



Physics Department – PhD in Theoretical Particle Physics

SISSA, Trieste

**Composite Axion Models
and
Q-Ball Superradiance**

Supervisor:
Aleksandr Azatov

Candidate:
Mohamed Mahdi

ACADEMIC YEAR 2025–2026

Abstract

The Standard Model is the most successful theory in physics, yet it leaves key questions open. One question is the strong CP problem: the strong interaction conserves parity and time reversal to better than 10^{-10} , a smallness that finds no explanation within the Standard Model. The Peccei–Quinn mechanism resolves this dynamically through a Goldstone boson called the axion; however, a realistic axion model must overcome the hierarchy problem of the Peccei–Quinn scale, the axion quality problem, and the cosmological domain-wall problem, while the heavy relics predicted by such models must not overclose the universe. Other questions such as dark matter and the baryon asymmetry also require solutions beyond the Standard Model. Some extensions of the Standard Model that attempt to address dark matter or the baryon asymmetry have rich scalar sectors that can generically support non-topological solitons. These solitons exchange charge and energy with their surrounding environment, so understanding their interaction is essential. This thesis takes up these two topics separately.

In the first part of this thesis, we construct explicit composite axion models, in which the Peccei–Quinn scale arises from a new confining interaction rather than an elementary scalar. We derive a general greatest-common-divisor criterion for computing the domain-wall number, and then construct two models with domain-wall number one in which the string-wall network self-annihilates: a deconstructed pair of chiral $SU(5)$ sectors, whose assumed confinement pattern we critically examine, and two vector-like sectors sharing a gauged Abelian symmetry (an anomaly-free direction orthogonal to the anomalous Peccei–Quinn direction). The leading operators that can explicitly break the Peccei–Quinn symmetry of both models appear at dimension nine, high enough to leave the strong CP solution intact. Both models predict stable colored relics, so we study the phenomenology of a short secondary inflationary epoch that can dilute them, leaving viable parameter space in which the string network re-enters the horizon near the QCD transition. This re-entry timing preserves the standard post-inflationary axion cosmology.

The second part of this thesis is independent of the first, and it concerns super-radiant scattering from Q-balls; in particular, it determines when linear scattering theory can be trusted. The charge and energy transferred to the Q-ball are fixed by unitarity in terms of a single particle–antiparticle conversion probability, and we show that the energy transferred coincides exactly with the energy cost μQ of the transferred charge. Nonlinear lattice simulations confirm the linear description for incident amplitudes below $\sim 10^{-2}$ of the soliton wall value. We also extend the linear formalism to the two-field Friedberg–Lee–Sirlin soliton, where the additional neutral field opens further conversion channels.

Together, these studies provide composite axion models that remain viable in a post-inflationary cosmology, and a quantitative amplitude threshold below which perturbative treatments of soliton scattering can be trusted.

Acknowledgments

I would like to begin by thanking my supervisor, Aleksandr Azatov, for his guidance throughout my PhD. I am grateful for his patience, his encouragement, and the many discussions that shaped how I think about physics and about this thesis in particular. His suggestions and comments were essential to bringing this work to completion, and I could not have asked for a better supervisor to learn from.

I would also like to thank all the professors of the Theoretical Particle Physics group at SISSA, and of the High Energy, Cosmology and Astroparticle Physics group at ICTP, for the environment they built and for what I learned from them over these years. I am grateful for the joint ICTP/SISSA PhD program itself, which gave me the chance to be part of both institutions.

I owe special thanks to Motoo Suzuki, whose collaboration on our paper, and later on parts of this thesis, was a real pleasure. His guidance and our discussions together taught me a great deal, and I am glad to have had the opportunity to work with him. I am equally grateful to my colleague Quoc-Trung Ho, with whom I shared these years of study, first at ICTP and then at SISSA. Our long discussions on our first paper, as well as his support throughout these years, made this journey a much better one.

Thanks are also due to my colleagues at SISSA, and in particular everyone who took part in our journal club. The discussions we had there did a great deal to shape my enthusiasm for this field.

To my friends from my undergraduate years, thank you for staying close despite the distance. Your calls and our conversations about life, work, and everything in between kept me grounded throughout this PhD. My particular thanks go to Banna, Shatla, Nafrawy, and Yasser.

Finally, I am deeply grateful to my parents and to my brother and sisters, for their constant love and support through all the years of my postgraduate studies. None of this would have been possible without them.

List of Publications

This thesis is based on the following publications, listed in chronological order:

- A. Azatov, Q. T. Ho and M. M. Khalil, *Q-ball perturbations with more details: Linear analysis vs lattice*, [Phys. Rev. D **111** \(2025\) 096010](#) [[arXiv:2412.13885](#)].
- A. Azatov, M. M. Khalil and M. Suzuki, *Towards a post-inflationary composite axion model*, [JHEP **03** \(2026\) 143](#) [[arXiv:2510.18538](#)].

Contents

Abstract	i
Acknowledgments	iii
List of Publications	v
I Post-Inflationary Composite Axion Models:	1
1 Introduction	3
1.1 The strong CP problem	3
1.2 The axion and its challenges	4
1.3 Objectives and Contributions	6
1.4 Outline of this part	7
2 A Review of Axion Physics	9
2.1 The vacuum structure of QCD	10
2.1.1 The $U(1)_A$ problem	10
2.1.2 Gauge vacua and the theta vacuum	11
2.1.3 Quarks and the physical parameter $\bar{\theta}$	15
2.1.4 The dilute instanton gas approximation	16
2.2 IR observables and the strong CP bound	18
2.2.1 The pion potential and the vacuum energy density	18
2.2.2 The neutron–proton mass splitting	19
2.2.3 The neutron electric dipole moment	21
2.2.4 The strong CP problem as a fine-tuning crisis	23
2.3 The Peccei–Quinn mechanism	23
2.3.1 Dynamical axion field	24
2.3.2 The original PQ model	25
2.3.3 The invisible axion models	26
2.3.4 Periodicity and the number of vacua	28

2.3.5	The quality of the Peccei–Quinn symmetry	29
2.4	Astrophysical bounds and detection	30
2.4.1	Couplings and production processes	31
2.4.2	Horizontal-branch stars: the photon coupling	32
2.4.3	Red giants and white dwarfs: the electron coupling	33
2.4.4	Neutron-star cooling: the nucleon coupling	33
2.4.5	Supernova 1987A	34
2.4.6	Laboratory detection	35
2.5	Axion cosmology	37
2.5.1	The axion potential in the early universe	38
2.5.2	Pre-inflationary axions: misalignment and isocurvature	40
2.5.3	Post-inflationary axions: the mode spectrum and vacuum re- alignment	42
2.5.4	Axion strings	43
2.5.5	Domain walls and the domain wall problem	44
2.6	Composite axion models	47
2.6.1	The Choi–Kim model	47
2.6.2	Forbidding the mass term: chiral composite axions	49
2.6.3	The Lazarides–Shafi mechanism	50
2.7	Summary	51
3	Composite Axion Models with Domain Wall Number One	53
3.1	A master formula for the domain wall number	53
3.1.1	Identification by a trivial rotation	54
3.1.2	Identification by an unbroken gauge symmetry	55
3.2	Automatic PQ symmetry	55
3.2.1	The original automatic model	56
3.2.2	A chiral moose realizing $N_{\text{DW}} = 1$ [1]	58
3.2.3	Critique: does the assumed confinement pattern actually occur?	60
3.3	Gauged PQ symmetry	62
3.3.1	The Barr–Seckel idea	62
3.3.2	Local versus global strings	63
3.3.3	A composite two-sector realization [1]	64
3.4	Heavy colored relics	68
3.4.1	Decay or dilution	68
3.4.2	Diluting the relic abundance	69
3.4.3	Axion string reentry	70
3.4.4	Axion abundance from the string-wall network	70

3.4.5	The viable parameter space	71
3.5	Summary	74
4	Conclusions	75
4.1	Summary of results	75
4.2	Limitations	76
II	Q-Ball Superradiance:	79
5	Introduction	81
5.1	Solitons in field theory	81
5.2	Q-balls in cosmological environment	82
5.3	Objectives and Contributions	82
5.4	Outline of this part	83
6	Non-topological Solitons and Their Stationary Solutions	85
6.1	The one-field Q-ball	85
6.1.1	Energy minimization at fixed charge	86
6.1.2	Stationary solutions	87
6.1.3	Existence window for the Q-ball frequency	88
6.1.4	Q-ball stability	89
6.1.5	Symmetries broken by the Q-balls	91
6.2	Q-balls of polynomial potentials	91
6.2.1	Numerical construction: the shooting method	92
6.3	Friedberg–Lee–Sirlin (FLS) soliton	93
6.3.1	Field equations and the two-field mechanical analogy	95
6.3.2	Numerical construction: multiple shooting	96
6.4	Summary	97
7	Scattering against Non-topological Solitons	99
7.1	Linear perturbations of a rotating background	100
7.1.1	The dichromatic ansatz and the two-channel equations	100
7.1.2	Mode classification	102
7.2	The two-channel S-matrix	103
7.2.1	Partial waves and the s-wave reduction	103
7.2.2	Construction of the S-matrix	104
7.2.3	Properties of the S-matrix	105
7.2.4	The $\lambda \rightarrow -\lambda$ exchange	106

7.3	Reflection and conversion probabilities	106
7.4	Charge and energy transfer	107
7.4.1	Kinematic interpretation	110
7.5	Lattice verification	110
7.6	Perturbations of the FLS soliton	114
7.6.1	Linearized equations and mode classification	114
7.6.2	Reflection and conversion probabilities	117
7.6.3	Charge and energy transfer	118
7.7	Summary	119
8	Conclusions	121
8.1	Summary of results	121
8.2	Limitations	122
A	The Shooting and Multiple Shooting Methods	125
A.1	Single shooting: the one-field Q-ball	125
A.2	Multiple shooting: the FLS soliton	127
A.2.1	Discretization and the augmented state	127
A.2.2	The Newton system and its block structure	129
A.2.3	Convergence of the Newton iteration	129
B	Partial-wave scattering off the Q-ball	131
B.1	The partial-wave S-matrix	131
B.2	From S_ℓ to a cross section	133
B.3	Numerical results	135
	Bibliography	149

Part I

Post-Inflationary

Composite Axion Models:

$N_{\text{DW}} = 1$ and the Quality Problem

Chapter 1

Introduction

1.1 The strong CP problem

One of the unresolved puzzles in the Standard Model (SM) concerns a parameter that experiments require to vanish to extraordinary precision [2]. The invariant parameter, built from the combination of the QCD vacuum angle θ and the phase of the quark mass matrix,

$$\bar{\theta} \equiv \theta + \arg \det M_q, \quad (1.1)$$

is a genuine physical parameter of the theory: it is P-odd and T-odd (hence CP-odd by CPT theorem), and it is not fixed by any symmetry of the SM to a particular value (as we establish in Section 2.1.3).

The parameter $\bar{\theta}$ gets contributions from two unrelated sectors; there is no physical reason for any cancellation between them, so the parameter should take an order-one value $\bar{\theta} \sim \mathcal{O}(1)$, and no symmetry of the Standard Model protects it against radiative corrections once generated, so a small value is not preserved automatically. The SM contains another CP-odd parameter, i.e., the CKM phase $\delta_{\text{CKM}} \approx 1.14$ rad, which is measured and of order one [3]. The most sensitive physical effect of the parameter $\bar{\theta}$ that can be measured is the electric dipole moment of the neutron (nEDM). Decades of experimental searches have yielded no observation of this effect $|d_n| \lesssim 10^{-26} e \text{ cm}$, resulting instead in the severe upper bound [2]

$$|\bar{\theta}| \lesssim 10^{-10}, \quad (1.2)$$

a limit that the next-generation n2EDM experiment aims to sharpen by a further order of magnitude $|d_n| \lesssim 10^{-27} e \text{ cm}$ [4].

A striking puzzle is posed here: nothing in the Standard Model favors a value of $\bar{\theta}$ so close to zero, yet experiments require it to vanish to better than one part in

10^{10} . The SM supplies no explanation of this cancellation; this *strong CP problem* is the starting point of Part I of this thesis, together with the models built around its most compelling solution, the axion. Beyond the axion, two other classes of solution have been proposed: spontaneously broken CP (Nelson–Barr models) and a massless up quark. We do not pursue either; the former since avoiding radiative regeneration of $\bar{\theta}$ there requires its own hierarchy of scales, while the latter, being strongly disfavored by lattice QCD results, will be mentioned briefly in Chapter 2.

1.2 The axion and its challenges

The most compelling solution to the strong CP problem is provided by the *Peccei–Quinn (PQ) mechanism* [5, 6]. The mechanism extends the SM to allow for a new spontaneously broken global symmetry, known as the PQ symmetry, that is anomalous with respect to QCD. The spontaneous symmetry breaking leaves behind a pseudo-Goldstone boson, the *axion* [7, 8], which promotes the θ -angle to a dynamical field. QCD effects generate a potential for this field whose minimum lies at $\bar{\theta} = 0$, dynamically solving the strong CP problem without requiring fine-tuning beyond the existence of the symmetry. For comprehensive modern reviews of axion physics and the PQ mechanism, see Refs. [9–11].

This elegant solution is, however, only as robust as the global symmetry it relies on. The PQ symmetry, as an ordinary global symmetry, is not expected to survive gravity unscathed. Quantum gravity is widely believed to respect no global symmetry at all, so Planck-suppressed operators that explicitly break the PQ symmetry are generically expected to appear in the low-energy effective theory [12–15]. This is hazardous for the solution of the strong CP problem: it tilts the axion potential away from the QCD minimum, and can reintroduce an effective $\bar{\theta}$ far above the bound just quoted. This is the *axion quality problem*¹, and it represents a major theoretical challenge in constructing realistic axion models.

Another challenge concerns the PQ scale itself. The PQ symmetry must be spontaneously broken; if the UV realization of the symmetry relies on an elementary scalar field, then the PQ scale inherits a hierarchy problem in the same technical sense as the electroweak scale [16]: the scalar mass is not protected against radiative corrections. The required tuning is, however, far milder in degree than the electroweak case. The sharper motivation for compositeness is therefore structural rather than a matter of degree: an elementary PQ scalar has no dynamical rea-

¹We present them here in the order most natural for motivating the axion; Chapter 3 addresses them in the order in which the chiral construction actually resolves them

son for the shape of its potential, whereas a confining sector generates one through dimensional transmutation, exactly as QCD generates f_π .

