School of Physical and Chemical Sciences, University of Canterbury,
Christchurch, New Zealand

An Introduction to the Axion Solution to the Strong CP Problem

Joseph A. Wilson

Supervisor Assoc. Prof. Jenni Adams

«Compiled on 06/11/20, 14:06»

UNIVERSITY OF CANTERBURY

Literature Review

Abstract

Literature review on something relating to axions.

Contents

0 Introduction

Overall

1 QCD and the Strong CP Problem

The Standard Model of particle physics is a quantum field theory which describes all known fundamental particles and interactions to very high precision—with notable exceptions including: the absence of gravity; dark matter; and the experimentally incorrect prediction that neutrinos are massless. Alongside these shortcomings, the standard model also exhibits issues of a more philosophical nature, such as the reasons for the values of the theory's many free parameters. (For example; why are there three generations of particles? Why do the particle masses and coupling constants have the values they do?) Among these mysteries is the *strong CP problem*, a fine-tuning problem regarding the apparent symmetry of the strong force under time reversal.[1] This chapter summarises the theory necessary to understand the statement and origin of the strong CP problem.

The standard model was born amid the explosion of phenomenological particle physics in the early 1960s, in an effort to explain the patterns in the rapidly growing catalogue of known particles. The introduction of the quark model efficiently explained the properties of the numerous hadrons in terms of six kinds of quarks possessing *colour symmetry*, and the theory was given the name *quantum chromodynamics* (QCD). This, together with the well-established quantum theory of electromagnetism, *quantum electrodynamics* (QED), became the first approximation to the standard model [1]. The modern statement of the standard model is that it is a Lorentz-invariant quantum field theory whose *gauge group* is the compact connected Lie group

$$\big(\mathrm{SU}(3) \oplus \mathrm{SU}(2) \oplus \mathrm{U}(1)\big)\big/\mathbb{Z}_6, \tag{1.1}$$

equipped with a particular action on matter fields. A quantum field theory is a quantised *gauge field theory*, which is a field theory whose Lagrangian is symmetric under the action of the gauge group (explained more in § 1.1). Each term in the gauge group of the standard model (1.1) corresponds to a fundamental force of nature: The factor $\mathrm{SU}(3)$ is the gauge group of quantum chromodynamics, the sector of the standard model describing the interactions of colour-charged quarks and gluons via the strong force. The factor $\mathrm{SU}(2) \oplus \mathrm{U}(1)$ corresponds to the unified electroweak force, and contains another $\mathrm{U}(1)$ sub-factor corresponding to the electromagnetic force of QED.

Quantum chromodynamics was formulated quite some time after quantum electrodynamics was understood, which is a reflection that QCD is more intricate than QED. Unlike QED, quantum chromodynamics exhibits *asymptotic freedom*, meaning that the effective strength of the strong interaction between colour-changed particles *increases* with increasing separation.

[1]Time reversal T and charge–parity CP symmetry are equivalent assuming the combined charge–parity–time symmetry; i.e., $T = CP \mod CPT$.

Thus, QCD is severely non-perturbative and evades even modern analytical treatment, except at low energies. Asymptotic freedom also results in quark confinement, which contributed to the confusion of particle physicists in the 1960s—quarks could never be observed in isolation. The dissimilarity of QCD to QED can be largely credited to the fact that the gauge group $\mathrm{SU}(3)$ is non-Abelian, implying that the gluon force carriers of QCD are *themselves* colour-charged, leading to asymptotic freedom [2]. This is in contrast to the Abelian $\mathrm{U}(1)$ theory of QED, in which the force carriers are uncharged photons. Despite the technical challenges which come with its non-perturbative nature, quantum chromodynamics remains a successfully predictive and elegant theory.

However, QCD comes with a challenge of theoretical concern: the *strong CP problem*, or "why does QCD fail to forbid CP violation?" Since we do not observe CP symmetry violation in QCD [3], we expect the theory of QCD prohibit it—but a careful inspection of QCD reveals that it does not. This problem proves to be of the fine-tuning variety, but it lacks even an anthropic solution: it does not change the universe in dramatic ways whether CP symmetry is broken in the QCD sector or not (as we so observe). Instead, the disparity indicates a deeper theoretical shortcoming, and provides good reason to pursue physics beyond the standard model.

This chapter fills in the background relevant to the formal statement of the strong CP problem, and assumes surface-level familiarity with differential geometry, Lagrangian mechanics, quantum mechanics and the notation of exterior calculus. The significance and role of the gauge group is outlined in the next section, where a geometrical overview of gauge field theory is presented. This overview aims to provide a basic understanding of the main mathematical objects—or 'moving parts'—of a gauge field theory, and of their physical interpretations. The chapter later introduces further concepts specific to QCD and the strong CP problem as they become relevant. Finally, QCD is defined and the origin of the so-called *strong CP problem* is highlighted.

1.1 Overview of Classical and Quantum Gauge Theory

Gauge theories take place on a spacetime manifold and consist of two mathematically distinct dynamical entities: a *matter field* and a *gauge field*. A classical gauge theory is completely specified by the prescription of four ingredients: the base manifold $\mathcal{M}$; the vector space V in which the matter field ψ takes its values; a *gauge group* G equipped with an action on the matter field; and equations of motion, usually supplied via a Lagrangian density $\underset{\sim}{\mathcal{L}}$. The dynamical gauge field is not prescribed at the outset—it arises naturally as a consequence of the matter field's symmetry under the action of the gauge group.

The underlying premise of gauge theory is that there may be physical redundancy in the mathematical description of a matter field at each point in spacetime. (For example, the complex phase of a total wavefunction is physically irrelevant.) This redundancy results in non-physical degrees of freedoms of the matter field at every point in spacetime: these are called *local gauge freedoms*. (Keeping with our example: a local rotation of phase is a gauge freedom.) Crucially, local gauge freedoms introduce ambiguity in the notion of the physical rate of change of a matter field about a point. This is because the point-wise independence of the matter field's gauge freedom means that differences in the field's value between nearby points is gauge-dependent. In other words, there is no preferred directional derivative of a matter field if a local gauge

Notations used in this chapter

$\mathsf{A}, \mathsf{B}, \dots$	matrices
$\mathcal{A}, \mathcal{B}, \dots$	objects with manifold structure
$\rightarrowtail$; $\twoheadrightarrow$; $\leftrightarrow$	injection; surjection; bijection
$F \rightarrowtail \mathcal{V} \overset{\pi}{\twoheadrightarrow} \mathcal{M}$	fibre bundle $\mathcal{V}$ over base space $\mathcal{M}$ with fibre F and projection $\pi : \mathcal{V} \twoheadrightarrow \mathcal{M}$
$\mathrm{T}_p\mathcal{M}$; $\mathrm{T}\mathcal{M}$	tangent space at $p \in \mathcal{M}$; tangent bundle of $\mathcal{M}$
$\mathrm{T}^r_s\mathcal{M}$	type $\binom{r}{s}$ tensors over $\mathcal{M}$, equal to $\left(\bigotimes_{i=1}^r \mathrm{T}\mathcal{M}\right) \otimes \left(\bigotimes_{i=1}^s \mathrm{T}^*\mathcal{M}\right)$
$\mathbb{T}\mathcal{M}$	tensor bundle of $\mathcal{M}$, equal to $\bigoplus_{r,s=0}^{\infty} \mathrm{T}^r_s\mathcal{M}$
$\underset{\sim}{A}$	spacetime tensor field; i.e., a section of $\mathbb{T}\mathcal{M}$
$\boldsymbol{A}$	vector field of some abstract vector space not contained in $\mathbb{T}\mathcal{M}$
$\underset{\sim}{\boldsymbol{A}}$	spacetime tensor field with values in some other abstract vector space
$\wedge^p V$	pth exterior power of the vector space V; i.e., the space of V-valued p-forms
$\Gamma(\mathcal{V})$	smooth sections $\mathcal{M} \to F$ of fibre bundle $F \rightarrowtail \mathcal{V} \twoheadrightarrow \mathcal{M}$
$\Omega^p(\mathcal{M})$	space of p-forms on $\mathcal{M}$, equal to $\Gamma(\wedge^p \mathrm{T}^*\mathcal{M})$
$\Omega^p(\mathcal{M}, V)$	space of V-valued p-forms, equal to $\Gamma(\mathcal{V} \otimes \wedge^p \mathrm{T}^*\mathcal{M})$ where $V \rightarrowtail \mathcal{V} \twoheadrightarrow \mathcal{M}$

freedom is present—until the choice of a *connection* is made. The triumph of gauge theory is that, by introducing the *gauge field* to act as a connection, this ambiguity is recast as a separate set of physical degrees of freedom. The implications of this are twofold: Firstly, the gauge field isolates a choice of derivative (namely, the covariant derivative with respect to the gauge field), allowing the inclusion of well-defined derivatives of the matter field in the theory's equations of motion. Secondly, the gauge field has a dynamical role in the theory, and it describes a new kind of field: force fields (and, after quantisation, force carrier bosons).

In geometrical language, the matter field $\boldsymbol{\psi} \in \Gamma(\mathcal{V})$ is a section of a vector bundle $V \rightarrowtail \mathcal{V} \twoheadrightarrow \mathcal{M}$. In other words, at any point $p \in \mathcal{M}$ in spacetime, the field continuously assigns a vector, $\boldsymbol{\psi}|_p \cong V$.[2] The gauge transformations of $\boldsymbol{\psi}|_p$ at a point p form a group G under composition—but this group does not describe local gauge transformations which act on the *entire* field $\boldsymbol{\psi}$. Rather, the total space of local gauge transformations is a principle G-bundle $G \rightarrowtail \mathcal{G} \twoheadrightarrow \mathcal{M}$, consisting of smooth maps $\mathcal{M} \to G$.[3] The action of $\mathcal{G}$ on the space of matter fields $\mathcal{V}$ may be denoted

$$\boldsymbol{\psi} \overset{g}{\mapsto} \boldsymbol{\psi}' = g \cdot \boldsymbol{\psi}$$

for $\boldsymbol{\psi} \in \Gamma(\mathcal{V})$ and $g \in \Gamma(\mathcal{G})$ (i.e., $\boldsymbol{\psi} : \mathcal{M} \to V$ and $g : \mathcal{M} \to G$). This specification of an action "$\cdot$" of $\mathcal{G}$ on the vector bundle $\mathcal{V}$ is equivalent to the choice of a linear representation $\rho : G \to \mathrm{End}(V)$ of G on V, applied globally on $\mathcal{M}$. Thus, we may write the action as a matrix product, $g \cdot \boldsymbol{\psi} = \rho(g)\boldsymbol{\psi} \equiv g_\rho \boldsymbol{\psi}$, remembering that both $\boldsymbol{\psi}$ and g (but not ρ) vary across $\mathcal{M}$.

