Università degli Studi di Padova

Dipartimento di Fisica e Astronomia
"Galileo Galilei"

Ph.D. Course in Physics

Cycle XXXVI

Angles of the Strong CP Problem

Coordinator

Ch.mo Prof. Giulio Monaco

Supervisor

Dr. Luca Vecchi

Ph.D. Candidate

Alessandro Valenti

September 2023

Abstract

The Strong CP problem is one of the most challenging and long-standing puzzles in physics beyond the Standard Model. In this thesis we review the formulation of the problem and thoroughly examine some of the most promising solutions put forth over the years. We first delve into both perturbative and non-perturbative aspects related to topological angles, including the issue of their perturbative renormalization. Then we identify and scrutinize two distinct classes of solutions addressing the Strong CP problem: UV solutions and those based on a Peccei-Quinn symmetry. Among the latter, we offer a brief overview of the QCD axion and its associated quality problem. To address this issue we propose an explicit model for a heavy axion, dubbed Grand Color axion, and study its phenomenology in detail. Next, we shift our focus to UV solutions. We start by illustrating the logic behind models of spontaneous P/CP violation, specializing then to the particular case of Nelson-Barr models. We meticulously investigate these models, pinpointing a naturalness issue inherently structured in this class of solutions. Finally, we present a set of scenarios designed to tackle this new problem and implement these ideas in a concrete UV realization.

Contents

Chapter 1

Introduction

Unveiling the fundamental laws of Nature is an extremely complex and challenging quest. Nevertheless, centuries of strenuous experimental and theoretical effort brought us to an understanding of our Universe which goes far beyond imagination. The result of this effort culminated in a simple yet amazingly successful theory known as the Standard Model of particle physics (SM).

Despite its incredible success, the Standard Model does not come without its own set of challenges and questions. Among these, the Strong CP problem is certainly one of the most intriguing puzzles. This introductory chapter starts with a concise review of the mathematical structure of the SM in section 1.1, providing the essential tools for the formal definition of the Strong CP problem in section 1.2. The last part of section 1.2 offers a glimpse about possible approaches to address the Strong CP problem, laying the groundwork for the detailed exploration undergone in the remaining chapters of this work as outlined in section 1.3.

1.1 Standard Model of Particle Physics

From a mathematical viewpoint, the Standard Model is formulated as a relativistic quantum field theory in four spacetime dimensions with gauge group

$$SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}. \tag{1.1.1}$$

The $SU(3)_{\mathrm{C}}$ factor mediates what are known as the strong interactions or Quantum Chromodynamics (QCD), while $SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ represents the unified electroweak interactions.

The matter content of the SM is comprised of a number of fermionic fields named quarks (charged under QCD) and leptons, and a single complex scalar field (Higgs field). Their charges are summarized in table 1.1. Additionally, each fermionic field comes in three generations or *flavours*.

Given the content of table 1.1, the Standard Model lagrangian is defined simply as the

	$SU(3)_{\rm C}$	$SU(2)_{\rm L}$	$U(1)_{\rm Y}$
q_L	$\mathbf{3}$	$\mathbf{2}$	$+\frac{1}{6}$
u_R	$\mathbf{3}$	$\mathbf{1}$	$+\frac{2}{3}$
d_R	$\mathbf{3}$	$\mathbf{1}$	$-\frac{1}{3}$
ℓ_L	$\mathbf{1}$	$\mathbf{2}$	$-\frac{1}{2}$
e_R	$\mathbf{1}$	$\mathbf{1}$	-1
H	$\mathbf{1}$	$\mathbf{2}$	$+\frac{1}{2}$

Table 1.1: Matter content of the Standard Model. q_L, u_R, d_R are the quarks doublets and singlets, while ℓ_L, e_R are the leptons doublets and singlets. H is the scalar Higgs field.

most general renormalizable lagrangian compatible with gauge and Lorentz symmetries:

$$\begin{aligned}\mathcal{L}_{\rm SM} = &-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu\,a} - \frac{1}{4}W^i_{\mu\nu}W^{\mu\nu\,i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ \bar{q}_{L,i}i\not{D}q_{L,i} + \bar{u}_{R,i}i\not{D}u_{R,i} + \bar{d}_{R,i}i\not{D}d_{R,i} + \bar{\ell}_{L,i}i\not{D}\ell_{L,i} + \bar{e}_{R,i}\not{D}e_{R,i} + (D_\mu H)^\dagger D^\mu H \\ &- (Y_d)_{ij}\,\bar{q}_{L,i}Hd_{R,j} - (Y_u)_{ij}\,\bar{q}_{L,i}\widetilde{H}u_{R,j} - (Y_e)_{ij}\,\bar{\ell}_{L,i}He_{R,j} + \text{h.c.} \\ &+ \mu^2|H|^2 - \frac{1}{2}\lambda|H|^4 \\ &+ \theta\frac{g^2}{32\pi^2}G\widetilde{G} + \theta_{\rm L}\frac{g_{\rm L}^2}{32\pi^2}W\widetilde{W} + \theta_{\rm Y}\frac{g_{\rm Y}^2}{32\pi^2}B\widetilde{B}.\end{aligned} \tag{1.1.2}$$

Some explanations are required. The first two lines include the covariant derivatives of the fields; their definitions are given in the conventions chapter. In particular, $G^a_{\mu\nu}, W^i_{\mu\nu}$ and $B_{\mu\nu}$ are the field strength of the $SU(3)_{\rm C}$, $SU(2)_{\rm L}$ and $U(1)_{\rm Y}$ gauge groups, respectively. The index associated to each fermion is their flavour index; in the following we will often leave it implicit. The third line identifies the Yukawa interactions, given in terms of three complex matrices Y_u, Y_d, Y_e and $\widetilde{H} = \epsilon H^*$, where ϵ is the Levi-Civita tensor in the $SU(2)_{\rm L}$ space. The fourth line encodes the Higgs field potential. The last line defines the topological terms, that we will discuss extensively in chapter 2.

The Higgs potential is minimized at $v = \sqrt{2}\mu/\sqrt{\lambda} \simeq 246$ GeV [1], where $H(x) = (0, v + h(x))^T/\sqrt{2}$. This gives rise to the electroweak symmetry breaking. The gauge group $SU(2)_{\rm L} \times U(1)_{\rm Y}$ is broken to $U(1)_{\rm EM}$, describing electromagnetic interactions. The would-be Nambu-Goldstone bosons (NGBs) arising from the spontaneous symmetry breaking of the gauge group are eaten by the gauge bosons associated to the broken generators, giving rise to the massive $W^\pm$ (charged under $U(1)_{\rm EM}$) and Z (neutral) bosons. Furthermore, electroweak symmetry breaking is responsible for the generation of the fermions' masses. Splitting the $SU(2)_{\rm L}$ doublets as $q_L = (u_L, d_L)^T$ and $\ell_L = (\nu_L, e_L)^T$, these are encoded in

$$\mathcal{L}_m = -\frac{v}{\sqrt{2}}Y_d\,\bar{d}_L d_R - \frac{v}{\sqrt{2}}Y_u\,\bar{u}_L u_R - \frac{v}{\sqrt{2}}Y_e\,\bar{e}_L e_R + \text{h.c.}\ . \tag{1.1.3}$$

The physical masses are extracted upon performing a singular value decomposition of the Yukawas in $\mathcal{L}_m$, achieved by rotating every fermionic field $f = d_L, d_R, u_L, u_R, e_L, e_R$ via a unitary matrix U_f. Such a rotation shifts the topological parameters of the gauge fields, as we will see later, and generates flavour- and CP-violating couplings between the up and down quarks:

$$\mathcal{L}_{\text{FV}} = -\frac{g_{\text{L}}}{2\sqrt{2}} V_{\text{CKM}}\, \bar{u}_L \gamma^\mu d_L W_\mu^+ + \text{h.c.} \tag{1.1.4}$$

Flavour and CP violation in the weak sector is thus described by the unitary Cabibbo-Kobayashi-Maskawa matrix $V_{\text{CKM}} = U_{u_L} U_{d_L}^\dagger$ and mediated by the heavy $W^\pm$ bosons.

Naturalness

The simple and elegant theory we just sketched is able to explain an incredible amount of phenomena that we observe in Nature. Yet, some key pieces of the puzzle are still missing. An obvious one is gravity. Naïve attempts of quantizing the Einstein-Hilbert action lead to a theory which is not extrapolable to arbitrary high energy scales. The experimental observation of neutrinos masses is also incompatible with the simple model (1.1.2). And the list goes on, with the questions of Dark Matter, Baryogenesis, etc.. These are all indisputable indications that the Standard Model needs to be extended. Now, whatever ultraviolet (UV) completion of the Standard Model is there, one expects (or rather, hopes) it to be able to reproduce the Standard Model parameters and particle content with little to no fine-tunings, much like the SM does with physics below the electroweak scale. For this to be true, the dimensionless combinations of the low-energy effective theory (namely the SM) parameters that one can build should all have $\mathcal{O}(1)$ values, *unless* they are protected by selection rules associated to a symmetry[1]. Theories failing these expectations are said to possess a *naturalness problem.* The perhaps most spectacular example of a naturalness problem is given by the cosmological constant problem. Taking as UV cutoff of the SM the reduced Planck mass $M_{\text{P}} \simeq 2.4 \times 10^{18}$ GeV, defining the scale around which a quantum theory of gravity is expected to live, one finds a fine-tuning of 120 orders of magnitude: $\Lambda_{\text{exp}} \sim 10^{-120} M_{\text{P}}^4$ [3]. Another, popular example is the weak-Planck hierarchy problem. The dimensionful parameter μ^2 in the Higgs potential is not protected by any symmetry and therefore one naïvely expects $\mu^2 \sim M_{\text{P}}^2$. The experimental value of the Higgs mass $m_H^2 = 2\mu^2 \simeq 125$ GeV [1] fails the naïve expectation and defines a naturalness problem.

It is important to stress that, strictly speaking, a naturalness problem is *not* a problem for the effective theory under consideration. From the effective theory perspective the incriminated "unnatural" parameters are not calculable and can only be experimentally determined. Explaining the smallness of these parameters without evoking fine-tunings is a challenge for the UV completion itself. Sometimes finding a clever mechanism to achieve this is relatively easy, as for the smallness of the pions' masses compared to the QCD confinement scale. Often

[1]This principle is quantitatively summarized in the 't Hooft naturalness' criterion [2], asserting that a theory is natural if, for all its parameters p which are small with respect to their fundamental scale p_{UV}, the limit $p \to 0$ corresponds to an enhancement of the symmetry of the system. This automatically guarantees the stability of these parameters under the Renormalization Group flow.

the challenge is harder and the solutions tend to grow in complexity. Regardless, the presence of naturalness problems likely points to the existence of some sophisticated mechanism within the underlying UV theory. And this is exactly what happens when one starts looking for solutions to another, popular naturalness problem: the Strong CP problem.

1.2 The Strong CP problem

The Strong CP problem is a naturalness problem immediately arising upon comparing the strength of weak and strong CP violation in the Standard Model[2]. In its renormalizable version, indeed, these are the only two observable sources of CP violation[3].

Weak CP violation is described in terms of the irreducible phase, δ, of the CKM matrix[4]

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{1.2.1}$$

and is therefore associated to processes mediated by the weak interactions. Actually, in physical observables δ never shows up alone but always in combination with the mixing angles in the so called *Jarlskog Invariant* [4]:

$$J = s_{12}^2 s_{13} s_{23} c_{12} c_{13}^2 c_{23} \sin\delta. \tag{1.2.2}$$

The reason is very clear: if any of the mixing angles is either 0 or $\pi/2$ (mod π), a pair of quarks becomes degenerate and one can always perform a chiral $SU(2)$ flavour rotation to eliminate the phase δ which becomes now unphysical. This property becomes explicit if we try to express J in terms of flavour invariants built using the Yukawas Y_u, Y_d, two of the fundamental CP-violating quantities in the Higgs basis. Their spurious flavour transformations $Y_u \to U_{q_L} Y_u U^\dagger_{u_R}, Y_d \to U_{q_L} Y_d U^\dagger_{d_R}$, together with the fact that we have three generations, lead to only one independent CP-odd invariant:

$$I_{\not{CP}} = \det\left[Y_u Y_u{}^\dagger, Y_d Y_d{}^\dagger\right]. \tag{1.2.3}$$

Expressing this quantity in the basis $Y_u = \widehat{Y_u}, Y_d = V_{\mathrm{CKM}}\widehat{Y_d}$, we get

$$I_{\not{CP}} = 2i(y_t^2 - y_c^2)(y_t^2 - y_u^2)(y_c^2 - y_u^2)(y_b^2 - y_s^2)(y_b^2 - y_d^2)(y_s^2 - y_d^2) s_{12}^2 s_{13} s_{23} c_{12} c_{13}^2 c_{23} \sin\delta \tag{1.2.4}$$

where we immediately notice that the expression vanishes as soon as a pair of quarks becomes degenerate, as expected. The last factor is precisely the quantity J defined earlier.

[2] See section 2.2 for the general definition of CP and 4.1 for its definition within the SM.

[3] CP violation associated to the $SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ topological angles is unobservable at low energies. By means of a chiral rotation the $SU(2)_{\mathrm{L}}$ angle can always be moved in front of the $U(1)_{\mathrm{Y}}$ topological term, that unlike non-Abelian group does not feature non-perturbative phenomena such as instantons (see chapter 2). Also, we are not including right-handed neutrinos and the associated U_{PMNS} which would bring additional CP-odd phases.

[4] We stick to the standard parametrization adopted in [1]. Here, $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$.

As of today, the Jarlskog invariant has been precisely measured in a multitude of experiments which set $J = 3.00^{+0.15}_{-0.09} \times 10^{-5}$. Independent measurements of the mixing angles allow to determine the value of the phase δ [1]:

$$\delta = 1.196^{+0.045}_{-0.043} . \tag{1.2.5}$$

Importantly, the suppression of J is due to the smallness of the mixing angles rather than the strength of CP violation in the weak sector, encoded in δ which is of $\mathcal{O}(1)$.

The origin of CP violation in the strong sector, instead, is associated to the other irreducible CP-odd parameter $\bar{\theta}$. This is given by a flavour invariant combination of the CP-odd quantities θ, the QCD topological angle, and the overall phase of the Yukawas:

$$\bar{\theta} = \theta - \arg \det Y_u Y_d. \tag{1.2.6}$$

This quantity is invariant under the anomalous $U(1)_{\mathrm{A}}$ rotation of the quarks $q_L \to e^{-i\alpha/2} q_L$, $u_R \to e^{i\alpha/2}$, $d_R \to e^{i\alpha/2} d_R$, sending $\theta \to \theta + 2N_g\alpha$ and $(Y_u, Y_d) \to e^{i\alpha}(Y_u, Y_d)$, and not protected by any symmetry except CP itself.

The $\bar{\theta}$ parameter is quite peculiar: precisely thanks to an anomalous $U(1)_{\mathrm{A}}$ rotation it can be entirely put in front of the QCD topological term, which is a total derivative ($G\widetilde{G} = \partial^\mu K_\mu$) and is therefore not associated to any vertex in ordinary perturbation theory (more on this in chapter 2). This renders $\bar{\theta}$ apparently unphysical. As we will discuss in chapter 2 in more detail, however, it is when we take into account non-perturbative phenomena that this parameter becomes observable. And in QCD this happens when it becomes strongly coupled. By far, the most important observable to which $\bar{\theta}$ is associated is the neutron electric dipole moment d_n. Precisely because of the non-perturbative nature of $\bar{\theta}$ it is difficult to make analytical predictions for d_n, which can be estimated as (see section 2.1.2):

$$d_n \sim c_n e \frac{m_u}{m_n^2} \bar{\theta} \tag{1.2.7}$$

In this expression m_n is the neutron mass, m_u the up-quark mass, e the electromagnetic coupling and c_n an incalculable constant expected to be in the range $|c_n| \in [1, 10]$. The most recent experimental searches for d_n set $|d_n|/e < 1.8 \times 10^{-26}$ cm at 90% C.L. [5], that upon employing (1.2.7) translates in the upper bound

$$|\bar{\theta}| < \mathcal{O}(0.5 - 5) \times 10^{-10}. \tag{1.2.8}$$

The comparison between (1.2.5) and (1.2.8) is what defines the *Strong CP problem*. The huge hierarchy between CP violation in the strong and the weak sector is a clear symptom that whatever completion of the SM is there, it must be either highly non-generic or extremely fine-tuned in order to explain a cancellation in $\bar{\theta}$ of order 10^{-10} while maintaining sizeable CP violation in the weak interactions[5].

Excluding the fine-tuning option, solutions to the Strong CP problem within the four-dimensional QFT realm[6] fall into two broad categories:

[5]It is worth pointing out, though, that even if a small $\bar{\theta}$ is not technically natural, its running within the SM first appears at 7-loops. Hence, it can be effectively treated as stable under the Renormalization Group (RG) flow (see section 2.2). Later we will show that this is not true in generic beyond the SM extensions, and a big $\bar{\theta}$ can be regenerated by the RG flow.

[6]We will not discuss String Theory solutions.

i) axion-like solutions;

ii) UV solutions.

This classification can be understood in terms of the spurious transformation properties of θ; namely its anomalous shift symmetry and P/CP-odd nature. Solutions exploiting the former lead to what is known as the axion mechanism, and comprise the first of the two categories. In this case the SM is extended in such a way that the anomalous axial transformation shifting θ becomes a real symmetry of the theory, the *Peccei-Quinn symmetry*, broken only by the topological term of QCD due to its anomalous nature. The spontaneous breaking of this symmetry at high-scales leads to an IR theory composed of the SM supplemented by a pseudo Nambu-Goldstone boson, the *axion*, whose potential generated at strong coupling by the topological term is responsible for stabilizing the axion at a value which effectively renders $\bar{\theta}$ unphysical. These elegant and dynamical solutions to the Strong CP problem are however jeopardized by the anomalous nature of the PQ symmetry itself, which must be of extremely high quality to ensure that the Strong CP problem is actually solved. For this reason, this class of solutions is known to be typically afflicted by a *quality problem*.

Solutions exploiting the P/CP property of θ, instead, define the class of *UV solutions* to the Strong CP problem. In this approach P/CP is assumed to be an exact symmetry of the theory at high energies, spontaneously broken at lower scales by the vacuum expetation value of some scalar field. While very elegant, the non-trivial challenge that such approaches face is that of generating a sizeable amount of weak P and CP violation and a chiral matter content while simultaneously ensuring (1.2.8).

1.3 Outline of the work

In the remaining chapters of this work we will try to convey the main ideas behind the two approaches just sketched by presenting some concrete works that have been put forth recently, focussing in particular on the original publications [6–9].

Chapter 2 is devoted to a more in-depth study of physics associated to topological angles, that in the previous section we just alluded to. This deviation is crucial in order to fully appreciate the nature of the problem itself. In section 2.1.1 we will review non-perturbative effects to which $\bar{\theta}$ is intrinsically tied, such as instantons and the neutron electric dipole moment. Then, in section 2.2, taken from the work [6], we will move to the issue of the perturbative renormalization of the topological angles. We will show how this question is both of theoretical interest and practical relevance, particularly in relation to UV solutions to the Strong CP problem.

Chapter 3 focuses on axion-like solutions to the Strong CP problem. In sections 3.1, 3.2 we will provide a review of the basic principles of the axion mechanism and its associated quality problem. The latter will naturally lead us to section 3.3, taken from [7], where we will study constructions in which the axion is more stable because of an increased size of its potential as induced by an additional strongly coupled group confining at a scale much bigger than the QCD one. For this approach to work an almost perfect alignment between the two potentials is required. As we will show, this can be obtained in a model in which

QCD and the new group are unified at high-scales in a bigger *Grand Color* group [7], whose construction we will present thoroughly including the details of the effective axion potential and the phenomenology of the model.

Chapter 4 deals with UV solutions to the Strong CP problem. After a brief review of the Spontaneous P/CP violation approach in section 4.1, we will quickly move to particular realizations denoted *Nelson-Barr models*. In section 4.2, taken from [8], we will study in detail the issue of generating an $\mathcal{O}(1)$ phase δ in models where CP is an exact UV symmetry, and we will quantitatively estimate the size of $\bar{\theta}$ generated after the spontaneous breaking of CP occurs. This will lead to the discovery of a new naturalness problem rooted in this kind of solutions, which we will overcome in section 4.3, taken from [9], by building an explicit UV completion of the Nelson-Barr effective theory in which CP violation is mediated to the SM in a *super-soft* way.

A final summary and possible future directions will be presented in chapter 5.

Chapter 2

Topological Angles

In a simple gauge theory, a topological angle θ is the coupling associated to the topological operator $Q(x)$:

$$Q(x) = \frac{g^2}{64\pi^2}\epsilon^{\mu\nu\rho\sigma}\delta^{ab}G^a_{\mu\nu}(x)G^b_{\rho\sigma}(x) \equiv \frac{g^2}{32\pi^2}G\widetilde{G}. \tag{2.1}$$

This operator is odd under conventional P and CP but gauge and Lorentz invariant, and should therefore be included in the lagrangian of any renormalizable gauge theory. However, a closer look reveals that (2.1) is actually negligible in ordinary perturbation theory. The reason is that it can be written as a total derivative

$$G\widetilde{G} = \partial_\mu K^\mu \tag{2.2}$$

where the current K^μ is called *Chern-Simons current*

$$K^\mu = 8\epsilon^{\mu\nu\rho\sigma}\,\mathrm{tr}\left(\frac{1}{2}A_\nu\partial_\rho A_\sigma - \frac{ig}{3}A_\nu A_\rho A_\sigma\right) = 4\epsilon^{\mu\nu\rho\sigma}\,\mathrm{tr}\left(\frac{1}{2}A_\nu G_{\rho\sigma} + \frac{ig}{3}A_\nu A_\rho A_\sigma\right). \tag{2.3}$$

In ordinary Feynman diagrams calculations, the fields are assumed to die off sufficently fast at infinity in order to work in the Fourier-transformed momentum space. This means that the operator $Q(x)$, being essentially a boundary term ($\int d^4x\, Q(x) = \int d^4x\, \partial_\mu K^\mu = \int_{|x|\to\infty} d^3x\, K_\perp$), does not contribute to the effective action and therefore to the perturbative expansion[1]. Beyond the perturbative approach, however, it is possible to find gauge field configurations for which the integral of $Q(x)$ is not identically vanishing. These field configurations are called *istantons* and are characterised by an integer number ν associated to the operator (2.1),

[1]For example, correlators for $G\widetilde{G}$ with two external legs are encoded in the partition function as

$$\mathcal{Z} \supset \int d^4x\, A^a_\mu(x) \int d^4y\, A^a_\nu(y)\, G^{\mu\nu}(x-y) \tag{2.4}$$

where $G^{\mu\nu}(x-y)$ is an ordinary Feynman diagram. Since it is impossible for $G^{\mu\nu}(x-y)$ to contain two derivatives contracted with a Levi-Civita tensor, this term cannot be associated to (2.1). Here the subtlety lies in the fact that in this expression integration by parts has been performed assuming that the boundary term gives no contribution.

$\int d^4x\, Q(x) = \nu$. Splitting the gauge field in a background given by these configurations plus quantum fluctuations dying off sufficiently fast at infinity, the path-integral is then divided in different sectors classified by the value of the integer ν:

$$\mathcal{Z} = \int \mathcal{D}\phi\, e^{iS + i\theta \int d^4x\, Q(x)} = \sum_{\nu} e^{i\theta\nu} \int \mathcal{D}\phi_\nu\, e^{iS_\nu} = \sum_{\nu} e^{i\theta\nu} \mathcal{Z}_\nu \tag{2.5}$$

where ϕ includes all the field entering in the action S including the gauge field fluctuations. In other words, the parameter θ is not unphysical but actually enters in a non-trival way in the partition function, behaving much like an angle (hence the name topological angle). The rôle of the instanton configurations in the path-integral and ultimately in determining the gauge theory vacuum will be the subject of the next section's investigation. We will find that θ actually does enter in physical observables, eventually leading in the case of QCD to the Strong CP problem as defined in chapter 1. For the very same reason it is also natural to wonder about the renormalization of the topological parameter, a question which has both theoretical and practical relevance. This aspect will be covered in section 2.2, which is taken from the original work [6].

2.1 Beyond Perturbation Theory

In this section we are interested in studying non-perturbative effects associated to the operator (2.1). This will naturally lead us to the concept of instantons and to the structure of the vacuum of gauge theories. Our discussion will be concise and mostly based on the expositions in [10–12]; other good references on the subject are [13–15]. Finally, we will discuss their applications to QCD and other non-perturbative phenomena relevant for phenomenology.

2.1.1 Instantons

Instantons are solutions to the classical equations of motion of the gauge fields that are localised in Euclidean time and space, nowhere singular and with a finite action. The reasons to consider an Euclidean spacetime instead of a Minkowski one are twofold: first, the path integral and its expansion are much more well-defined in Euclidean spacetime, where the action is positive definite and therefore the path-integral convergent. Second, we are interested in performing a semiclassical expansion of the path integral. In quantum mechanics, the WKB approximation tells us that this is done by considering solutions to the classical equations of motion with a Wick-rotated (imaginary) time. In quantum field theory, this amounts to consider our theory on a Euclidean spacetime and expand around field configurations with finite action (infinite actions would result in a zero in the path-integral).

The conversion from Minkowksi to Euclidean is done through a Wick rotation:

$$x_0 = -i x_4. \tag{2.1.1}$$

where x_4 is the Euclidean time coordinate. In this way scalar products become $a^\mu b_\mu = -(a \cdot b)_{\mathrm{E}}$, so that the metric is effectively Euclidean. Since we want the Euclidean action to enter in the path integral as e^{-S_E}, we define it as $S_E = -iS|_{x_0=-ix_4}$. Taking into account that $d^4x = -id^4x_\mathrm{E}$, the pure gauge action without the topological term becomes

$$S_E = \int d^4x_\mathrm{E}\,\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} \tag{2.1.2}$$

where now we do not need to distinguish upper and lower indices any more. Quantities involving an odd number of Levi-Civita tensors, instead, pick up an additional imaginary factor: in this way the topological operator becomes

$$\int d^4x\,Q(x) = -i\int d^4x_E\,\frac{g^2}{64\pi^2}\left(+\frac{i}{2}\epsilon_E^{\mu\nu\rho\sigma}G^a_{\mu\nu}G^a_{\rho\sigma}\right) = \int d^4x_E\,\frac{g^2}{32\pi^2}G\widetilde{G} \equiv \int d^4x_\mathrm{E}\,Q_E \tag{2.1.3}$$

and so enters in the Euclidean action as $S'_E = S_E - i\int d^4x_\mathrm{E}\,Q_E$. From now on we will drop the subscript which denotes Euclidean quantities.

In the semiclassical approximation, the starting points are solutions to the classical equation of motions which have finite action. Therefore we must look for gauge field configurations which satisfy

$$D_\mu G_{\mu\nu} = 0 \tag{2.1.4}$$

and such that the action (2.1.2) is finite. Let us start from the latter requirement. Since we are in Euclidean spacetime, we start by considering a sphere $S_3 \supset \mathbb{R}_4$ at some fixed radius r[2]. Because the measure in (2.1.2) scales as $\sim r^3$, finite action requires that the field strength must behave as $G_{\mu\nu} \sim r^{-3}$ for $r \to \infty$, and so naively $A_\mu \sim r^{-2}$. However, for $r \to \infty$ the gauge potential can also approach a pure gauge configuration:

$$A_\mu(x)|_{r\to\infty} = -\frac{i}{g}\left(\partial_\mu U\right)U^{-1} + o\left(\frac{1}{r^2}\right) \tag{2.1.5}$$

so that $G_{\mu\nu}(x)|_{r\to\infty} = 0$. Now, to be concrete consider the simple case of a $G = SU(2)$ gauge group. Since topologically $SU(2) \sim S_3$[3], what we are looking for are effectively maps from S_3 (the asymptotic boundary of Euclidean spacetime) to S_3 (gauge group transformations). These maps, however, are not all continuously deformable into each other: they are classified in equivalence classes which are encoded in what the mathematicians call the fundamental groups $\Pi_n(G)$. In the case at hand the relevant group is $\Pi_3(S_3) = \mathbb{Z}$. Therefore, solutions to the classical equations of motion can be labeled by an integer, the Pontryagin number ν. This number, also called Brouwer degree or winding number, can be expressed in terms of an integral over the volume form of the gauge group[4]:

$$\nu = \frac{1}{24\pi^2}\int_{S_3}\mathrm{tr}\,U^{-1}dUU^{-1}dUU^{-1}dU = \frac{1}{24\pi^2}\int_{S_3}d^3\theta\,\epsilon_{ijk}\,\mathrm{tr}\,U^{-1}\partial_iUU^{-1}\partial_jUU^{-1}\partial_kU. \tag{2.1.7}$$

[2]Note that the radius includes also the Euclidean time direction, $r = \sqrt{x_1^2+x_2^2+x_3^2+x_4^2}$.

[3]This is because an $SU(2)$ transformation can be written, thanks to the Pauli matrices identities, as $e^{i\alpha\cdot\sigma} = u_0\mathbb{1} + u\cdot\sigma$ subject to the constraint $u_0^2 + u\cdot u = 1$ coming from the requirement of unitarity. This identifies a sphere S_3 in the parameter space (u_0, u).

[4]It is straightforward to verify that this quantity is a topological invariant. Under coordinates transformations this is clearly invariant since it is a volume form (the transformation of the Levi-Civita tensor compensates the Jacobian factor from the measure). Under an infinitesimal gauge transformation, $\delta U = i\delta\alpha(x)^a t^a U \equiv \alpha U$. From $\delta U^{-1} = -U^{-1}\delta UU^{-1}$ and $\partial_iUU^{-1} = -U\partial_iU^{-1}$, we get $\delta(U^{-1}\partial_iU) = U^{-1}\partial_i\left(\delta UU^{-1}\right)U = U^{-1}\partial_i\alpha U$ and

To gain an understanding of what this number represents, consider the 1-dimensional case $S_1 \to U(1) \sim S_1$. Given the S_1 coordinate $\theta \in [0, 2\pi]$, we can have infinitely different maps given by

$$V^{(m)}(\theta) = e^{im\theta}. \tag{2.1.8}$$

where m is an integer. The Pontryagin number in this case is given by

$$\nu = \frac{1}{2\pi i}\int_{S_1} (V^{(m)})^{-1} dV^{(m)} = \frac{1}{2\pi i}\int_0^{2\pi} d\theta\, e^{-im\theta}\,(im)\,e^{im\theta} = m \tag{2.1.9}$$

and indeed $\Pi_1(S_1) = \mathbb{Z}$. Therefore, the index m is telling how many times the gauge group manifold S_1 is wrapped by the mapping. In the case of interest, $\Pi_3(S_3)$, the analogous quantity is given by the formula (2.1.7). But in this formula the quantities appearing in the trace are nothing else than the asymptotic values of the gauge field (2.1.5):

$$\nu = \frac{(ig)^3}{24\pi^2}\int_{S_3} d^3\theta\, n_\mu \epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\, A_\nu A_\rho A_\sigma \tag{2.1.10}$$

where n_μ is the vector orthogonal to S_3 as embedded in $\mathbb{R}_4$. Because on this configuration $G_{\mu\nu} = 0$, using the definition (2.3) together with Stokes theorem and (2.2) we get

$$\nu = \frac{(ig)^3}{24\pi^2}\int_{S_3} d^3\theta\, n_\mu \left(\frac{3i}{4g}K_\mu\right) = \frac{g^2}{32\pi^2}\int d^4x\, \partial_\mu K_\mu = \frac{g^2}{32\pi^2}\int d^4x\, G\widetilde{G}. \tag{2.1.11}$$

This result in telling us something very important. Even though in perturbation theory the operator $Q(x)$ (2.1) seemed to be completely negligible, its integrated value labels the topological sector in which the gauge field is living. These sectors are topologically distinct, in the sense that there is no homotopy[5] that maps a field configuration with one topological index to another with a different index. As we will see in a moment this has profound consequences on the physical structure of the gauge theory vacuum and ultimately on the relevance of the topological angle. Before doing this, however, we still need to find explicit instanton configurations; that are profiles satisfying (2.1.5). These can be found in terms of the natural map from S_3 to S_3:

$$U^{(1)} = \frac{x_4 + i\mathbf{x}\cdot\sigma}{r} \qquad\qquad r^2 = x_\mu x_\mu = \mathbf{x}^2 + x_4^2 \tag{2.1.12}$$

so

$$\delta\nu = \frac{3}{24\pi^2}\int_{S_3} d^3\theta\, \epsilon_{ijk}\,\mathrm{tr}\,\partial_i\alpha\partial_j U U^{-1}\partial_k U U^{-1} = -\frac{3}{24\pi^2}\int_{S_3} d^3\theta\, \epsilon_{ijk}\,\mathrm{tr}\,\partial_i\alpha\partial_j U\partial_k U^{-1} \tag{2.1.6}$$

Using integration by parts, the boundary term can be dropped since we are taking infinitesimal transformations and the remaining terms are of the form $\epsilon_{ijk}\partial_j\partial_k U$ that vanish identically.

[5] Transformations continuously connected to the identity. In this context they are also called “small gauge transformations”.

which results in[6]

$$\begin{cases} A_i & = \frac{i}{g}\frac{1}{r^2}\left[\mathbf{x}_i - \sigma_i\left(\mathbf{x}\cdot\sigma + ix_4\right)\right] \\ A_4 & = -\frac{1}{g}\frac{1}{r^2}\mathbf{x}\cdot\sigma \end{cases}. \tag{2.1.15}$$

If we plug this quantity in (2.3) we get $K_\mu = 16\,x_\mu/g^2r^4$, from which

$$\frac{g^2}{32\pi^2}\int d^4x\, G\widetilde{G} = \frac{g^2}{32\pi^2}\int_{S_3} d^3\theta\, n_\mu K_\mu = \frac{g^2}{32\pi^2}\frac{16r^2}{g^2r^5}\int_{S_3} r^3 d\Omega_3 = 1. \tag{2.1.16}$$

where we used $n_\mu = x_\mu/r$ and $\int_{S_3} d\Omega_3 = 2\pi^2$, confirming that this configuration has a non-trivial topology with winding number $\nu = 1$. Configurations with a generic winding number ν can be obtained by simply taking powers of $U^{(1)}$, $U^{(1)} \to U^{(\nu)} = (U^{(1)})^\nu$. Now, given (2.1.16) it is clear that A_μ cannot be pure gauge over the whole spacetime, but only at the boundary. A solution of the equations of motion over the whole spacetime can be found by employing the ansatz $A_\mu = f(r)(-i/g)(\partial_\mu U^{(\nu)})(U^{(\nu)})^{-1}$, plugging it in (2.1.4) and solving for $f(r)$. In this way, one finds [16]

$$A_\mu = \frac{r^2}{r^2+\rho^2}\left(-\frac{i}{g}\right)(\partial_\mu U^{(\nu)})(U^{(\nu)})^{-1}. \tag{2.1.17}$$

This solution becomes the pure gauge configuration we previously found as $r \to \infty$, and as a consequence of (classical) scale invariance depends on an arbitrary parameter ρ, whose inverse represents the size of the instanton. Given the profile of the globally defined solution, we can also compute its action. There is actually a nice trick that we can employ in order to avoid the cumbersome brute-force calculation, starting from the simple observation

$$\mathrm{tr}\,(G_{\mu\nu} \pm \widetilde{G}_{\mu\nu})^2 \geq 0. \tag{2.1.18}$$

Because $\mathrm{tr}\,G_{\mu\nu}G_{\mu\nu} = \mathrm{tr}\,\widetilde{G}_{\mu\nu}\widetilde{G}_{\mu\nu}$, it follows that

$$\mathrm{tr}\,G_{\mu\nu}G_{\mu\nu} \geq \left|\mathrm{tr}\,G_{\mu\nu}\widetilde{G}_{\mu\nu}\right|. \tag{2.1.19}$$

Therefore the action is minimized by field configuration which are self-dual, $G_{\mu\nu} = \widetilde{G}_{\mu\nu}$. It is straightforward to verify that our solution (2.1.17) is exactly self-dual[7]. Therefore, because

[6] The gauge field can also be written as

$$A_\mu = \eta_{i\mu\nu}\frac{x_\nu}{r^2}\sigma_i \tag{2.1.13}$$

where $\eta_{i\mu\nu}$ are the *t'Hooft symbols*

$$\eta_{i\mu\nu} = \epsilon_{i\mu\nu4} + \delta_{a\mu}\delta_{4\nu} - \delta_{4\mu}\delta_{a\nu}. \tag{2.1.14}$$

This rewriting illuminates the nature of the solution as map $S_3 \to S_3$. The symbols mix gauge and spacetimes indices. In particular, they are antisymmetric in $[\mu\nu]$ and therefore they live in the adjoint of $SO(4) \simeq SU(2)\times SU(2)$. But being also self-dual, $\eta_{i\mu\nu} = \epsilon_{\mu\nu\rho\sigma}\eta_{i\rho\sigma}/2$, they actually belong to the adjoint only of the first $SU(2)$ factor ($\mathbf{4}\otimes_{\mathrm{A}}\mathbf{4} \to (\mathbf{3},\mathbf{1}) + (\mathbf{1},\mathbf{3})$). Then the solution is nothing else than a map from the spacetime $SU(2)$, belonging to the Euclidean Lorentz group $SO(4)$, to the gauge group $SU(2)$, i.e. a map $S_3 \to S_3$.

[7] Actually, this could also be seen as the starting point to find instanton solutions, as originally done in [16]. Being the requirement $G_{\mu\nu} = \widetilde{G}_{\mu\nu}$ a first-order differential equation for A_μ, solving this is considerably easier than solving (2.1.4).

of (2.1.16) we already know the value of its action:

$$S_E = \int d^4x \, \frac{1}{2} \operatorname{tr} G_{\mu\nu} G_{\mu\nu} = \left| \int d^4x \, \frac{1}{4} G\widetilde{G} \right| = \frac{1}{4} \frac{32\pi^2}{g^2} |\nu| = \frac{8\pi^2}{g^2} |\nu| \, . \tag{2.1.20}$$

This result confirms the non-perturbative nature of the instanton solution, whose action scales as $1/g^2$. The action increases together with the winding number, meaning that in the path-integral the less suppressed contributions come from instantons with $\nu = 1$.

2.1.2 $\theta-$VACUA

The instanton solution found in the previous section was interpret as being a lump localized in space and time. However, this interpretation is by no means unique. There is a much more illuminating perspective which amounts to thinking about instantons as configurations which evolve in (Euclidean) *time*, as shown in figure 2.1. The boundaries I and III at $x_4^\pm \to \pm\infty$ are S_3's, connected by a cylinder II along x_4. The integral (2.1.10) becomes

$$\nu = \frac{(ig)^3}{24\pi^3} \left[\int_{\mathrm{I,III}} d^3\theta \, \epsilon_{4ijk} \operatorname{tr} A_i A_j A_k + \int_{-\infty}^{\infty} dx_4 \int_{\mathrm{II}} d^2\theta_i \, \epsilon_{i\nu\rho\sigma} \operatorname{tr} A_\nu A_\rho A_\sigma \right] . \tag{2.1.21}$$

To perform these integrals it is convenient to move to a gauge in which $A_4 = 0$, so that the

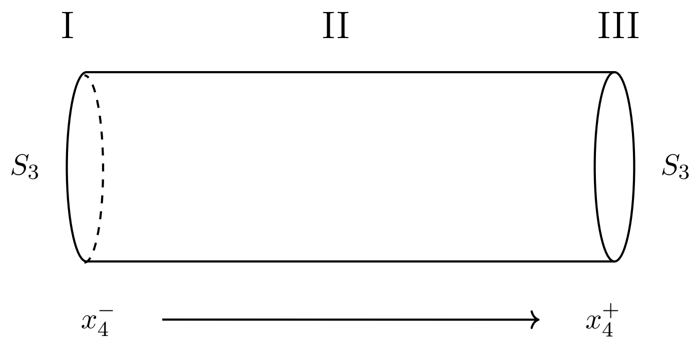

Figure 2.1: Instanton as an evolution in Euclidean time. The boundaries at $x_4^\pm \to \pm\infty$ are S_3's.

integral over II becomes null (it requires an index among $\{\nu, \rho, \sigma\}$ to be 4). Focussing on the charge $\nu = 1$ case, the required gauge transformation S must satisfy

$$S A_4 S^{-1} - \frac{i}{g} \frac{\partial S}{\partial x_4} S^{-1} = 0 \tag{2.1.22}$$

which upon using (2.1.17) becomes

$$\frac{\partial S}{\partial x_4} = S \frac{i\mathbf{x} \cdot \sigma}{r^2 + \rho^2} . \tag{2.1.23}$$

The solution to this differential equation reads [12]

$$\begin{cases} S &= e^{i\frac{\mathbf{x}\cdot\sigma}{\sqrt{\mathbf{x}^2+\rho^2}}\theta} \\ \theta &= \arctan\left(\frac{x_4}{\sqrt{\mathbf{x}^2+\rho^2}}\right)+\theta_0 \end{cases} \tag{2.1.24}$$

where θ_0 is an integration constant. Taking $\theta_0=(n+1/2)\pi$ leads to the correct instanton interpretation: indeed, for integer n this transformation sets $A_4=0$ everywhere and

$$A_i \longrightarrow \begin{cases} -\frac{i}{g}\partial_i h_n h_n^{-1} & x_4^+ \to \infty \\ -\frac{i}{g}\partial_i h_{n-1} h_{n-1}^{-1} & x_4^- \to -\infty \end{cases} \tag{2.1.25}$$

with

$$h_n = (h_1)^n = \left(e^{i\frac{x\cdot\sigma}{\sqrt{r^2+\rho^2}}\pi}\right)^n. \tag{2.1.26}$$

Both h_n and h_{n-1} are clearly elements of $SU(2)$, but are *not* homotopic. The field profiles in (2.1.25) are again standard mappings from S_3 to S_3[8] with topological number n and $n-1$, respectively, so that indeed

$$\nu = \frac{(ig)^3}{24\pi^3}\left[\int_{\mathrm{III}} d^3x\,\epsilon_{4ijk}\,\mathrm{tr}\,A_iA_jA_k - \int_{\mathrm{I}} d^3x\,\epsilon_{4ijk}\,\mathrm{tr}\,A_iA_jA_k\right] = n-(n-1)=1 \tag{2.1.27}$$

We conclude that a single instanton with charge ν can be seen as process in (Euclidean) time which changes the topological number of the asymptotic vacuum by ν units. In view of this interpreation, the picture of the gauge theory vacuum takes a new perspective: there are infinite vacuum states labelled by their winding number n, $|n\rangle$, and instantons take us from $|n\rangle$ to $|n+\nu\rangle$. Between the asymptotic vacua $G_{\mu\nu}$ is different from zero, so the field possesses a positive energy which must be overcome in order for the instanton process to happen. Classically this of course forbidden, but quantum mechanically we know that the amplitude for such tunnelling process is proportional to the Euclidean action $\sim \exp(-8\pi^2|\nu|/g^2)$. Indeed, denoting $|n\rangle$ the vacuum with topological number n, in the Euclidean path-integral representation of field theory

$$\langle n|\, e^{-HT}\, |m\rangle_J = \int DA_{\nu=n-m} e^{-S_E+\int d^4x\, J_\mu A_\mu} \tag{2.1.28}$$

where H is the Hamiltonian and J_μ the source associated to the gauge field with instanton number ν. Since $|n\rangle$ are not the true eigenstates of the Hamiltonian, the real vacuum should be given by a linear combination of them [17, 18]:

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{-in\theta}\,|n\rangle \tag{2.1.29}$$

[8]Within this mapping the spacetime S_3 consists of the whole $\mathbb{R}_3$ positioned at $x_4^\pm \to \pm\infty$, with the points at space infinity identified by the complex exponential. This parametrization is different from the one in (2.1.12), where the starting S_3 was identified by the condition $x_\mu x_\mu = r^2$ and the integral was performed on the sphere angles. Here the integration region is $\mathbf{x}_i \in (-\infty,\infty)$.

where θ is a real parameter which defines the *θ-vacua*. We can see that this is the correct combination by considering a gauge transformation g_m changing the topological number by m units[9], $g_m |n\rangle = |n+m\rangle$: since H is gauge invariant the vacuum must be an eigenstate of g_m, and indeed

$$g_m |\theta\rangle = \sum_n e^{-in\theta} g_m |n\rangle = \sum_n e^{-in\theta} |n+m\rangle = e^{im\theta} |\theta\rangle \,. \tag{2.1.30}$$

Therefore, in analogy with what happens in quantum mechanics with Bloch waves, θ defines a "pseudo-momentum" which labels physically inequivalent sectors that do not communicate. In particular, the vacuum-to-vacuum amplitude for two different θ-vacua reads

$$\begin{aligned}
\langle\theta'| e^{-HT} |\theta\rangle_J &= \sum_{n,m} e^{+im\theta'} e^{-in\theta} \langle m| e^{-HT} |n\rangle \\
&= \sum_{n,m} e^{i(m-n)\theta} e^{im(\theta-\theta')} \int DA_{n-m} e^{-S_E + \int d^4x\, J_\mu A_\mu} \\
&= 2\pi\delta(\theta-\theta') \sum_\nu e^{i\theta\nu} \int DA_\nu e^{-S_E + \int d^4x\, J_\mu A_\mu} \\
&= 2\pi\delta(\theta-\theta') \sum_\nu \int DA_\nu e^{-S'_E + \int d^4x\, J_\mu A_\mu}
\end{aligned} \tag{2.1.31}$$

where in the third line we have relabelled $\nu = m-n$ and exploited the identity $\sum_m e^{im(\theta-\theta')} = 2\pi\delta(\theta-\theta')$. The new action S'_E appearing in the last line is given by

$$S'_E = S_E - i\theta\nu = S_E - i\frac{g^2}{32\pi^2}\int d^4x\, G\widetilde{G} \tag{2.1.32}$$

where we have used (2.1.11) to represent the winding number ν in terms of the operator $Q(x)$. Thus, in this semiclassical approximation the complex structure of the gauge theory vacuum gives rise to an effective lagrangian term which is nothing else than the topological operator $Q(x)$ we started with, (2.1). The topological angle θ becomes now more than just a (apparently unphysical) coupling, but a fundamental selection rule labelling completely disjoint vacua of the theory as coded in $\delta(\theta-\theta')$ in (2.1.31), which holds up to arbitrary insertions of gauge-invariant operators. Of course, being just a semiclassical analysis, we expect this naïve picture to break down when quantum effects become important. Indeed, as we will see later, in QCD the topological angle θ becomes a true observable parameter rather than a label for the vacuum. Finally, for simplicity our analysis has focused on a $SU(2)$ gauge theory, but it actually generalises to any non-Abelian group (including, of course, QCD). This is guaranteed by Bott's theorem [19], which states that any continuous mapping $S_3 \to G$ can be continuously deformed into a mapping to any $SU(2)$ subgroup of G. Therefore as long that $G \supset SU(2)$, which is true for any non-Abelian compact Lie group, our analysis goes through completely unchanged. For Abelian gauge theories, instead, $\Pi_3(S_1) = 1$ and there is no analog of the winding number.

[9] As a clarification, we stress that gauge transformations changing the topological number of the vacuum do exist. These are not continuously connected to the identity and called "large gauge transformations". However, according to the previous instanton interpretation, these shift both the asymptotic winding numbers by the same amount and are thus fundamentally different from instantons.

Vacuum Energy

The instanton action found in section 2.1.1 can be used to determine the dependence of the vacuum energy on the topological angle θ. Consider the relation (2.1.31) that we just found,

$$\langle\theta'|\, e^{-HT}\,|\theta\rangle = 2\pi\delta(\theta-\theta')\sum_\nu e^{-i\theta\nu}\int DA_\nu e^{-S_E}. \tag{2.1.33}$$

in which now we are omitting the source J_μ, since it will not play any role. In the $T\to\infty$ limit the exponent in the left-hand side of the equation becomes the energy of the vacuum, since the contributions from the rest of the Hamiltonian eigenstates are exponentially suppressed:

$$\langle\theta|\, e^{-HT}\,|\theta\rangle = \sum_n e^{-E_nT}\,\langle\theta|n\rangle\,\langle n|\theta\rangle \xrightarrow{T\to\infty} e^{-E(\theta)T}|\,\langle\theta|\theta\rangle\,|^2 \equiv e^{-\mathcal{E}(\theta)V} \tag{2.1.34}$$

where in the last we have explicitly written a volume factor to mod out the energy density $\mathcal{E}(\theta)$. Therefore, in this approximation the vacuum energy density is given by

$$\mathcal{E}(\theta) = -\frac{1}{V}\log\left(\sum_\nu e^{-i\theta\nu}\int DA_\nu e^{-S_E}\right) = -\frac{1}{V}\log\mathcal{Z}(\theta). \tag{2.1.35}$$

The leading contribution in the sum over instantons comes from the one instanton and one anti-instanton configurations with $|\nu|=1$, since these have the less suppressed action $S_E = 8\pi^2/g^2$. Thus

$$\mathcal{E}(\theta) = -\frac{1}{V}\log\mathcal{Z}_0 - \frac{1}{V}\log\left(1 + K_{-1}e^{i\theta}e^{-\frac{8\pi^2}{g^2}} + K_{+1}e^{-i\theta}e^{-\frac{8\pi^2}{g^2}} + \dots\right) \tag{2.1.36}$$

where, at leading order in g^2, the factors $K_{\pm1}$ are given by integrating the one-loop fluctuation of the gauge field around the instanton/anti-instanton configuration. These factors always include a power of the volume, $K_{\pm1} = VK$, coming from translational invariance together with the fact that we must integrate also over the instanton position. We also are assuming that $K_{+1} = K_{-1}$, a result that is found in explicit calculations [13]. Therefore[10]

$$\mathcal{E}(\theta) - \mathcal{E}(0) \approx -2Ke^{-\frac{8\pi^2}{g^2}}\cos\theta. \tag{2.1.37}$$

This result is particularly interesting because it highlights two important features. First of all, the energy is minimised at the CP-conserving value $\theta = 0 \bmod 2\pi$. This is not a coincidence, but a general property of vector-like gauge theories (including pure Yang-Mills theories) as proved in an historical work by Vafa and Witten [21]. Making a nod to next chapter, if we imagine replacing θ by a dynamical field then its potential as generated by instantons naturally relaxes its vacuum expectation value (vev) to a CP-conserving value. This will basically set the stage for the axion mechanism as a dynamical solution to the Strong CP problem. Second, we can estimate the scale at which the instantons become relevant by substituting to g^2 its

[10] This result can be obtained in a more refined way using the Dilute Instanton Gas Approximation (DIGA), see [10, 13, 20].

1-loop value $g^2(\mu) = 8\pi^2/\beta_0 \log(\mu/\Lambda)$, where β_0 is the coefficient of the 1-loop $\beta-$function and Λ is the confinement scale. As a typical scale for μ we can take the inverse of instanton size ρ, so that

$$\mathcal{E}(\theta) - \mathcal{E}(0) \approx -2Ke^{-\beta_0 \log(1/\rho\Lambda)} \cos\theta = -2K(\rho\Lambda)^{\beta_0} \cos\theta. \tag{2.1.38}$$

Since in a proper path-integral treatment we should actually integrate over ρ, we see that the leading contribution comes from instantons of size $\rho \gtrsim 1/\Lambda$. Unfortunately, this is exactly the regime in which our semiclassical analysis is no more valid. The lesson is that instantons should be regarded as useful objects to gain a qualitative understanding of the non-perturbative effects which are going on (as we have seen), but should not be used for quantitative predictions. Indeed, in the regime of validity in which instanton calculus is valid, ordinary perturbation theory is much more predictive.

2.1.3 QCD

In this section we briefly elaborate on some consequences of what we discovered in the previous section, particularly in relation to QCD. More detailed treatments can be found in the references provided in the text.

The $U(1)_\mathrm{A}$ Problem

The $U(1)_\mathrm{A}$ problem was an apparent inconsistency in the Nambu-Goldstone bosons' spectrum of QCD. At energies far below the electroweak scale (but above the QCD confinement scale), the colored sector of the SM can be described in terms of a very simple lagrangian:

$$\mathcal{L}_\mathrm{C} = -\frac{1}{4}GG + \theta\frac{g^2}{32\pi^2}\, G\widetilde{G} + \bar{q}_{L,i} i\not{D} q_{L,i} + \bar{q}_{R,i} i\not{D} q_{R,i} - M_{ij}\bar{q}_{L,i}q_{R,j} + \text{h.c.} \tag{2.1.39}$$

where $q_{L,i}, q_{R,i}$ are the chiral components of the lightest quarks ($i = \{u,d,s\}, N_f = 3$), M is the (complex) mass matrix and θ is the QCD topological angle. This lagrangian features a $U(3)_\mathrm{L} \times U(3)_\mathrm{R} = SU(3)_\mathrm{V} \times SU(3)_\mathrm{A} \times U(1)_\mathrm{V} \times U(1)_\mathrm{A}$ flavour symmetry group, classically broken only by the presence of the masses[11]. The mass matrix M can always be made diagonal and real ($M_{ij} = m_i\delta_{ij}$) thanks to a $U(3)_\mathrm{L} \times U(3)_\mathrm{R}$ rotation, and generically leaves intact only the $U(1)_\mathrm{V}$ vectorial subgroup. In particular, the current non-conservation for the Abelian $U(1)_\mathrm{A}$ symmetry, generated by $q_i \to \exp(i\alpha\gamma^5)q_i$, reads

$$\begin{aligned} \partial_\mu J_5^\mu &= 2iM_{ij}\,\bar{q}_i\gamma^5 q_j \\ J_5^\mu &= \sum_i \bar{q}_i\gamma^\mu\gamma^5 q_i \end{aligned} \qquad \text{(classically).} \tag{2.1.40}$$

The QCD dynamics spontaneously breaks the axial symmetries by developing a vev for the fermion bilinear:

$$\langle \bar{q}_{L,i} q_{R,j} \rangle = f_\pi^2 \Lambda_\mathrm{C}\, \delta_{ij} \tag{2.1.41}$$

[11] And by the gauging of $U(1)_\mathrm{EM}$, that for the moment we are neglecting.

which leaves invariant only the $SU(3)_{\rm V} \times U(1)_{\rm V}$ subgroup, as predicted by Vafa and Witten [22]. The spontaneous breaking of the axial symmetries delivers $2N_f^2 - N_f^2 = 9$ Nambu-Goldstone modes, which are described as oscillations around the bilinear vev in an effective lagrangian that goes by the name of *chiral lagrangian* [23, 24]. Following the CCWZ construction [25, 26], the NGBs Π^a are packed in a matrix U defined in terms of the broken generators $t^{\bar{a}}$ of $SU(3)_{\rm A} \times U(1)_{\rm A}$ as

$$U = e^{i\Pi^{\bar{a}} t^{\bar{a}}/f_\pi}$$

$$\Pi^{\bar{a}} t^{\bar{a}} = \begin{pmatrix} \pi^0 + \frac{\eta_8}{\sqrt{3}} + \sqrt{\frac{2}{3}}\eta_1 & \sqrt{2}\pi^+ & \sqrt{2}K^+ \\ \sqrt{2}\pi^- & -\pi^0 + \frac{\eta_8}{\sqrt{3}} + \sqrt{\frac{2}{3}}\eta_1 & \sqrt{2}K^0 \\ \sqrt{2}K^- & \sqrt{2}K^0 & -\frac{2}{\sqrt{3}}\eta_8 + \sqrt{\frac{2}{3}}\eta_1 \end{pmatrix} \tag{2.1.42}$$

where the fields appearing in $\Pi \equiv \Pi^{\bar{a}} t^{\bar{a}}$ are the pions, the kaons and the η-mesons. In particular, η_1 is the NGB of the $U(1)_{\rm A}$ symmetry. This matrix transforms under chiral symmetries as $U \to LUR^\dagger$, so treating the quark masses as small compared to the chiral lagrangian scale, the leading order effective lagrangian consistent with chiral symmetries reads

$$\mathcal{L}_{\chi\rm PT}^{\text{no-anomaly}} = \frac{f_\pi^2}{4} \operatorname{tr} \partial_\mu U \partial^\mu U^\dagger + \frac{f_\pi^2}{4} B_0 \operatorname{tr} \left(U^\dagger M + U M^\dagger \right) + \text{h.o.}\,. \tag{2.1.43}$$

From this lagrangian we can read off the mesons' masses and, for instance, derive the famous Gell-Mann-Okubo relation $4m_K^2 = 3m_\eta^2 + m_\pi^2$ [27, 28] (in the limit $m_u = m_d$), where η is a combination of η_8 and η_1. But this also leads to a mass for the orthogonal combination, η', comparable to that of the other pseudo-NGB. In particular, Weinberg [29] was able to obtain the approximate bound

$$m_{\eta'} \lesssim \sqrt{3} m_\pi. \tag{2.1.44}$$

Experimentally, no such particle is found. While π, K and η all have masses compatible with (2.1.43), the candidate for η' is much more massive: $m_{\eta'} \simeq 958$ MeV [1], which compared to $m_\pi \simeq 134$ MeV strongly violates the bound (2.1.44). This puzzle defines what was known as the $U(1)_A$ *problem*, namely no NGB is there for the spontaneously broken $U(1)_{\rm A}$ symmetry. It is worth mentioning that also a second $U(1)_{\rm A}$ problem was present, related to the decay $\eta' \to 3\pi$. A chiral lagrangian calculation including $U(1)_{\rm EM}$ effects and exploiting the relation (2.1.40) shows that the amplitude for this process should vanish [30]. Experimentally, however, the width of this decay channel is consistently not null [1]. Thus we are led to a second paradox, again related to the $U(1)_{\rm A}$ symmetry.

