

# Primer on Axion Physics

Felix Yu

The canonical axion potential is reviewed, with an emphasis on the field theory underlying radial and angular modes of complex scalar fields. The explicit calculation of the instanton-induced breaking of the Goldstone field direction necessary to derive the canonical axion mass and decay constant relation is presented. The primer is intended to serve an audience with elementary quantum field theory expertise.

## 1. Introduction

Axions are hypothesized particles originally proposed to resolve the strong CP problem of the Standard Model. The essential mechanism of the axion solution is that the quantum chromodynamics (QCD) instanton-generated potential for the axion is minimized when the vacuum expectation value of the axion exactly cancels the (unknown) original  $\Theta$  parameter, leaving an effective  $\Theta_{\text{eff}}$  parameter that is vanishing. In this review, we will see the fundamental distinction between the axion potential and other, more common scalar potentials usually considered in quantum field theories, such as the Higgs potential of the Standard Model. We will also analyze the field theory from the modern viewpoint of Peccei-Quinn (PQ) symmetry, again making a distinction between the role of PQ symmetries relevant for axion physics and the more traditional electroweak gauge symmetry emphasized in Higgs physics. The goals of this primer are to derive and motivate the axion potential as a characteristic and benchmark model of pseudo-Nambu-Goldstone boson (pNGB) Lagrangians as well as to develop the basic phenomenological signals of axion and axion-like particles (ALPs) as ultralight dark matter candidates.

The mathematical framework for fundamental laws of Nature, in particular the Standard Model, is quantum field theory, which is based on the quantum action  $S = \int d^4x \mathcal{L}$ , where the Lagrangian  $\mathcal{L}$  (technically, the Lagrangian density, but we always call it the Lagrangian) is required to respect Poincaré invariance. Poincaré symmetry is Lorentz symmetry  $\oplus$  translation symmetry, and is required to establish the familiar energy-momentum conservation laws as well as the characterization of particles according to half-integer representations for spin-angular momentum. This motivates starting with the simplest quantum field

theory, namely that of a scalar field whose excitations are scalar particles with spin zero.

## 2. Scalar Field Theory

We begin with the field theory of a single scalar field  $\phi$ . Scalar fields assign a pure real number to every spacetime point, and thus they take the spacetime coordinate as their argument. For a scalar field with mass  $m$  and no interactions, we can write the Lagrangian (density) as

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \phi)^2 - \frac{1}{2} m^2 \phi^2 \quad (1)$$

Analyzing the Euler-Lagrange equation of motion dictated by this Lagrangian will recover the familiar Klein-Gordon equation governing free scalars with a mass  $m$ , assuming  $m > 0$ .

If there is a second scalar field with exactly the same mass, then the two fields can be combined into a single complex scalar field  $\Phi$ , where the real part of  $\Phi$  is the number assigned by the original  $\phi$  field and the imaginary part of  $\Phi$  represents the second scalar field. Clearly, the Lagrangian in this case can be written as

$$\mathcal{L} = \left| \partial_\mu \Phi \right|^2 - m^2 |\Phi|^2 \quad (2)$$

Note the prefactors of the kinetic and mass term are twice as large as the real scalar Lagrangian, reflecting the existence of two distinct real degrees of freedom.

To connect to Higgs physics and the phenomenon of spontaneous symmetry breaking, we can now add a quartic interaction. We have

$$\mathcal{L} = \left| \partial_\mu \Phi \right|^2 - m^2 |\Phi|^2 - \lambda |\Phi|^4 \quad (3)$$

We remark that the Lagrangian is always the kinetic energy minus potential energy, and so we can analyze the potential