It is worth emphasising that all gauge theories are isomorphic to a gauge theory possessing only one matter field $\boldsymbol{\psi}_{\text{total}}$ with one gauge group (G, ρ). For instance, if a theory consists of different fundamental particles and forces, represented by separate fields $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$ with gauge

[2] The reason many objects in gauge theory (such as fields $\boldsymbol{\psi}$) are defined as sections of fibre bundles (and not simply as smooth maps $\boldsymbol{\psi} : \mathcal{M} \to V$) is because fibre bundles are themselves smooth manifolds which admit the construction of connections (whereas the space of smooth functions lacks a manifold topology).

[3] We sometimes loosely refer to the fibre group G as the gauge group, but strictly we mean the fibre bundle $\mathcal{G}$ (following [4]).

groups $(G_\varphi, \rho_\varphi)$ and (G_ϕ, ρ_ϕ), then we may take the total matter field $\boldsymbol{\psi}_{\text{total}} = \boldsymbol{\varphi} \oplus \boldsymbol{\phi}$ to be the direct sum of the fields in the theory, and similarly equip it with the gauge group $G_{\text{total}} = G_\varphi \oplus G_\phi$ with representation $\rho_{\text{total}} = \rho_\varphi \oplus \rho_\phi$. This new theory, inheriting the original equations of motion, is physically identical to the original. In full generality, therefore, we treat our theory as possessing one matter and one gauge field, while freely speaking of its composite parts as separate fields where convenient.

A *connection* on $\mathcal{V}$ is a derivation[4] $\underset{\sim}{\nabla} : \Gamma(\mathcal{V}) \to \Omega^1(\mathcal{M}, \mathcal{V})$ from vector fields to vector-valued 1-forms. The interpretation is that $(\underset{\sim}{\nabla}\boldsymbol{\psi})(\underset{\sim}{X})$—that is, the action of the V-valued 1-form $\underset{\sim}{\nabla}\boldsymbol{\psi}$ on a vector $\underset{\sim}{X} \in \mathrm{T}\mathcal{M}$—gives the directional derivative of $\boldsymbol{\psi}$ along $\underset{\sim}{X}$ (with respect to the connection $\underset{\sim}{\nabla}$). Employing a basis $\{\boldsymbol{e}_a\}$ of $\mathcal{V}$ and local coordinates $\{x^\mu\}$ of $\mathcal{M}$, *any* connection is of the form

$$\begin{aligned}\underset{\sim}{\nabla}\boldsymbol{\psi} &= \underset{\sim}{\mathrm{d}}\boldsymbol{\psi} + \underset{\sim}{\boldsymbol{A}}\boldsymbol{\psi} \\ &= \left(\partial_\mu \psi^a{}_b + A_\mu{}^a{}_b \psi^b\right) \boldsymbol{e}_a \otimes \underset{\sim}{\mathrm{d}}x^\mu\end{aligned}$$

for some matrix-valued 1-form $\underset{\sim}{\boldsymbol{A}}$, where $\underset{\sim}{\mathrm{d}}$ is the exterior derivative. (More precisely, $\underset{\sim}{\boldsymbol{A}} \in \Omega^1(\mathcal{M}, \mathfrak{g})$ is a $\mathfrak{g}$-valued 1-form, where $\mathfrak{g}$ is the Lie algebra of the gauge group G.) If $\underset{\sim}{\nabla}\boldsymbol{\psi}$ is required to transform like the matter field $\boldsymbol{\psi}$ under local gauge transformations, that is, as

$$\underset{\sim}{\nabla}\boldsymbol{\psi} \overset{g}{\mapsto} g \cdot (\underset{\sim}{\nabla}\boldsymbol{\psi}) = (g \cdot \underset{\sim}{\nabla})(g_\rho \boldsymbol{\psi}) \overset{!}{=} g_\rho \underset{\sim}{\nabla}\boldsymbol{\psi},$$

then $\underset{\sim}{\nabla}$ is called a *covariant derivative* and the connection 1-form $\underset{\sim}{\boldsymbol{A}}$ consequently obeys

$$\underset{\sim}{\boldsymbol{A}} \overset{g}{\mapsto} g \cdot \underset{\sim}{\boldsymbol{A}} = g_\rho \underset{\sim}{\boldsymbol{A}} g_\rho^{-1} - \left(\underset{\sim}{\mathrm{d}} g_\rho\right) g_\rho^{-1}. \tag{1.2}$$

Such a connection $\underset{\sim}{\nabla}_A$ is not unique; it depends on the choice of the 1-form field $\underset{\sim}{\boldsymbol{A}}$. It is exactly this connection 1-form which is promoted to a dynamical object in a gauge theory and given the name *the gauge field.* At each point in spacetime, and the gauge field $\underset{\sim}{\boldsymbol{A}}$ linearly assigns to each direction in spacetime an infinitesimal transformation of the matter field (this is why $\underset{\sim}{\boldsymbol{A}}$ is Lie algebra valued).

In order for the gauge field to be incorporated in the theory's equations of motion, we would like to have a notion of its derivative. However, the covariant derivative $\underset{\sim}{\nabla} : \Gamma(\mathcal{V}) \cong \Omega^0(\mathcal{M}, \mathcal{V}) \to \Omega^1(\mathcal{M}, \mathcal{V})$ is not readily defined on 1-forms such as $\underset{\sim}{\boldsymbol{A}}$ until we canonically extend it to a *covariant exterior derivative*[5] $\underset{\sim}{\mathrm{d}}_\nabla : \Omega^p(\mathcal{M}, \mathcal{V}) \to \Omega^{p+1}(\mathcal{M}, \mathcal{V})$, given by

$$\underset{\sim}{\mathrm{d}}_\nabla \underset{\sim}{\boldsymbol{\varphi}} := \underset{\sim}{\mathrm{d}}\underset{\sim}{\boldsymbol{\varphi}} + \underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\boldsymbol{\varphi}}.$$

This enables the construction of, among other things, the curvature 2-form or *gauge field strength*

$$\begin{aligned}\underset{\sim}{\boldsymbol{F}} &:= \underset{\sim}{\mathrm{d}}_\nabla \underset{\sim}{\boldsymbol{A}} \\ &= \underset{\sim}{\mathrm{d}}\underset{\sim}{\boldsymbol{A}} + \underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\boldsymbol{A}} \equiv \left(\underset{\sim}{\mathrm{d}} A^a{}_b + \underset{\sim}{A}{}^a{}_c \wedge \underset{\sim}{A}{}^c{}_b\right) \boldsymbol{e}_a \otimes \boldsymbol{e}^b,\end{aligned}$$

[4] More precisely, $\underset{\sim}{\nabla}$ is a $\mathcal{C}^\infty(\mathcal{M})$-linear derivation, meaning $\underset{\sim}{\nabla}(f\boldsymbol{u} + g\boldsymbol{v}) = f\underset{\sim}{\nabla}\boldsymbol{u} + g\underset{\sim}{\nabla}\boldsymbol{v}$ for scalar fields $f, g \in \mathcal{C}^\infty(\mathcal{M})$ and $\underset{\sim}{\nabla}(\boldsymbol{u} \otimes \boldsymbol{v}) = \underset{\sim}{\nabla}(\boldsymbol{u}) \otimes \boldsymbol{v} + \boldsymbol{u} \otimes \underset{\sim}{\nabla}(\boldsymbol{v})$.

[5] The extension is uniquely defined by requiring the graded Leibniz property $\underset{\sim}{\mathrm{d}}_\nabla(\underset{\sim}{\varphi} \otimes \boldsymbol{\psi}) = \underset{\sim}{\mathrm{d}}\underset{\sim}{\varphi} \otimes \boldsymbol{\psi} + (-1)^p \underset{\sim}{\varphi} \wedge \underset{\sim}{\nabla}\boldsymbol{\psi}$ for $\underset{\sim}{\varphi} \in \Omega^p(\mathcal{M})$ and $\boldsymbol{\psi} \in \Gamma(\mathcal{V})$, analogous to the usual exterior derivative.

where $\{\boldsymbol{e}_a\}$ and $\{\boldsymbol{e}^a\}$ form a basis and dual basis of $\mathcal{V}$. The field strength is also equivalent to

$$\underset{\sim}{\boldsymbol{F}} = \underset{\sim}{\nabla} \wedge \underset{\sim}{\nabla} \equiv [\nabla_\mu, \nabla_\nu]\, \underset{\sim}{\mathrm{d}}x^\mu \wedge \underset{\sim}{\mathrm{d}}x^\nu.$$

The field strength $\underset{\sim}{\boldsymbol{F}}$ is useful because it is tensorial, in the sense that it transforms like the matter field; $\underset{\sim}{\boldsymbol{F}} \mapsto g \cdot \underset{\sim}{\boldsymbol{F}} = g_\rho \underset{\sim}{\boldsymbol{F}} g_\rho^{-1}$ under gauge transformations, even though $\underset{\sim}{\boldsymbol{A}}$ does not. The gauge field strength automatically satisfies the *Bianchi identity*, $\underset{\sim}{\mathrm{d}}_\nabla \underset{\sim}{\boldsymbol{F}} = 0$.

We have seen how the gauge field $\underset{\sim}{\boldsymbol{A}}$ and its strength $\underset{\sim}{\boldsymbol{F}}$ arise when we require the notion of a spacetime derivative for a matter field which possesses local gauge freedom, and are now acquainted with the dynamical objects of a gauge theory. The matter field $\boldsymbol{\psi}$, its derivative $\underset{\sim}{\nabla}_A \boldsymbol{\psi}$ and the strength of the gauge field $\underset{\sim}{\boldsymbol{F}}$ are constructions which all transform regularly under gauge transformations. All that remains to be specified in our theory are the equations of motion, which are to be expressed in terms of these three geometrical objects in a gauge invariant manner.

1.1.1 Lagrangians in Field Theories

For classical gauge theories, the equations of motion may be specified as the extremisers of an action

$$S[\boldsymbol{\psi}, \underset{\sim}{\nabla}_A \boldsymbol{\psi}, \underset{\sim}{\boldsymbol{F}}] = \int_{\mathcal{M}} \underset{\sim}{\mathcal{L}}[\boldsymbol{\psi}, \underset{\sim}{\nabla}_A \boldsymbol{\psi}, \underset{\sim}{\boldsymbol{F}}],$$

where $\underset{\sim}{\mathcal{L}}$ is a local Lagrangian density. In order that the equations of motion are physically well-defined, we require the Lagrangian to possess the relevant symmetries: gauge symmetry, so that the equations of motion are gauge invariant; and Lorentz symmetry (which is automatic if $\underset{\sim}{\mathcal{L}} = \mathcal{L}\, \underset{\sim}{\mathrm{vol}}$ is expressed as a volume form) [5, § 7.1]. The equations of motion are invariant under adjustments to the Lagrangian density by a total derivative $\underset{\sim}{\mathrm{d}}\underset{\sim}{K}$, since by Stokes' theorem these contribute only to terms on the boundary, where the fields are fixed by assumption. Therefore, a Lagrangian which possesses gauge symmetry is still generally permitted to transform as

$$\underset{\sim}{\mathcal{L}} \mapsto \underset{\sim}{\mathcal{L}} + \underset{\sim}{\mathrm{d}}\underset{\sim}{K} \tag{1.3}$$

under gauge transformations. If a Lagrangian transforms as (1.3) under a continuous gauge symmetry parametrised by n parameters α_i, then Noether's theorem implies the existence of n conserved current densities (viz. 3-forms[6]), one for each α_i,

$$\underset{\sim}{J}_{(i)} = \frac{\partial \underset{\sim}{\mathcal{L}}}{\partial \underset{\sim}{\nabla} \boldsymbol{\psi}} \frac{\partial \boldsymbol{\psi}}{\partial \alpha_i}\bigg|_{\mathrm{id}} - \frac{\partial \underset{\sim}{K}}{\partial \alpha_i}, \tag{1.4}$$

whose continuity equations read $\underset{\sim}{\mathrm{d}}\underset{\sim}{J}_{(i)} = 0$.