The severity of this hierarchy is tied to the phenomenologically required size of the PQ scale. Astrophysical and cosmological observations strongly constrain the allowed range of the PQ scale. The axion couplings are suppressed by inverse powers of the scale f_a , and axion emission from stellar environments can therefore alter stellar evolution. These astrophysical constraints, together with cosmological considerations, restrict the axion decay constant to the approximate window

$$10^8 \text{ GeV} \lesssim f_a \lesssim 10^{12} \text{ GeV}, \quad (1.3)$$

with the precise limits depending on the details of the axion model and cosmological history [10, 17, 18]. This high scale renders the axion extremely light and weakly interacting, giving rise to the class of models known as *invisible axion* models.

The large value of f_a also has important implications for the cosmological history of the axion field. Depending on whether PQ symmetry gets broken before or after inflation, the resulting axion abundance and topological defects can differ significantly. If the PQ breaking occurs before inflation, inflation homogenizes the axion field to one value $a/f_a = \theta_{\text{in}}$ (a free parameter of the scenario) over the observable Universe and any topological defects are diluted away. Conversely, if the PQ breaking occurs after inflation, different regions acquire independent axion field values, leading to the formation of axion strings, and the initial misalignment angle is effectively averaged over its distribution across causally disconnected patches, $\langle \theta_{\text{in}}^2 \rangle$ (thus the abundance becomes a prediction rather than a free parameter).

The post-inflationary scenario introduces an additional challenge through the formation of axion strings and domain walls. Axion strings form during the PQ phase transition, while domain walls appear later when QCD effects generate the axion potential and lift the continuous degeneracy of the axion direction to a discrete set of vacua [19, 20]. The number of physically distinct vacua is called the domain-wall number N_{DW} , and it determines the number of domain walls attached to each axion string. For $N_{\text{DW}} > 1$, the resulting string-wall network is stable (in the absence of explicit breaking) and eventually dominates the energy density of the Universe, in conflict with cosmological observations. Therefore, requiring

$$N_{\text{DW}} = 1 \quad (1.4)$$

constitutes another major challenge, *the domain-wall problem*, in constructing viable axion models, this time of cosmological origin.

These challenges motivate axion constructions in which the PQ symmetry emerges from a new strongly coupled sector. A composite axion provides a natural framework for addressing the hierarchy problem by generating the PQ scale dynamically through confinement, analogously to the generation of the pion decay constant f_π in QCD [21, 22]. In such models, the axion is not the phase of an elementary scalar field, but rather a composite pseudo-Nambu–Goldstone boson associated with the spontaneous breaking of an approximate global symmetry of the new strong sector. However, composite axion models generally introduce additional structure in the vacuum manifold, often leading to $N_{\text{DW}} > 1$. The Lazarides–Shafi mechanism provides a possible resolution by embedding the PQ symmetry into a larger gauge symmetry, such that vacua related by gauge transformations are physically equivalent and the effective domain-wall number is reduced below its naive value [23]. This mechanism therefore offers a route toward composite axion models that simultaneously address the hierarchy and domain-wall problems.

Finally, composite axion models introduce other challenges. The new confining sector generally contains additional composite states, some of which may be stable due to accidental symmetries. If these states are charged under QCD or otherwise interact strongly with the Standard Model (known as charged massive particles (CHAMPs)), they can survive as unwanted relics and lead to severe cosmological constraints, $\Omega_{\text{CHAMP}} h^2 / 0.12 \lesssim 10^{-10} - 10^{-8}$. Ensuring that these states can decay or are sufficiently diluted therefore represents another important requirement for realistic composite axion models [24].

1.3 Objectives and Contributions

The quality, hierarchy, and domain-wall problems are usually treated as separate difficulties, addressed, if at all, by separate machinery. Traditional composite axion models, built on vector-like confining gauge theories, often struggle to reconcile these requirements without additional model-building ingredients bolted on one problem at a time. The unifying idea pursued in this part of the thesis is that gauge invariance itself, rather than a global symmetry imposed by hand, is what protects the PQ symmetry and fixes the discrete vacuum structure responsible for the domain-wall problem. We realize this idea in two structurally distinct ways: in the first, the PQ symmetry arises as an accidental consequence of a confining sector whose matter content is chiral, so that chirality alone forbids any mass term for the confining fermions and, with it, any low-dimensional operator that could explicitly break the symmetry; in the second, we gauge the PQ symmetry directly, realizing it as the

anomaly-free Abelian direction orthogonal to an anomalous one, shared between two vector-like confining sectors, so that gauge invariance forbids explicit breaking outright rather than only up to some operator dimension.

This part of the thesis makes four constructive contributions, together with a fifth that is deliberately critical rather than positive. First, we derive a general criterion for the domain-wall number of a confining axion sector, $N_{\text{DW}} = \text{gcd}(k_{\text{PQ}}, k_{\mathcal{G}}, \dots)$, in terms of the PQ and gauge-anomaly coefficients of the embedding. Second, we use this criterion to construct a deconstructed pair of chiral $SU(5)$ sectors, a chiral extension of the automatic Peccei–Quinn mechanism of Ref. [25], with $N_{\text{DW}} = 1$, in which the string-wall network self-annihilates rather than persisting to threaten the cosmology. Third, we construct a composite realization of a gauged PQ symmetry built from two vector-like sectors, also with $N_{\text{DW}} = 1$, which to our knowledge is the first explicit ultraviolet completion of the gauged-PQ idea originally proposed by Barr and Seckel [26, 27]; in both constructions, the lowest-dimensional PQ-breaking operators are forbidden by gauge invariance, and the leading dangerous operators that survive appear only at dimension nine, comfortably high enough to leave the strong-CP solution intact against quantum-gravitational corrections. Fourth, we show that both models’ stable, colored composite relics can be diluted by a short secondary period of cosmic inflation, while still allowing the axion string-wall network to re-enter the horizon before the present, near the QCD transition; this re-entry timing is what preserves the standard post-inflationary axion cosmology and, with it, the predictive dark matter abundance that motivated the whole construction, subject to an $\mathcal{O}(1)$ modeling uncertainty in the radiated axion spectrum that we quantify in Section 3.4.5.

The fifth contribution is a critical re-examination, rather than a further construction. The chiral $SU(5)$ model of the second contribution rests on an assumed pattern of confinement and chiral symmetry breaking that we have not derived from first principles. In Section 3.2.3 we subject this assumption to scrutiny, and the outcome is only partly reassuring. We report this qualification explicitly, rather than leaving it to be discovered in the technical chapter, because a negative or qualified result about one’s own working assumptions is, in our view, as much a part of this thesis’ contribution as the positive constructions that precede it.

1.4 Outline of this part

The outline of Part I is organized as follows:

- Chapter 2 is an extended pedagogical review of the strong CP problem, the

Peccei–Quinn mechanism, and the axion physics. It also introduces the axion quality problem and the constraints on the PQ scale. After that, it discusses axion astrophysics and cosmology, leading to the domain-wall problem, before reviewing composite axion frameworks and the Lazarides–Shafi mechanism that motivate the constructions developed in the next chapter. A reader familiar with axion physics may proceed directly to Chapter 3.

- Chapter 3 is the technical core of this part. Concretely, this part of the thesis develops a general characterization of the domain-wall number in confining axion sectors, presents two explicit constructions realizing $N_{\text{DW}} = 1$ with parametrically high-quality PQ symmetries, and studies their cosmological viability as post-inflationary dark matter candidates.
- Chapter 4 summarizes the main results and discusses the principal open questions: the non-perturbative confinement dynamics on which the first construction rests, and the un-simulated string-population abundances entering the cosmological viability of the composite gauged model.

Chapter 2

A Review of Axion Physics

In this chapter we review the standard lore of the strong CP problem and its axion solution, providing the background needed for the remainder of Part I. Our aim is to give a clear and self-contained account of the established physics, organized around the key ideas that motivate the composite constructions developed later and the theoretical and cosmological challenges they are designed to address.

We begin with the foundational puzzle of the QCD vacuum. Section 2.1 explains how the parameter $\bar{\theta}$ arises from gauge-field topology and the chiral anomaly, while Section 2.2 works out its low-energy hadronic consequences through chiral perturbation theory, culminating in the neutron electric dipole moment bound. Section 2.3 then introduces the Peccei–Quinn mechanism and discusses the invisible-axion framework and the axion quality problem. This discussion follows the standard treatment of the subject, as reviewed for example in Refs. [9, 10, 28].

Having established the field-theoretic origin of the axion and its associated challenges, the chapter turns to the phenomenological and cosmological requirements that any realistic realization must satisfy. Section 2.4 collects the astrophysical and laboratory constraints on the axion decay constant f_a as reviewed in [18], while Section 2.5 examines the cosmological evolution of the axion field, the topological defects that exist in this theory and the resulting domain-wall problem [11, 29].

We then introduce in Section 2.6 the composite-axion framework, in which the PQ scale arises dynamically from a new confining interaction. Compositeness removes the need for an elementary PQ-breaking scalar, addressing the hierarchy problem, while chiral realizations provide a framework for understanding axion quality and domain-wall number through the underlying gauge structure. The Lazarides–Shafi mechanism is discussed in this context as a route toward avoiding the domain-wall problem and provides the conceptual foundation for the constructions developed in Chapter 3.

2.1 The vacuum structure of QCD

In this section we review how the structure of non-Abelian gauge theories introduces a CP-violating parameter into the strong interactions. We start from the historic $U(1)_A$ puzzle and its resolution through the chiral anomaly, which exposes the non-trivial topology of the QCD configuration space. That topology forces the definition of a continuous family of theta vacua, which in turn generates a new, physical CP-violating parameter whose consequences we must ultimately track down to low-energy hadronic scales.

2.1.1 The $U(1)_A$ problem

The Lagrangian of QCD, the gauge theory of the strong interactions, is deceptively simple to write down, yet it conceals one of the most subtle vacuum structures in quantum field theory. In the limit of N_f massless quark flavors, the QCD Lagrangian

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} + \sum_{j=1}^{N_f} \bar{q}_j i \not{D} q_j, \quad (2.1)$$

with the field strength $G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c$, possesses a large global symmetry $U(N_f)_L \times U(N_f)_R$, acting on left- and right-handed quarks independently,

$$q_{jL} \rightarrow (U_L)_{jk} q_{kL}, \quad q_{jR} \rightarrow (U_R)_{jk} q_{kR}. \quad (2.2)$$

For the two lightest flavors, $m_u, m_d \ll \Lambda_{\text{QCD}}$, this appears to be an excellent approximate symmetry of nature. The vector subgroups $SU(2)_I \times U(1)_B$ are manifest in the hadron spectrum, which is organized in isospin multiplets of well-defined baryon number. The axial subgroups $SU(2)_A \times U(1)_A$, however, do not show up in the spectrum: the quark condensate $\langle \bar{u}u \rangle = \langle \bar{d}d \rangle \neq 0$ breaks the axial subgroups spontaneously, so one expects four light pseudoscalar Nambu–Goldstone bosons. Three of them are the isotriplet pions, so the fate of the subgroup $SU(2)_A$ is understood. The fourth boson of this symmetry breaking was expected to be an isosinglet of mass comparable to m_π ($m_\eta \leq \sqrt{3}m_\pi$), yet the lightest isosinglet actually observed, i.e., the η meson, is far heavier than the pions, $m_\eta^2 \gg m_\pi^2$.¹ Weinberg [30] named this the $U(1)_A$ problem: naively $U(1)_A$ *should* be an approximate symmetry of the strong interactions, and yet nothing in the spectrum reflects it.

¹For $N_f = 3$ the same argument applies to the η' meson whose mass, 958 MeV, is even more dramatically at odds with the expectation.

The resolution, due to 't Hooft [31, 32], is that the naive power counting is simply wrong. The axial group $U(1)_A$ is not even an approximate symmetry of QCD, because the $U(1)_A$ axial current $J_A^\mu = \sum \bar{q}\gamma^\mu\gamma_5 q$, even in the chiral limit, suffers from the chiral anomaly [33–35]

$$\partial_\mu J_A^\mu = 2N_f \frac{g_s^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu}. \quad (2.3)$$

The right-hand side is a total derivative (as seen in Eq. (2.10)), so one may formally construct a conserved current by subtracting this Chern–Simons current from J_A^μ ; however, this current is not gauge invariant (as we will see in the next subsection). Therefore, it does not define a physical $U(1)_A$ symmetry acting on the gauge-invariant Hilbert space, and thus, Goldstone’s theorem implies the existence of a massless boson no longer. Because this resolution is the fundamental prerequisite for the strong CP problem, it merits a careful derivation, and it is to the topology of the gauge vacuum that we now turn.

2.1.2 Gauge vacua and the theta vacuum

The Yang–Mills term in the Lagrangian (2.1) can be written in terms of chromoelectric $E_i^a = G_{0i}^a$ and chromomagnetic fields $B_i^a = \frac{1}{2}\epsilon_{ijk}G_{jk}^a$,

$$-\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} = \frac{1}{2}E_i^a E_i^a - \frac{1}{2}B_i^a B_i^a = \frac{1}{2}\dot{A}_i^a \dot{A}_i^a - \frac{1}{2}B_i^a B_i^a, \quad (2.4)$$

where we have adopted the temporal gauge $A_0^a = 0$ in the last step. This gauge choice is preserved by any time-independent gauge transformation $g(\mathbf{x}) \in SU(3)$ as we can see from $A_0 \rightarrow A_0^g = g^{-1}(\mathbf{x})(A_0 + ig_s^{-1}\partial_0)g(\mathbf{x}) = 0$, where g_s is the QCD coupling. The magnetic term plays the role of a potential energy for the spatial gauge-field components A_i^a .

Since vacuum configurations minimize the energy, we restrict attention to finite-energy configurations, which require $B_i^a \rightarrow 0$ sufficiently rapidly at spatial infinity. The field space of finite-energy, time-independent configurations is $\mathcal{A} = \{A_i^a(\mathbf{x}) : B_i^a(\infty) = 0\}$. The physical configuration space is obtained from the general field space by quotienting out the time-independent gauge transformations $g(\mathbf{x})$ that preserve the finite-energy requirement, $g(\infty) = 1$.