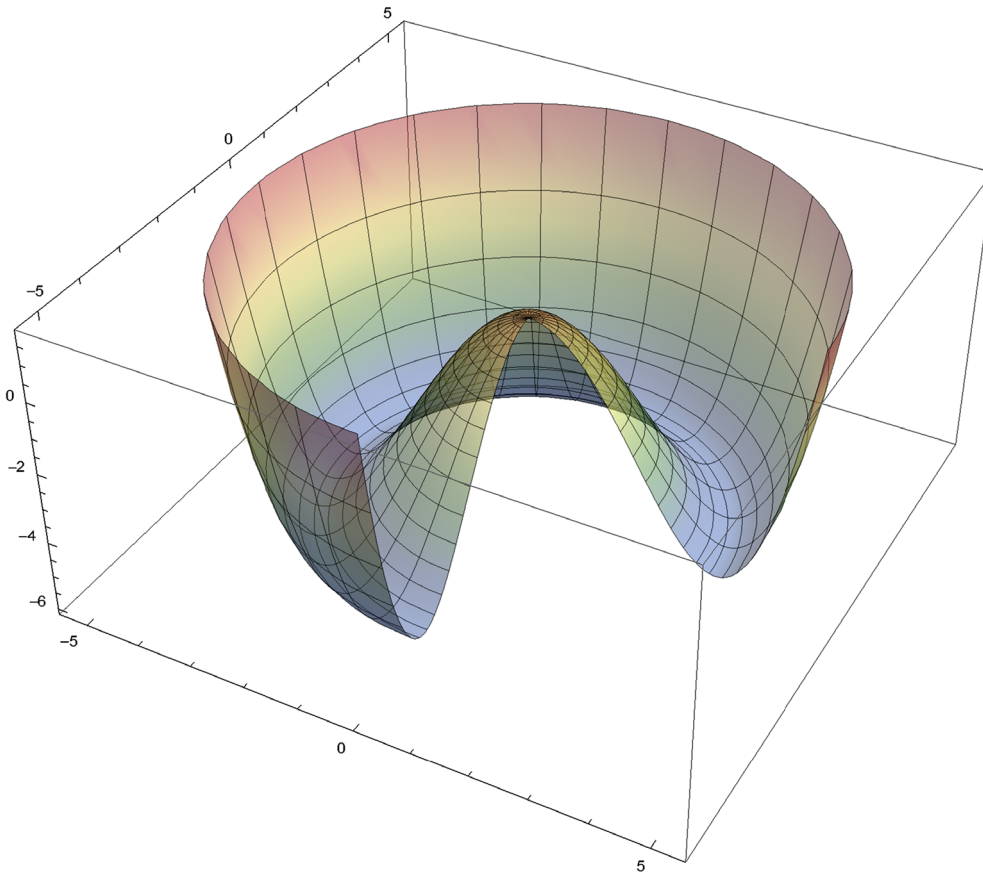
$$V = m^2 |\Phi|^2 + \lambda |\Phi|^4 \quad (4)$$

In order for the potential to be bounded from below, we require  $\lambda > 0$ , but we now have two possibilities for the sign of  $m^2$ . If  $m^2 > 0$ , the origin in field space is stable and the field has no vacuum expectation value (vev). If  $m^2 < 0$ , the origin in field space is unstable and the equation of motion for the scalar is extremized by a nonzero vev,  $|\Phi|^2 = v^2 = -m^2/\lambda$ . In either case, as we go to large excursions in field space, the potential energy grows because of  $\lambda > 0$ , but the behavior near the origin in field space will distinguish the two potentials. As an aside, if we coupled the complex scalar field to a  $U(1)$  gauge symmetry, replacing  $\partial_\mu$  by a covariant derivative  $D_\mu$  in the equation above, then the above

F. Yu  
PRISMA+ Cluster of Excellence & Mainz Institute for Theoretical Physics  
Johannes Gutenberg University  
55099 Mainz, Germany  
E-mail: yu001@uni-mainz.de

© 2023 The Authors. Annalen der Physik published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/andp.202300106



**Figure 1.** Schematic of the typical Higgs potential from a negative mass-squared term and a positive quartic term. The rotational symmetry is manifest for every point in field space and represents the Goldstone field direction.

Lagrangian would be the scalar part of the Abelian Higgs model, and the vev of the Higgs field would give a mass for the  $U(1)$  gauge boson.

Importantly, the vacuum state of the theory is controlled by one constraint equation,  $|\Phi|^2 = v^2$ , and therefore we have one degree of freedom that is unconstrained by the vacuum requirement. This is most evident when we change from a “Cartesian” field space parametrization to a “polar” parametrization,

$$\Phi = \left( \frac{v+h}{\sqrt{2}} \right) \exp(ia/v) \quad (5)$$

where  $h$  and  $a$  are the *radial* and *angular* modes of the complex field  $\Phi$ . When we write the original Lagrangian with this parametrization, we get

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 + \frac{1}{2}(\partial_\mu a)^2 + \frac{1}{4}\lambda v^4 - \lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 \quad (6)$$

where, we notice that the angular field  $a$  does not appear in the potential anymore, while the radial field  $h$  has a mass term  $m_h = \sqrt{2\lambda}v$  as well as cubic and quartic self-interactions. We also remark the appearance of a constant energy (density) term in the potential, which plays no role in the dynamics at hand but would

be relevant for the cosmological constant problem. The key point is that the angular field is a completely free and massless (pseudo)scalar field, and its (currently vanishing) dynamics are independent of the dynamics of the radial mode.