The solution of the $U(1)_{\rm A}$ problem(s) lies in the fact that the current non-conservation equation (2.1.40) is actually incorrect. It is altered by quantum effects, which lead to the so called ABJ anomaly relation [31]:

$$\partial_\mu J_5^\mu = 2iM_{ij}\, \bar{q}_i \gamma^5 q_j - 2N_f \frac{g^2}{32\pi^2} G\widetilde{G}. \tag{2.1.45}$$

Therefore, the flavour rotation needed to make M diagonal and real actually shifts θ, meaning that the correct $U(1)_{\mathrm{A}}$ invariant effective topological angle is given by

$$\bar{\theta} = \theta - \arg\det M. \tag{2.1.46}$$

The additional term in (2.1.45) proportional to the topological density can be interpreted as due to the non-invariance of the path-integral measure under $U(1)_{\mathrm{A}}$ transformations of the fermions [32]. Note, however, that we can in any case define another axial current by exploiting (2.2):

$$\bar{J}_5^\mu = J_5^\mu + 2N_f \frac{g^2}{32\pi^2} K^\mu \tag{2.1.47}$$

meaning that, apparently, we still have a conserved axial current apart from the explicit breaking due to the fermion masses. Thus we still expect a mass for the η' comparable to that of the other mesons: equation (2.1.45) alone cannot be the end of the story. The key missing piece is that the current $\bar{J}_5^\mu$ defined in (2.1.47) is *not* gauge invariant, because the Chern-Simons current, unlike J_5^μ, is not. If we consider only trivial gauge backgrounds, this would not be an issue and all the story we just outlined would go through unchanged. However we known that there are non-trivial background, given by instantons. These violate the conservation of the axial charge (in the limit $m_i = 0$) by $2N_f\nu$ units [33]:

$$\begin{gathered} \bar{Q}_5 = \int_{S_3} d^3\theta \, n_\mu \bar{J}_5^\mu = Q^5 + 2N_f \frac{g^2}{32\pi^2} \int_{S_3} d^3\theta \, n_\mu K^\mu \\ \downarrow \\ \Delta\bar{Q}_5 = 2N_f\nu. \end{gathered} \tag{2.1.48}$$

Therefore, the chiral symmetry defined by the current (2.1.47) is *not* really a symmetry of the theory. It is explicitly broken by the non-perturbative instantons. For this reason, we do not expect any Nambu-Goldstone mode coming from the further breaking due to the fermion condensate: the mass of the η' is dominated by instantons effects, which become relevant at the scale at which the group confines. Thus we expect $m_{\eta'} \sim \Lambda_{\mathrm{C}}$, which is now perfectly compatible with its measured value. The $U(1)_{\mathrm{A}}$ problem is solved.

The connection between instantons and the $\theta-$vacuum is actually more interesting in the presence of exactly massless fermions. Suppose there are N_A of them, and as in section 2.1.2 call g_ν the gauge transformation changing the topological number of the vacuum by ν units: $g_\nu \left|n\right\rangle = \left|n+\nu\right\rangle$. Then $[\bar{Q}_5, g_\nu] = 2\nu N_A g_\nu$, meaning that g_ν acts like a raising operator and so

$$\bar{Q}_5 \left|n\right\rangle = 2nN_A \left|n\right\rangle. \tag{2.1.49}$$

Thus, states with winding number n have a definite chirality. In particular, under a chiral rotation by an angle α,

$$e^{-i\alpha\bar{Q}_5} \left|\bar{\theta}\right\rangle = \sum_n e^{-in\bar{\theta} - i\alpha 2nN_A} \left|n\right\rangle = \left|\bar{\theta} + 2\alpha N_A\right\rangle \equiv \left|\bar{\theta}'\right\rangle. \tag{2.1.50}$$

Since $[H, \bar{Q}_5] = 0$, this means that the theory is invariant under the shift $\bar{\theta} \to \bar{\theta}'$. In other words, the parameter $\bar{\theta}$ becomes unphysical in the presence of massless fermions. Again with a nod to chapter 3, this is essentially the core of the massless up-quark proposal as a solution to the Strong CP problem.

It is also possible to show that associated to each instanton of charge ν there are chiral zero-modes of the fermion Dirac operator such that the net chirality is exactly ν [34]. This result is encoded in the Atiyah-Singer index theorem [35], which formalizes our previous qualitative argument. These zero-modes suppress topological transitions, but the concept of $\theta-$vacuum is still required on the basis of the cluster decomposition principle [17]. They also imply a non-vanishing vev for operators with non-zero chirality, hinting towards a proof of chiral symmetry breaking.

Strong Coupling: $\eta \to \pi\pi$ and Neutron EDM

The last topics we will discuss before moving to the next section are two historical observables which experimentally provide the most constraining bounds on the QCD topological angle: the CP-violating η-meson decay $\eta \to \pi\pi$ and the neutron electric dipole moment. These processes appear in the strong coupling regime and therefore it is not possible to perform a first-principle calculation starting from the lagrangian (2.1.39). We can, however, estimate the size of these effects using chiral perturbation theory in the large-N limit, where the effects of the chiral anomaly are suppressed and therefore we can treat the η' within this framework [36]. Although the discussion of the CP-violating η decay will be self-contained, the one of the neutron EDM will be just qualitative since its derivation within chiral perturbation theory is quite involved and eventually ill-defined. Our exposition will follow [37], and slightly differs from the historical ones [38, 39] which employ current algebra techniques.

1. $\eta \to \pi\pi$

The starting point is the CCWZ construction of the chiral lagrangian. We introduce the two coset representatives $\xi_{L,R} = \xi_{L,R}(\Pi)$ of $U(3)_L \times U(3)_R$, where Π are the Nambu-Goldstone fields, transforming under chiral rotations $g = (L, R)$ as

$$\xi_L \to L\, \xi_L\, h^\dagger(g, \Pi), \qquad \xi_R \to R\, \xi_R\, h^\dagger(g, \Pi). \tag{2.1.51}$$

The NGBs matrix given in (2.1.42) corresponds to $U = \xi_L \xi_R{}^\dagger$. The leading order mesonic lagrangian in the small momentum, small M and large-N expansion reads

$$\begin{aligned}\mathcal{L}^{\rm LO}_{\chi\rm PT} = \frac{f_\pi^2}{4}\,{\rm tr}\,\partial_\mu U \partial^\mu U^\dagger + \frac{f_\pi^2}{4} B_0\,{\rm tr}\left(U^\dagger M + U M^\dagger\right)& \\ - \frac{f_\pi^2}{4}\frac{a}{N}\left[\frac{i}{2}\left(\log\det U - \log\det U^\dagger\right) - \bar{\theta}\right]^2 + \text{h.o.}\,.&\end{aligned} \tag{2.1.52}$$

The difference with respect to (2.1.43) is given by the term in last line, which includes the effects of the anomaly. Its form is fixed by the spurious transformation property of the topological angle, $\bar{\theta} \to \bar{\theta} + 2\alpha N_f$, under anomalous chiral rotations $L = R^\dagger = e^{i\alpha}$, together with spurious P invariance (P: $U \to U^\dagger$, $\bar{\theta} \to -\bar{\theta}$) and large-$N$ counting. To make use of this lagrangian we must first minimise the potential of the NGBs, which because of the last term is misaligned from $U_0 = \langle U \rangle \sim \mathbb{1}$. In particular, since M is diagonal the minimum must be of the form $U_0 = \mathrm{diag}(e^{i\phi_u}, e^{i\phi_d}, e^{i\phi_s})$ and for convenience let us define Hermitian matrices χ, H such that $B_0 M^\dagger U_0 = \chi + iH$. The minimisation of the potential leads to the Dashen-Nuyts conditions [40, 41]

$$B_0 m_i \sin\phi_i = -\frac{a}{N}\left(\bar\theta + \sum_j \phi_j\right) \equiv -\frac{a}{N}\bar\theta' \tag{2.1.53}$$

which clearly imply $H = -\bar\theta'(a/N)\mathbb{1}$. Redefining $U \to \sqrt{U_0}U\sqrt{U_0}$, the lagrangian becomes

$$\begin{aligned}\mathcal{L}^{\mathrm{LO}}_{\chi\mathrm{PT}} = \frac{f_\pi^2}{4}\,\mathrm{tr}\,\partial_\mu U\partial^\mu U^\dagger + \frac{f_\pi^2}{4}\Bigg\{ \mathrm{tr}\left(U^\dagger\chi + \chi^\dagger U\right) - \frac{a}{N}\left[\bar\theta'^2 - \frac{1}{4}\log^2\left(\frac{\det U}{\det U^\dagger}\right)\right] \\ -i\bar\theta'\frac{a}{N}\left[\mathrm{tr}\left(U - U^\dagger\right) - \log\left(\frac{\det U}{\det U^\dagger}\right)\right]\Bigg\} + \mathrm{h.o.}\,.\end{aligned} \tag{2.1.54}$$

where now $\langle U \rangle = \mathbb{1}$, and so we can expand the fields Π around the origin. There a couple of interesting things to notice. First, in the chiral limit ($\chi = 0$), the η_1 meson still gets a mass $m^2_{\eta_1} = 3a/N$ from the second term in the braces. This originates from the anomaly, as we already discussed in the previous section. Second, the term in the last line generates CP-violating interactions between the mesons. Their strength is determined by $\bar\theta'$, which however at this point cannot be directly linked to $\bar\theta$ since it involves additional phases coming from U_0. To understand the dependence on $\bar\theta$ we should find the physical masses of the mesons to fix the phases. This is easily done in the isospin limit ($m_u = m_d \equiv m$), where the minimisation implies $\phi_u = \phi_u \equiv \phi$. In this limit we get

$$\begin{aligned} m_\pi^2 &= B_0 m\cos\phi \\ m_K^2 &= B_0(m\cos\phi + m_s\cos\phi_s)/2 \end{aligned} \tag{2.1.55}$$

and a mixing between $\eta_8 - \eta_1$

$$\mathcal{M}^2_{\eta_8-\eta_1} = \frac{B_0}{3}\begin{pmatrix} m\cos\phi + 2m_s\cos\phi_s & \sqrt{2}\,(m\cos\phi - m_s\cos\phi_s) \\ \sqrt{2}\,(m\cos\phi - m_s\cos\phi_s) & m_s\cos\phi_s + 2m\cos\phi + \frac{9}{B_0}\frac{a}{N}\end{pmatrix} \tag{2.1.56}$$

with eigenstates

$$\begin{aligned} \begin{pmatrix}\eta \\ \eta'\end{pmatrix} &= \begin{pmatrix}\cos\theta_{\mathrm{P}} & -\sin\theta_{\mathrm{P}} \\ \sin\theta_{\mathrm{P}} & \cos\theta_{\mathrm{P}}\end{pmatrix}\begin{pmatrix}\eta_8 \\ \eta_1\end{pmatrix} \\ \tan 2\theta_{\mathrm{P}} &= \frac{2\sqrt{2}(m\cos\phi - m_s\cos\phi_s)}{m\cos\phi - m_s\cos\phi_s + \frac{9}{B_0}\frac{a}{N}} = \frac{4\sqrt{2}(m_K^2 - m_\pi^2)}{2(m_K^2 - m_\pi^2) - 9a/N}. \end{aligned} \tag{2.1.57}$$

The physical masses satisfy $m_\eta^2 + m_{\eta'}^2 - 2m_K^2 = 3a/N$, so using the latest mesons data [1] and setting $N = 3$ we obtain $\sqrt{a} \simeq 850$ MeV and $\theta_{\mathrm{P}} \simeq -19°$. At this point we have all that we need to solve the system (2.1.53) and express $\bar\theta'$ as a function of $\bar\theta$ and the experimental meson masses. Unfortunately this leads to an overly complicated analytical expression. However, since for $\bar{\theta} = 0$ the absolute minimum occurs at $\phi_i = 0$ [42,43], we can approximate (2.1.53) to the lowest order in the angles ϕ_i to get

$$\frac{a}{N}\bar{\theta}' \simeq \frac{\bar{\theta}}{\frac{N}{a} + \sum_i \frac{1}{B_0 m_i}} \approx \frac{1}{2} m_\pi^2 \bar{\theta} \tag{2.1.58}$$

where in the last equality we only kept the contribution $2(B_0 m)^{-1} \simeq 2m_\pi^{-2}$ at the denominator, which is the dominant one according to (2.1.55) and to the numerical value of a. In this way we fixed the $\bar{\theta}$ dependence of the CP-violating couplings in the lagrangian (2.1.54), which we can now use to compute the width of $\eta \to \pi\pi$. In particular, the last line of (2.1.54) contains

$$\begin{aligned} \mathcal{L}_{\chi\mathrm{PT}}^{\mathrm{LO}} \supset & \frac{a}{N}\bar{\theta}' \frac{1}{\sqrt{3} f_\pi} \left(\sqrt{2}\sin\theta_{\mathrm{P}} - \cos\theta_{\mathrm{P}}\right) \eta\, \pi^+ \pi^- \\ \simeq & \frac{m_\pi^2}{2\sqrt{3} f_\pi} \bar{\theta} \left(\sqrt{2}\sin\theta_{\mathrm{P}} - \cos\theta_{\mathrm{P}}\right) \eta\, \pi^+ \pi^- \end{aligned} \tag{2.1.59}$$

from which we can immediately compute the decay rate [37–39]:

$$\Gamma(\eta \to \pi\pi) = \frac{m_\pi^4}{192\pi f_\pi^2 m_\eta} \bar{\theta}^2 \left(\sqrt{2}\sin\theta_{\mathrm{P}} - \cos\theta_{\mathrm{P}}\right)^2 \left(1 - \frac{4m_\pi^2}{m_\eta^2}\right)^{1/2} . \tag{2.1.60}$$

Using the latest constraints on the branching ratio of this process [1], we can finally set the bound

$$|\bar{\theta}| \lesssim 1.7 \times 10^{-4} \qquad\qquad (\eta \to \pi\pi)\,. \tag{2.1.61}$$

This bound is slightly stronger than the ones from [37–39] due to improvements in the measurements of the η widths. Hearteningly, though, the theoretical predictions do coincide[12].

2. NEUTRON EDM

In non-relativistic quantum mechanics, the neutron electric dipole moment is encoded in the Hamiltonian as

$$H \supset -d_n\, \mathbf{E} \cdot \mathbf{S}_n \tag{2.1.62}$$

where $\mathbf{E}$ stands for the electric field and $\mathbf{S}_n$ the spin operator of the neutron. In relativistic quantum field field theory, the analogous observable is induced by the lagrangian operator

$$\mathcal{L} \supset -\frac{i}{2} d_n\, \bar{n}\sigma^{\mu\nu}\gamma^5 n F_{\mu\nu} \tag{2.1.63}$$

written in terms of the electromagnetic field strength $F_{\mu\nu}$ and the neutron Dirac field $n(x)$ (see e.g. [44]). The $\bar{\theta}$-dependence of d_n is obtained, therefore, by adding the baryon fields to

[12] There is a mismatch of a factor of 2 with respect to [39] due to the different normalization of f_π ($f_\pi \to f_\pi/\sqrt{2}$), while the expressions for the mixing $\eta_1 - \eta_8$ coincide at leading order.

the chiral lagrangian and performing an analysis similar to what we have done for $\eta \to \pi\pi$. Unfortunately, in this case the procedure is less straightforward and much more ill-defined for multiple reasons. First of all, baryons are not (psedo-)NGBs and therefore acquire a mass of order the confining scale Λ_C. This means that to have a consistent description in terms of an effective theory we should treat them as heavy matter fields with three-momentum at most of order $\sim m_\pi$ (in an expansion known as heavy baryon χPT [45]). Thus we cannot consider creation/annihilation processes, which however are fortunately not required to discuss observables in which we have just one baryon in the initial and final states, i.e. EDMs. Second, the purely mesonic chiral lagrangian admits topological configurations, the Skyrmions, which seem to exactly match the baryons' spectrum [46, 47]. Apparently, thus, explicitly adding baryons to the chiral lagrangian results in some kind of double counting. Third, as we will see in a moment, the operator (2.1.63) is generated at 1-loop by the exchange of virtual mesons. The regularisation of the associated divergences requires the introduction of counter-terms which are incalculable, thus making the exact computation somehow pointless. For all these reasons we will just sketch the procedure needed in order to extract d_n from chiral perturbation theory; the complete calculation can be found in [37]. It is amazing, however, that just by dimensional analysis and symmetry arguments we can obtain an estimate for d_n which is remarkably close to the actual result. We know, from section 2.1.2, that $\bar\theta$ becomes unphysical if any of the quarks becomes massless. Therefore d_n should be suppressed by at least one power of the lightest quark mass, in this case m_u. Since $[d_n] = -1$, we must account for its mass dimension with the only other relevant scale at our disposal, m_n (the EDM is measured at momentum transfer $q^2 \simeq 0$). Finally d_n should be proportional to the electromagnetic coupling e, being a dipole, and of course to the only CP-odd parameter $\bar\theta$. Therefore, our estimate can be expressed as

$$d_n \sim c_n e \, \frac{m_u}{m_n^2} \, \bar\theta \tag{2.1.64}$$

where c_n is a constant that we cannot establish at this point. Plugging in the most recent data [1] and taking $c_n \sim \mathcal{O}(1)$ we get $d_n \sim 3 \times 10^{-16}$ cm. The most recent search [5] sets $|d_n|/e < 1.8 \times 10^{-26}$ cm at 90% C.L, translating in

$$|\bar\theta| \lesssim 10^{-10} \qquad\qquad (\text{neutron EDM})\,. \tag{2.1.65}$$

This simple estimate gives a bound much stronger than (2.1.61) and is surprisingly close to exact calculation result, as we will see in a moment.

The procedure to obtain d_n within the chiral lagrangian framework requires some additional technicalities with respect to the simple $\eta \to \pi\pi$ calculation. First, we must introduce the baryon octet

$$\begin{gathered} B(x) = \begin{pmatrix} \frac{\Sigma_0}{\sqrt{2}} + \frac{\Lambda_0}{\sqrt{6}} & \Sigma^+ & p \\ \Sigma^- & -\frac{\Sigma_0}{\sqrt{2}} + \frac{\Lambda_0}{\sqrt{6}} & n \\ \Xi^- & \Xi^0 & -\frac{2}{\sqrt{6}}\Lambda^0 \end{pmatrix} \\ B \xrightarrow{g} h(g,\Pi)\, B\, h^\dagger(g,\Pi). \end{gathered} \tag{2.1.66}$$

The kinetic terms of the baryons and now also of the mesons,

$$\mathcal{L}_{\text{kin}} = \frac{f_\pi^2}{4} \operatorname{tr} D_\mu U D^\mu U^\dagger + \operatorname{tr} \bar{B} i \gamma^\mu D_\mu B, \tag{2.1.67}$$

are given in terms of a covariant derivative accounting for *local* G transformations

$$\begin{aligned} D_\mu U &= \partial_\mu U + i l_\mu U - i U r_\mu \\ D_\mu B &= \partial_\mu B + [\Gamma_\mu, B] \end{aligned} \tag{2.1.68}$$

such that $D_\mu U \xrightarrow{g} L D_\mu U R^\dagger$ and $D_\mu B \xrightarrow{g} h D_\mu B h^\dagger$. The quantities appearing in (2.1.68) are the left- and right-handed currents, transforming as

$$\begin{aligned} l_\mu &\xrightarrow{g} L l_\mu L^\dagger + i \partial_\mu L L^\dagger \\ r_\mu &\xrightarrow{g} R r_\mu R^\dagger + i \partial_\mu R R^\dagger \end{aligned} \tag{2.1.69}$$

and the connection Γ_μ,

$$\Gamma_\mu = \frac{1}{2} \left[\xi_R{}^\dagger \left(\partial_\mu + i r_\mu \right) \xi_R + \xi_L{}^\dagger \left(\partial_\mu + i l_\mu \right) \xi_L \right]. \tag{2.1.70}$$

In the simple but relevant case of $U(1)_{\text{EM}}$ the currents read $l_\mu = r_\mu = eQA_\mu$, with $Q = \operatorname{diag}(2/3, -1/3, -1/3)$, and give rise to the electromagnetic interactions of the mesons and the baryons. The additional chirally-invariant operators relevant for the neutron EDM are

$$\begin{aligned} \mathcal{L} \supset &- \frac{1}{2} D \operatorname{tr} \bar{B} \gamma^\mu \gamma^5 \{\xi_\mu, B\} - \frac{1}{2} F \operatorname{tr} \bar{B} \gamma^\mu \gamma^5 [\xi_\mu, B] \\ &- b_1 \operatorname{tr} \bar{B} \tilde{\chi}_+ B - b_2 \operatorname{tr} \bar{B} B \tilde{\chi}_+ \end{aligned} \tag{2.1.71}$$

where ξ_μ and $\tilde{\chi}_\pm$ are defined in terms of $\xi_{L,R}$, the spurion $\tilde{\chi} = B_0 M$ and the currents as

$$\begin{cases} \xi_\mu = i \left[\xi_R{}^\dagger \left(\partial_\mu + i r_\mu \right) \xi_R - \xi_L{}^\dagger \left(\partial_\mu + i l_\mu \right) \xi_L \right] \\ \tilde{\chi}_\pm = \xi_R^\dagger \tilde{\chi}^\dagger \xi_L \pm \xi_L^\dagger \tilde{\chi} \xi_R \end{cases} . \tag{2.1.72}$$

$$(\xi_\mu, \tilde{\chi}_\pm) \xrightarrow{g} h(g, \Pi) \; (\xi_\mu, \tilde{\chi}_\pm) \; h^\dagger(g, \Pi).$$

The minimisation of the meson potential (2.1.52) required the redefinition $U \to \sqrt{U_0} U \sqrt{U_0}$, which can be seen as redefining the coset representatives as $\xi_{L,R} \to \sqrt{U_0} \xi_{L,R}$. Since U_0 and $\tilde{\chi}$ are both diagonal, and exploiting the relation $\tilde{\chi}^\dagger U_0 = \chi + iH = \chi - i(a/N)\bar{\theta}' \mathbb{1}$, this amounts to shifting

$$\tilde{\chi}_\pm \to \tilde{\chi}_\pm - i \frac{a}{N} \bar{\theta}' \left(U^\dagger \mp U \right) \tag{2.1.73}$$

which applied to (2.1.71) leads to CP-violating interactions between the mesons and the baryons

$$\mathcal{L}_{\not{C}\not{P}} \supset -i b_1 \frac{a}{N} \bar{\theta}' \operatorname{tr} \bar{B} (U - U^\dagger) B - i b_2 \frac{a}{N} \bar{\theta}' \operatorname{tr} \bar{B} B (U - U^\dagger). \tag{2.1.74}$$

The counterpart involving χ induces a splitting in the baryon masses which completely fixes b_1, b_2. The redefinition, instead, does not affect the CP-even F and D terms, whose couplings are fixed by semileptonic hyperons' decays. At this point we have, in principle, all the necessary ingredients to perform the 1-loop calculation leading to the effective operator (2.1.63). A sample diagram is given in figure 2.2, where at leading order only one CP-odd insertion is required, the other being the F/D coupling. The diagram is UV divergent and needs to be regularised appropriately in terms of (incalculable) counterterms. The full calculation is cumbersome and can be found in [37]. The numerical result, cutting off the loop integral at the

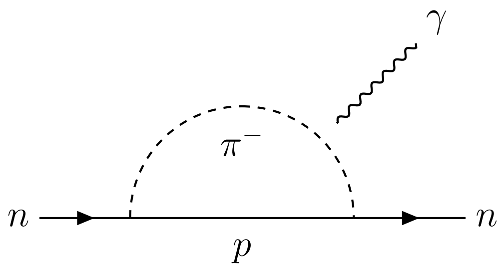

Figure 2.2: Example diagram contributing to the neutron EDM at 1-loop in the chiral lagrangian. The photon is to be attached both to the π^- and to the proton. Other diagrams include a virtual exchange of a K^+ and a Σ^-.

baryon mass scale, reads $d_n/e \simeq 3.3 \times 10^{-16}$ cm. This is compatible with results from other techniques available, including current algebra [38, 39], QCD sum rules [48], holography [49] and lattice QCD calculations [50, 51] and gives a bound on $\bar{\theta}$ very close to the one from our estimate (2.1.65). Interestingly, the exact same calculation can be repeated to compute the Λ hyperons EDM's, which however only provide the negligible bound $|\bar{\theta}| \lesssim 2$ [37].

2.2 Renormalization

The previous analysis showed how in quantum gauge theories the topological angles give rise to very interesting phenomena, even though they seemed associated to operators apparently unphysical from a perturbative point of view. In particular, in the case of QCD we showed how at strong coupling θ gives rise to CP-violating observables such as the neutron EDM or rare meson decays. From a quantum field theory point of view, then, it is only natural to wonder about its renormalization group evolution. Since θ does not parametrize vertices for the quantum fluctuations, however, perturbatively it can represent at most a counterterm necessary to reabsorb (finite and divergent) corrections to CP-violating diagrams with external background gauge fields (for instance associated to instanton configurations). In other words, in perturbation theory θ can only get *additively* renormalized because of contributions induced by the other couplings.

A first distinction we can make is between *finite* and *infinite* renormalization effects. Finite (threshold) corrections to θ are very common. In generic theories these occur at tree-level, when crossing a CP-violating fermion mass threshold, and even more often at loop level. In the case of the QCD θ angle, for example, ref. [52] found that the leading order finite correction is a 3-loop effect when matching to the effective field theory below the $W^\pm$ mass. Infinite corrections are more rare, and a necessary condition for these contributions to actually occur in a mass independent renormalization scheme is that the theory under consideration possesses polynomial CP-odd flavour-invariant combinations of the couplings (other than θ) that can contribute to[13]

$$\beta_\theta = \mu \frac{d\theta}{d\mu}. \tag{2.2.1}$$

In the Standard Model the first such combination, proportional to (1.2.4) [53, 54], appears at very high powers in the Yukawa couplings and therefore indicates that, if present, the UV divergence that renormalizes θ should occur at the prohibitive 7-loops order. In generic renormalizable theories this can occur earlier, as we will see in 2.2.2, but still not before the 2-loops order. However, it is not difficult to find non-renormalizable theories that induce divergent corrections to θ already at 1-loop. For example, at 1-loop the SMEFT operator $cg^2|H|^2 G\widetilde{G}/\Lambda^2$ gives the minimal subtraction result [55]

$$\beta_\theta = -4c\frac{m_H^2}{\Lambda^2} \qquad \text{(1-loop SMEFT)}. \tag{2.2.2}$$

Unfortunately, this effect is power-law suppressed and therefore practically irrelevant in the presence of a large gap between the IR scales and the UV cutoff. In this section, instead, we will present concrete examples of *renormalizable* theories that can induce infinite corrections to θ and we will identify the leading order structure of the beta function β_θ in any mass-independent renormalization scheme. We will also elaborate on the subtleties encountered in a perturbative treatment of θ and show how to concretely approach the calculation of β_θ in dimensional regularization. Finally we will discuss some interesting implications of a non-null β_θ.

This section and appendices 2.A, 2.B are taken from the original work [6].

2.2.1 θ in Perturbation Theory

As mentioned many times in this chapter, in a semi-classical approach to quantum field theory the bare fields are split into a classical finite-action background overlapping with the vacuum plus quantum fluctuations that vanish sufficiently fast at the boundary. The path integral includes a sum over all inequivalent background configurations, so that the generating

[13]Interestingly, reversing this argument we can rigorously prove that in some theories θ *is not* infinitely renormalized at any order. An example are Yang-Mills theories, where this property follows trivially because there are no flavour-invariant CP-odd phases other than θ, and hence there is nothing that can contribute perturbatively to β_θ. The same holds for QCD, since once the phases in the quark mass matrix are removed via anomalous chiral rotations of the fermions, CP violation is entirely encoded in $\bar{\theta}$. Another instance is provided by supersymmetric gauge theories, where the exact one-loop running of the holomorphic gauge coupling reveals that the theta angle does not run because of the absence of CP-odd (holomorphic) combination of the other marginal couplings.

functional explicitly reads[14]

$$
\begin{aligned}
\mathcal{Z}[J_0] &= \sum_{\phi_{0c}} e^{i\theta_0 \frac{g_0^2}{32\pi^2}\int G_{0c}\widetilde{G}_{0c} + i\int J_0\phi_{0c}} \int \mathcal{D}\delta\phi_0 \; e^{iS[\phi_{0c}+\delta\phi_0]+i\int J_0\delta\phi_0} \\
&= \sum_{\phi_{0c}} e^{i\theta_0 \frac{g_0^2}{32\pi^2}\int G_{0c}\widetilde{G}_{0c}} \; \widehat{\mathcal{Z}}[J_0,\phi_{0c}]
\end{aligned}
\tag{2.2.3}
$$

where now with respect to (2.5) we have been more careful in distinguishing the backgrounds ϕ_{0c} from the fluctuations $\delta\phi_0$, and we have explicitly indicated that the fields and couplings entering in (2.2.3) are the *bare* ones, as opposed to the renormalized ones. In particular, since the quantum fluctuations vanish at the boundary, $\int G_0\widetilde{G}_0 = \int G_{0c}\widetilde{G}_{0c}$ and so the angles θ can act only as counterterms in computations with external background fields. Now, an actual evaluation of (2.2.3) is necessarily regularization-scheme dependent. In the following we will specialize on dimensional regularization (Dim-Reg), in which space-time is continued to d dimensions with coordinates $x^\mu = \left\{x^{\bar{\mu}}, x^{\hat{\mu}}\right\}$, where $\bar{\mu}, \bar{\nu}, \cdots = 0,1,2,3$ and $\hat{\mu}, \hat{\nu}, \cdots$ denote the $(d-4)$-dimensional indices. In Dim-Reg the very definition of topological term forces us to define the Levi-Civita tensor and deal with the famous γ_5 problem. So far the only known consistent prescription is the 't Hooft-Veltman-Breitenlohner-Maison scheme [56, 57], where the Levi-Civita tensor is a formal object $\epsilon^{\bar{\mu}\bar{\nu}\bar{\alpha}\bar{\beta}}$ that carries only 4-dimensional indices. In other words, the $(d-4)$-dimensional indices $\hat{\mu}, \hat{\nu}, \cdots$ of an arbitrary vector do not contribute when contracted with this tensor. An important implication is that

$$
G_0\widetilde{G}_0 \equiv \frac{1}{2} G_0^{\bar{\mu}\bar{\nu}} G_0^{\bar{\alpha}\bar{\beta}} \epsilon_{\bar{\mu}\bar{\nu}\bar{\alpha}\bar{\beta}} = \partial_{\bar{\mu}} K_0^{\bar{\mu}}
\tag{2.2.4}
$$

is 4-dimensional divergence of a (x^μ-dependent) vector. Hence the regularized quantity $\int d^dx \; G_{0c}\widetilde{G}_{0c}$ contains a non-trivial residual $(d-4)$-dimensional integral and is not a topological term in d dimensions.

The d-dimensional continuation of (2.2.3) formally represents a set of regularized Green's functions. Therefore, such a path integral violates two of the familiar properties of the topological angle: its periodicity in 2π and its role as compensator (spurion) of the abelian axial symmetry. The technical reason for the first loss boils down to the fact that, as a consequence of (2.2.4), the bare angle θ_0 is not the coefficient of a topological operator in the regularized theory. The second loss occurs because anomalies are d-dependent; as a result, in d-dimensions a shift of the coefficient of $G\widetilde{G}$ does not fully compensate an axial rotation. An intuitive way of arriving to the same conclusions is provided by dimensional analysis: the engineering dimension of the bare coupling in Dim-Reg is $[\theta_0] = d-4$, and it is therefore impossible for θ_0 to be periodic in 2π or even to shift via the dimensionless parameter of the axial transformation while retaining its μ-independence in d-dimensions.

To recover the topological nature of the theta angle, as well as its role as a compensator for abelian axial transformations, one has to derive the renormalized 4-dimensional version of the path integral. In general this procedure requires a renormalization of the topological

[14] Throughout this section we implicitly assume that the gauge-fixing preserves the background gauge invariance.

angle as well. Renormalization renders (2.2.4) a genuinely (4-dimensional) topological term and the background-dependence in the 4-dimensional limit of the path integral (2.2.3) reduces to a dependence on the topological index ν. The renormalized coupling θ is periodic in 2π and transforms via a shift under abelian axial rotations. This ensures in particular that physical amplitudes are invariant under unitary field transformations.

β_θ from (extra)-ordinary diagrams

In the Minimal Subtraction scheme, the relations between the bare couplings θ_0, ξ_{0i} (the latter denoting all couplings except θ_0) and the renormalized couplings θ and ξ_i read (in $d = 4 - \epsilon$ dimensions)

$$\begin{aligned} \xi_{0i} &= \mu^{\rho_i \epsilon}[\xi_i + Z_{\xi_i}] \\ \theta_0 &= \mu^{-\epsilon}[\theta + Z_\theta]. \end{aligned} \tag{2.2.5}$$

By definition $Z_\theta = \sum_{n=1}^{\infty} \epsilon^{-n} Z_{\theta,n}(\xi_i)$ contains no finite term, and similarly for Z_{ξ_i}, so that the 4-dimensional beta functions $\beta_\xi \equiv \lim_{d\to 4} \mu d\xi/d\mu$ read

$$\begin{aligned} \beta_{\xi_i} &= \rho_j \xi_j \partial Z_{i,1}/\partial \xi_j - \rho_i Z_{i,1} \\ \beta_\theta &= \rho_i \xi_i \partial Z_{\theta,1}/\partial \xi_i + Z_{\theta,1} \end{aligned}. \tag{2.2.6}$$

In the above $\partial Z_\theta/\partial\theta = \partial Z_{\xi_i}/\partial\theta = 0$ because θ does not appear in Feynman diagrams. As customary for ordinary couplings, also the beta function of θ is controlled by the simple pole $Z_{\theta,1}/\epsilon$. Diagrammatically, the latter counterterm is defined to subtract the divergent contributions to the connected, CP-odd vertices with external background gauge fields in (2.2.3) :

$$\widehat{\mathcal{Z}}[J_0, \phi_{0c}] \supset -\frac{Z_{\theta,1}}{\epsilon} i \frac{g_0^2}{32\pi^2} \mu^{-\epsilon} \int d^d x \, G_{c0} \widetilde{G}_{c0}. \tag{2.2.7}$$

The divergent corrections to the external legs are removed via a renormalization of the background field, $A^\mu_{0c} = (g/g_0) A^\mu_c$, so that $g_0^2 G_{0c} \widetilde{G}_{0c} = g^2 G_c \widetilde{G}_c$. The divergence remaining in (2.2.7), if any, must be subtracted by $Z_{\theta,1}$.

A word regarding the actual computation of $Z_{\theta,1}$ is now needed. After all the topological term does not represent an ordinary vertex, so how can the combination of ordinary vertices generate divergent corrections to it? The key point is that (2.2.7), in order to be non-vanishing, must be a functional of classical backgrounds with *non-trivial* asymptotic behavior. This implies the calculation of $Z_{\theta,1}$ must be dealt with care. In particular, integration by parts cannot be performed lightly, as opposed to what is customary done when dealing with external sources for asymptotic states. An alternative way to calculate the divergent diagrams contributing to $Z_{\theta,1}$ may be via the trick proposed in [58]. The idea is to promote the CP-odd couplings to non-propagating fields, i.e. "axions", in the intermediate steps and then send them to constant values at the very end of the computation. In this way $\int d^d x \, \theta \, G\widetilde{G}$ would describe an ordinary vertex with $\theta(x)$ and a number of gluons, which is non-vanishing as long as the external $\theta(x)$ carries a momentum. It is now plausible that ordinary Feynman diagrams with external gluons and non-dynamical axions contain divergent contributions that need to be subtracted by $\int d^dx\, \theta\, G\widetilde{G}$, very much like ordinary diagrams were shown in [58] to be capable of generating finite threshold corrections to θ[15]. Thus we see that the divergent contributions to θ can be obtained via non-standard perturbative calculations, either by directly evaluating (2.2.7) in position space [59] or via the method proposed in [58]. There is actually a third approach to calculate directly β_θ, based on conventional perturbative methods typically employed in renormalizing composite operators. We will describe this approach in a moment. However, we stress that whatever method is adopted one has to pay particular attention to how chirality is implemented. The divergent contribution we are interested in is proportional to the Levi-Civita tensor and must arise from a fermion trace involving the γ_5 matrix. A consistent treatment of the latter has to be implemented, and at present in Dim-Reg the unique option seems to be given by the 't Hooft-Veltman-Breitenlohner-Maison scheme. Unfortunately, using this scheme might complicate the calculation a bit, due to the non-standard anti-commutation properties that γ_5 satisfies and the necessity of introducing ad-hoc symmetry-restoring counterterms in order to preserve non-anomalous (global and gauge) chiral symmetries.

β_θ from Operator Mixing

Exploiting the powerful framework of the local renormalization group (see appendix 2.B), it possible to show that in a generic renormalizable quantum field theory the renormalization group equations take the form [60]

$$\mu\frac{d}{d\mu}\begin{pmatrix}\partial_\mu J^\mu_{\dot{A}} \\ \mathcal{O}_{\bar{I}}\end{pmatrix} = -\begin{pmatrix}\gamma_{\dot{A}\dot{B}} & 0 \\ \gamma_{\bar{I}\dot{B}} & \frac{\partial}{\partial\xi_{\bar{I}}}\beta_{\bar{J}}\end{pmatrix}\begin{pmatrix}\partial_\mu J^\mu_{\dot{B}} \\ \mathcal{O}_{\bar{J}}\end{pmatrix} \tag{2.2.8}$$

where $J^\mu_{\dot{A}}$ are the (conserved or anomalous [61]) currents of the global symmetries of the theory, $\mathcal{O}_{\bar{I}}$ denote the marginal interactions, $\xi_{\bar{I}}$ the associated couplings, and $\beta_{\bar{I}} = \mu d\xi_{\bar{I}}/d\mu$. We consider a theory without classically relevant operators for simplicity, but their inclusion is straightforward and does not affect our discussion[16]. The equations (2.2.8) provide an interesting third way of computing the beta-function of θ. To see why, let us start by discussing the

[15] Now that the topological term behaves as an ordinary vertex, one may naively expect θ to be able to show up in matrix elements as well as in the beta functions. It turns out however that this cannot happen. One way to see it is that the renormalized S-matrix amplitudes must be periodic functions of θ. Yet, the new vertex described by the topological term is measured by the strength $\theta g^2/32\pi^2$. Hence the only way it can contribute to renormalized amplitudes and beta functions is via powers of $\theta g^2/32\pi^2$. However there is no way that a perturbative function of $\theta g^2/32\pi^2$ and the other couplings be invariant under $\theta \to \theta + 2\pi$ unless the dependence on $\theta g^2/32\pi^2$ is actually absent altogether. The situation is completely different when non-perturbative effects are taken into account, since in that case inverse powers of the gauge coupling cannot be excluded a priori and a dependence on $\theta g^2/32\pi^2$ can be turned into a periodic function of the sole θ.

[16] A complete basis of $\mathcal{O}_{\bar{I}}$'s includes two sets of marginal operators: $\mathcal{O}_{\bar{I}} = \{E_{\bar{I}''}, O_{\bar{I}'}\}$. The interactions $O_{\bar{I}'}$ are those that define the bare action. These are multiplied by sources $\xi_{\bar{I}'}$ whose background values represent the ordinary couplings of the theory. The $E_{\bar{I}''}$'s denote instead redundant marginal interactions. They include evanescent operators as well as operators that vanish via the equations of motion. They are not associated to any actual coupling of the theory. However they must be included in the functional integral multiplied by spacetime-dependent sources $\xi_{\bar{I}''}$ in order for $\mathcal{O}_{\bar{I}}$ to form a closed set of composite operators under renormalization. The background value of such sources vanishes. Because the $\xi_{\bar{I}''}$'s are not actual couplings, their (background-value) beta functions are proportional to the $\xi_{\bar{I}''}$'s themselves times functions of the true couplings $\xi_{\bar{I}'}$, i.e. the operators $E_{\bar{I}''}$ are multiplicatively renormalized. With this observation we see that

consequence of (2.2.8) on the CP-odd sector of QCD. The marginal CP-odd operators are the divergence of the singlet axial current, $\partial_\mu J^\mu_{\rm A}$, and the topological term $G\widetilde{G}$ (modulo operators that vanish via the equations of motion). By CP invariance the anomalous dimension matrix contains a block diagonal 2 by 2 subgroup involving these two operators only (the operators that vanish on-shell do not affect the following discussion). From (2.2.8) one confirms that this takes the form discussed in [62]. The vanishing $\gamma_{\bar{A}\bar{J}} = 0$ entry is understood as a result of the fact that in Dim-Reg the current $J^\mu_{\rm A}$ renormalizes multiplicatively, being the unique gauge-invariant axial current of the theory, and the same is true for its derivative. Equation (2.2.8) reveals that the $G\widetilde{G}$-$G\widetilde{G}$ element of the anomalous dimension, often denoted by $\gamma_{G\widetilde{G}}$ in the literature, is given by $\gamma_{G\widetilde{G}} = \partial\beta_{G\widetilde{G}}/\partial\xi_{G\widetilde{G}}$, with $\xi_{G\widetilde{G}}$ being the renormalized coupling

$$\xi_{G\widetilde{G}} \equiv \frac{\theta g^2}{32\pi^2}. \tag{2.2.9}$$

Crucially, θ cannot appear in any perturbative calculation, so all our expressions are evaluated at $\xi_{G\widetilde{G}} = 0 = \theta$. As expected, in the external source formalism the latter can only occur with derivatives [63]. Because $\mu d(\theta g^2)/d\mu = \beta_\theta g^2 + \theta\beta_{g^2}$ we find [6]:

$$\gamma_{G\widetilde{G}} = \frac{\partial}{\partial\xi_{G\widetilde{G}}}\beta_{G\widetilde{G}}\bigg|_{\xi_{G\widetilde{G}}=0} = \frac{\partial}{\partial\xi_{G\widetilde{G}}}\left[\beta_\theta\frac{g^2}{32\pi^2} + \xi_{G\widetilde{G}}\frac{\beta_{g^2}}{g^2}\right]\bigg|_{\xi_{G\widetilde{G}}=0} = \frac{\beta_{g^2}}{g^2}. \tag{2.2.10}$$

In the last equality it is crucial that β_θ nor β_{g^2} depend on θ, or equivalently $\xi_{G\widetilde{G}}$. The resulting relation between $\gamma_{G\widetilde{G}}$ and the beta function of the gauge coupling is consistently observed in explicit calculations [64,65][17] and proven in [66,67]. Note that β_θ disappears from this relation and (2.2.10) would still hold even if it was non-trivial. Yet, we know that $\beta_\theta = 0$ in QCD because of the argument given in the introduction: by CP invariance of the theory the only quantity that could appear in β_θ is θ itself, but this can never happen in perturbation theory.

Suppose now we extend QCD introducing in the Lagrangian new CP-violating dimension-4 operators $O_{\bar{I}}$ with $\bar{I} \neq G\widetilde{G}$, e.g. Yukawa couplings. A priori these may mix with the topological term as well as the derivative of the axial current. Yet, according to (2.2.8) the anomalous dimension matrix has a lower triangular form. This is because by dimensional analysis the renormalized axial and Chern-Simons currents cannot contain a component of the bare dimension-4 interactions $O_{\bar{I}0}$, they can only mix among each other following the pattern described above. On the other hand, nothing forbids the renormalized $O_{\bar{I}}$ to contain a linear combination of $\partial_\mu J^\mu_{{\rm A},0}, G_0\widetilde{G}_0$. The lower-triangular structure can also be understood as a consequence of the independence of the anomalous dimensions and beta functions on θ, i.e. $\partial\beta_{\bar{I}\neq G\widetilde{G}}/\partial\xi_{G\widetilde{G}} = 0$ [61]. Interestingly, equation (2.2.8) demonstrates that (2.2.10) remains valid even in the presence of the new interactions $O_{\bar{I}}$ ($\bar{I} \neq G\widetilde{G}$). More importantly, however, the new CP-violating couplings make it possible for θ to get additively renormalized,

(2.2.8) describes a familiar pattern of RG-mixing: the anomalous dimension matrix for $\{E_{\bar{I}''}, O_{\bar{I}'}\}$ has a lower triangular form in which the redundant operators renormalize among each other whereas the $O_{\bar{I}'}$ renormalize via a mixture of themselves, $\partial_\mu J^\mu_A$, and $E_{\bar{I}''}$.

[17] We use a different convention than these authors. For us the scaling dimension of an operator is $d_{\rm cl} + \gamma$, with $d_{\rm cl}$ the engineering dimension.

and (2.2.8) turns out to contain important information on β_θ. The component of $-\mu dO_{\bar{I}}/d\mu$ proportional to $G\widetilde{G}$ is (for $\bar{I} \neq G\widetilde{G}$) [6]:

$$\gamma_{\bar{I}G} = \frac{\partial}{\partial \xi_{\bar{I}}} \beta_{G\widetilde{G}} \bigg|_{\xi_{G\widetilde{G}}=0} = \frac{\partial}{\partial \xi_{\bar{I}}} \left[\beta_\theta \frac{g^2}{32\pi^2} + \xi_{G\widetilde{G}} \frac{\beta_{g^2}}{g^2} \right] \bigg|_{\xi_{G\widetilde{G}}=0} = \frac{g^2}{32\pi^2} \frac{\partial}{\partial \xi_{\bar{I}}} \beta_\theta. \tag{2.2.11}$$

Equation (2.2.11) may be interpreted as an indirect procedure for deriving β_θ. The latter may indeed be extracted by integrating the off-diagonal element $\gamma_{\bar{I}G}$ of the anomalous dimension matrix of the CP-violating operators, roughly measured by the amount of $G_0\widetilde{G}_0$ contained in $O_{\bar{I}}$, with respect to the couplings $\xi_{\bar{I}}$ that contribute additively to β_θ. Interestingly, this procedure does not require the background field method nor promoting the CP-odd phases to spurion fields. It can be carried out using ordinary perturbation theory because the anomalous dimension matrix is to be calculated assuming a non-vanishing momentum is flowing into the relevant operators, such that even $G\widetilde{G}$ describes an ordinary vertex.

2.2.2 β_θ in Renormalizable QFTs

Having shown how to extract the beta function of the topological angle, we will now identify the explicit form of the leading order β_θ for generic renormalizable quantum field theories. The most general lagrangian can be compactly written in terms of Weyl fermions ψ_i and real scalars ϕ_a as

$$\begin{aligned}\mathcal{L} = &- \frac{1}{4g^2_{AB}} F^A_{\mu\nu} F^{B\mu\nu} + \frac{1}{2} (D^\mu\phi)_a (D_\mu\phi)_a + \psi_i^\dagger i\bar{\sigma}^\mu (D_\mu\psi)_i \\ &- \left(\frac{1}{2} Y_{a\,ij} \psi_i \psi_j \phi_a + \text{hc} \right) - \frac{\lambda_{abcd}}{4!} \phi_a\phi_b\phi_c\phi_d + \frac{\theta^{AB}}{64\pi^2} \epsilon^{\mu\nu\rho\sigma} F^A_{\mu\nu} F^B_{\rho\sigma} \\ &+ \text{(relevant couplings)} + \text{(gauge fixing)} + \text{(ghosts)},\end{aligned} \tag{2.2.12}$$

where $(D_\mu\psi)_i = \partial_\mu\psi_i - iA^A_\mu T^A_{ij}\psi_j$ and $(D_\mu\phi)_a = \partial_\mu\phi_a - iA^A_\mu S^A_{ab}\phi_b$. The gauge generators T^A are hermitian whereas S^A are purely imaginary hermitian, and hence anti-symmetric. The fermions and scalars are in general in a reducible representation of the gauge group, and the indices $i, j, \cdots$ (ranging from 1 to some integer N_ψ) and $a, b, \cdots$ (ranging from 1 to N_ϕ) include both gauge and flavour components. The coupling λ is fully symmetric, and Y is symmetric in the fermionic indices. The gauge symmetry is an arbitrary product of abelian factors and simple groups $\mathcal{G}_{\text{gauge}} = \Pi_{\mathcal{G}}\mathcal{G}$. The indices $A, B, \cdots$ run over the adjoint representation of $\mathcal{G}_{\text{gauge}}$ and the (real) gauge coupling g^2_{AB} is to be interpreted as the direct sum of identities in the non-abelian part of that space plus a symmetric part for the abelian factors. Note that the normalization of the gauge fields is non-canonical, so that the gauge bosons' propagators are proportional to g^2_{AB}.

Gauge invariance restricts the form of some of the couplings in (2.2.12). In particular, the topological term, as the gauge couplign metric, must be proportional to the identity $\delta^{AB}_{\mathcal{G}}$ in any non-abelian components $\mathcal{G}$ and symmetric, but potentially with off-diagonal entries, in the abelian factors:

$$\theta^{AB} = \sum_{\mathcal{G}=\text{non}-\text{Ab.}} \theta_{\mathcal{G}} \delta^{AB}_{\mathcal{G}} + \sum_{\mathcal{G}_1,\mathcal{G}_2=\text{Ab.}} \theta^{AB}_{\mathcal{G}_1,\mathcal{G}_2}. \tag{2.2.13}$$

Another implications of gauge invariance is

$$S^A_{ab} Y_b = Y_a T^A + (T^A)^* Y_a. \tag{2.2.14}$$

which we will extensively use later. Note that none of these relations depend on the space-time dimension and hence can be assumed to hold in Dim-Reg as well.

CP and Flavour Symmetry

We want to identify which combinations of the couplings can appear in β_θ. In Dim-Reg the renormalization group evolution of θ is necessarily controlled by the marginal couplings of the theory. Classically relevant interactions are therefore not important in the present discussion and have not been explicitated in (2.2.12). To proceed we first have to familiarize with the approximate symmetries of our theory.

In the absence of interactions the lagrangian (2.2.12) enjoys a large global symmetry that includes a flavour symmetry that rotates the matter fields, one that rotates vectors, and CP:

$$\mathcal{G}_{\text{flav}} \times \mathcal{G}_{\text{glob}} \times \text{CP}. \tag{2.2.15}$$

The flavour symmetry $\mathcal{G}_{\text{flav}} \equiv U(N_\psi) \times O(N_\phi)$ rotates all fermions among each other and acts similarly on scalars, the group $\mathcal{G}_{\text{glob}}$ rotates the vectors leaving g^2_{AB} in (2.2.12) invariant and finally CP acts as usual up to unitary rotations. Precisely because of this freedom, the simultaneous actions of $\mathcal{G}_{\text{flav}} \times \mathcal{G}_{\text{glob}} \times$ CP can actually be interpreted as transformations belonging to a *generalized* CP symmetry, sending

$$\psi_i(x) \to U_{ij} \epsilon \psi^\dagger_j(\mathcal{P}x), \qquad \phi_a(x) \to O_{ab}\phi_b(\mathcal{P}x), \qquad A^A_\mu(x) \to R^{AB}\mathcal{P}^\nu_\mu A^B_\nu(\mathcal{P}x) \tag{2.2.16}$$

where U, O, R are matrices respectively of $U(N_\psi)$, $O(N_\phi)$, $\mathcal{G}_{\text{glob}}$ and we defined $\mathcal{P}^\nu_\mu x_\nu = x^\mu$. In the following this distinction will not make any difference, and for simplicity we will keep distinguishing CP from the flavour symmetry as in 2.2.15. However the notion of generalized CP will prove useful in the analyses of chapters 3, 4.

The symmetry (2.2.15) is explicitly broken by the gauge generators, the Yukawa and scalar couplings as well as anomalies. Yet, it can be formally restored by promoting T^A, S^A, Y_a, λ, θ to spurions with transformations designed to exactly compensate (2.2.16) so that the theory is manifestly invariant under the full group (2.2.15)[18]. Explicitly, the transformations of the spurions T^A, S^A, Y_a, λ under $\mathcal{G}_{\text{flav}} \times \mathcal{G}_{\text{glob}} \times$ CP are given by

$$\begin{aligned} T^A_{ij} &\to -R^{AB} U_{im} U^*_{jn} [T^B]^*_{mn} \\ S^A_{ij} &\to -R^{AB} O_{im} O_{jn} [S^B]^*_{mn} \\ Y_{a\,ij} &\to U^*_{im} U^*_{jn} O_{ab} [Y_{b\,mn}]^* \\ \lambda_{abcd} &\to O_{am} O_{bn} O_{cp} O_{dq} \lambda_{mnpq}. \end{aligned} \tag{2.2.17}$$

The renormalized coupling θ^{AB} should also be interpreted as a spurion. It is a complete singlet of $O(N_\phi)$, a 2-index symmetric tensor of $\mathcal{G}_{\text{glob}}$, and is also $U(N_\psi)$-invariant except for

[18]In Dim-Reg the addition of counterterms is generically necessary to render the theory formally invariant under the axial part of $U(N_\psi)$. We assume this is done order by order in perturbation theory.

its anomalous subgroup. Denoting with T_r^A the generators of each irreducible representation r of the gauge group modulo flavour degeneracies, and by U_r the flavour rotation among the fermions in r, the spurious $\mathcal{G}_{\rm flav} \times \mathcal{G}_{\rm glob} \times$ CP transformation of θ explicitly reads

$$\theta^{AB} \to -R^{AM}R^{BN}\left\{\theta^{MN} - 2\sum_r \arg\det U_r \;\; {\rm tr}\, T_r^M T_r^N\right\}. \tag{2.2.18}$$

Here the overall minus is necessary to compensate the P-odd nature of $\epsilon^{\mu\nu\rho\sigma}F^A_{\mu\nu}F^B_{\rho\sigma}$.

The beta function $\mu d\theta^{AB}/d\mu$ is not affected by the anomalous shift of the topological term and therefore is a complete singlet of the flavour symmetry $\mathcal{G}_{\rm flav}$ and CP-odd in the sense that

$$\mu\frac{d}{d\mu}\theta^{AB} \to -R^{AM}R^{BN}\mu\frac{d}{d\mu}\theta^{MN}. \tag{2.2.19}$$

The beta function is a linear combination of functions I^{AB} of the spurions (2.2.17) multiplied by real numerical coefficients, since its calculation involves no branch cuts. The quantities I^{AB} are obtained by contracting the indices of the spurions with the invariant tensors $\delta_{ij}, \delta_{ab}, g^2_{AB}$ that define the kinetic terms. In the language of Feynman diagram, this is just a consequence of the fact that any diagram contributing to the beta function is obtained by contracting vertices with propagators. In order to appear in β_θ the tensors should transform precisely as in (2.2.19). Yet, from (2.2.17) follows that under the full symmetry (2.2.15) any 2-index tensor function of (2.2.17) transforms as

$$I^{AB}(T,S,Y,\lambda) \to R^{AM}R^{BN}I^{MN}(T^*,S^*,Y^*,\lambda^*) \tag{2.2.20}$$

The minus signs of (2.2.17) cancel out because $\mathcal{G}_{\rm glob}$-covariance forces I^{AB} to be built out of an even number of (scalar plus fermion) gauge generators. Now, to reproduce the transformation of $\mu d\theta^{AB}/d\mu$ the tensors must satisfy $I^{AB}(T^*,S^*,Y^*,\lambda^*) = -I^{AB}(T,S,Y,\lambda)$, and of course be real. We thus conclude that it is the imaginary part of the invariant that has the correct CP-odd property, namely[19]

$$\mu\frac{d}{d\mu}\theta^{AB} = \sum_\alpha c_\alpha \;\; {\rm Im}\left[I^{AB}_\alpha\right], \tag{2.2.21}$$

where c_α are real numbers and α is some label. In other words, the CP-odd invariants that define β_θ are the *imaginary* parts of the $\mathcal{G}_{\rm flav}$-singlet, 2-index tensors of $\mathcal{G}_{\rm glob}$. A theory that does not possess any such quantity cannot renormalize θ. This is what happens in Yang-Mills theories as well as QCD. In the next subsection we will show the explicit form of the leading order contribution to β_θ in arbitrary renormalizable theories of the form (2.2.12).

[19]The relations $I^{AB}(T^*,S^*,Y^*,\lambda^*) = [I^{AB}(T,S,Y,\lambda)]^* = -I^{AB}(T,S,Y,\lambda)$,the first equality following from the fact that all coefficients are real and the second from the requirement that (2.2.20) reproduces (2.2.19), are equivalent to ${\rm Re}\,[I^{AB}(T^*,S^*,Y^*,\lambda^*)] = {\rm Re}\,[I^{AB}(T,S,Y,\lambda)] = 0$ and ${\rm Im}\,[I^{AB}(T^*,S^*,Y^*,\lambda^*)] = -\,{\rm Im}\,[I^{AB}(T,S,Y,\lambda)]$.

The 3-loop Diagrams

Following the method described above, we can identify all the structures that can potentially contribute to β_θ at any perturbative orders. We present an analysis up to 2-loops order in appendix 2.A. What we find is that there are no 1-loop-sized 2-index tensors that are CP-odd, and therefore that the 1-loop beta function must vanish. The first non-trivial contribution to β_θ potentially arises at 2-loops and is controlled by a unique structure [6]:

$$\mu\frac{d}{d\mu}\theta^{AB} = c\frac{\hbar^2}{(16\pi^2)^2}\,\mathrm{Im}\, I^{AB}_{(2)} + \mathcal{O}(\hbar^3), \tag{2.2.22}$$

where c is an ordinary number expected to be of order unity and

$$\begin{aligned} I^{AB}_{(2)} &= \mathrm{tr}\,\left\{(Y^*_a[T^A]^*Y_bY^*_aY_b - Y^*_aY_bY^*_a[T^A]^*Y_b)T^B\right\} \\ &= \mathrm{tr}\,\left\{(Y^*_aY_cY^*_bY_c - Y^*_cY_aY^*_cY_b)T^B\right\} S^A_{ab} \\ &= \frac{1}{2}\,\mathrm{tr}\,\{Y^*_aY_cY^*_bY_c - Y^*_cY_aY^*_cY_b\}\,(S^AS^B)_{ab} \end{aligned} \tag{2.2.23}$$

is the only invariant with non-vanishing CP-odd component (i.e. an imaginary part) at $\mathcal{O}(\hbar^2)$. Note that symmetry under the exchange $A \leftrightarrow B$ in the first and third lines of (2.2.23) is a consequence of cyclicity of the trace as well as symmetry of the Yukawa under the exchange of the fermionic indices and hermiticity of the gauge generators, whereas in the second line of (2.2.23) also (2.2.14) is needed.

The invariant $I^{(2)}_{AB}$ has been written in three different forms employing the identity (2.2.14). These expressions provide complementary information about the properties that a theory must possess in order to renormalize θ^{AB} at this order. For instance, from the second and third lines in (2.2.23) it is evident that there would be no 2-loop beta function if the scalars were gauge-singlets, i.e. if $S^A = 0$. To show this using the expression in the first line is less immediate: one has to use (2.2.14) with $S^A = 0$ to prove that $Y^*_aY_bT^A = T^AY^*_aY_b$, from which consistently follows that the first line of (2.2.23) vanishes. Actually, a more careful inspection reveals that the scalar fields have to belong to at least two different representations of $\mathcal{G}$. To prove this we distinguish between non-abelian and abelian gauge groups. In the case the indices A, B are associated to a non-abelian gauge group, from (2.2.13) we know that the relevant part of the beta function is the one proportional to $\delta^{\mathcal{G}}_{AB}$. If all the scalars belonged to the same representation, then contracting the latter with the expression in the third line of (2.2.23) would give a Casimir times the identity in the scalar index space. This implies that the invariant would vanish as $\mathrm{tr}\,\{Y^*_aY_cY^*_aY_c - Y^*_cY_aY^*_cY_a\} = 0$. In the case A, B refer to abelian gauge groups the generators can all be taken diagonal, i.e. $S^A_{ab} = q^A_a\delta_{ab}$ and so $(S^AS^B)_{ab} = q^A_aq^B_a\delta_{ab}$. As before, we see that when the charges q^A_a, q^B_a do not depend on the index a then the combination of scalar generators is again proportional to the identity δ_{ab} and the third line of (2.2.23) is identically zero.