The condition on the gauge transformation suggests a one-point compactification of physical space, $\mathbb{R}^3 \cup \{\infty\} \cong S^3$. Therefore, the relevant transformations are the maps $\mathcal{G} = \{g(\mathbf{x}) : S^3 \rightarrow SU(3)\}$. The field space $\mathcal{A}$ becomes a fiber bundle over

$\mathcal{A}/\mathcal{G}$ with fibers $\mathcal{G}$, and the associated long exact sequence reads

$$\pi_1(\mathcal{A}) \rightarrow \pi_1(\mathcal{A}/\mathcal{G}) \rightarrow \pi_0(\mathcal{G}) \rightarrow \pi_0(\mathcal{A}). \quad (2.5)$$

The field space $\mathcal{A}$, being an affine space, has trivial topology, $\pi_0(\mathcal{A}) = \pi_1(\mathcal{A}) = 0$. Consequently, the fundamental group of the configuration space is²

$$\pi_1(\mathcal{A}/\mathcal{G}) = \pi_0(\mathcal{G}) = \pi_3(SU(3)) = \mathbb{Z}, \quad (2.6)$$

so that maps from the compactified spatial slice into the gauge group fall into disconnected winding classes; this leads to large gauge transformations that cannot be continuously deformed to the identity, and they are classified by an integer winding number [36, 37].

Classical vacuum configurations minimize the magnetic energy and therefore satisfy $B_i^a = 0$ [38], so every such configuration is locally pure gauge

$$A_i^a = g^{-1}(\mathbf{x})(ig_s^{-1}\partial_i)g(\mathbf{x}). \quad (2.7)$$

Evaluated on a representative pure-gauge field from each homotopy class of $g(\mathbf{x}) : S^3 \rightarrow SU(3)$, the associated Chern–Simons functional takes integer values:

$$N_{\text{CS}}[A] = \frac{g_s^2}{32\pi^2} \int d^3\mathbf{x} \epsilon^{ijk} \left(A_i^a G_{jk}^a - \frac{g_s}{3} f^{abc} A_i^a A_j^b A_k^c \right). \quad (2.8)$$

These integers are not gauge-invariant observables: under a large gauge transformation of winding m , $N_{\text{CS}} \rightarrow N_{\text{CS}} + m$. Thus neighboring integer values correspond to different representatives related by large gauge transformations rather than physically distinct classical vacua. Consequently, only differences in Chern–Simons number carry gauge-invariant meaning.

Although the Chern–Simons number depends on the choice of gauge representative, its change between two time slices is gauge invariant. This change is given by the Pontryagin index

$$Q = \frac{g_s^2}{32\pi^2} \int d^4x G_{\mu\nu}^a \tilde{G}^{a\mu\nu} = N_{\text{CS}}(+\infty) - N_{\text{CS}}(-\infty), \quad (2.9)$$

which follows from the identity

$$G_{\mu\nu}^a \tilde{G}^{a\mu\nu} = \partial_\mu \epsilon^{\mu\nu\rho\sigma} \left(A_\nu^a G_{\rho\sigma}^a - \frac{g_s}{3} f^{abc} A_\nu^a A_\rho^b A_\sigma^c \right). \quad (2.10)$$

²Any simple compact group G has $\pi_3(G) = \mathbb{Z}$.

Although the classical vacuum representatives are separated by an energy barrier, Euclidean Yang–Mills theory admits finite-action solutions that interpolate between representatives whose Chern–Simons numbers differ by an integer. These solutions are the *instantons* [10, 39]. The simplest instanton, which carries unit Pontryagin index $Q = 1$, is known as the BPST instanton solution, and is given by

$$A_\mu(x) = \frac{(x - x_0)^2}{(x - x_0)^2 + \rho^2} g^{-1}(x) (i g_s^{-1} \partial_\mu) g(x), \quad g(x) = \bar{\sigma}^\mu (x - x_0)_\mu / |x - x_0|. \quad (2.11)$$

The family of instantons is described by a set of moduli: the center coordinate x_0 , the instanton size ρ , and the orientation moduli of the coset $SU(3)/H$ where H stabilizes the embedded $SU(2)$. These moduli act as collective coordinates, and they will be important when we work with instantons in Section 2.1.4.

We denote by $|n\rangle$ the semiclassical state built on the vacuum representative with Chern–Simons number n . These states fail to qualify as physical vacua on two counts: kinematically, a large gauge transformation maps one onto another, and dynamically, instantons induce tunneling between neighboring representatives. Physical vacuum states must therefore be eigenstates of the unitary operator Ω implementing a large gauge transformation of unit winding, whose action on the representatives is

$$\Omega |n\rangle = |n + 1\rangle. \quad (2.12)$$

In temporal gauge, Gauss’s law enforces strict invariance of physical states only under small gauge transformations; large transformations need only be represented by a phase. Since Ω commutes with every gauge-invariant operator, in particular the Hamiltonian, the true vacuum must diagonalize Ω along with H .

Consider a candidate vacuum $|\Psi\rangle = \sum_n c_n |n\rangle$ and demand that it be an Ω -eigenstate; unitarity fixes the eigenvalue to a phase $e^{i\theta}$,

$$\Omega |\Psi\rangle = \sum_n c_n |n + 1\rangle = \sum_m c_{m-1} |m\rangle \stackrel{!}{=} e^{i\theta} |\Psi\rangle = e^{i\theta} \sum_m c_m |m\rangle. \quad (2.13)$$

Matching coefficients of $|m\rangle$ gives the recursion $c_{m-1} = e^{i\theta} c_m$ for every m . This recursion has the unique solution $c_m = c_0 e^{-im\theta}$, fixed entirely by the single parameter θ up to the overall normalization c_0 . There is thus *one* family of Ω -eigenstates for each choice of $\theta \in [0, 2\pi)$, and setting $c_0 = 1$ gives the θ -vacuum [38, 40]

$$|\theta\rangle = \sum_n e^{-in\theta} |n\rangle, \quad \Omega |\theta\rangle = e^{i\theta} |\theta\rangle. \quad (2.14)$$

Because H commutes with Ω , the state $|\theta\rangle$ can simultaneously be chosen as an energy eigenstate. Moreover, no gauge-invariant operator has non-zero matrix elements between states of different eigenvalue $e^{i\theta}$, so distinct values of θ label distinct superselection sectors: physically inequivalent theories, each internally consistent, between which no local measurement or time evolution can ever induce a transition. The angle θ is therefore not a quantum number to be measured but a fixed parameter of the theory, exactly like a coupling constant, and one particular value describes the world we observe.

It remains to see explicitly how θ enters the path integral, since this is what ultimately produces a genuine new term in the effective Lagrangian rather than a mere bookkeeping device. Inserting (2.14) into the vacuum-to-vacuum transition amplitude over a large time interval t , and using (2.12), we find

$$\begin{aligned}\langle\theta|e^{-iHt}|\theta\rangle &= \sum_{m,n} e^{im\theta} e^{-in\theta} \langle m|e^{-iHt}|n\rangle \\ &= \sum_{\nu} e^{i\nu\theta} \sum_n \langle n+\nu|e^{-iHt}|n\rangle,\end{aligned}\tag{2.15}$$

where $\nu = m - n$ is the net change in winding number between the initial and final vacua. The amplitude from $|n\rangle$ to $|n+\nu\rangle$ becomes a path integral in which the boundary conditions force the field configurations to carry net topological charge exactly ν ,

$$\sum_n \langle n+\nu|e^{-iHt}|n\rangle = \int \mathcal{D}A_{\mu}^{(\nu)} e^{iS[A]}.\tag{2.16}$$

Substituting back into (2.15) yields the full path integral of the theory,

$$Z(\theta) = \sum_{\nu} e^{i\nu\theta} \int \mathcal{D}A_{\mu}^{(\nu)} e^{iS[A]}.\tag{2.17}$$

The θ -vacuum path integral is therefore identical to the ordinary QCD path integral with an extra term added directly to the Lagrangian,

$$\mathcal{L}_{\theta} = \theta \frac{g_s^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu}.$$

(2.18)

This is the central conclusion of the subsection. The existence of inequivalent θ -vacua implies that the most general gauge-invariant QCD Lagrangian contains the topological term with a new fundamental parameter θ that has no reason a priori to vanish. Because $G_{\mu\nu}^a \tilde{G}^{a\mu\nu}$ is a total derivative it does not alter the perturbative Feynman rules or the classical equations of motion; its effects are entirely non-perturbative as they exist only through configurations that do not fall off fast enough

at infinity, i.e., instantons. Its physical relevance in the full theory, however, depends on how the quarks respond to it, which is the question we take up next.

2.1.3 Quarks and the physical parameter $\bar{\theta}$

The analysis so far has been confined to the pure gauge sector. QCD, however, contains quarks, and it is crucial to establish how they interact with the topological vacuum, and in particular whether their presence can render the θ -term physically harmless or even unphysical.

The physical status of the vacuum angle hinges on the behavior of the theory under chiral transformations. Consider a global axial rotation of the quark fields,

$$q_j \rightarrow e^{i\alpha_j \gamma_5} q_j, \quad (2.19)$$

where $j = 1, \dots, N_f$ indexes the flavors. The classical, massless Dirac action is invariant under this rotation, but the quantum partition function is not. As was showed by Fujikawa [41], the fermionic path-integral measure $\mathcal{D}q \mathcal{D}\bar{q}$ transforms with a non-trivial anomalous Jacobian. For a uniform axial rotation across all flavors, $\alpha_j = \alpha/2N_f$, this Jacobian shifts the topological vacuum angle,

$$\theta \rightarrow \theta - \alpha. \quad (2.20)$$

The relation reveals at once that if even a single quark flavor were exactly massless, say a massless up quark $m_u = 0$, the corresponding axial rotation would be an exact, unbroken symmetry of the rest of the QCD Lagrangian. One could then choose $\alpha = \theta$ to absorb the topological term entirely into the fermion fields without altering any other physical parameter, rendering θ completely unphysical and dissolving the strong CP problem.

The quark mass matrix M_q explicitly breaks the axial symmetry; under the same chiral rotation (2.19) the mass term transforms as

$$\bar{q}_i M_q^{ij} q_j \rightarrow \bar{q}_i e^{i\alpha_i \gamma_5} M_q^{ij} e^{i\alpha_j \gamma_5} q_j, \quad (2.21)$$

and thus the argument of its determinant shifts,

$$\arg \det M_q \rightarrow \arg \det M_q + \alpha. \quad (2.22)$$

Comparing (2.20) and (2.22), neither θ nor $\arg \det M_q$ is invariant under a change of chiral basis, but their sum is. This allows us to define a single, basis-independent,

physically observable parameter,

$$\bar{\theta} \equiv \theta + \arg \det M_q. \quad (2.23)$$

Fermions therefore do not render the topological vacuum unphysical; they lock it to the flavor sector. Any non-zero $\bar{\theta}$ induces genuine P and T violation, and hence CP violation, in the hadronic sector, driven by the pseudotensor $\epsilon^{\mu\nu\rho\sigma}$ hidden inside $\tilde{G}^{a\mu\nu}$ in (2.18). The parameter $\bar{\theta}$ is thus the true physical quantity to be constrained by experiment, most sharply through the neutron electric dipole moment. Before we can extract that bound, however, we need a controlled way to compute how $\bar{\theta}$ affects the vacuum, and the natural first attempt is the semiclassical expansion around instantons.

2.1.4 The dilute instanton gas approximation

By an appropriate chiral rotation the invariant parameter $\bar{\theta}$ of (2.23) can be moved entirely into the topological gluon term (2.18). In this basis it is clear that $\bar{\theta}$ enters the quantum theory only through non-perturbative, instanton-mediated configurations, which generate a $\bar{\theta}$ -dependence of the vacuum energy and hence observable consequences. At weak coupling these effects can be evaluated in the dilute instanton gas approximation (DIGA), which models the vacuum as an ensemble of well-separated instantons and anti-instantons of topological charge $Q = \pm 1$.

In the semiclassical approximation of the path integral, the zero modes of the spectrum of fluctuation around instantons are traded for integrals over the collective coordinates of the solution: the center x_0 , the size ρ , and the color orientation. The integration over x_0 yields a constant factor of the spacetime volume, and the orientation average contributes also a constant factor of the group volume. Together with the one-loop determinant over the non-zero modes, these factors are absorbed into the normalization of the *instanton size distribution* $D(\rho)$, which is defined such that the mean number of instantons with size between ρ and $\rho + d\rho$ in a four-volume V_4 is $dN = V_4 D(\rho) d\rho$. At one loop [31, 32], we have

$$D(\rho) = \frac{1}{\rho^5} (\Lambda\rho)^{b_0} \prod_{j=1}^{N_f} (m_j\rho), \quad b_0 = 11 - \frac{2}{3}N_f, \quad (2.24)$$

where the factor ρ^{-5} is fixed by dimensional analysis, and b_0 is the one-loop contribution to the QCD beta function, $8\pi^2/g_s^2(\mu) = b_0 \log(\mu/\Lambda)$ so that the classical action $S_E = 8\pi^2 Q/g_s^2$ gives the contribution $e^{-8\pi^2/g_s^2(\rho^{-1})} \propto (\Lambda\rho)^{b_0}$ evaluated at the instanton scale $\mu = \rho^{-1}$. Each dynamical flavor supplies a factor $m_j\rho$, a direct conse-

quence of the Atiyah–Singer index theorem: in a unit-charge background the Dirac operator possesses an exact chiral zero mode for each flavor, and the Grassmann integral over that mode vanishes unless a chiral-symmetry-breaking mass insertion saturates it. Thus, a single massless quark would make $\bar{\theta}$ unobservable as we have seen.

Because dilute instantons are widely separated and weakly interacting, a configuration of n_+ instantons and n_- anti-instantons, each carrying the topological phase $e^{\pm i\bar{\theta}}$, can be treated as an ideal gas in the grand-canonical ensemble. Summing over all topological sectors, the partition function exponentiates [31, 32, 42–44],

$$Z(\bar{\theta}) = \sum_{n_+, n_- = 0}^{\infty} \frac{e^{i(n_+ - n_-)\bar{\theta}}}{n_+! n_-!} \left(V_4 \int_0^\infty d\rho D(\rho) \right)^{n_+ + n_-} = \exp \left(2V_4 \cos \bar{\theta} \int_0^\infty d\rho D(\rho) \right). \quad (2.25)$$

Using the correspondence $Z(\bar{\theta}) = e^{-V_4 E(\bar{\theta})}$ between the partition function and the Euclidean vacuum energy density,

$$E(\bar{\theta}) = 2 \left(1 - \cos \bar{\theta} \right) \int_0^\infty d\rho D(\rho). \quad (2.26)$$

This already captures the key features expected of the QCD vacuum: the energy density is 2π -periodic in $\bar{\theta}$, and the CP-preserving point $\bar{\theta} = 0$ is its absolute minimum. Moreover, if even a single flavor were massless $m_j \rightarrow 0$, $D(\rho)$ would vanish identically and the vacuum energy would be independent of $\bar{\theta}$, i.e., the strong CP problem would again disappear, in complete agreement with the operator-level argument of the previous subsection.