We remark that the pseudoscalar nature of the  $a$  field in Equation (6) is only meaningful in the context of the embedding as the angular mode of the original complex scalar field  $\Phi$ . In particular, comparing Equation (1) and the  $a$ -dependent term in Equation (6) shows that Equation (1) reduces to the  $a$ -dependent kinetic term if we neglect the mass  $m$  in Equation (1) and relabel  $\phi$  to  $a$ . While this relabeling seems entirely safe given that the Lagrangians now coincide, we have to recall that  $a$  is defined as a *compact* field direction via Equation (5) while the original free scalar field  $\phi$  lies on a noncompact field space. Formally, this distinction about compactness is reflected in the fact that the action  $S$  has an extra periodic redundancy for  $a \rightarrow a + 2\pi v$  but not for a massless  $\phi$ , as evident from the polar decomposition in Equation (5).

Importantly, the “zero point” in field space for  $a$  is entirely degenerate and lies along the entirety of the circle in **Figure 1**, as a result of the vev only being defined by its square. This is self-consistent with the Goldstone nature of  $a$  having a continuous shift symmetry in Equation (6), where any finite shift of  $a$  leaves the Lagrangian invariant. In fact, the absence of an origin in field

space is crucial for the axion solution to the strong CP problem, since it allows the axion to render any arbitrary  $\bar{\Theta}$  starting value to (effectively) zero via the PQ mechanism.

### 3. Instantons in Yang-Mills Theory and the $\Theta$ Vacuum Energy

In this section, we review the non-perturbative calculation of the instanton-induced vacuum energy of Yang-Mills theory. This calculation is taken from the tour-de-force publication of 't Hooft<sup>[1]</sup> and related literature on instantons.<sup>[2,3]</sup>

At the close of the last section, we demonstrated that the angular mode in a Higgsed potential is a free field with only a kinetic term. In more formal terms, the angular mode is a pseudoscalar field that has a continuous shift symmetry, meaning that any shift in the field has no effect on the Lagrangian. In particular, the field has no mass terms and no interactions, since masses and interactions violate the continuous shift symmetry. We remark that the generic nature of Goldstone bosons arising in ultraviolet extensions of the Standard Model can typically be mapped to the field theory at hand, although the key requirements to satisfy are that the ultraviolet fields exhibit Goldstones traversing a compact field space and the ultraviolet continuous symmetries are not gauged.<sup>[4]</sup> This will be relevant for the extension of the axion parameter space to ALPs.

The continuous shift symmetry of the angular mode is explicitly broken if the symmetry has an anomaly. Anomalies are, by definition, quantum origins of symmetry breaking, and in the context of the axion, originate from the fact that the continuous shift symmetry for the pseudoscalar field is broken to a periodic shift symmetry from its axial coupling to fermions carrying color charges. The fact that the continuous shift symmetry has a nonzero anomaly with respect to the SM color gauge group is the defining property of Peccei-Quinn symmetries, and this  $U(1)_{PQ} \times SU(3)_c^2$  anomaly explicitly breaks  $U(1)_{PQ}$  to a periodic shift symmetry and reflected in the coupling of the axion field to the dual field strength tensor of QCD. At low energies, the dual field strength tensor of QCD leads to an instanton-induced potential for the axion, which we will derive in this section. We remark that the periodicity condition Equation (5) and the color anomaly quantizes the vacuum structure of the axion field space according to a topological winding number, permitting domain wall solutions of the  $a$  field interpolating between energetically degenerate but topologically distinct vacuum states.

#### 3.1. Yang-Mills Theory and Instantons

Instantons are a feature of Yang-Mills theories. The Yang-Mills action for an  $SU(N)$  gauge theory in Minkowski metric is

$$\mathcal{L} = -\frac{1}{4} G_{\mu\nu}^a G^{\mu\nu,a} \quad (7)$$

where  $G_{\mu\nu}^a \equiv \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c$ . The simplest method for deriving the instanton solution follows Belavin, Polyakov, Schwartz, and Tyupkin (BPST).<sup>[5]</sup>

First, we define  $\tilde{G}^a = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}^a$  for the totally antisymmetric tensor  $\epsilon_{\mu\nu\rho\sigma}$  and  $\epsilon_{0123} = 1$ . Then, since  $G\tilde{G}$  is a total derivative, and since we are looking for classical field solutions to the action with minimized action, a unit instanton satisfies the quantization constraint

$$\int d^4 x_E \frac{g_s^2}{32\pi^2} G\tilde{G} = \int d^4 x_E \frac{g_s^2}{64\pi^2} \epsilon^{\mu\nu\rho\sigma} G_{\mu\nu}^a G_{\rho\sigma}^a = 1 \quad (8)$$

where  $x_E$  is the Euclidean spacetime coordinate. We will relate this quantization constraint to the Chern-Simons topological current shortly.

