretical Physics Institute, School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA

(Received 21 May 2022; accepted 12 July 2022; published 29 July 2022)

We show that models in which the strong CP problem is solved by introducing an axion field with a mass enhanced by non-QCD UV dynamics at a scale $\Lambda_{\rm SI}$ exhibit enhanced sensitivity to external sources of CP violation. In the presence of higher-dimensional CP-odd sources at a scale $\Lambda_{\rm CP}$, the same mechanisms that enhance the axion mass also modify the axion potential, shifting the potential minimum by a factor $\propto \Lambda_{\rm SI}^2/\Lambda_{\rm CP}^2$. This phenomenon of CP-violation enhancement, which puts stringent constraints on the scale of new physics, is explicitly demonstrated within a broad class of "small instanton" models with CP-odd sources arising from the dimension-six Weinberg gluonic and four-fermion operators. We find that for heavy axion masses $\gtrsim$100 MeV, arising from new dynamics at $\Lambda_{\rm SI} \lesssim 10^{10}$ GeV, CP violation generated up to the Planck scale can be probed by future electric dipole moment experiments.

DOI: 10.1103/PhysRevD.106.015030

I. INTRODUCTION

The Standard Model (SM) of particle physics has two sources of CP violation. The well-established and measured source of CP violation in the quark mixing sector, the Kobayashi-Maskawa phase [1], is responsible for a multitude of CP-violating phenomena observed in the quark flavor-changing transitions. At the same time, this phase induces electric dipole moments (EDMs) of neutrons and heavy atoms well below current experimental limits. The other source of CP violation, the nonperturbative parameter θ of quantum chromodynamics (QCD), is largely irrelevant for flavor physics, but tends to induce large EDMs. The nonobservation of EDMs that imply the smallness of theta, $|\theta| \lesssim 10^{-10}$ [2,3], contrasted with the naive expectation of $\theta \sim O(1)$, poses a naturalness problem for the Standard Model, the strong CP problem.

There are two generic approaches to resolve the strong CP problem. The first approach involves promoting the θ parameter to a new dynamical field, the QCD axion [4–10], which symbolically can be represented as

$$\frac{\theta}{32\pi^2} G^c_{\mu\nu}\tilde{G}^{c\mu\nu} \to \frac{1}{2}(\partial_\mu a)^2 + \frac{a}{32\pi^2 f_a} G^c_{\mu\nu}\tilde{G}^{c\mu\nu}, \qquad (1)$$

where $G^c_{\mu\nu}$ is the gluon field strength, $\tilde{G}^{c\mu\nu} \equiv \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}G^c_{\rho\sigma}$ with c the adjoint index and f_a is the decay constant of the axion field a. The QCD vacuum energy, which for small θ can be parametrically expressed as

$$E(\theta) \propto \theta^2 m_q \Lambda_{\rm QCD}^3 \to V(a) = \frac{1}{2} m_a^2 a^2, \qquad (2)$$

can be made to dynamically relax to the minimum of the potential $V(a)$. In this expression, $\Lambda_{\rm QCD}$ is the nonperturbative scale of the strong interactions, and m_q is the light quark mass. As a result, any initial value of $\theta = a/f_a$ will relax to the minimum of the axion potential. In the absence of additional sources of CP violation, this minimum is exactly at $\theta = 0$, as in Eq. (2). Therefore, the neutron EDM that scales as

$$d_n \propto \frac{m_q \theta}{\Lambda_{\rm QCD}^2} \qquad (3)$$

is also relaxed to zero.

Consider now additional sources of CP violation placed at some new physics scale $\Lambda_{\rm CP}$ that we will assume to be larger than the electroweak scale (for example, this could be due to supersymmetric theories with large CP-violating phases). Integrating out the new physics at this scale will, in general, result in a number of generic consequences:

[*]bedi0019@umn.edu
[†]tgher@umn.edu
[‡]pospelov@umn.edu

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

(1) The theta parameter may receive additive corrections to its value, $\theta \to \theta + \theta_{\rm rad}$. Since $G\tilde{G}$ is a dimension four operator, $\theta_{\rm rad}$ can depend only on the ratio of scales, and therefore has $\Lambda_{\rm CP}^0$ scaling. Potentially, this can be a large correction, but the axion mechanism will remove the theta term together with $\theta_{\rm rad}$.

(2) CP-violating new physics will generically induce higher-dimensional CP-odd operators, of which the most relevant are dimension six operators, $\mathcal{O}_6$ that are suppressed by the square of the new physics scale, and the resulting EDMs will have scaling $d_n(\mathcal{O}_6) \propto \Lambda_{\rm QCD}/\Lambda_{\rm CP}^2$ (or $m_q/\Lambda_{\rm CP}^2$, depending on the chiral properties of $\mathcal{O}_6$).

(3) In the presence of higher-dimensional CP-odd new physics operators, the axion potential minimum shifts away from zero inducing a low-energy value of theta, $\theta_{\rm ind} \propto \Lambda_{\rm QCD}^2/\Lambda_{\rm CP}^2$. This leads to an additional θ-induced contribution to d_n that has, for example, a comparable $m_q/\Lambda_{\rm CP}^2$ scaling [11–13].

An important conclusion can be drawn from these observations: the QCD axion mechanism ensures that for sufficiently large $\Lambda_{\rm CP}$, the observable EDMs can be made small and indeed within current bounds for $\Lambda_{\rm CP} \gtrsim 100$ TeV, one can allow for an arbitrarily large amount of (strong) CP violation above these scales. In this sense, the axion mechanism allows for a proper decoupling of new physics contributions to EDMs.

The second class of models does not introduce an axion, and instead appeals to symmetry arguments that help to argue why θ is zero or small. Historically, models with an exact CP symmetry or exact parity, that is spontaneously broken at some UV scale, have been argued to give a viable solution to the strong CP problem (see Refs. [14–21] for a representative set of ideas). Models based on mirror symmetries have also been used to implement this approach [22,23]. The most important feature of these models is the absence of a dynamical axion and the sensitivity of EDM observables to the value of θ generated at a UV scale. For example, the spontaneous breaking of CP symmetry may also result in complex quark Yukawa couplings that feed into $\theta_{\rm rad}$ (a representative set of calculations can be found in Refs. [24–28]). Since the θ term has $\Lambda_{\rm CP}^0$ scaling, this nondecoupling means that all possible sources of CP breaking have to be "controlled" to very high scales.

Recently, there has been renewed interest in models that solve the strong CP problem, which occupy an intermediate niche between the QCD axion solution and solutions based on discrete symmetries. In this class of models there is still a dynamical axion field and Peccei-Quinn symmetry at a high scale, but the axion mass is now enhanced compared to (2) by additional dynamical mechanisms at the small-instanton scale $\Lambda_{\rm SI}$. By small instantons we refer to instantons whose size $1/\Lambda_{\rm SI}$ is smaller than the inverse electroweak scale (see Fig. 1). For example, extending the strong gauge interactions and the corresponding axion to a larger group where the non-QCD partners confine at a much larger scale $\Lambda'_{\rm QCD}$ (identified with $\Lambda_{\rm SI}$) can lead to a significant parametric increase in the axion mass provided $\Lambda'_{\rm QCD} \gg \Lambda_{\rm QCD}$ [29–33]. Similarly, an axion "portal" between QCD and a mirror QCD with the alignment of θ and θ' can also result in a heavier axion for $\Lambda'_{\rm QCD} \gg \Lambda_{\rm QCD}$ [23,34]. Alternatively, if the QCD coupling running is modified to become strong above the TeV scale, the QCD axion mass would receive new contributions from "small"-size instantons [35–40]. This naturally occurs in models where at some UV scale, QCD propagates in five dimensions [41,42]. These models which significantly enhance the axion mass compared to the minimal QCD axion models have a distinctively different phenomenology. Indeed, given the conventional axion mass range 10^{-6}–10^{-3} eV, the enhancement mechanisms imply heavy axions could be in the 100 MeV range or above. These heavier axions avoid most of the astrophysical bounds, and make the axion amenable to searches at beam dump and collider experiments [34,39]. Moreover, such heavy axions will be less susceptible to possible distortions of the axion potential by the imperfections of the Peccei-Quinn global symmetry.