The three equivalent forms of (2.2.23) also help us identify the diagrams that contribute to the 2-loop beta function, e.g. if we adopted the background field method. These are illustrated in fig. 2.3, with the $\otimes$ indicating the insertion of the external background gauge field. The topology in Fig. 2.3a is responsible for generating the invariant in the first line

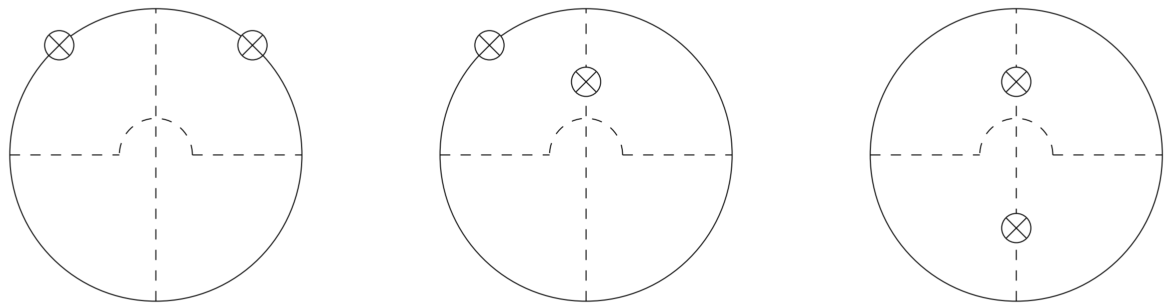

Figure 2.3: Topologies of the diagrams generating the CP-odd invariant (2.2.23). Crossed circles represent insertions of the external gauge field. Topology (a) refers to the form in (2.2.23) written in terms of two fermionic generators, topology (b) to the one in terms of one fermionic and one scalar generator, and topology (c) to the one in terms of two scalar generators. It is intended that each diagram within a given topology should be properly symmetrized. Figure taken from [6].

of (2.2.23), the one in fig. 2.3b is associated to the second line of (2.2.23) and finally the topology of Fig. 2.3c to the third form of the CP-odd invariant. The overall correction to β_θ must include a sum over all three topologies. We emphasize that these are effectively 3-loop diagrams, and yet they contribute to the 2-loop order beta function because θ appears in the action multiplied, in the canonically normalized field basis, by a loop factor $g^2\hbar/32\pi^2$. The coefficient c in (2.2.22) is model-independent and may be derived calculating the above diagrams in any model with a non-vanishing $I^{AB}_{(2)}$, as for example in the toy scenario discussed later. We will not explicitly compute it here because the calculation goes far beyond the author's technical abilities. Yet, there is circumstantial evidence that an explicit computation would find $c \neq 0$. Indeed, ref. [68] presents a calculation of the beta function of the gauge coupling at 3-loop order using the background field method. The class of diagrams considered there includes those of fig. 2.3[20]. Consistently, those authors find that the beta function of the gauge coupling is controlled by the set of CP-even invariants of appendix 2.A, including in particular the CP-even component of our $I^{AB}_{(2)}$ (namely its real part). This gives confidence that the evaluation of the diagrams in fig. 2.3 will not cancel against each other. The result $c = 0$ would thus be rather surprising, and would indicate the presence of an accidental cancellation.

2.2.3 Implications for the Strong CP problem

Adopting the logic outlined in the previous subsection, it easy to discover that in the renormalizable version of the Standard Model the first contribution to β_θ arises at least at 7-loops

[20] In practice the calculation is very different, though, because here we are interested in the CP-odd contributions proportional to the Levi-Civita tensor, whereas in [68] the authors were allowed to take $\epsilon_{\bar\mu\bar\nu\bar\alpha\bar\beta} = 0$.

and reads [53, 54]

$$\beta_\theta = \frac{\hbar^7}{(16\pi^2)^7} \det\left[Y_u {Y_u}^\dagger, Y_d {Y_d}^\dagger\right] \left\{a_1\, g'^2 + a_2 \ \mathrm{tr}\ \left(Y_u {Y_u}^\dagger - Y_d {Y_d}^\dagger\right)\right\} + \mathcal{O}(\hbar^8) \tag{2.2.24}$$

where the terms proportional to a_1, a_2 are there to comply with the selection rule associated to the symmetry $\{u, \tilde{H}, Y_u\} \leftrightarrow \{d, H, Y_d\}$, which would otherwise make β_θ vanish. This implies that the renormalization group evolution of θ is numerically negligible, leaving open the possibility that the absence of CP violation in low energy QCD be the result of some unknown mechanism at very short distances $\sim 1/\Lambda$, even close to the Planck scale, that sets $\bar{\theta}(\Lambda) \approx 0$. The numerical stability of $\theta = 0$ is however not a generic property in field theory, that as we have seen may potentially develop a non-trivial renormalization of the theta angle already at 2-loops. This demonstrates that UV solutions of the Strong CP problem cannot work if the effective field theory below Λ is completely generic and has unsuppressed Yukawa interactions. Actually, the radiative instability of UV solutions to the Strong CP problem is typically worse than naively expected. First of all, arbitrary extensions of the Standard Model would necessarily feature new mass scales, and it is well-known that threshold corrections can affect the topological angles at tree and loop order as we will see in chapter 4. What we want to stress here is another subtle effect: despite the fact that the beta function of θ is generically at 2-loop order, *physical* phases typically renormalize at 1-loop. By physical phases we mean those special combinations of the renormalized couplings, namely those combinations that are invariant under unitary rotations of the fields, that enter observables. We have seen that θ is not invariant under anomalous rotations of the fermions, see (2.2.18), and therefore the physical rescaling invariant quantity of interest must take the form

$$\bar{\theta} = \theta + f(\{\xi_i\}), \tag{2.2.25}$$

where f is a model-dependent function. When no f with the required properties exists then no invariant can be built out of θ and that parameter is unphysical. The physical angle runs according to

$$\beta_{\bar{\theta}} = \beta_\theta + \frac{\partial f}{\partial \xi_i} \beta_{\xi_i}. \tag{2.2.26}$$

The symmetry properties of $\beta_{\bar{\theta}}$ are precisely the same as those of β_θ. In particular, both must be proportional to the imaginary parts of the invariants I^{AB}. Yet, while β_θ and β_{ξ_i} are polynomial in the couplings, it may happen that $\partial f / \partial \xi_i$ brings inverse powers of the ξ_i's that are not compensated by β_{ξ_i}. In such a situation $\beta_{\bar{\theta}}$ develops negative powers of the couplings in front of the invariants I^{AB}, which in practice indicates that $\beta_{\bar{\theta}}$ arises at an order in perturbation theory that is lower compared to β_θ. We will see explicitly below how this can occur in a concrete toy model. In this respect, the Standard Model is special because this enhancement cannot appear. Indeed, in pure Standard Model $\bar{\theta} = \theta - \mathrm{Im\ tr} \log Y_u Y_d$ and we have

$$\beta_{\bar{\theta}} = \beta_\theta - \mathrm{Im\ tr}\, Y_u^{-1} \beta_{Y_u} - \mathrm{Im\ tr}\, Y_d^{-1} \beta_{Y_d}. \tag{2.2.27}$$

Since by the accidental flavour symmetries $\beta_Y = \beta'_Y Y$, for some polynomial β'_Y, then $\beta_{\bar{\theta}}$ is made up of the very same polynomial structures of couplings as β_θ. This fact for example ensures that the estimates of [53, 54] are robust.

A Toy Model

Let us consider an extension of the Standard Model in which we introduce two Weyl fermions with charges $\psi \sim \mathbf{3}$ and $\psi^c \sim \overline{\mathbf{3}}$ under $SU(3)_c$, a singlet scalar ϕ_1 and a scalar octet $\phi_2 \sim \mathbf{8}$. These fields are all neutral under the electroweak group. The Lagrangian reads

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\rm SM} + \frac{1}{2}(\partial\phi_1)^2 + \frac{1}{2}(D\phi_2)^2 + \psi^\dagger i\not{D}\psi + \psi^{c\dagger} i\not{D}\psi^c \\ -(m + y_1\phi_1)\,\psi\psi^c + y_2\,\psi\psi^c\phi_2 + \text{h.c.} - V(\phi_1, \phi_2^2, |H|^2)\end{aligned} \tag{2.2.28}$$

In addition to the CP-odd parameters of the Standard Model, this theory features two more phases. On top of that, the rescaling-invariant QCD theta angle contains a new contribution because θ is now affected by chiral rotations of ψ, ψ^c as well. A natural definition of the phases (the definition of the Jarlskog invariant is the standard one (1.2.4)) is given by

$$\begin{aligned}\varphi_y &= \arg\,(y_1 y_2^*)^2 \\ \varphi_m &= \arg y_1^2 y_2^2 (m^*)^4 \\ \bar\theta &= \theta - \arg\det Y_u Y_d - \arg m.\end{aligned} \tag{2.2.29}$$

All these CP-odd parameters run already at 1-loop:

$$\begin{aligned}\mu\frac{d\varphi_y}{d\mu} &= -\frac{4}{16\pi^2}\sin\varphi_y\left[|y_1|^2 + \frac{4}{3}|y_2|^2\right] \\ \mu\frac{d\varphi_m}{d\mu} &= +\frac{4}{16\pi^2}\left[|y_1|^2\left(\sin\varphi_y - 2\sin\frac{\varphi_y+\varphi_m}{2}\right) + \frac{4}{3}|y_2|^2\left(-\sin\varphi_y + 2\sin\frac{\varphi_y-\varphi_m}{2}\right)\right] \\ \mu\frac{d\bar\theta}{d\mu} &= -\frac{2}{16\pi^2}\left[|y_1|^2\sin\frac{\varphi_y+\varphi_m}{2} + \frac{4}{3}|y_2|^2\sin\frac{-\varphi_y+\varphi_m}{2}\right].\end{aligned} \tag{2.2.30}$$

As opposed to the Standard Model, this is also true for $\bar\theta$ because the phase of the exotic fermion mass has a non-trivial renormalization group evolution. Therefore, $\beta_{\bar\theta}$ arises at 1-loop. Instead, the result (2.2.22) applied to this model reads[21] [6]:

$$I_{(2)}^{AB} = 3\ \mathrm{Im}\left[(y_1 y_2^*)^2\right]\delta^{AB}. \tag{2.2.31}$$

[21]Given the expression (2.2.31), it is interesting to verify the earlier claim that the 2-loop β_θ requires at least two scalars with different gauge representations. Indeed, if we replace ϕ_2 with another scalar singlet (or ϕ_1 with another scalar in the adjoint) that invariant would not be generated by any diagram. The reason is that in such a scenario the kinetic term of the scalars would possess a $O(2)$ symmetry that rotates (ϕ_1, ϕ_2). As usual this flavour symmetry is explicitly broken by the interactions, but may be formally resurrected by promoting all the couplings to spurions, as described earlier. The pair (y_1, y_2) would be formally a doublet of $O(2)$ and β_θ should be a singlet. However (2.2.31) does not meet this requirement. Manifestly $O(2)$ invariants are $y_a y_a$, which is not invariant under axial rotations of the fermions, and $y_a y_a^*$, which is CP-even. There is no combination of the former that is simultaneously CP-odd and invariant under axial rotations, besides of course the one involving θ, which we know has no perturbative effect. Note that the quantity $y_a\epsilon_{ab}y_b = i\,\mathrm{Im}\,[y_1 y_2^*]$ is $SO(2)$-symmetric and CP-odd, but in order to have an invariant under $Z_2 \subset O(2)$ one would need an even power of it, thus resulting in a CP-even combination.

and so

$$\beta_\theta = \frac{c}{(16\pi^2)^2}(+3)\,\mathrm{Im}\left[(y_1 y_2^*)^2\right] + \mathcal{O}(\hbar^3)$$
$$= \frac{3c}{(16\pi^2)^2}|y_1|^2|y_2|^2\sin\varphi_y + \mathcal{O}(\hbar^3) \tag{2.2.32}$$

meaning that the latter beta function is always subdominant and can be neglected. Indeed it would be enough to have either a small y_1 or a small y_2 in order to suppress β_θ, but this would not stop $\bar{\theta}$ from running. Similarly, $\sin\varphi_y = 0$ would suppress β_θ but not $\beta_{\bar{\theta}}$ if $\sin\varphi_m \neq 0$. Only under the simultaneous conditions $\sin\varphi_y = \sin\varphi_m = 0$ the parameter $\bar{\theta}$ becomes approximately stable under the renormalization group, but this is obvious since in that situation our theory introduces no new CP-odd phases compared to the Standard Model.

There is now an important point to stress. The definitions (2.2.29) (and more generally (2.2.25)) are ambiguous because any combination of invariants is also invariant. Yet, the ambiguity in defining the physical theta angle arises only from a UV perspective. Its IR definition is basically fixed by experiments. In the Standard Model one usually takes $\theta - \arg\det Y_u Y_d$ because at leading order in a perturbative expansion in Yukawas (masses) this is precisely the quantity observed in low energy processes. The situation is similar here. Indeed, because the new fermions ψ, ψ^c are colored, phenomenologically we expect $|m| \gtrsim 1$ TeV; at scales below the mass threshold $|m|$ we can thus integrate them out such that our theory reduces to the Standard Model with topological angle given by

$$\bar{\theta}_{\mathrm{SM}} = \left[\theta - \arg\det Y_u Y_d - \arg m\right]_{\mu=|m|} + \mathcal{O}(\hbar), \tag{2.2.33}$$

where the small corrections are due to 1-loop threshold effects. We thus see that the UV ambiguity is resolved: within the accuracy we are working the rescaling-invariant topological angle defined in (2.2.29) is precisely the one constrained by low energy experiments. The distinction between $\bar{\theta}_{\mathrm{SM}}$ and $\bar{\theta}$ is 1-loop order, and so the difference in their beta functions starts at 2-loops. Hence the statement that $\bar{\theta}$ runs already at 1-loop in fact also applies to the physically relevant CP-odd parameter measured in experiments. An UV solution of the Strong CP problem that enforces the condition $\bar{\theta}(\Lambda) = 0$ would not be able to explain the absence of CP violation in low energy QCD, unless a non-trivial conspiracy among different UV parameters is present.

Finally, what would happen if we set $m = 0$? Such an alternative scenario may still be phenomenologically viable provided the exotic fermion gets a mass $y_1 v_1$ from the vacuum expectation value of the singlet $\langle\phi_1\rangle = v_1$. This model introduces a unique new phase φ_y whereas the definition of the QCD topological angle is apparently ambiguous. Again, the ambiguity is resolved by observing that the QCD angle at the matching scale reads

$$\bar{\theta}_{\mathrm{SM}} = \left[\theta - \arg\det Y_u Y_d - \arg y_1 v_1\right]_{\mu=|y_1 v_1|} + \mathcal{O}(\hbar). \tag{2.2.34}$$

Again, in order to have a small $|\bar{\theta}_{\mathrm{SM}}|$ there should be a cancellation between the phase of y_1, which runs at 1-loop, and $\theta - \arg\det Y_u Y_d$, which does not run before 7-loops. This cancellation is not stable unless $\sin\varphi_y = 0$, i.e. unless we recover the same amount of CP violation as in the Standard Model.

In conclusion, the lesson is clear: generic extensions of the Standard Model feature several flavour-invariant CP-odd phases with non-trivial renormalization group evolution, such that there are no simple UV conditions that ensure $|\bar{\theta}_{\rm SM}|$ stays small. Solutions of the Strong CP problem that are able to deal with this are rare and will be discussed in chapter 4.

Appendices

2.A CP-odd Invariants for β_θ Up to Two-Loops

In this appendix, taken from [6], we build the flavour-invariants for β_θ introduced in 2.2.2 and identify the CP-odd ones. We approach the problem perturbatively. By counting the powers of $\hbar$ in the flavour-singlet structures one sees that corrections to the 1PI vertex with two external gauge bosons proportional to $g^{2c_g} Y^{2c_Y} \lambda^{c_\lambda}$ correspond to diagrams with $n = c_g + c_Y + c_\lambda$ loops. Yet, the associated contribution to $Z_{\theta,1}$ is a 1-loop factor smaller because the definition of the θ vertex includes a factor $g^2\hbar/16\pi^2$. This implies that in order to find a correction of n-loop size to the beta function, hereafter denoted by $\beta_\theta^{(n)}$, one should calculate an $n+1$-loop diagram. On pure dimensional grounds, therefore, corrections to β_θ are expected to be controlled by

$$\begin{aligned}
&\beta_\theta^{(1)}: \qquad g^4, g^2\lambda, g^2Y^2 \\
&\beta_\theta^{(2)}: \qquad g^6, g^4\lambda, g^2\lambda^2, g^4Y^2, g^2Y^2\lambda, g^2Y^4 \\
&\ \ldots
\end{aligned} \tag{2.A.1}$$

In this appendix we will content ourselves with 1- and 2-loop size effects, namely $\beta_\theta^{(1,2)}$, though the formalism we adopt can be extended up to arbitrary order. Up to this order it is rather straightforward to argue that only the terms involving the Yukawa couplings have some real chance of being CP-odd, as we now show.

A general renormalizable gauge theory without scalar quartics and Yukawas always conserves CP (if the topological angles can be neglected, which is the case perturbatively). Hence there cannot be any CP-odd invariant I^{AB} built out of T^A, S^A only. Purely bosonic flavour-invariants cannot work either. They depend on S^A, λ_{abcd} and are automatically CP-even. Indeed, λ is real whereas S^A are purely imaginary. To build a CP-odd combination we would need an odd number of S^A, which cannot be covariant under rotations of the adjoint index because the only invariant tensor at our disposal for contractions is g^2_{AB}. This conclusion is expected on account of that the Feynman diagrams we are interested in must be proportional to the Levi-Civita tensor, and therefore fermion traces are strictly necessary to generate them. Actually, we can explicitly demonstrate that even the combinations $g^2\lambda$, $g^2\lambda^2$ and $g^4\lambda$ can be discarded. Recall that a fermion loop is necessary, so such invariant must include traces of the fermion generator. The only $g^2\lambda$ invariant we can have is $\operatorname{tr} T^AT^B\lambda_{aabb}$ and is manifestly real, i.e. CP-even, since the fermionic trace gives the direct sum of identities in the adjoint index multiplied by (real) fermion Casimirs. Similar considerations apply to the $g^2\lambda^2$ invariants, which are of the form $\operatorname{tr} T^A T^B \lambda^2$ with all possible contractions of the scalar indices. Whatever contraction is taken the invariant is CP-even. Finally, at order $g^4\lambda$ we can have structures of the form $\operatorname{tr} TTTT\lambda$, $\operatorname{tr} TT \operatorname{tr} TT\lambda$, and $\operatorname{tr} TTSS\lambda$. The former two are manifestly real because the potential CP-violating contribution would have to come exclusively from the fermionic generators, and we have recalled above that this is not possible. The last one may either contain $[S^M S^N]_{ab}\lambda_{abcc}$ or $[S^M S^N]_{aa}\lambda_{bbcc}$, which are both real. We did not include structures with a single scalar generator because those vanish: the (anti-symmetric) scalar indices in S^A should necessarily be contracted with those (symmetric) of λ. Similarly, three scalar generators would require the fermionic trace $\operatorname{tr} T = 0$, which vanishes in the absence of mixed gravity-gauge anomalies.

We conclude that, at least up to the perturbative order considered here, the Yukawa couplings are strictly necessary to build the CP-odd flavour-invariants I^{AB} of the theory (2.2.12). For convenience it is useful to introduce the basic covariant combinations in which they can appear:

$$\Upsilon^{A_1\cdots A_n}_{ab;ij} \equiv Y^*_{a\,ik}(T^{A_1})^*\cdots(T^{A_n})^* Y_{b\,kj}. \tag{2.A.2}$$

These objects transform under fermion rotations precisely as T^A, under scalar rotations as the product of two Yukawas, whereas under gauge boson rotations in an obvious way. We will use this compact notation to write the possible invariants appearing in $\beta^{(1,2)}_\theta$.

Absence of $\beta^{(1)}$: Two-Loops Diagrams

At lowest order we have a very limited number of flavour-invariant structures that can contribute. They are so few that we can write them explicitly:

$$g^2Y^2: \qquad \operatorname{tr} T^{(A}T^{B)}\Upsilon_{aa}, \quad \operatorname{tr} T^{(A}\Upsilon^{B)}_{aa}, \quad \operatorname{tr} T^{(A}T^{B)} \operatorname{tr}\Upsilon_{aa}, \quad S^{(A}_{mn}S^{B)}_{mn} \operatorname{tr}\Upsilon_{aa}. \tag{2.A.3}$$

Here and in the following $\operatorname{tr}\cdots$ denotes a trace over the fermionic indices and ( ) imply symmetrization. We did not include invariants in which the scalar indices of S^A_{ab} are contracted with those of the Yukawas because thanks to (2.2.14) these can be written in terms of the first and second invariants of (2.A.3). Furthermore, we did not include structures of the form $\operatorname{tr}\Upsilon^{(AB)}_{aa}$ because $\operatorname{tr}\Upsilon^{A_1\cdots A_n}_{ab} = \operatorname{tr}\Upsilon_{ab}T^{A_n}\cdots T^{A_1}$ due to the trace transposition property. This relation will also be exploited later in enumerating the invariants of higher order.

None of the invariants in (2.A.3) is CP-odd. To see this note that the matrices Υ_{aa}, Υ^A_{aa} are hermitian and therefore their trace is real. This immediately tells us that the last two structures are real. Similarly, the structures $\operatorname{tr} T\Upsilon^A$ are necessarily CP-even because they are the trace of the product of two hermitian matrices, which is real. The first structure in (2.A.3) is CP-even for the same reason, because $T^{(A}T^{B)}$ is also hermitian.

Terms in $\beta^{(2)}$: Three-Loops Diagrams

At the next order, the relevant structures are:

$$
\begin{aligned}
g^2Y^2\lambda: \quad & \operatorname{tr} TT\Upsilon_0\lambda, \quad \operatorname{tr} T\Upsilon_1\lambda, \quad SS\operatorname{tr}\Upsilon_0\lambda, \\
& \operatorname{tr} TT\operatorname{tr}\Upsilon_0\lambda \\
g^4Y^2: \quad & \operatorname{tr} TT\Upsilon_2, \quad \operatorname{tr} TTT\Upsilon_1, \quad \operatorname{tr} TTTT\Upsilon_0 \\
& \operatorname{tr} TT\operatorname{tr}\Upsilon_2, \quad SS\operatorname{tr}\Upsilon_2, \quad \operatorname{tr} TT\operatorname{tr} T\Upsilon_1, \quad \operatorname{tr} TTT\operatorname{tr}\Upsilon_1, \quad SSS\operatorname{tr}\Upsilon_1, \\
& \operatorname{tr} TTTT\operatorname{tr}\Upsilon_0, \quad \operatorname{tr} TT\operatorname{tr} TT\operatorname{tr}\Upsilon_0, \quad SSSS\operatorname{tr}\Upsilon_0 \\
g^2Y^4: \quad & \operatorname{tr}\Upsilon_2\Upsilon_0, \quad \operatorname{tr}\Upsilon_1\Upsilon_1, \quad \operatorname{tr}\Upsilon_1\Upsilon_0T, \\
& \operatorname{tr}\Upsilon_2\operatorname{tr}\Upsilon_0, \quad \operatorname{tr}\Upsilon_1\operatorname{tr}\Upsilon_1, \quad \operatorname{tr} TT\operatorname{tr}\Upsilon_0\Upsilon_0, \quad SS\operatorname{tr}\Upsilon_0\Upsilon_0 \\
& \operatorname{tr}\Upsilon_0\operatorname{tr}\Upsilon_0\operatorname{tr} TT, \quad SS\operatorname{tr}\Upsilon_0\operatorname{tr}\Upsilon_0
\end{aligned}
\tag{2.A.4}
$$

where to make our notation more compact the n in the expression Υ_n indicates the number of adjoint indices in Υ. The indices are left implicit because many contractions are possible, and again invariants where one scalar index must be contracted with a Yukawa have been omitted because of (2.2.14).

$g^2Y^2\lambda$ Terms

Most of the $g^2Y^2\lambda$ terms are manifestly real once we take into account that the symmetry of λ forces the scalar indices in $[\Upsilon_n]_{ab}$ to be symmetrized or contracted among themselves, i.e. $\Upsilon_{(ab)}$ or Υ_{aa}. In either case the resulting Υ tensor is hermitian in the fermion indices. Hence all traces are inevitably real and the invariants CP-even.

g^4Y^2 Terms

Here the scalar indices are contracted among themselves; it is the possible contractions of the gauge indices that increases the number of independent structures. However all the invariants turn out to be CP-even because of properties of the gauge generators and gauge invariance. Let us show this explicitly by considering the invariants of the form $\operatorname{tr} TT\Upsilon_2$. These are

$$
\begin{aligned}
& g^2_{CD}\operatorname{tr} T^{(A}T^{B)}\Upsilon^{CD}_{aa} \\
& g^2_{CD}\operatorname{tr} T^{C}T^{D}\Upsilon^{(AB)}_{aa}
\end{aligned}
\tag{2.A.5}
$$

and

$$
\begin{aligned}
& g^2_{CD}\operatorname{tr} T^{C}T^{(A}\Upsilon^{B)D}_{aa} \\
& g^2_{CD}\operatorname{tr} T^{C}T^{(A|}\Upsilon^{D|B)}_{aa} \\
& g^2_{CD}\operatorname{tr} T^{(A|}T^{C}\Upsilon^{|B)D}_{aa} \\
& g^2_{CD}\operatorname{tr} T^{(A|}T^{C}\Upsilon^{D|B)}_{aa}
\end{aligned}
\tag{2.A.6}
$$

where the contraction among adjoint indices is consistently performed with the metric g^2_{AB}, (2.2.12). In the first two structures, hermiticity of $T^{(A}T^{B)}$ and $\Upsilon^{(MN)}_{aa}$ ensures they are all

CP-even. The other structures are also CP-even. To see this let us look for a CP-odd version of the first invariant in (2.A.6), namely

$$g^2_{CD}\,\mathrm{tr}\,T^{(A|}T^C\Upsilon^{|B)D}_{aa} - g^2_{CD}\,\mathrm{tr}\,T^C T^{(A|}\Upsilon^{D|B)}_{aa} \tag{2.A.7}$$

The indices C, D run over all possible adjoint indices. On the other hand, A, B are restricted to a certain non-abelian group or to any two of the abelian factors. Consider first the case A, B refer to a certain non-abelian group. Then $T^{A,B}$ commute with all the $T^{C,D}$ that are associated to the other groups. In addition, $\theta^{AB} \propto \delta^{AB}_{\mathcal{G}}$ and hence we should not simply symmetrize but actually sum over A, B as well. As a result the above invariant identically vanishes. Consider next the case in which A, B refer to the abelian groups. Then $T^{A,B}$ commute with $T^{C,D}$ and the expression again identically vanishes. Similar considerations show that all invariants in (2.A.6) are CP-even.

For invariants of the form $\mathrm{tr}\,TTT\Upsilon_1$, we have

$$\begin{aligned} &g^2_{CD}\,\mathrm{tr}\,T^{(A|}T^C T^D\Upsilon^{|B)}_{aa} \\ &g^2_{CD}\,\mathrm{tr}\,T^C T^D T^{(A}\Upsilon^{B)}_{aa} \end{aligned} \tag{2.A.8}$$

and

$$\begin{aligned} &g^2_{CD}\,\mathrm{tr}\,T^{(A|}T^C T^{|B)}\Upsilon^D_{aa}] \\ &g^2_{CD}\,\mathrm{tr}\,T^C T^{(A|}T^D\Upsilon^{|B)}_{aa} \\ &g^2_{CD}\,\mathrm{tr}\,T^{(A}T^{B)}T^C\Upsilon^D_{aa} \\ &g^2_{CD}\,\mathrm{tr}\,T^C T^{(A}T^{B)}\Upsilon^D_{aa} \end{aligned} \tag{2.A.9}$$

The expressions with the Casimir $g^2_{CD}T^CT^D = \oplus_{\mathcal{G}}C_{\mathcal{G}}$ are manifestly real. The first and second in (2.A.9) are real because they are the trace of the product of two hermitian matrices. A potential CP-odd combination with the third structure is

$$g^2_{CD}\,\mathrm{tr}\,\left\{T^A, T^B\right\}[T^C, \Upsilon^D_{aa}]. \tag{2.A.10}$$

An identical one is obtained from the last structure in (2.A.9). Again, these identically vanish when imposing gauge-invariance. Namely, if A, B refer to indices of a non-abelian group, then we should include a sum $\delta^{AB}_{\mathcal{G}}$ and find that $\left\{T^A, T^B\right\}$ is proportional to the Casimir $C_{\mathcal{G}}$ of that group. Hence it commutes with all T^C and we get $g^2_{CD}\,\mathrm{tr}\,C_{\mathcal{G}}[T^C, \Upsilon^D_{aa}] = g^2_{CD}\,\mathrm{tr}\,[C_{\mathcal{G}}, T^C]\Upsilon^D_{aa} = 0$. The same result trivially applies also when A, B refer to the abelian groups.

The arguments just exposed can be applied to all the other invariants appearing in (2.A.4), including the ones with the scalar generators. Therefore our conclusion is that all the g^4Y^2 structures are CP-even.

g^2Y^4 Terms

The structures g^2Y^4 are more involved because the scalar indices can be contracted in several different ways. For example $\mathrm{tr}\,\Upsilon_2\Upsilon_0$ can be $\mathrm{tr}\,\Upsilon^{(AB)}_{ab}\Upsilon_{ab}$, $\mathrm{tr}\,\Upsilon^{(AB)}_{ab}\Upsilon_{ba}$, $\mathrm{tr}\,\Upsilon^{(AB)}_{aa}\Upsilon_{bb}$. Here it is crucial to observe that $[\Upsilon_{ab}]^\dagger = \Upsilon_{ba}$ if Υ has less than two gauge indices, or even more provided they are symmetrized. This way one finds that $\operatorname{tr}\Upsilon_2\Upsilon_0$ are all CP-even. Similar considerations apply to $\operatorname{tr}\Upsilon_1\Upsilon_1$ as well as all the multi-trace expressions in (2.A.4). Including the generator T, however, changes the game and leads finally to some interesting candidates. We are thus left with the structure $\operatorname{tr}\Upsilon_1\Upsilon_0 T$. The CP-odd invariants of this form are

$$\begin{aligned}&\operatorname{tr}\Upsilon^A_{ab}[T^B,\Upsilon_{ab}]\\&\operatorname{tr}\Upsilon^{(A}_{ab}[T^{B)},\Upsilon_{ba}]\\&\operatorname{tr}\Upsilon^{(A}_{aa}[T^{B)},\Upsilon_{bb}]\end{aligned}\tag{2.A.11}$$

where the first invariant is already symmetric in the adjoint indices, as follows from the trace transposition property. The same property also shows that the second and third invariants are actually equivalent. Furthermore, using (2.2.14) it is straightforward to prove that $[\Upsilon_{aa}, T^A] = 0$, so the second and third expressions actually vanish. This ensures that the only non-vanishing CP-odd invariant in (2.A.11) is the first one, which is therefore the only possible contribution to the $\beta-$function of θ^{AB} at 3-loops. This is precisely $I^{AB}_{(2)}$ of eq. (2.2.23).

2.B Topological Angles and Weyl Consistency Conditions

The insertion of composite operators in correlator functions can be systematically described by introducing in the bare Lagrangian appropriate spacetime-dependent sources for them (and adding the appropriate counterterms). After having integrated out the dynamical fields one is left with a generating functional for the time-ordered, connected correlation functions of the renormalized operators. The local renormalization group [69–72] is a very powerful incarnation of this general prescription. In that approach the Lagrangian is expanded in a complete basis of operators $\mathcal{O}$ and all couplings of the theory become functions of spacetime, including the metric that sources the energy momentum tensor. A local version of the renormalization group equation for the operators $\mathcal{O}$ can be derived, and once all couplings are sent to their constant (x^μ independent but μ-dependent) this becomes exactly (2.2.8). Furthermore, Jack and Osborn pointed out that the beta functions in a general P-conserving renormalizable theory must satisfy a constraint

$$\frac{\partial\widetilde{A}}{\partial\xi_{\bar{I}}} = T_{\bar{I}\bar{J}}\beta_{\bar{J}},\tag{2.B.1}$$

where $\widetilde{A}$ and $T_{\bar{I}\bar{J}}$ are functions of the couplings (except θ) that appear in the Weyl anomaly when all couplings are promoted to space-time dependent functions. Such conditions relate terms of the beta functions of different couplings and different orders in perturbation theory, and can serve as non-trivial consistency checks of multi-loop calculations. For example, eq. (2.B.1) has been employed to resolve an ambiguity in the 4-loop beta function of the strong gauge coupling in the Standard Model [73]. The very same tool can potentially be used to extract information about the renormalization group evolution in general renormalizable theories like (2.2.12) (see e.g. [74]). In that case however one cannot a priori ignore the topological angles. Irrespective of whether they vanish in the bare action, they may be needed as counterterms and therefore generically possess a beta function that qualitatively impacts (2.B.1). According to [61], the formal structure of (2.B.1) remains unchanged when considering P-violating theories. The absence of an explicit dependence on θ then implies that its $\bar{I} = \theta$ component becomes

$$0 = T_{\theta\theta}\beta_\theta + T_{\theta Y}\beta_Y + T_{\theta g}\beta_g + T_{\theta\lambda}\beta_\lambda, \tag{2.B.2}$$

where θ, Y, g, λ schematically denote the couplings in (2.2.12). The same considerations developed in subsection 2.2.2 allow us to identify the tensorial form of $T_{\bar{I}\bar{J}}$. At leading order we have [6],

$$T_{\theta^{AB}\theta^{CD}} = c_1 \frac{g^2_{AC} g^2_{BD}}{(16\pi^2)^3} \tag{2.B.3}$$

and

$$\begin{aligned} T_{\theta^{AB}Y} dY = {} & c_2 \frac{g^2_{AC} g^2_{BD}}{(16\pi^2)^3} \operatorname{tr}\left[(Y^\dagger (T^C)^* dY - dY^\dagger (T^A)^* Y) T^D\right] \\ & + c_3 \frac{g^2_{AC} g^2_{BD}}{(16\pi^2)^3} \operatorname{tr}\left[(Y^\dagger dY - dY^\dagger Y)\left\{T^C, T^D\right\}\right] \end{aligned} \tag{2.B.4}$$

with $c_{1,2,3}$ numerical coefficients. Note that $T_{\theta Y} dY$ must be CP-odd. In extracting its form a key role is played by the fact that CP-odd quantities can here depend on the derivative of the couplings, as opposed to the discussion about the form of β_θ. Analogously, one can see that the CP-odd structures $T_{\theta g} dg$ and $T_{\theta\lambda} d\lambda$ inevitably arise at a higher perturbative order. Within a 2-loop accuracy the contributions proportional to β_g, β_λ can therefore be neglected and (2.B.2) gives [6],

$$\begin{aligned} \mu \frac{d\theta^{AB}}{d\mu} = {} & \frac{c_2/c_1}{16\pi^2} \operatorname{tr}\left[(Y_a^\dagger (T^A)^* \beta_{Y_a} - \beta^\dagger_{Y_a} (T^A)^* Y_a) T^B\right] \\ & + \frac{c_3/c_1}{16\pi^2} \operatorname{tr}\left[(Y_a^\dagger \beta_{Y_a} - \beta^\dagger_{Y_a} Y_a)\left\{T^A, T^B\right\}\right] \\ & + \mathcal{O}(\hbar^3). \end{aligned} \tag{2.B.5}$$

Plugging in the 1-loop beta function of the Yukawa coupling Y_a, the fermionic trace proportional to c_3 does not contribute at 2-loops whereas the one proportional to c_2 consistently reproduces (2.2.22) provided $2c_2/c_1 = c$. We thus find a fourth independent method for computing the coefficient c of β_θ. Establishing its numerical value would require knowing the leading order $T_{\theta\theta}$ and $T_{\theta Y}$, which means performing a 1-loop and again a 3-loop calculation respectively.

The remaining components of (2.B.1) are also quite consequential and deserve to be carefully explored. A naive counting suggests that, in employing the consistency relations of Jack and Osborn with $\bar{I} = Y$, the 2-loop β_θ identified in this paper might be correlated to structures appearing in the 3-loop β_g, the 4-loop β_Y, and the 2-loop β_λ via the 6-loop contribution in $\widetilde{A}$ of order $g^4 Y^6$. Furthermore, an inspection of the full system (2.B.1) reveals that the highest perturbative order at which the g-Y-λ-θ beta functions enter in the consistency relations are respectively 5-4-3-2.

Chapter 3

Axion

In this chapter we begin exploring the landscape of solutions to the Strong CP problem by starting with the axion mechanism. In sections 3.1 and 3.2 we review the basic principles of the axion solution and its associated quality problem. Section 3.3 is devoted to the study of a particular heavy axion model designed to tackle this latter issue, and is taken from the original work [7].

3.1 Peccei-Quinn Symmetry and the Axion Mechanism

The axion solution to the Strong CP problem exploits one of the spurious symmetry of the QCD angle mentioned in chapter 1, namely its shift symmetry, by promoting it to an actual symmetry of the theory except for the explicit breaking due to the anomaly itself. In the presence of this anomalous symmetry, the *Peccei-Quinn symmetry* [75,76], physics is invariant under an arbitrary shift of $\bar{\theta}$. With such a shift we can always set $\bar{\theta} = 0$, meaning that no strong CP violation is present[1]. The Strong CP problem is then effectively solved.

The first obvious way to implement such a symmetry would be having a massless quark, presumably the lightest up-quark. We already noticed in section 2.1.2 how in the presence of a massless fermion the topological angle becomes unphysical. Concretely, this is due to the fact that a massless up quark would imply the presence of an anomalous $U(1)_{\mathrm{PQ}}$ symmetry acting as

$$u_R \longrightarrow e^{i\alpha} u_R. \tag{3.1.1}$$

This transformation would leave the SM lagrangian invariant except for shifting $\bar{\theta} \to \bar{\theta} + \alpha$. Taking $\alpha = -\bar{\theta}$ would nullify $\bar{\theta}$, thus removing strong CP violation effects. Unfortunately,

[1] The request of the existence of an anomalous $U(1)_{\mathrm{PQ}}$ symmetry can actually be reformulated as the request that the theory is invariant under a generalized notion of CP, in which CP is combined with the aforementioned anomalous symmetry. Concretely, if $\bar{\theta}$ shifts of an angle α under a $U(1)_{\mathrm{PQ}}$ transformation, under CP$'$= CP $\times\, U(1)_{\mathrm{PQ}}$ it transforms as $\bar{\theta} \to -(\bar{\theta} + \alpha)$ (the minus sign due to ordinary CP, see eq. 2.2.18). Thus, for the angle choice $\alpha = -2(\bar{\theta} + k\pi), k \in \mathbb{Z}$ our theory is effectively CP$'$ invariant. One can then chose to work in a basis in which $\bar{\theta} = 0$, where CP$'$ coincides with ordinary CP and absence of strong CP violation is more manifest.

this minimal and appealing solution is not viable since the possibility for the up-quark to be massless is currently ruled out at more than 20 standard deviations [77]. Interestingly, thought, this result is quite recent. While at first sight it seems obvious that $m_u = 0$ leads to wrong predictions within the chiral lagrangian, the renormalized mass entering in this framework is actually the sum of two contributions: one from the QCD lagrangian, that is the usual m_u, and the other from non-perturbative contributions e.g. due to instantons [78–81]. The size of the latter is approximately $\sim m_d m_s/\Lambda_{\mathrm{C}}$, which numerically is just about the size required in order to get correct predictions within χPT. Ruling out $m_u = 0$ was possible only thanks to precise lattice studies which were not available until recently [82].

Being the massless quark solution not a viable option, the anomalous $U(1)_{\mathrm{PQ}}$ must then be implement by some beyond the Standard Model sector. If this symmetry gets spontaneously broken at a sufficiently high scale f_a the new sector decouples, leaving as a remnant a pseudo-Nambu-Goldstone boson named *axion*. Due to to the anomalous nature of the Peccei-Quinn symmetry, such particle inherits a coupling to the QCD topological term:

$$\mathcal{L} \supset \frac{a}{f_a}\frac{g^2}{32\pi^2}G\widetilde{G}. \tag{3.1.2}$$

This coupling explicitly breaks the shift symmetry of the axion[2] and shows the axion mechanism at work: a redefinition of the axion field $a/f_a = \bar{a}/f_a - \bar{\theta}$ leaves the theory invariant except for completely eliminating the QCD topological angle, thus "washing away" the Strong CP problem.

As of today, the axion mechanism is by far the most popular solution to the Strong CP problem. Its popularity can be understood as due to three good reasons:

i) *Simplicity.*
The axion solution just requires the existence of a global $U(1)_{\mathrm{PQ}}$, anomalous under QCD, which gets spontaneously broken at some high scale f_a. The minimality of these assumptions is envied by all alternative solutions to the Strong CP problem, which instead demand more complicated structures in the UV and are therefore viewed as less plausible.

ii) *Predictivity.*
Irrespectively of how the Peccei-Quinn symmetry is implemented in the UV, the axion solution clearly predicts the existence of a pseudo-Nambu-Goldstone boson with an irreducible coupling to gluons (3.1.2). This coupling leads to interesting phenomenological signatures that can be tested via a wide array of probes, including collider, astrophysical, and cosmological observations. It is also responsible for generating a potential for the axion when we hit the QCD confinement scale, as we will see in 3.1.1. Crucially, the potential is such that $\bar{a}$ is minimized at the CP-conserving value $\langle \bar{a} \rangle = 0$ (equivalently $\langle a \rangle / f_a = -\bar{\theta}$) as a result of the Vafa-Witten theorem [21]. This aspect is actually fundamental, since in the absence of a stabilizing mechanism a random vacuum expectation value for the axion may be generated effectively reintroducing an $\mathcal{O}(1)$ topological angle.

[2]The shift symmetry of the axion is the non-linear realization of the spontaneously broken original $U(1)_{\mathrm{PQ}}$ symmetry.

iii) Compatibility with generic BSM scenarios.
Whatever fundamental source of CP violation is present at short distances, when we run down to the QCD confinement scale it gets completely encoded in the QCD topological angle and higher-dimensional LEFT operators[3]. The axion solution becomes "active" at that very scale, where its potential is generated, and by relaxing $\bar{\theta}$ to zero it completely removes the largest source of CP violation within the low energy effective field theory. This remarkable property is particularly appealing because in presence of the QCD axion fundamental questions such as the naturalness problem, the flavour puzzle, baryogenesis, dark matter, etc., may all be addressed without the need to worry about possibly large CP-odd phases in the new physics' sector.

In the following we will not be particularly interested in how effectively the PQ-symmetry is implemented, as long as the coupling (3.1.2) is present. However, for completeness we mention the two main historical implementations due to Kim-Shifman-Vainshtein-Zakharov (KSVZ) and Dine-Fischler-Srednicki-Zhitnitsky (DFSZ). For a more general overview of the possible scenarios see [84].

Beyond the SM matter content, the "minimal" KSVZ axion [39, 85] assumes the existence of a vector-like quark Ψ, neutral under the weak interactions, and of a complete singlet complex scalar field Φ . The most general renormalizable lagrangian compatible with a $U(1)_{\rm PQ}$ symmetry reads

$$\mathcal{L}_{\rm KSVZ} = \mathcal{L}_{\rm SM} + |\partial_\mu \Phi|^2 + \bar{\Psi} i \not{D} \Psi - y\, \Phi \bar{\Psi}_L \Psi_R + \text{h.c.} - V(|\Phi|, |H|) \tag{3.1.3}$$

where the $U(1)_{\rm PQ}$ is implemented as $\Psi_{L,R} \to e^{\pm i\alpha/2}\Psi_{L,R}, \Phi \to e^{i\alpha}\Phi$, clearly leaving (3.1.3) invariant except for shifting $\theta \to \theta - \alpha$. The Peccei-Quinn symmetry is spontaneously broken by the vacuum expectation value $\langle \rho \rangle = f_a$ of the field $\Phi = \rho\, e^{ia/f_a}$, which leads to a mass term $M_\Psi = y f_a$ for the vector-like fermion and to the massless NGB a, identified with the axion. With a chiral rotation $\Psi_{L,R} \to e^{\pm ia/2}\Psi_{L,R}$ the axion can be moved from the Ψ mass term to the QCD topological term, leading to a low-energy effective theory comprised of the SM plus an axion with the sole coupling (3.1.2).

The DFSZ model [86, 87] is basically a two-Higgs doublets (H_u, H_d) model supplemented by a complete singlet complex scalar fields Φ. There are many possible variations depending on which Higgs doublet couples to the leptons (H_d or $\tilde{H}_u = \epsilon H_u^*$) and to the scalar potential couplings between the doublets and Φ. The Peccei-Quinn charges are carried by H_u, H_d, Φ and the SM quarks and leptons, and the resulting axion is in general a combination of all the P-odd neutral scalar fields. The identification of the correct combination is straightforward and can be found e.g. in [84]. Because in this case the Peccei-Quinn charges are carried also by the SM fermions, an irreducible coupling of the axion also to the $U(1)_{\rm EM}$ topological term is present. This leads to the well-know ratio $E/N = 8/3$, where the coupling to photons is written as $(e^2/16\pi^2)(E/N)(\bar{a}/f_a)F\widetilde{F}$ (in the minimal KSVZ model $E/N = 0$). Interestingly,

[3] "Low Energy Effective Theory", that is the theory obtained integrating out the heavy SM weak bosons, the Higgs particle and the top quark. The remaining theory contains non-renormalizable operators suppressed by the heavy mass scales. The LEFT itself may inherits higher-dimensional operators directly from the SMEFT, obtained integrating out heavy BSM physics. For a complete classification of the LEFT operators up to dimension six and for the matching to the SMEFT see [83].

the DFSZ model can be seen as an extension of the very first proposal of an axion model, the Weinberg-Wilczek (WW) model [23, 88]. In this proposal no singlet Φ is present, but a PQ symmetry is still there thanks to the two Higgs doublets. However, since the Peccei-Quinn scale turns out to be about the same order of the electroweak scale, $f_a \sim v$, the axion couplings to the SM particles are not sufficiently suppressed and this model was soon ruled by laboratory searches [89–92]. This is the reason for which models in which f_a is unrelated to v are often called "invisible axion" models (contrary to the visible, but not found, WW axion).

3.1.1 QCD Axion Potential

Irrespectively of the UV completion delivering the axion, the axion effective theory always contains the irreducible coupling (3.1.2). Such coupling is responsible for generating a potential that, as predicted by the Vafa and Witten theorem [21], dynamically relaxes $\langle \bar{a} \rangle = 0$. This can be showed by observing that, at leading order in $1/f_a$, the axion can simply be treated as an external source. Therefore, the argument of [21] goes through basically unchanged except for replacing $\bar{\theta} \to \bar{a}/f_a$. The vacuum energy as given by the Euclidean path integral still satisfies [21]

$$e^{-VE\left(\frac{\bar{a}}{f_a}\right)} = \int d\mu\, e^{i\frac{g^2}{32\pi^2}\int d^4x_E \frac{\bar{a}}{f_a} G\widetilde{G}} \leq \left| \int d\mu\, e^{i\frac{g^2}{32\pi^2}\int d^4x_E \frac{\bar{a}}{f_a} G\widetilde{G}} \right| = e^{-VE(0)} \tag{3.1.4}$$

where $d\mu$ is the effective measure of the QCD Euclidean path integral, given by $d\mu = \det(\not{D} + M)\exp(-\int d^4x_E\, GG/4)$. Crucially, this is positive definite [22] and hence the absolute value can be brought inside the functional integral. Being the integrand a pure phase in the Euclidean formulation (as showed in 2.1.1), it has unit modulus and this concludes the proof of 3.1.4. The inequality $E(\bar{a}/f_a) \geq E(0)$ then ensures that the axion potential is indeed minimized at the CP-conserving value $\langle \bar{a} \rangle = 0 \bmod 2\pi$.

Concretely, a crude approximation to the axion potential is given by the one-instanton energy dependence of the vacuum on the topological angle (2.1.38), simply replacing $\bar{\theta} \to \bar{a}/f_a$. Its cosine shape confirms that $\langle \bar{a} \rangle = 0$. The axion mass can be obtained as

$$m_a^2 = \frac{\delta^2}{\delta \bar{a}^2} \log \mathcal{Z}\left(\frac{\bar{a}}{f_a}\right)\Big|_{\bar{a}=0} = \frac{1}{f_a^2}\frac{d}{d\bar{\theta}^2} \log \mathcal{Z}(\bar{\theta})\Big|_{\bar{\theta}=0} \equiv \frac{\chi_{\text{top}}}{f_a^2} \tag{3.1.5}$$

where we defined the QCD *topological susceptibility* χ_{top}, explicitly given by

$$\chi_{\text{top}} = \int d^4x\, \langle \frac{g^2}{32\pi^2} G\widetilde{G}(x), \frac{g^2}{32\pi^2} G\widetilde{G}(0) \rangle\,. \tag{3.1.6}$$

The quantity χ_{top} is very useful because it can be computed by means of chiral lagrangian techniques and non-perturbative lattice studies [93]. It provides a good approximation to the coefficient of the quadratic term in the axion potential, and most importantly fixes the distinctive scaling of the axion mass with its decay constant:

$$m_a^2 f_a^2 = \chi_{\text{top}}. \tag{3.1.7}$$

Being the topological susceptibility a calculable constant, this relation essentially leaves the QCD axion with only one free parameter.

The approximation (2.1.38) to the full potential of the axion is quite poor, since it is obtained in a semi-classical regime which is nowhere close to the real QCD dynamics. A much better approximation can be obtained within the chiral lagrangian framework. To this end we first remove the coupling (3.1.2) by means of the chiral rotation $q_R \to e^{-i\bar{a}/N_f f_a} q_R$, which leaves us with the lagrangian

$$\mathcal{L} = \mathcal{L}_{\text{kin}} - \bar{q}_L M e^{-i\bar{a}/N_f f_a} q_R + \text{h.c.}. \tag{3.1.8}$$

At this point we can, as usual, expand the chiral lagrangian in terms of the NGBs matrix $U(\Pi)$ (2.1.42) and the mass $M_a = M e^{-i\bar{a}/N_f f_a}$, which is now axion-dependent. The leading order potential, taking $N_f = 2$ for simplicity, reads

$$\begin{aligned} V^{\text{LO}}(\bar{a}, \pi_0) &= -\frac{f_\pi^2}{4} B_0 \,\text{tr}\left(U^\dagger M_a + U {M_a}^\dagger\right) \\ &= -\frac{f_\pi^2}{2} B_0 \left[m_d \cos\left(\frac{\bar{a}}{2f_a} + \frac{\pi_0}{f_\pi}\right) + m_u \cos\left(\frac{\bar{a}}{2f_a} - \frac{\pi_0}{f_\pi}\right)\right] \end{aligned} \tag{3.1.9}$$

where we neglected the dependence on the charged NGBs $\pi^\pm$, since they are irrelevant for the potential of the neutral fields as they cannot mix. The potential (3.1.9) can be rewritten as

$$\begin{aligned} V^{\text{LO}}(\bar{a}, \pi_0) &= -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4 m_u m_d}{(m_u + m_d)^2} \sin^2\left(\frac{\bar{a}}{2f_a}\right)} \cos\left(\frac{\pi_0}{f_\pi} - \phi_a\right) \\ \tan\phi_a &\equiv \frac{m_u - m_d}{m_u + m_d} \tan\left(\frac{\bar{a}}{2f_a}\right) \end{aligned} \tag{3.1.10}$$

which clearly shows that the π^0 vev is proportional to ϕ_a. For $f_a \ll f_\pi$ we can integrate out π^0, which at tree-level simply amounts to taking $\pi^0/f_\pi = \phi_a$. In this way we are left with the effective axion potential

$$V_{\text{eff}}^{\text{LO}}(\bar{a}) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4 m_u m_d}{(m_u + m_d)^2} \sin^2\left(\frac{\bar{a}}{2f_a}\right)}. \tag{3.1.11}$$

This potential is minimized at $\langle \bar{a} \rangle = 0$, as expected, and provides the tree-level expression for the axion mass

$$m_a^2 = \frac{m_u m_d}{(m_u + m_d)^2} \frac{m_\pi^2 f_\pi^2}{f_a^2} \Rightarrow m_a \simeq 5.7 \times \left(\frac{10^{12}\ \text{GeV}}{f_a}\right)\ \mu\text{eV}. \tag{3.1.12}$$

This value of the axion mass is consistent with (3.1.5) employing the most recent value $\chi_{\text{top}}^{1/4} = 75.5(5)$ MeV [93]. It also highlights two important features. First, in the limit[4] $m_d \gg m_u$, the axion mass is suppressed by a power of m_u. This is in accordance with the fact that

[4] Of course also the opposite limit is valid. Indeed the ratio $m_u m_d/(m_u + m_d)^2$ can be written as $\det M/(\,\text{tr}\, M)^2$, which vanished whenever a quark becomes massless. As usual m_u is used an example since experimentally the up-quark is found to be the lightest quark.

for $m_u = 0$ the shift symmetry of the axion becomes *exact*, since an axion shift can be entirely compensated by a chiral rotation of the massless up-quark. In other words, a non-anomalous $U(1)_{\rm PQ}$ is there and broken spontaneously only by the QCD confinement: the axion becomes an exact Nambu-Goldstone boson. Second, in the range $f_a \gtrsim 10^8$ GeV the axion is an *extremely light* pseudoscalar, even lighter than neutrinos. Therefore, experiments looking for axions are quite peculiar since its long de Broglie wavelenght makes it behave almost like a classical wave for which resonance cavities, for instance, are well-fit probes. We will not dive into the details of the various past and ongoing experiments looking for the QCD axion. A nice review can be found in [84]. Figure 3.1, taken from [94], shows the most recent experimental constraints coming from axion searches. These set a lower bound on the QCD axion decay constant given by $f_a \gtrsim 10^8 - 10^9$ GeV. Notably, the value $f_a \simeq 10^{12}$ GeV is quite peculiar since in this range the axion can naturally account for all the dark matter present in our universe via the *misalignment mechanism* [95–97].

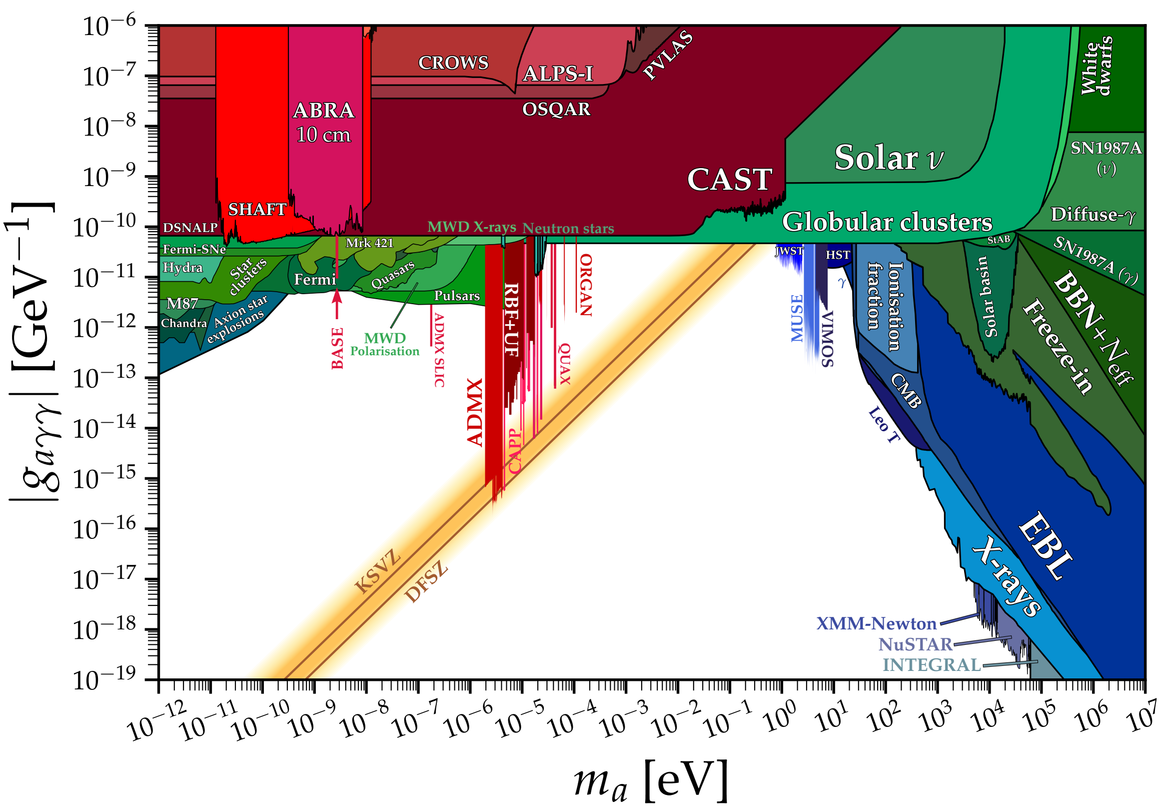

Figure 3.1: Collection of the most recent experimental constraints from axion searches in the plane $m_a - g_{a\gamma\gamma}$, where $g_{a\gamma\gamma}$ is the axion coupling to photons written as $\mathcal{L} \supset g_{a\gamma\gamma}\, aF\widetilde{F}/4$. This is related to f_a via $g_{a\gamma\gamma} = (E/N)e^2/8\pi^2 f_a$, where E/N is a constant depending on the specific UV completion of the axion. The yellow band refers to the QCD axion prediction (3.1.12) with some benchmark values for E/N (see [84] for details). Plot taken from [94].

3.2 Quality Problem

The axion solution to the Strong CP problem is apparently elegant, simple and predictive. However, there is a critical aspect that we have been neglecting so far. The entire mechanism relies crucially on the presence of global $U(1)_{\rm PQ}$ symmetry which must be exact to a very high degree (that we will quantify in a moment), except for being anomalous. Any operator apart from (3.1.2) explicitly violating this symmetry would generate additional contributions to the axion potential and destabilize its vev from the CP-conserving value $\langle \bar{a} \rangle = 0$, thus effectively reintroducing $\bar{\theta}$ and the associated Strong CP problem. From a quantum field theory perspective, global symmetries should be regarded as accidental and this particularly applies to anomalous symmetries, which cannot be consistently promoted to local symmetries at the quantum level. Therefore we expect that, if not forbidden for other reasons, generic operators violating by p units the PQ-charge conservation should be present. These may be generated as due to quantum gravity effects, which are conjectured to violate any symmetry which is not gauged [98], or more agnostically as due to whatever UV completion is there that, according to the totalitarian principle[5], should generate these operators if allowed.