The procedure from BPST is to consider the minimal action of the gauge fields arising from the inequality

$$\begin{aligned} 0 &\leq \int d^4 x_E \left( G_{\mu\nu}^a \pm \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}^a \right)^2 \\ &= \int d^4 x_E \left( G_{\mu\nu}^a \pm \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}^a + \frac{1}{4} \epsilon_{\mu\nu\rho\sigma} \epsilon_{\mu\nu\eta\theta} G_{\rho\sigma}^a G_{\eta\theta}^a \right) \\ &= \int d^4 x_E \left[ 2(G_{\mu\nu}^a)^2 \pm 2 \left( \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}^a G_{\mu\nu}^a \right) \right] \end{aligned} \quad (9)$$

In order to saturate the extremization condition, we must solve the first-order differential equation

$$G_{\mu\nu}^a = \pm \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}^a \quad (10)$$

for  $A_\mu$  field solutions, where + denotes self-dual fields and – indicates anti-self-dual fields. The explicit profile of the instanton solution is essentially a monopole in group space, and can be written using 't Hooft symbols.<sup>[1,6]</sup> For example, an  $SU(2)$  instanton can be written in temporal gauge ( $A_0(\vec{x}) = 0$ ) as the gauge field

$$A_i(\vec{x}) = iU_1(\vec{x})\partial_i U_1^\dagger(\vec{x}) \quad (11)$$

where the  $i = 1, 2, 3$  subscripts refer to  $\hat{x}$ ,  $\hat{y}$ , and  $\hat{z}$  components of the vector field and  $U_1(\vec{x})$  is the  $SU(2)$ -valued group matrix

$$U_1(\vec{x}) = \exp \left( i\pi \vec{x} \cdot \vec{\tau} / \sqrt{|\vec{x}|^2 + \rho^2} \right) \quad (12)$$

where  $\vec{\tau} = \vec{\sigma}/2$  are the  $SU(2)$  generators and  $\rho$  is a parameter characterizing the instanton size. This prescription generalizes to higher winding numbers by replacing  $U_1$  by  $U_n \equiv U_1^n$  for the desired winding number  $n$ . An additional complication arises when considering instanton solutions for groups larger than  $SU(2)$ , but a fortunate simplification arises from the fact that all  $SU(N)$  groups share the same homotopy classification  $\Pi_3(SU(N)) = \mathbb{Z}$ , which essentially means that all instanton solutions of a given winding number in an  $SU(N)$  group are smoothly deformable into each other. Thus, it is sufficient to characterize instanton solutions in an arbitrary  $SU(N)$  group by first considering a convenient  $SU(2)$  subgroup and then augmenting the solutions by an appropriate “index of embedding” that keeps track of the dimensionality of the larger group space.<sup>[6,7]</sup>

We remark that a standard introduction to instantons motivates Equation (11) as gauge-equivalent to 0. From this perspec-

tive, we note that (still in temporal gauge) trivial  $A_i = 0$  solutions to the field equations obviously minimize the Yang-Mills action. Yet since the action is defined by an integral over (Euclidean) spacetime, we should simultaneously consider all gauge field configurations that are equivalent to the trivial  $A_i = 0$  solution, which motivates us to consider “pure gauge” field configurations for  $A_i$  as written in Equation (11). Imposing the (anti-)self-duality condition from Equation (10) then leads to the BPST (anti-)instanton solution.

### 3.2. The $\Theta$ -Vacuum in Yang-Mills

The main consequence of instanton solutions in Yang-Mills theory is its impact on the vacuum structure. We first illustrate this point by noting that the unit instanton solution satisfying Equation (8) corresponds to the flow of the Chern-Simons current<sup>[6]</sup>

$$K^\mu = 2\epsilon^{\mu\nu\alpha\beta} \left( A_\nu^a \partial_\alpha A_\beta^a + \frac{g}{3} f^{abc} A_\nu^a A_\alpha^b A_\beta^c \right) \quad (13)$$

where the charge of the  $K^\mu$  current is the usual integral of  $K^0$  over 3D space,

$$\mathcal{K} = \frac{g^2}{32\pi^2} \int d^3x K^0(x) \quad (14)$$

While the Chern-Simons current is not gauge-invariant and hence not a physical current, its divergence is gauge-invariant and actually the familiar  $G\tilde{G}$  operator,

$$\partial_\mu K^\mu = G\tilde{G} \quad (15)$$

which allows us to identify the unit instanton as carrying Chern-Simons topological charge. Consequently, the vacuum state of Yang-Mills theory is an infinitely degenerate vacuum state with “pre-vacua” labeled by Chern-Simons charge, as required from the homotopy classification.