FIG. 1. Schematic diagram of the different scales referred to in the text. The scale of CP violation, $\Lambda_{\rm CP}$ due to dimension six operators, is a UV scale near the Planck scale, M_P, and $\Lambda_{\rm SI}$ is the small-instanton scale (assumed to be above the electroweak scale, v and QCD strong coupling scale, $\Lambda_{\rm QCD}$) where new dynamics enhances the axion mass. The PQ symmetry breaking scale f_a is assumed to be an independent parameter that can either be above (as shown in the figure) or below the scale $\Lambda_{\rm SI}$.

Besides the enhanced axion mass it is therefore also interesting to consider whether EDM observables could be enhanced in these models. In this paper we investigate heavy axion models in the presence of additional sources of CP violation, which are parametrized as higher-dimensional operators that arise from SM fields and are not related to Planck scale gravitational corrections associated with the axion quality problem. The central question we would like to address is whether there is a similar decoupling as for the standard QCD axion, where all observables from, for example, dimension six operators, scale as $\Lambda_{\rm QCD}^2/\Lambda_{\rm CP}^2$, or if there is an enhancement of CP violation mediated by the induced θ which is similar to models attempting to solve the strong CP problem using exact parity or CP symmetries.

To answer this question we compute the topological susceptibility and mixed correlators in heavy axion models that arise from two sources of CP violation: the dimension six Weinberg gluonic operator and a CP-odd four-fermion operator. Such CP-odd operators induce a linear term in θ (or equivalently a) in the axion potential leading to a shift $\theta_{\rm ind}$ in the potential minimum. Similar contributions were proposed in [36], and were estimated on dimensional grounds for fermionic and scalar operators in [37,38,43].

Instead of relying on dimensional analysis our computation employs a simple, noninteracting instanton (or anti-instanton) background that ignores strong coupling effects, where we are able to extract qualitative results which show that the induced theta, $\theta_{\rm ind} \propto \Lambda_{\rm SI}^2/\Lambda_{\rm CP}^2$.

This induced shift is qualitatively different from the usual QCD axion scenario and solutions based on exact discrete symmetries due to the presence of the new scale $\Lambda_{\rm SI}$. While there is still decoupling in the $\Lambda_{\rm CP} \to \infty$ limit, our results show that the induced θ can enhance the magnitude of observable EDMs, even to the point that if $\Lambda_{\rm SI}^2/\Lambda_{\rm CP}^2$ is too large, the strong CP problem will reappear. Thus, models with a dynamically enhanced axion mass are subject to bounds depending on the amount of CP violation that is present at energy scales that may significantly exceed 100 TeV. Interestingly, the enhanced EDMs are potentially observable in future EDM experiments.

This paper is organized as follows: in Sec. II we investigate vacuum correlators in an instanton (or anti-instanton) background with different sources of CP violation that shift the axion potential minimum. In Sec. III, we consider different heavy QCD axion models with small instantons, deriving the resulting size of the induced θ and subsequent constraints on the CP-violating scale, $\Lambda_{\rm CP}$. We reach our conclusions in Sec. IV.

II. INSTANTON CORRELATION FUNCTIONS

We begin with briefly reviewing QCD dynamics and the instanton solution that will be used to compute various instanton correlation functions. The pure Yang-Mills part of the QCD Lagrangian is given by

$$\mathcal{L}_{\rm QCD} = -\frac{1}{4g^2} G^a_{\mu\nu} G^{a\mu\nu} + \frac{\theta}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{4}$$

where g is the QCD gauge coupling, θ is the QCD vacuum angle, and $a = 1, \ldots, 8$ labels the gauge adjoint representation. The BPST instanton solution [44] is given by

$$A^a_\mu(x) = \frac{2\eta^a_{\mu\nu}(x - x_0)_\nu}{(x - x_0)^2 + \rho^2}, \tag{5}$$

where the instanton is located at x_0 and has a size ρ. The $\eta^a_{\mu\nu}$ denote the group-theoretic 't Hooft η symbols [45]. The topological charge is defined to be

$$Q = \frac{1}{32\pi^2} \int d^4x G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{6}$$

where $Q = 1$ for the one instanton solution (5). We will next compute correlation functions in the instanton (or anti-instanton) background (5) that will be useful in obtaining contributions to EDM observables such as the neutron EDM.

A. Topological susceptibility

The vacuum-to-vacuum amplitude in QCD can be written as

$$\langle 0|0\rangle = \sum_Q \int \mathcal{D}A^{(Q)}_\mu e^{-S_E}, \tag{7}$$

where the Euclidean action for (4) in an instanton background of charge Q [46] is given by

$$S_E = \frac{8\pi^2}{g^2}|Q| + iQ\theta. \tag{8}$$

The topological susceptibility is then introduced as [8,11,47]

$$\chi(0) = -i \lim_{k\to 0} \int d^4x e^{ikx} \times \langle 0|T\left\{\frac{1}{32\pi^2} G\tilde{G}(x), \frac{1}{32\pi^2} G\tilde{G}(0)\right\}|0\rangle, \tag{9}$$

where $G\tilde{G}$ is shorthand notation for $G^a_{\mu\nu}\tilde{G}^{a\mu\nu}$.

Since the amplitude in the $|Q| > 1$ instanton background becomes more exponentially suppressed, only the $Q = \pm 1$ configurations dominate the path integral. Henceforth, we refer to S_E in (8) only for $|Q| = 1$. In the instanton background (5) we then obtain the two-point correlator

$$\langle 0|T\{G\tilde{G}(x), G\tilde{G}(0)\}|0\rangle_{Q=+1} = \int \mathcal{D}A_\mu G\tilde{G}(x) G\tilde{G}(0) e^{-\frac{8\pi^2}{g_0^2}}, \tag{10}$$

$$= \int d^4x_0 \frac{d\rho}{\rho^5} C[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-\frac{8\pi^2}{g^2(1/\rho)}} \times \frac{192\rho^4}{((x-x_0)^2+\rho^2)^4} \frac{192\rho^4}{(x_0^2+\rho^2)^4}, \tag{11}$$

where the running coupling $g(1/\rho)$ encodes corrections from the quantum fluctuations. In (10) we have replaced the path integral over the fluctuation A_μ with an integration in (11) over the collective coordinates (see Ref. [45]) where, assuming an $SU(N)$ gauge group,[1] the coefficient

$$C[N] = \frac{C_1 e^{-C_2 N}}{(N-1)!(N-2)!}, \tag{12}$$

and C_1, C_2 are order one constants ($C_1 = 0.466$, $C_2 = 1.679$ using Pauli-Villars regularization [49]). The gauge coupling running is given by

$$\frac{8\pi^2}{g^2(1/\rho)} = \frac{8\pi^2}{g_0^2} - b_0 \log(M_{\rm UV}\rho), \tag{13}$$

where $b_0 = 4N - N/3 = 11N/3$ is the pure $SU(N)$ Yang-Mills β-function coefficient and $g_0 = g(M_{\rm UV})$ with UV cutoff $M_{\rm UV}$.

In principle, we could consider an ensemble of instantons and anti-instantons [50–52] to compute correlation functions. However, the qualitative aspects of such an ensemble can be simply captured by one instanton and one anti-instanton [53,54], where the (anti-)instantons are assumed to be noninteracting with each other and can be justified in the weak coupling regime. Thus, we will compute correlation functions by adding the contribution from an instanton background to that in an anti-instanton background. The total contribution to the topological susceptibility, obtained by performing the x integration first that arises from (9), followed by the x_0 integration in (11), is then given by

$$\chi(0) = -2i \int \frac{d\rho}{\rho^5} C[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-\frac{8\pi^2}{g^2(1/\rho)}}. \tag{14}$$

Assuming an asymptotically free theory, the integral in (14) is divergent for large instantons but can be evaluated with a IR cutoff $\rho_{\rm IR}$ on the instanton size. Assuming $N = 3$ with $\rho_{\rm IR} = 1/\Lambda_{\rm QCD}$ we obtain $\chi(0) \propto \Lambda_{\rm QCD}^4$.

[1]In principle, we should also include the normalized Haar measure of the group, as computed in [40,48]. We will omit this measure since its value is simply one for (11) [or an $O(1)$ number in more generic cases], and therefore our qualitative results remain unchanged.