On the other hand, the Lagrangian of a *quantum* field theory enjoys an enlarged criterion of gauge symmetry: it is also permitted to transform as

$$\int_{\mathcal{M}} \underset{\sim}{\mathcal{L}} \mapsto \int_{\mathcal{M}} \underset{\sim}{\mathcal{L}} + n \cdot 2\pi\hbar, \tag{1.5}$$

[6] A *current density* naturally transforms as a 3-form. The partial derivative $\partial \underset{\sim}{\mathcal{L}} \big/ \partial \underset{\sim}{\nabla} \boldsymbol{\psi}$ of a volume form $\underset{\sim}{\mathcal{L}}$ with respect to a 1-form $\underset{\sim}{\nabla} \boldsymbol{\psi}$ is itself a 3-form, and is defined by canonically extending the scalar partial derivative to be an anti-derivation on differential forms. For details, see [6].

where $n \in \mathbb{Z}$ may vary discretely under different gauges. This is because of the origin of a QFT's equations of motion in the Feynman path integral. Explicitly, the quantum mechanical amplitude that the fields $\underset{\sim}{\psi}$ and $\underset{\sim}{A}$ satisfy prescribed boundary conditions on $\partial\Omega$ surrounding some region of spacetime $\Omega \subseteq \mathcal{M}$ is given by the path integral

$$\mathscr{A} = \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\psi}, \underset{\sim}{A}] \exp\left\{\frac{i}{\hbar} S[\underset{\sim}{\psi}, \underset{\sim}{\nabla}_A \underset{\sim}{\psi}, \underset{\sim}{A}]\right\}, \tag{1.6}$$

where the intended meaning of $\mathcal{D}[\cdots]$ is an integration over all field configurations on Ω. If the Lagrangian were to undergo a discrete gauge transformation (1.5), the amplitude $\mathscr{A} \mapsto \mathscr{A}\exp(n \cdot 2\pi i)$ would be left invariant. In other words, physical consistency of a QFT does not require the single-valuedness of the action S, but only of $\exp(iS/\hbar)$. This leads to an enlargement of the space of possible Lagrangian densities to include, in particular, *topological terms*. These prove to be especially relevant to QFTs and to the strong CP problem itself.

1.1.2 The Yang–Mills Lagrangian and the Topological θ-Term

An important component of a gauge theory's equations of motion are the terms in the Lagrangian which describe the dynamics of the gauge field $\underset{\sim}{A}$. These specify how the gauge field behaves in the vacuum (e.g., describing the classical electromagnetic field, or the quantum theory of photons, in the absence of matter). Thus, we are interested in the possible *consistent* Lagrangians which may be constructed from the gauge field $\underset{\sim}{A}$ without the matter field $\underset{\sim}{\psi}$.

Along with requiring Lorentz and gauge invariance, consistency also requires that the quantum field theory associated to a Lagrangian be *renormaliseble*. Loosely speaking, a classical field theory is renormalisable if it can be "quantised without introducing irrecoverable infinities". (This restricts the form of the Lagrangian considerably, but how this happens is beyond this review's scope.) Under these constraints, the only admissible QFT Lagrangians which may be constructed from the gauge field $\underset{\sim}{A}$ alone are linear combinations of

$$\langle \underset{\sim}{F} \wedge \star \underset{\sim}{F} \rangle \equiv \frac{1}{2} \left\langle \boldsymbol{F}_{\mu\nu}, \boldsymbol{F}_{\rho\sigma} \right\rangle_{\mathrm{Ad}} g^{\mu\rho} g^{\nu\sigma} \underset{\sim}{\mathrm{vol}} \quad \text{and} \quad \langle \underset{\sim}{F} \wedge \underset{\sim}{F} \rangle \equiv \frac{1}{4} \left\langle \boldsymbol{F}_{\mu\nu}, \boldsymbol{F}_{\rho\sigma} \right\rangle_{\mathrm{Ad}} \epsilon^{\mu\nu\rho\sigma} \underset{\sim}{\mathrm{vol}},$$

where $\star$ is the Hodge dual [5, § 7.1.2]. The inner product $\langle\ ,\ \rangle_{\mathrm{Ad}}$ on the Lie algebra $\mathfrak{g}$ of the gauge group G is chosen such that it is gauge-invariant.[7] Such an inner product $\langle\ ,\ \rangle_{\mathrm{Ad}}$ on $\mathfrak{g}$ is not unique; it depends on a choice of *coupling constants*. In particular, if the gauge group G is the direct sum of n simple Lie groups,[8] then $\langle\ ,\ \rangle_{\mathrm{Ad}}$ is specified by the choice of exactly n coupling constants, one corresponding to each factor of G [5, § 2.5]. Physically, the coupling constants determine the relative interaction strengths of the forces associated to each factor of G, and must enter the theory as free parameters determined experimentally. (For instance, the gauge group of the standard model (1.1) has three such coupling constants for the strong $\mathrm{SU}(3)$, weak $\mathrm{SU}(2)$, and electromagnetic $\mathrm{U}(1)$ interactions.)

The term $\langle \underset{\sim}{F} \wedge \star \underset{\sim}{F} \rangle$ is known as the *Yang–Mills Lagrangian*, and is a major component in the standard model, describing boson force carrier propagation and self-interaction (such as

[7] Recall that $\underset{\sim}{A}$, and hence $\underset{\sim}{F}$, are $\mathfrak{g}$-valued forms, so an inner product on $\mathfrak{g}$ is needed to produce a scalar. For this scalar to be gauge invariant, the inner product must additionally be "Ad-invariant" [5, § 7.3].

[8] Any compact connected Lie group is either of this form, or is a finite quotient of such a group, if $\mathrm{U}(1)$ is counted as "simple". [5, § 2.4.3].

gluon self-interactions). The Yang–Mills Lagrangian yields the equation of motion $\underset{\sim}{\mathrm{d}}_\nabla \star \underset{\sim}{\boldsymbol{F}} = 0$. For the Abelian gauge group $G = \mathrm{U}(1)$ of electromagnetism, this equation, together with the Bianchi identity $\underset{\sim}{\mathrm{d}}_\nabla \underset{\sim}{F} = 0$, are the source-free Maxwell equations. (Since the Lie algebra of $\mathrm{U}(1)$ is $\mathbb{R}$, the 2-form $\underset{\sim}{F}$ of QED is a scalar-valued.) Expressing $\underset{\sim}{F} = \underset{\sim}{\mathrm{d}}ct \wedge \underset{\sim}{E} + \star(\underset{\sim}{\mathrm{d}}ct \wedge \underset{\sim}{B})$ in terms of the non-relativistic electric and magnetic field 1-forms by choosing a spacetime split, the Yang–Mills term is the familiar electromagnetic energy density $\underset{\sim}{F} \wedge \star \underset{\sim}{F} = \underset{\sim}{B}^2 + \underset{\sim}{E}^2/c^2$.

The other term $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ is known as the *Chern–Simons term* or the *topological θ-term*, for reasons which will become apparent after a survey of its properties.

- The Chern–Simons term is odd under both time-reversal symmetry T and parity P (notice $\epsilon^{\mu\nu\rho\sigma} \mapsto -\epsilon^{\mu\nu\rho\sigma}$ under T or P) but not under charge conjugation C. This may also be seen by introducing a spacetime split, whereby $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \mathrm{tr}(\boldsymbol{E} \cdot \boldsymbol{B})\, \underset{\sim}{\mathrm{vol}}$, since $\underset{\sim}{\boldsymbol{E}}$ is a vector of odd-parity and $\underset{\sim}{\boldsymbol{B}}$ is a pseudovector of even-parity. Hence, it may give rise to CP-violating dynamics.

- It is *topological* because it does not depend on the geometry of spacetime via the metric $g^{\mu\nu}$ (instead, all spacetime indices are contracted with $\epsilon^{\mu\nu\rho\sigma}$). Hence, an action $\int_{\mathcal{M}} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ depends only on the integrand's topology over $\mathcal{M}$. (Recall that $\underset{\sim}{\boldsymbol{F}}$ is $\mathfrak{g}$-valued, so for sufficiently interesting gauge groups, the space of gauge fields may have non-trivial topology.)

- Furthermore, $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ is a total derivative of the *Chern–Simons 3-form* $\underset{\sim}{\omega}_3$,

$$\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \underset{\sim}{\mathrm{d}}\underset{\sim}{\omega}_3 = \underset{\sim}{\mathrm{d}}\,\mathrm{tr}\left(\underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\mathrm{d}}\underset{\sim}{\boldsymbol{A}} + \frac{2}{3} \underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\boldsymbol{A}} \right),$$

  meaning that the action $\int_{\mathcal{M}} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ depends only on the topology of $\underset{\sim}{\boldsymbol{A}}$ on the spacetime boundary $\partial\mathcal{M}$. As such, it does not affect the classical equations of motion. However, it has important implications in the quantum theory.

- The integral is a discrete topological invariant

$$n = \frac{1}{8\pi^2} \int_{\mathcal{M}} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \frac{1}{8\pi^2} \int_{\partial\mathcal{M}} \underset{\sim}{\omega}_3 \in \mathbb{Z}, \tag{1.7}$$

  known as the *Pontryagin number*, the *second Chern class* [7, § 1] or simply the 'winding number' [4, § 2.2] of the gauge field configuration $\underset{\sim}{\boldsymbol{A}}$. Importantly, this means that the Chern–Simons term is *not* totally gauge invariant if there are topologically distinct gauge fields $\underset{\sim}{\boldsymbol{A}}$ with varying winding number.