To quantify the sensitivity of the axion mechanism to these operators, let us consider an explicit realization à la KSVZ where the axion appears as the phase of a complex $U(1)_{\rm PQ}$ field. A generic lagrangian operator of dimension n violating the PQ-charge by p units may be written as (assuming $n-p$ to be even)

$$\mathcal{L}_{\not{\rm PQ}} = \frac{\lambda_*}{2 f_{\rm UV}^{n-4}} |\Phi|^{n-p} \Phi^p + \text{ h.c.} \tag{3.2.1}$$

This operator induces a shift in the effective potential (3.1.11) which takes the form

$$\delta V(\bar{a}) = |\lambda_*| \frac{f_a^n}{f_{\rm UV}^{n-4}} \cos\left(p\bar{a} + \delta_{\lambda_*}\right) \tag{3.2.2}$$

where $\delta_{\lambda_*} = \arg \lambda_*$ is the phase responsible of destabilizing the minimum of the axion from $\langle \bar{a} \rangle = 0$. Without the need of explicitly minimizing the full potential, requiring that the Strong CP problem is still solved, $|\bar{\theta}_{\rm eff}| = |\langle \bar{a} \rangle|/f_a \lesssim 10^{-10}$, can be roughly translated in the constraint[6]

$$\left| \frac{\delta V(\bar{a}=0)}{V_{\rm eff}(\bar{a}=0)} \right| \lesssim 10^{-10}. \tag{3.2.3}$$

Taking the QCD axion effective potential to be parametrically of the size $V_{\rm eff}(\bar{a}) \sim f_\pi^2 m_\pi^2$ and $\delta V(\bar{a}) \sim \lambda_* f_a^n / f_{\rm UV}^{n-4}$, for generic $\delta_{\lambda_*} \sim \mathcal{O}(1)$ the above inequality can be turned into a lower bound on the dimension n of the PQ-breaking operator:

$$n - 4 \gtrsim \frac{\log\left(10^{10} |\lambda_*| f_a^2 / m_a^2\right)}{\log\left(f_{\rm UV}/f_a\right)}. \tag{3.2.4}$$

[5] "Everything not forbidden is compulsory", typically associated to M. Gell-Mann [99] (although there are some debates on the paternity of the principle as applied to quantum physics [100]).

[6] Around the unperturbed minimum $\langle \bar{a} \rangle = 0$, the potential can be approximated as $V(\bar{a}) \simeq m_a^2 \bar{a}^2/2 + \delta V'|_{\bar{a}=0} \bar{a} + \dots$. Therefore, at leading order $\langle \bar{a} \rangle \sim \delta V'|_{\bar{a}=0}/m_a^2 \sim [|\lambda_*| \sin\delta_{\lambda_*} f_a^{n-1}/f_{\rm UV}^{n-4}]/[f_\pi^2 m_\pi^2/f_a^2]$ and thus $\bar{\theta}_{\rm eff} \sim [|\lambda_*| \sin\delta_{\lambda_*} f_a^n / f_{\rm UV}^{n-4}]/[f_\pi^2 m_\pi^2] \sim \delta V(\bar{a}=0)/V_{\rm eff}(\bar{a}=0)$.

Employing the lower bound $f_a \gtrsim 10^9$ GeV and taking as UV the scale of quantum gravity $f_{\mathrm{UV}} = M_{\mathrm{P}}$[7] with $\lambda_* = 1$, we get the bound

$$n \gtrsim 10 \qquad \begin{pmatrix} f_a = 10^9 \text{ GeV} \\ f_{\mathrm{UV}} = M_{\mathrm{P}} \end{pmatrix} \tag{3.2.5}$$

Therefore, operators of dimension 9 or lower must be forbidden in order for the QCD axion to solve the Strong CP problem: the anomalous $U(1)_{\mathrm{PQ}}$ must be of extremely high *quality*. This serious issue, often referred to as the *axion quality problem* [101–105], represents a major drawback of the simple QCD axion solution and essentially reintroduces a fine-tuning that, if not addressed, is just a more sophisticated version of the original Strong CP problem. Ultimately, this is due to the smallness of the QCD axion potential as compared to contributions from additional uncontrollable sources of $U(1)_{\mathrm{PQ}}$-breaking, which are enhanced as a result of the current constraints on f_a being so stringent. The standard axion solution is thus apparently a very delicate one. In the next section we will discuss how this problem may be alleviated within quantum field theory and study in detail the particular proposal of the original work [7], from which the bulk of the section and appendix 3.A are taken. However, it is worth mentioning that the quality problem may actually be solved within quantum gravity itself. For example, for fundamental axions such as the ones arising from string theory compactifications [106], the axion shift symmetry is perturbatively exact as a result of the higher-form gauge symmetry associated to its dual field. Non-perturbative breaking is typically exponentially suppressed by the classical action, and in cases where the effect is calculable as in that of Euclidean wormholes [107, 108], the suppression may be enough to avoid the associated quality problem if a sufficient number of axions is present [109] (see also [110] for an independent UV result). Nevertheless, without a complete control over the quantum theory of gravity at play these should be regarded only as special points that under no circumstances are to be preferred over the ones where the problem is actually present, unless an explanation to exclude the latter is found.

3.3 Heavy Axions and Grand Color

The QCD axion solution to the Strong CP problem is seriously threatened by the Peccei-Quinn quality problem. There are two possible paths to address this issue within the four-dimensional QFT realm. One can either find a mechanism to suppress the potentially dangerous Planckian perturbations, or find new corrections to the potential such that the QCD contribution gets effectively enhanced. In either case, unfortunately, one or both of the attractive properties $i), iii)$ of section 3.1 are lost.

Known theories that address the quality problem by suggesting mechanisms to suppress quantum gravity perturbations typically invoke a number of ad-hoc global or gauge symmetries such that the PQ symmetry is accidentally protected up to operators of very high

[7]The Planck mass appearing here is the reduced Planck mass, $M_{\mathrm{P}} = 1/\sqrt{8\pi G} = 2.435 \times 10^{18}$ GeV. This is the scale defining perturbation theory around $g_{\mu\nu} = \eta_{\mu\nu}$ for the Einstein-Hilbert theory, and has the correct dimensionality to keep $[S] = \hbar$ in (3.2.1).

dimension. If the dimension is high enough, the induced effective potential (3.2.2) is so suppressed that the resulting $|\bar{\theta}_{\text{eff}}|$ lays well below the experimental constraint, obliteraing the axion quality problem. Many incarnations of this idea have been proposed, either relying on large discrete symmetries [111–116], which can be made "local" as first suggested in [117], or on new gauge $U(1)$ [103, 105, 118–121] or non-Abelian symmetries [101, 122–124]. It is also possible to achieve such protection in models where the axion is a composite pNGB of some new confining dynamics [125–132]. Whatever the realization, this approach to the quality problem relies on such an intricate structure at short distances that the conceptual simplicity promised by the QCD axion gets partially if not completely overshadowed.

Rather than explaining why Planck-scale perturbations are suppressed compared to the QCD-induced potential, one may alternatively build a model with a heavy axion, in which the QCD potential is replaced by a much larger and more stable one. Here the challenge is to identify a framework in which the new $U(1)_{\text{PQ}}$-breaking effects are perfectly aligned with the QCD-induced potential so that the overall energy is still minimized at a value $|\langle \bar{a} \rangle|/f_a \lesssim 10^{-10}$, compatibly with current data. One option is realized in scenarios where the large axion potential still comes from QCD but, as opposed to the standard mechanism, it is due to new short-distance effects [133]. If the QCD coupling grows relatively large at some UV scale, indeed, small instantons may become relevant, and such effects are naturally aligned with the low energy QCD potential. Unfortunately, a strongly-coupled UV framework of this type is intrinsically sensitive to the misaligning effect of whatever new CP-odd phases are present at the UV cutoff [102, 134]. To firmly establish the viability of this program it is thus necessary to analyze a concrete realization. Currently, the only explicit and tractable model of this type is the one of [135]. This work shows that under reasonable assumptions the Strong CP problem may in fact be solved, though the required setup introduces a few copies of the color gauge group along with a corresponding axion for each copy, and is therefore not as minimal as one might have hoped. Another viable avenue is to postulate a scenario in which $U(1)_{\text{PQ}}$ is anomalous under an additional non-abelian group C'. Provided the anomaly coefficient is the same as the one of QCD, the axion coupling to the two sectors reads

$$\mathcal{L}_{\text{axion}} \supset \left(\bar{\theta} + \frac{a}{f_a}\right) \frac{g_{\mathrm{C}}^2}{32\pi^2} G^a_{\mu\nu} \widetilde{G}^{a\,\mu\nu} + \left(\bar{\theta}' + \frac{a}{f_a}\right) \frac{g_{\mathrm{C}'}^2}{32\pi^2} G'^a_{\mu\nu} \widetilde{G}'^{a\,\mu\nu}, \tag{3.3.1}$$

with $G'^a_{\mu\nu}$ indicating the field strength of the C' vectors. If it is possible to further identify a structural condition that ensures

$$\bar{\theta} = \bar{\theta}' \tag{3.3.2}$$

up to corrections smaller than 10^{-10}, then a unique axion $\bar{a}/f_a = \bar{\theta} + a/f_a$ can be defined. Its potential may be naturally dominated by the C' dynamics and be such that $\langle \bar{a} \rangle = 0$, analogously to QCD. In this framework one can obtain a sizable axion potential if the new non-abelian sector becomes strong at scales $\Lambda_{\mathrm{C}'} \gg \Lambda_{\mathrm{C}}$, and the quality problem is improved. The non-trivial task is explaining (3.3.2).

We may justify (3.3.2) by invoking a Z_2 symmetry [136]. To realize this program a full copy of the SM is however needed, and in particular the new confining group must be a mirror copy of QCD, i.e. $\mathrm{C}' = SU(3)_{\mathrm{C}'}$. The mirror symmetry must be softly broken in order to ensure that the mirror sector be sufficiently heavy to have escaped detection. If the soft breaking is achieved via CP- and flavour-conserving interactions, any possible correction to (3.3.2) is controlled by loops of the SM Yukawas and higher-dimensional operators. We already stressed in chapter 2 how the former corrections are extremely small, while the latter can be taken under control as well provided the soft breaking scale is sufficiently small compared to the UV cutoff. These mirror models are currently the most studied incarnation of heavy axion models [137–141].

Yet, there may be a simpler and more minimal way to justify (3.3.2) that does not require invoking a discrete mirror symmetry. One may in fact embed color $SU(3)_{\mathrm{C}}$ into a larger *Grand Color* group at short distances, and then postulate the latter be broken into the SM times the new confining group C'. In this setup the structure of eq. (3.3.1) emerges at the symmetry-breaking threshold, with (3.3.2) easily satisfied at tree-level even when C' is not an $SU(3)$. This class of models was first suggested in [142], without however providing a concrete realization. Explicit models have been recently proposed in [143, 144], but slightly deviate from the idea of [142] due to the presence of mass terms for the exotic fermions. The latter make it difficult to ensure (3.3.2) remains protected against radiative effects. In the following we will propose and study an explicit realization of the Grand Color scenario that does not require the introduction of ad-hoc mass terms and robustly satisfies (3.3.2). Anticipating later results, in our model the axion's mass turns out to have a very distinctive scaling [7]:

$$m_a^2 \sim \frac{y_u y_d}{N} \frac{f^4}{f_a^2} \tag{3.3.3}$$

where $y_{u,d}$ are the up- and down-quark Yukawas renormalized by the new confining dynamics $\mathrm{C}' = Sp(N-3)$ at $\Lambda_{\mathrm{C}'} \sim 4\pi f/\sqrt{N}$. The choice $C' = Sp(N-3)$ is actually crucial to obtain a realistic model of electroweak symmetry breaking, as explained in greater detail in section 3.3.2. Expression (3.3.3) differs from the one predicted by existing heavy axion models and its magnitude falls somewhat in between mirror and UV-instanton models mentioned before. It is significantly enhanced compared to the one predicted by potentials dominated by small instantons. As a result, an improvement in axion quality is achieved with a significantly smaller f, and hence a reduced sensitivity to physics at the cutoff scale. The scaling in (3.3.3) is however suppressed compared to what is found in Z_2-symmetric models, and so for a similar axion mass our f needs to be larger.

3.3.1 A Grand Color Group

The gauge group of the model we consider is $SU(N)_{\mathrm{GC}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}'}$ and the entire matter content, SM included, is reported in table 3.1. The Grand Color is an $SU(N)_{\mathrm{GC}}$ gauge group and the SM quarks are in the fundamental and anti-fundamental representations. In order to cancel gauge anomalies, hypercharge must be partly embedded into the Grand Color and an abelian factor $U(1)_{\mathrm{Y}'}$, while to avoid triviality N must be odd [145]. Yet, the leptonic sector remains basically the same as in the SM, whereas the scalar sector must include at least two additional fields, Φ in the adjoint and Ξ in the 2-index anti-symmetric of $SU(N)_{\mathrm{GC}}$, in order to break Grand Color in a phenomenologically viable way.

	$SU(N)_{\rm GC}$	$SU(2)_{\rm L}$	$U(1)_{\rm Y'}$
Q	$\mathbf{N}$	$\mathbf{2}$	$\frac{1}{2N}$
U	$\overline{\mathbf{N}}$	$\mathbf{1}$	$-\frac{1}{2}-\frac{1}{2N}$
D	$\overline{\mathbf{N}}$	$\mathbf{1}$	$+\frac{1}{2}-\frac{1}{2N}$
ℓ	$\mathbf{1}$	$\mathbf{2}$	$-\frac{1}{2}$
e	$\mathbf{1}$	$\mathbf{1}$	$+1$
H	$\mathbf{1}$	$\mathbf{2}$	$+\frac{1}{2}$
Φ	$\mathbf{Adj}$	$\mathbf{1}$	0
Ξ	$\mathbf{N}\otimes_A\mathbf{N}$	$\mathbf{1}$	$\frac{1}{N}$

Table 3.1: Minimal field content of the model [7]. The fermions are Weyls'. The scalars Φ, Ξ are solely needed in order to break Grand Color into the SM gauge group.

The most general renormalizable Lagrangian for the fields in table 3.1 includes the standard kinetic terms and topological angles, a scalar potential, and a Yukawa interaction with the Higgs doublet H of the same form as in the SM[8]

$$\mathcal{L}_{\rm Yuk} = Y_u^*\, QHU + Y_d^*\, Q\widetilde{H}D + Y_e^*\, \ell\widetilde{H}e + {\rm hc}, \tag{3.3.4}$$

plus the operators $QQ\Xi^\dagger$ and $UD\Xi$. As explained in more detail later, though, the presence of the latter interactions would spoil the key relation (3.3.2). These couplings can be forbidden in several ways, for example gauging B-L, promoting Ξ to a composite scalar, or — perhaps less elegantly — invoking a global symmetry. Which of these mechanisms is actually at work is not a concern. In the following we will simply assume that (3.3.4) represent the full set of renormalizable Yukawa interactions.

The breaking of Grand Color is obtained in two steps [7]:

$$\begin{aligned} SU(N)_{\rm GC}\times SU(2)_{\rm L}\times U(1)_{\rm Y'} &\xrightarrow{\langle\Phi\rangle} SU(3)_{\rm C}\times SU(N-3)\times SU(2)_{\rm L}\times U(1)_{\rm Y'}\times U(1)_{\rm GC} \\ &\xrightarrow{\langle\Xi\rangle} SU(3)_{\rm C}\times Sp(N-3)\times SU(2)_{\rm L}\times U(1)_{\rm Y}. \end{aligned} \tag{3.3.5}$$

In the first step the vev of a scalar Φ breaks $SU(N)_{\rm GC}$ into $SU(3)_{\rm C}\times SU(N-3)\times U(1)_{\rm GC}$. The abelian factor is normalized such that the fundamental representation of $SU(N)_{\rm GC}$ decomposes as

$$\mathbf{N}\to(\mathbf{3},\mathbf{1})_{\frac{1}{6}-\frac{1}{2N}}\oplus(\mathbf{1},\mathbf{N-3})_{-\frac{1}{2N}}. \tag{3.3.6}$$

The second step consists in breaking $SU(N-3)\times U(1)_{\rm GC}\times U(1)_{\rm Y'} \xrightarrow{\langle\Xi\rangle} Sp(N-3)\times U(1)_{\rm Y}$ through the vev of the Ξ component in the antisymmetric of $SU(N-3)$, which according to

[8]The conjugation in the definition of the Yukawas is required in order to be consistent with (1.1.2) (and thus with the convention of [1]).

(3.3.6) carries a $U(1)_{\rm GC}$ charge equal to $-1/N$[9]. It follows that the unbroken $U(1)_{\rm Y}$ charges are the sum of $U(1)_{\rm Y'}$ and the $U(1)_{\rm GC}$ generators. For simplicity we take both scalar vevs of order $f_{\rm GC}$. Importantly, because none of the new scalars Φ, Ξ has Yukawa couplings one can in principle promote both of them to composite operators. In that case there would no hidden fine-tuning in requiring the Grand Color breaking scale be much smaller than the UV cutoff, i.e. $f_{\rm GC} \ll f_{\rm UV}$.

Below the Grand Color breaking scale $f_{\rm GC}$, the fields Q, U, D split into the direct sum of the SM quarks plus exotic chiral fermions as shown in table 3.2. The exotic fermions $\psi_{q,u,d}$ inherit the Yukawa couplings to H from (3.3.4) and are therefore formally the same as the SM ones up to renormalization effects. Crucially, however, there is no interaction between the SM fermions and the ψ's apart from higher-dimensional operators suppressed by $f_{\rm GC}$. This implies that the flavour symmetries of the two sectors are effectively distinct: loops of the $Sp(N-3)$-charged sector will never be able to induce flavour-violating processes in the SM. In addition, the field-basis invariant $SU(3)_{\rm C}$ and $Sp(N-3)$ topological angles, inherited

		$SU(3)_{\rm C}$	$Sp(N-3)$	$SU(2)_{\rm L}$	$U(1)_{\rm Y}$
$Q = \begin{pmatrix} q \\ \psi_q \end{pmatrix}$	q	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{2}$	$\frac{1}{6}$
	ψ_q	$\mathbf{1}$	$\mathbf{N-3}$	$\mathbf{2}$	0
$U = \begin{pmatrix} u \\ \psi_u \end{pmatrix}$	u	$\bar{\mathbf{3}}$	$\mathbf{1}$	$\mathbf{1}$	$-\frac{2}{3}$
	ψ_u	$\mathbf{1}$	$\mathbf{N-3}$	$\mathbf{1}$	$-\frac{1}{2}$
$D = \begin{pmatrix} d \\ \psi_d \end{pmatrix}$	d	$\bar{\mathbf{3}}$	$\mathbf{1}$	$\mathbf{1}$	$\frac{1}{3}$
	ψ_d	$\mathbf{1}$	$\mathbf{N-3}$	$\mathbf{1}$	$\frac{1}{2}$

Table 3.2: Decomposition of the quarks below the scale $f_{\rm GC}$. Here $\psi_q = (\psi_{q_u}, \psi_{q_d})$ is an electroweak doublet. The SM hypercharge $U(1)_{\rm Y}$ is the sum of $U(1)_{\rm Y'}$ and $U(1)_{\rm GC} \subset SU(N)_{\rm GC}$.

by Grand Color as shown in (3.3.1), at tree-level satisfy $\bar\theta = \bar\theta' = \theta - \arg\det Y_u Y_d$, where θ denotes the $SU(N)_{\rm GC}$ angle. Radiative effects can spoil this tree-level relation, and it is mandatory for us to show that the misaligning affects are under control. There are three different sources of radiative effects that can potentially invalidate (3.3.2): the scalar sector, the Yukawa couplings, and non-renormalizable interactions. The vevs of Φ and Ξ are the order parameters of Grand Color breaking and their insertion is necessary to generate a difference in the two topological angles. Other than that, however, the scalar sector cannot appreciably contribute to a violation of (3.3.2) since the most general renormalizable potential $V(H, \Phi, \Xi)$ is automatically CP-conserving and its parameters can always be chosen so that CP does not get broken spontaneously. Furthermore, all radiative corrections due to the Yukawa sector (3.3.4) at and below $f_{\rm GC}$ are known to be completely negligible [52–54]. Had we allowed the presence of unsuppressed flavour-violating coefficients for $QQ\Xi^\dagger$, $UD\Xi$, this nice property

[9] To avoid any confusion, by $Sp(N-3)$ we denote the group of symplectic unitary $N-3$ matrices. Consistently, non-triviality of the theory implies $N-3$ is even.

would not have held anymore[10].

The bottom line is that in this model eq. (3.3.2) remains satisfied up to the desired accuracy at the renormalizable level. The most dangerous non-renormalizable interactions are dimension-5 and dimension-6 operators that contribute differently to the topological angles once the scalars Φ, Ξ acquire a vev:

$$\frac{\bar{c}_5}{f_{\rm UV}}\frac{g_{\rm GC}^2}{32\pi^2}\Phi\, G_{\rm GC}\widetilde{G}_{\rm GC},\ \frac{\bar{c}_6}{f_{\rm UV}^2}\frac{g_{\rm GC}^2}{32\pi^2}\Phi^\dagger\Phi\, G_{\rm GC}\widetilde{G}_{\rm GC},\ \frac{\bar{c}_6'}{f_{\rm UV}^2}\frac{g_{\rm GC}^2}{32\pi^2}\Xi^\dagger\Xi\, G_{\rm GC}\widetilde{G}_{\rm GC}. \tag{3.3.7}$$

The dominant effect comes from the first interaction, but this can be avoided by charging Φ under an additional gauge symmetry, or postulating that Φ be the scalar responsible for breaking $U(1)_{\rm PQ}$, in which case $f_a \sim f_{\rm GC}$. The last two operators are more model-independent and imply $|\bar{\theta}-\bar{\theta}'| \sim f_{\rm GC}^2/f_{\rm UV}^2$. Taking the Planck scale as the UV cutoff, satisfying the relation eq. (3.3.2) up to corrections of order 10^{-10} imposes the constraint $f_{\rm GC} \lesssim 10^{13}$ GeV. This bound can be further relaxed if Φ, Ξ are composite operators.

Overall, the picture that emerges is qualitatively similar to the Z_2-symmetric scenarios: color and the exotic confining dynamics have basically the same topological angle if no Yukawa couplings are introduced beyond Y_u, Y_d and the breaking of Grand Color is sufficiently soft. Under these conditions a unique axion $\bar{a}/f_a = \bar{\theta} + a/f_a$ from the breaking of a $U(1)_{\rm PQ}$ with a Grand Color anomaly would automatically relax to zero the topological angles of both color and $\mathrm{C}' = Sp(N-3)$. By making the latter confine at a scale $f \gg f_\pi$ much larger than QCD we will see the axion mass can be enhanced and the axion quality improved while still robustly solving the Strong CP problem. Actually, we will have to require f larger than the weak scale because the exotic fermions carry electroweak charges.

We stress that the precise origin of the axion is not relevant. What matters is that its couplings to the $SU(3)_{\rm C} \times Sp(N-3)$ topological terms be the same. In addition, we will work under the hypothesis that

$$f_a > f \tag{3.3.8}$$

so that the tools of effective field theory can be employed in the following to study the axion potential[11]. Scenarios in which $f_a > f_{\rm GC}$ automatically lead to equal couplings to the $SU(3)_{\rm C} \times Sp(N-3)$ topological terms. It is perhaps worth showing that the same may also be true for $f_a < f_{\rm GC}$. To see this let us for example UV complete the axion sector via an interaction $\mathcal{L}_{\rm PQ} \supset yFF^c\Theta$, with F (F^c) fermions in the fundamental (anti-fundamental) of $SU(N)_{\rm GC}$ carrying $U(1)_{\rm PQ}$ charge $+1$ and Θ a scalar of charge -2 responsible for breaking

[10] In this respect our approach differs qualitatively from [143, 144], where the beyond the SM fermions filling the Grand Color multiplet are decoupled by giving them large masses. Such a decoupling may also be achieved in our model, where $QQ\Xi^\dagger$, $UD\Xi$ would generate a vector-like mass matrix M for the $Sp(N-3)$ fermions below $f_{\rm GC}$. Unfortunately, decoupling would typically violate (3.3.2) because M introduces a new physical CP-odd phase that contributes to $\bar{\theta}$ at tree-level. In order to preserve $|\bar{\theta}' - \bar{\theta}| < 10^{-10}$ one would therefore be forced to demand that $|\arg\det M| < 10^{-10}$. Here we avoid this fine-tuning by forbidding the couplings $QQ\Xi^\dagger$, $UD\Xi$. This way the extra fermions remain chiral, like the SM fermions, and get trapped into the heavy $Sp(N-3)$ hadrons.

[11] The opposite regime, with $f_a < f$, may nevertheless provide a solution to the Strong CP problem but requires a completely different study.

$U(1)_{\rm PQ}$ spontaneously at a scale $\sim f_a$. In such a model the axion acquires the very same couplings to the $SU(3)_{\rm C} \times Sp(N-3)$ topological terms even with $f_a < f_{\rm GC}$ because below the Grand Color breaking scale F, F^c split into the direct sum of fermions that are both in the fundamental representation of $SU(3)_{\rm C}$ and $Sp(N-3)$ and so have the same Dynkin index. The phase in y does not affect this conclusion. Finally, we note that for definiteness we decided to work within a KSVZ axion model, but it should be clear that a DFSZ model would equally do.

3.3.2 The Axion Potential at Leading Order

At scales below $f_{\rm GC}$ our model reduces to the SM plus an $Sp(N-3)$ gauge theory with three families of fermions $\psi_q = (\psi_{q_u}, \psi_{q_d})$, $\psi_{u,d}$ charged as shown in table 3.2, with Yukawa couplings (3.3.4), and an axion $\bar{a}$ equally coupled to color $SU(3)_{\rm C}$ and $Sp(N-3)$. All the scalars contained in Φ, Ξ acquire masses proportional to $f_{\rm GC}$ and decouple.

The fate of the exotic fermions and the axion potential is completely determined assuming that $Sp(N-3)$ confines at a scale $f < f_{\rm GC}$ larger than $v \approx 246$ GeV. This hypothesis is certainly realized provided $N \geq 9$. In order to get rid of an otherwise large mixing between the axion and the η' singlet of the $Sp(N-3)$ dynamics we remove the axion from the topological term via a rotation of the $\psi_{q,u,d}$. This can for example be achieved via a phase re-definition $\psi_u \to e^{i\bar{a}/3f_a}\psi_u$, where the factor of 3 denotes the number of generations, which puts the axion in front of the up-type Yukawas, i.e. $Y_u \to e^{i\bar{a}/3f_a} Y_u$.

The physics at confinement is better described in terms of a strong $Sp(N-3)$ dynamics with an approximate $SU(12)$ global symmetry under which the column vector $\Psi = (\psi_{q_u}, \psi_{q_d}, \psi_u, \psi_d)$ transforms as the fundamental representation. At confinement the chiral condensates $\langle \psi_{q_u}\psi_{q_d}\rangle = -\langle \psi_{q_d}\psi_{q_u}\rangle = \langle \psi_u \psi_d\rangle \sim 4\pi f^3/\sqrt{N}$ break $SU(12) \to Sp(12)$ [146, 147]. To demonstrate this we first observe that, because all bound states of $Sp(N-3)$ are bosonic, 't Hooft anomaly matching implies that $SU(12)$ must be broken. Finally, by the Vafa-Witten theorem we know that the vectorial subgroup, namely $Sp(12)$, should remain unbroken [148]. Crucially, the electroweak symmetry is part of the unbroken group. The choice $\mathrm{C}' = Sp(N-3)$ is essential to achieve this key property.

The pattern $SU(12) \to Sp(12)$ delivers 65 would-be Nambu-Goldstone bosons Π. These are not exact because the weak gauging of $SU(2)_{\rm L} \times U(1)_{\rm Y}$ and the Yukawa couplings constitute a small explicit breaking of $SU(12)$. In particular 51 of the would-be NGBs acquire positive mass squared of order $g^2 f^2, g'^2 f^2$ from loops of the $SU(2)_{\rm L} \times U(1)_{\rm Y}$ vectors [146]. The other 14, denoted by Π_0, are gauge-neutral and can in principle mix with $\bar{a}$, similarly to the π_0 in the standard QCD axion. It is the dynamics of these Π_0 that controls vacuum alignment and in particular the vacuum expectation value of the axion. The vev of the charged NGBs, instead, vanish and can be ignored in our discussion. Incidentally, some of the charged NGBs are electroweak doublets and mix with the fundamental H. The heavy linear combinations are integrated out, whereas we assume that the mass parameter of H is such that there exists a unique light eigenstate with a small and negative mass squared. This will play the role of the Higgs doublet of the SM, $H_{\rm SM}$. The fine-tuning we just invoked is nothing but the usual hierarchy problem[12].

[12]Note that from this observation follows that the true SM Yukawa couplings in low-energy observables differ

The dynamics of the neutral NGBs can be effectively described observing that the electroweak symmetry leaves intact a smaller $SU(3)_q \times SU(3)_u \times SU(3)_d \times U(1)_{\rm B}$ global subgroup of $SU(12)$, associated to the independent flavour rotations of ψ_q, ψ_u, ψ_d as well as the $Sp(N-3)$ baryon number under which ψ_u, ψ_d have charge opposite to ψ_q. The vacuum condensates break this symmetry down to $SO(3)_q \times SU(3)_{u-d}$. As a result the 14 neutral NGBs can be effectively parametrized in terms of three matrices: a special, unitary and symmetric matrix $\Sigma_L \in SU(3)_q/SO(3)_q$, a special, unitary matrix $\Sigma_R \in SU(3)_u \times SU(3)_d/SU(3)_{u-d}$, and finally $\eta_{\rm B}$, the NGB of $U(1)_{\rm B}$. The boson $\eta_{\rm B}$ remains an exactly massless state because the baryon number is not explicitly broken. We will discuss its phenomenology in subsection 3.3.4. The remaining 13 neutral scalars, along with the axion, acquire a potential from the Yukawa interactions of $\psi_{q,u,d}$. We stress that these couplings are the same as those of the SM quarks at the threshold $f_{\rm GC}$, though below that scale they renormalize differently. At the scale f relevant for the present discussion the $\psi_{q,u,d}$ couplings, which to avoid over-complicating our notation will still be denoted by $Y_{u,d}$, are expected to be somewhat larger than the SM Yukawas by a flavour-universal factor due to loops of the $Sp(N-3)$ dynamics. Expanding in powers of $Y_{u,d}$ the most general potential reads [7]:

$$V_{\rm neutral} = \frac{c_{ud}}{N} f^4 \,{\rm tr}\, Y_u^* \Sigma_R Y_d^\dagger \Sigma_L e^{i\frac{\bar{a}}{N_g f_a}} + {\rm h.c.} + \mathcal{O}(Y^4, v^2/f^2), \tag{3.3.9}$$

with $N_g = 3$ the number of generations. The dominant contribution arises from a loop of H and the $Sp(N-3)$ dynamics. The factor of N has been identified using a large N scaling and recalling that $f^2 \propto N$. The parameter c_{ud} is a real incalculable quantity. Subleading corrections contain $|H_{\rm SM}|^2$ and/or higher order insertions of the Yukawa couplings. The former cannot affect qualitatively the potential; such corrections are necessarily small because we are interested in the chiral regime $v \lesssim f$ (see also 3.3.4). The latter will be argued to be negligible in subsection 3.3.3.

Contrary to the standard QCD axion, the theory under consideration contains a light fundamental scalar with Yukawa couplings to the fermions Ψ and it is not possible to directly apply the results of [21] in order to argue that $\langle \bar{a} \rangle = 0$. The minimization problem is therefore conceptually different from QCD. In particular, in QCD [21] imply that the vev of the pions must vanish and the low energy dynamics contains no CP violation other than the one encoded in $\bar{\theta}$. In our scenario, on the other hand, the axion effective potential can depend non-trivially on the vacuum configuration of the Π_0's. We will have to prove $\langle \bar{a} \rangle = 0$ by brute force. This is what we will do in the next section.

Before turning to the minimization of the potential, though, we stress that (3.3.9) possesses a $Z_{N_g} \subset SU(4N_g)$ symmetry under which $\Sigma_{R,L} \to e^{\pm i 2\pi n/N_g} \Sigma_{R,L}$. This discrete symmetry signals the presence of a set of inequivalent vacua sharing the same perturbative mass spectrum and axion vev, which may indicate a cosmological domain-wall problem if the temperature of the Universe ever exceeded f[13]. This issue adds to the more familiar domain-wall problem of axion models, which takes place at the scale $f_a > f$.

compared to $Y_{u,d}$, not only due to different RG effects, but also because of some mixing angle. We will neglect these corrections since our results are anyway affected by uncertainties of $\mathcal{O}(1)$ from incalculable coefficients.

[13] These domain-walls are stable because $Z_3 \subset U(1)_{\rm B}$, and explicit breaking of $U(1)_{\rm B}$ occurs via effective operators of an extremely high dimensionality, since the baryon number is very well protected by our gauge symmetries.

Minimisation

The potential (3.3.9) involves 14 fields (13 neutral NGBs Π_0 and the axion $\bar{a}$), and its minimization is highly non-trivial. To perform this task we find it convenient to first discuss the properties of the more general structure

$$V_{\text{neutral}}^{\text{LO}} = V_0\left(\Pi_0/f\right) e^{i\bar{a}/f_a N_g} + \text{h.c.}, \tag{3.3.10}$$

where the number of light fermion generations N_g as well as the explicit expression of V_0 are left arbitrary. Interestingly, both the potentials of our model and that of the standard QCD axion have precisely this form. Therefore some of the results discussed here have a rather general validity. In particular, in appendix 3.A we demonstrate that the absolute minimum of (3.3.10) is found by maximizing $|V_0|$, whereas the axion vacuum is determined by $\langle\bar{a}\rangle/f_a = N_g(\pi - \phi) \bmod 2\pi$, where $\phi = \arg V_0$ at the extremum.

In the case at hand V_0 is given in equation (3.3.9), and in the basis in which $Y_u = \widehat{Y}_u$ is diagonal can be written as

$$\begin{aligned} V_0 &= \frac{c_{ud}}{N} f^4 \operatorname{tr} \widehat{Y}_u \Sigma_R \widehat{Y}_d V_{\text{CKM}}^\dagger \Sigma_L \\ &= \frac{c_{ud}}{N} f^4\, [\widehat{Y}_u]_i [A]_{ii}, \qquad\qquad A = \Sigma_R \widehat{Y}_d V_{\text{CKM}}^\dagger \Sigma_L. \end{aligned} \tag{3.3.11}$$

$|V_0|$ is maximized when A is aligned as much as possible along $\widehat{Y}_u$, with the corresponding entries satisfying $|[A]_{33}| > |[A]_{22}| > |[A]_{11}|$. Suppose for the time being that it is possible to find a configuration $\Sigma_{L,R}$ that fully diagonalizes A, so that the diagonal entries read $[A]_{ii} = |[A]_{ii}|e^{i\phi_i}$, where ϕ_i are phases subject to $\phi_1 + \phi_2 + \phi_3 = 2\pi n$ because of the constraint $\det A = \det \widehat{Y}_d \in \mathbb{R}$. Under this hypothesis $|V_0|$ would be maximized when $\phi_i = \phi_j$ is common to all entries, such that the trace becomes a coherent sum of terms, and the minimum configuration would read $\phi_i = 2\pi n/N_g$. The phase of V_0 at the minimum would finally be $\phi = \arg\, c_{ud} + 2\pi n/N_g$, and from eq. (3.A.7) we would infer that $\langle\bar{a}\rangle/f_a = N_g(\pi - (\arg\, c_{ud} + 2\pi n/N_g)) = N_g(\pi - \arg\, c_{ud})$, or [7]

$$\frac{\langle\bar{a}\rangle}{f_a} = \begin{cases} 0 \bmod 2\pi & \text{if } N_g = \text{even} \\ 0 \bmod 2\pi & \text{if } N_g = \text{odd and } c_{ud} < 0\,. \\ \pi \bmod 2\pi & \text{if } N_g = \text{odd and } c_{ud} > 0 \end{cases} \tag{3.3.12}$$

This shows that, as long as A can be diagonalized, the system has a natural tendency to relax the axion to a CP-conserving vev. Thus the axion vev is usually vanishing, though for odd N_g and positive c_{ud} we get $\langle\bar{a}\rangle/f_a = \pi$. Despite being CP-conserving, the latter option is not phenomenologically acceptable because incompatible with the Gell-Mann-Okubo relations [38]. In the standard QCD axion the result of [21] ensures that $\langle\bar{a}\rangle/f_a = 0$, which implies that c_{ud} must be negative. In our model later on we will offer some argument indicating that c_{ud} should be negative.

Unfortunately, it is possible to prove that as soon as $N_g \geq 3$ the matrix A cannot be exactly diagonalized because Σ_L, being unitary-symmetric, does not contain enough degrees of freedom to diagonalize $A^\dagger A$. In scenarios with $N_g \geq 3$ the logic leading to (3.3.12) can thus at most be approximate. And yet, we find (at least for the physically relevant case $N_g = 3$) that (3.3.12) remains valid. Despite the impossibility of diagonalizing A, in fact, the relation $\det A = A_{11}A_{22}A_{33} + \Delta$ holds up to a very small perturbation Δ. Equation (3.3.12) then applies because in a perturbative expansion for small off-diagonal elements the dynamical phases of the three diagonal elements of A are determined at leading order to be $2\pi n/N_g$. That is, the corresponding fluctuations fall into a deep potential well, which cannot be destabilized by the next to leading corrections due to Δ. As a result the overall phase of V_0 is still determined by $\phi = \arg c_{ud} + 2\pi n/N_g$ and the axion vev by (3.3.12), as if A could be exactly diagonalized.

Even though the above arguments seem rather convincing, an explicit calculation would help lifting any doubt on (3.3.12). Furthermore, an explicit analysis is necessary to compute the masses of the NGBs and the axion. In the following we will thus verify (3.3.12) and calculate the axion mass for $N_g = 1$, where in fact A is trivially diagonalized, as well as for $N_g = 2$, where it can be fully diagonalized by the NGB matrices. Subsequently we will consider the phenomenologically relevant case $N_g = 3$. Along the way we will argue in favor of $c_{ud} < 0$.

Warming Up With $N_g = 1$ and $N_g = 2$

The $N_g = 1$ case is almost trivial, since the spectrum of NGBs is composed of a charged composite Higgs, that is not relevant to vacuum alignment, and the exact flat direction $\eta_{\rm B}$. The potential simply reduces to a potential for the axion:

$$V^{\rm LO}_{\rm neutral} = 2\,\frac{c_{ud}}{N} f^4\, y_u y_d \cos\left(\frac{\bar{a}}{f_a}\right) \qquad (N_g = 1) \tag{3.3.13}$$

where y_u, y_d are the up and down quark Yukawas. This potential is minimised at $\langle \bar{a} \rangle / f_a = 0 \bmod 2\pi$ if $c_{ud} < 0$ or $\langle \bar{a} \rangle / f_a = \pi \bmod 2\pi$ if $c_{ud} > 0$, as expected from (3.3.12). The axion mass is given by

$$m_a^2 = 2\,\frac{|c_{ud}|}{N}\, y_u y_d \frac{f^4}{f_a^2} \qquad (N_g = 1). \tag{3.3.14}$$

It is possible to show that for $N_g = 1$ the parameter c_{ud} must be negative. The argument is a bit involved and will only be sketched here.

Our argument starts by considering a modified $N_g = 1$ scenario in which $Y_u = Y_d$ and only the neutral component of the fundamental Higgs is dynamical. This is certainly not our model, but its effective potential is just a simple generalization of ours because the UV diagrams that generate it, in terms of fundamental fermions and H, are virtually identical to those in our model modulo corrections of order v^2/f^2. In particular, the sign of the overall coefficient c_{ud} is exactly the same in the two scenarios because determined by equal correlators in the unperturbed $Sp(N-3)$ theory. The conclusion that the axion vev vanishes only for $c_{ud} < 0$ remains valid. But crucially, in the modified model the fermionic determinant arising from the integration of Ψ is real and positive definite because the fermionic spectrum is effectively doubled [22]. Therefore the result of [21] can be generalized [149] to argue that the axion must be minimized at zero, and hence indirectly that $c_{ud} < 0$. This for us is proof that the coefficient c_{ud} in (3.3.13) is negative.

The minimization of the $N_g = 2$ case is more interesting. In this case the potential (3.3.9) depends on a CKM matrix that can be written in terms of the Cabibbo angle:

$$V_{\text{CKM}} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \qquad (N_g = 2). \tag{3.3.15}$$

As anticipated earlier, we find two inequivalent vacua related by a Z_2 symmetry. These are given by

$$\begin{aligned} c_{ud} > 0 : &\quad \begin{cases} \langle \Sigma_L \rangle &= (\pm) \begin{pmatrix} i\cos\theta_c & i\sin\theta_c \\ i\sin\theta_c & -i\cos\theta_c \end{pmatrix}, \\ \langle \Sigma_R \rangle &= (\pm) \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \end{cases} \qquad \langle \bar{a} \rangle = 0, \\ c_{ud} < 0 : &\quad \begin{cases} \langle \Sigma_L \rangle &= (\pm) \begin{pmatrix} i\cos\theta_c & i\sin\theta_c \\ i\sin\theta_c & -i\cos\theta_c \end{pmatrix} \\ \langle \Sigma_R \rangle &= (\mp) \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \end{cases}, \qquad \langle \bar{a} \rangle = 0. \end{aligned} \tag{3.3.16}$$

With two generations the vacuum configurations precisely diagonalise $A = \Sigma_R \widehat{Y}_d V^\dagger_{\text{CKM}} \Sigma_L$, and in both cases $\langle \bar{a} \rangle / f_a = 0$ consistently with (3.3.12). In this vacuum all scalar excitations (except for the exact flat direction η_{B}) are massive.

The $N_g = 2$ case is so simple to handle analytically that we were able to find an explicit expression for the effective axion potential. This is obtained by solving the equation of motion for the neutral NGBs Π_0 and plugging it back into V_{neutral}. It is a reliable approximation of the axion self-couplings in the limit $f \ll f_a$ in which the neutral NGBs are much heavier than the axion. We find

$$V_{\text{eff}}\left(\frac{\bar{a}}{f_a}\right) = -2\,\frac{|c_{ud}|}{N} f^4 \,\operatorname{tr} \widehat{Y}_u \widehat{Y}_d \sqrt{1 - 4\frac{\det \widehat{Y}_u \widehat{Y}_d}{\left(\operatorname{tr} \widehat{Y}_u \widehat{Y}_d\right)^2} \sin^2\left(\frac{\bar{a}}{2f_a}\right)} \qquad (N_g = 2), \tag{3.3.17}$$

which is consistently minimized at $\langle \bar{a} \rangle / f_a = 0 \bmod 2\pi$. This result is reminiscent of the potential of the QCD axion in 2-flavour QCD. The axion mass immediately follows:

$$m_a^2 = 2\frac{|c_{ud}|}{N}\,\frac{\det \widehat{Y}_u \widehat{Y}_d}{\operatorname{tr} \widehat{Y}_u \widehat{Y}_d}\,\frac{f^4}{f_a^2} \qquad (N_g = 2). \tag{3.3.18}$$

Equation (3.3.17) is very valuable because we will not be able to obtain an explicit expression for $N_g = 3$. It is therefore useful to extract as much information as possible from it. First, we observe that in the limit of a heavy second generation equations (3.3.17), (3.3.18) reduce to (3.3.13), (3.3.14). This is a highly non-trivial check of the consistency of our results. It is a consequence of the fact that a large Yukawa coupling for the second generation implies that a number of NGBs becomes much heavier than those associated to the light first generation. Up to corrections suppressed by the heavy NGB mass, therefore, the potential should reduce to the $N_g = 1$ case, which is what we see here explicitly. The very same logic constrains the structure of the $N_g = 3$ potential, as we will verify numerically. A second important lesson we can draw from (3.3.17), and more readily (3.3.14), is that the non-trivial dependence of the axion potential should be controlled by $\det \widehat{Y}_u \widehat{Y}_d$. This is also a very general result, independent of N_g. Indeed, if any of the eigenvalues of Y_u or Y_d were to vanish the UV Lagrangian would be invariant under an additional anomalous axial symmetry which could be combined with $U(1)_{\rm PQ}$ to obtain an exact unbroken one. In that situation the axion would become an exact flat direction. Hence the axion potential must be proportional to at least a power of all the eigenvalues. The quantity $\det \widehat{Y}_u \widehat{Y}_d$ is the simplest object with this property. The axion mass squared cannot be simply proportional to the determinant unless $N_g = 1$, however. By counting the units of $\hbar$ we need at least $2N_g - 2$ additional coupling constants in the denominator, so dimensional analysis forces m_a^2 to be inversely proportional to an appropriate combination of Y_u, Y_d as found in (3.3.18). This holds true also in the SM, as noted in section (3.1.1). The significant hierarchy in the SM fermion masses and the decoupling properties mentioned in the previous paragraph, together indicate that the axion mass in our model is always numerically close to (3.3.14).

The Real World: $N_g = 3$

Having checked the simplified scenarios $N_g = 1, 2$, we can now turn to the realistic case $N_g = 3$. The previous calculations support the correctness of (3.3.12), and as such we expect the leading order potential (3.3.9) to be minimised at $\langle \bar{a} \rangle / f_a = 0\,(\pi)$ for $c_{ud} < 0\,(> 0)$. Unfortunately, in the case $N_g = 3$ the potential involves 14 fields and it is not possible to approach the problem analytically. For this reason we employ a customised MATHEMATICA algorithm which enables us to numerically find the minimum of the potential up to a very high accuracy. The minimisation procedure is repeated many times in order to statistically validate the result. The Yukawa couplings that appear in the potential are renormalized at the scale $\sim 4\pi f/\sqrt{N}$ by the $Sp(N-3)$ dynamics. As a benchmark we employ the PDG data for $V_{\rm CKM}$ [1] and the numerical values of $\widehat{Y}_u, \widehat{Y}_d$ that correspond to the SM quark Yukawas evaluated at the TeV scale [150]. Changing the numerical value of these couplings does not affect our results qualitatively. What we find is exactly (3.3.12): the distinct vacua configurations are related by a Z_3 symmetry $\Sigma_{R,L} \to e^{\pm i 2\pi n/3} \Sigma_{R,L}$; all vacua give rise to the same A, which is diagonal up to small off-diagonal elements; the axion is minimised at $\langle \bar{a} \rangle / f_a = 0$ or π depending on the sign of c_{ud}.

In the study of the $N_g = 1$ toy model we gave an argument supporting the claim that c_{ud} is negative. At sufficiently large N there is no distinction between the coefficients c_{ud} for the $N_g = 1$ and $N_g = 3$ scenarios. We are therefore motivated to conjecture that $c_{ud} < 0$ also for $N_g = 3$, and from now on work under this hypothesis. Given the central role played by this hypothesis, it would be interesting to find an independent proof, for instance using lattice QCD techniques.

We are now interested in studying the spectrum of the NGBs and the axion. To read off the masses we first canonically normalize the kinetic term $f^2 \,\mathrm{tr}\, \partial_\mu \Sigma_{L,R} \partial^\mu \Sigma_{L,R}$, altered by the vev of the NGBs, and then perform an $SO(14)$ rotation to diagonalise the Hessian of the potential. As a result of this operation, the masses of the 13 neutral NGBs are found to be [7]

$$m^2_{\Pi_0} \simeq \begin{pmatrix} 6.4 \times 10^{-2} \\ 2.5 \times 10^{-2} \\ 2.4 \times 10^{-2} \\ 2.4 \times 10^{-2} \\ 2.4 \times 10^{-2} \\ 2.4 \times 10^{-2} \\ 2.3 \times 10^{-2} \\ 2.7 \times 10^{-5} \\ 1.2 \times 10^{-5} \\ 2.5 \times 10^{-6} \\ 1.6 \times 10^{-6} \\ 1.6 \times 10^{-6} \\ 4.0 \times 10^{-8} \end{pmatrix} \times \frac{|c_{ud}|}{N} f^2 \tag{3.3.19}$$

for all the three Z_3-symmetric vacua configurations. The eigenstate corresponding to the axion is the lightest one for any $f < f_a$. The mass that we extract numerically respects the scaling suggested in the previous section, namely [7]

$$m_a^2 \simeq 2\frac{|c_{ud}|}{N} y_u y_d \frac{f^4}{f_a^2} \simeq 1.8 \times 10^{-10} \times \frac{|c_{ud}|}{N}\frac{f^4}{f_a^2} \qquad (N_g = 3). \tag{3.3.20}$$

The NGBs' masses are substantially unaffected by the mixing with the axion as long as $f/f_a \ll 1$ (even though the mixing will turn out to be important for phenomenology, see subsection 3.3.4). In the extreme limit $f/f_a \to 1$ the mixing impacts the NGBs masses by a few 10%, with the lightest being affected the most.

We conclude this section emphasizing that the axion mass (3.3.20) is parametrically enhanced with respect to the standard QCD axion as long as $f \gtrsim 10^2 f_\pi$. Therefore, the goal outlined at the beginning is fulfilled.

3.3.3 Subleading Corrections and Heavy Axion Quality

The analysis carried out so far demonstrates that the Grand Color axion has a large mass and a leading order potential minimized at $\langle \bar{a} \rangle = 0$. Higher order corrections cannot affect this result unless they introduce new sizable sources of CP violation or flavour-violation. We will argue next that subleading effects due to renormalizable interactions do not spoil our solution of the Strong CP problem and that the effect of non-renormalizable operators can be taken under control.

Renormalizable Interactions

In the renormalizable version of our scenario the effective axion potential $V_{\text{eff}}(\bar{a}/f_a)$ depends on flavour-invariant combinations of the parameters $Y_{u,d}$, $\langle \Sigma_{L,R} \rangle$. The Yukawas parametrize *explicit* CP violation, whereas the NGB vacuum potentially represents an independent source of *spontaneous* CP violation. A non-vanishing vev for the axion is induced by CP-odd, flavour-conserving combinations of these parameters.

Crucially, in our model all CP-odd invariants must necessarily be proportional to explicit CP violation. This follows from the fact that spontaneous CP violation does not take place. As a first simple check of this statement, let us inspect the $N_g = 2$ toy model, where we have an explicit analytic solution. Here the CKM matrix is real, i.e. there is no explicit CP violation, and it is readily seen that equations (3.3.16) preserve the generalised CP transformation $\Sigma_{R,L} \to -\Sigma^*_{R,L}$. A far less trivial check is obtained for $N_g = 3$. In that case we verified that, when we switch off the CKM phase, the explicit numerical solution of the leading order potential in (3.3.9) also satisfies the relation $\langle \Sigma_{R,L} \rangle = e^{\pm i2\pi n/3} \langle \Sigma_{R,L} \rangle^*$. In other words, in the absence of explicit violation, CP is not spontaneously broken.

An important consequence of what we just demonstrated is that any CP-odd flavour invariant in V_{eff} must be proportional to the explicit CP violation in $Y_{u,d}$. A key property of our theory, inherited from the SM, is that explicit CP violation should disappear whenever two of the eigenvalues of the SM quark mass matrix squared are degenerate, or any mixing angle goes to zero, or when the CKM phase vanishes. This is very important. We have already seen that the non-trivial part of the effective axion potential must be proportional to $\det Y_u Y_d$. Here we find that any explicit CP-violating interaction of the axion must contain a further suppression that disappears in the above limits. Such a suppression is so significant that explicit CP violation in V_{eff} becomes effectively innocuous: CP-odd flavour-invariant combinations of the Yukawas and $\langle \Sigma_{L,R} \rangle$ may arise in V_{eff} only at very high order in an expansion in $Y_{u,d}$ and are numerically extremely small. It is therefore not surprising that, even including the CKM phase, our $\mathcal{O}(Y^2)$ potential does not induce an axion vev. In fact, at $\mathcal{O}(Y^2)$ the effect of explicit CP violation cannot be visible, the only flavour-invariant candidate $\operatorname{tr} Y_u^* \langle \Sigma_R \rangle Y_d^\dagger \langle \Sigma_L \rangle$ is real and hence there is nothing that can be on the right-hand side of $\langle \bar{a} \rangle = 0$.

In summary, subleading corrections to the effective axion potential are either even in $\bar{a}$ or odd, the latter being proportional to the explicit CP violation. Terms even in $\bar{a}$ cannot destabilize our solution because the leading order theory has no flat directions. The terms odd in $\bar{a}$ are however dangerous if they include a tadpole. In that case the axion vev is shifted from the origin. Still, the shift must be proportional to the tiny explicit CP-violating phase that controls the tadpole and the vacuum expectation value would thus be safely below $|\langle \bar{a} \rangle|/f_a \lesssim 10^{-10}$. Our axion dynamically solves the Strong CP problem like in the standard QCD scenario. Higher dimensional operators with new CP-violating or flavour-violating couplings can however introduce new CP-odd flavour invariants which can be numerically more relevant that those of the renormalizable theory. These effects are discussed next.

Higher-Dimensional Operators

At the root of the axion quality problem is the fact that a huge f_a makes the axion potential extremely sensitive to cutoff-suppressed $U(1)_{\text{PQ}}$-breaking interactions. As stated at the beginning of this section, this sensitivity may be alleviated by increasing the axion mass. However, there is no free lunch. To enhance the axion mass, the confinement scale f has to be rather large as well. As a result, cutoff-suppressed $U(1)_{\rm PQ}$-*conserving* contributions to the axion potential, as long as they are CP- or flavour-violating, may become important and in principle spoil the solution of the Strong CP problem.

More precisely, consider the flavour-violating but $U(1)_{\rm PQ}$-preserving operator

$$\frac{\bar{c}_{ijkl}}{f_{\rm UV}^2} Q_i Q_j U_k D_l, \tag{3.3.21}$$

where i, j, k, l are flavour indices. This has precisely the same axial $U(1)_{\rm A}$ charges as the leading order potential in (3.3.9) and thus represents a modification $\delta V_0 \sim \bar{c}\, 16\pi^2 f^6/(N^2 f_{\rm UV}^2)$ of the quantity $V_0 \sim \operatorname{tr} Y_u Y_d\, f^4/N$ defined in (3.3.11). This is not aligned with (3.3.9) for generic $\bar{c}_{ijkl}$, and must therefore be small. According to (3.A.2) the axion vev is of order $\langle \bar{a} \rangle / f_a \sim \operatorname{Im} \delta V_0/|V_0|$, where the imaginary part of the flavour invariant δV_0 may come either from $\bar{c}_{ijkl}$ directly or from phases of the leading NGB vev, which become physical when contracted with a flavour-violating $\bar{c}_{ijkl}$. The requirement that the effective topological angle be less than 10^{-10} becomes

$$f \lesssim 10^{-7} f_{\rm UV} \tag{3.3.22}$$

which for a maximal UV cutoff of order $f_{\rm UV} = M_{\rm P}$ reads $f \lesssim 10^{11}$ GeV. Somewhat similar considerations apply to operators that do not violate the axial $Sp(N-3)$ symmetry $U(1)_{\rm A}$ but still violate flavour, like

$$\frac{\bar{c}_{ijkl}}{f_{\rm UV}^2} (\Psi_i \Psi_j)(\Psi_k \Psi_l)^\dagger. \tag{3.3.23}$$

This operator modifies the NGB vev and in turn shifts the axion minimum. As a conservative bound we impose (3.3.22). Moreover, we could have operators that do not violate flavour, but contribute new CP-odd flavour-conserving phases to the axion effective potential. A typical example is the Weinberg operator

$$\frac{\bar{c}_W}{M_{\rm UV}^2} \frac{g_{\rm GC}^3}{16\pi^2} G_{\rm GC} G_{\rm GC} \widetilde{G}_{\rm GC}. \tag{3.3.24}$$

The new CP-odd phase should be smaller than 10^{-10} to guarantee a solution of the Strong CP problem. From this requirement the weaker upper bound $\bar{c}_W f^2/f_{\rm UV}^2 \lesssim 10^{-10}$ follows.

Finally, let us come back to the original motivation: the axion quality problem. As in section 3.2, imposing that a $U(1)_{\rm PQ}$-violating operator of dimension n does not significantly alter the axion vev implies

$$\lambda_* f_a^4 \left(\frac{f_a}{f_{\rm UV}} \right)^{n-4} \lesssim 10^{-10} m_a^2 f_a^2, \tag{3.3.25}$$

and by comparing this to the lower bound of the standard QCD axion, i.e. $(n-4)_{\rm std}$, we get

$$(n-4) = (n-4)_{\rm std} - \frac{\log m_a^2/m_{a,\rm std}^2}{\log f_{\rm UV}/f_a}. \tag{3.3.26}$$

The quality problem is logarithmically sensitive to the mass ratio $m_a^2/m_{a,\text{std}}^2$ and becomes more and more sensitive to this quantity as f_a gets larger. In our model we find an appreciable improvement as long as $f \gg 10^2 f_\pi$, see (3.3.20). As a numerical example, for $f_a = 10^{10}$ GeV and $f = 10^8$ GeV we get an axion mass of a few GeV. Taking $f_{\text{UV}} = M_{\text{P}}$ and $\lambda_* = 1$ this corresponds to [7]

$$n \gtrsim 7 \qquad \begin{pmatrix} f = 10^8 \text{ GeV} \\ f_a = 10^{10} \text{ GeV} \\ f_{\text{UV}} = M_{\text{P}} \end{pmatrix} \tag{3.3.27}$$

which is a significant improvement compared to $n_{\text{std}} \gtrsim 10$.

In conclusion, the sensitivity to Peccei-Quinn-violating interactions suppressed by the UV cutoff is reduced compared to the standard scenario as a result of the larger axion mass. However, this very enhancement leads to a novel sensitivity to *Peccei-Quinn-preserving* deformations[14]. This "heavy axion quality problem" [7] is generically shared by all models that attempt to increase the axion mass via a new strong coupling at $f \gg f_\pi$. Future measurements of the dipole moments of the neutron, as well as of atoms and molecules will potentially be able to set upper bounds on the axion mass of these scenarios.

3.3.4 Phenomenology

The phenomenology of our model is extremely rich. Here we present just a qualitative assessment.

The $Sp(N-3)$ dynamics generates many massive hadrons, all of which are unstable because there is no unbroken flavour symmetry that protects them. Heavy hadrons of mass $\propto 4\pi f/\sqrt{N}$ as well as baryons quickly decay into NGBs. The electroweak-charged NGBs decay into the SM Higgs boson, $W^\pm, Z$, and neutral Π_0's. The latter are much more long-lived, and decay dominantly into QCD hadrons and/or photons via the mixing with the axion and the η' of the $Sp(N-3)$ dynamics. Less relevant decay channels for Π_0's are into SM fermions via non-renormalizable interactions generated at the scale f_{GC}. Very likely, yet, only the lightest hadrons were significantly produced in the early Universe because in order to robustly avoid a domain-wall problem associated to the Z_3-degeneracy of the NGB potential reheating must probably have occurred after $Sp(N-3)$ confinement (see discussion below eq. (3.3.9)).