We now calculate the vacuum energy of  $SU(2)$  Yang-Mills theory, which can be obtained using the functional integral method. We start with

$$\exp(-\mathcal{E} V \cdot T) = \int \mathcal{D}A \exp(-S_E[A]) \quad (16)$$

where  $S_E[A]$  is the Euclidean action corresponding to Equation (7) and  $\mathcal{E}$  is the vacuum energy density and  $V \cdot T$  is the volume of Euclidean spacetime. The functional integral is a sum over topological sectors with winding numbers  $w$  as

$$\begin{aligned} \exp(-\mathcal{E} V \cdot T) &= 1 \times \int [\mathcal{D}A]_{w=0} \exp \left[ - \int \frac{1}{2} \delta A \cdot (-\partial^2) \delta A \right] \\ &+ \exp(-8\pi^2/g^2) \times 2 \int [\mathcal{D}A]_{w=\pm 1} \\ &\times \exp \left[ - \int \frac{1}{2} \delta A \cdot (-\partial^2) \delta A \right] \\ &+ \dots \end{aligned} \quad (17)$$

where the ... indicate a sum over higher topological sectors that are further suppressed by higher powers of the instanton action. The remaining path integral is calculated over variations  $\delta A$  centered around the instanton background with fixed  $w$ <sup>[1-3]</sup> and essentially distills into an evaluation of the determinantal operator associated with the eigenvalues of  $-\partial^2$ , which must be regularized and renormalized. Eschewing the technical aspects of the calculation, the evaluation of the path integral modifies the  $1/g^2$  coupling in instanton action prefactor into a running coupling, where  $1/g^2(\rho)$  is to be evaluated at the appropriate instanton size.

In addition, the overall path integral suffers divergences from the integration of symmetries of the instanton solution. Namely, there are eight transformations that leave the classical solution invariant: four translations, one dilatation and three  $SU(2)$  rotations. The three  $SU(2)$  rotations are handled by an integration over Euler angles specifying the orientation of the unit instanton solution in group space. We parametrize the translations by a shift of  $x_E^j \rightarrow x_E^j + x_0^j, j = 1, 2, 3, 4$ , which serve to shift the center of the instanton solution in Equations (11) and (12) but do not change its topological index, and we have already introduced the instanton size parameter  $\rho$  in Equation (12). Simply stated, the path integral is a discrete sum over all instantons of a given topological index, and thus the zero mode redundancies of solutions with the same topological index must be parametrized as explicit integration variables. Hence, the zero modes for translation symmetry and dilatation symmetry are left as indefinite integrals over  $d^4x_0$  and  $d\rho$ , giving

$$\begin{aligned} &\exp(-\mathcal{E} V \cdot T) \\ &= \text{const.} \left( 1 + 2 \int d^4x_0 \int \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp(-8\pi^2/g^2(\rho)) + \dots \right) \end{aligned} \quad (18)$$

where  $C = 2^{10} \pi^6 e^{7.0539\dots}$  from the classic 't Hooft calculation.<sup>[1]</sup> The  $\frac{1}{\rho^5}$  comes from dimensional analysis, and the  $1/g^8$  factor comes from the Jacobian for change of variables from the integral over zero modes. Subsequent terms corresponding to  $n_+$  instantons and  $n_-$  anti-instantons, denoted by ... in Equation (18), are given by

$$\frac{1}{n_+! n_-!} \left[ \int d^4x_0 \int \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp(-8\pi^2/g^2(\rho)) \right]^{n_+ + n_-} \quad (19)$$

which is the basis for the dilute gas approximation where the instantons are well-separated. We remark that as  $\rho$  becomes large,  $g^2(\rho)$  grows and the semiclassical approximation breaks down.

We can now reshuffle the sum over instantons and anti-instantons to a summation over net topological number,  $n_+ - n_-$ , with

$\exp(-\mathcal{E} V \cdot T) = \text{const.} \times$

$$\left( 1 + 2 \sum_{\substack{n_+, n_- \\ n_+ - n_- = 1}} \frac{1}{n_+! n_-!} \left[ \int d^4 x_0 \int \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp(-8\pi^2/g^2(\rho)) \right]^{n_+ + n_-} + \dots \right) \simeq \exp \left( 2 \int d^4 x_0 \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp(-8\pi^2/g^2(\rho)) \right) \quad (20)$$

In the last line, we have used the identity

$$\exp(2x) = \sum_{n_1} \sum_{n_2} \frac{1}{n_1!} \frac{1}{n_2!} e^{n_1 x} e^{n_2 x} \quad (21)$$