The choice of the symbol θ in the name "θ-term" reflects the angular nature of any coefficient θ attached to the Chern–Simons term, as in

$$\underset{\sim}{\mathcal{L}}_\theta[\underset{\sim}{\boldsymbol{A}}] = \frac{\theta}{8\pi^2} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle\, \hbar. \tag{1.8}$$

The action of this Lagrangian is $\theta n \hbar$ whenever $\underset{\sim}{\boldsymbol{A}}$ has winding number n. Since this enters the path integral as $e^{i\theta n} = e^{i(\theta + 2\pi)n}$, the coefficient θ, henceforth the *θ-parameter*, is only distinguishable modulo 2π and is hence an angular quantity. As a whole, (1.8) is referred to as the θ-term.

Glossary of technical terms

- *degenerate* — quantum eigenstates which share the same eigenvalues (usually energy) but which are physically distinguishable
- *axial, chiral* — a transformation acting differently on left- and right-handed fermions.
- *anomalous symmetry* — a symmetry of the classical Lagrangian, but not of the measure $\mathcal{D}[\boldsymbol{\psi}, \underset{\sim}{\boldsymbol{F}}]$ in the path integral (1.6), and hence not a symmetry of the associated quantum theory.
- *spontaneously broken symmetry* — a symmetry of the Lagrangian which fails to manifest in the ground state solutions.
- *Nambu–Goldstone boson* — a scalar boson which arises due to a spontaneously broken symmetry. The dimension of the broken symmetry group is the number of resulting Nambu–Goldstone bosons.

The standard model employs the Yang–Mills Lagrangian $\langle \underset{\sim}{\boldsymbol{F}} \wedge \star \underset{\sim}{\boldsymbol{F}} \rangle$, but does not find use for the other possible θ-term. Historically, the θ-term was dismissed as unphysical, since at first sight it appears to be a gauge-dependent boundary term. It was only with the discovery of *instantons* and *the non-trivial vacuum* of QCD in the mid 1970s [8] that it was realised that the θ-term should be considered, and does not necessarily vanish [9].

1.1.3 The θ-Term is a Consequence of the Non-trivial Vacuum

The θ-term is more than just a mathematical possibility which lacks physical reason for its inclusion in the Lagrangian. It is in fact central to non-Abelian gauge theories as a direct consequence of their *non-trivial vacuum structure*, which is responsible for interesting non-perturbative dynamical effects. The non-trivial vacuum is not an obvious feature of non-Abelian theories, as it is absent in Abelian theories such as QED, out of which QCD emerged.

For an Abelian gauge group $G = \mathrm{U}(1)$ with total gauge group bundle $\mathcal{G} = \mathcal{U}(1)$, the space of asymptotically-identity gauge transformations $\Gamma^{\mathrm{id}}(\mathcal{U}(1))$ is continuously connected to the identity.[9] This means that any two gauge-equivalent gauge field configurations can be *continuously* gauge transformed into each other. Hence, all gauge transformations preserve the topology of the gauge field, and the winding number is zero for all $\underset{\sim}{\boldsymbol{A}} \in \Omega^1(\mathcal{M}, \mathfrak{u}(1))$. Consequently, the θ-term (with constant θ) is neither relevant in classical electromagnetism nor in QED, because it vanishes identically.

However, for non-Abelian gauge groups, the space of gauge transformations $\Gamma^{\mathrm{id}}(\mathcal{G})$ may have a non-trivial topology, containing gauge transformations which are not diffeomorphic. This allows for the possibility of gauge-equivalent field configurations which cannot be *continuously* transformed into one another. In the case where $\mathcal{M}$ is $(1+3)$-dimensional spacetime and $\mathcal{G} = \mathcal{SU}(3)$ is the total gauge group of QCD, the space $\Gamma^{\mathrm{id}}(\mathcal{SU}(3))$ consists of path-disconnected regions labelled by some $\nu \in \mathbb{Z}$. (In topological language, the third homotopy group $\pi_3(\Gamma^{\mathrm{id}}(\mathcal{SU}(3))) \cong \mathbb{Z}$ is the group of integers.) This means that there are topologically inequivalent gauge transformations of any given $\underset{\sim}{\boldsymbol{A}}$, with the winding number labelling each

[9] We consider the space of asymptotically-identity gauge transformations $\Gamma^{\mathrm{id}}(\mathcal{G})$, whose elements are maps $g : \mathcal{M} \to G$ with $g|_{\partial\mathcal{M}} = \mathrm{id}$ (or equivalently with $g(x) \to \mathrm{id}$ as $x \to \infty$), because $\underset{\sim}{\boldsymbol{A}}$ must be fixed on the boundary in order to prescribe boundary conditions. «Who realised this? 't Hooft? Wigner?»

distinct topological class. Elements $g_0 : \mathcal{M} \to G$ in the identity-connected component of $\Gamma^{\mathrm{id}}(\mathcal{G})$ are named *small* gauge transformations, and all others *large*. A large gauge transformation $g_\nu \in \Gamma^{\mathrm{id}}(\mathcal{G})$ shifts the winding number n of a gauge field $\underset{\sim}{\boldsymbol{A}}$ to $n + \nu$, giving rise to $\mathbb{Z}$-many gauge-equivalent fields $\{g_\nu \cdot \underset{\sim}{\boldsymbol{A}}\}_{\nu \in \mathbb{Z}}$ which belong to distinct homotopy classes. Gauge field configurations with nonzero winding number are known as *instantons* [9, 10].

Despite the name, large gauge transformations are not truly gauge symmetries, in the sense that they are not genuine automorphisms of the equations of motion. Indeed, large gauge transformations may transition between states which can be distinguished by physical measurement. For instance, if the gauge field $\underset{\sim}{\boldsymbol{A}}$ has winding number n, then the action of the θ-term

$$S[\underset{\sim}{\boldsymbol{A}}] = \int_{\mathcal{M}} \underset{\sim}{\mathcal{L}}_\theta[\underset{\sim}{\boldsymbol{A}}] = \theta n \hbar$$

is proportional to n. A large gauge transformation g_ν then shifts this action $S[\underset{\sim}{\boldsymbol{A}}] \to S[g_\nu \cdot \underset{\sim}{\boldsymbol{A}}] = S[\underset{\sim}{\boldsymbol{A}}] + \theta\nu\hbar$. From the path integral (1.6), this induces relative phases $e^{i\theta\nu}$ varying across the domain of integration which may interfere and alter the amplitude. In other words, instantons are measurable.

The θ-Vacuum

Of particular interest are the implications of instantons for the QCD vacuum. In a Yang–Mills theory whose Lagrangian includes the term $\underset{\sim}{\mathcal{L}}_{\mathrm{YM}} = \langle \underset{\sim}{\boldsymbol{F}} \wedge \star \underset{\sim}{\boldsymbol{F}} \rangle$, a *vacuum state* is one in which the field strength (and all other fields) vanish; $\underset{\sim}{\boldsymbol{F}} = \underset{\sim}{\boldsymbol{0}}$. Not only is this consistent with an identically vanishing gauge field $\underset{\sim}{\boldsymbol{A}}_0 = \underset{\sim}{\boldsymbol{0}}$, but also with gauge transformations (1.2) of $\underset{\sim}{\boldsymbol{A}}_0$,

$$g \cdot \underset{\sim}{\boldsymbol{A}}_0 = (\underset{\sim}{\mathrm{d}} g_\rho)(g_\rho)^{-1},$$

which are called "pure gauge" configurations. Small gauge transformations are not measurable and describe the same vacuum state, whereas large ones $g_n \cdot \underset{\sim}{\boldsymbol{A}}_0$ are distinguishable, hence describing *distinct vacua* $|n\rangle$ labelled by winding number. Since $g_\nu \cdot |n\rangle = |n + \nu\rangle$, these states are not gauge-invariant. The only vacuum states which are invariant under the gauge (up to an overall phase) are those of the form

$$|\theta\rangle = \sum_{n \in \mathbb{Z}} e^{i\theta n} |n\rangle \,,$$

for some constant parameter θ. Thus, the theory possesses a topological circle of distinct gauge-invariant vacua $|\theta\rangle$ labelled by the *vacuum angle* $\theta \in [0, 2\pi)$.

The existence of the enriched vacuum $|\theta\rangle$ is equivalent to the inclusion of the θ-term in an effective Lagrangian, in the following way: Denote by ${}_+\langle n|m\rangle_-$ the quantum amplitude that the vacuum state $|n\rangle$ at $t \to -\infty$ evolves to $|m\rangle$ at $t \to \infty$. The vacuum-to-vacuum amplitude is

$${}_+\langle \theta|\theta\rangle_- = \sum_{n,m} e^{i\theta(n-m)} {}_+\langle m|n\rangle_- = \sum_\nu e^{i\theta\nu} \sum_n {}_+\langle n|n+\nu\rangle_-, \tag{1.9}$$

with all summations over $\mathbb{Z}$. In the path integral formulation (1.6), the amplitude ${}_+\langle n|n+\nu\rangle_-$ can be expressed explicitly as

$${}_+\langle n|n+\nu\rangle_- = \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}}; \nu] \exp\left\{ \frac{i}{\hbar} \int_\Omega \underset{\sim}{\mathcal{L}} \right\}, \tag{1.10}$$

where $\mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu]$ means that the path integral is over all instanton gauge fields with winding number ν, since only those induce a transition $|n\rangle \mapsto |n+\nu\rangle$. Combining (1.9) and (1.10), the amplitude of evolution from the gauge invariant vacuum to itself is

$$\begin{aligned} {}_+\langle\theta|\theta\rangle_- &= \sum_\nu e^{i\theta\nu} \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu] \exp\left\{\frac{i}{\hbar}\int_\Omega \underset{\sim}{\mathcal{L}}\right\} \\ &= \sum_\nu \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu] \exp\left\{\frac{i}{\hbar}\int_\Omega \underset{\sim}{\mathcal{L}} + i\theta\nu\right\}, \end{aligned}$$

which, using (1.7) to write $i\theta\nu$ as the θ-term in the Lagrangian, is

$$= \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}}] \exp\left\{\frac{i}{\hbar}\int_\Omega \Big[\underbrace{\underset{\sim}{\mathcal{L}} + \frac{\theta}{8\pi^2}\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}}\rangle \hbar}_{\underset{\sim}{\mathcal{L}}_{\text{eff}}}\Big]\right\}.$$

The effects of the non-trivial vacuum are thus encapsulated in the effective Lagrangian

$$\underset{\sim}{\mathcal{L}}_{\text{eff}} = \underset{\sim}{\mathcal{L}} + \frac{\theta}{8\pi^2}\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}}\rangle \hbar.$$

It is in this sense that the θ-term *arises* in the Lagrangian due to the non-trivial QCD vacuum.

1.1.4 Dirac Fermion Fields and the Chiral Anomaly

«Need flowing introduction.»

An important kind of matter field in the standard model is the *Dirac fermion field* $\boldsymbol{\varphi}$, which transforms under the spin-$\frac{1}{2}$ representation of the Lorentz group. Dirac fermions take their values in a 4-component complex space, denoted $\mathbb{C}^4_{\text{D}}$. The entire matter content of the standard model, including quarks and leptons, is comprised purely of such fermion fields.