The hadrons can be directly produced at the LHC and future colliders. In addition, the $Sp(N-3)$ dynamics can be indirectly probed via precision measurements. As a rough measure of the current impact of these constraints we impose the qualitative bound $f \gtrsim$ TeV. In this regime the low energy signatures are mainly controlled by $\bar{a}$ and Π_0 (and, if present, η_B; see below). The effective field theory is governed by the couplings to the topological terms of the

[14] The same point was discovered independently in [151]. Their numerical estimates are based on 1-instanton calculations and therefore not reliable for our model nor for mirror-symmetric scenarios. Yet their conclusions are qualitatively general and agree with ours.

gluon and the photon

$$\begin{aligned}\mathcal{L}_{\rm EFT} \supset & \frac{1}{2}(\partial\bar{a})^2 - \frac{m_a^2}{2}\bar{a}^2 + \frac{g_{\rm C}^2}{32\pi^2}\frac{\bar{a}}{f_a}G\widetilde{G} + \bar{c}_{a\gamma\gamma}\frac{e^2}{32\pi^2}\frac{\bar{a}}{f_a}F\widetilde{F} \\ & + \frac{1}{2}(\partial\Pi_{0,i})^2 - \frac{m^2_{\Pi_0,i}}{2}\Pi^2_{0,i} + \frac{g_{\rm C}^2}{32\pi^2}\bar{c}^i_{\Pi_0 gg}\frac{\Pi_{0,i}}{f}G\widetilde{G} + \bar{c}^i_{\Pi_0\gamma\gamma}\frac{e^2}{32\pi^2}\frac{\Pi_{0,i}}{f}F\widetilde{F}.\end{aligned} \tag{3.3.28}$$

The effective couplings to the Z and $W^{\pm}$ bosons are phenomenologically less relevant. Assuming that the $U(1)_{\rm PQ}$ has no electroweak anomaly, and momentarily ignoring the mixing with Π_0, the coefficient $\bar{c}_{a\gamma\gamma}$ can be computed by moving the axion from the $SU(N)_{\rm GC}$ topological term to the Yukawas with an anomalous chiral rotation, and re-placing it only in front of the QCD topological term below the Grand Color breaking. Recalling that the SM hypercharge is given by a combination of $U(1)_{\rm Y'}$ and $U(1)_{\rm GC}$, we get

$$\bar{c}_{a\gamma\gamma} = -\frac{1}{2}(N-3)\left(1-\frac{1}{3N}\right). \tag{3.3.29}$$

Consistently with expectations, this expression vanishes for $N=3$, when our model reduces to a standard KSVZ scenario. We will consider $N=13$ for definiteness, noting that a number of colors > 17 would typically induce a Landau pole for $SU(2)_{\rm L}$ below the Planck scale whereas for $N<9$ the condition $f \gtrsim$ TeV would not be attained. We verified that with these parameters the condition $f < f_{\rm GC} \lesssim 10^{13}$ GeV is also satisfied.

The neutral NGBs have no bare coupling to the SM vectors. However, they acquire them from the mixing with $\bar{a}$ and the heavy η' of $Sp(N-3)$. These contributions are parametrically of order $\bar{c}_{\Pi_0 gg} \sim f/f_a$ and $\bar{c}_{\Pi_0\gamma\gamma} \sim \max\left\{m^2_{\Pi_0}/m^2_{\eta'}, f/f_a\right\}$. As a result the decay rate into gluons, controlled by the $\bar{a}-\Pi_0$ mixing, is of order

$$\Gamma_{\Pi_0\to gg} \gtrsim \Gamma_{\bar{a}\to gg}\frac{f_a}{f}, \tag{3.3.30}$$

where we took into account the different scaling of the masses with f, f_a (see (3.3.20) and (3.3.19)). Importantly, the rate is always greater than $\Gamma_{\bar{a}\to gg}$ in the regime (3.3.8). This parametric estimate is confirmed by an accurate numerical analysis, which also reveals that some Π_0 can have rates several orders of magnitude larger than shown in (3.3.30). Note also that the NGB mixing with the axion does not appreciably modify (3.3.29). As long as $f \ll f_a$ the effect is parametrically suppressed, and we numerically verified that as f approaches f_a the change in the axion coupling to photons is still at most $\mathcal{O}(10\%)$.

The axion mass is given in eq. (3.3.20) and in the allowed regime $f \gtrsim$ TeV is always larger than the standard one. Furthermore, as we saw around (3.3.22), a conservative condition for the Strong CP problem to be solved is $f \lesssim 10^{11}$ GeV, where we identified the UV cutoff with the Planck scale. Combining the two bounds we see that our scenario populates the light-pink bend in the $m_a - f_a$ plot of Fig. 3.2 labeled by "Grand Color axion", defined by the implicit relation 10^3 GeV $\leq f \leq 10^{11}$ GeV — where $f = f(m_a, f_a)$ is given by (3.3.20). We included a hard cut at $f < f_a$ to indicate the regime of validity of the effective field theory approach adopted in this paper, see eq. (3.3.8). In the grey region $f > f_a$ our results do not necessarily apply, though without a detailed analysis this region cannot be excluded. The dotted grey

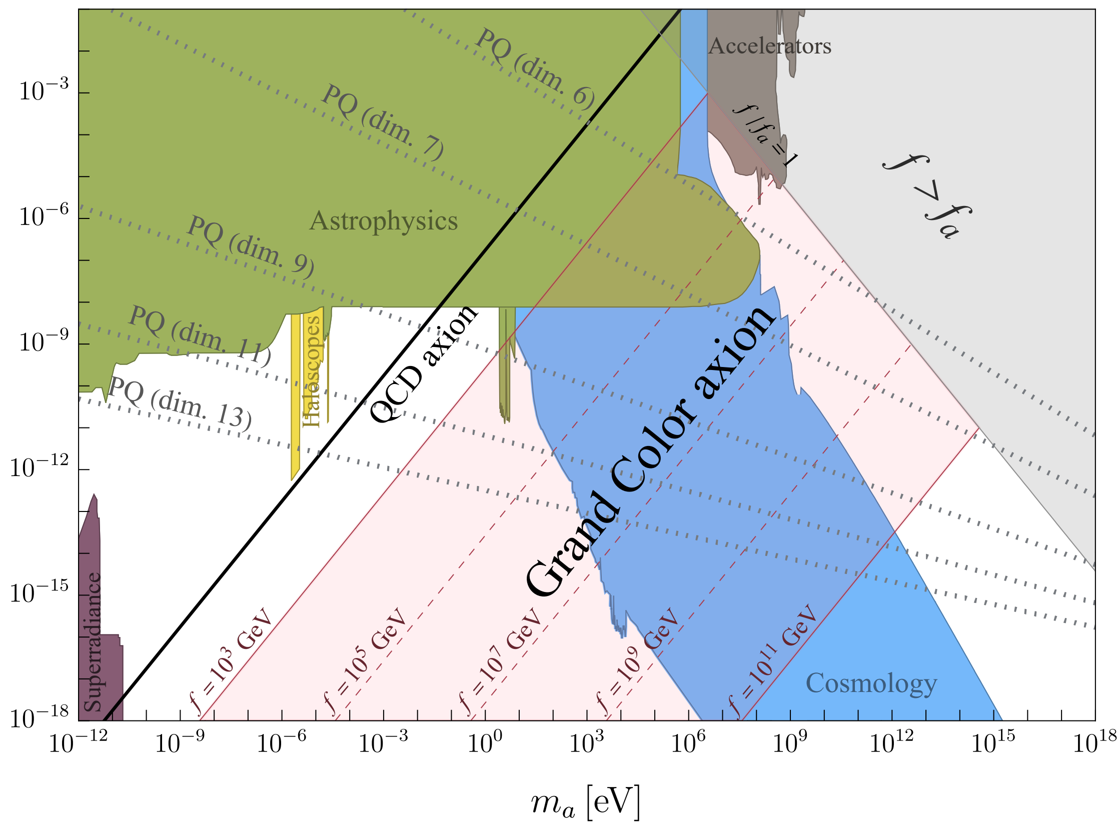

Figure 3.2: Collection of the main bounds on f_a vs m_a, as discussed in subsection 3.3.4. For definiteness we assumed $N = 13$ and $|c_{ud}| = 1$. The light-pink bend denoted by "Grand Color axion" identifies the region populated by our scenario. Plot taken from [7].

lines in Fig. 3.2 show contour regions of eq. (3.3.26) with $f_{\rm UV} = M_{\rm P}$. The axion quality problem is progressively more alleviated as we move towards the upper-right corner. The regime with high quality is the one in which, for a fixed f_a, the axion mass is maximal, i.e. the confinement scale reaches the extreme value $f \to f_a$ compatible with (3.3.8).

The "Grand Color axion" bend is mostly probed by cosmological observations from BBN and CMB physics (blue), astrophysics (green), and collider experiments (brown). In particular, the left boundary of the blue region is taken from the collection of bounds in [94]. The rightmost part of the cosmology bound is due mainly to the $N_{\rm eff}$ bounds from [152] and the requirement that the total axion decay rate satisfies $\Gamma_{\rm tot} \geq 3H(T_{\rm BBN})$, where $H(T_{\rm BBN})$ is the Hubble rate when the Universe reached temperatures of order $T_{\rm BBN} = 4$ MeV [153], in order not to interfere with Big-Bang-Nucleosynthesis. The hadronic decay rate has been calculated adapting the results of [154] to our model. The axion decay rate takes into account (3.3.29), includes also the mixing with the QCD mesons, and is modified with respect to the standard case because of the much larger value of m_a. For the astrophysics bounds we refer to [94]. The collider bounds on top of figure 3.2 are taken from [155] and [156]. The white regions are currently viable.

In the far bottom-left of the "Grand Color axion" bend one should make sure that the heavier NGBs are sufficiently long-lived to avoid disrupting the primordial abundance of light elements. However, because of their large masses, vacuum misalignment typically over-produces them in the form of a Bose condensate of $\bar{a}, \Pi_0$ unless the initial misalignment angles are extremely small. Moreover, in that region the axion quality problem is not ameliorated compared to the standard scenario. For these reasons we believe the upper-right region is more motivated.

The upper-right corner of the "Grand Color axion" bend is also more interesting because in that regime ongoing and future accelerators as well as future CMB observations are able to explore our model. In particular, because f_a can approach the TeV scale this scenario can be probed for example at HL-LHC [141], Kaon and Hyperon factories [155], DUNE [156], NA62 [157], Belle II [158, 159] and MATHUSLA [160]. The phenomenological signatures are much richer than the standard QCD axion because of a variety of axion-like-particles Π_0 with anomalous couplings, see (3.3.28). Future CMB surveys will also significantly improve measurements of $N_{\rm eff}$ and will be able to constrain a wider region of parameter space on the right of the blue area. In contrast to the accelerator searches mentioned above, however, here the neutral NGBs are not expected to play any relevant role because are much heavier and decay faster than the axion, see (3.3.30). Yet, they might still lead to non-standard cosmological signatures, albeit quite indirect. The energy stored in a Bose condensate of neutral NGBs via vacuum misalignment may be estimated as $\rho_{\Pi_0}(T) \sim \theta_r^2 m_{\Pi_0}^2 f^2 (T/T_r)^3$, with θ_r denoting the misalignment angle at the end of inflation and T_r the reheating temperature. For certain values of masses and decay rates the temperature $T_{\rm m}$ at which $\rho_{\Pi_0}(T_{\rm m})$ dominates over radiation is actually larger than the decay temperature $T_{\rm d} \sim \sqrt{\Gamma_{\Pi_0} M_{\rm P}}$. When this happens the Universe undergoes an early period of matter domination which might result in a depletion of the primordial densities of visible and dark matter.

One last comment should be added before closing. In addition to $\Pi_0, \bar{a}$, our model generically features a virtually massless photophobic axion-like particle $\eta_{\rm B}$ from the breaking of $U(1)_{\rm B} \subset SU(12)$. It acquires no potential from the Yukawa interactions, since these are $U(1)_{\rm B}$-symmetric, and only has electroweak anomalous couplings. Its effective Lagrangian reduces to

$$\mathcal{L}_{\rm EFT} \supset \frac{1}{2}(\partial \eta_{\rm B})^2 + \frac{\eta_{\rm B}}{f_{\rm B}} \left(\frac{g_{\rm L}^2}{32\pi^2} W\widetilde{W} - \frac{g_{\rm Y}^2}{32\pi^2} Y\widetilde{Y} \right), \tag{3.3.31}$$

where $f_{\rm B} = (2/\sqrt{3}) f/N$, plus derivative interactions with itself and the other NGBs. While its coupling to photons vanish, 1-loop generated interactions to the SM fermions lead to a constraint of order $f \gtrsim 300 \times N$ TeV [161] whereas, under the hypothesis $T_f < f$, the impact on $N_{\rm eff}$ is minimal. The bound on f just quoted is much stronger than those considered before. To obliterate the problem we may get rid of $\eta_{\rm B}$ by gauging $U(1)_{\rm B\text{-}L}$, as suggested for other reasons in subsection 3.3.1. In order for $\eta_{\rm B}$ to be gauged away, all the fundamental scalars should be neutral and right-handed neutrinos should be added in order to ensure gauge anomalies cancellation.

Appendices

3.A General Considerations on the Leading Order Potential

In this appendix we derive the minimization condition for the leading order potential of the neutral scalars. This can be compactly written as in eq. (3.3.10). The extremality conditions read

$$\begin{cases} \frac{\delta V_0}{\delta \Pi_m} e^{i\langle \bar{a}\rangle/f_a N_g} + \frac{\delta V_0^*}{\delta \Pi_m} e^{-i\langle \bar{a}\rangle/f_a N_g} = 0 \\ V_0 e^{i\langle \bar{a}\rangle/f_a N_g} - V_0^* e^{-i\langle \bar{a}\rangle/f_a N_g} = 0. \end{cases} \tag{3.A.1}$$

Combining these equations (we will show below that $|V_0| \neq 0$) we get $e^{i\langle \bar{a}\rangle/f_a N_g} = \pm V_0^*/|V_0|$, so that the above system can be rewritten as

$$\begin{cases} \frac{\delta |V_0|}{\delta \Pi_m} = 0 \\ \sin\left(\frac{\langle \bar{a}\rangle}{N_g f_a}\right) = \mp \frac{\operatorname{Im} V_0}{|V_0|} \end{cases} \tag{3.A.2}$$

from which it is clear that the vacuum is obtained extremizing $|V_0|$, and a necessary condition for the Strong CP problem to be solved is that V_0 is real at the minimum.

To see whether the vacuum is actually at the minimum or at the maximum of $|V_0|$ we must study the Hessian, which reads

$$\mathcal{H} = \begin{pmatrix} M^2_{\Pi\Pi} & M^2_{a\Pi} \\ M^2_{a\Pi} & M^2_{aa} \end{pmatrix} \tag{3.A.3}$$

with

$$[M^2_{\Pi\Pi}]_{mn} = \frac{1}{f^2}\frac{\delta^2 V_0}{\delta \Pi_m \delta \Pi_n} e^{i\langle \bar{a}\rangle/f_a N_g} + \frac{1}{f^2}\frac{\delta^2 V_0^*}{\delta \Pi_m \delta \Pi_n} e^{-i\langle \bar{a}\rangle/f_a N_g} \tag{3.A.4}$$

$$\left[M^2_{a\Pi}\right]_m = \frac{i}{f f_a N_g}\frac{\delta V_0}{\delta \Pi_m} e^{i\langle \bar{a}\rangle/f_a N_g} - \frac{i}{f f_a N_g}\frac{\delta V_0^*}{\delta \Pi_m} e^{-i\langle \bar{a}\rangle/f_a N_g} \tag{3.A.5}$$

$$M^2_{aa} = -\frac{1}{f_a^2 N_g^2} V_0 e^{i\langle \bar{a}\rangle/f_a N_g} - \frac{1}{f_a^2 N_g^2} V_0^* e^{-i\langle \bar{a}\rangle/f_a N_g}.$$

The NGB-axion mixing and the pure axion term are order f/f_a and f^2/f_a^2. For simplicity we will work in the limit $f \ll f_a$, so we can treat them as perturbations, but our conclusions will apply in general. In this way, if $M^2_{\Pi\Pi}$ has no flat directions the lightest eigenvalue approximately reads

$$M_a^2 = M_{aa}^2 - M_{a\Pi}^2 [M_{\Pi\Pi}^2]^{-1} M_{a\Pi}^2 + \mathcal{O}(f^4/f_a^4) \leq M_{aa}^2. \tag{3.A.6}$$

Thus a necessary condition for stability is $M^2_{aa} \geq M^2_a \geq 0$, which translates into [7]

$$V_0 e^{i\langle \bar{a}\rangle/N_g f_a} = -|V_0|. \tag{3.A.7}$$

This tells that the absolute minimum of the potential is reached when $V^{\rm LO}_{\rm neutral} = 2V_0 e^{i\langle\bar{a}\rangle/N_g f_a} = -2|V_0|$, and hence the vacuum is obtained by maximizing $|V_0|$. This justifies the earlier assumption $V_0 \neq 0$, for $V_0 = 0$ would be energetically disfavored. The conclusion just derived has been obtained for $f \ll f_a$ but in fact has general validity because a mass mixing always pushes the lightest eigenstate to lower values, so the condition $M^2_{aa} \geq 0$ is anyway necessary. The presence of flat directions in $M^2_{\Pi\Pi}$ would not alter the conclusion either. Indeed, for the same reason we just explained these directions cannot mix with the axion otherwise they would turn tachyonic after the mixing is removed. It follows that Π flat directions cannot affect the axion mass nor the argument leading to (3.A.7).

In summary, we demonstrated in complete generality that the minimum of $V^{\rm LO}_{\rm neutral}$ is obtained by maximizing $|V_0|$ with respect to the Nambu-Goldstone fields. The axion vev follows. Indeed, writing $V_0 = |V_0(\langle\Pi\rangle)| e^{i\phi(\langle\Pi\rangle/f)}$, eq. (3.A.7) indicates that the value of the axion at the minimum is determined by $\langle\bar{a}\rangle/N_g f_a + \phi = \pi \bmod 2\pi$.

The solution of the Strong CP problem requires $N_g(\pi - \phi)$ be a multiple of 2π, and this cannot be assessed unless an explicit form of V_0 is given. In subsection 3.3.2 we analyzed in detail the potential of our scenario. Here, as a quick check of our results, we study the leading order potential for the standard axion in 2-flavour QCD. Note that the same structure (3.3.11) applies to the standard QCD axion provided $\widehat{Y}_u$ is interpreted as the quark mass matrix and A the pion matrix. In the QCD case the axion is rotated such that the quark masses are positive, i.e. $m_{u,d} > 0$, and the axion field appears as in (3.3.10) with $N_g = 2$, see (3.1.9). Switching off the charged pion components we have $V_0 = C[m_u e^{i\pi_0/f_\pi} + m_d e^{-i\pi_0/f_\pi}]$, with C some constant. For $m_u \neq 0$ we find that $|V_0|$ has two extrema, one at $\langle\pi_0\rangle = 0$ and the other at $\langle\pi_0\rangle = \pi/2$. At the two extrema the function $|V_0|$ is respectively given by $m_u + m_d$ and $|m_u - m_d|$. Hence the absolute maximum of $|V_0|$ is attained when $\langle\pi_0\rangle = 0$ and the Strong CP problem, as well-known, is solved, i.e. $\langle\bar{a}\rangle/f_a = 2\pi \sim 0$. When $m_u = 0$ the function $|V_0| = |C| m_d$ is constant and $\phi = \arg V_0$ arbitrary. The argument above tells us that the axion is now a flat direction, as expected.

Chapter 4

UV Solutions

This chapter is devoted to the study of UV solutions to the Strong CP problem, namely solutions based on spontaneous P/CP violation. In section 4.1 we briefly review the main ideas behind this approach. Then, in sections 4.2, 4.3, we specialize to the case of Nelson-Barr models. We first perform a comprehensive analysis of the Nelson-Barr low-energy effective theory, identifying the main criticalities associated with these models. Subsequently, in section 4.3, we propose a set of highly predictive UV scenarios designed to address these challenges. Section 4.2 and appendices 4.A, 4.B are taken from the original work [8], while section 4.3 is taken from the original work [9].

4.1 Spontaneous P/CP Violation

UV solutions to the Strong CP problem assume that some generalized notion of CP is an exact symmetry of whatever completion of the SM is there in the far UV. Being the QCD topological angle CP-odd, this condition completely eliminates the Strong CP problem at the UV completion scale. The phenomenological fact that at energies around and below the electroweak scale we observe P and CP violation, however, implies that this symmetry cannot remain exact, and thus should be spontaneously broken at some scale $\Lambda_{\not{C}\not{P}}$. At low energies the theory is expected to reproduce the SM parameters and its particle content without fine-tunings. Yet, as noted in chapter 1, P and CP violation in the SM are highly non-generic. Therefore, the challenge that these models encounter is that of explaining $\mathcal{O}(1)$ P and CP violation in the weak sector while simultaneously suppressing it in the strong sector.

Crucially, these solutions stand up against the Peccei and Quinn one as there is no quality problem associated to generalized CP. Indeed, the latter can be embedded into a gauge symmetry in extra-dimensional theories of gravity [162, 163], and as such quantum gravity effects cannot lead to effective operators encoding its explicit violation [117, 164].

From a quantum field theory point of view, the existence of an exact generalized CP symmetry requires the action to be invariant under the most general CP transformation (2.2.16),

$$\psi_i(x) \to U_{ij}\epsilon\psi_j^\dagger(\mathcal{P}x), \qquad \phi_a(x) \to O_{ab}\phi_b(\mathcal{P}x), \qquad A_\mu^A(x) \to R^{AB}\mathcal{P}_\mu^\nu A_\nu^B(\mathcal{P}x), \tag{4.1.1}$$

for some matrices U, O, R respectively of $U(N_\psi)$, $O(N_\phi)$, $\mathcal{G}_{\rm glob}$ (see chapter 2). This requirement strongly constraints the form of the renormalizable and non-renormalizable interactions of the theory. Notably, in the Standard Model it is not possible to find any definition of CP which leaves the action invariant. As such its definition is ambiguous. Nevertheless, typically one refers to P as the would-be symmetry flipping the chirality of the fermions: in terms of four-component fermions, $\psi_L \leftrightarrow \psi_R$ (equivalently $\psi \to \gamma^0\psi$). This definition is clearly violated in the full SM because of the weak sector begin chiral. However, much below the electroweak scale the heavy $W^\pm, Z$ and Higgs' bosons can be integrated out, leading to the well-known $SU(3)_{\rm C} \times U(1)_{\rm EM}$ theory (2.1.39). This is invariant under the previous definition of P except for the topological angles and non-renormalizable operators, and thus justifies the origin of the conventional notion of P. Actually, the low-energy theory is invariant under an additional symmetry called charge-conjugation (C), sending $\psi \to \psi^c = -i\gamma^2\bar{\psi}^*$. Since ψ^c has the same Lorentz transformation properties as ψ, we can generalize the previous notion of P by embedding it with C to get another (almost) conserved symmetry, CP: $\psi \to -i\gamma^0\gamma^2\bar{\psi}^*$. Its action is actually equivalent to time-reversal (T) due to the CPT theorem [165]. In terms of Weyl fermions CP is much simpler as it amounts to taking $U = \mathbb{1}$ in (4.1.1) in the mass basis of the quarks. Unlike P, however, the weak gauging of the full SM does not violate this symmetry. The entire breaking comes from the Yukawas:

$$
\begin{aligned}
Y_d\,\bar{Q}_L H d_R + Y_d^\dagger\,\bar{d}_R H^\dagger Q_L &\xrightarrow{\rm CP} Y_d^*\,\bar{Q}_L H d_R + (Y_d^*)^\dagger\,\bar{d}_R H^\dagger Q_L \\
Y_u\,\bar{Q}_L \widetilde{H} u_R + Y_u^\dagger\,\bar{u}_R \widetilde{H}^\dagger Q_L &\xrightarrow{\rm CP} Y_u^*\,\bar{Q}_L \widetilde{H} u_R + (Y_u^*)^\dagger\,\bar{u}_R \widetilde{H}^\dagger Q_L
\end{aligned}
\tag{4.1.2}
$$

where we also let $H \xrightarrow{\rm CP} H^*$. Thus, CP in the SM is only broken by the fact that the Yukawas are complex (modulo $SU(N_\psi)$ rotations). This is consistent with the fact that real Yukawas would identically imply $I_{\not{C}\not{P}} = \det\left[Y_u Y_u^\dagger, Y_d Y_d^\dagger\right] = 0$, or equivalently $\delta = 0$ as the diagonalization of the Yukawas would require only orthogonal matrices (instead of unitary) so that the resulting CKM matrix would be real.

Given the conventional notions of P and CP, models of spontaneous P or CP violation are typically distinguished according to which of these two symmetries (or a close-enough generalization) are promoted to actual symmetries of the theory upon extending the SM content.

Spontaneous P violation typically requires the introduction of chiral-partners of the SM fermions, whose exact charges depend on the specific implementation of P. The most straightforward way to implement such a symmetry is to double the SM fermions as to associate a right-handed partner to every left-handed fermion and vice-versa, basically making them vector-like. In this way parity can be enforced as an exact symmetry and $\bar{\theta} = 0$ above $\Lambda_{\not{P}}$. Since these partners have not been observed they must be quite heavy, and the mass splitting can be obtained by introducing an additional Higgs field. However, the gauge degrees of freedom of the weak sector are not enough to make the two Higgs vevs' simultaneously real, and thus an unacceptable $\mathcal{O}(1)$ $\bar{\theta}$ is generated by the splitting. Another possibility is to extend the SM gauge group to $SU(3)_{\rm C} \times SU(2)_{\rm L} \times SU(2)_{\rm R} \times U(1)_{\rm Y'}$, in such a way that the SM right-handed fermions are embedded in $SU(2)_{\rm R}$ doublets and play the role of the partners.

Parity also exchanges $SU(2)_\mathrm{L} \leftrightarrow SU(2)_\mathrm{R}$, and the most general Yukawas of the quarks read

$$-\mathcal{L}_\mathrm{Yuk} = h\,\bar{Q}_L \Phi Q_R + \widetilde{h}\,\bar{Q}_L \widetilde{\Phi} Q_R + \text{h.c.} \tag{4.1.3}$$

where the Higgs field Φ sits in the bifundamental of $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ ($\widetilde{\Phi} = \sigma^2 \Phi^* \sigma^2$). The matrices $h, \widetilde{h}$ are Hermitian as a consequence of P, and at this stage $\bar{\theta} = \theta + \arg\det h\widetilde{h} = 0$. Unfortunately, to obtain a phenomenologically acceptable low-energy theory additional scalar fields are required[1] and one encounters the same problem as before, namely their vevs' induce a tree-level $\bar{\theta}$ [166, 167]. The way out of this problem is then to introduce new partners to each SM fermion (and to the Higgs), instead of grouping the right-handed ones in $SU(2)_\mathrm{R}$ doublets. Depending on the eventual introduction of an additional $U(1)_{\mathrm{Y}''}$ and on the exact charges of the partners many possibilities open [168]. The by far most popular one is the Babu-Mohapatra model [169], in which no $U(1)_{\mathrm{Y}''}$ is present and the partners have exactly the same charges of the SM fermions except $SU(2)_\mathrm{L} \leftrightarrow SU(2)_\mathrm{R}$. This model has been quite extensively investigated in literature [170–173], and despite the fact in the "see-saw" regime its viability seems comprised due to big radiative corrections to $\bar{\theta}$ [173], the opposite regime might still offer a valid alternative. As a variation of this minimal and popular idea one could introduce an additional $U(1)_{Y''}$. In this way the partners' sector is completely decoupled from the SM (except for color gauge interactions and a quartic Higgs portal), and the radiative corrections to $\bar{\theta}$ are basically identical as in the SM and therefore extremely suppressed. However, the partners are exactly stable due to the unbroken $U(1)_{\mathrm{EM}'}$ and they lead to an abundance of fractionally charged hadrons conflicting with observations [174]. Thus either this symmetry must be spontaneously broken, leading again to an effective Babu-Mohapatra model, or a non-standard thermal history of the universe is required [175]. In an additional variation also color is doubled, and tuning the scale of breaking to the vectorlike subgroup identified with QCD can lead to a completely different phenomenology with respect to the usual Babu-Mohapatra model [176, 177].

The spontaneous CP violation approach promotes a generalization of the conventional notion of CP to an exact symmetry of the theory. The fundamental difference with respect to P solutions is that in this case the entire SM fermionic spectrum need not be "paired". This allows the spontaneous breaking of CP to occur in a completely independent secluded sector, being then mediated to the SM thanks to some messenger fields as depicted in figure 4.1. The non-trivial task that this approach faces, however, is that the mediation must be able to generate an $\mathcal{O}(1)$ CKM phase while simultaneously keeping $\bar{\theta}$ under control[2]. Within ordinary 4-dimensional theories, the virtually unique scenarios that have a chance of accomplishing this are Nelson-Barr models [178, 179]. These models exploit the key observation that the CKM phase appears in flavour-violating observables whereas the topological angles control CP violation in flavour-invariant processes. Thus, they are are able to produce a sizable

[1] The vev $\langle\Phi\rangle$ can always be made diagonal by means of an $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ rotation. Therefore, the surviving low-energy gauge symmetry is $SU(3)_\mathrm{C} \times SU(2)_\mathrm{V} \times U(1)_{\mathrm{Y}'}$ or $SU(3)_\mathrm{C} \times U(1)_{\mathrm{Y}'} \times U(1)_{\mathrm{Y}''}$, neither of which is phenomenologically acceptable.

[2] The problem of generating an $\mathcal{O}(1)$ CKM phase is not there for spontaneous P violation models, as enforcing P does not necessarily requires the Yukawas to be real (e.g. in the model (4.1.3) it is sufficient for them to be Hermitian). Enforcing CP implies that there exist a basis in which all the couplings are real, and thus the CKM phase must be dynamically generated.

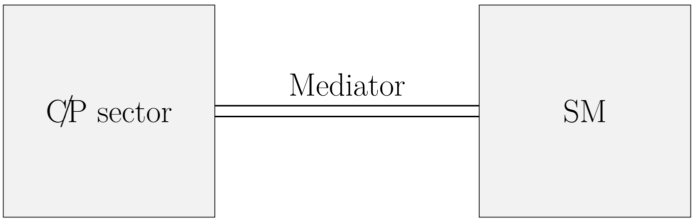

Figure 4.1: Schematic representation of the logic behind models of spontaneous CP violation. The spontaneous breaking of CP happens in a secluded sector and is communicated to the SM through a messenger field.

CKM phase and a small QCD angle by communicating the spontaneous violation of CP via flavour-violating couplings. Typically, this occurs through a complex mixing of the quarks with colored mediators. When this is the case, the up- and down- quark mass terms at tree-level take the schematic form [179]

$$
\begin{aligned}
\mathcal{L}_u &\sim \begin{pmatrix}(\mathbf{2},+1/6)_{\mathrm{SM}} & (\mathbf{1},+2/3)_{\mathrm{BSM}} & (\mathbf{2},+1/6)_{\mathrm{BSM}}\end{pmatrix} \begin{pmatrix} y_u H & 0 & \xi_{1/6} \\ \xi_{2/3} & M_{2/3} & 0 \\ 0 & 0 & M_{1/6} \end{pmatrix} \begin{pmatrix} (\mathbf{1},-2/3)_{\mathrm{SM}} \\ (\mathbf{1},-2/3)_{\mathrm{BSM}} \\ (\mathbf{2},-1/6)_{\mathrm{BSM}} \end{pmatrix} \\
\mathcal{L}_d &\sim \begin{pmatrix}(\mathbf{2},+1/6)_{\mathrm{SM}} & (\mathbf{1},-1/3)_{\mathrm{BSM}} & (\mathbf{2},+1/6)_{\mathrm{BSM}}\end{pmatrix} \begin{pmatrix} y_d \widetilde{H} & 0 & \xi_{1/6} \\ \xi_{1/3} & M_{1/3} & 0 \\ 0 & 0 & M_{1/6} \end{pmatrix} \begin{pmatrix} (\mathbf{1},+1/3)_{\mathrm{SM}} \\ (\mathbf{1},+1/3)_{\mathrm{BSM}} \\ (\mathbf{2},-1/6)_{\mathrm{BSM}} \end{pmatrix}
\end{aligned}
\tag{4.1.4}
$$

where we identified the fermionic fields[3] by their $SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ charges, distinguishing the SM ones from the BSM mediators (vector-like with respect to the SM gauge group). The matrices $y_{u,d}, M_{1/6}, M_{1/3}, M_{2/3}$ are by assumption CP-even and can thus be taken as real, whereas $\xi_{2/3}, \xi_{1/3}, \xi_{1/6}$ represent the complex mixings mediating CP violation to the SM. Because of this particular structure, it is straigthforward to check that the determinant of the full mass matrices involves only the CP-even subcomponents. Hence it is real, ensuring that at tree-level $\bar{\theta} = 0$ with an in general $\mathcal{O}(1)$ CKM phase as result of the mass matrix being complex [179]. Two main concerns arise. First, the property that at tree-level this structure is capable of setting $\bar{\theta} = 0$ may be spoiled by quantum effets, which may alter the form of the mass matrix and reintroduce an unacceptably big topological angle. Second, the qualitative intuition that it is indeed possible to feed enough CP violation in the CKM phase only through the mixing should be quantitatively investigated. In the remaining sections of this chapter, based on the original works [8, 9], we will prove that certain class of realizations are indeed capable of robustly solving the Strong CP problem while simultaneously being phenomenologically acceptable. In passing we will discuss how a new potential naturalness problem emerges in these models, and subsequently we will provide an explicit UV completion facing this issue and at the same time being highly predictive. For additional recent works on the topic see [170, 180–188].

[3]The fermions are written as Weyl spinors to make contact with the next section.

4.2 Nelson-Barr Models

The basic assumption underlying the Nelson-Barr approach to the Strong CP problem is that CP is a good symmetry of the UV. This means that in the effective field theory below some UV cutoff there exists a field basis in which all topological angles vanish and all couplings are real. CP is spontaneously broken by a color-neutral sector, whose relevant degrees of freedom are CP-odd scalars Σ with non-vanishing vacuum expectation value, and then communicated to the SM via a mediator sector of fermionic particles ψ characterized by a (CP-conserving) mass scale m_ψ. To keep the radiative corrections to the QCD theta angle under control, the CP-odd scalars should interact with the SM only via couplings that carry spurionic charges under the SM flavour symmetry. These couplings cannot involve the doublet quark representation, which would lead to an unacceptably big $\bar{\theta}$ [189]. Within a perturbative framework, these requests are so significant that basically leave two (non-exclusive) options: either Σ couple to the singlet up or the singlet down quark representations, i.e. via the effective mixings $\xi_{1/3}$ or $\xi_{2/3}$ of (4.1.4). We will refer to these scenarios as models of $d-$ and $u-$mediation [8]. It is also possible to consider linear combinations of these two.

Explicitly, the Lagrangian of the minimal models with $d-$mediation consists of the obvious kinetic terms, including the gauge interactions, a potential $V(\Sigma, H)$ for the scalars, the Yukawas and the mass terms. Adopting the notation of [8], the latter read[4]

$$\begin{aligned} -\mathcal{L}^d_{\text{Yuk}} &= y_u\, qHu + y_d\, q\widetilde{H}d \\ &\quad + y\,\psi\Sigma d + m_\psi\,\psi\psi^c + \text{h.c.} \qquad (d-\text{mediation}), \end{aligned} \tag{4.2.1}$$

where y_u, y_d, y, m_ψ are CP-even matrices in flavour space, the theta angles vanish by CP, and ψ (ψ^c) has SM charges conjugate (equal) to d. Summation over the family and gauge indices is understood. A completely analogous Lagrangian can be written for minimal models with $u-$mediation. The only difference is that d in (4.2.1) is replaced by the electroweak singlet up representation and ψ (ψ^c) should have SM charges conjugate (equal) to u:

$$\begin{aligned} -\mathcal{L}^u_{\text{Yuk}} &= y_u\, qHu + y_d\, q\widetilde{H}d \\ &\quad + y\,\psi\Sigma u + m_\psi\,\psi\psi^c + \text{h.c.} \qquad (u-\text{mediation}), \end{aligned} \tag{4.2.2}$$

We emphasize that in either case Σ should be a family of at least two scalars. If there were just one scalar, the CP-odd phase in its vacuum expectation value could be removed from (4.2.1) via a re-definition of ψ and ψ^c, so no CKM phase could be induced. The fermions ψ, ψ^c could come in a family of fields, as well, though this is not strictly necessary (we will discuss this possibility in subsection 4.2.3). The scenarios in (4.2.1), (4.2.2) are particular incarnations of the Nelson-Barr class first proposed in [190], as clear from the comparison with (4.1.4).

As we will demonstrate in 4.2.1, scenarios of $d-$ and $u-$mediation ensure that the corrections to the QCD theta parameter arising from the messenger sector are sufficiently small. These are *irreducible corrections*, since they are due to the very same messenger sector that is responsible for transferring CP violation to the SM, i.e. generating the CKM phase. To

[4] In this and the following section fermions are taken as 2-component Weyl spinors.

avoid larger contributions to $\bar{\theta}$, some symmetry must be invoked to forbid $q\widetilde{H}\psi^c$ in (4.2.1) (or $qH\psi^c$ in models with $u-$mediation). On the other hand, interactions of the CP-violating sector with the leptons are basically unconstrained phenomenologically.

Other, *reducible corrections* to $\bar{\theta}$ generically arise from loops involving additional states, but these can all be naturally suppressed [178]. As a concrete example let us consider the corrections to $\bar{\theta}$ from loops involving excitations of the CP-violating sector in the scenarios (4.2.1). All symmetries of the theory allow a quartic scalar interaction with the Higgs doublet $\lambda_{mn}\Sigma_m^\dagger\Sigma_n|H|^2 \subset V(\Sigma, H)$. Once this is taken into account it is easy to verify that one obtains 1-loop corrections of the type [190]

$$\bar{\theta} = \frac{c_y}{16\pi^2}\,\mathrm{Im}\left[\langle\Sigma\rangle^\dagger\,\lambda\frac{1}{m_\Sigma^2}y^*y^t\,\langle\Sigma\rangle\right] \tag{4.2.3}$$

for some real number $|c_y| \sim 1$. These would be unacceptably large for generic couplings of order one. However it is possible to naturally suppress (4.2.3) taking $|y| \ll 1$. Completely analogous considerations hold for other *reducible* contributions to $\bar{\theta}$, such as those from the gauge sector of the original model of [178]. In the following we will study the models in equations (4.2.1) and (4.2.2) in the simplifying limit $|y| \ll 1$. In this limit the fluctuations of the CP-breaking sector can be ignored, see (4.2.3), and the corrections to $\bar{\theta}$ arise entirely from loops of the SM and the mediator sector. The latter effects are *irreducible* and therefore fundamentally characterize this approach to the Strong CP problem.

4.2.1 Reproducing the Standard Model: a Coincidence of Scales

Models of $d-$ and $u-$mediation are very efficient at taming the irreducible corrections to $\bar{\theta}$, as we will see in 4.2.2. However, any construction of the Nelson-Barr type must satisfy a highly non-trivial condition in order to reproduce the SM at scales below the messengers mass. Specifically, the two scales $y\langle\Sigma\rangle, m_\psi$ should be comparable to each other

$$\mathrm{Im}\,y\,\langle\Sigma\rangle \sim \mathrm{Re}\,y\,\langle\Sigma\rangle \sim m_\psi, \tag{4.2.4}$$

within a few orders of magnitude, depending on the model. We will see shortly that $|y\,\langle\Sigma\rangle\,| \gtrsim |m_\psi|$ is necessary to generate a sizable CKM phase, whereas the more subtle $|y\,\langle\Sigma\rangle \lesssim |m_\psi|$ is required to reproduce the quark mass spectrum within a reliable perturbative description. To appreciate the origin of (4.2.4) we first focus on the model in eq. (4.2.1) with $|y| \ll 1$, as anticipated above.

The Yukawa part of the Lagrangian (4.2.1), under our assumption that the only remnant of the CP-violating sector is the vacuum expectation value of Σ, can be written as

$$\begin{aligned}\mathcal{L}^d_{\text{Yuk}}\Big|_{\text{frozen}} &= -y_u\,qHu - y_d\,q\widetilde{H}d - \xi^\dagger\psi d - m_\psi\,\psi\psi^c + \text{h.c.}\\ &= -y_u\,qHu - \begin{pmatrix} q & \psi\end{pmatrix}\begin{pmatrix} y_d\widetilde{H} & 0\\ \xi^\dagger & m_\psi\end{pmatrix}\begin{pmatrix} d\\ \psi^c\end{pmatrix} + \text{h.c.},\end{aligned} \tag{4.2.5}$$

where we introduced the column vector ξ:

$$\xi_i^* = y_{mi}\,\langle\Sigma_m\rangle\,. \tag{4.2.6}$$

In this setup CP violation is due entirely to ξ. In other words, there exists a field basis in which y_u, y_d, m_ψ are real, the theta angles vanish, and the only complex quantity is ξ.

Because ψ, ψ^c are colored, collider bounds force them to lie above the TeV scale (see 4.2.4). It then makes sense to diagonalize the mass matrix neglecting the electroweak scale in a first approximation. Performing the following $SU(4)$ transformation

$$\begin{pmatrix} d \\ \psi^c \end{pmatrix} \to \begin{pmatrix} 1 - \frac{\xi\xi^\dagger}{|\xi|^2}\,(1 - \frac{m_\psi}{M}) & \frac{\xi}{M} \\ -\frac{\xi^\dagger}{M} & \frac{m}{M} \end{pmatrix} \begin{pmatrix} d \\ \psi^c \end{pmatrix} \tag{4.2.7}$$

the Yukawa sector becomes

$$\begin{aligned} \mathcal{L}^d_{\text{Yuk}}\Big|_{\text{frozen}} &\to - Y_u\, qHu - Y_d\, q\widetilde{H}d - Y\, q\widetilde{H}\psi^c - M\,\psi\psi^c + \text{h.c.} \\ &= - Y_u\, qHu - \begin{pmatrix} q & \psi \end{pmatrix} \begin{pmatrix} Y_d\widetilde{H} & Y\widetilde{H} \\ 0 & M \end{pmatrix} \begin{pmatrix} d \\ \psi^c \end{pmatrix} + \text{h.c.} \end{aligned} \tag{4.2.8}$$

with (matrix multiplication is understood)

$$\begin{cases} M^2 = \xi^\dagger\xi + m_\psi^2 \\ Y_u = y_u \\ Y_d = y_d \left[1 - \frac{\xi\xi^\dagger}{|\xi|^2}\,(1 - \frac{m_\psi}{M})\right] \\ Y = y_d \frac{\xi}{M} = Y_d \frac{\xi}{m_\psi}. \end{cases} \tag{4.2.9}$$

After the unitary rotation (4.2.7) the heavy degrees of freedom have real masses M and complex Yukawa couplings Y. At scales much smaller than M we can integrate them out and recover the SM including higher-dimensional operators suppressed by inverse powers of M. The couplings Y_u and Y_d are the SM Yukawas, up to small loop effects. At the renormalizable level, CP violation is encoded in the complex matrix Y_d, or more precisely in a tree-level CKM phase, and a radiatively generated theta angle $\bar{\theta}$. They will be our focus next.

The CKM Phase and Perturbativity

A CKM phase compatible with the experimentally determined value can be reproduced as long as ξ/m_ψ has imaginary and real entries of comparable order of magnitude satisfying

$$\left|\frac{\xi}{m_\psi}\right| = |Y_d^{-1}Y| \gtrsim 1. \tag{4.2.10}$$

Before proving (4.2.10) let us attempt to derive the CKM phase in the limit $|\xi| \ll |M|$, where analytical calculations can be easily performed.

To do this we move to the basis in which Y_u is diagonal, where the CKM matrix is the unitary matrix that diagonalizes $Y_d Y_d^\dagger = y_d(1 - \xi\xi^\dagger/M^2)y_d^\dagger$. For $\xi = 0$ the latter is just the orthogonal matrix that diagonalizes $y_d y_d^\dagger$. Equivalently, the SM masses and mixing angles are determined by $Y_u = y_u$ and $Y_d = y_d$ and the CKM phase vanishes. We can thus express the corresponding real CKM matrix using the Wolfenstein parametrization [191] with η set to zero. For non-zero $|\xi\xi^\dagger| \ll M^2$ the mixing angles are corrected at order $|\xi|^2/M^2$ whereas a small $\eta \propto \mathrm{Im}\,\xi\xi^\dagger/M^2$ is generated. Since the invariant J in (1.2.4), measuring the strength of weak CP violation in the SM, is linear in η, the real part of ξ can only enter at subleading $|\xi|^4/M^4$ order. Now, plugging (4.2.9) in (1.2.4) and Taylor expanding J in powers of $\xi\xi^\dagger/M^2$ we obtain, at leading order, [8]

$$\begin{aligned} J &= A(1-\rho)\frac{m_s}{m_b}\lambda_C^4\frac{\mathrm{Im}\,\xi_2\xi_3^\dagger}{M^2}\left[1+\mathcal{O}\left(\frac{|\xi|^2}{M^2},\lambda_C\right)\right] \\ &\simeq 4.1\times 10^{-5}\,\frac{\mathrm{Im}\,\xi_2\xi_3^\dagger}{M^2}\left[1+\mathcal{O}\left(\frac{|\xi|^2}{M^2},\lambda_C\right)\right]. \end{aligned} \tag{4.2.11}$$

In deriving (4.2.11) we also exploited the numerical relations $m_d \sim m_b\lambda_C^4$, $m_s \sim m_b\lambda_C^2$, $m_u/m_t \sim \lambda_C^7$, $m_c/m_t \sim \lambda_C^4$ (the masses are renormalized at ~ 1 TeV), and expanded in the Cabibbo angle. The factor of m_s/m_b in (4.2.11) originates from the fact that the phases in $Y_d Y_d{}^\dagger$ are controlled by the off-diagonal elements of the hermitian matrix $y_d\,\mathrm{Im}\,[\xi\xi^\dagger/M^2]y_d{}^\dagger$, and so they disappear when $m_s/m_b, m_d/m_b \to 0$.

Equation (4.2.11) suggests that the experimental value of J reported in section 1.2 can be reproduced provided $\mathrm{Im}\,\xi_2\xi_3^\dagger/M^2 \sim 0.73$. But there is a serious problem with the estimate (4.2.11): it is not possible to satisfy $|\xi_2||\xi_3|/M^2 \sim 0.73$ compatibly with the constraint $|\xi|^2 = M^2 - m_\psi^2 \leq M^2$. We should conclude that the value of $|\xi|/M$ needed to apparently reproduce the observed CKM phase with (4.2.11) is too large for the perturbative expansion used to be reliable [192]. Some non-perturbative technique must be employed to determine whether these models can or cannot generate the CKM phase. And this is complicated by the fact that with $|\xi| \sim M$ the CKM mixing angles are not just functions of y_d, but can be significantly affected by ξ/M.

Fortunately there is a way to basically "integrate out" y_d from the problem and obtain necessary and sufficient conditions on ξ for Nelson-Barr models to reproduce the CKM phase. The argument goes as follows. We want to explicitly compute J using the tree-level approximation $Y_{u,d}^{\mathrm{SM}} = Y_{u,d}$. Employing (4.2.9) and simple algebraic manipulations we get

$$\begin{aligned} \det\,[H_u, H_d] &= \det\left[h_u, h_d - YY^\dagger\right] \\ &= Y^\dagger\left[h_u, [h_u, h_d]^2\right]Y - Y^\dagger Y\; Y^\dagger h_u\,[h_u, h_d]\,h_u Y \\ &\quad - Y^\dagger h_u^2 Y\; Y^\dagger\,[h_u, h_d]\,Y + Y^\dagger h_u Y\; Y^\dagger\,\{h_u, [h_u, h_d]\}\,Y, \end{aligned} \tag{4.2.12}$$

where we defined $H_{u,d} = Y_{u,d}Y_{u,d}{}^\dagger$ and $h_{u,d} = y_{u,d}y_{u,d}{}^\dagger$. In the first line of (4.2.12), $H_d = h_d - YY^\dagger$ follows from (4.2.9). The second equality is a consequence of the fact that, for any traceless matrix C, $\det C = \mathrm{tr}\,C^3/3$. In our case $C = [h_u, h_d] - \left[h_u, YY^\dagger\right]$, with $\det\,[h_u, h_d] = 0$ because $h_{u,d}$ are CP-even. The non-vanishing terms in $\det C$ are traces containing powers of $\left[h_u, YY^\dagger\right]$, and can therefore be written as $Y^\dagger f Y$ for some anti-symmetric function f of $h_{u,d}$. The next step towards our necessary and sufficient conditions relies on the observation that the functions f can be equivalently re-written in terms of $H_{u,d}$ by re-using $H_d = h_d - YY^\dagger$. Note that this replacement should be carried out uniquely in f, and not in the $Y, Y^\dagger$ of (4.2.12), otherwise we would obviously get back to the left-hand side of (4.2.12). This replacement leads us to an important relation, which we write as follows

$$\det\left[H_u, H_d\right] = I_{2,1} + Y^\dagger Y\ I_{1,2} + Y^\dagger H_u^2 Y\ I_{1,0} - Y^\dagger Y\ I_{1,1} \quad (d\text{−mediation}) \tag{4.2.13}$$

where

$$\begin{aligned} I_{2,1} &= Y^\dagger \left[H_u, [H_u, H_d]^2\right] Y \\ I_{1,2} &= Y^\dagger H_u \left[H_u, H_d\right] H_u Y \\ I_{1,1} &= Y^\dagger \left\{H_u, [H_u, H_d]\right\} Y \\ I_{1,0} &= Y^\dagger \left[H_u, H_d\right] Y. \end{aligned} \tag{4.2.14}$$

Because, within a tree-level approximation, $Y_{u,d}$ are the SM Yukawa matrices, we can express them in a convenient form, say taking a diagonal $Y_u = \widehat{Y}_u$ and writing $Y_d = V^*_{\text{CKM}} \widehat{Y}_d$, where V_{CKM} is the CKM matrix in the Wolfenstein parametrization[5]. Then eq. (4.2.13) must be interpreted as a *constraint* on the coupling Y, or equivalently on ξ/m_ψ. As promised, the dependence on y_d is included but implicit, i.e. y_d has been integrated out.

One can then easily solve (4.2.13) via numerical integration. The parameter ξ is defined by 2 angles, a modulus and three phases. However the overall phase can be removed by a vector-like rotation of the mediators, so in practice only two of its phases are physical. Scanning over 300 angles and phases we obtain the $|\xi|/m_\psi$ vs η plot shown in the upper part of figure 4.2. For models of $d-$mediation (where (4.2.13) has been derived) we find that $|\xi|/m_\psi \gtrsim 2$ is necessary. The analytical approximation (4.2.11) works very well for $|\xi| \ll M$ but becomes inadequate rather quickly as ξ increases and fails to account for the large spread seen in the upper part of figure 4.2. The numerical analysis also shows that ξ_1 is not very important and that ξ_2, ξ_3 should be comparable in size and have large phases. The irrelevance of ξ_1 is an expected consequence of our choice of basis, since the last equation in (4.2.9) says that ξ_1 appears in (4.2.13) multiplied by the smallest of the down-type Yukawas or larger powers of the Cabibbo angle. The main players are clearly ξ_2, ξ_3, as anticipated by (4.2.11). To obtain the observed J from (4.2.13), the CP-odd contributions due to terms $\propto \operatorname{Im}\xi\xi^\dagger/m_\psi^2$ in the structures $Y_2^* f_{23} Y_3$ should win over the terms $\propto \eta$ in terms like $Y_3^* \lambda_C^2 f_{23} Y_3$ (see, e.g., the expression of the invariant $I_{2,1}$ in appendix 4.A). This requires [8]

$$\frac{\operatorname{Im}\xi_2\xi_3^*}{\xi_3\xi_3^*} \frac{[\widehat{Y}_d]_2}{[\widehat{Y}_d]_3} \gtrsim \lambda_C^2. \tag{4.2.15}$$

Because we approximately have $[\widehat{Y}_d]_2/[\widehat{Y}_d]_3 \sim \lambda_C^2$, we see that (4.2.15) is satisfied for $|\xi_2| \sim |\xi_3|$. This condition is visible in the lower-left plot of fig. 4.2. Barring accidental correlations between y_u, y_d and ξ, these findings imply that in a generic basis all entries in ξ should be comparable and satisfy $|\operatorname{Re}\xi| \sim |\operatorname{Im}\xi| \gtrsim m_\psi$. This proves the claim (4.2.10).

[5] The conjugation is a consequence of the definition of the Yukawa interactions (4.2.8), which for typographical convenience differ here from the ones of equation (1.1.2) and (3.3.4) exactly by a complex conjugation.

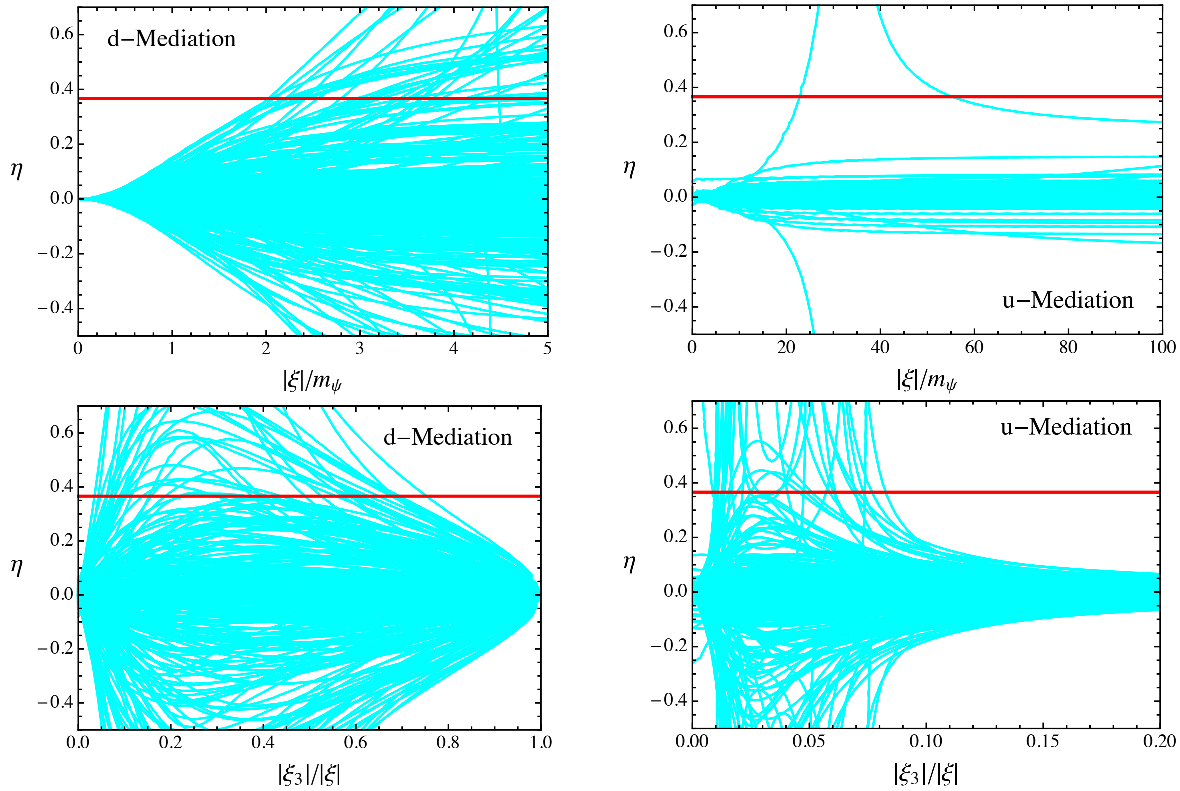

Figure 4.2: Value of the CP-odd parameter η of the Wolfenstein parametrization of the CKM (recall that $J = A^2\eta\lambda_C^6[1+\mathcal{O}(\lambda_C^2)]$) in models of $d-$mediation (left) and $u-$mediation (right). In the upper plots we generated 300 models with random values for the direction (2 angles) and the 2 physical phases of ξ, and kept an arbitrary dependence on the modulus $|\xi|/m_\psi$. Note the difference in the scale of the x-axis between the two plots. In the lower plots we also scanned over the modulus in the range $0 < |\xi|/m_\psi < 5$ ($d-$mediation), $20 < |\xi|/m_\psi < 40$ ($u-$mediation), and kept $|\xi_3|/|\xi|$ arbitrary. The red line indicates the real world value $J \simeq 3.0 \times 10^{-5}$ and the cyan lines are the models' predictions. Plots taken from [8].

In $u-$mediation, a completely analog procedure leads to a relation similar to (4.2.13) with the replacement $u \leftrightarrow d$. The field basis we adopt is now the one with Y_d diagonal and $Y_u = V^t\widehat{Y}_u$. Only for very few choices of angles and $|\xi|/m_\psi \gtrsim 20$ we can reproduce the CKM phase, as visible from the top-right plot of fig. 4.2. The basic reason can be traced back to the larger mass hierarchy of the up-quark sector. As a consequence, for example, in the $u-$mediation version of (4.2.11) one finds a more significant m_c/m_t suppression replaces m_s/m_b. As before, ξ_1 is irrelevant, but to satisfy the analog of (4.2.15) one must have $|\xi_3|/|\xi_2| \lesssim m_c/(m_t\lambda_C^2) \sim 0.07 \sim \lambda_C^2$. This expectation is confirmed by the bottom-right plot of fig. 4.2. We conclude that models of $u-$mediation can reproduce the observed CKM phase provided their UV completion features some sort of anti-correlation between y_u and ξ.