From Equation (20), we note that  $V \cdot T = \int d^4 x_0$  is the volume of Euclidean spacetime, and we thus conclude that the vacuum energy density from the instanton configurations is

$$\mathcal{E} = -2 \int \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp[-8\pi^2/g^2(\rho)] \quad (22)$$

Now, we consider the contribution to the vacuum energy density when the Yang-Mills Lagrangian (again in Minkowski metric) includes a  $\Theta$  term,

$$\mathcal{L} = -\frac{1}{4} G_{\mu\nu}^a G^{\mu\nu,a} + \Theta \frac{g_s^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a, \mu\nu} \quad (23)$$

The only effect on the previous computation is a new factor of  $e^{i\Theta(n_+ - n_-)}$  in Equation (19), since the  $G\tilde{G}$  term integrates to a net topological winding number. Carrying this through, the sum over net winding number in Equation (20) now evaluates to

$$\exp(-\mathcal{E} V \cdot T) \simeq \exp \left[ (e^{i\Theta} + e^{-i\Theta}) \int d^4 x_0 \int \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp[-8\pi^2/g^2(\rho)] \right] \quad (24)$$

and we obtain the vacuum energy density expression

$$\mathcal{E} = -2 \cos \Theta \int \frac{d\rho}{\rho^5} C \frac{1}{g^8} \exp[-8\pi^2/g^2(\rho)] \quad (25)$$

Given that the integral over instanton size is evaluated with an infrared cutoff, the vacuum energy density induced by instanton configurations in the dilute instanton gas approximation has a cosine dependence on  $\Theta$ . We can now distill the PQ mechanism into simple energetics: by coupling the QCD axion to this non-perturbatively generated vacuum potential, the axion field acquires a vacuum expectation value that cancels the  $\Theta$  parameter in Equation (23), solving the strong CP problem.

#### 4. The Strong CP Problem in the Standard Model and the Axion Solution

In the Standard Model, the non-observation of an electric dipole moment (EDM) for the neutron puts severe constraints on the  $\Theta$  parameter,

$$\bar{\Theta} \equiv \Theta + \arg \det Y_u Y_d \quad (26)$$

where  $\Theta$  is the same parameter as in the Yang-Mills Lagrangian Equation (23) and  $Y_u$  and  $Y_d$  are the (unknown) original Yukawa matrices for up and down quarks in the Standard Model (SM).

The current constraint on the neutron EDM comes from the nEDM collaboration, with  $|d_n| < 1.8 \times 10^{-26} e \text{ cm}$ .<sup>[8]</sup> Following Ref. [9], this upper bound constrains the  $\bar{\Theta}$  parameter since  $d_n = C_{\text{EDM}} e \bar{\Theta}$ , where  $C_{\text{EDM}} = 2.4 \times 10^{-16}$  in the SM. As a result, the current nEDM upper bound leads to

$$\bar{\Theta} < 7.5 \times 10^{-11} \quad (27)$$

which quantifies the strong CP problem. Returning to Equation (26), we emphasize that  $\bar{\Theta}$  has two profoundly distinct origins. First, the  $\Theta$  parameter of QCD has a domain  $(-\pi, \pi]$ , and thus would generally be expected to be an  $\mathcal{O}(1)$  number. Second, the  $Y_u$  and  $Y_d$  matrices are similarly not measured in the Standard Model: we only know the eigenvalues of these matrices after rotating to the fermion mass basis, and we also know that the combination of the unitary matrices that diagonalize  $Y_u$  and  $Y_d$  gives the Cabibbo-Kobayashi-Maskawa matrix that has a Jarlskog invariant  $CP$ -violating measure of  $3.08 \times 10^{-5}$ .<sup>[10]</sup> Hence, the strong CP problem stems from the fact that two unrelated origins for  $\bar{\Theta}$  must be radically aligned and canceled to achieve an acceptable value allowed by nEDM constraints, which is exacerbated by the large domain for each underlying phase parameter. As an aside, we can mention that the massless up quark solution for the strong CP problem, albeit excluded by lattice measurements,<sup>[11]</sup> is reflected in the structure of Equation (26). Namely, if the up quark were massless, the  $\arg \det Y_u Y_d$  term becomes undefined, since the argument of zero is arbitrary, reflecting an enhanced axial symmetry in the fermion mass basis that can be used to rotate  $\Theta$ . As another aside, an alternative class of models solving the strong CP problem are Nelson-Barr models,<sup>[12,13]</sup> where  $CP$  violation is introduced via spontaneous



breaking at a high scale. A recent review of Nelson-Barr models can be found in Ref. [14].