The Dirac matrices $\boldsymbol{\gamma}^\mu$ form a basis for the algebra of spacetime, the Clifford algebra $\mathcal{C}l_{1,3}(\mathbb{C})$, satisfying $\boldsymbol{\gamma}^{(\mu}\boldsymbol{\gamma}^{\nu)} = \eta^{\mu\nu}$, and are used to write 4-component Dirac fermions in the spin-$\frac{1}{2}$ representation. An inner product on fermions $\langle\boldsymbol{\psi},\boldsymbol{\varphi}\rangle \equiv \bar{\boldsymbol{\psi}}\boldsymbol{\varphi} \in \mathbb{R}$ is provided by the Dirac adjoint $\bar{\boldsymbol{\psi}} := \boldsymbol{\psi}^\dagger\boldsymbol{\gamma}^0$, so that Lorentz-invariant quantities may be naturally constructed. Finally, fermions may be separated into left-handed $\boldsymbol{\varphi}_+$ and right-handed $\boldsymbol{\varphi}_-$ components with the projection operators $\boldsymbol{\varphi}_\pm = \frac{1}{2}(1 \pm \boldsymbol{\gamma}^5)$ where $\boldsymbol{\gamma}^5 := i\boldsymbol{\gamma}^0\boldsymbol{\gamma}^1\boldsymbol{\gamma}^2\boldsymbol{\gamma}^3$. In theories which violate parity, left- and right-handed fermions may experience different interactions. (For instance, left-handed neutrinos interact in the standard model, while right-handed neutrinos are completely inert.)

The simplest fermion equation of motion, the Dirac equation, derives from the *Dirac Lagrangian density*

$$\underset{\sim}{\mathcal{L}}_{\text{Dirac}} = \bar{\boldsymbol{\varphi}}(i\hbar c\boldsymbol{\gamma}^\mu\partial_\mu - mc^2)\boldsymbol{\varphi}\ \underset{\sim}{\text{vol}},$$

which describes a single non-interacting spin-$\frac{1}{2}$ fermion of mass m. The Dirac Lagrangian may be localised in the presence of a gauge symmetry, giving

$$\underset{\sim}{\mathcal{L}}_{\text{Dirac}} = \bar{\boldsymbol{\psi}}(i\boldsymbol{\gamma}^\mu\nabla_\mu - m)\boldsymbol{\psi}\ \underset{\sim}{\text{vol}}, \tag{1.11}$$

where $\underset{\sim}{\nabla}\psi \equiv (\nabla_\mu \psi)\,\underset{\sim}{\mathrm{d}}x^\mu$ is the covariant derivative with respect to the gauge field, and where we have now begun to employ units in which $\hbar = c = 1$ for brevity. The matter field $\psi = \varphi_{(1)} \oplus \cdots \oplus \varphi_{(n)}$ may now more generally be comprised of multiple fermion fields, which may be rotated into one another by the gauge group (an example being *flavour symmetry* in QCD). The localised Dirac Lagrangian describes multiple fermion types and their interactions with the gauge bosons. The electromagnetic interactions of charged fermionic matter are described by (1.11) in the case of a $\mathrm{U}(1)$ gauge symmetry, where the gauge field $\underset{\sim}{\boldsymbol{A}} \equiv \underset{\sim}{A}$ is the electromagnetic vector potential.

The Anomalous Axial Symmetry

In quantum theories of fermions, the θ-term makes another important appearance, arising in the context of the *chiral anomaly*. The Dirac Lagrangian classically possesses two fundamental $\mathrm{U}(1)$ symmetries which rotate fermion phases *vectorially* or *axially*. The classical axial symmetry is violated upon quantisation—an effect known as the chiral anomaly.

To illustrate, the vectorial symmetry $\mathrm{U}(1)_V$ is the invariance of the Dirac Lagrangian under global Abelian transformations known as *vector fermion rotations*. The action of $\mathrm{U}(1)_V$ is a simple phase rotation $\psi \mapsto e^{i\alpha}\psi$, where α is the parameter. The conserved Noether current density associated to this symmetry is

$$\underset{\sim}{J}_V = \hbar c\,\bar{\psi}\star\underset{\sim}{\gamma}\,\psi, \qquad \text{i.e.,} \quad j_V^\mu = \hbar c\,\bar{\psi}\gamma^\mu\psi,$$

where $\underset{\sim}{\gamma} = \gamma_\mu\,\underset{\sim}{\mathrm{d}}x^\mu$. (These 3-form and vector representations are related by $\underset{\sim}{J}_V = \star\underset{\sim}{j}_V$.) The associated continuity equation reads $\underset{\sim}{\mathrm{d}}\underset{\sim}{J}_V = 0$ (i.e., $\partial_\mu j_V^\mu = 0$), corresponding to the conservation of charge.

On the other hand, the axial symmetry $\mathrm{U}(1)_A$ is invariance under *axial fermion rotations*, which transform left- ψ_+ and right-handed ψ_- fermion components oppositely:

$$\psi \overset{\mathrm{U}(1)_A}{\mapsto} e^{i\gamma^5\theta}\psi, \qquad \text{i.e.,} \quad \psi_\pm \overset{\mathrm{U}(1)_A}{\mapsto} e^{\pm i\theta}\psi_\pm. \tag{1.12}$$

«Isn't the mass term not invariant here?» The associated Noether current density,

$$\underset{\sim}{J}_A = \hbar c\,\bar{\psi}\star\underset{\sim}{\gamma}\,\gamma_5\,\psi, \qquad \text{i.e.,} \quad j_A^\mu = \hbar c\,\bar{\psi}\gamma^\mu\gamma_5\,\psi,$$

is conserved classically. The charge associated to $\underset{\sim}{J}_A$ is the number of left-handed particles minus the number of right-handed, named the *baryon number* in QCD. However, the $\mathrm{U}(1)_A$ symmetry is *anomalous*, meaning it does not survive the quantisation procedure, and is not an exact symmetry of the quantum theory. Specifically, an anomalous symmetry does not leave invariant the integral measure $\mathcal{D}[\cdots]$ in (1.6) of the path integral in the quantum theory [4]. Instead, the continuity equation $\underset{\sim}{\mathrm{d}}\underset{\sim}{J} = 0$ fails by the presence of none-other than the θ-term,

$$\underset{\sim}{\mathrm{d}}\underset{\sim}{J}_A \propto \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle\,,$$

with the constant of proportionality depending on the details of the theory (the number of fermion species, etc). This means that the axial current $\underset{\sim}{J}_A$ is not conserved—and hence CP symmetry is violated—in the presence of instantons, where $\int \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ is nonzero.

By Noether's theorem (1.4), this is equivalent to the addition of a total derivative to the Lagrangian $\underset{\sim}{\mathscr{L}}_{\text{eff}} = \underset{\sim}{\mathscr{L}} + \underset{\sim}{\mathrm{d}}\underset{\sim}{K}$. This total derivative is precisely the θ-term, because

$$\underset{\sim}{\mathrm{d}}\underset{\sim}{J}_A = \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \frac{\partial \underset{\sim}{\mathrm{d}}\underset{\sim}{K}}{\partial \theta} \qquad \Longrightarrow \qquad \underset{\sim}{\mathrm{d}}\underset{\sim}{K} = \theta \, \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle \,,$$

where here θ is the axial rotation parameter appearing in (1.12). The axial current $\underset{\sim}{J}'_A$ of this new Lagrangian $\underset{\sim}{\mathscr{L}}_{\text{eff}}$ is indeed conserved. In other words, an axial rotation does not leave the Lagrangian invariant, but instead generates an effective θ-term [11, § 8].

«What does this mean?? Relevant because Peccei-Quinn mechanism depends on chiral symmetry breaking. Write something about the gauge invariant QCD vacuum $|\theta\rangle$. "Recognizing the complicated nature of the QCD vacuum, effectively adds an extra term to the QCD Lagrangian." - [lecture notes on axions] »

1.2 QCD and the Strong CP Problem

We are almost prepared to express the theory of quantum chromodynamics so that the strong CP problem manifests itself. All that remains is to introduce the final piece of QCD—the fermion mass terms, and the implications of the chiral anomaly.

Quantum chromodynamics describes the strong interactions among hadronic matter. Fields which interact via the strong force are called *colour-charged*, and in QCD, the colour-charged fields are the *quarks*.

In QCD with N_c colours, a quark is a colour-charged Dirac fermion, represented by a matter field $\boldsymbol{\psi} = \boldsymbol{\varphi} \otimes \boldsymbol{c}$ with values in $\mathbb{C}^4_{\mathrm{D}} \otimes \mathbb{C}^{N_c}_{\mathrm{C}}$, where $\boldsymbol{\varphi}(x) \in \mathbb{C}^4_{\mathrm{D}}$ is a plain Dirac fermion and $\boldsymbol{c}(x) \in \mathbb{C}^{N_c}_{\mathrm{C}}$ is a N_c-component vector in *colour space*. A quark field $\boldsymbol{\psi}$ can either be viewed as a fermion with components in colour space, or equivalently as a N_c-tuple of fermions. The gauge group is $\mathrm{SU}(N_c)$, and its action on $\boldsymbol{\psi}$ is to transform colour space under the fundamental representation; i.e., $\boldsymbol{\psi} \mapsto \boldsymbol{\psi}' = \boldsymbol{\varphi} \otimes \mathsf{U}\boldsymbol{c}$ with $\mathsf{U} \in \mathrm{SU}(N_c)$. The gauge field $\underset{\sim}{\boldsymbol{A}}$, named the *gluon field*, is an $\mathfrak{su}(N_c)$-valued 1-form, which can equivalently be viewed as a collection of $\dim \mathfrak{su}(N_c) = N_c^2 - 1$ independent 1-form fields, or eight distinct gluon types in the case $N_c = 3$ as in the standard model. QCD can be constructed with N_f quark types—or *flavours*—by taking the matter field to be a direct product of N_f quark fields, each sharing the same $\mathrm{SU}(N_c)$ gauge action.