1. Fermion contributions

The introduction of fermions modifies the path integral and the collective coordinate integration. In the massless fermion limit, the pure vacuum-to-vacuum transition amplitude is zero. Instead, the instanton now causes transitions from left-handed to right-handed fermions violating the $U(1)$ chiral symmetry so that, for example, $\langle 0|\bar\psi_{Ri}\psi_{Li}|0\rangle \neq 0$. Thus, instantons only contribute to correlation functions in which each fermion flavor and chirality appears at least once.

The effect of massless fermions is usually formulated as an "effective" Lagrangian [45,49]

$$\mathcal{L}_f = \int d^4x_0 \frac{d\rho}{\rho^5} C[N] e^{0.292 N_f} \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-S_E} \times \rho^{3N_f} \det[\bar\psi_R(x_0)\psi_L(x_0)] + \text{H.c.}, \tag{15}$$

where the determinant is taken over the N_f fermion flavors, and $\psi^\alpha_{L,R}(x_0)$ are the fermion zero modes. The constant $e^{0.292N_f}$ assumes Pauli-Villars regularization and the gauge coupling running (13) now includes the fermion contributions $b_0 \to b_0 - 2/3 N_f$.

Note that because of the explicit appearance of the fermion zero modes $\psi_{L,R}(x_0)$ in (15), there is only a contribution to the axion potential if the external fermion zero mode legs are closed. There are two ways this can occur. The first way is to assume that the fermions have an explicit mass m_f [corresponding to a nonzero Higgs vacuum expectation value (VEV), $v \approx 246$ GeV] that connects left- and right-handed fermion fields. The determinant in the effective action then gives a contribution $\propto (\rho m_f)^{N_f}$ for N_f fermion flavors. This is the case for the usual contributions from "large" instantons with $\rho \sim \rho_{\rm IR} = 1/\Lambda_{\rm QCD}$ and $m_f \lesssim \Lambda_{\rm QCD}$. However, since we are interested in "small" instantons corresponding to instanton sizes ($\sim 1/\Lambda_{\rm SI}$) much smaller than the inverse of the electroweak scale, a second possibility is to close the external fermion zero-mode legs in (15) with $N_f/2$ Higgs bosons. This contribution will be proportional to the product of Yukawa couplings (times a loop factor) and is larger than the Higgs VEV contribution that now scales as $\sim (m_f/\Lambda_{\rm SI})^{N_f}$ (assuming $\Lambda_{\rm SI} \gg v$). Instead of proceeding with the 't Hooft determinant operator in the effective Lagrangian (15) we will follow the approach taken in Refs. [38,40] and directly compute the vacuum-to-vacuum amplitude by including the Higgs-fermion Yukawa interaction in the path integral.

Consider a Higgs field H which couples to N_f flavors of massless fermions with the following Euclidean action:

$$S_H = S_H^{(0)} - i \int d^4x \sum_{i=1}^{N_f} \frac{y_i}{\sqrt{2}} H(x) \bar\psi_i(x) \psi_i(x), \tag{16}$$

where $S_H^{(0)}$ is the quadratic (free) part of the Higgs action and y_i are the Yukawa couplings. The Yukawa couplings, or equivalently the fermion masses, have been redefined to be real with their phase included in $\bar{\theta} = \theta + \mathrm{Arg\,Det}M_q$, where M_q is the quark mass matrix. The vacuum-to-vacuum amplitude now takes the form

$$\langle 0|0\rangle_{\Delta Q=1} = \int d^4x_0 \frac{d\rho}{\rho^5} C[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-S_E} \int \mathcal{D}H e^{-S_H^{(0)}} \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-S_\psi^{(0)} + i\int d^4x \sum_{i=1}^{N_f} \frac{y_i}{\sqrt{2}} H(x)\bar{\psi}_i(x)\psi_i(x)},$$
$$= \int d^4x_0 \frac{d\rho}{\rho^5} C[N] e^{0.292 N_f} \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-S_E} (N_f - 1)!! \left(\prod_{i=1}^{N_f} \frac{y_i \rho}{\sqrt{2}}\right) \mathcal{I}^{N_f/2} \tag{17}$$

where the action S_E is defined in (8) with $\theta \to \bar{\theta}$. The first line in (17) shows the collective coordinate integration arising from the gauge field part of the path integral and the second line contains the Higgs and massless fermion contributions to the path integral with $S_\psi^{(0)}$ the quadratic (free) part of the fermion action. Integrating over the fermionic fields introduces the factor $e^{0.292 N_f}$ and the running gauge coupling now contains fermionic contributions via $b_0 \to b_0 - 2/3 N_f$. Finally, the path integral over the Higgs field gives a nonzero contribution to the amplitude provided all Higgs fields are contracted where $(N_f - 1)!!$ is the number of Higgs contractions and the quantity $\mathcal{I}$ is given by [38,40]

$$\mathcal{I} = -\int d^4x_1 \int d^4x_2 \bar{\psi}_i^{(0)}(x_1)\psi_i^{(0)}(x_1)\bar{\psi}_j^{(0)}(x_2)\psi_j^{(0)}(x_2) \times \Delta_H(x_1 - x_2),$$
$$= \frac{\rho^4}{4\pi^8} \int d^4x_1 \int d^4x_2 \int d^4k \frac{1}{k^2 + m_H^2} \frac{e^{-ik(x_1 - x_0)}}{((x_1 - x_0)^2 + \rho^2)^3} \times \frac{e^{ik(x_2 - x_0)}}{((x_2 - x_0)^2 + \rho^2)^3},$$
$$\approx \begin{cases} \frac{1}{12\pi^2\rho^2} & m_H\rho \ll 1, \\ \frac{1}{5\pi^2 m_H^2 \rho^4} & m_H\rho \gg 1. \end{cases} \tag{18}$$

In the second line of (18) we have substituted for the scalar Feynman propagator $\Delta_H(x_1 - x_2)$ and the fermions have been replaced with their respective zero mode expressions given in [49]. Note that for an instanton background we have two zero modes $\bar{\psi}_{i,L}^{(0)}, \psi_{j,R}^{(0)}$ (and $\bar{\psi}_{i,R}^{(0)}, \psi_{j,L}^{(0)}$ in an anti-instanton background) where the subscripts L, R, which are suppressed hereon, denote left- and right-handed fields, respectively. Thus, combining (18) and (17) gives the final expression (assuming $m_H\rho \ll 1$)

$$\langle 0|0\rangle_{\Delta Q=1} = \int d^4x_0 \frac{d\rho}{\rho^5} C_f[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-S_E}, \tag{19}$$

with S_E defined in (8) (assuming $\theta \to \bar{\theta}$), and

$$C_f[N] \equiv (N_f - 1)!! \left(\frac{2}{3}\right)^{N_f/2} \left(\prod_{i=1}^{N_f} \frac{y_i}{4\pi}\right) e^{0.292 N_f} C[N]. \tag{20}$$

The expression (19) shows how the instanton density in the vacuum-to-vacuum amplitude is modified in the presence of massless fermions and a Higgs-fermion Yukawa interaction. As expected, the amplitude vanishes if any Yukawa coupling is zero. Thus, the topological susceptibility (14) in the presence of massless fermions is obtained by the substitutions $C[N] \to C_f[N]$, $\theta \to \bar{\theta}$, and $b_0 \to b_0 - 2/3 N_f$.

In the case of "large" instantons associated with the scale $1/\Lambda_{\mathrm{QCD}}$, the expression for the vacuum-to-vacuum amplitude differs from (19). As already mentioned, each light fermion ($m_f \lesssim \Lambda_{\mathrm{QCD}}$) introduces an $e^{0.292}\rho m_f$ factor.[2] This can be seen via the first line in (18) where Δ_H can be replaced by v^2, which just gives $\mathcal{I} = v^2$, and hence:

$$\langle 0|0\rangle_{\Delta Q=1} = \int d^4x_0 \frac{d\rho}{\rho^5} C[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-S_E} \times \left(\prod_{i=1}^{N_L} \rho m_i\right) e^{0.292 N_L}, \tag{21}$$

where the product runs only over N_L light fermions and $m_i = y_i v/\sqrt{2}$.