For any realistic theory with $b_0 > 0$, however, the size integral in Eq. (2.26) is severely infrared-divergent: the integrand grows as $\rho^{b_0 + N_f - 5}$ (for three light flavors as ρ^7). This is the *large-instanton problem* of zero-temperature QCD. Because the running coupling $g_s(\rho^{-1})$ grows without bound in the infrared ($\mu = \rho^{-1} \rightarrow 0$), the large configurations that dominate the integrand lie precisely where the weak-coupling, semiclassical saddle-point underlying DIGA breaks down. The dilute gas therefore provides the correct qualitative intuition for the $\bar{\theta}$ -dependence of the vacuum, but it cannot be used for a quantitative calculation of $E(\bar{\theta})$ at zero temperature [10].³ A genuinely controlled zero-temperature computation demands a different strategy, one that abandons weak-coupling expansions and works instead with the physical spectrum of the theory. That is the task of the next section.

³We return to this framework in Section 2.5.1, where DIGA becomes reliable once more: at high temperature the thermal bath of the early universe supplies a natural infrared cutoff via Debye screening, suppressing large instantons.

2.2 IR observables and the strong CP bound

To sidestep the infrared breakdown of the semiclassical expansion, the low-energy properties of QCD must be formulated in terms of the true physical degrees of freedom, the light hadrons. This is the domain of *chiral perturbation theory*, the low-energy effective field theory of QCD. By following the parameter $\bar{\theta}$ through the hadronic sector we can extract concrete, non-perturbative predictions. In this section we trace those consequences from the vacuum energy and meson masses through the baryon mass splittings to the neutron electric dipole moment, and so establish the empirical bound on the strong CP parameter.

2.2.1 The pion potential and the vacuum energy density

To study the low-energy implications of $\bar{\theta}$ in the hadronic sector, it is convenient to rotate it out of the gauge sector and into the flavor sector. A suitable flavor-singlet axial transformation of the light quarks removes the topological term from the action and moves $\bar{\theta}$ into the phase of the quark mass matrix, $M_q \rightarrow M_q e^{i\bar{\theta}/2} = M_q(\bar{\theta})$.⁴

In the low-energy regime of QCD the pseudo-Goldstone bosons of the spontaneously broken chiral $SU(2)_L \times SU(2)_R$ symmetry are collected in the unitary matrix $U = \exp(i\pi^a \tau_a / f_\pi)$, where τ_a are the Pauli matrices and $f_\pi \approx 92.4$ MeV is the pion decay constant. At leading order the effective Lagrangian for two light flavors is given by the nonlinear sigma model [46–50],

$$\mathcal{L}_{\chi\text{PT}} = \frac{f_\pi^2}{4} \text{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{m_\pi^2 f_\pi^2}{2 \text{Tr} M_q} \text{Tr}[U^\dagger M_q(\bar{\theta}) + \text{h.c.}], \quad (2.27)$$

where $m_\pi \approx 135$ MeV is the physical neutral pion mass at $\bar{\theta} = 0$ and the quark mass is given by

$$M_q = \bar{m}_q \mathbf{1} + \delta m_q \tau_3, \quad \bar{m}_q = \frac{m_u + m_d}{2}, \quad \delta m_q = \frac{m_u - m_d}{2}. \quad (2.28)$$

Restricting ourselves to the configuration along the neutral pion direction, we set $U = \exp(i\hat{\pi}_0 \tau_3)$ with $\hat{\pi}_0 \equiv \pi_0 / f_\pi$. Evaluating the trace in the potential term of (2.27) gives the static pion potential

$$V(\hat{\pi}_0; \bar{\theta}) = m_\pi^2 f_\pi^2 \left[1 - \cos(\hat{\pi}_0) \cos(\bar{\theta}/2) - \frac{\delta m_q}{\bar{m}_q} \sin(\hat{\pi}_0) \sin(\bar{\theta}/2) \right]. \quad (2.29)$$

⁴We have chose the rotation to be distributed equally over the two light flavors for convenience, $\alpha_u = \alpha_d = \bar{\theta}/2$ in the 2×2 block, so det picks up $e^{i\bar{\theta}}$ as required. Other choices like the one used in Ref. [45], distributes the phase in inverse proportion to the quark masses are equally as valid.

A non-zero $\bar{\theta}$ generates a term linear in the pion field, which destabilizes the naive vacuum at $\langle \pi_0 \rangle = 0$. Minimizing with respect to $\hat{\pi}_0$ fixes the $\bar{\theta}$ -dependent vacuum expectation value of the pion field,

$$\tan \langle \hat{\pi}_0 \rangle = \frac{\delta m_q}{\bar{m}_q} \tan(\bar{\theta}/2). \quad (2.30)$$

The physical vacuum thus realigns itself to minimize the explicit chiral breaking. Substituting this saddle-point configuration back into (2.29) isolates the genuine $\bar{\theta}$ -dependence of the vacuum energy density [45],

$$E(\bar{\theta}) \equiv V(\langle \hat{\pi}_0 \rangle; \bar{\theta}) = m_\pi^2 f_\pi^2 [1 - D(\bar{\theta})], \quad (2.31)$$

where we define

$$D(\bar{\theta}) = \left\{ \cos^2(\bar{\theta}/2) + \frac{\delta m_q^2}{\bar{m}_q^2} \sin^2(\bar{\theta}/2) \right\}^{1/2}. \quad (2.32)$$

This is the rigorous, non-perturbative counterpart of the semiclassical instanton result (2.26): a controlled, 2π -periodic function of $\bar{\theta}$, minimized at the CP-preserving value $\bar{\theta} = 0$. Expanding it to quadratic order will, in Section 2.3, directly supply the axion mass, so we record for later use this expansion around the origin,

$$E(\bar{\theta}) \simeq \frac{1}{2} m_\pi^2 f_\pi^2 \frac{m_u m_d}{(m_u + m_d)^2} \bar{\theta}^2, \quad \bar{\theta} \rightarrow 0. \quad (2.33)$$

The realignment also modifies the spectrum. For example, at the maximal value $\bar{\theta} = \pi$, taking the physical ratio $m_d \simeq 2m_u$ for illustration [3], the radical in Eq. (2.31) reduces to $1/3$, and the physical pion mass, when we expand the potential around the shifted minimum, is strongly suppressed to $m_\pi^2(\bar{\theta} = \pi) = m_\pi^2/3$. Such a rearrangement of the light spectrum does not stay confined to the mesons; because baryons couple to the same chiral order parameter, it also feeds into the nucleon sector, to which we now turn.

2.2.2 The neutron–proton mass splitting

To map $\bar{\theta}$ onto baryonic observables, we evaluate the nucleon sector with the pion fields set to their realigned vacuum value $\langle \hat{\pi}_0 \rangle$ from (2.30). In the chiral limit the nucleon field $N = (p, n)$ is described by the Lagrangian [51]

$$\mathcal{L}_{\text{Nuc}, 0} = \bar{N} \left(i \not{D} - m_N^{(0)} - g_A \gamma^\mu \gamma^5 u_\mu \right) N, \quad (2.34)$$

where the covariant derivative is defined as $D_\mu = \partial_\mu + \Gamma_\mu$. The chiral connection Γ_μ and the chiral vielbein u_μ are built from the square root of the chiral field, $\xi^2 = U$,

$$\Gamma_\mu = \frac{1}{2}(\xi^\dagger \partial_\mu \xi + \xi \partial_\mu \xi^\dagger), \quad u_\mu = \frac{1}{2i}(\xi^\dagger \partial_\mu \xi - \xi \partial_\mu \xi^\dagger). \quad (2.35)$$

The vector coupling of the connection is fixed by geometry to $g_V = 1$ while the axial coupling was measured to have the value $g_A \approx 1.27$ [3]. The mass $m_N^{(0)}$ is the chiral-limit nucleon mass from QCD effects. Away from the chiral limit, the nucleon Lagrangian allows for more terms that break the axial symmetry,

$$\mathcal{L}_{\text{Nuc}} = \mathcal{L}_{\text{Nuc}, 0} - \frac{m_N^{(1)}}{4\bar{m}_q} \bar{N} \text{Tr}[\chi(\bar{\theta})] N - \frac{\Delta m_N}{4\delta m_q} \bar{N} \left(\chi(\bar{\theta}) - \frac{1}{2} \text{Tr}[\chi(\bar{\theta})] \right) N, \quad (2.36)$$

where $\chi(\bar{\theta}) \equiv \xi^\dagger M_q(\bar{\theta}) \xi + \text{h.c.}$ The quark masses lead to two contributions that break the axial symmetry: one is $m_N^{(1)}$ which preserves the isospin symmetry, and the other is Δm_N which breaks both isospin and axial symmetries. The nucleon masses under the full Lagrangian (2.36) become

$$m_p \equiv m_N^{(0)} + \frac{m_N^{(1)}}{4\bar{m}_q} \cdot \text{Tr} \langle \chi(\bar{\theta}) \rangle + \frac{\Delta m_N}{4\delta m_q} \cdot \left(\langle \chi_u(\bar{\theta}) \rangle - \frac{1}{2} \text{Tr}[\langle \chi(\bar{\theta}) \rangle] \right), \quad (2.37)$$

$$m_n \equiv m_N^{(0)} + \frac{m_N^{(1)}}{4\bar{m}_q} \cdot \text{Tr} \langle \chi(\bar{\theta}) \rangle + \frac{\Delta m_N}{4\delta m_q} \cdot \left(\langle \chi_d(\bar{\theta}) \rangle - \frac{1}{2} \text{Tr}[\langle \chi(\bar{\theta}) \rangle] \right), \quad (2.38)$$

where $\langle \chi(\bar{\theta}) \rangle$ is the value of $\chi(\bar{\theta})$ at the shifted vacuum (2.30),

$$\begin{aligned} \langle \chi(\bar{\theta}) \rangle &= 2M_q \cos \langle \hat{\pi}_0 \rangle \cos(\bar{\theta}/2) + 2M_q \tau_3 \sin \langle \hat{\pi}_0 \rangle \sin(\bar{\theta}/2) \\ &= \frac{1}{D(\bar{\theta})} \cdot \left[2M_q \cos^2(\bar{\theta}/2) + 2M_q \tau_3 \frac{\delta m_q}{\bar{m}_q} \sin^2(\bar{\theta}/2) \right]. \end{aligned} \quad (2.39)$$

The splitting at arbitrary $\bar{\theta}$ is [52]

$$m_p - m_n = \Delta m_N(\bar{\theta}) = \frac{\Delta m_N}{4\delta m_q} \cdot \text{Tr}[\langle \chi(\bar{\theta}) \rangle \tau_3] = \frac{\Delta m_N}{D(\bar{\theta})}, \quad (2.40)$$

where $\Delta m_N \sim m_u - m_d < 0$ is the standard QCD splitting. At $\bar{\theta} = \pi$ the splitting is enhanced, becoming proportional to the mass sum $(m_u + m_d)$ rather than the difference. For $m_d \simeq 2m_u$ this is roughly a threefold increase, and once the QED contribution is included the ratio $\Delta m_N(\pi)/\Delta m_N(0)$ grows even larger.

Because the phase space for neutron beta decay depends steeply on the splitting, with $\Gamma_n \propto Q^5$, $Q = -\Delta m_N - m_e$, even an $\mathcal{O}(1)$ value of $\bar{\theta}$ would drastically

accelerate neutron decay and disrupt *Big Bang nucleosynthesis*. Macroscopic nuclear observables thus already constrain $\bar{\theta}$, well before one invokes any dedicated CP-violating observable. The sharpest such observable, and the one that sets the actual bound, is the static electric dipole moment of the neutron, which we compute next.

2.2.3 The neutron electric dipole moment

While the mass splitting provides an implicit bound, the definitive constraint on $\bar{\theta}$ comes from a direct CP-violating observable: the static electric dipole moment of the neutron, d_n . We shall isolate the one-pion couplings to the nucleons from the Lagrangian (2.36), and extract d_n accordingly.

The nucleon can couple to one pion through a vertex generated from the axial coupling,

$$u_\mu = \frac{1}{2f_\pi} \partial_\mu \pi^a \tau_a + \mathcal{O}(\pi^2) \quad (2.41)$$

or the mass splitting,

$$\chi = \langle \chi \rangle + \frac{2A}{f_\pi} \pi^a \tau_a + \dots, \quad (2.42)$$

where

$$A = \bar{m}_q \cos \langle \hat{\pi}_0 \rangle \sin(\bar{\theta}/2) - \delta m_q \sin \langle \hat{\pi}_0 \rangle \cos(\bar{\theta}/2) = \frac{\bar{m}_q^2 - \delta m_q^2}{\bar{m}_q} \frac{\sin \bar{\theta}}{2D(\bar{\theta})}. \quad (2.43)$$

The Lagrangian of the pion-nucleon coupling is

$$\begin{aligned} \mathcal{L}_{\pi NN} &= -\frac{g_A}{2f_\pi} \partial_\mu \pi^a \bar{N} \gamma^\mu \gamma^5 \tau_a N - \frac{\Delta m_N}{4\delta m_q} \frac{2A}{f_\pi} \bar{N} \pi^a \tau_a N \\ &= g_{\pi NN} \bar{N} i \gamma^5 \pi^a \tau_a N + \bar{g}_{\pi NN} \bar{N} \pi^a \tau_a N. \end{aligned} \quad (2.44)$$

The CP-even pseudoscalar coupling $g_{\pi NN}$ is obtained by using integration by parts and employing the nucleon's equation of motion,

$$g_{\pi NN} = g_A m_N / f_\pi \approx 12.9. \quad (2.45)$$

The CP-odd scalar coupling $\bar{g}_{\pi NN}$ is given to leading order in $\bar{\theta}$ by

$$\bar{g}_{\pi NN} = \frac{\Delta m_N}{m_d - m_u} \frac{m_u m_d}{m_u + m_d} \frac{\bar{\theta}}{f_\pi} \approx 0.016 \bar{\theta}, \quad (2.46)$$

consistent with the range $|\bar{g}_{\pi NN}| \simeq (0.015\text{--}0.027) |\bar{\theta}|$ obtained from QCD sum rules and lattice determinations of $\Delta m_N(0)$ [53].