The Lagrangian of pure QCD is a sum of the Dirac and Yang–Mills Lagrangians,

$$\underset{\sim}{\mathscr{L}}_{\text{QCD}} = \bar{\boldsymbol{\psi}}(i\gamma^\mu \nabla_\mu - \mathsf{m})\boldsymbol{\psi}\, \underset{\sim}{\mathrm{vol}} - \langle \underset{\sim}{\boldsymbol{F}} \wedge \star \underset{\sim}{\boldsymbol{F}} \rangle + \frac{\theta}{8\pi^2} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle \,,$$

where $\boldsymbol{\psi} \equiv \boldsymbol{\psi}^{(1)} \oplus \cdots \oplus \boldsymbol{\psi}^{(N_f)}$ is the matter field separated into its quark flavours, and $\mathsf{m} = m_1 \oplus \cdots \oplus m_{N_f}$ is a diagonal matrix of quark masses. With indices written explicitly, and $\hbar$ and c temporarily reinstated for completeness, the Lagrangian density may be spelled out as

$$\mathscr{L}_{\text{QCD}} = \sum_{q=1}^{N_f} \bar{\psi}^{(q)}_{a\mathfrak{c}} \big(i\hbar c \gamma^{\mu a}{}_b \nabla_\mu - m_q c^2 \delta^a{}_b\big) \psi^{b\mathfrak{c}}_{(q)} - \frac{\hbar}{4} F^{\mathfrak{a}}{}_{\mathfrak{b}\mu\nu} F_{\mathfrak{a}}{}^{\mathfrak{b}\mu\nu} + \frac{\theta\hbar}{8\pi^2} F^{\mathfrak{a}}{}_{\mathfrak{b}\mu\nu} F_{\mathfrak{a}}{}^{\mathfrak{b}}{}_{\rho\sigma} \epsilon^{\mu\nu\rho\sigma},$$

where Latin and Fraktur indices denote $\mathbb{C}^4_{\mathrm{D}}$ fermion components and $\mathbb{C}^3_{\mathrm{C}}$ colour components, respectively. «Is this all correct?»

Yukawa Couplings and the Measurable $\bar{\theta}$-parameter

The QCD sector of the standard model is an extension of pure QCD with three colours and six quarks. The quarks are partitioned into *up type* and *down type*, and again into three *generations*, each with varying masses and charges under the other components of the standard model gauge group (1.1).

generation	I	II	III
up type	u $2.2\,\mathrm{MeV}/c^2$	c $1.3\,\mathrm{GeV}/c^2$	t $170\,\mathrm{GeV}/c^2$
down type	d $4.7\,\mathrm{MeV}/c^2$	s $0.1\,\mathrm{GeV}/c^2$	b $4.2\,\mathrm{GeV}/c^2$

Figure 1.1: Quark masses in the standard model.

Pure QCD contains a mass term $\bar{\psi}\mathsf{m}\psi$, giving each quark $\psi^{(q)}$ an intrinsic mass m_q. This is *not* the mechanism by which quarks exhibit mass in the standard model—it cannot be, since is $\bar{\psi}\mathsf{m}\psi$ not invariant under axial rotations. «Why must axial rotations be a symmetry at all?» Instead, quarks obtain mass via the *Higgs mechanism*, whereby ψ is coupled to the Higgs field H in *Yukawa interaction terms* in the Lagrangian [5, § 7.6.6].

$$\underset{\sim}{\mathcal{L}}_{\text{mass}} = \Re\big(H\bar{\psi}_+\mathsf{m}\,\psi_-\big) = \frac{1}{2}\big(H\bar{\psi}_+\mathsf{m}\,\psi_- + H^\dagger\bar{\psi}_-\mathsf{m}^\dagger\psi_+\big)$$

The Lagrangian of the standard model QCD sector is thus

$$\underset{\sim}{\mathcal{L}}^{\text{SM}}_{\text{QCD}} = \big[\bar{\psi}i\gamma^\mu\nabla_\mu\psi + \Re\big(H\bar{\psi}_+\mathsf{m}\,\psi_-\big)\big]\,\underset{\sim}{\mathrm{vol}} - \langle\underset{\sim}{\boldsymbol{F}}\wedge\star\underset{\sim}{\boldsymbol{F}}\rangle + \frac{\theta}{8\pi^2}\langle\underset{\sim}{\boldsymbol{F}}\wedge\underset{\sim}{\boldsymbol{F}}\rangle\,. \tag{1.13}$$

Under independent axial rotations of each of the quark fields, the Yukawa mass term is invariant provided the quark masses also shift phase. In addition, each independent $\mathrm{U}(1)_A$ rotation generates a corresponding θ-term, due to the chiral anomaly (§ 1.1.4). Thus, the full QCD Lagrangian (1.13) is invariant under transformations of the form

$$\psi_{(q)} \mapsto e^{i\gamma^5 a_q/2}\psi_{(q)}, \qquad m_q \mapsto e^{-ia_q}m_q, \qquad \theta \mapsto \theta + \sum_{q=1}^{N_f} a_q, \tag{1.14}$$

where α_q parametrise the N_f independent $\mathrm{U}(1)_A$ rotations. To aid physical interpretation, these $\mathrm{U}(1)_A$ freedoms are exploited in order to *normalise* the Yukawa mass terms by making the mass phases real. «This is only a guess.»

The fact that the θ-parameter may be redefined by axially rotating the quark fields means that θ is not directly observable. However, this gauge freedom can be fixed by defining

$$\bar{\theta} = \theta + \arg\det\mathsf{m} = \theta + \arg\prod_{q=1}^{N_f} m_q,$$

which is invariant under (1.14), as the two right-hand terms transform by $\pm\sum a_q$. The Lagrangian of the QCD sector of the standard model, complete with the $\bar{\theta}$-term, is thus

$$\underset{\sim}{\mathcal{L}}^{\text{SM}}_{\text{QCD}} = \big[\bar{\psi}i\gamma^\mu\nabla_\mu\psi + \Re\big(H\bar{\psi}_+\mathsf{m}\,\psi_-\big)\big]\,\underset{\sim}{\mathrm{vol}} - \langle\underset{\sim}{\boldsymbol{F}}\wedge\star\underset{\sim}{\boldsymbol{F}}\rangle + \frac{\bar{\theta}}{8\pi^2}\langle\underset{\sim}{\boldsymbol{F}}\wedge\underset{\sim}{\boldsymbol{F}}\rangle\,. \tag{1.15}$$

The Statement of the Strong CP Problem

Proceeding with the assumption that $\bar{\theta} \neq 0$, one finds that the strong force now violates CP symmetry. A physical prediction of the standard model modified with a CP-violating QCD sector (1.15) is that the neutron is expected to possess an electric dipole moment of approximate magnitude $|d_n| \approx 10^{-18}\, e\,\mathrm{cm}$. In reality, current measurements [12, 13] of the neutron's electric dipole moment yield a tight upper bound of $|d_n| \lesssim 10^{-26}\, e\,\mathrm{cm}$, which in turn implies a stringent constraint on the $\bar{\theta}$-parameter, $|\bar{\theta}| \lesssim 10^{-10}$ [3]. Thus, the $\bar{\theta}$-term is not considered to be part of the standard model.

However, $\bar{\theta}$ can only be zero if apparently unrelated parameters of the standard model perfectly cancel each other: the vacuum angle θ, a QCD parameter; and the quark mass phases $\arg\det \mathsf{m}$, deriving from multiple electroweak parameters. One therefore expects $\bar{\theta}$ to be $\mathcal{O}(1)$ in Nature, and its extremely small value is hence a problem of fine-tuning. The strong CP problem is then the question, "why is $\bar{\theta}$ so small?"

At first sight, the strong CP problem may not appear to be a problem at all. After all, QCD is a theory whose Lagrangian possibly—but not necessarily—admits a term $\propto \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ which gives rise to CP-violating interactions in the strong force, predicting an electric dipole moment of the neutron. Empirical data is consistent with the neutron's electric dipole moment (and hence the CP-violating term) being zero. From a phenomenological perspective, it is satisfactory to simply leave the $\bar{\theta}$-term out of the theory's Lagrangian and end the story there. Indeed, a tautological way to 'resolve' the strong CP problem is to simply require that CP be a symmetry of the strong force. However, this only begs the question of why CP symmetry appears to be preserved in some sectors of the standard model while it is broken in others.

Furthermore, there is a strong argument that the inclusion of the CP-violating term is "natural." That is, we lack reason to exclude it on a theoretical basis: it is Lorentz and gauge invariant, etc.; it is an implication of the non-trivial vacuum structure of QCD (instantons); and it arises via the chiral anomaly for fermions. From an empirical perspective, if no CP-violating interactions were observed in Nature, then $\bar{\theta}$ could be justifiably set to zero on the basis of symmetry. However, the weak interaction is explicitly parity-violating.[10] Hence, the fact that the $\bar{\theta}$-term violates CP is not theoretically satisfactory reason for its exclusion.

The strong CP problem differers from other fine-tuning problems in the standard model in the sense that it is of almost no consequence to everyday physics. Variation of the θ-parameter hardly affects nuclear physics at all because its effects are suppressed by the quark masses [14]. On the other hand, variations of the cosmological constant, for example, predict universes drastically different to our own, and similarly for the value of the weak scale, or the quark and lepton masses. Such fine-tuning problems at least have anthropic solutions—but the strong CP problem does not.[11] The strong CP problem is therefore a compelling theoretical indication that the standard model remains incomplete.

[10] In fact, the standard model is asymmetric under all combinations of charge conjugation, C; parity P; and time-reversal T modulo the prevailing combined CPT symmetry.

[11] Given a theory linking the presence of dark matter to the smallness of θ, an anthropic solution may exist if it turns out that dark matter is necessary for, e.g., galaxy formation (investigated in [14]).

1.3 The Massless Quark Solution

The simplest resolution to the strong CP problem is to stipulate that at least one quark is in fact massless. If this were true, then $\det \mathsf{m}$ would vanish, and the parameter $\bar{\theta} = \theta + \arg\det \mathsf{m}$ would be rendered unphysical. The *massless quark solution* is the claim that $\bar{\theta} \approx 0$ because the up quark is massless, $m_u = 0$.

At first sight, this economical resolution to the strong CP problem appears to be in contradiction with the experimentally determined masses of the quarks, all of which are non-zero (including the up quark, with mass $\sim 2.2\,\mathrm{MeV}$). However, it was realised in the mid-1980s that the mass of the up quark has *two* contributions in the standard model Lagrangian: not only the Yukawa mass m_u (the 'bare mass') as introduced above, but also a non-perturbative contribution m_{eff} from topological effects (i.e., instantons) [15]. Only the bare quark masses contribute to the value of $\bar{\theta}$ via the quark mass matrix m. Importantly, it was plausible that this secondary source of the up quark's mass could be of order $m_{\text{eff}} \approx 2.2\,\mathrm{MeV}$, allowing m_u to vanish while still preserving the up quark's overall mass.

The massless up quark hypothesis remained controversial until *lattice gauge theory* had advanced sufficiently to made numerical simulations of non-perturbative effects in QCD possible. Strong consensus that the instanton contribution m_{eff} is not sufficiently large was reached in late 2019 [15–17]. Instead, another mechanism is required to explain the smallness of the $\bar{\theta}$-parameter.