Equation (4.2.10) is our first step towards (4.2.4). Analyzing the effective theory below

M carefully, one finds there is also an *upper bound* on $|\xi|/|m_\psi|$. This is in fact a necessary condition if the effective theory has to reproduce the observed SM particle spectrum. When $|m_\psi|/|\xi| \to 0$ the heavy state becomes a combination of ψ and one component of the d's, with a large Dirac mass $M \sim |\xi|$. The two orthogonal d components have independent Yukawa couplings with only two q's, whereas the remaining q together with ψ^c form a massless Dirac with an anomalous axial symmetry. But this is just the SM with a massless down-quark. We can confirm this observing that

$$\det Y_d = \det y_d \frac{m_\psi}{M}, \tag{4.2.16}$$

or perhaps more easily noting that in the limit $|m_\psi|/|\xi| \to 0$ the SM Yukawa matrix Y_d in (4.2.9) becomes rank 2, $\lim_{|m_\psi|/|\xi|\to 0} Y_d = y_d \left[1 - \xi\xi^\dagger/|\xi|^2\right]$. We see that the limit $m_\psi/|\xi| \to 0$ is phenomenologically unacceptable. Clearly there should be a lower bound on $m_\psi/|\xi|$. Let us see what this is. For small but non-vanishing $|m_\psi|/|\xi| \ll 1$ the mass of the down quark is of order $m_d \sim (\widehat{y}_d v/\sqrt{2})(m_\psi/|\xi|)$, where $\widehat{y}_d$ denotes a typical eigenvalue of y_d and we used $M \sim |\xi|$. It is not possible to establish a firm bound on $m_\psi/|\xi|$ this way, however, because the value of $\widehat{y}_d$ is model-dependent and can in principle range between $[\widehat{Y_d}]_d$ and the non-perturbative $\sim 4\pi$. A robust, model-independent bound on $m_\psi/|\xi|$ can instead be derived from the UV description above M. Inspecting the field basis (4.2.8) we see that the coupling Y between the heavy fermionic state and the SM quark doublet becomes parametrically large when $|\xi| \gg |m_\psi|$, see the third line in (4.2.9). When m_ψ is too small it becomes non-perturbative, say $|Y| > 4\pi$, and we lose predictivity. Because fig. 4.2 showed that $|\xi_i| \sim |\xi_j|$ is necessary to reproduce the CKM phase, the constraint $|Y| \ll 4\pi$ may be expressed as [8]

$$\left|\frac{\xi}{m_\psi}\right| = |Y_d^{-1} Y| \ll \frac{4\pi v}{\sqrt{2} m_b} \sim 10^3 \qquad (d\text{−mediation}). \tag{4.2.17}$$

Accidentally, this upper bound is numerically comparable to what one finds requiring the low energy theory reproduces the observed down-quark mass with $\widehat{y}_d \sim [\widehat{Y}_d]_b$, despite the two bounds have very different meaning. In the next subsection we will be able to quantify the perturbativity bound by inspecting the value of $\bar{\theta}$ predicted by these models.

In models with u−mediation $|\xi|/|m_\psi| \lesssim m_t/m_u \sim 10^5$ is at least necessary to obtain a realistic spectrum if $\widehat{y}_u \sim [\widehat{Y}_u]_t$, from the effective field theory point of view. A stronger bound on $|\xi|/m_\psi$ applies however because we found that the condition $|\xi_3| \lesssim \lambda_C^2 |\xi|$ is necessary to reproduce the CKM phase, see fig. 4.2. Hence, the UV description is non-perturbative unless $|Y| \ll 4\pi$, or more explicitly $[\widehat{Y}_u]_c |\xi|/m_\psi \ll 4\pi$ and simultaneously $\lambda_C^2 [\widehat{Y}_u]_t |\xi|/m_\psi \ll 4\pi$. The latter provides the most stringent bound, which reads [8]

$$\left|\frac{\xi}{m_\psi}\right| = |Y_u^{-1} Y| \ll \frac{4\pi v}{\sqrt{2} m_t\, \lambda_C^2} \sim 300 \qquad (u\text{−mediation}). \tag{4.2.18}$$

A concrete manifestation of this non-perturbativity problem is seen in potentially large radiative corrections to the $\bar{\theta}$ parameter when matching to the SM at scales $\sim M$, which we analyze later.

Combining our findings, we conclude that Nelson-Barr models of $d-$ and $u-$mediation can reproduce the amount of CP violation encoded in the CKM matrix as well as the observed fermion masses provided the two effective scales of the model, the CP-even mass m_ψ and the CP-odd mass mixing ξ between the mediators and the SM quarks, have comparable size [8]:

$$\begin{aligned} 2 \lesssim \left|\frac{\xi}{m_\psi}\right| \ll 10^3 \qquad & (d-\text{mediation}), \\ 20 \lesssim \left|\frac{\xi}{m_\psi}\right| \ll 300 \qquad & (u-\text{mediation}). \end{aligned} \tag{4.2.19}$$

In particular, while scenarios of $d-$mediation can generate the CKM phase with generic complex vectors ξ with entries of comparable order, models of $u-$mediation require some sort of anti-correlation between ξ and y_u. This makes such scenarios less generic than models of $d-$mediation.

The coincidence (4.2.19) is a structural property of Nelson-Barr models, rooted in the way CP violation is communicated to the SM. Such coincidence cannot be addressed via the effective field theory formalism employed here. Rather, it represents a constraint on the UV completion. Crucially, such relation is key to the viability of these models, since without it the low energy theory does not reduce to the SM. In the absence of a robust explanation of (4.2.4) this solution of the Strong CP problem is severely fine-tuned, and hence not convincing. In other words, a new *naturalness* problem arises in these models. If not addressed, this is nothing else than a rephrasing of the original Strong CP problem.

4.2.2 Irreducible Contributions to $\bar{\theta}$

Provided (4.2.19) are satisfied, the effective field theory below M reproduces the SM up to irrelevant operators suppressed by inverse powers of M. The measured SM Yukawa couplings, including all radiative corrections, are given by

$$\begin{aligned} Y_u^{\text{SM}} &= F_u Y_u \\ Y_d^{\text{SM}} &= F_d Y_d, \end{aligned} \tag{4.2.20}$$

where $F_{u,d} = 1 + \mathcal{O}(YY^\dagger, Y_{u,d}Y_{u,d}^\dagger)$ are 3 by 3 matrices functions of Y_u, Y_d, Y, M. The structure shown in (4.2.20) may be understood taking advantage of the spurionic flavour charges of the SM Yukawas, for instance in the field basis in (4.2.8). That is, we interpret Y, Y_u, Y_d as fields transforming under fictitious (spurious) flavour symmetries that leave (4.2.8) invariant:

$$\begin{aligned} Y_u \to& U_q^* Y_u {U_u}^\dagger \\ Y_d \to& U_q^* Y_d {U_d}^\dagger \\ Y \to& U_q^* Y, \end{aligned} \tag{4.2.21}$$

where $U_{q,u,d}$ are $SU(3)$ matrices. The Yukawas $Y_{u,d}^{\text{SM}}$ in the effective field theory must be dimensionless combinations of the couplings of our theory that transform precisely as $Y_{u,d}$. This takes us to (4.2.20).

The SM topological angle, obtained by matching (4.2.8) with the SM at the scale M, is given by the usual quantity:

$$\begin{aligned}\bar{\theta} &= \theta - \operatorname{Im}\log\det Y_u^{\rm SM}\det Y_d^{\rm SM} \\ &= \theta - \operatorname{Im}\log\det F_u\det F_d\end{aligned} \tag{4.2.22}$$

where we used that $\det Y_{u,d}$ are real, see (4.2.9) and (4.2.16). There are no tree-level contributions because the complex mass matrix of the colored fermions has real determinant (see for example (4.2.5)), which is the defining characteristic of Nelson-Barr models [178, 179]. Of course the same is true in (4.2.8), since the unitary matrix in (4.2.7) has unit determinant. In our language, this just follows from the absence of tree-level flavour-invariant, CP-odd combinations of the parameters.

We can estimate the radiative contributions to $\bar{\theta}$ using the same trick as above. Contributions to $\bar{\theta}$ must be obviously CP-odd combination of our couplings but, importantly, also invariant under spurious $SU(3)$ rotations. This is because as we have seen its expression must be written in the flavour-invariant combinations $\theta, \det F_u, \det F_d$. In addition, there can be a dependence on the electroweak scale $v \simeq 246$ GeV, but for the moment let us work at leading order in the latter and set $v = 0$.

The leading CP-odd flavour-invariant combination of our couplings is the one with fewer insertions of Yukawas. Up to a factor of order unity, and the appropriate power of the loop factor $1/16\pi^2$ needed to match the powers of couplings, such combination coincides with the value of $\bar{\theta}$ at the matching scale. With $Y = 0$ we have the SM and the first correction to $\bar{\theta}$, as we now know very well, is extremely suppressed. Potentially large corrections must involve Y. Equation (4.2.21) shows that all the invariants are products of basic invariants built out of two powers of Y and several Y_u, Y_d (see appendix 4.A for a few examples). The CP-odd flavour-invariant with the smallest number of $Y_{u,d}$ insertions is (see $I_{1,0}$ in 4.A) [8]

$$\begin{aligned}\bar{\theta}\Big|_{\rm analy} &= c_{\rm analy}\left(\frac{1}{16\pi^2}\right)^3 \operatorname{Im} Y^\dagger\left[Y_d Y_d^\dagger, Y_u Y_u^\dagger\right] Y \\ &\sim \left(\frac{1}{16\pi^2}\right)^3 \lambda_C^2 \widehat{Y}_t^2 \widehat{Y}_b^3 \widehat{Y}_s \frac{\operatorname{Im}\xi_i\xi_j^*}{m_\psi^2} \\ &\sim 6\times 10^{-18}\,\frac{\operatorname{Im}\xi_i\xi_j^*}{m_\psi^2} \qquad\qquad (d-\text{mediation}).\end{aligned} \tag{4.2.23}$$

For the numerical estimate of (4.2.23) we went in the basis in which Y_u is diagonal, where Y_d is unitary up to CKM rotations, took $\lambda_C \sim 0.23$ for the Cabibbo angle and renormalized the couplings at 1 TeV. The factor λ_C^2 arises because the result is proportional to the 23 element of the CKM. In terms of familiar Feynman diagrams, this contribution to the QCD topological angle arises from 3-loop corrections to $F_{u,d}$ as well as direct corrections to θ, with virtual fermions and the Higgs. Had we considered scenarios with unsuppressed couplings between the q's and the messengers we would have found unacceptably large 2-loop corrections to $\bar{\theta}$ [189]. Imposing $|\bar{\theta}| < 10^{-10}$ on (4.2.23) one obtains a lower bound on $|m|/|\xi|$ a bit looser than (4.2.17). All other (subleading) flavour-invariants lead to weaker constraints. Non-perturbative values of the coupling Y would imply unacceptably large corrections to $\bar{\theta}$.

Because in eq. (4.2.23) we neglected powers of v, that expression represents only the leading *non-decoupling* contribution to $\bar{\theta}$, which dominates if M is sufficiently large compared to the weak scale. When matching the UV theory to the SM at scales $\sim M$, however, one also finds additional threshold contributions that decouple as $16\pi^2 v^2/M^2 \to 0$. To estimate the leading *decoupling* effect we should allow $\bar{\theta}$ to depend on v, in which case its expression should respect the same selection rules (4.2.21) plus the additional spurious symmetry $v \to -v$, $(Y, Y_u, Y_d) \to -(Y, Y_u, Y_d)$. Now, an important complication found when estimating the decoupling contributions is that these can be non-analytic in the couplings $Y_{u,d}$. Indeed, after electroweak symmetry breaking the Yukawas may appear not only as couplings, but also as masses. On the other hand, $\bar{\theta}$ is necessarily analytic in Y because such a coupling does not control the large mass M directly, but rather a small mixing angle of order Yv/M. We thus learn that the most general expression for $\bar{\theta}$, again leading in Y, reads

$$\begin{aligned}\bar{\theta} &= \operatorname{Im} Y^\dagger f\left(Y_d Y_d^\dagger, Y_u Y_u^\dagger, \frac{v^2}{M^2}\right) Y \\ &= \operatorname{Im} Y^\dagger f_0\left(Y_d Y_d^\dagger, Y_u Y_u^\dagger, 0\right) Y + \frac{v^2}{M^2} \operatorname{Im} Y^\dagger f_1\left(Y_d Y_d^\dagger, Y_u Y_u^\dagger, \frac{v^2}{M^2}\right) Y + \mathcal{O}\left(\frac{v^4}{M^4}\right),\end{aligned} \tag{4.2.24}$$

where f is an unknown anti-symmetric 3 by 3 matrix, $f^t = -f$. The leading effect controlled by the term f_0 is precisely (4.2.23). The main one proportional to v^2/M^2 is the decoupling effect we want to estimate. As explained above, f_1 can have a residual non-analytic dependence on v^2 that we cannot Taylor expand. Because this dependence is in principle arbitrarily complicated, it is not possible to find an explicit form of f_1 based solely on symmetry arguments. We can however reliably estimate the order of magnitude. Anti-symmetry of f requires that f_1 depends on both $Y_d Y_d^\dagger, Y_u Y_u^\dagger$. Some of this dependence could be hidden in logarithms of the masses; and these are precisely the quantities that are not constrained by our selection rules. Importantly, though, f_1 should be proportional to at least one power of $Y_d Y_d^\dagger$ and one power of $Y_u Y_u^\dagger$. The reason is that if the whole dependence of f_1 on, say, the up-type Yukawa was in the unknown non-analytic terms, then $\bar{\theta}$ would be singular in the limit $Y_u Y_u^\dagger \to 0$. And this cannot be the case because such IR divergences do not appear in matching the Wilson coefficients of an effective theory. We conclude that the leading decoupling contributions scale similarly to (4.2.23) [8]

$$\begin{aligned}\bar{\theta}\Big|_{\text{nonanaly}} &\sim c_{\text{nonanaly}} \left(\frac{1}{16\pi^2}\right)^2 \frac{v^2}{M^2} \lambda_C^2 \widehat{Y}_t^2 \widehat{Y}_b^3 \widehat{Y}_s \frac{\operatorname{Im} \xi_i \xi_j^*}{m_\psi^2} \\ &\sim c_{\text{nonanaly}}\, 5 \times 10^{-17} \left(\frac{\text{TeV}}{M}\right)^2 \frac{\operatorname{Im} \xi_i \xi_j^*}{m_\psi^2} \qquad (d-\text{mediation}),\end{aligned} \tag{4.2.25}$$

up to logarithms of the masses that we cannot estimate using spurion techniques, which have been included in c_{nonanaly}. In appendix 4.B we perform a more standard loop analysis in the mass basis and confirm this result, see (4.B.15).

The powers of $1/16\pi^2$ in (4.2.25) are different from those of (4.2.23) because the decoupling effect is proportional to a mass squared (or analogously to $\propto v^2/M^2$), rather than a coupling squared. One can easily appreciate why the number of loops (and hence $1/16\pi^2$'s) in (4.2.25) is one less than (4.2.23) by re-instating the powers of $\hbar$ and observing that the Yukawa couplings scale as $\sim \hbar^{-1/2}$, whereas $v^2 \sim \hbar$, $M \sim \hbar^0$. As a result, at least at the leading order, the main qualitative difference between non-decoupling (4.2.23) and decoupling (4.2.25) contributions is the formal replacement $1/(16\pi^2) \to v^2/M^2$, which indicates that decoupling effects are parametrically less relevant than the non-decoupling ones when $M \gtrsim 4\pi v \sim 3$ TeV. More precisely, the experimental bound $\bar{\theta} \lesssim 10^{-10}$ reads $|\xi|/m_\psi \lesssim (800-900)(M/\text{TeV})$ for numbers c_{nonanaly} of order unity.

Yet, we argued above that c_{nonanaly} may contain large logs of the mass ratios (dimensional analysis is enough to show this does not occur in c_{analy}). Those that may in principle impact our estimate are the logs of the largest available mass, namely M^2. Inspecting the relevant diagrams one sees there are at most 3 powers of large logs. This is compatible with the SM computation of [54], which is one loop higher. As a conservative estimate, we may thus take $c_{\text{nonanaly}} \sim \log^3 M^2/m_b^2$ and the bound becomes $|\xi|/m_\psi \lesssim 30(M/\text{TeV})$. In subsection 4.2.4 we will see that electroweak constraints lead to bounds on the very same quantity that are comparable to this one, and obviously much more accurate theoretically than our order one estimate (4.2.25). A genuine 2-loop computation would be necessary to determine the value of c_{nonanaly} as well as whether (4.B.15) can realistically compete with the bounds of subsection 4.2.4.

In models with $u-$mediation, repeating an analysis completely analogous to the one leading to (4.2.23), we find [8]

$$\begin{aligned}\bar{\theta}\Big|_{\text{analy}} &\sim \left(\frac{1}{16\pi^2}\right)^3 \ \operatorname{Im} Y^\dagger \left[Y_d Y_d^\dagger, Y_u Y_u^\dagger\right] Y \\ &\sim \frac{\lambda_C^2 \widehat{Y}_b^2 \widehat{Y}_t^3 \widehat{Y}_c}{(16\pi^2)^3} \frac{\operatorname{Im} \xi_2 \xi_3^*}{m_\psi^2} \\ &\sim 5 \times 10^{-15} \frac{\operatorname{Im} \xi_2 \xi_3^*}{m_\psi^2} \qquad\qquad (u-\text{mediation}).\end{aligned} \tag{4.2.26}$$

As it was for models of $d-$mediation, $|\bar{\theta}| < 10^{-10}$ gives a constraint consistent with the perturbative bound, see (4.2.18), but a bit milder (recall that $|\xi_3| \sim \lambda_C^2 |\xi|$ here). Analogously to (4.2.25), decoupling corrections are of the same order as (4.2.26) up to the replacement $1/(16\pi^2) \to v^2/M^2$ and possibly large logs. The bound on $|\xi/m_\psi|$ from the electroweak T parameter analyzed in subsection 4.2.4 is stronger and more accurate.

4.2.3 More Families of Mediators

The previous results can be generalized to the case in which the mediators appear in different families with index $a, b = 1, \cdots, n_\psi$. We limit our analysis to scenarios where m_ψ has non-degenerate eigenvalues and all the mediators ψ_a, in the basis in which m_ψ is diagonal, mix with the SM fermions. This is equivalent to saying that

$$\xi_{ia}^* = y_{mia} \langle \Sigma_m \rangle \tag{4.2.27}$$

is non-vanishing for any i when m_ψ is diagonal. The special rotation that removes the mass mixing (before electroweak symmetry breaking) now is

$$\begin{pmatrix} d \\ \psi^c \end{pmatrix} \to \begin{pmatrix} A & \xi(M^\dagger)^{-1} \\ -m_\psi^{-1}\xi^\dagger A & m_\psi{}^\dagger(M^\dagger)^{-1} \end{pmatrix} \begin{pmatrix} d \\ \psi^c \end{pmatrix} \tag{4.2.28}$$

where the condition $AA^\dagger = 1-\xi(MM^\dagger)^{-1}\xi^\dagger = [1+\xi(m_\psi m_\psi{}^\dagger)^{-1}\xi^\dagger]^{-1}$ is necessary to ensure this transformation is unitary (the second equality is a consequence of the first and our definition of M, see below). After the above rotation is performed the Yukawa and mass terms look formally as in (4.2.8), where summation over indices is always understood. The masses and couplings of (4.2.8), in matrix notation, now explicitly read

$$\begin{cases} MM^\dagger = \xi^\dagger\xi + m_\psi m_\psi{}^\dagger \\ Y_u = y_u \\ Y_d = y_d A \\ Y = y_d\xi(M^\dagger)^{-1} = Y_d A^{-1}\xi(M^\dagger)^{-1}. \end{cases} \tag{4.2.29}$$

In models with $u-$mediation the very same results hold except for the replacement $d \leftrightarrow u$.

An analysis similar to the one performed in 4.2.1 says that $\mathrm{Im}\,(AA^\dagger)_{ij} \gtrsim 1$ is necessary to reproduce the experimental value of J. This of course means that $\mathrm{Im}\,\xi\xi^\dagger \gtrsim |m_\psi|^2$. Also, $|m_\psi| \ll |\xi|$ would signal a non-perturbative regime. To see this we multiply $AA^\dagger$ on the left by $\xi^\dagger$ and on the right by ξ, and use the definition of $MM^\dagger$, to obtain

$$\xi^\dagger AA^\dagger\xi = m_\psi m_\psi{}^\dagger[1-(MM^\dagger)^{-1}(m_\psi m_\psi{}^\dagger)]. \tag{4.2.30}$$

From this follows that if one eigenvalue of $m_\psi m_\psi{}^\dagger$ is much smaller than $|\xi^\dagger\xi|$ the matrix $AA^\dagger$ develops a null vector or, in other words, the rank of A becomes smaller than 3. To see this let us go in the basis in which m_ψ is diagonal and suppose that $[m_\psi]_{\hat a} = 0$ for some $a = \hat a$. Equation (4.2.30) then reads $[\xi^*]_{i\hat a}[AA^\dagger]_{ij}[\xi]_{j\hat a} = 0$, which implies $[\xi]_{j\hat a}$ is a null eigenvector because of our hypothesis (4.2.27). We may rephrase this stating that $\det m_\psi = 0$ implies $\det A = 0$. In the same limit the last equation in (4.2.29) shows that Y becomes non-perturbative. These considerations demonstrate that the coincidence $\xi \sim m_\psi$ (see (4.2.4)) must be realized even in the general case with more families of mediators. With $n_\psi > 1$ it is more appropriate to express it as [8]

$$\begin{cases} 2 \lesssim |Y_d^{-1}Y| \ll 10^3 & (d-\text{mediation}) \\ 20 \lesssim |Y_u^{-1}Y| \ll 300 & (u-\text{mediation}). \end{cases} \tag{4.2.31}$$

As in 4.2.1, $|Y_d{}^{-1}Y|$ (or $|Y_u{}^{-1}Y|$) must be at least of order one in order for J to be reproduced and simultaneously cannot be much larger than quoted otherwise the theory becomes non-perturbative. The perturbative bounds are analogous to (4.2.17) and (4.2.18). However, as opposed to what happened with a single family, we will now show that the constraint on $\bar\theta$ sets more stringent upper limits than (4.2.31).

In the case $n_\psi > 1$ the derivation of $\bar\theta$ deserves some care because the new family index allows us to build more flavour-invariants involving the mass matrix M. It is not immediately obvious that large corrections to the theta angle can be avoided. For brevity we will analyze the non-decoupling effects only; the decoupling corrections can be estimated as in the previous section. We will prove that there are no (non-decoupling) 2-loop corrections to $\bar{\theta}$ and that (non-decoupling) 3-loop contributions are under control, exactly as it was in the one-family models. The basic ingredients, along with their spurious transformations, are

$$\begin{aligned} Y_u &\to U_q^* Y_u U_u{}^\dagger \\ Y_d &\to U_q^* Y_d U_d{}^\dagger \\ Y &\to U_q^* Y U_{\psi^c}{}^\dagger \\ M &\to U_\psi^* M U_{\psi^c}{}^\dagger, \end{aligned} \tag{4.2.32}$$

where $U_{q,u,d}$ and U_{ψ,ψ^c} are $SU(3)$ and $SU(n_\psi)$ flavour matrices, respectively. As before, we are interested in corrections proportional to Y. To warm up, it is straightforward to see that there is no correction to $\bar{\theta}$ at 1-loop order. Indeed, the unique combination $\mathcal{O}(Y^2)$ that is invariant under rotations of the SM fermions is $Y^\dagger Y$. Similarly, the mass can only enter via $M^\dagger M$, which is invariant under ψ rotations. Now, the class of flavour-singlets one can build out of these two objects are just traces of $Y^\dagger Y$ and (dimensionless functions of) $M^\dagger M$. Anything of this form will be automatically real and CP-even, however, since such matrices are hermitian. Hence there cannot be 1-loop corrections to $\bar{\theta}$.

Let us then move to the 2-loop order, considering first only combinations of Y, M. For this task it is convenient to make use of some group theory. Since the building blocks are hermitian $n_\psi \times n_\psi$ matrices they can be expanded in a basis of $SU(n_\psi)$ generators T^A. Explicitly,

$$h^A T^A \equiv Y^\dagger Y - \frac{1}{n_\psi} \operatorname{tr} Y^\dagger Y, \tag{4.2.33}$$

and similarly for $M^\dagger M$. The trace parts are real and contribute to $\bar{\theta}$ at subleading order. Here we are interested in the leading contribution to the topological angle, so they can be safely neglected. With this notation, our 2-loop effects must be of the form

$$h^A h^B \; F^{AB}(M^\dagger M), \tag{4.2.34}$$

where the dimensionless function F^{AB} is symmetric and traceless. The key observation is that CP violation in the $SU(n_\psi)$ space acts as $T^A \to (T^A)^* = \eta^{AB} T^B$ on the generators and thus as $h^A \to \eta^{AB} h^B$ on the adjoints, where η^{AB} can be chosen to be diagonal and satisfying $\eta^{AC}\eta^{CB} = \delta^{AB}$. Its explicit form depends on n_ψ. For example $\eta^{AB} = \mathrm{diag}(1, -1, 1)$ in $SU(2)$. From the algebra follows that the completely symmetric tensor d^{ABC} is CP-even, whereas the structure function f^{ABC} is CP-odd. We conclude that all $SU(n_\psi)$-invariant combinations of adjoints are automatically CP-even unless the expression contains an odd number of f^{ABC}. In particular, the combination (4.2.34) cannot contain the structure functions and is therefore CP-even: there is no 2-loop effect at $\mathcal{O}(Y^4)$. The absence of (non-decoupling) 2-loop contributions involving both Y and $Y_{u,d}$ is even easier to understand. This class of flavour-invariants consists of traces of $Y_{u,d} Y_{u,d}{}^\dagger$, $Y F(M^\dagger M) Y^\dagger$, which are again real by hermiticity, and hence CP-even.

We have thus demonstrated that there are no non-decoupling 2-loop contributions to $\bar{\theta}$ in these models. The first effects arise at 3-loops. Those due to Y, M are constrained by the $SU(n_\psi)$ arguments given above. To get a non-vanishing combination at most one index of the structure function can be contracted with h^A or a function of the masses, as these are all necessarily symmetric. The unique option is

$$h^A h^B h^C F^{A'B'C'} f^{AA'A''} f^{BB'B''} f^{CC'C''} d^{A''B''C''}, \tag{4.2.35}$$

with $F^{A'B'C'}$ a dimensionless function of $M^\dagger M$. (Note that the SM invariant $\mathrm{Im}\ \det[Y_u {Y_u}^\dagger, Y_d {Y_d}^\dagger]$ is precisely of this form.) A parametric estimate gives [8]

$$\bar{\theta}\Big|_{n_\psi \geq 3} \sim \begin{cases} \left(\frac{1}{16\pi^2}\right)^3 \widehat{Y}_b^4 \widehat{Y}_s^2 \left(|{Y_d}^{-1} Y|\right)^6 \sim 10^{-21} \left(|{Y_d}^{-1} Y|\right)^6 & (d-\text{mediation}) \\ \left(\frac{1}{16\pi^2}\right)^3 \widehat{Y}_t^4 \widehat{Y}_c^2 \left(|{Y_u}^{-1} Y|\right)^6 \sim 10^{-12} \left(|{Y_u}^{-1} Y|\right)^6 & (u-\text{mediation}), \end{cases} \tag{4.2.36}$$

where we assumed all numerical factors are of order unity apart from the usual powers of $1/4\pi$. The bound $\bar{\theta} \lesssim 10^{-10}$ translates into much more stringent constraints than quoted in (4.2.31), because of the large powers of Y involved. Importantly, though, this 3-loop contribution does *not* exist if $n_\psi \leq 2$, for the very same reason the SM with less than 3 generations has no Jarlskog invariant: the totally symmetric tensor vanishes and (4.2.36) cannot be built in those cases. More model-independent contributions, which exist also for $n_\psi = 2$, must involve the SM Yukawas. The larger ones are proportional to the up-type Yukawa. The key building blocks are $Y^\dagger Y = h^A T^A + \text{trace}$, $Y^\dagger Y_u {Y_u}^\dagger Y = h_u^A T^A + \text{trace}$, and $F(M^\dagger M) = F^A T^A + \text{trace}$. There is a unique way the indices of the CP-odd function f^{ABC} can be contracted: $f^{ABC} h^A h_u^B F^C(M^\dagger M)$. Including an appropriate number of the loop $1/16\pi^2$ factors, the latter CP-odd invariant can equivalently be written as [8]

$$\begin{aligned} \bar{\theta}\Big|_{n_\psi \geq 2} &\sim \left(\frac{1}{16\pi^2}\right)^3 \mathrm{Im}\ \mathrm{tr}\left[Y^\dagger Y_u {Y_u}^\dagger Y, Y^\dagger Y\right] F(M^\dagger M) \\ &\sim \begin{cases} \left(\frac{1}{16\pi^2}\right)^3 \lambda_C^2 \widehat{Y}_t^2 \widehat{Y}_b^3 \widehat{Y}_s \left(|{Y_d}^{-1} Y|\right)^4 \sim 6 \times 10^{-18} \left(|{Y_d}^{-1} Y|\right)^4 & (d-\text{mediation}) \\ \left(\frac{1}{16\pi^2}\right)^3 \widehat{Y}_t^4 \widehat{Y}_c^2 \left(|{Y_u}^{-1} Y|\right)^4 \sim 10^{-12} \left(|{Y_u}^{-1} Y|\right)^4 & (u-\text{mediation}). \end{cases} \end{aligned} \tag{4.2.37}$$

The numerical bound on $|{Y_d}^{-1} Y|$ (and $|{Y_u}^{-1} Y|$) is a bit stronger than in the case of a single family of ψ, ψ^c, see (4.2.23) and (4.2.26), again because of the larger power of the new Yukawa.

4.2.4 Experimental Signatures

The mediators ψ, ψ^c are constrained by direct and indirect collider searches. In this section we study both $d-$ and $u-$mediation and provide a qualitative assessment of the most stringent bounds for scenarios with one and two families of mediators, including those arising from radiative effects controlled by Y. We will see the latter are actually very relevant phenomenologically.

As shown in (4.2.9) for a single family and (4.2.29) for more families, Y can be written as the $d-$type (or $u-$type) Yukawa multiplied on the right by a flavour-violating matrix. The latter is simply $[Y_d{}^{-1}Y]_i = \xi_i/m_\psi$ (or $[Y_u{}^{-1}Y]_i = \xi_i/m_\psi$) if a single family is considered, or more generally $[Y_d{}^{-1}Y]_{ia}$ (or $[Y_u{}^{-1}Y]_{ia}$). In models of $d-$mediation we have argued that $Y_d^{-1}Y$ must have entries of comparable size in order to reproduce the CKM. Therefore, for simplicity we treat it as a single coupling

$$[Y_d{}^{-1}Y]_{ia} = |Y_d{}^{-1}Y| \qquad i = 1,2,3 \tag{4.2.38}$$

when quantifying the numerical bounds below. In other words, we will assume the only hierarchies involved in our calculations are those due to the quark masses and powers of the Cabibbo angle, and instead ignore possible cancellations in the sum of different $[Y_d{}^{-1}Y]_{ia}$'s.

The results of 4.2.1 reveal that in scenarios of $u-$mediation the CKM is reproduced provided $[Y_u{}^{-1}Y]_{ia}$ has a $i=3$ component suppressed by $\sim \lambda_C^2$ compared to the others. Our analysis will therefore be performed assuming that

$$[Y_u{}^{-1}Y]_{ia} = |Y_u{}^{-1}Y| \times \begin{cases} 1 & i=1,2 \\ \lambda_C^2 & i=3 \end{cases} \tag{4.2.39}$$

where $|Y_u{}^{-1}Y|$ is the parameter we will constrain, similarly to $d-$mediation. A more rigorous study of the phenomenology would require a numerical scan, but this is beyond the scope of our qualitative analysis.

Finally, when estimating the bounds on scenarios with two families for simplicity we take

$$M_1 = M_2 = M. \tag{4.2.40}$$

We will come back to the implications of this simplifying assumption below.

$d-$Mediation

In models of $d-$mediation the ψ, ψ^c behave similarly to a heavy b-quark: they are pair-produced and decay promptly into quarks and vector bosons or the Higgs boson. Current direct searches imply $m_\psi \gtrsim 1400$ GeV for a single family [193].

The relevant couplings of the ψ, ψ^c to the SM can be read off directly from (4.2.8). With the flavour indices shown explicitly, including those of ψ, ψ^c (a,b), this is $\mathcal{L}^d_{\text{Yuk}} \supset -Y_{ia} q_i \widetilde{H} \psi^c_a - M_{ab}\psi_a\psi^c_b$. Integrating out the heavy fermion at tree-level, below the scale M_a we find (after a field re-definition of the quark doublet) a correction to the SM Lagrangian [8]:

$$\begin{aligned} \delta\mathcal{L}_{\text{SM}}^{(\text{tree})} &= \frac{1}{v^2}\left[\bar{c}_{ik}(Y_d)_{kj} q_i \widetilde{H} d_j |H|^2 + \text{h.c.}\right] \\ &- \frac{1}{2v^2}\bar{c}_{ji}\left[q_i^\dagger \bar{\sigma}^\mu q_j \; H^\dagger i\overleftrightarrow{D_\mu} H + q_i^\dagger \sigma^a \bar{\sigma}^\mu q_j \; H^\dagger \sigma^a i\overleftrightarrow{D_\mu} H\right] \\ \bar{c}_{ij} &= \frac{v^2}{2}\left(Y\frac{1}{M^\dagger M}Y^\dagger\right)_{ij}. \end{aligned} \tag{4.2.41}$$

An equivalent description of the following effects can be given in terms of the (flavour-violating) couplings in the mass basis (see appendix 4.B.1).

The second line of (4.2.41) gives rise to (flavour-violating) corrections to the Z-couplings of the down left-handed quarks. In the mass basis Y reduces to $\widehat{Y}_d$ multiplied by an anarchic matrix on the right. The constraint on the flavour-diagonal components of $\bar{c}$ is therefore dominated by that on the bottom. In the mass basis it reads

$$|\bar{c}_{33}| \lesssim 0.008, \tag{4.2.42}$$

(see e.g. [194] and note that in our model $\bar{c}_{ii} < 0$). Other precision electroweak measurements are very weakly affected because of the small couplings involved.

All couplings in (4.2.41) contribute to $\Delta F = 1$ and $\Delta F = 2$ transitions. Because of the strongly hierarchical structure inherited from the SM Yukawas, however, flavour-violating observables are far less crucial than in most scenarios of physics beyond the SM. The most constrained $\Delta F = 2$ operators involve the left-handed down sector:

$$\mathcal{L}_{\rm EFT} \supset -C^d_{ij;kl}\ ({d_L}^\dagger)_i \bar{\sigma}^\mu (d_L)_j\ ({d_L}^\dagger)_k \bar{\sigma}_\mu (d_L)_l. \tag{4.2.43}$$

The coefficient is corrected at tree-level $\delta C^d_{ij;kl} = \bar{c}_{ji}\bar{c}_{lk}/(2v^2)$, from the second line of (4.2.41) via a Z exchange, and also receives a 1-loop contribution from the exchange of the electroweak gauge bosons, of order $\delta C^d \sim \bar{c} g^2 \lambda_C^2/(16\pi^2 v^2)$. The main bound on (4.2.43) in this model is currently due to $B_s - \overline{B}_s$ mixing, and conservatively reads $|C^d_{32;32}| \lesssim 6.7 \times 10^{-12}$ GeV^{-2} (in the mass basis and when renormalized at $M \sim 1$ TeV), see e.g. [195]. The resulting bound on $|Y_d^{-1} Y|/M$ is weaker than (4.2.42). Operators in the $\Delta F = 2$ class are also induced by the first line of (4.2.41); however their coefficients are down by larger factors of the SM Yukawas. More importantly, at 1-loop the effective field theory below M features additional $\Delta F = 2$ interactions of the form $\mathcal{L}_{\rm EFT} \supset -C_{ij;kl}\ {q_i}^\dagger \bar{\sigma}^\mu q_j\ {q_k}^\dagger \bar{\sigma}_\mu q_l$, with coefficients (again in the mass basis in which $M^\dagger M$ is diagonal) [8]

$$C^{(\text{1-loop})}_{ij;kl} = \frac{1}{8}\frac{1}{(4\pi)^2} Y^*_{ia} Y_{ja} Y^*_{kb} Y_{lb} \frac{\log M_b^2/M_a^2}{M_b^2 - M_a^2}. \tag{4.2.44}$$

The dominant constraints on these come again from $B_s - \overline{B}_s$ mixing and directly compete with (4.2.43). Importantly, (4.2.44) has a different parametric dependence than the corrections δC^d mentioned above, i.e. it is controlled by $|Y_d^{-1} Y|/\sqrt{M}$ rather than $|Y_d^{-1} Y|/M$, and starts to dominate for masses $M \gtrsim 4\pi v \sim 3$ TeV. It even becomes more important than (4.2.42) at around $M \gtrsim 18$ TeV.

Let us next move to $\Delta F = 1$ observables. Among the most relevant operators are

$$\mathcal{L}_{\rm EFT} \supset (C_9)_{ij} (\overline{d}_L)_i \gamma^\mu (d_L)_j\ \bar{\ell}\gamma_\mu \ell + (C_{10})_{ij} (\overline{d}_L)_i \gamma^\mu (d_L)_j\ \bar{\ell}\gamma_\mu\gamma^5 \ell, \tag{4.2.45}$$

with ℓ any of the charged leptons and $(C_9)_{ij} = +\bar{c}_{ji}(1 - 4\sin^2\theta_w)/(2v^2)$, $(C_{10})_{ij} = -\bar{c}_{ji}/(2v^2)$. These follow from the second line of (4.2.41), integrating out the Z, and contribute to rare K, B meson decays, most notably $B \to X_s \ell\bar{\ell}$, $B_s \to \ell\bar{\ell}$, and ϵ'/ϵ. Conservatively requiring the new physics contribution lies within one sigma from the SM prediction (see, e.g., [196, 197]) we obtain bounds that are somewhat comparable numerically and a bit stronger than those derived from $B_s - \overline{B}_s$ mixing. However, they have the very same parametric dependence on

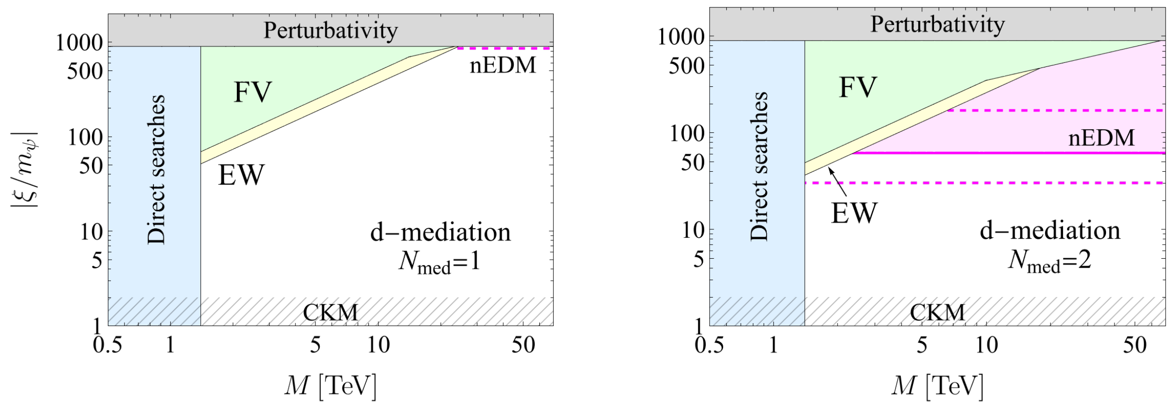

Figure 4.3: Summary of the most relevant experimental constraints for models of $d-$mediation with one (left) or two (right) families of ψ, ψ^c. Direct searches are shown in light blue, electroweak bounds in yellow ("EW", see (4.2.42)), constraints from flavour violation in light green ("FV", in particular those due to the $\Delta F = 1$ transitions discussed below (4.2.45) are visible at lower mass and the $B_s - \overline{B}_s$ mixing constraint imposed on $C^{(1-\text{loop})}$ dominates at larger M), and the bound from the neutron electric dipole moment in magenta ("nEDM") with a rough estimate of the error band (dashed lines). As seen in 4.2.1, $|Y_d{}^{-1}Y|$ is also subject to a lower bound (see hatching), required in order for these models to be able to reproduce the CKM phase, and an upper bound, from perturbativity (see the grey region). See the text for details. Plots taken from [8].

$|Y_d^{-1}Y|/M$ as (4.2.42), and all turn out to be weaker than the latter. Loop-induced $\Delta F = 1$ observables, say $B \to X_s \gamma$, lead to subleading bounds.

Finally we should not forget the constraints coming from the non-observation of the neutron electric dipole moment d_n. We estimated the dominant non-decoupling contributions to the $\bar{\theta}$ parameter for one family of mediators in (4.2.23) and for two families in (4.2.37). Because we did not compute the order one number in front of those expressions we allow for an unknown overall factor ranging within $[0.1, 10]$. For one family only the most conservative bound, $|Y_d{}^{-1}Y| \lesssim 860$, wins over the perturbative constraint (4.2.17). On the other hand, for two families of mediators $\bar{\theta} \lesssim 10^{-10}$ roughly translates into $|Y_d{}^{-1}Y| \leq 60^{+100}_{-30}$, which is always stronger than (4.2.17). Other contributions to d_n are induced by higher-dimensional operators. However, these give small corrections [189]. For example, quark dipole interactions first arise at 2-loops and are suppressed by the small light quark masses.

As anticipated at the beginning, we collect all the constraints in a single plot assuming that $Y_d{}^{-1}Y$ has anarchic entries of comparable size $|Y_d{}^{-1}Y|$. When plotting the bounds on the 2-family model we also take degenerate masses as in (4.2.40). The mass degeneracy effectively increases the Wilson coefficients $\bar{c}_{ij}$ by a factor of 2, thus making the electroweak and flavour constraints stronger than in the single family model. The results are shown in figure 4.3 for one family of mediators (left) and two families (right). We see that direct searches as well as electroweak precision tests and flavour data have already started to constrain our scenarios, though a sizeable portion of accessible parameter space around $M \gtrsim 1$ TeV is still available. Remarkably, in the 2-family model the most significant constraint on the coupling constant comes from the 3-loop contribution to the neutron electric dipole moment. To make a more quantitative assessment of the allowed parameter space it would therefore be necessary to calculate the numerical coefficient in front of (4.2.37).

$u-$MEDIATION

In models of $u-$mediation collider searches currently imply $m_\psi \gtrsim 1200$ GeV [198]. In this regime it is appropriate to describe the electroweak and flavour constraints in terms of an effective field theory, as done for $d-$mediation.

The strongest constraint from electroweak observables arises due to radiative corrections to the $\widehat{T}$ parameter. In the effective field theory this corresponds to the dimensionless coefficient of $|H^\dagger D_\mu H|^2/v^2$. The main contributions are of order $Y^2\widehat{Y}_t^2$ and Y^4. Considering an arbitrary number of mediators' families, in the basis in which M is diagonal we obtain [8]:

$$\begin{aligned}\widehat{T} = & \frac{3}{16\pi^2} Y_{3a}^* Y_{3a} \frac{m_t^2}{M_a^2}\left(\log\frac{M_a^2}{m_t^2} - 1\right) \\ & + \frac{3}{64\pi^2}\frac{v^2}{M^2} Y_{ia}^* Y_{ja} Y_{jb}^* Y_{ib} \frac{\log M_b^2/M_a^2}{M_b^2 - M_a^2}.\end{aligned} \tag{4.2.46}$$

Since S is very small we find that the 95% CL bound of [199] simply reduces to $\Delta T = 4\pi\widehat{T}/e^2 < 0.1$. Other electroweak observables lead to weaker bounds. In particular, corrections to the couplings of the up-type quarks to Z, most importantly those of the charm, are very small.

The dominant $\Delta F = 2$ effects show up in the radiative K^0, B_d, B_s meson oscillations. The associated operator is again the one in (4.2.43) with the Wilson coefficient (4.2.44), where of course now $Y \propto Y_u$. Using respectively $|C^d_{21;21}| \lesssim 2.0 \times 10^{-15}$ GeV^{-2}, $|C^d_{32;32}| \lesssim 6.7 \times 10^{-12}$ GeV^{-2}, $|C^d_{31;31}| \lesssim 8.0 \times 10^{-13}$ GeV^{-2} [195] we find comparable bounds on $|Y_u{}^{-1}Y|/\sqrt{M}$, though $B_s - \overline{B}_s$ mixing slightly wins. This constraint is comparable to the one coming from (4.2.46) at large M. Among the $\Delta F = 1$ observables, by far the most constraining turns out to be $B \to X_s\gamma$. In the effective field theory the associated operator is first generated at 1-loop [8, 200]

$$\mathcal{L}_{\text{EFT}} \supset -\frac{1}{9}\frac{e}{16\pi^2}\left(Y\frac{1}{M^\dagger M}Y^\dagger\right)_{23} m_b\ \overline{s_L}\sigma^{\mu\nu}b_R\ F_{\mu\nu}. \tag{4.2.47}$$

The bound on $|Y_u^{-1}Y|/M$ derived from [200] is weaker than the one due to the electroweak T parameter as well as $B_s - \overline{B}_s$ mixing (already at $M \gtrsim 100$ GeV).

Similarly to models of $d-$mediation, the constraints due to the neutron electric dipole are dominated by (4.2.26) for one generations of mediators and by (4.2.37) for two generations. We find that in the former case only the most conservative bound on $\bar{\theta}$ is more stringent than the perturbativity requirement (4.2.18), whereas for two generations it translates, under our hypothesis (4.2.39), into $|Y_u{}^{-1}Y| \leq 13^{+22}_{-7}$.

All bounds are collected in figure 4.4. The main conclusions are similar to those drawn for theories of $d-$mediation. Yet, the parameter space of models of $u-$mediation is significantly

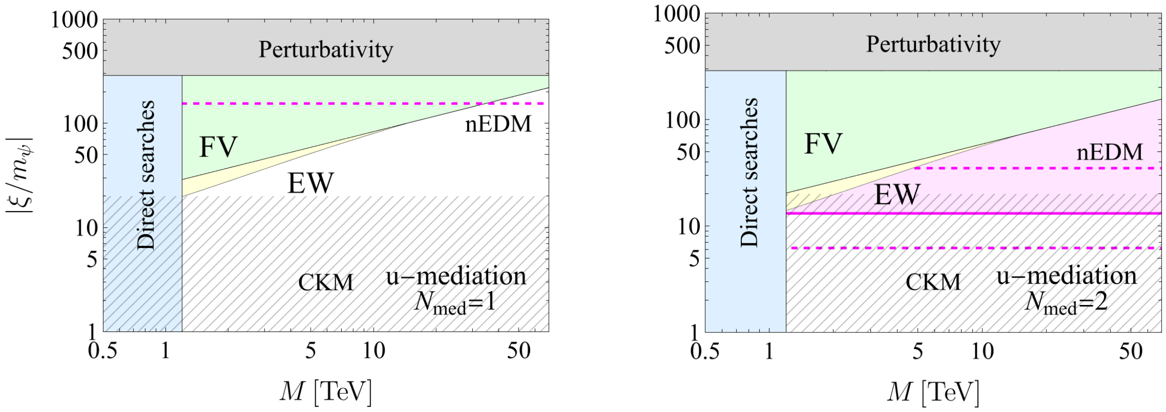

Figure 4.4: Summary of the most relevant experimental constraints on models of $u-$mediation with one (left) or two (right) families of ψ, ψ^c. To reproduce the CKM we imposed the structure (4.2.39). We show direct searches in light blue, electroweak in yellow ("EW", see (4.2.46)), flavour violation in light green ("FV", from $B_s - \overline{B}_s$ mixing), the neutron electric dipole moment ("nEDM") in magenta. The lower and upper bounds from 4.2.1 are identified by the hatching and the grey region. See the text for details. Plots taken from [8].

reduced by the lower bound $|Y_u{}^{-1}Y| \gtrsim 20$ necessary to reproduce the CKM phase, see figure 4.2. In particular, $u-$mediation with two or more families appears to be basically excluded; a thorough numerical scan and an explicit computation of (4.2.37) should tell us if some region of the parameter space is still allowed. Another, far less concrete, reason why $u-$mediation is less attractive is perhaps that $d-$mediation can be easily embedded into a grand-unified picture, where in terms of $SU(5)$ representations would be more appropriately called $\overline{\mathbf{5}}$-mediation. On the other hand, $u-$type quarks come in the same multiplet as the doublets q, which we saw should not mix with the mediators otherwise $\bar{\theta}$ gets too large, and this generates a tension between models of **10**-mediation and the stringent constraint coming from the neutron EDM non-observation.

4.3 Addressing the Nelson-Barr Hierarchy Problems

In Nelson-Barr models, the smallness of the neutron electric dipole moment strongly constraints the couplings of the mediators to the SM. Radiative corrections to $\bar{\theta}$ are below the current bounds if the couplings of the $d-(u-)$ quarks to the CP-violating sector are sufficiently small (4.2.3),

$$y \ll 1, \tag{4.3.1}$$

and CP-odd couplings between the messengers and the quark doublets are forbidden. From an effective field theory perspective, this non-generic structure can be interpreted as arising from selection rules associated to a spurious $U(1)_{\text{A}}$ symmetry. Below the scale of spontaneous CP breaking the scalars decouple and the relevant BSM degrees of freedom are just the messengers

ψ, ψ^c. Therefore, the effective spurious $U(1)_{\mathrm{A}}$ can be seen as involving the messengers, their CP-even mass m_ψ and ξ as in table 4.1.

	$U(1)_A$
ψ	z_ψ
ψ^c	z_{ψ^c}
m_ψ	$-z_\psi - z_{\psi^c}$
ξ, Σ	$-z_\psi$

Table 4.1: Spurious $U(1)_{\mathrm{A}}$ charges in the minimal Nelson-Barr model.

However, these are not the only sources of contamination to $\bar{\theta}$. The spontaneous breaking of CP itself may introduce a sensitivity to unknown physics at high scales that can potentially jeopardize the solution to the Strong CP problem. To avoid this the breaking must be sufficiently *soft*. That is, it is necessary to make sure that the CP-odd scalars Σ responsible for the spontaneous breaking of CP have vacuum expectation values much smaller compared to the UV cutoff:

$$|\langle\Sigma\rangle| \ll f_{\mathrm{UV}}. \tag{4.3.2}$$

Very explicitly, this is necessary to prevent uncontrollable higher-dimensional operators suppressed by the cutoff f_{UV}, e.g. $c_{ab}\Sigma_a^\dagger\Sigma_b\epsilon^{\mu\nu\alpha\beta}G_{\mu\nu}G_{\alpha\beta}g_s^2/(16\pi^2 f_{\mathrm{UV}}^2)$, to spoil the solution.

Finally, a crucial condition was derived in 4.2.1 that these models must strictly satisfy. The SM amount of CP violation and quark masses can be reproduced in a controllable setup only if a peculiar coincidence is present (4.2.19) [8]:

$$\begin{aligned} 2 \lesssim \left|\frac{\xi}{m_\psi}\right| \ll 10^3 \qquad & (d-\text{mediation}), \\ 20 \lesssim \left|\frac{\xi}{m_\psi}\right| \ll 300 \qquad & (u-\text{mediation}). \end{aligned} \tag{4.3.3}$$

Thus, we are apparently left with an effective theory that introduces many new additional fine-tunings instead of solving the one for which it was put forth. First, the existence of the hierarchy (4.3.2) has to be explained, otherwise no solution of the Strong CP problem is offered. Indeed, if no justification of the hierarchy is provided the QCD theta angle would be incalculable, because susceptible of corrections from unknown corrections from the UV, and its smallness would merely be the consequence of a hidden assumption (even though, fortunately, the requirement (4.3.2) may easily be addressed by Supersymmetry [201] or a strong dynamics [189]). Similarly, the coincidence of scales (4.3.3) cannot be explained by the effective field theory of table 4.1. It must be the consequence of some property of the UV completion. This coincidence is especially remarkable because ξ is CP-odd whereas m_ψ must be CP-even. How is it possible that quantities with an a priori qualitatively different UV origin, like the CP-even m_ψ, y and the vacuum expectation value of a separate sector, turn out to be comparable to each other in size? "Explaining" (4.2.4) is tantamount to finding a class of UV completions for the effective field theory described above that naturally accommodates it. This is as key as (4.3.2) if one wishes to truly solve the Strong CP problem via spontaneous CP violation. In the case no such UV completions can be found one would have to conclude that Nelson-Barr scenarios are just a very elaborated way to trade the smallness of $\bar{\theta}$ with a subtle fine-tuning of the parameters.

From a model-building viewpoint, the real challenge in constructing realistic Nelson-Barr scenarios is therefore making sure that (4.3.3) is realized within a framework that also explains (4.3.2) and (4.3.1). One might naively argue that (4.3.3) would be easily accommodated in a theory where all couplings are of order unity and all scales beyond the SM of comparable size. A more careful look, however, reveals that this cannot be the case. In non-SUSY versions additional corrections to $\bar{\theta}$ always arise from loops involving excitations of the CP-violating sector and would be unacceptably large for generic couplings of order unity, see (4.3.1). But, if y is taken to be small, why is the potential of the CP-violating sector minimized at a scale $\langle\Sigma\rangle \sim m_\psi/y$ that "knows" about the couplings of the mediators to the SM? In SUSY versions of Nelson-Barr one can avoid the corrections controlled by y if CP violation takes place at scales much larger than SUSY-breaking [201]. Yet, then there are necessarily several scales involved and the question remains: why would (4.2.4) be satisfied? One may alternatively justify (4.2.4) postulating that m_ψ itself arises from the vacuum expectation value of a new scalar field, $m_\psi = y_S \langle S\rangle$. This would be an interesting approach in both non-SUSY as well as SUSY realizations. Now, granting the reasonable assumption $y \sim y_S$, the coincidence would be explained by making sure that both scalars acquire comparable vacuum expectation values, i.e. for $\langle S\rangle \sim \langle\Sigma\rangle$. The conceptual hurdle then is achieving this in such a way that m_ψ be a CP-even parameter, which is mandatory if the strong CP is to be solved. Is there a way to guarantee this? In the rest of this section we will show that it is indeed possible to find elegant UV completions of the Nelson-Barr scenarios that simultaneously address the key requirements (4.3.2), (4.3.1), and (4.2.4) by presenting the model of the original work [9]. In particular, we will show how these requirements are naturally explained in scenarios in which CP violation is dynamical and responsible for generating both ξ and m_ψ, the messengers are chiral (that is they have $z_{\psi^c} \neq -z_\psi$ in table 4.1), and the global $U(1)_{\rm A}$ is gauged. At low energy the UV completion reproduces the scenarios of $d-$mediation studied in section 4.2. This guarantees the Strong CP problem is robustly solved. In addition, though, our constructions add new constraints and phenomenological signatures which we discuss in 4.3.2. Overall, the main message is that realistic scenarios of spontaneous CP violation are very predictive and compelling solutions of the Strong CP problem.

4.3.1 The Basic Setup

We begin our analysis with a preliminary discussion of the key ingredients. In the following we will assume that CP is an exact symmetry in the UV. This means that there exists a field basis in which all couplings are real and the topological angles vanish. CP is then spontaneously broken in the effective field theory, as described below. We will focus on scenarios with $d-$mediation.

As anticipated, we tackle the hierarchy problem (4.3.2) within non-Supersymmetric models. This is achieved replacing the vacuum expectation value of the fundamental scalars Σ by the condensate of a set of SM-neutral fermion bilinears

$$\langle\chi_\alpha\chi^c_\beta\rangle \sim 4\pi f^3\delta_{\alpha\beta}. \tag{4.3.4}$$

Here $\chi_\alpha, \chi^c_\alpha$ are two or more families of fermions charged under, say, the fundamental and anti-fundamental of an exotic strong $SU(n)$[6] (α, β are the flavour indices, and to save typing we will often omit them). The latter dynamics undergoes dynamical chiral symmetry breaking at a mass scale $\sim 4\pi f$, roughly the equivalent of $\Lambda_{\rm C} \sim 1$ GeV in real-world QCD. The powers of 4π in (4.3.4) are borrowed from naive-dimensional analysis arguments analogous to those adopted in QCD.

In the chiral limit the condensate does not violate CP. To see this observe that the results of [22] imply that in our scenarios the vector-like symmetries as well as parity remain unbroken, and hence $\langle\chi_\alpha\chi^c_\beta\rangle = C\delta_{\alpha\beta}$ with $C^\dagger = C$ a real number by P invariance. Because CP acts as $C \to C^*$ it follows that the condensate is also CP invariant. The situation is completely different when the chiral symmetry is explicitly broken by tiny effects, such as higher-dimensional operators like $(\chi\chi^c)^2$, since in that case some of the approximate Nambu-Goldstone bosons emerging from chiral symmetry breaking may acquire a vacuum expectation value $\sim f$ and break CP (and/or P) spontaneously. We will later show that in our models $\langle\chi\chi^c\rangle$ generically has large imaginary entries, with magnitude (4.3.4), even if all the parameters of the Lagrangian are CP-even by hypothesis. This elegant mechanism of spontaneous CP violation of course requires at least two families of χ, χ^c, since there would be no Nambu-Goldstones otherwise. The identification $\Sigma \to \chi\chi^c$ obviously implies that the Yukawa couplings y of (4.3.1) should be replaced by non-renormalizable interactions:

$$y\psi\Sigma d \to \frac{\psi d\chi\chi^c}{f^2_{\rm UV}}, \tag{4.3.5}$$

for some high UV cutoff scale $f_{\rm UV}$.

The above basic setup accomplishes two goals at once. First, it ensures that the hierarchy $f \ll f_{\rm UV}$ is naturally explained via dimensional transmutation. In other words, our constraint (4.3.2) is satisfied. If we are careful enough, this means we do not have to worry about possible UV effects spoiling our solution of the Strong CP problem. Second, in a picture where (4.3.5) controls the main interaction between the SM quarks and the CP-violating sector, the non-renormalizable nature of (4.3.5) automatically guarantees that the excitations of the CP-violating sector, the hadrons of the χ, χ^c dynamics, have very tiny couplings of order

$$y \sim 4\pi\frac{f^2}{f^2_{\rm UV}} \ll 1 \tag{4.3.6}$$

with d and ψ. Hence, potentially sizeable loop corrections to the theta parameter due to the CP-violating sector are completely negligible. We have thus automatically satisfied (4.3.1) as well. So far, so good. The first serious model-building challenge is, as anticipated, making sure that the coincidence (4.2.4) is explained. To achieve this, we look for a model in which

[6] Later on we will choose $n = 3$, so here we do not keep track of the large n counting.

the mass of ψ, ψ^c is given by a new coupling $y_S \sim y$ times the vacuum expectation value of a (composite) scalar $\langle S\rangle$, of the same order as $\langle \Sigma \rangle$.