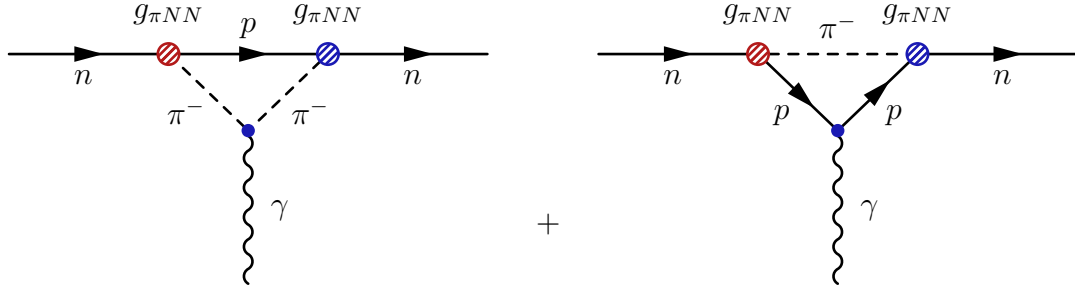


Figure 2.1: The one-loop vertex diagrams contributing to the neutron electric dipole moment via the CP-odd pion-nucleon vertex (red). In the left diagram, a virtual charged-pion loop dresses the nucleon propagator, and the external photon is radiated from the internal pion line. The right diagram has the photon vertex on the internal proton line.

The interplay of the two couplings in (2.44) generates an electric dipole form factor $F_3(q^2)$ in the electromagnetic vertex,

$$\langle N(p') | J_\mu^{\text{EM}} | N(p) \rangle \supset \bar{u}_n(p') \left[\frac{F_3(q^2)}{2m_N} i\gamma_5 \sigma_{\mu\nu} q^\nu \right] u_n(p), \quad (2.47)$$

with $d_n = F_3(0)/2m_N$. Evaluating the one-loop diagrams of Fig. 2.1 in the heavy-baryon expansion $m_\pi \ll m_N$ gives, at leading non-analytic order [52],

$$d_n = \frac{e g_{\pi NN} \bar{g}_{\pi NN}}{4\pi^2 m_N} \left[\log \frac{m_N}{m_\pi} + C \right], \quad (2.48)$$

where C is a scheme-dependent constant absorbing the local counterterm; only the chiral logarithm is universal. The logarithm signals a long-distance, pion-cloud-dominated contribution: d_n is controlled by low-energy pion loops rather than short-distance operators. Numerically, $d_n \sim 10^{-16} \bar{\theta} e \cdot \text{cm}$ up to an $\mathcal{O}(1)$ hadronic uncertainty, concretely, a factor of two to three due to the spread of sum-rule and lattice determination of $\bar{g}_{\pi NN}$ [10, 53]. The most recent experimental limit from the nEDM collaboration is [2]

$$|d_n| \lesssim 1.8 \times 10^{-26} e \cdot \text{cm}, \quad (2.49)$$

which, mapped onto the chiral-loop result, translates into an exceptionally stringent

bound on the fundamental parameter,

$$\boxed{|\bar{\theta}| \lesssim 10^{-10}}. \quad (2.50)$$

This value is the primary boundary condition of the strong CP problem. It is worth pausing to ask why so small a number should be regarded as a problem at all, which is the subject of the final subsection.

2.2.4 The strong CP problem as a fine-tuning crisis

This extreme smallness of the invariant angle, $|\bar{\theta}| \lesssim 10^{-10}$, constitutes a genuine naturalness crisis within the Standard Model. As a parameter, $\bar{\theta} = \theta + \arg \det M_q$ receives contributions from two entirely decoupled sectors: the topological structure of the non-Abelian $SU(3)_C$ gauge vacuum, which fixes θ , and the electroweak Yukawa sector responsible for the complex quark mass matrices, which fixes the phase $\arg \det M_q$.

A strict CP symmetry acting on the whole Lagrangian would set $\bar{\theta} = 0$ at tree level, but CP is known to be badly broken in the electroweak sector: the CKM matrix carries an unsuppressed CP-violating phase, $\delta_{\text{CKM}} \approx 1.14$ rad [3], of order unity, which successfully describes all observed CP violation in the K - and B -meson systems. Attributing the smallness of $\bar{\theta}$ to an accidental cancellation between the QCD vacuum phase and the weak Yukawa phases, accurate to one part in ten billion, is a severe fine-tuning. Moreover, the alternative escape due to a massless up quark ($m_u = 0$), which would make $\bar{\theta}$ unphysical through an unbroken axial $U(1)_A$, is strongly disfavored by lattice determinations that establish $m_u \simeq 2.16$ MeV in $\overline{\text{MS}}$ at 2 GeV [54].

The absence of any structural explanation for $\bar{\theta} \rightarrow 0$ is what defines the strong CP problem, and it provides the direct motivation for a dynamical solution. The most successful such proposal, the Peccei–Quinn mechanism, is the subject of the next section, and it will make essential reuse of the vacuum energy (2.31) derived above.

2.3 The Peccei–Quinn mechanism

The strong CP problem is a problem precisely because $\bar{\theta}$ is a fixed input with no dynamics of its own. The Peccei–Quinn (PQ) resolution removes exactly this rigidity: it introduces a new field $a(x)$, the axion, which enters QCD only through the combination $\bar{\theta} + a/f_a$, so that the effective vacuum angle becomes a dynamical quan-

tity free to relax to the CP-conserving point. Since the $\bar{\theta}$ -dependence of the QCD vacuum energy has already been derived in Eq. (2.31), the axion potential follows immediately by replacing the constant vacuum angle with the dynamical field; no new strong-interaction input is required.

We first set out the mechanism in general terms, and then work through its concrete realizations: the ruled-out original model and the invisible DFSZ and KSVZ constructions, as reviewed in [10, 28]. We close the section by determining the periodicity of the axion potential and the number of its degenerate vacua, and by confronting the axion quality problem that motivates the composite models of Section 2.6.

2.3.1 Dynamical axion field

The core idea of the Peccei–Quinn mechanism is that the CP-violating parameter $\bar{\theta}$ can be compensated by a dynamical field through the introduction of a new global symmetry $U(1)_{\text{PQ}}$. The symmetry is spontaneously broken at a scale v_{PQ} , and the Nambu–Goldstone boson of this breaking is the axion $a(x)$, which shifts under the symmetry as $a(x) \rightarrow a(x) + \alpha v_{\text{PQ}}$. Crucially, $U(1)_{\text{PQ}}$ is anomalous under QCD, exactly as the axial rotations of Section 2.1.3. The low-energy effective theory must therefore reproduce this anomaly, which it does through the coupling already familiar from (2.18),

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{SM}} + \frac{1}{2} \partial_\mu a \partial^\mu a + \left(\bar{\theta} + N \frac{a}{v_{\text{PQ}}} \right) \frac{g_s^2}{32\pi^2} G_{\mu\nu}^b \tilde{G}^{b\mu\nu}, \quad (2.51)$$

where N is the model-dependent anomaly coefficient of the PQ symmetry, to be associated later with the domain wall number. Absorbing the anomaly coefficient into the definition of the physical axion decay constant, $f_a \equiv v_{\text{PQ}}/N$, the axion enters the theory only through the shifted angle $\bar{\theta} + a/f_a$. The UV field of the axion Φ can be coupled nontrivially to SM, and can thus allow for other types of interaction such as the model-dependent derivative couplings $\partial_\mu a \bar{f} \gamma^\mu \gamma_5 f$ and the photon coupling $a F \tilde{F}$. However, the axion interaction through $\bar{\theta} + a/f_a$ is universal.

The results obtained in Section 2.2 lead to understanding of the dynamics of the axion. The QCD vacuum energy depends on its parameters only through the angle multiplying $G\tilde{G}$, so the axion potential follows from the vacuum energy (2.31) by direct substitution,

$$\boxed{V(a) = E \left(\bar{\theta} + \frac{a}{f_a} \right)}. \quad (2.52)$$

Because E is minimized at vanishing argument, the axion relaxes to

$$V'(a) = 0 \implies \langle a \rangle = -\bar{\theta} f_a, \quad (2.53)$$

and the effective vacuum angle $\bar{\theta} + \langle a \rangle / f_a$ vanishes identically. CP conservation in the strong interactions is thus restored dynamically, for any initial value of $\bar{\theta}$ and without adjusting any parameter: the same vacuum energy that quantified the strong CP problem now solves it provided nothing else breaks the PQ symmetry.

Expanding $V(a)$ about its minimum, with the curvature recorded in (2.33), gives the axion mass

$$m_a^2 = \frac{E''(0)}{f_a^2} = \frac{m_\pi^2 f_\pi^2}{f_a^2} \frac{m_u m_d}{(m_u + m_d)^2}, \quad (2.54)$$

which is numerically $m_a \simeq 5.7 \mu\text{eV} (10^{12} \text{ GeV} / f_a)$ [45, 55] (this result is obtained from the NLO calculation rather than the LO calculation (2.54)). The entire low-energy phenomenology of the axion is therefore controlled by the single scale f_a , and the question becomes what physics determines it. The remainder of this section examines the possible answers, beginning with the most economical one.

2.3.2 The original PQ model

The original implementation introduces only electroweak-scale degrees of freedom, in the form of a second Higgs doublet [5–8]. With the single doublet of the Standard Model, the Yukawa couplings $Y_u \bar{q}_L \tilde{H} u_R$ and $Y_d \bar{q}_L H d_R$ involve the same field through $\tilde{H} = \epsilon H^*$, whose phase enters the two terms with opposite signs, so no assignment of PQ charges leaves both couplings invariant. Once u_R and d_R couple to two independent doublets H_u, H_d , the obstruction disappears, and the Yukawa sector is invariant under

$$q_L \rightarrow e^{i\alpha} q_L, \quad (u_R, d_R) \rightarrow e^{-i\alpha} (u_R, d_R), \quad (H_u, H_d) \rightarrow e^{2i\alpha} (H_u, H_d). \quad (2.55)$$

We parameterize the doublets about their vacuum expectation values v_u, v_d as

$$H_u = \frac{v_u}{\sqrt{2}} e^{ia_u/v_u} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad H_d = \frac{v_d}{\sqrt{2}} e^{-ia_d/v_d} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (2.56)$$

suppressing the radial modes. One combination of the phase fields a_u, a_d is eaten by the Z boson, while the orthogonal combination is the physical axion a . After symmetry breaking, the axion-dependent phases in the quark mass matrices can equivalently be removed by chiral rotations of the left-handed quark fields $u_L \rightarrow e^{-ia_u/v_u} u_L$, $d_L \rightarrow e^{ia_d/v_d} d_L$, applied to each of the N_g generations, which results in

the the anomalous shift of the QCD vacuum angle,

$$\bar{\theta} \rightarrow \bar{\theta} - N_g \frac{a_u}{v_u} + N_g \frac{a_d}{v_d} = \bar{\theta} - N_g \frac{a}{v_{\text{PQ}}}, \quad v_{\text{PQ}} = \frac{v_u v_d}{v_{\text{EW}}}, \quad (2.57)$$

where $v_{\text{EW}} = (v_u^2 + v_d^2)^{1/2} = 246 \text{ GeV}$.⁵ The anomaly coefficient is thus the number of quark generations, $N = N_g = 3$, and the decay constant $f_a = v_{\text{PQ}}/N_g$ is chained to the electroweak scale. The master formula (2.54) then fixes the mass,

$$m_a = N_g \frac{m_\pi f_\pi}{v_{\text{PQ}}} \frac{\sqrt{m_u m_d}}{m_u + m_d} \simeq \frac{140 \text{ keV}}{\sin 2\beta}, \quad (2.58)$$

with $\tan \beta = v_d/v_u$. Because this axion is built entirely from electroweak-scale fields, its couplings to ordinary matter are suppressed only by $1/v_{\text{EW}}$. Such a weak-scale axion is copiously produced in rare meson decays. Therefore, the predicted branching ratio $\text{BR}(K^+ \rightarrow \pi^+ + a) \sim 10^{-5}$ [56] was soon found to lie far above the experimental limit $\text{BR}(K^+ \rightarrow \pi^+ + \text{nothing}) < 3.8 \times 10^{-8}$ set at KEK [57].⁶ The stellar energy-loss, which we shall see in Section 2.4, excludes it independently [60]; more recent population-based neutron-star bounds push the allowed QCD-axion mass well below the meV scale [61], excluding it a fortiori. The original model is therefore experimentally ruled out, but as we discuss next, the mechanism itself survives.

2.3.3 The invisible axion models

What saves the PQ mechanism as a solution of the strong CP problem is the crucial observation that the field responsible for PQ breaking need not participate in electroweak symmetry breaking, allowing the two scales to be separated [62–65]. If the PQ-breaking scalar is a Standard-Model singlet, v_{PQ} can be pushed to very high values, and by (2.54) the axion becomes correspondingly light, weakly coupled, and long-lived. Such an axion is dubbed “invisible” in the precise sense of evading every then-existing search. Crucially, the relaxation mechanism and the mass formula (2.54) hold unchanged; only the value of f_a moves. We now discuss the two surviving archetypes in turn.

⁵The overall sign of the shift is a convention, absorbed into $a \rightarrow -a$; we conform to the sign of Eq. (2.51).

⁶This bound has been further tightened by recent experiments at NA62 to be at the 10^{-11} level [58, 59].

The DFSZ model

The Dine–Fischler–Srednicki–Zhitnitsky construction [62, 63] extends the two doublet structure of the original model by an SM-singlet scalar Φ with a large vacuum expectation value $v_\Phi \gg v_{\text{EW}}$, coupled to the doublets through the potential term $\Phi^2 H_u H_d$. Invariance fixes the PQ charges: $(q_L, u_R, d_R) \rightarrow (e^{i\alpha} q_L, e^{-i\alpha} u_R, e^{-i\alpha} d_R)$, $(H_u, H_d) \rightarrow e^{2i\alpha} (H_u, H_d)$, and $\Phi \rightarrow e^{-2i\alpha} \Phi$. The same coupling does double duty in the spectrum: of the three phase fields a_u, a_d, a_Φ , one combination is eaten by the Z boson, and the $\Phi^2 H_u H_d$ term gives mass to the remaining unwanted CP-odd combination rather than leaving an additional Nambu–Goldstone boson. Therefore, exactly one light state survives. For $v_\Phi \gg v_{\text{EW}}$ this state resides dominantly in the phase of the singlet, and the decay constant is set by the singlet scale, $f_a \simeq v_\Phi/N$ with $N = 6$. The distinctive feature of the DFSZ axion is that it nevertheless retains direct, renormalizable couplings to all Standard Model quarks and leptons; this is because the SM fermions acquire their masses from H_u, H_d , and the physical axion is still partly composed of a_u and a_d .