2 The Axion Solution

The Peccei–Quinn Mechanism

Perhaps the most famous resolution to the strong CP problem is the Peccei–Quinn theory of the *axion*, first proposed in 1977 [18]. In essence, the axion solution involves extending the standard model in order to promote the original $\bar{\theta}$-parameter to a field $\bar{\theta}(x) = \bar{\theta}_{\text{SM}} + a(x)$ in such a way that it is dynamically relaxed to zero. In doing so, a new massive boson described by the scalar axion field $a(x)$ is necessarily introduced. There are different inequivalent ways to extend the standard model to realise the axion solution, but all Peccei–Quinn axion models share the same necessary features:

- The extended Lagrangian $\underset{\sim}{\mathcal{L}}_{\text{PQ}}$ possesses an additional global chiral symmetry $\mathrm{U}(1)_{\text{PQ}}$. The exact action of this *Peccei–Quinn symmetry* depends on the particular axion model, and is not of central importance. The defining feature of $\mathrm{U}(1)_{\text{PQ}}$ is that it is chiral, so that a $\mathrm{U}(1)_{\text{PQ}}$ transformation by α radians anomalously induces a θ-term $\alpha \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ in the effective Lagrangian (via the chiral anomaly).

- The Peccei–Quinn symmetry $\mathrm{U}(1)_{\text{PQ}}$ is *spontaneously broken*, and the single resulting Nambu–Goldstone boson is named the axion field, $a(x)$. Being a Nambu–Goldstone boson, the axion transforms as $a(x) \mapsto a(x) + \alpha$ under a $\mathrm{U}(1)_{\text{PQ}}$ rotation of α radians. The chiral anomaly results in a potential for the axion $a(x)$ (giving rise to axion mass) with a potential minimum occurring where $a(x) = -\bar{\theta}_{\text{SM}}$.

If the extra $\mathrm{U}(1)_{\text{PQ}}$ symmetry was indeed an exact gauge symmetry, then the strong CP problem is trivially solved, because a $\mathrm{U}(1)_{\text{PQ}}$ rotation by $\bar{\theta}$ radians cancels the $\bar{\theta}$-term in the effective Lagrangian, meaning the dynamics of the theory are equivalent to one in which $\bar{\theta} = 0$. However, the main result of Peccei and Quinn [18] is that $\mathrm{U}(1)_{\text{PQ}}$ need not be an exact symmetry of the theory: if $\mathrm{U}(1)_{\text{PQ}}$ is spontaneously broken, then $\bar{\theta}$ is still driven to zero because it obtains a potential from the chiral anomaly with a minimum at $\bar{\theta} = 0$, [19].

2.1 Axion Models

A Toy Axion Model

To illustrate the Peccei–Quinn mechanism explicitly, consider a minimal toy axion model, which involves the addition of two fields to the standard model: a complex scalar field Φ known as the *parent field* (so named since its phase, after spontaneous breaking, is the axion field); and an additional fermion $\boldsymbol{q}$. The extended Lagrangian takes the form

$$\underset{\sim}{\mathcal{L}}_{\text{PQ}} = \underset{\sim}{\mathcal{L}}_{\text{SM}}^{\bar{\theta}} + [\langle \underset{\sim}{\mathrm{d}}\Phi, \underset{\sim}{\mathrm{d}}\Phi \rangle + \mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-)]\underset{\sim}{\mathrm{vol}} + \underset{\sim}{\mathcal{L}}_q, \tag{2.1}$$

where $\underset{\sim}{\mathcal{L}}_{\text{SM}}^{\bar{\theta}}$ is the standard model Lagrangian (including the $\bar{\theta}$-term); $\langle \underset{\sim}{\mathrm{d}}\Phi, \underset{\sim}{\mathrm{d}}\Phi \rangle$ is a kinetic term for the parent field; $\mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-)$ is a Yukawa coupling term; and $\underset{\sim}{\mathcal{L}}_q$ stands for any other terms involving the new fermion $\boldsymbol{q}$. The action of $\mathrm{U}(1)_{\text{PQ}}$ on these fields is

$$\Phi \mapsto e^{i2\alpha}\Phi, \qquad \boldsymbol{q} \mapsto e^{i\gamma^5\alpha}\boldsymbol{q} \quad \text{or} \quad \begin{cases} \bar{\boldsymbol{q}}_+ \mapsto e^{i\alpha}\bar{\boldsymbol{q}}_+ \\ \boldsymbol{q}_- \mapsto e^{i\alpha}\boldsymbol{q}_- \end{cases},$$

which indeed leaves $\langle \underset{\sim}{\mathrm{d}}\Phi, \underset{\sim}{\mathrm{d}}\Phi \rangle$ and $\mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-)$ invariant. However, since $\boldsymbol{q}$ undergoes an axial rotation, the entire (effective) Lagrangian $\underset{\sim}{\mathcal{L}}_{\text{PQ}}$ is only invariant with the simultaneous subtraction of a term $\alpha \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ arising from the chiral anomaly. Therefore, the entire action of $\mathrm{U}(1)_{\text{PQ}}$ is to transform $\bar{\theta} \mapsto \bar{\theta} - \alpha$, as well as the fields. At this stage, the theory predicts CP violation in the strong sector in areas where the axions and instantons interact such that $\bar{\theta}(x) \neq 0$.

The final component of the axion solution is to make the parent field Φ spontaneously break. This may be done by adding a “Mexican hat” potential to $\underset{\sim}{\mathcal{L}}_{\text{PQ}}$ of the form

$$V(|\Phi|) = \lambda\big(|\Phi|^2 - f_a^2\big)^2,$$

where the parameter f_a is interpreted as the *axion scale*: the energy below which axion dynamics are relevant.«???» Below this energy scale, the parent field Φ relaxes to some non-unique minimum of the form

$$\Phi = |\Phi| e^{i \arg \Phi} = f_a e^{ia/f_a},$$

where a varies across space. At sufficiently low energies, the effective degree of freedom is the phase of Φ, not Φ itself.[1] The resulting phase $a(x)$ is named the Nambu–Goldstone boson associated with the spontaneous breaking of $\mathrm{U}(1)_{\text{PQ}}$ by the parent field Φ, and is identified as the axion field. After spontaneous breaking, the Yukawa term involves a complex mass, which can be normalised by a $\mathrm{U}(1)_{\text{PQ}}$ rotation by $-a/f_a$

$$\mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-) = f_a \mathbb{R}(\bar{\boldsymbol{q}}_+ e^{ia/f_a} \boldsymbol{q}_-) \mapsto f_a \mathbb{R}(\bar{\boldsymbol{q}}_+ \boldsymbol{q}_-).$$

This axial rotation of $\boldsymbol{q}$ induces a corresponding rotation of $\theta \mapsto \theta - a/f_a$, so that the (effective, normalised) Lagrangian (2.1) becomes

$$\underset{\sim}{\mathcal{L}}_{\text{PQ}} = \underset{\sim}{\mathcal{L}}_{\text{SM}} + [f_a^2 \langle \underset{\sim}{\mathrm{d}}a, \underset{\sim}{\mathrm{d}}a \rangle + f_a \mathbb{R}(\bar{\boldsymbol{q}}_+ \boldsymbol{q}_-)]\underset{\sim}{\mathrm{vol}} + \underset{\sim}{\mathcal{L}}_q + \left(\bar{\theta} - \frac{a}{f_a}\right)\frac{1}{8\pi^2} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle, \tag{2.2}$$

after spontaneous symmetry breaking of Φ.

Peccei and Quinn showed that the last term in (2.2) provides an effective potential $V(a)$ for the axion whose minimum occurs at $V(\bar{\theta} f_a) = 0$, giving the axion a vacuum expectation value

[1] We assume that λ is sufficiently large that the radial degree of freedom ρ in $\Phi = (f_a + \rho)e^{ia/f_a}$ can be neglected at the energy scale f_a.

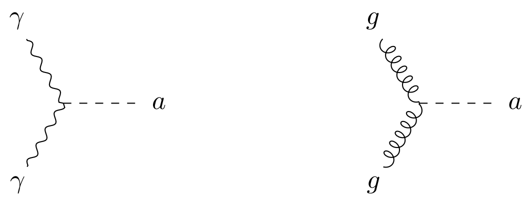
$\langle a \rangle = \bar{\theta} f_a$ and a mass $m_a = \partial^2 V / \partial a^2 \,|_{\langle a \rangle}$. The axion field a is not physical, since $a \mapsto a + \alpha$ under a $\mathrm{U}(1)_{\mathrm{PQ}}$ rotation; however, the deviation from the expectation value $a_{\mathrm{phys}} := a - \langle a \rangle$ is physical. Expressing the Lagrangian in terms of the physical axion field reveals that the $\bar{\theta}$-term vanishes, thus solving the strong CP problem [19]. Focusing on terms involving a_{phys}, the effective Lagrangian is

$$\underset{\sim}{\mathcal{L}}_{\mathrm{PQ}} = \underset{\sim}{\mathcal{L}}_{\mathrm{SM}} + \underset{\sim}{\mathcal{L}}'_q + \left[f_a^2 \left\langle \underset{\sim}{\mathrm{d}} a_{\mathrm{phys}}, \underset{\sim}{\mathrm{d}} a_{\mathrm{phys}} \right\rangle + \frac{1}{2} m_a^2 a_{\mathrm{phys}}^2 \right] \underset{\sim}{\mathrm{vol}} + \frac{a_{\mathrm{phys}}}{f_a} \frac{1}{8\pi^2} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle \,, \tag{2.3}$$

where $\underset{\sim}{\mathcal{L}}'_q$ includes all terms involving $\boldsymbol{q}$. Different axion models give rise to different $\underset{\sim}{\mathcal{L}}'_q$, but otherwise share the Lagrangian (2.3). The axion mass m_a depends on the axion scale f_a as

$$m_a \approx 6\,\mathrm{eV} \left(\frac{10^6 \mathrm{GeV}}{f_a} \right)$$

and is otherwise independent of the axion model, using accepted values of standard model parameters [20].



(a) (b)

Figure 2.1: Axion–photon and axion-gluon interaction vertices.

The precise axion–matter interactions entering through the term $\underset{\sim}{\mathcal{L}}'_q$ are model dependent, but generally have coupling strengths inversely proportional to the axion scale, f_a [21]. However, all Peccei–Quinn axions interact with the gauge field though the last term in the Lagrangian (2.3). The last term is proportional to $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \underset{\sim}{F} \wedge \underset{\sim}{F} + \langle \underset{\sim}{\boldsymbol{G}} \wedge \underset{\sim}{\boldsymbol{G}} \rangle$, where $\underset{\sim}{\boldsymbol{F}} = \underset{\sim}{F} \oplus \underset{\sim}{\boldsymbol{G}}$ is the total gauge field decomposed into the electromagnetic $\underset{\sim}{F}$ and gluonic $\underset{\sim}{\boldsymbol{G}}$ gauge fields. In perturbation theory, this corresponds to a Feynman vertex in which an axion and two photons $a\gamma\gamma$, or an axion and two gluons agg meet, as in figure 2.1a. The $a\gamma\gamma$ interaction is strong where $\underset{\sim}{F} \wedge \underset{\sim}{F} = (\boldsymbol{E} \cdot \boldsymbol{B})\, \underset{\sim}{\mathrm{vol}}$ is large. This implies that axions may be generated from photons, and vice-versa, in the presence of strong electromagnetic fields, suggesting methods for experimentally detecting axions in the lab or in magnetised stellar bodies. The agg vertex gives rise to interactions between axions and strongly-interacting hadronic matter (in particular pions) [20].