Since in the present framework y arises from a non-renormalizable operator, also the mass of ψ, ψ^c has to arise from a dimension-6 operator similar to (4.3.5). We therefore introduce another composite scalar made up of new SM-neutral fermions with $SU(n)$ charges, i.e. $S \to \lambda\lambda$, and look for models in which the explicit mass term of Table 4.1 is UV-completed by

$$m_\psi \psi\psi^c \to \frac{\psi\psi^c\lambda\lambda}{f_{\rm UV}^2}. \tag{4.3.7}$$

As long as the UV dynamics is sufficiently generic (but of course CP-invariant), the coefficients in front of (4.3.5) and (4.3.7) should be comparable in size, implying $y \sim y_S$ as needed. In addition, $|\langle\lambda\lambda\rangle| \sim |\langle\chi\chi^c\rangle|$ would be generically satisfied because the λ's are charged under the very same confining $SU(n)$ carried by ψ, ψ^c. With these assumptions we get

$$m_\psi \sim \frac{\langle\lambda\lambda\rangle}{f_{\rm UV}^2} \sim \frac{|\langle\chi\chi^c\rangle|}{f_{\rm UV}^2} \sim |y\langle\Sigma\rangle|. \tag{4.3.8}$$

This is the desired result (4.3.3). We are making progress, but the main hurdle comes next: why does $\arg m_\psi$ vanish, that is why is $\langle\lambda\lambda\rangle$ real while $\langle\chi\chi^c\rangle$ is complex?

Since ψ, ψ^c are colored, an hypothetical phase in their mass would immediately translate into a correction to $\bar\theta$. If no further assumption is made, $\langle\lambda\lambda\rangle$ is expected to carry its own broken chiral symmetries and be complex, as we argued for $\langle\chi\chi^c\rangle$. Fortunately this can be avoided. We identified four sufficient conditions that guarantee that the complex phases in $\langle\lambda\lambda\rangle$ are sufficiently small to be compatible with $|\bar\theta| \lesssim 10^{-10}$. These are [9]:

(a) λ must appear in a single family;

(b) there must be a gauge $U(1)$ under which the $SU(n)$ sector is chiral;

(c) the gauge $U(1)$ must commute with the non-abelian flavour symmetry of the χ, χ^c's;

(d) the scale of spontaneous CP violation has to satisfy

$$\frac{f}{f_{\rm UV}} \lesssim 10^{-5}. \tag{4.3.9}$$

We will see later how these conditions can be implemented in concrete models. Here we explain why $\langle\lambda\lambda\rangle$ is approximately CP-even when they are satisfied.

To start, (a) implies λ does not carry non-abelian global symmetries which would otherwise generically imply large complex phases arise from the vacuum expectation value of the associated Nambu-Goldstone fields. The unique spontaneously broken global symmetry λ is allowed to carry is an *axial abelian* one, if present. In fact, at the beginning we argued that such a symmetry must be there in order to reproduce the desired structure. It should not be a surprise to find then that the global charge carried by λ must be precisely the $U(1)_{\rm A}$ of table 4.1. To see this assume that such a $U(1)_{\rm A}$ exists. It follows that the new strong sector has to be charged under it if we want to allow (4.3.5). In particular, $\chi\chi^c$ has to be chiral under the $U(1)_{\rm A}$. From this observation one might directly conclude that λ must also be charged. Indeed an accurate $U(1)_{\rm A}$ can only survive in the IR if such a symmetry has no $U(1)_{\rm A} \times SU(n) \times SU(n)$ anomaly. This of course requires the fermion λ to be charged. There is an alternative argument that forces λ to be chiral under the very same $U(1)_{\rm A}$. Because m_ψ is to be given by (4.3.7) we better make sure that ψ, ψ^c are themselves chiral under $U(1)_{\rm A}$, otherwise a Dirac mass term for ψ, ψ^c would be allowed and we would not be able to convincingly explain (4.3.3). The bilinear $\lambda\lambda$ has thus to be charged as well. However we want to put it, the necessary low-energy structure of these models combined with (4.3.5) and (4.3.7) lead us to conclude that the $SU(n)$ sector must be chiral under the global $U(1)_{\rm A}$.

Under the hypothesis (a) we see that the only Nambu-Goldstone mode that can potentially induce a sizable complex phase in $\langle\lambda\lambda\rangle$ is the one associated to the $U(1)_{\rm A}$. However, condition (b) renders the latter unphysical: the phase in $\langle\lambda\lambda\rangle$ due to the vacuum expectation value of the (would-be) $U(1)_{\rm A}$ Nambu-Goldstone boson is eaten by the abelian gauge field. Importantly, for this to fully hold (c) must be satisfied. According to (c), the gauge charges of $\chi_\alpha, \chi^c_\alpha$ must be the same for each flavour α. This ensures that the *gauge* $U(1)$ acts on the $SU(n)$ sector exactly as the *global* $U(1)_{\rm A}$. As a consequence, the longitudinal component of the gauge boson exactly coincides with the $U(1)_{\rm A}$ Nambu-Goldstone. In the unitary gauge this is removed and cannot show up in $\langle\lambda\lambda\rangle$.

Whenever (a), (b), (c) are satisfied the only Nambu-Goldstone bosons that can contribute to the phase of $\langle\lambda\lambda\rangle$ are those of the non-abelian flavour symmetry acting on $\chi_\alpha, \chi^c_\alpha$. But because λ is neutral under such a symmetry, their effect is proportional to the small chiral symmetry-breaking couplings that are responsible for triggering spontaneous CP violation, see below (4.3.4). These come from operators of at least dimension-6 in our scenarios (see next section) and therefore lead to

$$\arg m_\psi \sim \frac{f^2}{f_{\rm UV}^2}. \tag{4.3.10}$$

The experimental constraint $|\bar\theta| \lesssim 10^{-10}$ becomes an interesting upper bound (4.3.9) on the scale of CP breaking. The vacuum expectation values of the Nambu-Goldstone modes have a completely negligible impact on m_ψ if (d) is assumed.

Interestingly, (4.3.9) also guarantees that the contamination of other CP-odd phases does not spoil our solution of the Strong CP problem. In particular, one may fear uncontrollable complex contributions to $\langle\lambda\lambda\rangle$ (and, less relevantly, to $\langle\chi\chi^c\rangle$) arising from the vacuum expectation value of any of the CP-odd *massive* hadrons η of the exotic dynamics. In general the potential of the composite scalars is the sum of a zeroth order term from the renormalizable part of the λ, χ, χ^c interactions, plus a small perturbation: $V = V_0 + V_1$. In our models all perturbations are due to higher-dimensional operators of at least dimension six because of the chiral nature of the $U(1)$, and therefore $V_1/V_0 \sim f^2/f^2_{\rm UV}$. We have seen above that an hypothetical complex phase in $\langle\lambda\lambda\rangle$ must come at next to leading order, and therefore be controlled entirely by the small perturbation V_1. This leads us again to (4.3.10). The condition (d) prevents these effects from appreciably affecting the $\bar\theta$ parameter. We stress that the results of [22] are central to our conclusions. In a theory with fundamental scalars Σ, S

we would not have at our disposal such powerful theorems and it would be difficult to find general conditions guaranteeing $m_\psi = y_S \langle S \rangle$ be CP-even.

We thus have found a picture in which the key requirements (4.3.2), (4.3.1), and (4.2.4) discussed at the beginning of this section are structurally realized and the Strong CP problem can be robustly addressed. Remarkably, this basic set up has very important phenomenological consequences. Together with (4.3.8) and (4.3.4), eq. (4.3.9) implies an upper bound on the messengers' mass:

$$m_\psi \sim 4\pi f \frac{f^2}{f_{\rm UV}^2} \lesssim 10^{-14} f_{\rm UV}. \tag{4.3.11}$$

The scale at which CP violation is communicated to the SM, which is set by the messengers' mass, is therefore *super-soft* [9]:

$$m_\psi \ll 4\pi f \ll 4\pi f_{\rm UV}. \tag{4.3.12}$$

This feature has important theoretical implications. With a super-soft CP-violating scale the effect of possible additional heavy physics decouples from $\bar\theta$. That is, because of (4.3.12) the existence of new physics unrelated to the Strong CP problem, characterized by masses $m \gg m_\psi$ and sizeable couplings to the SM but not to the CP-violating sector, is not severely constrained in our scenarios. Precisely, the impact of heavy particles on $\bar\theta$ decouples as powers of $m_\psi^2/m^2 \ll 1$. New physics above the TeV may thus safely be invoked to address other puzzles in physics beyond the SM, like the origin of the SM flavour hierarchy, or to partially stabilize the weak-Planck scale hierarchy without spoiling our solution of the Strong CP problem. Said differently, there need not be a desert between $4\pi f$ and the TeV scale.

There are also interesting phenomenological consequences, however. Since our effective field theory arguments are at most reliable up to the Planck scale we expect $f_{\rm UV} \lesssim 2.4 \times 10^{18}$ GeV. It follows that [9]

$$m_\psi \lesssim \text{few 10's TeV}. \tag{4.3.13}$$

This constraint must be interpreted at an order of magnitude level, given it depends on an unknown UV scale and the coefficients of dimension-6 interactions, as well as the incalculable value of the condensates $\langle \chi_\alpha \chi_\beta^c \rangle$, $\langle \lambda\lambda \rangle$. Yet, the implication is clear: our solutions predict new colored fermions not far from the TeV. This important constraint makes these scenarios testable and very predictive[7]. We may also reverse the argument and observe that, since the messengers are colored fermions, the lack of experimental evidence of such particles says that $m_\psi \gtrsim 10^3$ GeV. Via (4.3.11) this implies $f_{\rm UV} \gtrsim 10^{17}$ GeV. Assuming order one coefficients in (4.3.8), we see the UV cutoff must lie close to the Planck scale.

The main ingredients of the model have now been identified. In the following we will construct a concrete realization and discuss some of the main phenomenological implications.

4.3.2 A Concrete Realization

An anomaly-free realization of the ideas in 4.3.1 requires more fields than the minimal ones necessary to address the Strong CP problem, namely more than just $\psi, \psi^c, \chi_\alpha, \chi_\alpha^c, \lambda$. Some of

[7] An analogous connection with the TeV was made in the context of mirror-world models in [138].

the extra states may lead to interesting phenomenological signatures, which will be analyzed later on.

A. Field Content

The particle content beyond the SM involves only fermionic (Weyl) fields and is summarized in table 4.2. The non-abelian gauge groups are all asymptotically free and the Landau poles of the abelian sector are many orders of magnitude above the Planck scale. The embedding of the fields charged under the SM in complete grand-unified $SU(5) \supset SU(3)_\mathrm{C} \times SU(2)_\mathrm{L} \times U(1)_\mathrm{Y}$ multiplets is straightforward (see the caption of table 4.2). Let us discuss the role of the various fields in turn.

	$SU(3)_\mathrm{C}$	$SU(2)_\mathrm{L}$	$U(1)_\mathrm{Y}$	$SU(3)$	$U(1)$
ψ_1	$\mathbf{3}$	$\mathbf{1}$	$-\frac{1}{3}$	$\mathbf{1}$	$+1$
ψ_2	$\mathbf{3}$	$\mathbf{1}$	$-\frac{1}{3}$	$\mathbf{1}$	-1
ψ_1^c	$\overline{\mathbf{3}}$	$\mathbf{1}$	$+\frac{1}{3}$	$\mathbf{1}$	$-\frac{1}{3}$
ψ_2^c	$\overline{\mathbf{3}}$	$\mathbf{1}$	$+\frac{1}{3}$	$\mathbf{1}$	$+\frac{1}{3}$
ψ_1'	$\mathbf{1}$	$\mathbf{2}$	$+\frac{1}{2}$	$\mathbf{1}$	$+1$
ψ_2'	$\mathbf{1}$	$\mathbf{2}$	$+\frac{1}{2}$	$\mathbf{1}$	-1
$\psi_1'^c$	$\mathbf{1}$	$\mathbf{2}$	$-\frac{1}{2}$	$\mathbf{1}$	$-\frac{1}{3}$
$\psi_2'^c$	$\mathbf{1}$	$\mathbf{2}$	$-\frac{1}{2}$	$\mathbf{1}$	$+\frac{1}{3}$
$\chi_{\alpha=1,2}$	$\mathbf{1}$	$\mathbf{1}$	0	$\mathbf{3}$	$+\frac{1}{2}$
$\chi^c_{\alpha=1,2}$	$\mathbf{1}$	$\mathbf{1}$	0	$\overline{\mathbf{3}}$	$+\frac{1}{2}$
λ	$\mathbf{1}$	$\mathbf{1}$	0	$\mathbf{8}$	$-\frac{1}{3}$
$N_{I=1,2,3,4}$	$\mathbf{1}$	$\mathbf{1}$	0	$\mathbf{1}$	$-\frac{2}{3}$
$N'_{I=1,2,3,4}$	$\mathbf{1}$	$\mathbf{1}$	0	$\mathbf{1}$	$-\frac{1}{6}$

Table 4.2: Field content beyond the SM [9]. All fields are Weyl fermions. Note that the messengers $(\psi_a, \psi_a') \oplus (\psi_a^c, \psi_a'^c)$ form complete $\mathbf{5}_a \oplus \overline{\mathbf{5}}_a$ multiplets of a grand-unified $SU(5)$ SM gauge group, but are chiral under $U(1)$.

The CP-violating Sector. The minimal CP-violating sector realizing the program spelled in 4.3.1 is composed of two families of χ, χ^c in the fundamental representation of a new confining $SU(3)$ gauge group and a single Weyl fermion λ in the adjoint representation. These fermions are all charged under the axial gauge $U(1)$, with charges chosen such that the anomaly $SU(3) \times SU(3) \times U(1)$ is absent.

This theory, when supplemented with small (CP-conserving by hypothesis) perturbations of the type $(\chi\chi^c)(\chi\chi^c)^\dagger / f_{\mathrm{UV}}^2$, breaks CP spontaneously. To assess the qualitative behavior of the non-perturbative dynamics let us first neglect all couplings except for the $SU(3)$ gauge interaction. Then χ, χ^c, λ enjoy a global anomaly-free symmetry $SU(2)_\chi \times SU(2)_{\chi^c} \times U(1)_{\rm V} \times U(1)_{\rm A}$, where $U(1)_{\rm V}$ is just the χ, χ^c baryon number, whereas the global $U(1)_{\rm A}$ acts on χ, χ^c, λ precisely as the $U(1)$ of table 4.2. While there is no definite proof, there are good reasons to expect that in the IR this theory develops two condensates $\langle \chi\chi^c \rangle \sim \langle \lambda\lambda \rangle$. Indeed, according to the arguments of [22] the vectorial subgroup $SU(2)_{\chi+\chi^c} \times U(1)_{\rm V}$ must remain intact. Hence the only allowed condensates are $\langle \lambda\lambda \rangle, \langle \chi\chi^c \rangle$. An heuristic argument based on the most attractive channel [202] suggests that $\langle \lambda\lambda \rangle$ is likely to form first, immediately followed by $\langle \chi\chi^c \rangle$. In the following we will therefore assume that both condensates form, and that they have comparable sizes.

After chiral symmetry breaking the $U(1)_{\rm A}$ Goldstone mode is eaten by the $U(1)$ gauge via the Higgs mechanism. The remaining 3 Nambu-Goldstone bosons π^q are described by the $SU(2)$ matrix $U = e^{i\pi^q\sigma^q/f}$, with σ^q the Pauli matrices. Below the chiral symmetry breaking scale, the condensates may be parametrized as

$$\begin{aligned} \langle \chi_\alpha \chi^c_\beta \rangle &= c_\chi 4\pi f^3 U_{\alpha\beta} \\ \langle \lambda\lambda \rangle &= c_\lambda 4\pi f^3 \end{aligned} \tag{4.3.14}$$

or some unknown parameters c_χ, c_λ expected to be of order unity. The latter are guaranteed to be CP-even, as demonstrated in subsection 4.3.1. The renormalizable theory we just described conserves CP and has a degenerate vacuum parametrized by any value of the π^q's. However, additional small interactions can break explicitly the accidental $SU(2)_\chi \times SU(2)_{\chi^c}$ symmetry and generate a potential for U. For example, a set of unavoidable dimension-6 interactions of the type

$$\frac{c_{\alpha\beta;\gamma\delta}}{f^2_{\rm UV}} (\chi_\alpha \chi^c_\beta)(\chi_\gamma \chi^c_\delta)^\dagger \tag{4.3.15}$$

breaks $SU(2)_\chi \times SU(2)_{\chi^c}$ completely if the (real) coefficients $c_{\alpha\beta;\gamma\delta}$ are generic. This operator in fact describes the dominant source of explicit chiral symmetry breaking in the model of table 4.2. Once (4.3.15) is included in the picture, the Nambu-Goldstone bosons acquire a small potential (see (4.3.14)) $V_{\rm NGB}(\pi) \sim 16\pi^2 f^4 (f/f_{\rm UV})^2\, c_{\alpha\beta;\gamma\delta} U_{\alpha\beta} U^*_{\gamma\delta}$, and the vacuum degeneracy is lifted. The actual vacuum configuration of the U field depends on the unknown CP-even parameters $c_{\alpha\beta;\gamma\delta}$. What matters to us here is not the exact expression of the vacuum state, however, but simply that CP is generically broken by complex entries of order (4.3.4). To prove this one has to observe that CP is spontaneously broken if $\langle U \rangle^* \neq \langle U \rangle$, i.e. if $\langle \pi_1 \rangle \neq 0$ or $\langle \pi_3 \rangle \neq 0$. It is then easy to see that, for generic CP-even coefficients $c_{\alpha\beta;\gamma\delta}$, $V_{\rm NGB}$ is indeed minimized at[8]

$$\mathrm{Im}\, \langle U \rangle \sim 1. \tag{4.3.16}$$

A numerical scan of the coefficients $c_{\alpha\beta;\gamma\delta}$ in the range $[-10, 10]$ confirms that this holds in more than 50% of the parameters space: spontaneous CP violation is generic in our model.

[8]As usual, all these expressions are to be interpreted as holding in the field basis in which all couplings are CP-even, according to the hypothesis of exact CP invariance of the UV.

The Mediator Sector. The messengers ψ, ψ^c of table 4.2 are vector-like under the SM subgroup $SU(3)_{\rm C} \times U(1)_{\rm Y}$, but chiral under the new gauge $U(1)$. They appear in four different representations in order to avoid an anomaly in $SU(3)_{\rm C} \times SU(3)_{\rm C} \times U(1)$ and $U(1)_{\rm Y} \times U(1)_{\rm Y} \times U(1)$. This implies the existence of a non-minimal number of mediators' families. Of course other constructions are possible, but we will stick to this one in the following.

The particles ψ', ψ'^c are introduced in table 4.2 with the sole scope of removing the $U(1)_{\rm Y} \times U(1) \times U(1)$ anomaly left by the ψ, ψ^c system. The ψ', ψ'^c mix with the SM lepton doublets similarly to how ψ, ψ^c mix with quarks. In this sense they can be seen as mediators for CP violation in the leptonic sector. Their presence does not spoil the solution of the Strong CP problem, as will be argued below. If preferred, it is possible to find alternative scenarios where the leptonic messengers are replaced by other fields with SM charges. In order for such alternatives to be phenomenological viable, though, the new states must have sufficiently large decay rates into SM particles. This is certainly the case with the field content of Table 4.2.

Extra states. The last degrees of freedom that need to be discussed are the N, N' particles. This SM neutral sector removes the remaining $U(1) \times U(1) \times U(1)$ and $U(1) \times$ grav $\times$ grav anomalies. One may find many different ways to cancel these anomalies, and the field content of table 4.2 is just an arbitrary choice. A different choice will be discussed later. It is worth to make a few comments, though. First, one cannot replace the spectator sector by increasing the number of χ, χ^c, λ families. If this avenue is pursued, the $SU(3)$ sector would contain flavors with different $U(1)$ charges; as a consequence the spontaneously broken abelian global symmetry and the gauged $U(1)$ would no more coincide. This would be a disaster because it would imply that the phase from $\lambda\lambda$ could not be removed by the Higgs mechanism and the Strong CP problem would not be solved. Second, we were unable to replace the SM-neutral N, N' with a slightly more involved (but still unstable) messenger sector of SM-charged states. This does not mean it is not possible, of course. Overall, our opinion is that a separate SM- and $SU(3)$-neutral spectator sector is a rather generic feature of our models.

B. Interactions

There is a very limited set of new couplings at the renormalizable level: the obvious kinetic terms (including those of the exotic gauge fields) and, since our model contains an abelian group, a kinetic mixing between hypercharge and the $U(1)$ gauge. The latter is however not relevant phenomenologically because the exotic vector acquires a large mass, as we will see below. On the other hand, there are no renormalizable couplings among the exotic fermions of table 4.2 and the SM fermions, and the former are all chiral. Non-gauge interactions involving the exotic fermions all arise from operators of dimension six or higher suppressed by powers of the UV cutoff $f_{\rm UV} \sim 10^{17-18}$ GeV, see the discussion around (4.3.9). Operators of dimension larger than 6 lead to effects suppressed by at least $v/f_{\rm UV} \lesssim 10^{-15}$ or $(f/f_{\rm UV})^3 \lesssim 10^{-15}$ and have no phenomenological impact (there is a single exception to this conclusion, i.e. the mass of N', which we discuss around (4.3.25).). Therefore, here we limit ourselves to a discussion of the leading dimension-6 interactions.

We are interested in four-fermion operators O involving the fermions of table 4.2. These appear in the effective Lagrangian as $cO/f_{\rm UV}^2$. Our working hypothesis of exact CP invariance at the Planck scale implies that all Wilson coefficients c are real in a suitable field basis. For definiteness we will take them to be of order unity. No particular flavour structure is favoured by our models.

Operators involving only the fermions $\psi, \psi^c, \psi', \psi'^c, N, N'$ and SM fields are strongly irrelevant. In particular they do not bring CP-violating phases and do not contribute appreciably to $\bar{\theta}$. On the other hand, operators containing χ, χ^c, λ can become important at low energies because after symmetry breaking they can be turned into *relevant* interactions. Let us thus focus on interactions with χ, χ^c, λ. The charge assignments of table 4.2 imply these belong to five distinct classes. First we find the operators that, after symmetry breaking, generate CP-violating mass terms as in (4.3.5):

$$(\chi_\alpha \chi^c_\beta)^\dagger \psi_1 d, \qquad \chi_\alpha \chi^c_\beta \psi_2 d, \qquad (\chi_\alpha \chi^c_\beta)^\dagger \psi'_1 \ell, \qquad \chi_\alpha \chi^c_\beta \psi'_2 \ell. \tag{4.3.17}$$

Note that the charge assignments do not allow analogous (and dangerous) dim-6 interactions among $\chi\chi^c$, ψ, ψ^c and the quark doublet q. In particular, the $U(1)$ charges are carefully chosen to forbid $\psi\psi^c\chi\chi^c$, which would generate a complex mass for the quark mediators and re-introduce a Strong CP problem. In the second class we have operators inducing CP-even masses of the type (4.3.8) after symmetry breaking. These are

$$\psi_1 \psi_1^c \lambda\lambda, \qquad \psi_2 \psi_2^c (\lambda\lambda)^\dagger, \qquad \psi'_1 \psi'^c_1 \lambda\lambda, \qquad \psi'_2 \psi'^c_2 (\lambda\lambda)^\dagger. \tag{4.3.18}$$

Third, we have operators involving only χ, χ^c, λ, like the ones in eq. (4.3.15). These are especially important because they represent the dominant source of explicit breaking of the chiral $SU(2)_\chi \times SU(2)_{\chi^c}$ symmetry group of the strong $SU(3)$ sector. Since we have already argued that CP gets spontaneously broken we do not need to analyze them in detail here. The fourth class of operators containing χ, χ^c, λ consists of

$$\chi_\alpha^\dagger \bar{\sigma}^\mu \chi_\beta \; J_\mu, \qquad \lambda^\dagger \bar{\sigma}^\mu \lambda \; J_\mu, \tag{4.3.19}$$

with J_μ indicating any gauge singlet vector current (of dimension 3) constructed out of the other fields, SM fields included. After chiral symmetry breaking this class of operators generate tiny (in general CP-violating) interactions between the $SU(3)$ hadrons and the rest of the world. Fortunately, they also do not affect $\bar{\theta}$ appreciably because they induce corrections suppressed by powers of $(f/f_{\rm UV})^2 \lesssim 10^{-10}$. Finally, the last class of operators involving the CP-violating sector parametrize flavour-violating interactions with the spectator sector:

$$\chi_\alpha \chi^c_\beta \lambda N_I. \tag{4.3.20}$$

This generates, after $SU(3)$ confinement, a mass for N_I, as we will discuss below. On the other hand, N'_I remain massless at dimension-6. We will estimate their mass in a moment.

C. Phenomenology

We now analyze in some detail the phenomenology of our model. Yet, before embarking in this journey, it proves useful to summarize the different mass scales in the theory. These are pictorially shown in figure 4.5.

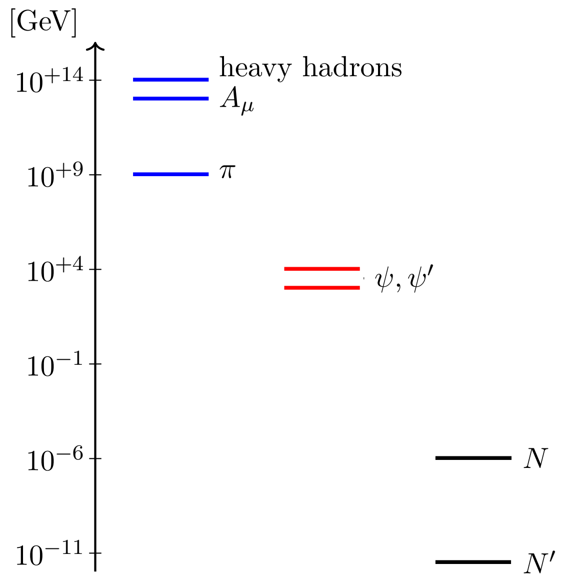

Figure 4.5: Schematic representation of the mass scales involved. See the text for details. Taken from [9].

The highest scale is the UV cutoff, parameterized by $f_{\rm UV} \sim 10^{17-18}$ GeV. One may identify it with the Planck scale, but we decided to keep our discussion more general. All masses of the particles beyond the SM masses arise from the dynamically generated scale [9]

$$m_{\not{C}\not{P}} \sim 4\pi f \sim 10^{13-14}\ \text{GeV} \tag{4.3.21}$$

of the confining $SU(3)$ sector. The natural hierarchy $f/f_{\rm UV}$ appears in many of the following expressions, so it is convenient to introduce the more compact notation

$$\epsilon \equiv \frac{f}{f_{\rm UV}}. \tag{4.3.22}$$

As we argued around eq. (4.3.9), $\epsilon \lesssim 10^{-5}$ ensures there are no large contributions to the QCD $\bar\theta$ parameter due to the vacuum expectation value of CP-odd resonances.

Let us now see what masses arise from chiral symmetry breaking. First, heavy $SU(3)$ hadrons all have masses of order $m_{\not{C}\not{P}}$. The $SU(3)$ dynamics also generates pseudo Nambu-Goldstone bosons, the key players in the spontaneous breaking of CP (see (4.3.16)). Their masses are induced dominantly by the interactions in (4.3.15) and are expected to be of order (see the corresponding potential $V_{\rm NGB}$) [9]

$$m_\pi \sim \epsilon\, m_{\not{C}\not{P}} \lesssim 10^{8-9}\ \text{GeV}. \tag{4.3.23}$$

Furthermore, chiral symmetry breaking generates a mass for the $U(1)$ vector, $m_{\rm A} \sim g_{\rm A} f \sim (g_{\rm A}/4\pi) m_{\not{C}\not{P}}$, with $g_{\rm A}$ the $U(1)$ gauge coupling. For definiteness we will assume that $g_{\rm A}$ is not far from order unity, so that $m_{\rm A} \gg m_\pi$. This assumption does not have any significant impact on our analysis, though.

Similarly, the fermions ψ, ψ^c and separately ψ', ψ'^c form, after chiral symmetry breaking, two families of Dirac pairs with CP-even masses generated by the interactions (4.3.18) and CP-odd mixings (4.3.17) with the SM d, ℓ representations. Overall, these are of order $m_\psi, m_{\psi'} \sim \epsilon^2\, m_{\not{C}\not{P}} \lesssim 1 - 10$ TeV (see also (4.3.8)).

The spectator sector of table 4.2 lives at scales parametrically smaller than the TeV. N gets a mass after a seesaw-like mixing with heavy fermionic hadrons $\sim \chi\chi^c\lambda$ via (4.3.20). A rough estimate says that [9]

$$m_N \sim \epsilon^4\, m_{\not{C}\not{P}} \lesssim 10^2 - 10^3 \text{ eV}. \tag{4.3.24}$$

Finally, the dominant contribution to the mass of N' arises from dimension-9 interactions like $N'N'\chi\chi^c\lambda\lambda$. After chiral symmetry breaking we get [9]

$$m_{N'} \sim \epsilon^5\, m_{\not{C}\not{P}} \lesssim 10^{-3} - 10^{-2} \text{ eV}. \tag{4.3.25}$$

Note that the larger powers of ϵ in (4.3.24), (4.3.25) compared to the other beyond the SM particles make $m_{N,N'}$ more sensitive to the actual $\mathcal{O}(1)$ couplings involved in our estimates.

We next turn to a study of how the Strong CP is solved, and then cosmological and collider signatures of our scenarios.

The CKM Phase and the Strong CP problem. It is easy to see why the model introduced in 4.3.2 solves the Strong CP problem. CP is spontaneously violated by the vacuum expectation value $\langle\chi\chi^c\rangle$. The excitations of the CP-violating sector (heavy hadrons and Nambu-Goldstone bosons) couple to the SM via small Planck-suppressed couplings (4.3.6) and can be ignored, as anticipated in (4.3.1). For all practical purposes the CP-violating sector is frozen and parametrized by the two condensates $\langle\chi\chi^c\rangle$ and $\langle\lambda\lambda\rangle$. Within the effective field theory at scales $\ll m_\pi \ll 4\pi f$ the relevant degrees of freedom are the SM particles and $\psi, \psi^c, \psi', \psi'^c, N, N'$. There is a unique CP-odd spurion, namely $\langle\chi\chi^c\rangle$, which couples to the colored sector solely via (4.3.17). The CP-violating theory we are describing is essentially that of table 4.1 with ξ replaced by $\langle\chi\chi^c\rangle/f_{\rm UV}^2$ and mediators' masses satisfying $|m_\psi| \sim |\xi|$ (see (4.3.8)). This theory reproduce the SM at scales $\ll m_\psi$, including the CKM phase, and a $\bar\theta$ very comfortably below the current bounds, as we showed in section 4.2.

Yet, our UV completion adds new ingredients to the effective field theory of table 4.1. It predicts the constraint (4.3.13) and additional (SM-charged) unstable particles ψ', ψ'^c, neutral states N, N', and Planck-suppressed CP-conserving interactions. The additional states cannot play any role in transferring CP violation to QCD. In particular, the leptonic CP-violating coupling in (4.3.17) is completely irrelevant for the Strong CP problem, since the additional CP-odd flavour-invariants felt by the colored particles are the same quark invariants found in the absence of ψ', ψ'^c with additional suppressing factors controlled by the tiny lepton Yukawa couplings. Furthermore N, N' are also not important for what concerns the Strong CP problem because they only interact with the SM via (CP-conserving) gauge-interactions and (CP-conserving) irrelevant couplings. In general, non-renormalizable operators cannot affect $\bar{\theta}$ appreciably because CP violation is super-soft and their contributions are therefore suppressed by powers of $m_\psi^2/(4\pi f_{\rm UV})^2$.

We conclude that our picture satisfies all the low energy requirements spelled out at the beginning of section 4.3, including (4.3.2), (4.3.1), and (4.2.4), and robustly solves the Strong CP problem.

Cosmology. We next study the cosmology of our models. Referring back to figure 4.5, we begin with a discussion of the heavy states beyond the SM, and then proceed to lower masses.

Most of the heavy hadrons are unstable and decay into pions or into SM particles and ψ, ψ', N, N' via (4.3.17), (4.3.18), (4.3.19). Similar considerations apply to the heavy $U(1)$ vector. Due to the axial $U(1)$, however, the baryons $\chi\chi\chi$ are practically stable. Hence, if thermalized, the exotic baryons would have decoupled at $T \sim m_{\not{C}\not{P}}/25$ and subsequently dominated the expansion rate until very recently. To avoid conflict with the physics of BBN, we assume that the temperature of our Universe has never exceeded [9]

$$T_{\rm RH} \ll \frac{m_{\not{C}\not{P}}}{25} \sim 10^{11-13}\ \text{GeV}. \tag{4.3.26}$$

This guarantees that baryons were never thermalized and their abundance was always safely within acceptable values. The constraint (4.3.26) also serves another purpose. If the Universe was hot enough to go through the CP-violating phase transition we may have ended up generating topological defects via the Kibble-Zurek mechanism. These may then have come to dominate the expansion of the Universe, which is also phenomenologically unacceptable. Thanks to the condition (4.3.26), though, the transition to the CP-violating vacuum occurred before or during inflation, so that any abundance of topological defects would have been diluted to an acceptable amount.

Pions, on the other hand, represent no cosmological hazard. They decay via (4.3.17) into messengers and SM fermions with lifetimes $\tau_\pi \sim 4\pi/(y^2 m_\pi)$, where $y \sim 10^{-9}$ was introduced in (4.3.6). Even if produced abundantly at re-heating, they would have disappeared from the plasma when the early Universe was at $T \sim$ few TeV.

Going further down in mass we encounter the states $\psi, \psi^c, \psi', \psi'^c, N, N'$. Obviously, since $\psi, \psi^c, \psi', \psi'^c$ mix with the SM quarks and leptons, they are unstable and decayed very quickly as soon as they decoupled.

The situation is a bit more complicated for N, N'. These particles are cosmologically stable and couple to the plasma only via higher-dimensional operators suppressed by the UV cutoff as well as gauge $U(1)$ interactions. For temperatures satisfying (4.3.26) they never thermalized. Under the reasonable assumption that they were not directly produced by the inflaton, a tiny population of N, N' was nevertheless generated at, and soon after, re-heating by the annihilation of nearly thermalized $\psi, \psi^c, \psi', \psi'^c$. The latter processes are mediated by the effective 4-fermion interaction

$$\mathcal{L}_{\rm eff} \supset -\frac{1}{f^2} J_\mu^{\rm A} \left[q_N N^\dagger \bar{\sigma}^\mu N + q_{N'} N'^\dagger \bar{\sigma}^\mu N'\right] \tag{4.3.27}$$

where $J_\mu^{\rm A} = \sum_i q_i\ \Psi_i{}^\dagger \bar{\sigma}_\mu \Psi_i$ is the $U(1)$ current of the thermalized charged fermions $\Psi_i = \psi, \psi^c, \psi', \psi'^c$ and q_i their $U(1)$ charges. We would like to provide a quantitative estimate of the energy density carried by the spectators. In order to do so it is enough to focus on the N's because N' have a much smaller mass and their energy density is suppressed by a factor $m_{N'}/m_N \sim \epsilon \lesssim 10^{-5}$ compared to that of N.

An approximate estimate of the yield $Y_N = n_N/s$ is obtained via the Boltzmann equation $dY_N/dt = \Gamma_{\rm coll}/s$. If the Ψ_i's are taken to have had a thermal distribution, for simplicity, the collision term is given by

$$\Gamma_{\rm coll} = q_N^2 \left(\sum_i g_i q_i^2\right) \frac{1}{18\pi^5 f^4} \left(\frac{7\pi^4 T^4}{120}\right)^2 \tag{4.3.28}$$

where g_i is the multiplicity of each Ψ_i (helicity excluded). From table 4.2 we have $q_N^2 \sum_i g_i q_i^2 = 400/81$. Finally, the present-day energy in units of the entropy, $\rho_N/s = \sum_{I=1}^4 m_{N_I} Y_N$ (there is a family of 4 N's in table 4.2), is approximately [9]

$$\begin{aligned} \frac{\rho_N}{s} &\sim \sum_{I=1}^4 m_{N_I} \frac{\Gamma_{\rm coll}}{Hs}\bigg|_{T_{\rm RH}} \\ &\sim \frac{\rho_{\rm DM}}{s} \left(\frac{\sum_{I=1}^4 m_{N_I}}{4 \text{ KeV}}\right) \left(\frac{10^{13} \text{ GeV}}{f}\right)^4 \left(\frac{T_{\rm RH}}{10^{11} \text{ GeV}}\right)^3 . \end{aligned} \tag{4.3.29}$$

Since this expression is dominated by the large-T regime, where details of re-heating can be important, our computation should be viewed as a qualitative estimate of the actual density. Nevertheless, the message is clear: in the small $T_{\rm RH}$ regime (4.3.26), where eq. (4.3.29) was consistently derived, our spectator sector represents generically a subleading component of dark matter. Yet, with some luck it could be a portion of the missing matter of the Universe, with N a good cold dark matter candidate and N' a negligible component of dark radiation.

Collider Signatures. The particles $\psi, \psi^c, \psi', \psi'^c$ are subject to collider, electroweak, and flavor constraints. A detailed analysis of the quark mediators was done in section 4.2. The bottom line is that most of parameter space is allowed for masses above the TeV.

The physics of the lepton mediators has not been discussed before in this context, but it is easy to show that these states lead to weaker constraints on the parameters of our scenarios compared to the quark mediators. As for the quark mediators, the ψ'-ℓ mixing can be removed via a rotation of (ℓ, ψ'^c). The massive eigenstate couples to the Higgs and the lepton singlets with coupling that up to a unitary rotation is oriented along the direction of the SM lepton Yukawa coupling Y_e, i.e. $Y' \propto Y_e$.

The ψ', ψ'^c are produced in pairs via Drell-Yan, and then decay into leptons and vector bosons or the Higgs. Current constraints are looser than for quark mediators and are not relevant to our models, where $m_{\psi'} \sim m_\psi$ is the natural expectation. Deviations of the Z couplings to leptons are constrained at the permille level. The corresponding bounds are not much stronger than those of the quark mediators, however, because the new coupling Y' is proportional to the SM lepton Yukawa and therefore highly hierarchical. The most significant constraint from flavor-violation comes from the non-observation of $\mu \to e\gamma$ and is well under control for $|Y_e^{-1} Y'| \lesssim 300 \times m_{\psi'}/\text{TeV}$. CP violation, including the electric dipole moment of the electron, is strongly suppressed. Overall, we conclude that lepton mediators are allowed to live at the TeV scale.

Baryogenesis. Explaining the observed baryon asymmetry may at first sight appear difficult in our scenarios, as baryogenesis necessitates of both new sizable CP-violating phases and new interactions with the SM. Yet, successfull baryogenesis above the weak scale does not require new couplings to the *colored sector*, and therefore does not immediately jeopardize our solution of the Strong CP problem. In fact, it may be realized simply adding new physics with CP-violating couplings to the leptons. Low-energy leptogenesis thus appears to be the most natural and safe option in these models.

A slight modification of our model has all the necessary ingredients. Suppose we replace the spectator sector of table 4.2 with this more complicated set of fermions neutral under the SM and the new $SU(3)$ but chiral under the $U(1)$ [9]:

	$U(1)$
$N_{1,2}$	$-\frac{1}{3}$
$N'_{1,\cdots,5}$	$-\frac{2}{3}$
$X_{1,2,3}$	$+\frac{1}{2}$
$X'_{1,\cdots,5}$	$-\frac{1}{6}$

The main difference compared to the spectator sector of table 4.2 is that with this modified field content one finds a renormalizable coupling $\psi_2'^c HN$ as well as dimension-6 interactions that generate complex Majorana masses of order the TeV for N, N', X (note that N' mixes with N). The state X' instead obtains a mass from dimension-9 interactions and is expected to be of order (4.3.25). In an appropriate portion of the parameter space the modified model may thus generate the observed baryon asymmetry via resonant decays $N, N' \to \psi'^{c\dagger} H^\dagger, \psi'^c H \to \ell^\dagger H^\dagger, \ell H$ as studied in [203]. The other fields of the spectator sector would behave qualitatively as in the scenario discussed before; X would be a cold dark matter candidate and X' a negligible part of radiation. The crucial difference compared to our earlier model is that here X is much heavier and for this reason has an abundance that is roughly a factor 10^9 larger compared to (4.3.29). This implies that the re-heating temperature is allowed to be three orders of magnitude smaller, $T_{\rm RH} \sim 10^8$ GeV. We can thus have a viable dark matter candidate comfortably within the allowed regime (4.3.26). The phenomenology of this modified version of our scenario is quite rich and would deserve further scrutiny. Our purpose here is merely to demonstrate that there is no structural obstruction to incorporating a mechanism for baryogenesis in our scenarios.

Appendices

4.A Flavour Invariants

The invariants $I_{n,m}$ of (4.2.14) can be systematically expanded in powers of the Cabibbo angle employing the numerical relations $m_u/m_t \sim \lambda_C^7$, $m_c/m_t \sim \lambda_C^4$ and $m_d/m_b \sim \lambda_C^4$, $m_s/m_b \sim \lambda_C^2$. In the field basis in which the up Yukawa is diagonal ($Y_u = \widehat{Y}_u$ and $Y_d = V_{\rm CKM}^* \widehat{Y}_d$) the leading terms are [8]

$$
\begin{aligned}
I_{2,1} &= Y^\dagger \left[H_u, [H_u, H_d]^2\right] Y \\
&= 2\widehat{Y}_t^4 \widehat{Y}_c^2 \widehat{Y}_b^4 \widehat{Y}_s^2 \left[1 + \mathcal{O}\left(\lambda_C^2\right)\right] \\
&\times \left[-A^2 \eta \left(\frac{|\xi_2|^2}{m_\psi^2} + 2\frac{|\xi_3|^2}{m_\psi^2} \right) \lambda_C^6 - A^2 \eta \frac{\widehat{Y}_d}{\widehat{Y}_s} \lambda_C^5 \frac{\operatorname{Re} \xi_2 \xi_1{}^\dagger}{m_\psi^2} - A^2 \eta \frac{\widehat{Y}_s}{\widehat{Y}_b} \lambda_C^4 \frac{\operatorname{Re} \xi_3 \xi_2{}^\dagger}{m_\psi^2} \right. \\
&+ \left. A(1-\rho) \frac{\widehat{Y}_s}{\widehat{Y}_b} \lambda_C^4 \frac{\operatorname{Im} \xi_3 \xi_2{}^\dagger}{m_\psi^2} + A \frac{\widehat{Y}_d}{\widehat{Y}_b} \lambda_C^3 \frac{\operatorname{Im} \xi_3 \xi_1{}^\dagger}{m_\psi^2} + A^2 \rho \frac{\widehat{Y}_d}{\widehat{Y}_s} \lambda_C^5 \frac{\operatorname{Im} \xi_1 \xi_2{}^\dagger}{m_\psi^2} \right],
\end{aligned} \tag{4.A.1}
$$

$$
\begin{aligned}
I_{1,2} &= Y^\dagger H_u \left[H_u, H_d\right] H_u Y \\
&= 2\widehat{Y}_t^4 \widehat{Y}_c^2 \widehat{Y}_b^4 \left[A \frac{\widehat{Y}_s}{\widehat{Y}_b} \lambda_C^2 \frac{\operatorname{Im} \xi_3 \xi_2{}^\dagger}{m_\psi^2} \right] \left[1 + \mathcal{O}\left(\lambda_C^2\right)\right],
\end{aligned} \tag{4.A.2}
$$

$$
\begin{aligned}
I_{1,1} &= Y^\dagger \left\{H_u, [H_u, H_d]\right\} Y \\
&= 2\widehat{Y}_t^4 \widehat{Y}_b^4 \left[A \frac{\widehat{Y}_s}{\widehat{Y}_b} \lambda_C^2 \frac{\operatorname{Im} \xi_3 \xi_2{}^\dagger}{m_\psi^2} \right] \left[1 + \mathcal{O}\left(\lambda_C^2\right)\right],
\end{aligned} \tag{4.A.3}
$$

$$
\begin{aligned}
I_{1,0} &= Y^\dagger \left[H_u, H_d\right] Y \\
&= 2\widehat{Y}_t^2 \widehat{Y}_b^4 \left[A \frac{\widehat{Y}_s}{\widehat{Y}_b} \lambda_C^2 \frac{\operatorname{Im} \xi_3 \xi_2}{m_\psi^2} \right] \left[1 + \mathcal{O}\left(\lambda_C^2\right)\right].
\end{aligned} \tag{4.A.4}
$$

In calculating (4.2.13) the subleading terms are crucial because huge cancellations occur. For example, the invariants $I_{1,1}, I_{1,0}$ are not proportional to m_c^2 at leading order, but their sum in (4.2.13) is. Similarly, several other important cancellations take place.

There are five more CP-odd flavour invariants one can build out of H_u, H_d, and Y, but $I_{1,2}, I_{1,1}, I_{1,0}$ are the most relevant. Importantly, the largest ones in size (see $I_{1,0}, I_{1,1}$) have a similar dependence on $A(\widehat{Y}_s/\widehat{Y}_b)\lambda_C^2 \ \operatorname{Im} \xi_3 \xi_2{}^\dagger / m_\psi^2$. This explains the similarity in the factors in front of (4.2.23), (4.2.25), (4.2.37).

4.B Mass Basis

4.B.1 Diagonalization

After electroweak symmetry breaking there appear a new mass mixing of order Yv/M in the down sector, see (4.2.8). We introduce the 4-family vectors

$$D = \begin{pmatrix} q_d \\ \psi \end{pmatrix} \qquad D^c = \begin{pmatrix} d \\ \psi^c \end{pmatrix} \tag{4.B.1}$$

and diagonalize the mass matrix via $SU(4)$ rotations $D \to U_D D$, $D^c \to U_{D^c} D^c$,

$$U_D^t \begin{pmatrix} \frac{Y_d v}{\sqrt{2}} & \frac{Y v}{\sqrt{2}} \\ 0 & M \end{pmatrix} U_{D^c} = \widehat{M}_d. \tag{4.B.2}$$

The resulting couplings to $W^\pm, Z$ and the Higgs boson h (defined as $H^t = (0, v+h)/\sqrt{2}$ in the unitary gauge) read:

$$\begin{aligned}
\mathcal{L}_W &= -\frac{g}{\sqrt{2}} V_{i\alpha} [{q_u}^\dagger]_i \bar{\sigma}^\mu D_\alpha W^+_\mu + \text{h.c.} \\
\mathcal{L}_Z &= -\frac{g}{2c_w} \left[[{q_u}^\dagger]_i \bar{\sigma}^\mu [q_u]_i - \mathcal{Z}_{\alpha\beta} {D_\alpha}^\dagger \bar{\sigma}^\mu D_\beta - 2 s_w^2 J^\mu_{\text{EM}} \right] Z_\mu \\
\mathcal{L}_h &= -\frac{[\widehat{M}_u]_{ij}}{v} [q_u]_i u_j h - [\mathcal{Y}]_{\alpha\beta} D_\alpha D^c_\beta h + \text{h.c.}
\end{aligned} \tag{4.B.3}$$

where J^μ_{EM} is the flavour-diagonal QED current of q_u, u, D, D^c, and

$$\begin{aligned}
V_{i\alpha} &= [{U_{q_u}}^\dagger]_{ij} [U_D]_{j\alpha} \\
\mathcal{Z}_{\alpha\beta} &= [{U_D}^\dagger]_{\alpha i} [U_D]_{i\beta} \\
&= \delta_{\alpha\beta} - [U_D^*]_{4\alpha} [U_D]_{4\beta} \\
[\mathcal{Y}]_{\alpha\beta} &= \frac{[\widehat{M}_d]_{\alpha\beta}}{v} - \frac{M}{v} [U_D]_{4\alpha} [U_{D^c}]_{4\beta}.
\end{aligned} \tag{4.B.4}$$

Explicit expressions for $V, \mathcal{Z}, \mathcal{Y}$ can be derived as an expansion in Yv/M (see for example [187, 188]). We show here only the leading order:

$$\begin{aligned}
U_D &= \begin{pmatrix} U & \frac{Y^* v}{\sqrt{2}M} \\ -\frac{Y^t v}{\sqrt{2}M} U & 1 \end{pmatrix} \left[1 + \mathcal{O}(Y^2 v^2 / M^2) \right] \\
U_{D^c} &= \begin{pmatrix} U' & 0 \\ 0 & 1 \end{pmatrix} \left[1 + \mathcal{O}(Y^2 v^2 / M^2) \right]
\end{aligned} \tag{4.B.5}$$

with $U^t (Y_d {Y_d}^\dagger) U^* = U'^\dagger ({Y_d}^\dagger Y_d) U' = 2 \widehat{M}_d^2 / v^2$.

4.B.2 Decoupling Contributions to $\bar\theta$

Here we describe the irreducible, decoupling contributions to $\bar\theta$ in models of $d-$mediation. All estimates in this section are obtained working in the electroweak basis in which Y_u is diagonal. In this case one effectively has $U_{q_u} = U_u = 1$ and the expressions in (4.B.4) are all controlled by $V = U_D$ and U_{D^c}; also, at leading order the matrix U should be identified with the 3-dimensional CKM matrix, see (4.B.5).

First off, it is easy to see that there are no 1-loop corrections to $\bar\theta$, similarly to the SM. CP violation at 2-loops may arise due to loops of the $W^\pm$, Z or the Higgs h. These can again contribute via corrections to the Yukawas (4.2.20) or direct contributions to θ. Inspecting the couplings of (4.B.3) we see that 2-loop diagrams with fermions and $W^\pm$ (we will refer to these as $W^\pm - W^\pm$ diagrams) generate corrections directly to θ of the form [8]

$$\begin{aligned}\bar\theta\Big|_{\text{nonanaly},WW} =& \left(\frac{g^2}{16\pi^2}\right)^2 \text{Im}\left([V]_{i\alpha}[V^*]_{i\beta}[V]_{j\beta}[V^*]_{j\alpha}\right)\ F^{ij;\alpha\beta}_{1WW} \\ &+ \left(\frac{g^2}{16\pi^2}\right)^2 \text{Im}\left([V]_{i\alpha}[V^*]_{i\beta}[V]_{4\beta}[V^*]_{4\alpha}\right)\ F^{i;\alpha\beta}_{2WW},\end{aligned} \tag{4.B.6}$$

with $i,j = 1,2,3$ and $\alpha,\beta = 1,2,3,4$. Here $V_{i\alpha}$ a generalization of the CKM matrix and $F_{1WW,2WW}$ are real functions of the masses (squared) indicated by the corresponding indices. A sum over families is understood, though the imaginary prefactor is non-zero only when $i \neq j$ and $\alpha \neq \beta$. The functions $F_{1WW,2WW}$ are constrained by a number of physical considerations. By unitarity of the 4 by 4 matrix $V_{\alpha\beta}$, eq. (4.B.6) vanishes unless $F_{1WW,2WW}$ depend on both m^2_α and m^2_β. Indeed, if the dependence on m^2_α was absent one could sum over α and obtain a vanishing expression because of a generalized GIM mechanism. Similar logic applies to β. On the other hand, if no dependence on m^2_j (or equivalently m^2_i) exists we may replace $[V]_{j\beta}[V^*]_{j\alpha} = \delta_{\alpha\beta} - [V]_{4\beta}[V^*]_{4\alpha}$ in the above expression. The $\delta_{\alpha\beta}$ does not contribute but the reminder does not vanish. In practice, we can split eq. (4.B.6) into a piece that has a non-trivial dependence on both $i \neq j$ (F_{1WW}) and one that does not depend on one of the two, say on j (F_{2WW}).

Incidentally, the contribution $\propto F_{2WW}$ has exactly the same structure obtained in 2-loop corrections to $\bar\theta$ from diagrams with fermions, one $W^\pm$ and one Z ($W^\pm - Z$ for short), up to an irrelevant correction [8]:

$$\bar\theta\Big|_{\text{nonanaly},WZ} = \left(\frac{g^2}{16\pi^2}\right)^2 \text{Im}\left([V]_{i\alpha}[V^*]_{i\beta}[V]_{4\beta}[V^*]_{4\alpha}\right)\ F^{i;\alpha\beta}_{WZ}. \tag{4.B.7}$$

To prove this observe that the $W^\pm - Z$ loop is proportional to (see (4.B.3))

$$\begin{aligned}&\sum_\sigma \text{Im}\left([V]_{i\beta}[V^*]_{i\gamma}[\mathcal{Z}]_{\sigma\gamma}[\mathcal{Z}^*]_{\sigma\beta}\right)\ F'^{i;\beta\gamma\sigma}_{WZ} \\ &= \text{Im}\left([V]_{i\beta}[V^*]_{i\gamma}[V]_{4\beta}[V^*]_{4\gamma}\right)\left[\sum_\sigma |V_{4\sigma}|^2 F'^{i\beta\gamma\sigma}_{WZ} - F'^{i\beta\gamma\gamma}_{WZ} - F'^{i\beta\gamma\beta}_{WZ}\right] \\ &\equiv \text{Im}\left([V]_{i\beta}[V^*]_{i\gamma}[V]_{4\beta}[V^*]_{4\gamma}\right) F^{i;\beta\gamma}_{WZ}\end{aligned} \tag{4.B.8}$$

For the same reason explained above, F_{WZ} must depend on $m_i^2, m_\beta^2, m_\gamma^2$ otherwise the sum vanishes. The resulting structure is the one in (4.B.7), as promised. The loop diagrams with only fermions and Z vanish because

$$\mathrm{Im}\left([\mathcal{Z}]_{\alpha\beta}[\mathcal{Z}^*]_{\alpha\gamma}[\mathcal{Z}]_{\sigma\gamma}[\mathcal{Z}^*]_{\sigma\beta}\right) = 0 \qquad \text{(no sum)}, \tag{4.B.9}$$

as can be seen from the explicit expression of $[\mathcal{Z}]_{\alpha\beta}$ given in (4.B.3).

Diagrams with virtual Higgs bosons also contribute, but we will see below they are subleading. Before showing this we discuss a key constraint that the functions $F_{1WW,2WW,WZ}$ are subject to. Because there cannot occur IR divergences in matching the UV to the SM effective field theory below the scale M, $F_{1WW,2WW,WZ}$ must be well-behaved when $M^2 \gg m_{i,j}^2, m_{\alpha,\beta}^2, m_W^2$. Hence

$$F_{1WW}^{ij;\alpha\beta} = \min\left(\frac{m_i^2}{m_W^2}, \frac{m_j^2}{m_W^2}\right) \min\left(\frac{m_\alpha^2}{m_W^2}, \frac{m_\beta^2}{m_W^2}\right) \widetilde{F}_{1WW}\left(\frac{m_i^2}{m_\alpha^2}, \frac{m_j^2}{m_\alpha^2}, \frac{m_\beta^2}{m_\alpha^2}, \frac{m_W^2}{m_\alpha^2}\right) \tag{4.B.10}$$

$$F_{2WW}^{i;\alpha\beta} = \frac{m_i^2}{m_W^2} \min\left(\frac{m_\alpha^2}{m_W^2}, \frac{m_\beta^2}{m_W^2}\right) \widetilde{F}_{2WW}\left(\frac{m_i^2}{m_\alpha^2}, \frac{m_\beta^2}{m_\alpha^2}, \frac{m_W^2}{m_\alpha^2}\right) \tag{4.B.11}$$

$$F_{WZ}^{i;\alpha\beta} = \frac{m_i^2}{m_W^2} \min\left(\frac{m_\alpha^2}{m_W^2}, \frac{m_\beta^2}{m_W^2}\right) \widetilde{F}_{WZ}\left(\frac{m_i^2}{m_\alpha^2}, \frac{m_\beta^2}{m_\alpha^2}, \frac{m_W^2}{m_\alpha^2}, \frac{m_Z^2}{m_\alpha^2}\right), \tag{4.B.12}$$

with the powers of $m_{i,\alpha}^2/m_W^2$ ensuring that $F_{1WW,2WW,WZ}$ be regular when *any* of the quark masses involved vanishes. Of course this does not mean that the result is inevitably proportional to m_u^2 or m_d^2, because the CP-odd factor $\mathrm{Im}\,[V]_{i\alpha}[V^*]_{i\beta}[V]_{j\beta}[V^*]_{j\alpha}$ is non-vanishing only for specific combinations of indices, the dominants of which do not necessarily involve the first generation. Actually (4.B.12) demonstrates that the dominant contributions to $\bar{\theta}$ arise from diagrams in which the heavier generations run in the loop. Finally, the end result for $\bar{\theta}$ should clearly be regular as $g \to 0$. This tells us that $\widetilde{F}_{1WW,2WW,WZ}$ are analytic in the vector boson masses. The dominant contributions can be calculated for $m_W^2/M^2 \to 0$.

As a check of our arguments, one can verify that (4.B.6) together with (4.B.12) reproduce the structure of the non-analytic contributions to $\bar{\theta}$ found in the SM [52]. The unitarity of the CKM matrix however forces a non-trivial cancellation at 2-loops. Such degeneracy is lifted in diagrams with an additional strong coupling (or photon) loop, and so the actual end result is down by a factor $g_s^2/16\pi^2$ compared to (4.B.6). In our case no such cancellation takes place because there is no “heavy top quark" to compensate for the fourth $d-$type family. For this reason $\bar{\theta}$ is already corrected at 2-loops.

At this point we have all the tools to estimate the size of the non-analytic contributions to $\bar{\theta}$. We begin considering contributions to F_{1WW}, where both $i, j \neq i$ appear. The largest mass factor from (4.B.12) is obtained when $i = 3, j = 2$ and $\alpha = 4, \beta = 3$. The proportionality to $(m_c^2/m_W^2)(m_b^2/m_W^2)$ renders such corrections rather innocuous. In addition, (4.B.6) turns out to be very small as well. Using the approximate expressions for V derived in the previous appendix, and recalling that Y can be written as a function of the SM Yukawa Y_d and ξ/m_ψ as in (4.2.9), we find that $\mathrm{Im}\,[V]_{34}[V^*]_{33}[V]_{23}[V^*]_{24} \sim (m_b/M)\lambda_C^2(m_s/M)(|\xi|^2/m_\psi^2)$. Even including potentially large logs, the resulting contribution to $|\bar{\theta}|$ is at most numerically comparable to (4.2.23) for $M \sim 1$ TeV. The effect becomes subleading with $(\mathrm{TeV}/M)^2$ as soon as the mediator mass is above the TeV, which has to be the case because of direct searches.

Next we turn to an estimate of the $W^\pm - W^\pm$ loops controlled by F_{2WW}, or analogously of the $W^\pm - Z$ loop in eq. (4.B.7), which we argued to be comparable parametrically. In this case the largest mass enhancement is obtained with $i = 3$, $\alpha = 4$, $\beta = 3$, when the fermions running in the loop are the top, the bottom, and the heavy fermion. This is much larger than the effect proportional to F_{1WW} that we just analyzed. On the other hand,

$$\mathrm{Im}\,[V]_{34}[V^*]_{33}[V]_{43}[V^*]_{44} \sim \lambda_C^2 \frac{m_b m_s}{M^2} \frac{\mathrm{Im}\,\xi_2 \xi_3{}^\dagger}{m_\psi^2} \tag{4.B.13}$$

is comparable to the imaginary part found above. The final result is thus expected to be of order [8]

$$\bar{\theta}\Big|_{\text{nonanaly},WZ} = c_{\text{nonanaly}} \left(\frac{g^2}{16\pi^2}\right)^2 \frac{m_t^2}{m_W^2}\frac{m_b^2}{m_W^2}\,\lambda_C^2 \frac{m_b m_s}{M^2}\frac{\mathrm{Im}\,\xi_i\xi_j{}^\dagger}{m_\psi^2} \tag{4.B.14}$$

$$\sim c_{\text{nonanaly}}\; 10^{-16} \left(\frac{\text{TeV}}{M}\right)^2 \frac{\mathrm{Im}\,\xi_i\xi_j{}^\dagger}{m_\psi^2}. \tag{4.B.15}$$

We discussed it below (4.2.25). Here we just observe that an independent way to understand the necessity of factors of quark masses in front of (4.2.25) is to note that when all SM fermions are degenerate we can put V_{ij} in diagonal form, in which case (4.B.6) vanishes.