B. Weinberg gluonic operator

The Weinberg operator is a purely gluonic, CP odd, dimension six term given by $\mathcal{O}_W = GG\tilde{G}$ [56] that leads to the Lagrangian term

$$\mathcal{L} \supset \frac{1}{\Lambda_W^2} GG\tilde{G}, \tag{22}$$

[2]In QCD, $\chi(0) \propto m_f$, whereas the $\chi(0)$ resulting from (21) $\propto m_f^{N_L}$. The difference can be understood in terms of instanton-(anti-)instanton interactions—either via mixing between the fermion zero modes of the instanton with those of the anti-instanton [55] or using 't Hooft vertices with fermion legs joined between an instanton and anti-instanton [51].

where Λ_W is an effective UV scale. The operator (22) can induce a shift in the axion potential minimum, which can be computed by considering the mixed correlator [13,57]

$$\chi_W(0) = -i \lim_{k\to 0} \int d^4x e^{ikx} \times \langle 0|T\left\{\frac{1}{32\pi^2} G\tilde{G}(x), \frac{1}{\Lambda_W^2} GG\tilde{G}(0)\right\}|0\rangle. \quad (23)$$

In the instanton background (5) we obtain

$$\mathcal{O}_W = f_{abc} G^a_{\mu\kappa} G^b_{\kappa\nu} \tilde{G}^{c\nu\mu}(x) = -\frac{1536\rho^6}{((x-x_0)^2+\rho^2)^6}, \quad (24)$$

where f_{abc} are the structure constants. Note that for an $SU(N)$ gauge group, the $SU(2)$ instanton solution is embedded in the top left corner of the $N \times N$ matrix of $SU(N)$ generators. Thus, the sum in (24) only gives nonzero contributions for a, b, $c = 1, 2, 3$. Furthermore,

$$\begin{aligned} \langle 0|T\{G\tilde{G}(x), GG\tilde{G}(0)\}|0\rangle_{Q=+1} &= \int \mathcal{D}A_\mu G\tilde{G}(x) GG\tilde{G}(0) e^{-\frac{8\pi^2}{g_0^2}}, \\ &= \int d^4x_0 \frac{d\rho}{\rho^5} C[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-\frac{8\pi^2}{g^2(1/\rho)}} \\ &\quad \times \frac{192\rho^4}{((x-x_0)^2+\rho^2)^4} \frac{-1536\rho^6}{(x_0^2+\rho^2)^6}. \end{aligned} \quad (25)$$

Again performing the integrals first over x and then x_0 gives

$$\chi_W(0) = 2i \frac{384\pi^2}{5\Lambda_W^2} \int \frac{d\rho}{\rho^7} C[N] \left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-\frac{8\pi^2}{g^2(1/\rho)}}, \quad (26)$$

where we have also included the anti-instanton contribution.

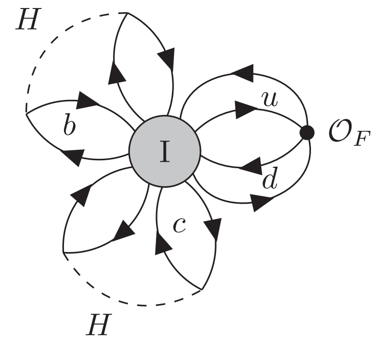
In the presence of fermions, $\chi_W(0)$ is obtained by making the substitutions $C[N] \to C_f[N]$ for small instantons [or by introducing the factor $(\rho m_f)^{N_L}$, as in (21) for large instantons], $\theta \to \bar{\theta}$ and $b_0 \to b_0 - 2/3N_f$ in the running gauge coupling $g(1/\rho)$.

C. Four-fermion operators

Another class of dimension six operators which can affect the axion solution are the four-fermion operators. Such operators are suppressed by an effective mass scale Λ_F and given by

$$\mathcal{L} \supset \sum_{ijkl} \frac{\lambda_{ijkl}}{\Lambda_F^2} \bar{\psi}_i \psi_j \bar{\psi}_k \psi_l, \quad (27)$$

where λ_{ijkl} are complex coefficients with flavor indices i, j, k, l. Note that the spinor and electroweak structure has been suppressed in (27), although it is straightforward to incorporate these details. Of particular interest is the spinor structure of (27) resulting in CP violation. These are operators of the type $\mathcal{O}_{F,ijkl} = \bar{\psi}_i i\gamma_5 \psi_j \bar{\psi}_k \psi_l$ which are anti-Hermitian with the corresponding λ_{ijkl} purely imaginary.



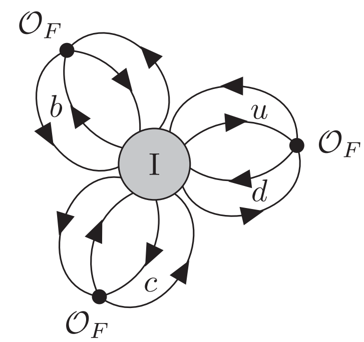


FIG. 2. The t'Hooft vertex that includes the insertion of four-fermion operators. Fermion legs are closed with one four-fermion operator $\mathcal{O}_F$ and two Higgs-fermion Yukawa interactions (a) and three four-fermion operators $\mathcal{O}_F$ (b).

The CP-violating effect arising from (27) can be obtained by including the four-fermion interactions in the path integral (17). These operators allow for new ways to close the fermion legs in the 't Hooft vertex, as depicted in Fig. 2. The largest contribution arises from just one insertion of $\mathcal{O}_F$, as shown in Fig. 2(a), while more insertions of the four-fermion operator, such as in Fig. 2(b) are suppressed by powers of Λ_F. Similar to the definition (23) for $\chi_W(0)$ we can define a fermion mixed correlator

$$\chi_{F,ijkl}(0) = -i \lim_{k\to 0} \int d^4x e^{ikx} \times \langle 0|T\left\{\frac{1}{32\pi^2} G\tilde{G}(x), \frac{\lambda_{ijkl}}{\Lambda_F^2} \mathcal{O}_{F,ijkl}(0)\right\}|0\rangle. \quad (28)$$

The only operators contributing to the fermion path integral are those with two pairs of flavor indices ($i = j \neq k = l$ or $i = l \neq k = j$), i.e., $\mathcal{O}_{F,iijj}$, and $\mathcal{O}_{F,ijji}$, both of which are hereon generically referred to as $\mathcal{O}_{F,ij}$ with the corresponding coupling constant $\lambda_{ij} \equiv \lambda_{iijj}$ (or λ_{ijij}). The explicit expression for such a generic operator $\mathcal{O}_{F,ij}$ can be computed as

$$\chi_{F,ij}(0) = -2i\int d^4x_0 \frac{d\rho}{\rho^5} C[N] e^{0.292N_f}\left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N} e^{-\frac{8\pi^2}{g^2(1/\rho)}}$$
$$\times \frac{2\lambda_{ij}}{y_i y_j}(N_f-3)!!\left(\prod_{k=1}^{N_f}\frac{y_k\rho}{\sqrt{2}}\right)\mathcal{I}^{N_f/2-1}\frac{1}{\Lambda_F^2}\bar{\psi}_i^{(0)} i\gamma_5\psi_i^{(0)}\bar{\psi}_j^{(0)}\psi_j^{(0)}(0)\frac{1}{32\pi^2}\int d^4x G\tilde{G}(x),$$
$$= 2i\frac{2(-i\lambda_{ij})}{y_i y_j}\int\frac{d\rho}{\rho^5}\frac{C_f[N]}{N_f-1}\left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N}\frac{12}{5\rho^2\Lambda_F^2}e^{-\frac{8\pi^2}{g^2(1/\rho)}}, \tag{29}$$

where we have also included the effect of the anti-instanton. The part of $\mathcal{O}_{F,ij}$ contributing to the path integral in the instanton background is $i\psi^\dagger_{L,i}\psi_{R,i}\psi^\dagger_{L,j}\psi_{R,j}$, while in the anti-instanton background (where $G\tilde{G}\to -G\tilde{G}$) it is $-i\psi^\dagger_{R,i}\psi_{L,i}\psi^\dagger_{R,j}\psi_{L,j}$. These two contributions add up[3] to give the factor of $2i$ in (29).

The result (29) can also be understood in terms of the results (17) and (18) from the fermionic path integral, up to the overall ratio of couplings. If we assume that $\mathcal{O}_F$ is generated by a heavy scalar of mass Λ_F, interacting with Standard Model quarks via Yukawa interactions, (18) implies a factor of $12\pi^2\rho^2/5\pi^2\Lambda_F^2\rho^4 = 12/5\rho^2\Lambda_F^2$ relative to the expression (19), which matches the factor inside the integral. The factor $1/(N_f-1)$ arises from having a fewer number of contractions-$(N_f-3)!!$ compared to (17), assuming only one insertion of the operator $\mathcal{O}_{F,ij}$.