The KSVZ model

By contrast, another construction due to Kim, Shifman, Vainshtein, and Zakharov introduces no additional Higgs doublet [64, 65]. It requires a single SM-singlet complex scalar Φ and a heavy vector-like quark Q ,⁷

$$\mathcal{L} = |D_\mu \Phi|^2 - V(\Phi) - (Y_Q \bar{Q}_L \Phi Q_R + \text{h.c.}), \quad V(\Phi) = \lambda \left(|\Phi|^2 - \frac{v_\Phi^2}{2} \right)^2, \quad (2.59)$$

with $U(1)_{\text{PQ}}$ acting only on the new fields, $Q_L \rightarrow e^{i\alpha} Q_L$, $Q_R \rightarrow e^{-i\alpha} Q_R$, $\Phi \rightarrow e^{2i\alpha} \Phi$. Parameterizing $\Phi = \frac{1}{\sqrt{2}}(v_\Phi + \rho) e^{ia/v_\Phi}$, the axion is purely the phase of the singlet. The ordinary quarks carry no PQ charge, so the only PQ-charged fermion generating the axion coupling to QCD is the heavy quark, which shifts the vacuum angle by $\bar{\theta} \rightarrow \bar{\theta} + a/v_\Phi$. For a single heavy quark the anomaly coefficient is $N = 1$, so $f_a = v_\Phi$, and the master formula (2.54) gives

$$m_a = \frac{\sqrt{m_u m_d}}{m_u + m_d} \frac{m_\pi f_\pi}{f_a}. \quad (2.60)$$

This inverse proportionality to the PQ-breaking scale is the hallmark of every invisible-axion model. The scale f_a is now a free parameter, bounded only by the astrophysical and cosmological considerations of Sections 2.4–2.5, and the KSVZ axion can be made light and weakly coupled by raising f_a subject to its bounds,

⁷Additional heavy quarks only increase the anomaly coefficient of the PQ symmetry.

evading every laboratory and stellar-cooling constraint that excluded the original model. Having fixed the mass and couplings of each model, we now examine a global feature of the axion potential that all of them share, and that will prove decisive for cosmology: the number of degenerate minima.

2.3.4 Periodicity and the number of vacua

The axion potential inherits its periodicity from the QCD vacuum energy, which is 2π -periodic in the effective vacuum angle $\bar{\theta}$. The relevant question is therefore how many times this fundamental period is traversed as the axion ranges over its own field space. This number determines the multiplicity of the axion vacua.

Consider first the original Peccei–Quinn model. For a single quark generation, the anomalous shift (2.57) gives $\bar{\theta} \rightarrow \bar{\theta} - a/v_{\text{PQ}}$, so that the axion potential $V(a) = E(\bar{\theta} - a/v_{\text{PQ}})$ contains a single minimum over the fundamental interval $a/v_{\text{PQ}} \in [0, 2\pi)$. With three generations, however, each generation contributes identically to the anomaly, and the QCD periodicity is traversed three times as the axion covers the same field range. The potential therefore develops three degenerate minima, as shown in Figure 2.2, which correspond to representatives $\bar{\theta} = 0, 2\pi, 4\pi$. Although these values are physically equivalent because QCD depends only on $\bar{\theta}$ modulo 2π , they correspond to distinct points of the axion vacuum manifold.

More generally, the number of times the QCD vacuum energy is traversed is determined by the QCD anomaly coefficient of the PQ symmetry. This quantity is the *domain-wall number*,

$$N_{\text{DW}} = \mathcal{A}(U(1)_{\text{PQ}} \cdot SU(3)_c^2), \quad (2.61)$$

which counts the number of degenerate axion vacua. Equivalently, the QCD anomaly explicitly breaks the continuous PQ symmetry,

$$U(1)_{\text{PQ}} \rightarrow \mathbb{Z}_{N_{\text{DW}}},$$

and the N_{DW} vacua are precisely the points in axion field space related by this residual discrete symmetry.

The distinction between these vacua is invisible in local axion physics. For example, if we define the axion decay constant as $f_a = v_{\text{PQ}}/N_{\text{DW}}$, the field variable a/f_a spans N_{DW} 2π -periods. Local physics in a region of spacetime is built around one of these vacua, so it sees but one period $a/f_a \in [0, 2\pi)$ which is convenient for computing quantities such as the axion mass since we use only (2.54). This rescaling

did not eliminate the global multiplicity of the vacuum manifold, as it delegated it to a number of periods of a/f_a . Whether the N_{DW} vacua represent genuinely distinct physical states, or are identified by an additional symmetry, determines whether stable domain walls form. This question is the origin of the cosmological domain-wall problem and motivates the discussion of its resolution through mechanisms such as the Lazarides–Shafi mechanism in Section 2.6.

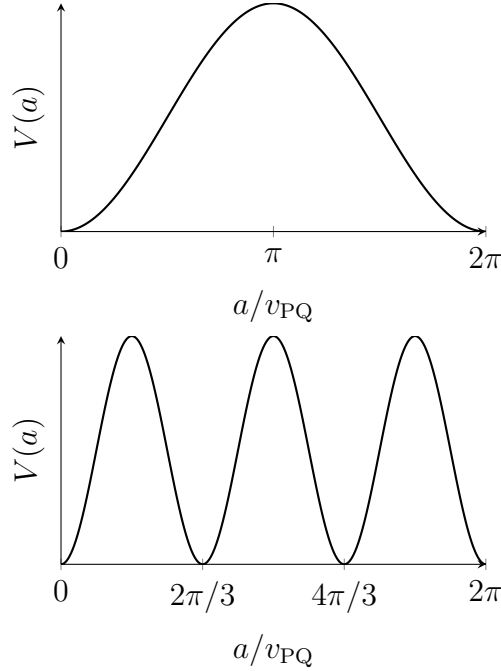


Figure 2.2: Axion potential of the original PQ model generated by one (top) vs three (bottom) quark generations. The vacuum is traversed one (three) times in the 2π period of a/v_{PQ} . Adjacent regions of spacetime that occupy different vacua will have a defect (domain wall) that interpolates between the vacua.

2.3.5 The quality of the Peccei–Quinn symmetry

The entire argument of Sections 2.3.1–2.3.4 rests on $U(1)_{\text{PQ}}$ being broken *only* by the QCD anomaly. Suppose instead that some additional physics, in particular physics associated with quantum gravity, explicitly breaks $U(1)_{\text{PQ}}$ through a non-renormalizable operator of mass dimension n like $\mathcal{O} = |\Phi|^{2k}(\Phi + \Phi^*)$, suppressed by the appropriate power of the Planck scale M_{Pl} . Such a term adds a second, generically misaligned contribution to the axion potential,

$$V(a) \sim \Lambda_{\text{QCD}}^4 f\left(\frac{a}{f_a}\right) + \frac{f_a^n}{M_{\text{Pl}}^{n-4}} g\left(\frac{a}{f_a}\right), \quad (2.62)$$

with f, g generic $\mathcal{O}(1)$, 2π -periodic functions whose minima have no reason to coincide.

Because the second term knows nothing of $\bar{\theta}$, it pulls the axion away from the CP-conserving point $\langle a \rangle = -\bar{\theta} f_a$ of (2.53), reintroducing an effective $\bar{\theta} + \langle a \rangle / f_a \neq 0$ and directly undoing the purpose of the mechanism. This is the single most important qualifier to attach to everything said so far: the PQ solution is only ever as good as the smallness of whatever residual explicit breaking of $U(1)_{\text{PQ}}$ exists. Demanding that the induced shift stay within the experimental bound $|\bar{\theta}| \lesssim 10^{-10}$ requires

$$\frac{f_a^{n-4}}{M_{\text{Pl}}^{n-4}} \frac{f_a^4}{\Lambda_{\text{QCD}}^4} \lesssim 10^{-10}. \quad (2.63)$$

For $f_a \sim 10^9$ GeV this forces the lowest PQ-violating operator to have dimension $n \geq 9$. This is the *axion quality problem*. The concern is not merely formal. It is widely believed, and sometimes elevated to the status of a folk theorem, that quantum gravity violates every global symmetry not protected by an unbroken gauge symmetry: virtual black holes and Euclidean wormholes can carry global charge into or out of a region with no compensating gauge flux to forbid it, so any global symmetry, however good an approximation at low energies, should be broken explicitly by Planck-suppressed operators of every dimension consistent with the unbroken *gauge* symmetries [12–14].

Since $U(1)_{\text{PQ}}$ is, in every model above, a global symmetry with no gauge structure protecting it, this is precisely the threat just quantified. Minimal KSVZ and DFSZ constructions provide no structural protection against such operators, and must simply assume, with no justification beyond convenience, that quantum gravity respects $U(1)_{\text{PQ}}$ far better than these general arguments suggest. Addressing this threat structurally, rather than by assumption, is one of our central motivations. Before specializing to any ultraviolet completion, however, we first collect the astrophysical and cosmological constraints that every invisible axion, composite or elementary, must satisfy.

2.4 Astrophysical bounds and detection

The axion of every invisible model of Sections 2.3–2.6 couples to the photon, the electron, and the nucleons with strengths suppressed by $1/f_a$ but never exactly zero: these are the couplings through which the QCD anomaly communicates with the axion, dressed by whatever electromagnetic and Yukawa charges the PQ-breaking sector happens to carry. This section is a study of those three couplings, probed from two complementary directions by the same elementary processes. In the hot plasma of a stellar interior the processes run forward, producing axions that stream

out of the core; the energy-loss argument of Ref. [18] then bounds each coupling, and hence f_a from below, by requiring that this drain does not shorten the observed lifetime of any stellar evolutionary phase. In the laboratory the same processes are run in reverse, converting axions back into photons inside strong magnetic fields, which is the only route to a discovery rather than an exclusion.

We first define the couplings and the two production processes from which every bound in this section follows. We then apply the energy-loss argument in environments of increasing density, each of which isolates one coupling: the photon coupling in horizontal-branch stars, the electron coupling in red giants and white dwarfs, and the nucleon coupling in neutron stars and in Supernova 1987A. We close with the laboratory program that inverts these processes. The complementary bound on the opposite side of the axion window is cosmological in origin and is deferred to Section 2.5.

2.4.1 Couplings and production processes

The two couplings of practical relevance are the axion–photon coupling, generated by the electromagnetic anomaly of $U(1)_{\text{PQ}}$,

$$\mathcal{L}_{a\gamma\gamma} = -\frac{g_{a\gamma\gamma}}{4} a F^{\mu\nu} \tilde{F}_{\mu\nu}, \quad g_{a\gamma\gamma} = \frac{\alpha}{2\pi f_a} \left(\frac{E}{N} - 1.92 \right), \quad (2.64)$$

where E/N is the ratio of the electromagnetic to color anomaly of the PQ symmetry and the number 1.92 encodes the model-independent contribution of axion–pion mixing [45], and the axion–fermion coupling,

$$\mathcal{L}_{aff} = \frac{C_f}{2f_a} \partial_\mu a \bar{f} \gamma^\mu \gamma_5 f, \quad (2.65)$$

where f runs over the electron and the nucleons, the two cases of astrophysical relevance, with coefficients C_e and C_N . These are present at tree level whenever the fermion carries PQ charge, as the electron does in DFSZ, and generated otherwise at higher order: in KSVZ the nucleon couplings are induced through axion–pion mixing and remain $\mathcal{O}(1)$, while the electron coupling arises only at one loop from the photon coupling (2.64), suppressed by α^2 and hence irrelevant for the bounds below.

Two production processes, built from these couplings, dominate in any stellar plasma, and both will reappear, inverted, in the detection techniques of Section 2.4.6.

The first is the *Primakoff process*,

$$\gamma + Ze \rightarrow a + Ze, \quad (2.66)$$

in which a thermal photon converts into an axion in the electrostatic field of a nucleus or electron; the external field supplies the second photon leg of the vertex (2.64), so the rate is controlled by $g_{a\gamma\gamma}$. The second is *bremssstrahlung*,

$$f + f' \rightarrow f + f' + a, \quad f \in \{e, N\}, \quad (2.67)$$

in which an axion is radiated from an electron or nucleon line during a Coulomb or nuclear collision, at a rate controlled by C_e or C_N . Both cross sections scale as $1/f_a^2$, so the energy-loss rate scales as $1/f_a^2 \propto m_a^2$: axions become *more* dangerous as they become heavier, in contrast to a particle of fixed coupling. The observable consequence is always the same. An open axion channel accelerates the consumption of nuclear fuel and shortens the corresponding evolutionary phase, and the duration of that phase can be read off from the relative population of stars occupying it within a coeval system such as a globular cluster. We now apply this template, in order of increasing density, to the environments that isolate each coupling.

2.4.2 Horizontal-branch stars: the photon coupling

The core helium-burning, or horizontal-branch, phase of low-mass stars is especially sensitive to the Primakoff process (2.66). The core temperature and density during helium burning are high enough for axion emission to compete with the nuclear luminosity, and the duration of the phase is directly measurable from the observed ratio of horizontal-branch to red-giant-branch stars in globular clusters, conventionally quantified by the R -parameter. Requiring that axion losses not shorten the horizontal-branch lifetime beyond what this ratio allows gives [18, 66]

$$g_{a\gamma\gamma} \lesssim 0.66 \times 10^{-10} \text{ GeV}^{-1}, \quad (2.68)$$

which, for an $\mathcal{O}(1)$ anomaly ratio $E/N = 1.92$, translates into $f_a \gtrsim 10^7 \text{ GeV}$, unless the two unrelated terms cancel significantly which can also be seen as fine tuning. This constraint defines the lower edge of the “axion window.” The electron coupling is probed instead by the later evolutionary stages of the same low-mass stars, to which we turn next.

2.4.3 Red giants and white dwarfs: the electron coupling

For models in which the axion couples to electrons at tree level, most importantly DFSZ, the bremsstrahlung channel (2.67) with $f = e$ is best constrained by two related classes of low-mass star.

The first is the tip of the red-giant branch. In a low-mass red giant, axion bremsstrahlung on electrons in the degenerate helium core would drain energy and delay the onset of helium ignition, allowing the core to grow more massive and the tip of the red-giant branch to reach a higher luminosity than standard models predict. The observed tip luminosity in globular clusters therefore bounds the electron coupling, with current analyses giving $g_{aee} \equiv C_e m_e / f_a \lesssim 1.5 \times 10^{-13}$ [67, 68]. Some of these analyses even report a mild preference for a non-zero value [68], a hint that remains statistically weak but that motivates continued study.

The second class is cooling white dwarfs. A white dwarf cools by photon and neutrino emission from a degenerate, non-burning interior, and the additional channel (2.67) would accelerate that cooling. The effect is observable in two independent ways: in the shape of the white-dwarf luminosity function, which counts white dwarfs per luminosity interval and would be depleted at the faint end by extra cooling, and in the measured secular period drift $\dot{P}/P$ of pulsating white dwarfs, which tracks the cooling rate directly. Both probes bound C_e and, for DFSZ-like models, translate into $f_a \gtrsim 10^8\text{--}10^9$ GeV [69–71]. Intriguingly, several white-dwarf datasets are fit slightly better with a small non-zero axion coupling [10], the same weak “cooling anomaly” hinted at by the red-giant tip, which is one reason the electron coupling remains an active target. The nucleon coupling, by contrast, is probed most powerfully in far denser environments, beginning with neutron stars.