In this primer, we focus of course on the axion solution to the strong CP problem. As alluded to in the previous section, we can use the energetics of the  $\Theta$  parameter to cause the axion field to develop a tadpole and cancel the original  $\Theta$  as a result. It is necessary and sufficient to couple the axion field to a colored fermion  $\psi$  via an axial-vector current,

$$\mathcal{L} = \left( \frac{\partial_\mu a}{f_a} \right) \bar{\psi} \gamma^\mu \gamma^5 \psi \quad (28)$$

where we have introduced the decay constant  $f_a$  on dimensional grounds. We recognize that the classical shift symmetry for the axion is preserved by this interaction term since the axion is derivatively coupled. Crucially, the derivative coupling of the axion is precisely a current coupling to the Adler-Bell-Jackiw chiral anomaly for  $\psi$ ,<sup>[15,16]</sup> and by using the equation of motion for  $\psi$ , we can realize the desired coupling of  $a$  to the dual field strength tensor of the gluons,

$$\mathcal{L} = N \frac{a}{f_a} \frac{g_s^2}{32\pi^2} G_{\mu\nu} \tilde{G}_{\mu\nu} \quad (29)$$

where  $N$  denotes the PQ anomaly with respect to the color group.

Comparing the  $\Theta$  term with Equation (29), we remark that a finite shift in the  $a$  field redefines the  $\Theta$  term, and hence  $a$  acts as a *spurion* for  $\Theta$ . It is crucial to recall that since  $a$  is a Goldstone field at the classical level, the lack of an origin in field space for  $a$  is the necessary requirement for this spurion interpretation. Using the spurion argument (see Refs. [17, 18]), we can now observe that below the scale of QCD confinement, instantons induce a periodic potential for the axion leading the axion field to acquire a tadpole offset that minimizes the overall vacuum energy and adjusts the  $\Theta$  parameter to zero:

$$\mathcal{L} = \Lambda_{\text{QCD}}^4 \cos \left( \frac{Na}{f_a} + \Theta \right) \quad (30)$$

where  $\Lambda_{\text{QCD}}$  is the scale of QCD confinement, about 200 MeV, and is formally related to the topological susceptibility of QCD. After shifting the axion field to absorb the constant  $\Theta$  vacuum angle (and setting  $N = 1$  to avoid domain wall problems), we perform a Taylor expansion of the cosine to obtain an axion mass squared of  $m_a^2 f_a^2 \approx \Lambda_{\text{QCD}}^4 \approx m_\pi^2 f_\pi^2$ .

Since the axion decay constant  $f_a$  is generally required to be very large,  $\mathcal{O}(10^9)$  GeV from supernovae constraints,<sup>[10]</sup> the axion mass is typically very small, with

$$m_a = 5.70(7) \left( \frac{10^9 \text{ GeV}}{f_a} \right) \text{ meV} \quad (31)$$

Given that there are no QCD states below this mass, the only possible Standard Model final states allowed for axion decays are photons and perhaps neutrinos. The neutrino decay is typically discarded since they may not be kinematically allowed and the rate is also suppressed by the neutrino mass divided by  $f_a$ . On the other hand, the diphoton rate is dictated by the PQ anomaly

with respect to the electromagnetic field strength,

$$\mathcal{L} = \frac{a}{f_a} \left( \frac{E}{N} - 1.92 \right) \frac{e^2}{16\pi^2} F_{\mu\nu} \tilde{F}^{\mu\nu} \quad (32)$$

where we have rescaled the  $f_a$  scale by  $1/N$  to move the  $N$  dependence purely into the electromagnetic coupling and the  $-1.92$  value arises from the axion mixing with the QCD mesons. In benchmark axion models, the EM anomaly is fixed by the color and EM charges of the fermions coupling to the axion, such as  $8/3$  for the Dine-Fischler-Srednicki-Zhitnitsky model<sup>[19,20]</sup> and  $E/N = 0$  for the Kim-Shifman-Vainshtein-Zakharov model.<sup>[21,22]</sup> We remark that the lifetime of the axion exceeds the age of the universe for  $m_a < 20$  eV with  $E/N = 0$ ,<sup>[10]</sup> and thus light axions are a prime cold dark matter candidate.

## 5. Conclusion

We have introduced the basic physics of axions as the angular mode associated with a complex scalar field charged under a Peccei-Quinn symmetry. We have seen that the Peccei-Quinn anomaly with the color gauge group leads to an instanton-induced potential that breaks the classical, continuous shift symmetry into a periodic shift symmetry, resulting in a light axion mass proportional to the QCD confinement scale and inversely proportional to the axion decay constant. The EM anomaly and the mixing with QCD mesons dictate the axion couplings to photons, leading to a rich set of experimental signals across decades in couplings and masses. Thus, since the diphoton coupling is essentially fixed given the axion mass, axions serve as a one-parameter model for dark matter. We close by remarking that ALPs relax the relationship between the mass and the axion decay constant by using a non-QCD topological susceptibility to determine their relationship, which helps to populate the entire  $(m_a, f_a)$  plane probed by axion experimental efforts.