Furthermore, notice that y_i and y_j have been explicitly factored out of (29) to write the result in terms of $C_f[N]$ defined in (20). For $-i\lambda_{ij}\sim 1$, this shows that the effect of the four-fermion operator, being $\propto 1/y_i y_j$, is most enhanced for the up and down quarks compared to that from the Weinberg gluonic operator or the second and third generation quarks. However, the four-fermion operator coefficient λ_{ij} can be chirally suppressed by Yukawa couplings [43]. For example, such four-fermion operators with a chiral suppression can arise from the overlap of fermion profiles in extra dimension models [58]. Thus, we will henceforth assume that $-i\lambda_{ij}\propto y_i y_j$ so that the effect of the four-fermion operator is similar to that of the Weinberg gluonic operator as well as the contributions from the other generations of quarks.

Assuming $-i\lambda_{ij} = y_i y_j/2$, we then have $N_f(N_f-1)$ contributions of the fermion susceptibility (29) for both types of operators $\mathcal{O}_{F,iijj}$, and $\mathcal{O}_{F,ijji}$, each. Thus, for $N_f = 6$ we obtain

$$\chi_F(0) \equiv 2N_f(N_f-1)\chi_{F,ij}(0),$$
$$= 2i\frac{144}{5\Lambda_F^2}\int\frac{d\rho}{\rho^7}C_f[N]\left(\frac{8\pi^2}{g^2(1/\rho)}\right)^{2N}e^{-\frac{8\pi^2}{g^2(1/\rho)}}. \tag{30}$$

Using (30) we will place limits on a generic scale Λ_F that represents all of these fermion effects.

Finally, note that in supersymmetric theories the operator $\mathcal{O}_F$ can arise from a dimension-four term in the superpotential [59]. After integrating out the scalar superpartners this leads to a four-fermion term with

$$\frac{1}{\Lambda_F^2}\sim\frac{g^2}{16\pi^2}\frac{1}{\Lambda_{\rm UV}m_{\rm SUSY}}, \tag{31}$$

where $\Lambda_{\rm UV}$ is the UV scale of the superpotential term and $m_{\rm SUSY}$ is the supersymmetry-breaking scale of the scalar superpartners. The bounds on Λ_F can thus be interpreted as bounds on the scalar superpartner masses.

III. INDUCED THETA

Using the results in Sec. II we can now obtain an estimate for the shift in the axion potential minimum due to CP-odd operators. In the presence of the Weinberg operator the axion potential is modified by a linear term in the axion field

$$V(a) = \chi_W(0)\left(\frac{a}{f_a}\right) + \frac{1}{2}\chi(0)\left(\frac{a}{f_a}\right)^2, \tag{32}$$

where we have promoted the theta angle to the axion field, $\bar{\theta}\to a/f_a$. This leads to a shift in the potential minimum by an amount

$$\left\langle\frac{a}{f_a}\right\rangle \equiv \theta_{\rm ind} = -\frac{\chi_W(0)}{\chi(0)}. \tag{33}$$

In the case of four-fermion operators the linear potential term again causes a shift in the potential minimum given by (33), with $\chi_W(0)$ replaced by $\chi_F(0)$.

The induced θ then directly contributes to EDM observables such as the neutron EDM where

[3] Instead, for CP-even operators of the type $\bar{\psi}_i\psi_i\bar{\psi}_j\psi_j$ there is a cancellation between the two contributions since $\psi^\dagger_{L,i}\psi_{R,i}\psi^\dagger_{L,j}\psi_{R,j}$ and $\psi^\dagger_{R,i}\psi_{L,i}\psi^\dagger_{R,j}\psi_{L,j}$ both appear with the same sign.

$$d_n \propto \frac{m_q}{\Lambda_{\rm QCD}^2}|\theta_{\rm ind}| = \frac{m_q}{\Lambda_{\rm QCD}^2}\frac{\chi_{W,F}(0)}{\chi(0)}. \tag{34}$$

The experimental limit arising from the neutron EDM gives the constraint

$$|\theta_{\rm ind}| \lesssim 10^{-10}, \tag{35}$$

which can now be used to obtain constraints on various heavy axion scenarios.[4]

A. QCD

We first consider the effect of dimension six operators in QCD with N_L light fermions (i.e., $m_f \lesssim \Lambda_{\rm QCD}$). The induced θ (33) that arises from including the Weinberg operator is given by

$$\theta_{\rm ind}^{\rm QCD} \approx \xi_W \frac{b_0 - 4 + N_L}{b_0 - 6 + N_L}\frac{\Lambda_{\rm QCD}^2}{\Lambda_W^2}, \tag{36}$$

where $\xi_W = 384\pi^2/5$, b_0 is the β-function coefficient and the CP-violation scale $\Lambda_{\rm CP}$ is identified with Λ_W. Note that in (36) the product of all light quark masses cancel and the induced θ becomes small (or decouples) as $\Lambda_W \to \infty$. Imposing the constraint (35) for QCD ($b_0^{\rm QCD} = 9$, $N_L = 3$ and $\Lambda_{\rm QCD} \approx 300$ MeV), gives the limit $\Lambda_W \gtrsim 10^6$ GeV on the effective scale of the Weinberg operator.

For the case of the CP-odd four-fermion operator, the 't Hooft vertex now has two fewer factors of ρm_f compared to the topological susceptibility resulting from (21). This gives a bound similar to Λ_W when there is no chirality suppression in the four-fermion operator, otherwise the Λ_F bound is much weaker. A calculation for $\theta_{\rm ind}$ using the chiral anomaly can be found in [60], which agrees with our estimate of the bound on Λ_F within an order of magnitude.

As such, current constraints on the neutron EDM correspond to new CP-violating physics at $\sim 10^6$ GeV. Thus, future neutron EDM experiments can probe new CP-violating sources at scales ranging from $\sim 10^6$–10^9 GeV, beyond which the SM contribution due to the CKM phase becomes comparable in size.

B. 4D small instantons

1. Product gauge group

A heavy axion can be generated by extending the QCD gauge group into a product gauge group $SU(3)^k = SU(3)_1 \times SU(3)_2 \times \ldots \times SU(3)_k$ which is spontaneously broken at a scale $\Lambda_{\rm SI}$ [39,40]. Small instantons at the scale $\Lambda_{\rm SI}$ associated with the product gauge groups lead to this enhancement. The SM quarks are assumed to be charged under only $SU(3)_1$. In addition, there are k axions, labeled by i, which couple to the k SU(3) $G\tilde{G}$ terms with decay constants f_{a_i}, eliminating the k theta terms.

At the scale $\Lambda_{\rm SI}$ the QCD gauge coupling α is matched to the $SU(3)_k$ gauge couplings α_i via the relation

$$\frac{1}{\alpha(\Lambda_{\rm SI})} = \sum_{i=1}^{k}\frac{1}{\alpha_i(\Lambda_{\rm SI})}. \tag{37}$$

This relation implies that each individual coupling α_i must be larger than the QCD coupling at the scale $\Lambda_{\rm SI}$. Therefore, the larger couplings $\alpha_i(\Lambda_{\rm SI})$ can make the small instanton effects dominate over the usual QCD large instantons. This effect is most dominant in the limit $k \gg 1$, where the axion masses scale as $m_{a_1} \sim \sqrt{\Pi_f y_f}\Lambda_{\rm SI}^2/f_{a_1}$ (with y_f the quark Yukawa couplings) and $m_{a_i} \sim \Lambda_{\rm SI}^2/f_{a_i}$ for $i = 2\ldots, k$, showing that the lightest axion mass (m_{a_1}) can remain much heavier that the QCD axion mass for $\Lambda_{\rm SI} \gg \Lambda_{\rm QCD}$.