2.4.4 Neutron-star cooling: the nucleon coupling

Neutron stars provide the densest laboratories accessible to the axion–nucleon coupling. In the degenerate nuclear matter of a neutron-star core, axions are radiated by the bremsstrahlung process (2.67) with $f = N$, at a rate that competes with the standard neutrino emission governing the star’s thermal evolution. An open axion channel would accelerate the cooling and lower the surface temperature of a neutron star of given age below the value predicted by standard cooling models.

The most-studied case is the young neutron star in the supernova remnant Cassiopeia A, whose surface temperature has been monitored over roughly two decades and appears to be dropping unusually fast. Standard explanations invoke enhanced neutrino emission from the onset of neutron superfluidity, but the same rapid cool-

ing also bounds any additional axion losses, giving limits on the nucleon coupling that, for QCD-axion models, correspond to $f_a \gtrsim \text{few} \times 10^8 \text{ GeV}$ [72, 73]. The cooling trend itself is not universally accepted, however: part of the apparent decline has been attributed to calibration drift in the Chandra ACIS instrument rather than to the star’s intrinsic cooling, so this bound, while comparable in strength to the globular-cluster bounds above, rests on an observation that is still debated. A more robust, if slightly weaker, constraint follows from the thermal-luminosity distribution of the wider population of isolated neutron stars and magnetars, which gives $f_a \gtrsim \text{few} \times 10^8 \text{ GeV}$ without relying on any single contested source [61].

A separate and intriguing data point comes from the stacked X-ray spectra of the seven thermally emitting ‘Magnificent Seven’ isolated neutron stars, which show an unexplained hard excess in the 2–8 keV band [74]. This excess is compatible with an axion-like particle produced by nucleon bremsstrahlung in the stellar core and reconverted to photons in the star’s magnetosphere, giving an allowed region $m_a \lesssim 2 \times 10^{-5} \text{ eV}$, $g_{a\gamma\gamma} g_{ann} \in (2 \times 10^{-21}, 10^{-18}) \text{ GeV}^{-1}$. Because the observable constrains the product $g_{a\gamma\gamma} g_{ann}$ rather than either coupling alone, translating this into a bound on the nucleon coupling requires an independent assumption on $g_{a\gamma\gamma}$; we quote it here as a suggestive hint rather than list it alongside the exclusion bounds above. These arguments probe the nucleon coupling in the same f_a range as the classic globular-cluster bounds, and they are sharpened decisively by the neutrino burst of Supernova 1987A, to which we now turn.

2.4.5 Supernova 1987A

The most powerful astrophysical constraint comes from a single event: the core collapse of Supernova 1987A, whose neutrino burst was recorded directly by the Kamiokande-II [75], IMB [76], and Baksan [77] detectors, with a duration and total energy in good agreement with standard core-collapse simulations. Axions produced in the proto-neutron-star core through the bremsstrahlung process (2.67) would carry away energy in addition to the neutrinos. Provided the axion mean free path is long enough for it to escape the core freely rather than being reabsorbed, this extra channel would shorten the observed neutrino signal below the several-second duration actually recorded.

The condition is conventionally phrased as the energy-loss criterion of Ref. [18]: a novel channel is excluded if, at the fiducial core conditions of density $\rho \sim 3 \times 10^{14} \text{ g cm}^{-3}$ and temperature $T \sim 30 \text{ MeV}$, its emissivity per unit mass exceeds roughly $\varepsilon_a \lesssim 10^{19} \text{ erg g}^{-1} \text{ s}^{-1}$, comparable to the neutrino emissivity that sets the observed cooling time. Applied to the axion, this criterion excludes a broad window

of axion–nucleon couplings corresponding, very roughly, to

$$10^8 \text{ GeV} \lesssim f_a \lesssim \text{few} \times 10^9 \text{ GeV}. \quad (2.69)$$

The two edges of the window have quite different physical origins. Above the upper edge the coupling is too weak for the argument to bite: too few axions are produced to affect the signal, so large f_a is unconstrained. Below the lower edge the axion mean free path becomes shorter than the core radius, and the axion is *trapped*: it is reabsorbed and rethermalized before escaping, so that instead of streaming out instantly it must diffuse out slowly, exactly as the neutrinos themselves diffuse out over several seconds. In this trapping regime the axion once again fails to be an efficient energy sink, and the exclusion window closes rather than extending to arbitrarily small f_a .⁸ The transition between free-streaming and trapped emission reproduces, for the axion, the same physics that governs neutrino transport in the collapsing core.

The precise boundaries of the window depend on the detailed nuclear physics of the bremsstrahlung rate, in particular on the pion-exchange dynamics and multiple-scattering suppression in dense matter, and they remain the subject of continued refinement with modern supernova simulations [78, 79]. The SN1987A bound is contested; it rests on single-event statistics, and the modeling is sensitive to uncertain assumptions about the supernova’s explosion mechanism [80]. Together, the arguments of Sections 2.4.2–2.4.5 are the origin of the commonly quoted lower bound $f_a \gtrsim 10^8\text{--}10^9 \text{ GeV}$ on every invisible-axion model, a bound that directly overlaps the parameter space of the composite constructions of Chapter 3. What the energy-loss argument can never do, however, is discover the axion. For that, the processes of this section must be run in reverse.

2.4.6 Laboratory detection

Every laboratory search inverts a production process of Section 2.4.1. The relevant interaction is the *inverse Primakoff process*: in a strong static magnetic field, which supplies one photon leg of the vertex (2.64) exactly as the stellar electrostatic field does in (2.66), an axion converts into a real photon,

$$a \xrightarrow{B} \gamma. \quad (2.70)$$

The major detection techniques are all realizations of this conversion; what dis-

⁸However, the region $f_a \lesssim 10^8 \text{ GeV}$ is independently excluded by other bounds.

tinguishes them is the assumed source of axions: the Sun, the laboratory itself, or the galactic halo. We take them in that order, moving from an astrophysical source, through none at all, to a cosmological one.

Helioscopes. The Sun is the nearest axion source: its core emits a Primakoff flux (2.66) exactly analogous to the one bounded in horizontal-branch stars. A helioscope is a long, strong dipole magnet pointed at the Sun, reconverting this flux into X-ray photons through (2.70) at the far end of the magnet bore. The CERN Axion Solar Telescope (CAST) is the leading realized example, setting $g_{a\gamma\gamma} < 0.66 \times 10^{-10} \text{ GeV}^{-1}$ for $m_a \lesssim 0.02 \text{ eV}$ [81], matching the horizontal-branch bound (2.68) with an entirely independent technique, while its proposed successor IAXO [82] would improve on this by more than an order of magnitude and reach into the QCD-axion band for m_a up to roughly 1 eV.

Light-shining-through-walls. A second class of experiments dispenses with any external source. An intense laser beam in a magnetic field converts photons to axions by the Primakoff process; the axions traverse an optically opaque wall unattenuated, and are reconverted to photons by (2.70) in a second magnetic-field region beyond it; the production and detection vertices of the stellar argument realized back to back on a laboratory bench. Realized in the ALPS and OSQAR experiments [83, 84], this technique is currently far less sensitive than the alternatives, but it makes no astrophysical or cosmological assumption whatsoever about the axion’s origin or abundance.

Microwave-cavity haloscopes. As we will see in Section 2.5, the axion is a natural candidate for the galactic dark matter; anticipating this result, the halo threads every laboratory with a non-relativistic axion field of narrow velocity dispersion, $v/c \sim 10^{-3}$. A haloscope, proposed by Sikivie [85], places a tunable, high- Q microwave cavity inside a strong magnet: when the cavity resonance matches the axion mass, $\nu = m_a c^2/h$, halo axions convert resonantly into microwave photons through (2.70), with the power enhanced by the quality factor. The ADMX collaboration has reached DFSZ sensitivity over the band $2.7 \mu\text{eV} \lesssim m_a \lesssim 4.2 \mu\text{eV}$ [86, 87] with recent results at $3.3 \mu\text{eV}$ [88]. Couplings that cavity haloscopes cannot reach, in particular the CP-odd axion–gluon coupling, which induces an oscillating nucleon EDM, are instead targeted by nuclear-magnetic-resonance experiments such as CASPER [89], and by a broad next-generation program surveyed in [90].

The reach of these searches should be measured against the window established above. Through $m_a \propto 1/f_a$, the ADMX band corresponds to $f_a \approx (1.4\text{--}2) \times$

10^{12} GeV: haloscopes have therefore already reached the QCD-axion line at the very top of the viable window, but only in narrow mass slices, and only under the assumption that axions constitute the entirety of the local dark matter. The middle and lower reaches of the window, $f_a \sim 10^8\text{--}10^{11}$ GeV, corresponding to masses from tens of μeV to tens of meV, remain largely untested in the laboratory, with next-generation haloscopes and IAXO advancing from the two ends. Over most of the window, therefore, the astrophysical bounds of this section remain the most powerful experimental input into the parameter space of the composite constructions of Chapter 3. Those bounds constrain the axion’s couplings; its cosmological history, which we develop next, constrains its abundance, closes the window from above, and determines whether the haloscope’s central assumption, axion dark matter, is in fact realized.

2.5 Axion cosmology

The axion field leaves cosmological imprints that must satisfy the constraints of early-universe observation. Its qualitative behavior hinges on a single timing question: did the Peccei–Quinn phase transition, at which $U(1)_{\text{PQ}}$ spontaneously breaks, occur before or after cosmic inflation? Did the maximum temperature after reheating exceed v_{PQ} ?

If the transition occurs before inflation, and the symmetry is not subsequently restored during reheating, a single causal patch of the pre-inflationary universe is inflated to cover the entire observable universe today. The axion field is then completely homogeneous on observable scales, characterized by a single number, the initial misalignment angle θ_{in} , and no topological defects survive, since any that formed before inflation are diluted to cosmologically irrelevant abundance. This is the *pre-inflationary* scenario of Section 2.5.2.

If the transition occurs after inflation, causally disconnected patches of the post-inflationary universe, regions that were never in causal contact and so could not have agreed on a common value of the axion field, each select their own statistically independent value of θ_{in} . Wherever such patches meet the field must interpolate, and topological defects, first strings and later domain walls, inevitably form, each with its own contribution to the axion abundance. This is the *post-inflationary* scenario of Sections 2.5.3–2.5.5.

The treatment in this section follows the standard framework of Refs. [29, 91–93]. We begin with the ingredient common to both scenarios, the temperature dependence of the axion potential, then treat the pre- and post-inflationary cases in

turn, and finally arrive at the domain-wall problem and its resolutions.

2.5.1 The axion potential in the early universe

The early universe is homogeneous, isotropic, and spatially flat on the largest scales, so the metric takes the Robertson–Walker form, $ds^2 = dt^2 - R^2(t) d\mathbf{x} \cdot d\mathbf{x}$, with Hubble rate $H(t) = \dot{R}(t)/R(t)$. The axion equation of motion in this background, following from (2.51) and (2.54), is

$$\ddot{a} + 3H(t)\dot{a} - \frac{1}{R^2(t)}\nabla^2 a + \frac{\partial V}{\partial a} = 0, \quad (2.71)$$

with the Hubble term acting as time-dependent friction.

At the high temperatures of interest the potential takes the cosine form of the dilute instanton gas,

$$V(a, T) = m_a^2(T) f_a^2 \left[1 - \cos(a/f_a) \right], \quad (2.72)$$

matching onto the exact chiral result (2.52) as $T \rightarrow 0$, with the zero-temperature mass fixed by (2.54). This is precisely where the dilute instanton gas of Section 2.1.4, set aside there because it was dominated by uncontrollably large instantons at zero temperature, is properly cut off by the finite temperature. A thermal bath of gluons endows the gauge field with a Debye screening mass,

$$m_D^2(T) = g_s^2(T) T^2 \left(1 + \frac{N_f}{6} \right), \quad (2.73)$$

which acts as an infrared regulator on the instanton-size integral (2.24): an instanton larger than $\rho \sim m_D^{-1}(T)$ is exponentially suppressed, because its long-wavelength gauge field is screened by the thermal plasma just as a long-wavelength electric field is screened in an ordinary plasma of charges.

Repeating the partition-function construction of (2.24)–(2.25) with this thermal cutoff,

$$m_a^2(T) f_a^2 = 2 \int d\rho D(\rho) e^{-2\pi^2(1+N_f/6) T^2 \rho^2} \sim \Lambda^4 \left(\frac{\Lambda}{T} \right)^{7+N_f/3} \prod_q \frac{m_q}{\Lambda}, \quad (2.74)$$

the integral is now cut off at exactly the scale that invalidated the $T = 0$ calculation, so the semiclassical approximation is self-consistently under control for $T \gtrsim \Lambda_{\text{QCD}}$ [94]. This thermal computation, and not the zero-temperature one of Section 2.1.4 where chiral perturbation theory was the only reliable tool, underlies

the temperature-dependent axion mass in (2.72),

$$m_a(T) \simeq \Lambda_{\text{QCD}} \left(\frac{\Lambda_{\text{QCD}}}{f_a} \right) \left(\frac{\Lambda_{\text{QCD}}}{T} \right)^4, \quad T \gtrsim \Lambda_{\text{QCD}}. \quad (2.75)$$

Lattice calculations find an effective exponent close to, but not exactly, 4 depending on the temperature range [95], so we adopt the value 4 for convenience, and match the mass in the zero-temperature limit to $m_a(0)$ fixed by (2.31). The mass is exponentially suppressed at $T \gg \Lambda_{\text{QCD}}$, and it is turned on rapidly as T approaches Λ_{QCD} , defining an onset temperature T_{osc} at which $m_a(T_{\text{osc}}) \sim 3H(T_{\text{osc}})$,

$$T_{\text{osc}} \sim \Lambda_{\text{QCD}} \left(\frac{M_{\text{Pl}}}{f_a} \right)^{1/6} \sim 1 \text{ GeV} \left(\frac{10^{12} \text{ GeV}}{f_a} \right)^{1/6}. \quad (2.76)$$

The temperature T_{osc} marks the onset of coherent axion oscillations. To follow the field through that onset in both scenarios at once, it is convenient to decompose it into comoving Fourier modes,

$$a(\mathbf{x}, t) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} a_{\mathbf{k}}(t) e^{i\mathbf{k} \cdot \mathbf{x}}. \quad (2.77)$$

Linearizing the potential about the minimum, $V'_a(a) = m_a^2(t) f_a \sin(a/f_a) \simeq m_a^2(t) a$, each mode obeys

$$\ddot{a}_{\mathbf{k}} + 3H(t) \dot{a}_{\mathbf{k}} + \left[\frac{k^2}{R^2(t)} + m_a^2(t) \right] a_{\mathbf{k}} = 0. \quad (2.78)$$

The friction term is removed by the rescaling $\psi_{\mathbf{k}} \equiv R^{3/2}(t) a_{\mathbf{k}}$, which recasts (2.78) as an undamped oscillator with a time-dependent frequency,

$$\ddot{\psi}_{\mathbf{k}} + \omega_k^2(t) \psi_{\mathbf{k}} = 0, \quad \omega_k^2(t) = \frac{k^2}{R^2(t)} + m_a^2(t) - \frac{9}{4} H^2(t) - \frac{3}{2} \dot{H}(t). \quad (2.79)$$

In the radiation era, $H = 1/2t$, the two curvature terms combine into $+3/16t^2$, which is negligible as soon as either $m_a(t)$ or $k/R(t)$ rises above $H(t)$, leaving $\omega_k^2 \simeq k^2/R^2 + m_a^2$. Once the frequency varies adiabatically, $\dot{\omega}_k \ll \omega_k^2$, the WKB solution applies,

$$a_{\mathbf{k}}(t) \simeq \frac{C_{\mathbf{k}}}{R^{3/2}(t) \sqrt{\omega_k(t)}} \cos\left(\int^t \omega_k dt' + \varphi_{\mathbf{k}}\right), \quad (2.80)$$

and the adiabatic invariant $R^3 \omega_k A_k^2$ is conserved, with A_k the mode amplitude. Eqs. (2.78)–(2.80) contain all the dynamics we need. The two cosmological scenarios differ only in which modes are populated, and with what amplitude, a question settled entirely by whether the PQ transition preceded or followed inflation. We

take the pre-inflationary case first.