## Acknowledgements

This work is 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 39083149).

Open access funding enabled and organized by Projekt DEAL.

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

axions, ALP phenomenology, QCD instantons

Received: March 13, 2023

Revised: July 18, 2023

Published online:

[1] G. 't Hooft, *Phys. Rev. D* **1976**, 14, 3432; [erratum: *Phys. Rev. D* 1978, 18, 2199].

- [2] C. G. Callan, R. F. Dashen, Jr, D. J. Gross, *Phys. Lett. B* **1976**, 63, 334.
- [3] C. G. Callan, R. F. Dashen, Jr, D. J. Gross, *Phys. Rev. D* **1978**, 17, 2717.
- [4] We remark that the generic nature of Goldstone bosons arising in ultraviolet extensions of the Standard Model can typically be mapped to the field theory at hand, although the key requirements to satisfy are that the ultraviolet fields exhibit Goldstones traversing a compact field space and the ultraviolet continuous symmetries are not gauged (Goldstones arising from noncompact continuous symmetries are called dilatons).
- [5] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, Y. S. Tyupkin, *Phys. Lett. B* **1975**, 59, 85.
- [6] M. Shifman, *Advanced Topics in Quantum Field Theory*, Cambridge University Press, Cambridge **2022**.
- [7] C. Sasaki, H. Murayama, *Nucl. Phys. B* **1998**, 532, 498.
- [8] C. Abel, S. Afach, N. J. Ayres, C. A. Baker, G. Ban, G. Bison, K. Bodek, V. Bondar, M. Burghoff, E. Chanel, Z. Chowdhuri, P.-J. Chiu, B. Clement, C. B. Crawford, M. Daum, S. Emmenegger, L. Ferraris-Bouchez, M. Fertl, P. Flaux, B. Franke, A. Fratangelo, P. Geltenbort, K. Green, W. C. Griffith, M. van der Grinten, Z. D. Grujic, P. G. Harris, L. Hayen, W. Heil, R. Henneck, et al., *Phys. Rev. Lett.* **2020**, 124, 081803.
- [9] M. Pospelov, A. Ritz, *Nucl. Phys. B* **2000**, 573, 177.
- [10] Particle Data Group, R. L. Workman, V. D. Burkert, V. Crede, E. Klempt, U. Thoma, L. Tiator, K. Agashe, G. Aielli, B. C. Allanach, C. Amsler, M. Antonelli, E. C. Aschenauer, D. M. Asner, H. Baer, Sw. Banerjee, R. M. Barnett, L. Baudis, C. W. Bauer, J. J. Beatty, V. I. Belousov, J. Beringer, A. Bettini, O. Biebel, K. M. Black, E. Blucher, R. Bonventre, V. V. Bryzgalov, O. Buchmuller, M. A. Bychkov, et al., *PTEP* **2022**, 2022, 083C01.
- [11] Z. Fodor, C. Hoelbling, S. Krieg, L. Lellouch, T. Lippert, A. Portelli, A. Sastre, K. K. Szabo, L. Varnhorst, *Phys. Rev. Lett.* **2016**, 117, 082001.
- [12] A. E. Nelson, *Phys. Lett. B* **1984**, 136, 387.
- [13] S. M. Barr, *Phys. Rev. Lett.* **1984**, 53, 329.
- [14] M. Dine, P. Draper, *JHEP* **2015**, 08, 132.
- [15] S. L. Adler, *Phys. Rev.* **1969**, 177, 2426.
- [16] J. S. Bell, R. Jackiw, *Nuovo Cim. A* **1969**, 60, 47.
- [17] J. E. Kim, G. Carosi, *Rev. Mod. Phys.* **2010**, 82, 557; [erratum: *Rev. Mod. Phys.* 2019, 91, 049902].
- [18] A. Kivel, J. Laux, F. Yu, *JHEP* **2022**, 11, 088.
- [19] M. Dine, W. Fischler, M. Srednicki, *Phys. Lett. B* **1981**, 104, 199.
- [20] A. R. Zhitnitsky, *Sov. J. Nucl. Phys.* **1980**, 31, 260.
- [21] J. E. Kim, *Phys. Rev. Lett.* **1979**, 43, 103.
- [22] M. A. Shifman, A. I. Vainshtein, V. I. Zakharov, *Nucl. Phys. B* **1980**, 166, 493.