For concreteness, let us consider the case with small k, where there is some perturbative control and the instanton (or anti-instanton) background still gives us qualitatively accurate results. Assuming the product gauge group is broken by scalars with a VEV, v_ϕ, the effective cutoff for the instanton size then becomes $2\pi v_\phi$, in contrast to the naive expectation, $\Lambda_{\rm SI}$ [40]. The constraint (35) can then be used to obtain limits on the scales associated with the sources of CP violation from the Weinberg and four-fermion operators. Since the QCD instanton contribution to $\chi_{W,F}(0)$ is suppressed by at least $\Lambda_{\rm QCD}^2/\Lambda_{W,F}^2$, the small instanton contribution from the UV gauge group dominates and results in

$$\theta_{\rm ind} \approx \xi_{W,F}\frac{2}{b_{0,i}-6}\frac{(2\pi v_\phi)^2}{\Lambda_{W,F}^2} \approx \xi_{W,F}\frac{8\pi^2}{b_{0,i}-6}\frac{\Lambda_{\rm SI}^2}{\Lambda_{W,F}^2}, \tag{38}$$

where $\xi_F = 24N_f/5$, ξ_W is defined under (36), and we have assumed $\Lambda_{\rm SI} \approx v_\phi$ in the second expression in (38). The constraint (35) then implies $\Lambda_{\rm SI}/\Lambda_W \lesssim 10^{-8}$ and $\Lambda_{\rm SI}/\Lambda_F \lesssim 10^{-7}$ or $\Lambda_{\rm SI} \lesssim 10^{10}(10^{11})$ GeV for $\Lambda_W(\Lambda_F) = M_P$ where $M_P = 2.4 \times 10^{18}$ GeV is the (reduced) Planck mass,[5] $b_{0,1} = 13/2$ and $b_{0,k} = 21/2$. For $i = 2, \ldots, k-1$, the same expression (38) holds with $b_{0,i} = 10$, and $v_\phi \to \sqrt{2}v_\phi$, which does not change the bounds significantly.[6] Note that if UV couplings are included in (22) then the effective

[4]For simplicity, we will present limits that arise from the individual operators $\mathcal{O}_W$ and $\mathcal{O}_F$ separately. Our results can be straightforwardly generalized by summing the contributions in (33) if both operators are present.

[5]The difference in these two bounds results from the size of the different prefactors $\xi_{W,F}$, where ξ_W results from the large number of color contractions in (23), while ξ_F arises from the smaller flavor multiplicity of the four-fermion operator (27).

[6]It is possible that the axion mass could instead be dominated by QCD large instantons. But in this case the CP violation arising from small instantons of the product gauge group gives the much weaker constraint that $\Lambda_{\rm SI}/\Lambda_W \lesssim 10^{-8} \times m_{a,\rm QCD}/m_{a_1}$. For instance, assuming $m_{a,\rm QCD}/m_{a_1} = 10^3$ implies that $\Lambda_{\rm SI} \lesssim 10^{13}$ GeV for $\Lambda_W = M_P$.

scale Λ_W can be larger than M_P. Assuming $f_a > \Lambda_{\rm SI}$, the limits on $\Lambda_{W,F}$ correspond to a maximum possible axion mass enhancement of $\sim 10^7$ for $k = 3$ relative to the QCD axion [39,40]. As such, axion masses $m_a \gtrsim 100$ MeV with $f_a \lesssim 10^7$ GeV [61,62] can be explored in future experimental searches.

However, when $f_a < \Lambda_{\rm SI}$, we need to UV complete the dimension five axion-$G\tilde{G}$ coupling and explain the PQ breaking. This can be done in a minimal KSVZ-type scenario [7,8], by introducing a single heavy Dirac fermion Ψ, with mass m_Ψ, charged under the $U(1)_{PQ}$ symmetry, which changes the instanton measure by a factor of $e^{0.292}\rho m_\Psi$. Combining this with the contribution arising from the running of the gauge coupling between m_Ψ and $\Lambda_{\rm SI}$, the topological susceptibility (or any similar correlator) is modified to

$$\chi(0) \to \chi(0)\frac{b_0-4}{b_0-11/3}\left(\frac{m_\Psi}{\Lambda_{\rm SI}}\right)^{-2/3} e^{0.292}\left(\frac{m_\Psi}{\Lambda_{\rm SI}}\right), \\ \approx \chi(0)\left(\frac{f_a}{\Lambda_{\rm SI}}\right)^{1/3}, \tag{39}$$

where the Yukawa coupling between Ψ and the PQ scalar is assumed to be order one, i.e., $m_\Psi \approx f_a$. Since $m_a^2 \propto \chi(0)$ this suppresses the axion mass enhancement by an amount $(f_a/\Lambda_{\rm SI})^{1/6}$ [63]. Thus for the experimentally interesting region of $m_a \gtrsim 100$ MeV and $f_a \lesssim 10^7$ GeV, the axion mass enhancement is reduced by up to a factor of 10 when $f_a < \Lambda_{\rm SI}$.

A similar result is also obtained for an enlarged color group [32,35,36] where $\Lambda_{\rm SI}$ is identified with the scale where the enlarged symmetry group is broken and the appropriate b_0 is used. In all these cases, there is again a nondecoupling effect that depends on the ratio $\Lambda_{\rm SI}/\Lambda_{W,F}$.

2. Mirror QCD

A heavy axion can also be obtained by assuming that there exists a $\mathbb{Z}_2$ mirror copy of QCD that becomes strong at a scale $\Lambda'_{\rm QCD}(\equiv\Lambda_{\rm SI}) \gg \Lambda_{\rm QCD}$ [22,23,30,31,34]. The axion is $\mathbb{Z}_2$ neutral and couples to both QCD and mirror QCD, via the interaction

$$\frac{1}{32\pi^2}\frac{a}{f_a}\varepsilon^{\mu\nu\rho\sigma}(G^c_{\mu\nu}G^c_{\rho\sigma} + G'^c_{\mu\nu}G'^c_{\rho\sigma}), \tag{40}$$

where $G'_{\mu\nu}$ is the mirror QCD field strength. The axion now receives contributions from the mirror QCD instantons (which are small in size relative to those from QCD) and gives rise to limits on higher dimensional operators with scales $\Lambda_{W,F}$ involving gluons and fermions in the mirror sector.

The mirror QCD expression for the induced θ due to the Weinberg operator can be obtained by substituting $\Lambda_{\rm SI}$ in (36). This leads to the bounds $\Lambda_{\rm SI}/\Lambda_W \lesssim 10^{-7}$ or $\Lambda_{\rm SI} \lesssim 10^{11}$ GeV for $\Lambda_W = M_P$, assuming the mirror Higgs VEV $v' \gg \Lambda_{\rm SI}$ such that QCD′ is a pure Yang-Mills theory at $\Lambda_{\rm SI}$ with $b_0^{\rm QCD'} = 11$. These bounds for the Weinberg operator do not change appreciably if this assumption is relaxed.

The induced θ from the four-fermion operator can be obtained by considering $N_L \geq 2$ light flavors in QCD′. Applying the QCD result (21) for QCD′ then gives

$$\theta_{{\rm ind},F} \approx \frac{2N_L(N_L-1)}{5\pi^2}\frac{b_0-4+N_L}{b_0-8+N_L}\frac{\Lambda_{\rm SI}^4}{v'^2\Lambda_F^2}, \\ \approx \frac{2N_L(N_L-1)}{5\pi^2}\frac{b_0-4+N_L}{b_0-8+N_L}\frac{\Lambda_{\rm SI}^2}{\Lambda_F^2}, \tag{41}$$

where we have taken $v' \approx \Lambda_{\rm SI}$ in the last expression in (41). Assuming $b_0 = 9$ and $N_L = 3$, implies $\Lambda_{\rm SI}/\Lambda_F \lesssim 10^{-5}$, or $\Lambda_{\rm SI} \lesssim 10^{13}$ GeV for $\Lambda_F = M_P$. Again, the difference in the $\Lambda_{W,F}$ bounds arises from the different color and flavor multiplicity factors.

C. 5D small instantons

Another way for the QCD coupling to become large at a UV scale and increase the effect of small instantons is to consider a 5D model where QCD gluons propagate in a fifth dimension of size R. The axion can be identified with a UV boundary localized field that couples to QCD via a coupling proportional to $1/f_a$, with f_a an independent parameter of the theory. This allows the decay constant to be either above or below the small instanton scale and allows for more general possibilities. Above the scale $1/R$ the QCD coupling increases in strength until the coupling becomes strong at the cutoff scale Λ_5 which is defined by the relation [41]

$$\Lambda_5 R = \frac{6\pi\epsilon}{\alpha(1/R)}, \tag{42}$$

where $\alpha = g^2/(4\pi)$ and $\epsilon \leq 1$ is a perturbativity parameter.[7] The small instanton scale can be identified as $\Lambda_{\rm SI} \equiv \Lambda_5$. The 4D effective action is approximately given by [41]

$$S_{\rm eff} \approx \frac{2\pi}{\alpha_s(1/R)} - \frac{R}{\rho} + b_0 \ln\frac{R}{\rho}, \tag{43}$$

where the power-law term R/ρ arises from summing over the 5D Kaluza-Klein gluons. Thus, small instantons of size $1/\Lambda_{\rm SI} \lesssim \rho \lesssim R$ can now reduce the effective action and contribute greatly to the path integral.