2.1.1 The Original Peccei–Quinn–Weinberg–Wilczek Axion

The first proposed axion model, the Peccei–Quinn–Weinberg–Wilczek (PQWW) axion, [22] implements the $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry by supposing that the standard model possesses two Higgs fields H_1 and H_2 which couple differently to up and down quarks, instead of just one which couples to all quarks. «Additionally, a complex scalar field Φ is introduced as in the toy model, which is spontaneously broken and contains the axion as its angular degree of freedom.» «

WRONG: Apparently the complex scalar field is also a Higgs doublet!? [Marsh - Axion Cosmology, pg 7] » Denoting by $\boldsymbol{\psi}_\pm = \boldsymbol{\psi}_\pm^{(u)} \oplus \boldsymbol{\psi}_\pm^{(d)}$ the left- and right-handed quarks arranged into up-type and down-type parts, the two Higgs fields

$$\underset{\sim}{\mathcal{L}}_{\text{Yukawa}} = \Re\big(H_1 \bar{\boldsymbol{\psi}}_+ \mathsf{m}_u \boldsymbol{\psi}_-^{(u)} + H_2 \bar{\boldsymbol{\psi}}_+ \mathsf{m}_d \boldsymbol{\psi}_-^{(d)}\big),$$

where m_u and m_d are (non-square) matrices of Yukawa coupling constants. The first Higgs field H_1 gives mass to the up-type quarks, and H_2 to down-type quarks. The presence of the two Higgs fields lets $\underset{\sim}{\mathcal{L}}_{\text{Yukawa}}$ be invariant under two independent chiral rotations of the up and down quarks, hence accomplishing the additional $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry.

In this model, the axion scale f_a is necessarily on the order of the electroweak scale, $f_{\mathrm{EW}} \approx 246\,\mathrm{GeV}$ (which is the vacuum expectation value of the Higgs field). The resulting axions are too massive ($m_a \approx 25\,\mathrm{keV}$) and too strongly interacting to agree with experiment. In particular, the PQWW axion is ruled out by the non-observation of the kaon decay $K^+ \to \pi^+ + a$ in electron beam-dump experiments[2] [19, 23]. Any successful axion model must have a higher energy scale f_a (i.e., lighter mass m_a) to be compatible with the constraints which excluded the PQWW model [22].

2.1.2 Light Invisible Axions

Axions with a larger energy scale $f_a \gg f_{\mathrm{EW}}$ are light ($m_a \sim 1/f_a$), long lived (e.g., the rate of $a \to 2\gamma$ goes as $(f_a)^5$) and weakly interacting (couplings are suppressed by $1/f_a$). In particular, electromagnetic interactions are weak, rendering them invisible. Such axion models generally fall into two classes: [19, 22]

- *The Kim–Shifman–Vainshtein–Zakharov (KSVZ) Axion*
  The KSVZ model introduces an additional massive quark $\boldsymbol{q}$ as well as the parent field Φ.

- *The Dine–Fischler–Srednicki–Zhitnitsky (DFSZ) Axion*
  The DFSZ model contains two Higgs doubles like the PQWW model, but also contains a separate scalar parent field Φ.

2.2 Laboratory Bounds

[2] Beam-dump experiments involve firing high-energy protons into a high-absorption material in order to isolate neutral particles which are created from the decelerating protons and which propagate through the absorber.

Terminology from cosmology

- *thermalisation* — the process of a particle species reaching thermal equilibrium (i.e., uniform energy and abundance) over cosmological scales, mitigated by self-interactions or processes involving other matter species which diffuse energy.
- *freeze-out* — the point beyond which the rates of thermalising processes become negligible due to a sufficiently large rate of cosmic expansion, resulting in persistent non-equilibrium distributions of a particle species. Since the universe cools as it expands, freeze-out may be viewed as the change of phase of a species from a gas to a cooler condensate.

3 Axions in Cosmology

With the advent of precision cosmology, particle physicists can use the entire universe as a laboratory for ever more sensitive experiments.

[Role of axions: possible dark matter candidates]

Many cosmological tests of axion-like particles involve predicting relative particle abundances in the universe at various epochs, such as the baryon-to-photon or neutrino-to-photon ratios. Such arguments begin with a minimal 'thermodynamical' model of the universe as a homogeneous, isotropic, expanding background upon which different particle species exist uniformly distributed and in thermal equilibrium. The universe's macrostate is characterised by each species' abundance and energy distribution. Interactions and processes between species, which at any time may depend on present particle abundances and energies, define differential relations which can be solved to determine the abundance of each species at any point in the universe's evolution.

The assumptions of isotropy and homogeneity specify a spacetime with a Friedmann–Lemaître–Robertson–Walker (FLRW) metric,

$$\underset{\sim}{g} = -c^2 \underset{\sim}{\mathrm{d}}t^2 + a(t)^2 \left(\frac{\underset{\sim}{\mathrm{d}}r^2}{1 - kr^2} + r^2 \underset{\sim}{\Theta} \right)$$

where $k \in \{+1, 0, -1\}$ reflects the type of spatial curvature, $\underset{\sim}{\Theta} = \underset{\sim}{\mathrm{d}}\theta^2 + \sin\theta\, \underset{\sim}{\mathrm{d}}\varphi^2$ is the metric of the unit sphere, and $\underset{\sim}{\mathrm{d}}x^2 \equiv \underset{\sim}{\mathrm{d}}x \otimes \underset{\sim}{\mathrm{d}}x$. Standard cosmological models have flat spatial curvature, $k = 0$. The scale factor $a(t)$ describes the cosmological evolution of the universe and defines the *Hubble parameter,* $H := \dot{a}/a$, or cosmic expansion rate. If the rate Γ of a particle interaction is large (corresponding to large probability per unit spacetime volume for the interaction to occur), then it will provide a mechanism for thermalisation of the species involved—or in the case of production and decay processes, will drive the species to abundance or extinction. If the rate is smaller than the rate of cosmic expansion, $\Gamma \ll H$, then the interaction or process will freeze-out and become negligible. If a species' most dominant interactions freeze-out, then it becomes thermally isolated from other fields and its abundance remains fixed.

3.1 Axion Interactions and Processes

Axions decay into photons via the axion-photon vertex at a rate

$$\Gamma_{a\to 2\gamma} = \frac{g_{a\gamma\gamma}^2 m_a^3}{64\pi} \approx 10^{-24}\,\mathrm{s}^{-1}\left(\frac{m_a}{\mathrm{eV}}\right)^5,$$

where the coupling strength $g_{a\gamma\gamma}$ can be approximately written in terms of the mass given, with an $\mathcal{O}(1)$ model dependence. Thus, for axions to exist in significant abundance, they must be sufficiently light, less the decay process dominates. The inverse decay $2\gamma \to a$ is also possible, with a rate $\Gamma_{2\gamma\to a} \propto 1/T$ which increases as the universe cools, so that axions recouple at late epochs [20, 22].

Axions also interact by the strong force with hadronic matter via the gluon-axion vertex, 2.1b. In particular, this gives rise to Primakoff and Compton electron scattering processes shown in figure 3.1. These two scattering processes, along with photon decay and inverse decay, are the dominant interactions relevant to cosmology.

(a) Primakoff

(b) Compton

Figure 3.1: Dominant axion processes with electrons (or positrons with arrows reversed).

Bibliography

[1] S. Weinberg, "The Making of the standard model", The European Physical Journal C-Particles and Fields **34**, 5–13 (2004).

[2] D. J. Gross, "The discovery of asymptotic freedom and the emergence of QCD", Proceedings of the National Academy of Sciences **102**, 9099–9108 (2005).

[3] P. D. Group et al., "Review of Particle Physics", Progress of Theoretical and Experimental Physics **2020**, 083C01 (2020) `10.1093/ptep/ptaa104`.

[4] D. Tong, *Cambridge, Department of Applied Mathematics and Theoretical Physics, Lecture Notes: Gauge Theory*, URL: `http://www.damtp.cam.ac.uk/user/tong/gaugetheory.html`. Last visited on 2020/05/24, 2018.

[5] M. Hamilton, *Mathematical Gauge Theory: With Applications to the Standard Model of Particle Physics*, Universitext (Springer International Publishing, 2017).

[6] E. Mitsou, *Differential form description of the Noether-Lagrange machinery, vielbein/gauge-field analogies and energy-momentum complexes*, 2013.

[7] E. Witten, "Quantum field theory and the Jones polynomial", Communications in Mathematical Physics **121**, 351–399 (1989).

[8] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, "Pseudoparticle solutions of the Yang-Mills equations", Physics Letters B **59**, 85–87 (1975).

[9] G. Gabadadze and M. Shifman, "QCD vacuum and axions: What's happening?", International Journal of Modern Physics A **17**, 3689–3727 (2002).

[10] S. Vandoren and P. van Nieuwenhuizen, *Lectures on instantons*, 2008.

[11] B. Gripaios, *Lectures on Physics Beyond the Standard Model*, 2015.

[12] C. Abel et al., "Measurement of the Permanent Electric Dipole Moment of the Neutron", Phys. Rev. Lett. **124**, 081803 (2020).

[13] M. Dine, *TASI Lectures on The Strong CP Problem*, 2000.

[14] M. Dine, L. S. Haskins, L. Ubaldi, and D. Xu, "Some remarks on anthropic approaches to the strong CP problem", Journal of High Energy Physics **2018** (2018) `10.1007/jhep05(2018)171`.

[15] C. Alexandrou et al., *Ruling out the massless up-quark solution to the strong CP problem by computing the topological mass contribution with lattice QCD*, 2020.

[16] M. Dine, P. Draper, and G. Festuccia, "Instanton effects in three flavor QCD", Phys. Rev. D **92**, 054004 (2015).

[17] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, 2016.

[18] R. D. Peccei and H. R. Quinn, "CP Conservation in the Presence of Pseudoparticles", Phys. Rev. Lett. **38**, 1440–1443 (1977).

[19] R. D. Peccei, "QCD, Strong CP and Axions", (1996).

[20] D. Cadamuro, S. Hannestad, G. Raffelt, and J. Redondo, "Cosmological bounds on sub-MeV mass axions", Journal of Cosmology and Astroparticle Physics **2011**, 003–003 (2011).

[21] L. D. Duffy and K. van Bibber, "Axions as dark matter particles", New Journal of Physics **11**, 105008 (2009).

[22] D. J. Marsh, "Axion cosmology", Physics Reports **643**, 1–79 (2016).

[23] E. Riordan et al., "Search for short-lived axions in an electron-beam-dump experiment", Physical Review Letters **59**, 755 (1987).