Loops with fermions and 2 virtual Higgses, or one Higgs and one $W^\pm$, or one Higgs and one Z are respectively controlled by (no sum over indices is implied)

$$\mathrm{Im}\,\left([\mathcal{Y}]_{\alpha\beta}[\mathcal{Y}^*]_{\alpha\gamma}[\mathcal{Y}]_{\sigma\gamma}[\mathcal{Y}^*]_{\sigma\beta}\right) \tag{4.B.16}$$

$$\mathrm{Im}\,\left([V]_{i\beta}[V^*]_{i\gamma}[\mathcal{Y}]_{\gamma\sigma}[\mathcal{Y}^*]_{\beta\sigma}\right) \tag{4.B.17}$$

$$\mathrm{Im}\,\left([\mathcal{Z}]_{\alpha\beta}[\mathcal{Z}^*]_{\alpha\gamma}[\mathcal{Y}]_{\gamma\sigma}[\mathcal{Y}^*]_{\beta\sigma}\right) = |V_{4\alpha}|^2\;\mathrm{Im}\,\left([V_{4\beta}[V^*]_{4\gamma}[\mathcal{Y}]_{\gamma\sigma}[\mathcal{Y}^*]_{\beta\sigma}\right). \tag{4.B.18}$$

We inspected these structures and found that a non-vanishing correction to θ or $F_{u,d}$ in (4.2.20) can only be obtained if subleading terms in the mixing $\sim Yv/M$ between ψ and the SM are taken into account. As a result corrections to $\bar{\theta}$ due to loops of the Higgs boson are always smaller than in (4.B.15).

Chapter 5

Summary and Outlook

The Standard Model of particle physics is one of the most successful theories of all time. The precision to which it has been tested is unmatched in science, and after more than fifty years it is still incredible how such a simple model can predict so closely what we observe in Nature. Yet, we know that this cannot be the end of the story. Dark Matter, Baryogenesis, a quantum description of gravity, are all problems that cannot be addressed within this simple theory. The Standard Model needs to be extended. Still, the ultimate theory is expected to reproduce the Standard Model parameters and its particle content in a *natural* way, namely without fine-tunings. However, experiments clearly indicate that the pattern of CP violation in the Standard Model is highly non-generic. This suggests that CP violation in its completion must also be non-generic. It must apparently feature sizeable CP-odd phases, in order to explain CP violation in the weak interactions, while simultaneously justify the absence of CP violation in the strong sector. What properties should the UV completion of the Standard Model have in order to accommodate these experimental facts without having to fine-tune its parameters? Identifying theories with the correct characteristics defines what is known as the *Strong CP problem*.

Among all the naturalness problems, the Strong CP problem is particularly special: it is associated to a topological angle, that of QCD. Thus, it would not be a problem at all if the strong interactions remained perturbative. In reality these become non-perturbative around the GeV scale and, even though understanding CP violation in that regime is not a trivial task, we are still able to make quantitative predictions thanks to modern tools such as chiral lagrangian techniques and lattice computations. Today's theoretical understanding is however possible only thanks to a number of seminal works that elucidated the connection between the topological angle, the QCD vacuum structure and the $U(1)_{\mathrm{A}}$ problem, as revised in chapter 2. In particular, the intimate connection with the $U(1)_{\mathrm{A}}$ problem underlines how the Strong CP problem is real and *cannot* be solved by the strong dynamics itself: it requires some external explanation.

Another remarkable property of the QCD angle is that although its smallness is not technically natural, its infinite renormalization can appear only at 7-loops order or more in the Standard Model. Thus, if thanks to some mechanism it set to a small enough value at the matching scale between the Standard Model and its completion, a big angle is not regenerated below that threshold and the Strong CP problem may be effectively solved. This is not true in generic BSM scenarios. As we showed in chapter 2, infinite corrections to the QCD angle may appear already at the 2-loops order in sufficiently generic theories and may invalidate the mechanism fixing its value at the threshold scale. This result is particularly relevant for UV solutions to the Strong CP problem, where the perturbative stability of $\bar{\theta}$ as set at the threshold is crucial to provide realistic models. For this reason, an important future extension would be to actually compute the numerical coefficient in front of the 2-loops beta function. Interestingly, this computation can be carried out in any model as long as the tensorial structure does not vanish, even though it requires the calculation of a highly non-trivial 3-loops diagram. In addition, we also showed how quite generally the running of $\bar{\theta}$ starts already at 1-loop order. Identifying theories in which both the running of θ and $\bar{\theta}$ occurs at minimum at 2-loops order may therefore be the first step towards a sistematical classification of possible UV solutions to the Strong CP problem.

A multitude of solutions to the Strong CP problem have been proposed. By far, the most popular one is the QCD axion, studied in chapter 3. In this simple setup the QCD angle is replaced by a dynamical field, called *axion*, that naturally relaxes its vacuum expectation value to a CP-conserving minimum when QCD becomes strongly coupled. This elegant solution, however, crucially relies on the existence of an anomalous $U(1)_{\rm PQ}$ symmetry that must be exact up a very high degree except for its anomalous coupling to the QCD topological term. The associated fine-tuning introduces a *quality problem* that, if not addressed, renders the axion solution to Strong CP problem ineffective. This issue may be faced in two ways: one can either find a mechanism to suppress the potentially dangerous perturbations, or find new corrections to the axion potential such that the QCD contribution gets effectively enhanced. In chapter 3 we investigated the latter option by postulating that the dominant contribution to the axion potential arises from an additional gauge group confining at scales much higher than QCD. The model-building challenge posed by this interesting idea led to an elegant scenario with a very minimal field content, in which the two groups unify at high-scales into a Grand Color group. The resulting axion is parametrically heavier than the standard one, sizeably ameliorating the quality problem and opening a new region in the usual $m_a - f_a$ parameter space with mass around the GeV and a TeV decay constant. Interestingly, this “visible axion” as well as other striking phenomenological signatures of the model will be testable in the near future at collider experiments and via cosmological observations. Therefore, it could prove useful to study in greater detail the phenomenology of the model, or even to explore scenarios beyond the minimal realization presented in this chapter, for example compatible with a grand unification scheme.

The other possible option to tackle the Strong CP problem is to impose an UV condition such that at the matching scale between the Standard Model and its completion the QCD angle is set to a value consistent with experimental observations. This gives rise to the class of *UV solutions* to the Strong CP problem, as opposed to IR-effective ones such as the QCD axion. The basic assumption of these solution is an exact generalized CP symmetry in the UV, which gets spontaneously broken at lower scale. In chapter 4 we gave a brief overview of the most popular constructions of this kind, focussing then on models falling in the Nelson-Barr class. We scrupulously analysed the viability of these models by deriving the conditions for which the Strong CP problem is solved while simultaneously the correct amount of low-energy CP violation, encapsulated in the CKM phase, is reproduced and the most recent experimental bounds are complied with. This required a careful estimation of loop corrections to $\bar{\theta}$ up to 3-loops as well as a detailed analysis of the new particles' influence on the SM quark masses and on low and high-energy observables. As a result, we found that these models must satisfy a non-trivial coincidence between a priori unrelated CP-odd and CP-even mass scales. This immediately leads to the question of how naturally this coincidence can be generated in concrete UV completions. In the last part of chapter 4 we showed how this property can emerge by gauge invariance and a CP-conserving, but otherwise generic, physics at the Planck scale. We presented a construction in which the spontaneous breaking of CP is due to a confining non-Abelian dynamic, automatically realizing the crucial coincidence without any additional fine-tuning. The model is remarkably predictive, requiring vector-like quarks and leptons around the TeV scale and a dark sector featuring interesting cosmological signatures, among which viable dark matter candidates. We also showed how baryogenesis can be easily accounted for in this scheme. As a future direction, thus, it would be interesting to perform a more accurate study of the cosmological history of our model to single out the best implementation capable of simultaneously reproducing the amount of dark matter and of matter-antimatter asymmetry observed in our universe.

After more than fifty years, the Strong CP problem still remains one of most puzzling questions in physics beyond the Standard Model. The theoretical effort put forth in trying to overcome this challenge led to a vast landscape of models tackling the problem from different angles, and it is still a source of inspiration for many new and original approaches. While most scenarios have already started to be constrained by current experiments, only future observations will be able to tell us which is the correct path to follow and, maybe, help to shed light on one of the greatest puzzles in theoretical particle physics.

Conventions

The metric in Minkowski space is the mostly-minus one

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}.$$

The Levi-Civita tensor $\epsilon^{\mu\nu\rho\sigma}$ is totally antisymmetric with $\epsilon^{0123} = +1$.

Given a compact Lie group G with Lie algebra $\mathfrak{g}$, the generators T^a_R in some representation R are finite-dimensional Hermitian matrices satisfying the algebra

$$\left[T^a_R, T^b_R\right] = i f^{abc} T^c_R$$

where f^{abc} are the structure constants of the group. In the adjoin representation $(T^a)^{bc} = -i f^{abc}$. The generators are chosen so that

$$\operatorname{tr} T^a_R T^b_R = T(R)\delta^{ab}$$

where $T(R)$ is the index of the representation. This satisfies the equality

$$T(R)d(G) = C_2(R)d(R)$$

where $d(G)$ is the dimension of the group (the number of generators), $d(R)$ is the dimension of the representation R and $C_2(R)$ is the quadratic Casimir, defined as

$$T^a_R T^a_R = C_2(R)\mathbb{1}_R.$$

In a non-abelian gauge theory, the gauge fields $A_\mu(x) \equiv A^a_\mu(x)T^a$ sit in the adjoin representation and transform under a gauge transformation $U(x) = e^{i\alpha(x)^a T^a}$ as

$$A_\mu \to U A_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1}.$$

where g is the gauge coupling. A field $\psi(x)$ in the representation R of G transforms as $\psi \to e^{i\alpha(x)\cdot T_R}\psi$, so that the covariant derivative, which satisfies $D_\mu(U\psi) = U D_\mu \psi$, is given by

$$D_\mu \psi = \left(\partial_\mu - i g A^a_\mu T^a_R\right)\psi.$$

The gauge field-strength can be defined in terms of the covariant derivative as

$$F_{\mu\nu} = \frac{i}{g}\left[D_\mu, D_\nu\right] = \partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right]$$

or in components

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$$

and transforms under a gauge transformation U as

$$F_{\mu\nu} \to U F_{\mu\nu} U^{-1}.$$

Acknowledgments

Il primo ringraziamento va a Luca, che in questi tre anni è stato molto più che un semplice supervisore. Il bagaglio di conoscenze che mi ha trasmesso è inestimabile ed è grazie a lui se ho capito cosa vuol dire fare ricerca, e in particolar modo fare Fisica, per davvero. Allo stesso modo sento di dover ringraziare tutto il personale del dipartimento ed i colleghi con cui ho avuto modo di interagire durante questi anni, che hanno contribuito a rendere questa esperienza estremamente piacevole e formativa. In particolar modo Paride, con il quale ho avuto confronti sempre estremamente illuminanti e utili per prendere le decisioni giuste.

I ringraziamenti più importanti, tuttavia, vanno alla mia famiglia. Senza il sostegno di papà Santino, mamma Laura e Michela questo percoso, culminato con questo dottorato, non sarebbe stato possibile. Il vostro supporto costante ed incondizionato è stato per me fondamentale e per questo ve ne sarò eternamente grato. Vorrei anche ringraziare anche tutti i miei nonni, zii, cugini e amici per il loro affetto. E Polly, che purtroppo non ha potuto essere presente alla conclusione di questo percorso ma mi ha accompagnato durante questa parte importante della mia vita.

Infine, un ringraziamento speciale va a Martina. La tua presenza e il tuo continuo e infallibile amore è stato ciò che ha reso questi anni speciali e significativi per me. Sono sicuro che anche in futuro, qualunque difficoltà incontreremo, il nostro rapporto rimarrà sempre speciale ed il nostro legame sarà sempre più forte.

Bibliography

[1] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi:10.1093/ptep/ptac097.

[2] Gerard 't Hooft. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B*, 59:135–157, 1980. doi:10.1007/978-1-4684-7571-5_9.

[3] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)]. arXiv:1807.06209, doi:10.1051/0004-6361/201833910.

[4] C. Jarlskog. Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation. *Phys. Rev. Lett.*, 55:1039, 1985. doi:10.1103/PhysRevLett.55.1039.

[5] C. Abel et al. Measurement of the Permanent Electric Dipole Moment of the Neutron. *Phys. Rev. Lett.*, 124(8):081803, 2020. arXiv:2001.11966, doi:10.1103/PhysRevLett.124.081803.

[6] Alessandro Valenti and Luca Vecchi. Perturbative running of the topological angles. *JHEP*, 01:131, 2023. arXiv:2210.09328, doi:10.1007/JHEP01(2023)131.

[7] Alessandro Valenti, Luca Vecchi, and Ling-Xiao Xu. Grand Color axion. *JHEP*, 10:025, 2022. arXiv:2206.04077, doi:10.1007/JHEP10(2022)025.

[8] Alessandro Valenti and Luca Vecchi. The CKM phase and $\overline{\theta}$ in Nelson-Barr models. *JHEP*, 07(203):203, 2021. arXiv:2105.09122, doi:10.1007/JHEP07(2021)203.

[9] Alessandro Valenti and Luca Vecchi. Super-soft CP violation. *JHEP*, 07(152):152, 2021. arXiv:2106.09108, doi:10.1007/JHEP07(2021)152.

[10] Sidney Coleman. *Aspects of Symmetry: Selected Erice Lectures*. Cambridge University Press, Cambridge, U.K., 1985. doi:10.1017/CBO9780511565045.

[11] L. H. Ryder. *QUANTUM FIELD THEORY*. Cambridge University Press, 6 1996. doi:10.1017/CBO9780511813900.

[12] Ta-Pei [0000-0002-1137-0969] Cheng and Ling-Fong [0000-0002-8035-3329] Li. *Gauge Theory of Elementary Particle Physics*. Oxford University Press, Oxford, UK, 1984.

[13] A. I. Vainshtein, Valentin I. Zakharov, V. A. Novikov, and Mikhail A. Shifman. ABC's of Instantons. *Sov. Phys. Usp.*, 25:195, 1982. doi:10.1070/PU1982v025n04ABEH004533.

[14] Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, 8 2013. doi:10.1017/CBO9781139644174.

[15] Thomas Schäfer and Edward V. Shuryak. Instantons in QCD. *Rev. Mod. Phys.*, 70:323–426, 1998. arXiv:hep-ph/9610451, doi:10.1103/RevModPhys.70.323.

[16] A. A. Belavin, Alexander M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin. Pseudoparticle Solutions of the Yang-Mills Equations. *Phys. Lett. B*, 59:85–87, 1975. doi:10.1016/0370-2693(75)90163-X.

[17] Curtis G. Callan, Jr., R. F. Dashen, and David J. Gross. The Structure of the Gauge Theory Vacuum. *Phys. Lett. B*, 63:334–340, 1976. doi:10.1016/0370-2693(76)90277-X.

[18] R. Jackiw and C. Rebbi. Vacuum Periodicity in a Yang-Mills Quantum Theory. *Phys. Rev. Lett.*, 37:172–175, 1976. doi:10.1103/PhysRevLett.37.172.

[19] R. Bott. An Application of Morse theory to the topology of Lie groups. *Bull. Soc. Math. Fr.*, 84:251–281, 1956.

[20] Curtis G. Callan, Jr., Roger F. Dashen, and David J. Gross. Toward a Theory of the Strong Interactions. *Phys. Rev. D*, 17:2717, 1978. doi:10.1103/PhysRevD.17.2717.

[21] Cumrun Vafa and Edward Witten. Parity Conservation in QCD. *Phys. Rev. Lett.*, 53:535, 1984. doi:10.1103/PhysRevLett.53.535.

[22] C. Vafa and Edward Witten. Restrictions on Symmetry Breaking in Vector-Like Gauge Theories. *Nucl. Phys. B*, 234:173–188, 1984. doi:10.1016/0550-3213(84)90230-X.

[23] Steven Weinberg. Phenomenological Lagrangians. *Physica A*, 96(1-2):327–340, 1979. doi:10.1016/0378-4371(79)90223-1.

[24] J. Gasser and H. Leutwyler. Chiral Perturbation Theory to One Loop. *Annals Phys.*, 158:142, 1984. doi:10.1016/0003-4916(84)90242-2.

[25] Sidney R. Coleman, J. Wess, and Bruno Zumino. Structure of phenomenological Lagrangians. 1. *Phys. Rev.*, 177:2239–2247, 1969. doi:10.1103/PhysRev.177.2239.

[26] Curtis G. Callan, Jr., Sidney R. Coleman, J. Wess, and Bruno Zumino. Structure of phenomenological Lagrangians. 2. *Phys. Rev.*, 177:2247–2250, 1969. doi:10.1103/PhysRev.177.2247.

[27] Murray Gell-Mann. The Eightfold Way: A Theory of strong interaction symmetry. 3 1961. doi:10.2172/4008239.

[28] Susumu Okubo. Note on unitary symmetry in strong interactions. *Prog. Theor. Phys.*, 27:949–966, 1962. doi:10.1143/PTP.27.949.

[29] Steven Weinberg. The U(1) Problem. *Phys. Rev. D*, 11:3583–3593, 1975. doi:10.1103/PhysRevD.11.3583.

[30] J. S. Bell and D. G. Sutherland. Current algebra and eta $\to 3\pi$. *Nucl. Phys. B*, 4:315–325, 1968. doi:10.1016/0550-3213(68)90316-7.

[31] Stephen L. Adler. Axial vector vertex in spinor electrodynamics. *Phys. Rev.*, 177:2426–2438, 1969. doi:10.1103/PhysRev.177.2426.

[32] Kazuo Fujikawa. Path Integral Measure for Gauge Invariant Fermion Theories. *Phys. Rev. Lett.*, 42:1195–1198, 1979. doi:10.1103/PhysRevLett.42.1195.

[33] Gerard 't Hooft. Symmetry Breaking Through Bell-Jackiw Anomalies. *Phys. Rev. Lett.*, 37:8–11, 1976. doi:10.1103/PhysRevLett.37.8.

[34] Gerard 't Hooft. Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle. *Phys. Rev. D*, 14:3432–3450, 1976. [Erratum: Phys.Rev.D 18, 2199 (1978)]. doi:10.1103/PhysRevD.14.3432.

[35] M. F. Atiyah and I. M. Singer. Dirac Operators Coupled to Vector Potentials. *Proc. Nat. Acad. Sci.*, 81:2597–2600, 1984. doi:10.1073/pnas.81.8.2597.

[36] Edward Witten. Current Algebra Theorems for the U(1) Goldstone Boson. *Nucl. Phys. B*, 156:269–283, 1979. doi:10.1016/0550-3213(79)90031-2.

[37] Antonio Pich and Eduardo de Rafael. Strong CP violation in an effective chiral Lagrangian approach. *Nucl. Phys. B*, 367:313–333, 1991. doi:10.1016/0550-3213(91)90019-T.

[38] R. J. Crewther, P. Di Vecchia, G. Veneziano, and Edward Witten. Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics. *Phys. Lett. B*, 88:123, 1979. [Erratum: Phys.Lett.B 91, 487 (1980)]. doi:10.1016/0370-2693(79)90128-X.

[39] Mikhail A. Shifman, A. I. Vainshtein, and Valentin I. Zakharov. Can Confinement Ensure Natural CP Invariance of Strong Interactions? *Nucl. Phys. B*, 166:493–506, 1980. doi:10.1016/0550-3213(80)90209-6.

[40] Roger F. Dashen. Some features of chiral symmetry breaking. *Phys. Rev. D*, 3:1879–1889, 1971. doi:10.1103/PhysRevD.3.1879.

[41] J. Nuyts. Is CP-invariance violation caused by an SU(3) singlet? *Phys. Rev. Lett.*, 26:1604–1605, 1971. [Erratum: Phys.Rev.Lett. 27, 361 (1971)]. doi:10.1103/PhysRevLett.26.1604.

[42] P. Di Vecchia and G. Veneziano. Chiral Dynamics in the Large n Limit. *Nucl. Phys. B*, 171:253–272, 1980. doi:10.1016/0550-3213(80)90370-3.

[43] Edward Witten. Large N Chiral Dynamics. *Annals Phys.*, 128:363, 1980. doi:10.1016/0003-4916(80)90325-5.

[44] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory.* Addison-Wesley, Reading, USA, 1995.

[45] Elizabeth Ellen Jenkins and Aneesh V. Manohar. Baryon chiral perturbation theory using a heavy fermion Lagrangian. *Phys. Lett. B*, 255:558–562, 1991. doi:10.1016/0370-2693(91)90266-S.

[46] T. H. R. Skyrme. A Nonlinear field theory. *Proc. Roy. Soc. Lond. A*, 260:127–138, 1961. doi:10.1098/rspa.1961.0018.

[47] T. H. R. Skyrme. A Unified Field Theory of Mesons and Baryons. *Nucl. Phys.*, 31:556–569, 1962. doi:10.1016/0029-5582(62)90775-7.

[48] Maxim Pospelov and Adam Ritz. Theta vacua, QCD sum rules, and the neutron electric dipole moment. *Nucl. Phys. B*, 573:177–200, 2000. arXiv:hep-ph/9908508, doi:10.1016/S0550-3213(99)00817-2.

[49] Lorenzo Bartolini, Francesco Bigazzi, Stefano Bolognesi, Aldo L. Cotrone, and Andrea Manenti. Neutron electric dipole moment from gauge/string duality. *Phys. Rev. Lett.*, 118(9):091601, 2017. arXiv:1609.09513, doi:10.1103/PhysRevLett.118.091601.

[50] M. Abramczyk, S. Aoki, T. Blum, T. Izubuchi, H. Ohki, and S. Syritsyn. Lattice calculation of electric dipole moments and form factors of the nucleon. *Phys. Rev. D*, 96(1):014501, 2017. arXiv:1701.07792, doi:10.1103/PhysRevD.96.014501.

[51] Jack Dragos, Thomas Luu, Andrea Shindler, Jordy de Vries, and Ahmed Yousif. Confirming the Existence of the strong CP Problem in Lattice QCD with the Gradient Flow. *Phys. Rev. C*, 103(1):015202, 2021. arXiv:1902.03254, doi:10.1103/PhysRevC.103.015202.

[52] I. B. Khriplovich. Quark Electric Dipole Moment and Induced θ Term in the Kobayashi-Maskawa Model. *Phys. Lett. B*, 173:193–196, 1986. doi:10.1016/0370-2693(86)90245-5.

[53] John R. Ellis and Mary K. Gaillard. Strong and Weak CP Violation. *Nucl. Phys. B*, 150:141–162, 1979. doi:10.1016/0550-3213(79)90297-9.

[54] I. B. Khriplovich and A. I. Vainshtein. Infinite renormalization of Theta term and Jarlskog invariant for CP violation. *Nucl. Phys. B*, 414:27–32, 1994. arXiv:hep-ph/9308334, doi:10.1016/0550-3213(94)90419-7.

[55] Elizabeth E. Jenkins, Aneesh V. Manohar, and Michael Trott. Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence. *JHEP*, 10:087, 2013. arXiv:1308.2627, doi:10.1007/JHEP10(2013)087.

[56] Gerard 't Hooft and M. J. G. Veltman. Regularization and Renormalization of Gauge Fields. *Nucl. Phys. B*, 44:189–213, 1972. doi:10.1016/0550-3213(72)90279-9.

[57] P. Breitenlohner and D. Maison. Dimensional Renormalization and the Action Principle. *Commun. Math. Phys.*, 52:11–38, 1977. doi:10.1007/BF01609069.

[58] Howard Georgi, Theodore Tomaras, and A. Pais. STRONG CP VIOLATION WITHOUT INSTANTONS. *Phys. Rev. D*, 23:469–472, 1981. doi:10.1103/PhysRevD.23.469.

[59] V. A. Novikov, Mikhail A. Shifman, A. I. Vainshtein, and Valentin I. Zakharov. Calculations in External Fields in Quantum Chromodynamics. Technical Review. *Fortsch. Phys.*, 32:585, 1984.

[60] Florent Baume, Boaz Keren-Zur, Riccardo Rattazzi, and Lorenzo Vitale. The local Callan-Symanzik equation: structure and applications. *JHEP*, 08:152, 2014. arXiv:1401.5983, doi:10.1007/JHEP08(2014)152.

[61] Boaz Keren-Zur. The local RG equation and chiral anomalies. *JHEP*, 09:011, 2014. arXiv:1406.0869, doi:10.1007/JHEP09(2014)011.

[62] D. Espriu and R. Tarrach. Renormalization of the Axial Anomaly Operators. *Z. Phys. C*, 16:77, 1982. doi:10.1007/BF01573750.

[63] G. M. Shore. New methods for the renormalization of composite operator Green functions. *Nucl. Phys. B*, 362:85–110, 1991. doi:10.1016/0550-3213(91)90557-E.

[64] S. A. Larin. The Renormalization of the axial anomaly in dimensional regularization. *Phys. Lett. B*, 303:113–118, 1993. arXiv:hep-ph/9302240, doi:10.1016/0370-2693(93)90053-K.

[65] Taushif Ahmed, Long Chen, and Michał Czakon. Renormalization of the flavor-singlet axial-vector current and its anomaly in dimensional regularization. *JHEP*, 05:087, 2021. arXiv:2101.09479, doi:10.1007/JHEP05(2021)087.

[66] Peter Breitenlohner, Dieter Maison, and K. S. Stelle. Anomalous Dimensions and the Adler-bardeen Theorem in Supersymmetric Yang-Mills Theories. *Phys. Lett. B*, 134:63–66, 1984. doi:10.1016/0370-2693(84)90985-7.

[67] Martin Lüscher and Peter Weisz. Renormalization of the topological charge density in QCD with dimensional regularization. *Eur. Phys. J. C*, 81(6):519, 2021. arXiv:2103.15440, doi:10.1140/epjc/s10052-021-09296-1.

[68] A. G. M. Pickering, J. A. Gracey, and D. R. T. Jones. Three loop gauge beta function for the most general single gauge coupling theory. *Phys. Lett. B*, 510:347–354, 2001. [Erratum: Phys.Lett.B 535, 377 (2002)]. arXiv:hep-ph/0104247, doi:10.1016/S0370-2693(01)00624-4.

[69] H. Osborn. Derivation of a Four-dimensional c Theorem. *Phys. Lett. B*, 222:97–102, 1989. doi:10.1016/0370-2693(89)90729-6.

[70] I. Jack and H. Osborn. Analogs for the c Theorem for Four-dimensional Renormalizable Field Theories. *Nucl. Phys. B*, 343:647–688, 1990. doi:10.1016/0550-3213(90)90584-Z.

[71] H. Osborn. Weyl consistency conditions and a local renormalization group equation for general renormalizable field theories. *Nucl. Phys. B*, 363:486–526, 1991. doi:10.1016/0550-3213(91)80030-P.

[72] I. Jack and H. Osborn. Constraints on RG Flow for Four Dimensional Quantum Field Theories. *Nucl. Phys. B*, 883:425–500, 2014. arXiv:1312.0428, doi:10.1016/j.nuclphysb.2014.03.018.

[73] C. Poole and A. E. Thomsen. Weyl Consistency Conditions and γ_5. *Phys. Rev. Lett.*, 123(4):041602, 2019. arXiv:1901.02749, doi:10.1103/PhysRevLett.123.041602.

[74] Joshua Davies, Florian Herren, and Anders Eller Thomsen. General gauge-Yukawa-quartic β-functions at 4-3-2-loop order. *JHEP*, 01:051, 2022. arXiv:2110.05496, doi:10.1007/JHEP01(2022)051.

[75] R. D. Peccei and Helen R. Quinn. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.*, 38:1440–1443, 1977. doi:10.1103/PhysRevLett.38.1440.

[76] R. D. Peccei and Helen R. Quinn. Constraints Imposed by CP Conservation in the Presence of Instantons. *Phys. Rev. D*, 16:1791–1797, 1977. doi:10.1103/PhysRevD.16.1791.

[77] Y. Aoki et al. FLAG Review 2021. *Eur. Phys. J. C*, 82(10):869, 2022. arXiv:2111.09849, doi:10.1140/epjc/s10052-022-10536-1.

[78] Howard Georgi and Ian N. McArthur. INSTANTONS AND THE mu QUARK MASS. 3 1981.

[79] David B. Kaplan and Aneesh V. Manohar. Current Mass Ratios of the Light Quarks. *Phys. Rev. Lett.*, 56:2004, 1986. doi:10.1103/PhysRevLett.56.2004.

[80] Kiwoon Choi, C. W. Kim, and W. K. Sze. Mass Renormalization by Instantons and the Strong CP Problem. *Phys. Rev. Lett.*, 61:794, 1988. doi:10.1103/PhysRevLett.61.794.

[81] Tom Banks, Yosef Nir, and Nathan Seiberg. Missing (up) mass, accidental anomalous symmetries, and the strong CP problem. In *2nd IFT Workshop on Yukawa Couplings and the Origins of Mass*, pages 26–41, 2 1994. arXiv:hep-ph/9403203.

[82] Constantia Alexandrou, Jacob Finkenrath, Lena Funcke, Karl Jansen, Bartosz Kostrzewa, Ferenc Pittler, and Carsten Urbach. Ruling Out the Massless Up-Quark Solution to the Strong ***CP*** Problem by Computing the Topological Mass Contribution with Lattice QCD. *Phys. Rev. Lett.*, 125(23):232001, 2020. arXiv:2002.07802, doi:10.1103/PhysRevLett.125.232001.

[83] Elizabeth E. Jenkins, Aneesh V. Manohar, and Peter Stoffer. Low-Energy Effective Field Theory below the Electroweak Scale: Operators and Matching. *JHEP*, 03:016, 2018. arXiv:1709.04486, doi:10.1007/JHEP03(2018)016.

[84] Luca Di Luzio, Maurizio Giannotti, Enrico Nardi, and Luca Visinelli. The landscape of QCD axion models. *Phys. Rept.*, 870:1–117, 2020. arXiv:2003.01100, doi:10.1016/j.physrep.2020.06.002.

[85] Jihn E. Kim. Weak Interaction Singlet and Strong CP Invariance. *Phys. Rev. Lett.*, 43:103, 1979. doi:10.1103/PhysRevLett.43.103.

[86] A. R. Zhitnitsky. On Possible Suppression of the Axion Hadron Interactions. (In Russian). *Sov. J. Nucl. Phys.*, 31:260, 1980.

[87] Michael Dine, Willy Fischler, and Mark Srednicki. A Simple Solution to the Strong CP Problem with a Harmless Axion. *Phys. Lett. B*, 104:199–202, 1981. doi:10.1016/0370-2693(81)90590-6.

[88] Frank Wilczek. Problem of Strong P and T Invariance in the Presence of Instantons. *Phys. Rev. Lett.*, 40:279–282, 1978. doi:10.1103/PhysRevLett.40.279.

[89] Frank Wilczek. Decays of Heavy Vector Mesons Into Higgs Particles. *Phys. Rev. Lett.*, 39:1304, 1977. doi:10.1103/PhysRevLett.39.1304.

[90] T. W. Donnelly, S. J. Freedman, R. S. Lytel, R. D. Peccei, and M. Schwartz. Do Axions Exist? *Phys. Rev. D*, 18:1607, 1978. doi:10.1103/PhysRevD.18.1607.

[91] Lawrence J. Hall and Mark B. Wise. FLAVOR CHANGING HIGGS - BOSON COUPLINGS. *Nucl. Phys. B*, 187:397–408, 1981. doi:10.1016/0550-3213(81)90469-7.

[92] William A. Bardeen, R. D. Peccei, and T. Yanagida. CONSTRAINTS ON VARIANT AXION MODELS. *Nucl. Phys. B*, 279:401–428, 1987. doi:10.1016/0550-3213(87)90003-4.

[93] Giovanni Grilli di Cortona, Edward Hardy, Javier Pardo Vega, and Giovanni Villadoro. The QCD axion, precisely. *JHEP*, 01:034, 2016. arXiv:1511.02867, doi:10.1007/JHEP01(2016)034.

[94] Ciaran O'Hare. cajohare/axionlimits: Axionlimits. https://cajohare.github.io/AxionLimits/, July 2020. doi:10.5281/zenodo.3932430.

[95] L. F. Abbott and P. Sikivie. A Cosmological Bound on the Invisible Axion. *Phys. Lett. B*, 120:133–136, 1983. doi:10.1016/0370-2693(83)90638-X.

[96] Michael Dine and Willy Fischler. The Not So Harmless Axion. *Phys. Lett. B*, 120:137–141, 1983. doi:10.1016/0370-2693(83)90639-1.

[97] John Preskill, Mark B. Wise, and Frank Wilczek. Cosmology of the Invisible Axion. *Phys. Lett. B*, 120:127–132, 1983. doi:10.1016/0370-2693(83)90637-8.

[98] Eran Palti. The Swampland: Introduction and Review. *Fortsch. Phys.*, 67(6):1900037, 2019. arXiv:1903.06239, doi:10.1002/prop.201900037.

[99] M. Gell-Mann. The interpretation of the new particles as displaced charge multiplets. *Nuovo Cim.*, 4(S2):848–866, 1956. doi:10.1007/BF02748000.

[100] Helge Kragh. Physics and the totalitarian principle. *arXiv preprint arXiv:1907.04623*, 2019.

[101] Howard M. Georgi, Lawrence J. Hall, and Mark B. Wise. Grand Unified Models With an Automatic Peccei-Quinn Symmetry. *Nucl. Phys. B*, 192:409–416, 1981. doi:10.1016/0550-3213(81)90433-8.

[102] Michael Dine and Nathan Seiberg. String Theory and the Strong CP Problem. *Nucl. Phys. B*, 273:109–124, 1986. doi:10.1016/0550-3213(86)90043-X.

[103] Stephen M. Barr and D. Seckel. Planck scale corrections to axion models. *Phys. Rev. D*, 46:539–549, 1992. doi:10.1103/PhysRevD.46.539.

[104] Marc Kamionkowski and John March-Russell. Planck scale physics and the Peccei-Quinn mechanism. *Phys. Lett. B*, 282:137–141, 1992. arXiv:hep-th/9202003, doi:10.1016/0370-2693(92)90492-M.

[105] Richard Holman, Stephen D. H. Hsu, Thomas W. Kephart, Edward W. Kolb, Richard Watkins, and Lawrence M. Widrow. Solutions to the strong CP problem in a world with gravity. *Phys. Lett. B*, 282:132–136, 1992. arXiv:hep-ph/9203206, doi:10.1016/0370-2693(92)90491-L.

[106] Peter Svrcek and Edward Witten. Axions In String Theory. *JHEP*, 06:051, 2006. arXiv:hep-th/0605206, doi:10.1088/1126-6708/2006/06/051.

[107] Steven B. Giddings and Andrew Strominger. Axion Induced Topology Change in Quantum Gravity and String Theory. *Nucl. Phys. B*, 306:890–907, 1988. doi:10.1016/0550-3213(88)90446-4.

[108] Ki-Myeong Lee. Wormholes and Goldstone Bosons. *Phys. Rev. Lett.*, 61:263–266, 1988. doi:10.1103/PhysRevLett.61.263.

[109] Luca Martucci, Nicolò Risso, Alessandro Valenti, and Luca Vecchi. In preparation. 2023.

[110] Mehmet Demirtas, Naomi Gendler, Cody Long, Liam McAllister, and Jakob Moritz. PQ axiverse. *JHEP*, 06:092, 2023. arXiv:2112.04503, doi:10.1007/JHEP06(2023)092.

[111] Alex G. Dias, V. Pleitez, and M. D. Tonasse. Naturally light invisible axion in models with large local discrete symmetries. *Phys. Rev. D*, 67:095008, 2003. arXiv:hep-ph/0211107, doi:10.1103/PhysRevD.67.095008.

[112] Linda M. Carpenter, Michael Dine, and Guido Festuccia. Dynamics of the Peccei Quinn Scale. *Phys. Rev. D*, 80:125017, 2009. arXiv:0906.1273, doi:10.1103/PhysRevD.80.125017.

[113] Keisuke Harigaya, Masahiro Ibe, Kai Schmitz, and Tsutomu T. Yanagida. Peccei-Quinn symmetry from a gauged discrete R symmetry. *Phys. Rev. D*, 88(7):075022, 2013. arXiv:1308.1227, doi:10.1103/PhysRevD.88.075022.

[114] A. G. Dias, A. C. B. Machado, C. C. Nishi, A. Ringwald, and P. Vaudrevange. The Quest for an Intermediate-Scale Accidental Axion and Further ALPs. *JHEP*, 06:037, 2014. arXiv:1403.5760, doi:10.1007/JHEP06(2014)037.

[115] Keisuke Harigaya, Masahiro Ibe, Kai Schmitz, and Tsutomu T. Yanagida. Peccei-Quinn Symmetry from Dynamical Supersymmetry Breaking. *Phys. Rev. D*, 92(7):075003, 2015. arXiv:1505.07388, doi:10.1103/PhysRevD.92.075003.

[116] Gongjun Choi and Tsutomu T. Yanagida. High quality axion in supersymmetric models. *JHEP*, 12:067, 2022. arXiv:2209.09290, doi:10.1007/JHEP12(2022)067.

[117] Lawrence M. Krauss and Frank Wilczek. Discrete Gauge Symmetry in Continuum Theories. *Phys. Rev. Lett.*, 62:1221, 1989. doi:10.1103/PhysRevLett.62.1221.

[118] Hajime Fukuda, Masahiro Ibe, Motoo Suzuki, and Tsutomu T. Yanagida. A "gauged" $U(1)$ Peccei–Quinn symmetry. *Phys. Lett. B*, 771:327–331, 2017. arXiv:1703.01112, doi:10.1016/j.physletb.2017.05.071.

[119] Michael Duerr, Kai Schmidt-Hoberg, and James Unwin. Protecting the Axion with Local Baryon Number. *Phys. Lett. B*, 780:553–556, 2018. arXiv:1712.01841, doi:10.1016/j.physletb.2018.03.054.

[120] Quentin Bonnefoy, Emilian Dudas, and Stefan Pokorski. Axions in a highly protected gauge symmetry model. *Eur. Phys. J. C*, 79(1):31, 2019. arXiv:1804.01112, doi:10.1140/epjc/s10052-018-6528-z.

[121] Luc Darmé and Enrico Nardi. Exact accidental U(1) symmetries for the axion. *Phys. Rev. D*, 104(5):055013, 2021. arXiv:2102.05055, doi:10.1103/PhysRevD.104.055013.

[122] Clifford Cheung. Axion Protection from Flavor. *JHEP*, 06:074, 2010. arXiv:1003.0941, doi:10.1007/JHEP06(2010)074.

[123] Luca Di Luzio, Enrico Nardi, and Lorenzo Ubaldi. Accidental Peccei-Quinn symmetry protected to arbitrary order. *Phys. Rev. Lett.*, 119(1):011801, 2017. arXiv:1704.01122, doi:10.1103/PhysRevLett.119.011801.

[124] Hye-Sung Lee and Wen Yin. Peccei-Quinn symmetry from a hidden gauge group structure. *Phys. Rev. D*, 99(1):015041, 2019. arXiv:1811.04039, doi:10.1103/PhysRevD.99.015041.

[125] Lisa Randall. Composite axion models and Planck scale physics. *Phys. Lett. B*, 284:77–80, 1992. doi:10.1016/0370-2693(92)91928-3.

[126] Bogdan A. Dobrescu. The Strong CP problem versus Planck scale physics. *Phys. Rev. D*, 55:5826–5833, 1997. arXiv:hep-ph/9609221, doi:10.1103/PhysRevD.55.5826.

[127] Michele Redi and Ryosuke Sato. Composite Accidental Axions. *JHEP*, 05:104, 2016. arXiv:1602.05427, doi:10.1007/JHEP05(2016)104.

[128] Benjamin Lillard and Tim M. P. Tait. A High Quality Composite Axion. *JHEP*, 11:199, 2018. arXiv:1811.03089, doi:10.1007/JHEP11(2018)199.

[129] M. B. Gavela, M. Ibe, P. Quilez, and T. T. Yanagida. Automatic Peccei–Quinn symmetry. *Eur. Phys. J. C*, 79(6):542, 2019. arXiv:1812.08174, doi:10.1140/epjc/s10052-019-7046-3.

[130] Marco Ardu, Luca Di Luzio, Giacomo Landini, Alessandro Strumia, Daniele Teresi, and Jin-Wei Wang. Axion quality from the (anti)symmetric of SU($\mathcal{N}$). *JHEP*, 11:090, 2020. arXiv:2007.12663, doi:10.1007/JHEP11(2020)090.

[131] Luca Vecchi. Axion quality straight from the GUT. *Eur. Phys. J. C*, 81(10):938, 2021. arXiv:2106.15224, doi:10.1140/epjc/s10052-021-09745-x.

[132] Roberto Contino, Alessandro Podo, and Filippo Revello. Chiral models of composite axions and accidental Peccei-Quinn symmetry. *JHEP*, 04:180, 2022. arXiv:2112.09635, doi:10.1007/JHEP04(2022)180.

[133] Bob Holdom and Michael E. Peskin. Raising the Axion Mass. *Nucl. Phys. B*, 208:397–412, 1982. doi:10.1016/0550-3213(82)90228-0.

[134] Ryuichiro Kitano and Wen Yin. Strong CP problem and axion dark matter with small instantons. *JHEP*, 07:078, 2021. arXiv:2103.08598, doi:10.1007/JHEP07(2021)078.

[135] Prateek Agrawal and Kiel Howe. Factoring the Strong CP Problem. *JHEP*, 12:029, 2018. arXiv:1710.04213, doi:10.1007/JHEP12(2018)029.

[136] V. A. Rubakov. Grand unification and heavy axion. *JETP Lett.*, 65:621–624, 1997. arXiv:hep-ph/9703409, doi:10.1134/1.567390.

[137] Zurab Berezhiani, Leonida Gianfagna, and Maurizio Giannotti. Strong CP problem and mirror world: The Weinberg-Wilczek axion revisited. *Phys. Lett. B*, 500:286–296, 2001. arXiv:hep-ph/0009290, doi:10.1016/S0370-2693(00)01392-7.

[138] Anson Hook. Anomalous solutions to the strong CP problem. *Phys. Rev. Lett.*, 114(14):141801, 2015. arXiv:1411.3325, doi:10.1103/PhysRevLett.114.141801.

[139] Hajime Fukuda, Keisuke Harigaya, Masahiro Ibe, and Tsutomu T. Yanagida. Model of visible QCD axion. *Phys. Rev. D*, 92(1):015021, 2015. arXiv:1504.06084, doi:10.1103/PhysRevD.92.015021.

[140] Savas Dimopoulos, Anson Hook, Junwu Huang, and Gustavo Marques-Tavares. A collider observable QCD axion. *JHEP*, 11:052, 2016. arXiv:1606.03097, doi:10.1007/JHEP11(2016)052.

[141] Anson Hook, Soubhik Kumar, Zhen Liu, and Raman Sundrum. High Quality QCD Axion and the LHC. *Phys. Rev. Lett.*, 124(22):221801, 2020. arXiv:1911.12364, doi:10.1103/PhysRevLett.124.221801.

[142] Savas Dimopoulos. A Solution of the Strong CP Problem in Models With Scalars. *Phys. Lett. B*, 84:435–439, 1979. doi:10.1016/0370-2693(79)91233-4.

[143] Tony Gherghetta, Natsumi Nagata, and Mikhail Shifman. A Visible QCD Axion from an Enlarged Color Group. *Phys. Rev. D*, 93(11):115010, 2016. arXiv:1604.01127, doi:10.1103/PhysRevD.93.115010.

[144] M. K. Gaillard, M. B. Gavela, R. Houtz, P. Quilez, and R. Del Rey. Color unified dynamical axion. *Eur. Phys. J. C*, 78(11):972, 2018. arXiv:1805.06465, doi:10.1140/epjc/s10052-018-6396-6.

[145] Edward Witten. An SU(2) Anomaly. *Phys. Lett. B*, 117:324–328, 1982. doi:10.1016/0370-2693(82)90728-6.

[146] Michael E. Peskin. The Alignment of the Vacuum in Theories of Technicolor. *Nucl. Phys. B*, 175:197–233, 1980. doi:10.1016/0550-3213(80)90051-6.

[147] John Preskill. Subgroup Alignment in Hypercolor Theories. *Nucl. Phys. B*, 177:21–59, 1981. doi:10.1016/0550-3213(81)90265-0.

[148] D. A. Kosower. SYMMETRY BREAKING PATTERNS IN PSEUDOREAL AND REAL GAUGE THEORIES. *Phys. Lett. B*, 144:215–216, 1984. doi:10.1016/0370-2693(84)91806-9.

[149] Alessandro Valenti and Luca Vecchi. Vector-like symmetries and parity conservation in gauge theories with Yukawa couplings. 6 2023. arXiv:2306.13088.

[150] Boaz Keren-Zur, Paolo Lodone, Marco Nardecchia, Duccio Pappadopulo, Riccardo Rattazzi, and Luca Vecchi. On Partial Compositeness and the CP asymmetry in charm decays. *Nucl. Phys. B*, 867:394–428, 2013. arXiv:1205.5803, doi:10.1016/j.nuclphysb.2012.10.012.

[151] Ravneet S. Bedi, Tony Gherghetta, and Maxim Pospelov. Enhanced EDMs from small instantons. *Phys. Rev. D*, 106(1):015030, 2022. arXiv:2205.07948, doi:10.1103/PhysRevD.106.015030.

[152] David I. Dunsky, Lawrence J. Hall, and Keisuke Harigaya. Dark Radiation Constraints on Heavy QCD Axions. 5 2022. arXiv:2205.11540.

[153] P. F. de Salas, M. Lattanzi, G. Mangano, G. Miele, S. Pastor, and O. Pisanti. Bounds on very low reheating scenarios after Planck. *Phys. Rev. D*, 92(12):123534, 2015. arXiv:1511.00672, doi:10.1103/PhysRevD.92.123534.

[154] Daniel Aloni, Yotam Soreq, and Mike Williams. Coupling QCD-Scale Axionlike Particles to Gluons. *Phys. Rev. Lett.*, 123(3):031803, 2019. arXiv:1811.03474, doi:10.1103/PhysRevLett.123.031803.

[155] Evgueni Goudzovski et al. New physics searches at kaon and hyperon factories. *Rept. Prog. Phys.*, 86(1):016201, 2023. arXiv:2201.07805, doi:10.1088/1361-6633/ac9cee.

[156] Kevin J. Kelly, Soubhik Kumar, and Zhen Liu. Heavy axion opportunities at the DUNE near detector. *Phys. Rev. D*, 103(9):095002, 2021. arXiv:2011.05995, doi:10.1103/PhysRevD.103.095002.

[157] Fatih Ertas and Felix Kahlhoefer. On the interplay between astrophysical and laboratory probes of MeV-scale axion-like particles. *JHEP*, 07:050, 2020. arXiv:2004.01193, doi:10.1007/JHEP07(2020)050.

[158] Sabyasachi Chakraborty, Manfred Kraus, Vazha Loladze, Takemichi Okui, and Kohsaku Tobioka. Heavy QCD axion in b→s transition: Enhanced limits and projections. *Phys. Rev. D*, 104(5):055036, 2021. arXiv:2102.04474, doi:10.1103/PhysRevD.104.055036.

[159] Emilie Bertholet, Sabyasachi Chakraborty, Vazha Loladze, Takemichi Okui, Abner Soffer, and Kohsaku Tobioka. Heavy QCD axion at Belle II: Displaced and prompt signals. *Phys. Rev. D*, 105(7):L071701, 2022. arXiv:2108.10331, doi:10.1103/PhysRevD.105.L071701.

[160] John Paul Chou, David Curtin, and H. J. Lubatti. New Detectors to Explore the Lifetime Frontier. *Phys. Lett. B*, 767:29–36, 2017. arXiv:1606.06298, doi:10.1016/j.physletb.2017.01.043.

[161] Nathaniel Craig, Anson Hook, and Skyler Kasko. The Photophobic ALP. *JHEP*, 09:028, 2018. arXiv:1805.06538, doi:10.1007/JHEP09(2018)028.

[162] Ki-woon Choi, David B. Kaplan, and Ann E. Nelson. Is CP a gauge symmetry? *Nucl. Phys. B*, 391:515–530, 1993. arXiv:hep-ph/9205202, doi:10.1016/0550-3213(93)90082-Z.

[163] Michael Dine, Robert G. Leigh, and Douglas A. MacIntire. Of CP and other gauge symmetries in string theory. *Phys. Rev. Lett.*, 69:2030–2032, 1992. arXiv:hep-th/9205011, doi:10.1103/PhysRevLett.69.2030.

[164] John Preskill and Lawrence M. Krauss. Local Discrete Symmetry and Quantum Mechanical Hair. *Nucl. Phys. B*, 341:50–100, 1990. doi:10.1016/0550-3213(90)90262-C.

[165] R. F. Streater and A. S. Wightman. *PCT, spin and statistics, and all that.* 1989.

[166] Stephen M. Barr, D. Chang, and G. Senjanovic. Strong CP problem and parity. *Phys. Rev. Lett.*, 67:2765–2768, 1991. doi:10.1103/PhysRevLett.67.2765.

[167] Yue Zhang, Haipeng An, Xiangdong Ji, and Rabindra N. Mohapatra. General CP Violation in Minimal Left-Right Symmetric Model and Constraints on the Right-Handed Scale. *Nucl. Phys. B*, 802:247–279, 2008. arXiv:0712.4218, doi:10.1016/j.nuclphysb.2008.05.019.

[168] Lawrence J. Hall and Keisuke Harigaya. Implications of Higgs Discovery for the Strong CP Problem and Unification. *JHEP*, 10:130, 2018. arXiv:1803.08119, doi:10.1007/JHEP10(2018)130.

[169] K. S. Babu and Rabindra N. Mohapatra. A Solution to the Strong CP Problem Without an Axion. *Phys. Rev. D*, 41:1286, 1990. doi:10.1103/PhysRevD.41.1286.

[170] Abdelhamid Albaid, Michael Dine, and Patrick Draper. Strong CP and SUZ_2. *JHEP*, 12:046, 2015. arXiv:1510.03392, doi:10.1007/JHEP12(2015)046.

[171] Raffaele Tito D'Agnolo and Anson Hook. Finding the Strong CP problem at the LHC. *Phys. Lett. B*, 762:421–425, 2016. arXiv:1507.00336, doi:10.1016/j.physletb.2016.09.061.

[172] Nathaniel Craig, Isabel Garcia Garcia, Giacomo Koszegi, and Amara McCune. P not PQ. *JHEP*, 09:130, 2021. arXiv:2012.13416, doi:10.1007/JHEP09(2021)130.

[173] Jordy de Vries, Patrick Draper, and Hiren H. Patel. Do Minimal Parity Solutions to the Strong CP Problem Work? 9 2021. arXiv:2109.01630.

[174] Valerio De Luca, Andrea Mitridate, Michele Redi, Juri Smirnov, and Alessandro Strumia. Colored Dark Matter. *Phys. Rev. D*, 97(11):115024, 2018. arXiv:1801.01135, doi:10.1103/PhysRevD.97.115024.

[175] David Dunsky, Lawrence J. Hall, and Keisuke Harigaya. Higgs Parity, Strong CP, and Dark Matter. *JHEP*, 07:016, 2019. arXiv:1902.07726, doi:10.1007/JHEP07(2019)016.

[176] Quentin Bonnefoy, Lawrence Hall, Claudio Andrea Manzari, and Christiane Scherb. A Colorful Mirror Solution to the Strong CP Problem. 3 2023. arXiv:2303.06156.

[177] Michele Redi and Andrea Tesi. Neutrinos, Dark Matter and Higgs Vacua in Parity Solutions of the strong CP problem. 7 2023. arXiv:2307.03161.

[178] Ann E. Nelson. Naturally Weak CP Violation. *Phys. Lett. B*, 136:387–391, 1984. doi:10.1016/0370-2693(84)92025-2.

[179] Stephen M. Barr. Solving the Strong CP Problem Without the Peccei-Quinn Symmetry. *Phys. Rev. Lett.*, 53:329, 1984. doi:10.1103/PhysRevLett.53.329.

[180] Michael Dine and Patrick Draper. Challenges for the Nelson-Barr Mechanism. *JHEP*, 08:132, 2015. arXiv:1506.05433, doi:10.1007/JHEP08(2015)132.

[181] Oz Davidi, Rick S. Gupta, Gilad Perez, Diego Redigolo, and Aviv Shalit. Nelson-Barr relaxion. *Phys. Rev. D*, 99(3):035014, 2019. arXiv:1711.00858, doi:10.1103/PhysRevD.99.035014.

[182] Sebastian Ohmer. Spontaneous CP Violation and the Strong CP Problem in Left-Right Symmetric Theories. *Phys. Rev. D*, 99(11):115031, 2019. arXiv:1808.02075, doi:10.1103/PhysRevD.99.115031.

[183] Jason Evans, Chengcheng Han, Tsutomu T. Yanagida, and Norimi Yokozaki. Complete solution to the strong CP problem: Supersymmetric extension of the Nelson-Barr model. *Phys. Rev. D*, 103(11):L111701, 2021. arXiv:2002.04204, doi:10.1103/PhysRevD.103.L111701.

[184] Jakob Schwichtenberg, Paul Tremper, and Robert Ziegler. A grand-unified Nelson–Barr model. *Eur. Phys. J. C*, 78(11):910, 2018. arXiv:1802.08109, doi:10.1140/epjc/s10052-018-6388-6.

[185] Gongjun Choi and Tsutomu T. Yanagida. Solving the strong CP problem with horizontal gauge symmetry. *Phys. Rev. D*, 100(9):095023, 2019. arXiv:1909.04317, doi:10.1103/PhysRevD.100.095023.

[186] Gilad Perez and Aviv Shalit. High quality Nelson-Barr solution to the strong CP problem with $\theta = \pi$. *JHEP*, 02:118, 2021. arXiv:2010.02891, doi:10.1007/JHEP02(2021)118.

[187] A. L. Cherchiglia and C. C. Nishi. Consequences of vector-like quarks of Nelson-Barr type. *JHEP*, 08:104, 2020. arXiv:2004.11318, doi:10.1007/JHEP08(2020)104.

[188] A. L. Cherchiglia, G. De Conto, and C. C. Nishi. Flavor constraints for a vector-like quark of Nelson-Barr type. *JHEP*, 11:093, 2021. arXiv:2103.04798, doi:10.1007/JHEP11(2021)093.

[189] Luca Vecchi. Spontaneous CP violation and the strong CP problem. *JHEP*, 04:149, 2017. arXiv:1412.3805, doi:10.1007/JHEP04(2017)149.

[190] Luis Bento, Gustavo C. Branco, and Paulo A. Parada. A Minimal model with natural suppression of strong CP violation. *Phys. Lett. B*, 267:95–99, 1991. doi:10.1016/0370-2693(91)90530-4.

[191] Lincoln Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51:1945, 1983. doi:10.1103/PhysRevLett.51.1945.

[192] Gudrun Hiller and Martin Schmaltz. Strong Weak CP Hierarchy from Nonrenormalization Theorems. *Phys. Rev. D*, 65:096009, 2002. arXiv:hep-ph/0201251, doi:10.1103/PhysRevD.65.096009.

[193] Albert M Sirunyan et al. A search for bottom-type, vector-like quark pair production in a fully hadronic final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 102:112004, 2020. arXiv:2008.09835, doi:10.1103/PhysRevD.102.112004.

[194] Roberto Contino, Margherita Ghezzi, Christophe Grojean, Margarete Muhlleitner, and Michael Spira. Effective Lagrangian for a light Higgs-like scalar. *JHEP*, 07:035, 2013. arXiv:1303.3876, doi:10.1007/JHEP07(2013)035.

[195] Luca Silvestrini and Mauro Valli. Model-independent Bounds on the Standard Model Effective Theory from Flavour Physics. *Phys. Lett. B*, 799:135062, 2019. arXiv:1812.10913, doi:10.1016/j.physletb.2019.135062.

[196] Jason Aebischer, Wolfgang Altmannshofer, Diego Guadagnoli, Méril Reboud, Peter Stangl, and David M. Straub. B-decay discrepancies after Moriond 2019. *Eur. Phys. J. C*, 80(3):252, 2020. arXiv:1903.10434, doi:10.1140/epjc/s10052-020-7817-x.

[197] M. Bauer, S. Casagrande, U. Haisch, and M. Neubert. Flavor Physics in the Randall-Sundrum Model: II. Tree-Level Weak-Interaction Processes. *JHEP*, 09:017, 2010. arXiv:0912.1625, doi:10.1007/JHEP09(2010)017.

[198] Morad Aaboud et al. Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector. *JHEP*, 07:089, 2018. arXiv:1803.09678, doi:10.1007/JHEP07(2018)089.

[199] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, R. Kogler, K. Moenig, M. Schott, and J. Stelzer. The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC. *Eur. Phys. J. C*, 72:2205, 2012. arXiv:1209.2716, doi:10.1140/epjc/s10052-012-2205-9.

[200] Pere Arnan, Lars Hofer, Federico Mescia, and Andreas Crivellin. Loop effects of heavy new scalars and fermions in $b \to s\mu^+\mu^-$. *JHEP*, 04:043, 2017. arXiv:1608.07832, doi:10.1007/JHEP04(2017)043.

[201] Gudrun Hiller and Martin Schmaltz. Solving the Strong CP Problem with Supersymmetry. *Phys. Lett. B*, 514:263–268, 2001. arXiv:hep-ph/0105254, doi:10.1016/S0370-2693(01)00814-0.

[202] Stuart Raby, Savas Dimopoulos, and Leonard Susskind. Tumbling Gauge Theories. *Nucl. Phys. B*, 169:373–383, 1980. doi:10.1016/0550-3213(80)90093-0.

[203] Apostolos Pilaftsis and Thomas E. J. Underwood. Resonant leptogenesis. *Nucl. Phys. B*, 692:303–345, 2004. arXiv:hep-ph/0309342, doi:10.1016/j.nuclphysb.2004.05.029.