Using an approximate expression for the integrals in (14) and (26) with the effective action (43), the induced θ from 5D small instantons is

[7] Note that in the 5D model, small instantons can be made to dominate when perturbativity still holds. This implies that our instanton (or anti-instanton) approximation used for the correlators will give more accurate quantitative results relative to QCD.

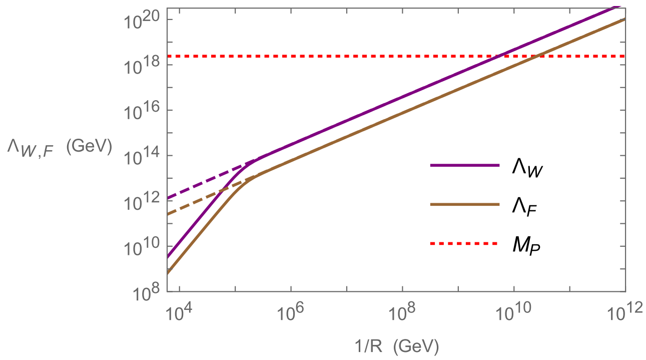


FIG. 3. Lower limit on the effective scale of the dimension six Weinberg (four-fermion) operator, depicted in purple (orange) as a function of the extra dimension scale $1/R$, assuming $\epsilon = 0.30$. The Planck scale is shown as a dotted line for reference. The dashed lines represent the limit from the approximation (44). The deviation from (44) arises since for small $1/R$, large QCD instantons begin to dominate the instanton integral $\chi(0)$.

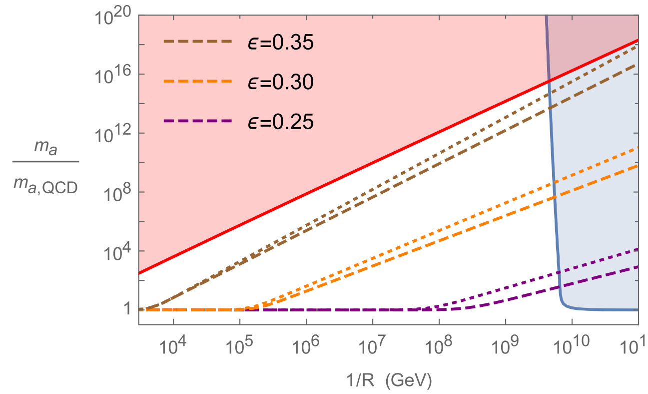


FIG. 4. The ratio of the enhanced axion mass to the QCD axion mass as a function of the extra dimension scale, $1/R$. The dotted contour lines assume $f_a > \Lambda_{\mathrm{SI}}$ and depict the ratio for different values of the perturbitivity parameter ϵ, up to the maximum possible enhancement in the red shaded region. The dashed contours assume $f_a = 10^6$ GeV and include the suppression (39) when $f_a < \Lambda_{\mathrm{SI}}$. The blue shaded region to the right shows the excluded $1/R$ range due to the Weinberg gluonic operator.

$$\theta_{\mathrm{ind}} \approx \xi_{W,F} \frac{\Lambda_{\mathrm{SI}}^2}{\Lambda_{W,F}^2}, \tag{44}$$

where ξ_W and ξ_F are defined under (36) and (38), respectively. The induced θ no longer necessarily decouples in the limit $\Lambda_{\mathrm{SI}}, \Lambda_{W,F} \to \infty$. Imposing the constraint (35) leads to the limit $\Lambda_{\mathrm{SI}}/\Lambda_W(\Lambda_F) \lesssim 10^{-7}(10^{-6})$. For $\Lambda_W(\Lambda_F) = M_P$ this implies an upper bound $\Lambda_{\mathrm{SI}} \lesssim 10^{11}(10^{12})$ GeV on the 5D strong coupling scale. The limit on $\Lambda_{W,F}$ from an exact numerical evaluation of θ_{ind} is shown in Fig. 3. We see that the limit on $\Lambda_{W,F}$ deviates from (44) for small $1/R$ (and hence small Λ_5). The limits on the ratio $\Lambda_{\mathrm{SI}}/\Lambda_{W,F}$ imply that for the case when $\Lambda_{W,F} \sim \Lambda_5(=\Lambda_{\mathrm{SI}})$, the dimension six terms would need to be generated from some new physics in the UV completion of the 5D model with an additional suppression in the otherwise order-one coefficients.

The corresponding range of axion mass enhancement is depicted in Fig. 4. Note that both effects of small instantons—the enhancement of the axion mass and the shift in the axion potential minimum due to CP-violating operators—are dominant only for large $1/R$, since eventually large (QCD) instantons dominate the susceptibility at small values of $1/R$.

Furthermore, when $f_a < \Lambda_{\mathrm{SI}}$ the axion mass enhancement is reduced by the factor $(f_a/\Lambda_{\mathrm{SI}})^{1/6}$ as obtained from (39). This means that in the experimentally viable region of $m_a \gtrsim 100$ MeV and $f_a \lesssim 10^7$ GeV [61,62], the axion mass enhancement is reduced by up to an order of magnitude, as can be seen in Fig. 4, where we have taken $f_a = 10^6$ GeV as a representative value.

D. Enhanced EDMs

Compared to QCD, the small instanton contributions provide an enhancement to the EDMs due to CP-violating sources. In particular, using (38), (41), and (44) we see that the neutron EDM (34) is enhanced by a factor of $\Lambda_{\mathrm{SI}}^2/\Lambda_{\mathrm{QCD}}^2$ compared to the θ induced from new CP-odd sources in QCD [see (36)]. Therefore, measuring the neutron EDM can be interpreted as a probe of the small instanton scale, Λ_{SI}. For example, if $\Lambda_{W,F} = M_P$, this corresponds to modified strong dynamics at scales of order $\Lambda_{\mathrm{SI}} \sim 10^8$–$10^{11}$ GeV, where the lower limit represents a neutron EDM value equivalent to the Standard Model CKM contribution. Furthermore, if the CP-violating sources appear at scales lower than the Planck scale, then any new contribution due to small instantons will appear at even lower scales 10^4 GeV $\lesssim \Lambda_{\mathrm{SI}} \lesssim 10^8$ GeV, where the model-dependent lower limit corresponds to the scale of axion mass enhancement.

Finally note that when $f_a \lesssim \Lambda_{\mathrm{SI}}$, the UV completion of the dimension five axion-gluon coupling does not affect the predictions for the induced θ. Since the neutron EDM (34) depends only on the ratio of the mixed correlators with the topological susceptibility, the suppression factor in (39) cancels, leaving the results for the induced θ unchanged.

IV. CONCLUSION

The QCD axion solution provides an elegant mechanism for solving the strong CP problem in such a way that an arbitrarily large amount of CP violation at UV scales Λ_{CP} can be sufficiently decoupled as $\Lambda_{\mathrm{CP}} \to \infty$. This is in contrast with solutions to the strong CP problem that invoke exact discrete symmetries. For these solutions there is a nondecoupling of the additional sources of CP violation, which means that arbitrarily large amounts of CP violation cannot be tolerated at UV scales in models with exact parity or CP symmetry.

Heavy axion models represent a qualitatively different class of solution to the strong *CP* problem in which new dynamics at some UV scale $\Lambda_{\rm SI}$ magnifies the effect of small instantons (which are normally exponentially suppressed), giving rise to a new contribution and enhancement of the axion mass. This has led to renewed interest in axion searches outside the usual QCD axion mass window. However, in the presence of additional sources of *CP* violation, the enhanced effect of small instantons could also lead to enhanced EDM observables such as the neutron EDM as well as possible nondecoupling effects.

We have estimated these effects by calculating the topological susceptibility and mixed correlators in the presence of two *CP*-violating dimension six operators: the Weinberg gluonic operator and a *CP*-odd four-fermion operator. The calculation is performed using an instanton (or anti-instanton) background where Standard Model fermion chiral zero modes in the 't Hooft vertex are closed with the Higgs boson. Identifying the scale of the additional sources of *CP* violation with $\Lambda_{\rm CP}$ we find that the axion potential minimum shifts by an amount $\theta_{\rm ind} \propto \Lambda_{\rm SI}^2/\Lambda_{\rm CP}^2$ in several heavy axion models, where $\Lambda_{\rm SI}$ is the scale where small instanton effects dominate. This result reveals that unlike the minimal QCD axion models, the amount of decoupling is limited, although not as restrictive as models with exact discrete symmetries. Imposing the neutron EDM derived limit $|\bar{\theta}| \lesssim 10^{-10}$, we obtain the constraint $\Lambda_{\rm SI}/\Lambda_{\rm CP} \lesssim 10^{-8}$, which is stronger than the naive estimate of 10^{-5} due to sizable prefactors that depend on the particular heavy axion model. In particular, for a benchmark value of $\Lambda_{\rm CP} \simeq M_P$ requires $\Lambda_{\rm SI} \lesssim 10^{10}$ GeV (as can be seen in Fig. 4 for the 5D small instanton model).

The modification of the decoupling behavior is a direct consequence of the new dynamical scale $\Lambda_{\rm SI}$. Our results therefore imply that EDM observables such as the neutron EDM can be enhanced in heavy axion models up to the current experimental limit $d_n \lesssim 10^{-26}$ e · cm. This compares with the SM CKM prediction ($\sim 10^{-32}$–10^{-31} e · cm). Thus, besides axion searches, EDM observables provide another probe of UV scales in heavy axion models associated with new dynamics, assuming that this class of models plays any role in solving the strong *CP* problem.

ACKNOWLEDGMENTS

T. G. thanks Raymond Co and Keisuke Harigaya for useful discussions. This work is supported in part by the Department of Energy under Grant No. DE-SC0011842 at the University of Minnesota. T. G. is also supported by the Simons Foundation.

[1] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
[2] C. Abel *et al.*, Phys. Rev. Lett. **124**, 081803 (2020).
[3] B. Graner, Y. Chen, E. G. Lindahl, and B. R. Heckel, Phys. Rev. Lett. **116**, 161601 (2016); **119**, 119901(E) (2017).
[4] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[5] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[6] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[7] J. E. Kim, Phys. Rev. Lett. **43**, 103 (1979).
[8] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. **B166**, 493 (1980).
[9] M. Dine, W. Fischler, and M. Srednicki, Phys. Lett. **104B**, 199 (1981).
[10] A. R. Zhitnitsky, Sov. J. Nucl. Phys. **31**, 260 (1980).
[11] I. I. Y. Bigi and N. G. Uraltsev, Nucl. Phys. **B353**, 321 (1991).
[12] M. Pospelov and A. Ritz, Phys. Rev. D **63**, 073015 (2001).
[13] M. Pospelov and A. Ritz, Ann. Phys. (Amsterdam) **318**, 119 (2005).
[14] A. E. Nelson, Phys. Lett. **136B**, 387 (1984).
[15] S. M. Barr, Phys. Rev. Lett. **53**, 329 (1984).
[16] K. S. Babu and R. N. Mohapatra, Phys. Rev. Lett. **62**, 1079 (1989).
[17] K. S. Babu and R. N. Mohapatra, Phys. Rev. D **41**, 1286 (1990).
[18] R. N. Mohapatra and A. Rasin, Phys. Rev. Lett. **76**, 3490 (1996).
[19] R. Kuchimanchi, Phys. Rev. Lett. **76**, 3486 (1996).
[20] B. Holdom, Phys. Rev. D **61**, 011702 (1999).
[21] G. Hiller and M. Schmaltz, Phys. Lett. B **514**, 263 (2001).
[22] Z. Berezhiani, L. Gianfagna, and M. Giannotti, Phys. Lett. B **500**, 286 (2001).
[23] A. Hook, Phys. Rev. Lett. **114**, 141801 (2015).
[24] J. R. Ellis and M. K. Gaillard, Nucl. Phys. **B150**, 141 (1979).
[25] I. B. Khriplovich, Phys. Lett. B **173**, 193 (1986).
[26] M. E. Pospelov, Phys. Lett. B **391**, 324 (1997).
[27] P. H. Frampton and O. C. W. Kong, Phys. Lett. B **402**, 297 (1997).
[28] J. de Vries, P. Draper, and H. H. Patel, arXiv:2109.01630.
[29] S. Dimopoulos, Phys. Lett. **84B**, 435 (1979).
[30] V. A. Rubakov, JETP Lett. **65**, 621 (1997).
[31] H. Fukuda, K. Harigaya, M. Ibe, and T. T. Yanagida, Phys. Rev. D **92**, 015021 (2015).
[32] T. Gherghetta, N. Nagata, and M. Shifman, Phys. Rev. D **93**, 115010 (2016).
[33] T. Gherghetta and M. D. Nguyen, J. High Energy Phys. 12 (2020) 094.
[34] A. Hook, S. Kumar, Z. Liu, and R. Sundrum, Phys. Rev. Lett. **124**, 221801 (2020).
[35] B. Holdom and M. E. Peskin, Nucl. Phys. **B208**, 397 (1982).

[36] B. Holdom, Phys. Lett. **154B**, 316 (1985); **156B**, 452(E) (1985).
[37] M. Dine and N. Seiberg, Nucl. Phys. **B273**, 109 (1986).
[38] J. M. Flynn and L. Randall, Nucl. Phys. **B293**, 731 (1987).
[39] P. Agrawal and K. Howe, J. High Energy Phys. 12 (2018) 029.
[40] C. Csáki, M. Ruhdorfer, and Y. Shirman, J. High Energy Phys. 04 (2020) 031.
[41] T. Gherghetta, V. V. Khoze, A. Pomarol, and Y. Shirman, J. High Energy Phys. 03 (2020) 063.
[42] T. Gherghetta and A. Pomarol, J. High Energy Phys. 11 (2021) 136.
[43] R. Kitano and W. Yin, J. High Energy Phys. 07 (2021) 078.
[44] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, Phys. Lett. **59B**, 85 (1975).
[45] G. 't Hooft, Phys. Rev. D **14**, 3432 (1976); **18**, 2199(E) (1978).
[46] R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[47] E. Witten, Nucl. Phys. **B156**, 269 (1979).
[48] S. F. Cordes, Nucl. Phys. **B273**, 629 (1986).
[49] A. I. Vainshtein, V. I. Zakharov, V. A. Novikov, and M. A. Shifman, Usp. Fiz. Nauk **136**, 553 (1982) [Sov. Phys. Usp. **25**, 195 (1982)].
[50] C. G. Callan, Jr., R. F. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978).
[51] D. Diakonov and V. Y. Petrov, Nucl. Phys. **B245**, 259 (1984).
[52] E. V. Shuryak, Nucl. Phys. **B319**, 521 (1989).
[53] D. Diakonov and V. Y. Petrov, Nucl. Phys. **B272**, 457 (1986).
[54] D. Diakonov, M. V. Polyakov, and C. Weiss, Nucl. Phys. **B461**, 539 (1996).
[55] D. Diakonov and V. Y. Petrov, Phys. Lett. **147B**, 351 (1984).
[56] S. Weinberg, Phys. Rev. Lett. **63**, 2333 (1989).
[57] C. Weiss, Phys. Lett. B **819**, 136447 (2021).
[58] Q. Bonnefoy, P. Cox, E. Dudas, T. Gherghetta, and M. D. Nguyen, J. High Energy Phys. 04 (2021) 084.
[59] M. Pospelov, A. Ritz, and Y. Santoso, Phys. Rev. Lett. **96**, 091801 (2006).
[60] H. An, X. Ji, and F. Xu, J. High Energy Phys. 02 (2010) 043.
[61] D. Cadamuro and J. Redondo, J. Cosmol. Astropart. Phys. 02 (2012) 032.
[62] J. Jaeckel and M. Spannowsky, Phys. Lett. B **753**, 482 (2016).
[63] R. T. Co, T. Gherghetta, and K. Harigaya, arXiv:2206.00678.