2.5.2 Pre-inflationary axions: misalignment and isocurvature

If the PQ transition occurs before inflation, the axion field is stretched to scales far outside the horizon during inflation. Modes that are initially sub-horizon oscillate and are diluted as $a_{\mathbf{k}} \propto R^{-3/2}$ until they exit the horizon; thereafter, their amplitudes approach a constant while their physical momenta k/R continue to decrease exponentially. Thus inflation exponentially suppresses spatial gradients and inhomogeneities on fixed physical scales, effectively leaving the axion field spatially homogeneous over the observable Universe. The homogeneous mode is held in place by Hubble friction for as long as the axion potential is negligible.

$$a(t) = a_0 \equiv f_a \theta_{\text{in}}, \quad \theta_{\text{in}} \in [-\pi, \pi), \quad H(t) \gg m_a(t), \quad (2.81)$$

where the *initial misalignment angle* θ_{in} is a free parameter: with the entire observable universe descending from a single PQ patch, the theory makes no prediction for its value.

As the universe cools toward T_{osc} the growing mass overtakes the friction and the zero mode ($k = 0$) begins to oscillate. Its evolution is the $k = 0$ case of the WKB solution (2.80) with $\omega_0(t) \simeq m_a(t)$, so the amplitude redshifts as

$$A_0(t) \simeq \frac{a_0}{R^{3/2}(t)} \left(\frac{m_a(t_{\text{osc}})}{m_a(t)} \right)^{1/2}, \quad (2.82)$$

normalized to a_0 at the onset time t_{osc} defined by $m_a(t_{\text{osc}}) \simeq 3H(t_{\text{osc}})$, with $R(t_{\text{osc}}) \equiv 1$. Averaging over one oscillation, the energy and number densities of the coherent condensate are

$$\rho_a = \frac{1}{2} \langle \dot{a}_0^2 + m_a^2 a_0^2 \rangle = \frac{1}{2} m_a^2(t) A_0^2(t), \quad n_a = \frac{\rho_a}{m_a} = \frac{1}{2} m_a(t) A_0^2(t) \propto R^{-3}(t). \quad (2.83)$$

The oscillating zero mode is thus a coherent condensate of non-relativistic axions whose comoving number is conserved from the onset onward, which is precisely the adiabatic-invariant statement below (2.80). Evaluating n_a at t_{osc} gives [91–93]

$$n_a(t_{\text{osc}}) \sim m_a(T_{\text{osc}}) f_a^2 \theta_{\text{in}}^2, \quad (2.84)$$

and redshifting this conserved number to the present, where $m_a(0)$ fixes the energy

density $\rho_a(0) = m_a(0) n_a(0)$, yields the misalignment relic abundance [17]⁹

$$\Omega_a h^2 \simeq 0.45 \theta_{\text{in}}^2 \left(\frac{f_a}{10^{12} \text{ GeV}} \right)^{1.19} \left(\frac{\Lambda_{\text{QCD}}}{1 \text{ GeV}} \right). \quad (2.85)$$

This is the central formula of the pre-inflationary scenario. Requiring $\Omega_a h^2$ not to exceed the observed dark-matter density $\Omega_{\text{cdm}} h^2 \simeq 0.12$ [96] bounds f_a from above for generic $\theta_{\text{in}} = \mathcal{O}(1)$; conversely, a smaller θ_{in} , for which the scenario offers no dynamical explanation, allows f_a to be pushed arbitrarily high.

For the sake of completeness, we will consider the question of isocurvature perturbations due to inflation. The inflationary expansion that homogenizes the field also seeds fluctuations in it. During inflation the axion is an almost massless scalar, so like any light field it acquires de Sitter fluctuations of order $H_I/2\pi$ per Hubble time [97–99], which for the angle $\theta = a/f_a$ shift the effective misalignment as

$$\theta_{\text{in}}^2 \longrightarrow \theta_{\text{in}}^2 + \left(\frac{H_I}{2\pi f_a} \right)^2. \quad (2.86)$$

Because these are fluctuations of the axion and not of the inflaton, they carry no perturbation in the total energy density at the time they are generated, and instead reappear later as spatial variations in the axion (dark-matter) fraction relative to radiation. These are *isocurvature* perturbations [100], and they imprint a distinctive signature on the cosmic microwave background that the adiabatic, inflaton-sourced spectrum does not. Planck measures the adiabatic power to be $\mathcal{P}_{\text{ad}} \simeq 2.2 \times 10^{-9}$ and detects no isocurvature component, bounding its fraction to $\beta_{\text{iso}} \equiv \mathcal{P}_{\text{iso}}/(\mathcal{P}_{\text{ad}} + \mathcal{P}_{\text{iso}}) < 0.038$, so that $\mathcal{P}_{\text{iso}} \lesssim 8 \times 10^{-11}$ [101]. Writing the isocurvature power in terms of the axion parameters,

$$\mathcal{P}_{\text{iso}} = \left(\frac{\delta\Omega_a}{\Omega_{\text{cdm}}} \right)^2 = \frac{4}{\theta_{\text{in}}^2} \left(\frac{H_I}{2\pi f_a} \right)^2 \left(\frac{\Omega_a}{\Omega_{\text{cdm}}} \right)^2, \quad (2.87)$$

turns the Planck bound into a stringent upper limit on the inflationary Hubble scale H_I for any given f_a, θ_{in} .

⁹The exponent of scale dependence is slightly different from the DIGA induced value $\rho_a(0) \sim m_a(0)m_a(T_{\text{osc}})f_a^2\theta_{\text{in}}^2T_{\text{osc}}^{-3} \sim \theta_{\text{in}}^2T_{\text{osc}}^{-7} \sim \theta_{\text{in}}^2f_a^{7/6}$ since it becomes less reliable as temperature decreases.

2.5.3 Post-inflationary axions: the mode spectrum and vacuum realignment

If the PQ transition occurs *after* inflation instead, which is the scenario of central interest to this thesis, the field is genuinely inhomogeneous on scales up to the horizon at the transition, and no single mode dominates: every $a_{\mathbf{k}}$ in (2.78) must be followed. A mode is super-horizon when $\lambda_k(t) = 2\pi R(t)/k \gg H^{-1}(t)$, which is equivalent to $k/R \ll H$. Conversely, the mode is sub-horizon when $k/R \gg H$; the zero mode of Section 2.5.2 is the $k \rightarrow 0$ limit of the former.

Before axion-mass dominance, $T > T_{\text{osc}}$. In this epoch, the axion mass $m_a(t)$ is still negligible, so $\omega_k^2 \simeq k^2/R^2$. Super-horizon modes have $k/R \ll H$, hence the WKB condition fails, the modes are overdamped, and they freeze at constant amplitude with negligible energy density, exactly as the zero mode froze in Section 2.5.2. On the other hand, sub-horizon modes have $k/R \gg H$, so their own momentum plays the role of a mass and drives oscillation even before the potential turns on. The WKB solution (2.80) with $\omega_k \simeq k/R$ gives amplitude $A_k \propto R^{-3/2}(k/R)^{-1/2} \propto R^{-1}$, and hence

$$\rho_{a,k} = \frac{1}{2} \langle \dot{a}_k^2 + (k^2/R^2) a_k^2 \rangle \simeq \frac{k^2}{R^2} A_k^2 \propto R^{-4}(t), \quad (2.88)$$

redshifting as radiation, as befits relativistic quanta.

After axion-mass dominance, $T < T_{\text{osc}}$. Once $m_a(t) > 3H(t)$ the frequency becomes $\omega_k^2 \simeq m_a^2 + k^2/R^2$. The modes require additional classification. Super-horizon modes, with $k/R \ll H < m_a$, oscillate at $\omega_k \simeq m_a$ and automatically redshift as matter, $\rho \propto m_a(T)R^{-3}$. Sub-horizon modes, however, are classified further into two bands. The middle band is defined through $H \ll k/R < m_a$; the frequency of each of these non-relativistic modes is approximately given by the axion mass, so that their energy density $\rho \propto m_a(T)R^{-3}$. The third band of relativistic sub-horizon modes, $H < m_a \ll k/R$, has the frequency approximately as k/R , then they redshift as radiation $\rho \propto R^{-4}$.

In the post-inflationary axion scenario, the abundance of cold axions averages over all initial angles, $\Omega_a h^2 \propto \langle \theta_{\text{in}}^2 \rangle \sim \pi^2/3$.¹⁰ The abundance is therefore a definite number rather than a free parameter, which is exactly what makes the post-inflationary scenario predictive; however, it also makes its topological defects unavoidable.

¹⁰This average is obtained from using $\sin(a/f_a) \approx a/f_a$ for the entirety of the interval $a/f_a \in [-\pi, \pi)$. The correct average needs to integrate using the full sine function instead.

2.5.4 Axion strings

The independence of the patches carries a topological consequence with no pre-inflationary counterpart. Each patch sits somewhere on the vacuum circle $a \in [0, 2\pi N f_a)$, and wherever several patches meet the field can wind: sampled around a small loop, a/f_a may run once, or several times, around the full circle. Continuity then forbids the field from lying in the vacuum everywhere inside the loop, and it is forced back up to the symmetric point $\langle |\Phi| \rangle = 0$ along a line threading the loop, as sketched in Figure 2.3. By the Kibble mechanism [19] this line is a stable one-dimensional defect, an *axion string*, classified by the winding number of $\pi_1(S^1) = \mathbb{Z}$. The tension of a string is dominated not by its thin core but by the slowly falling

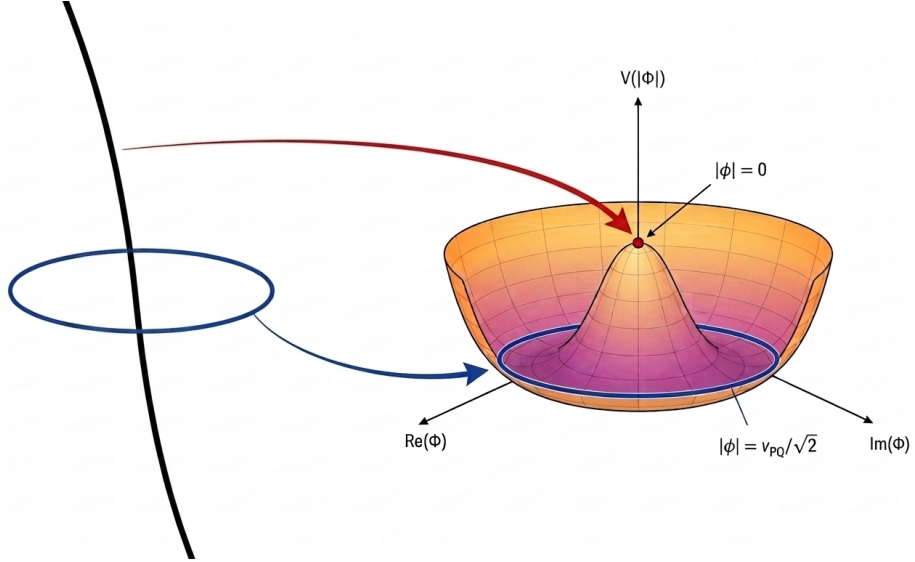


Figure 2.3: An axion string. Around any loop encircling the core, the phase of Φ winds once through 2π (left); on the vacuum manifold this is a full circuit of the brim of the Mexican-hat potential (right). Continuity then forces the field inside the loop up to the symmetric point $|\Phi| = 0$, and the locus of such points traces out the string core.

gradient energy of the winding field around it. A straight string of unit winding has an azimuthal gradient $|\nabla a| \simeq f_a/r$ at transverse distance r , so integrating the gradient energy $\frac{1}{2}|\nabla a|^2$ over the transverse plane gives [102]

$$\mu = \pi N^2 f_a^2 \log(f_a/H), \quad (2.89)$$

where the logarithm is cut off in the ultraviolet by the core size f_a^{-1} and in the infrared by the typical inter-string separation, of order the horizon H^{-1} . The tension thus grows logarithmically as the network coarsens.

Consider first a hypothetical network of *non-interacting* strings, simply carried along by the expansion. With the number of strings per comoving volume held fixed

and each physical length growing as R , the string length per unit physical volume scales as $R/R^3 = R^{-2}$, so

$$\rho_{\text{str}}^{\text{free}} = \mu \times \frac{\text{length}}{\text{volume}} \propto R^{-2}(t) \quad (\text{free stretching}), \quad (2.90)$$

one power slower than radiation, because the string's one-dimensional stretching cancels one factor of the volume dilution. A network diluting only as R^{-2} would eventually dominate, the string analog of the domain-wall problem below.

Simulations, however, find that the real network avoids this fate by relaxing almost immediately into a *scaling* regime [103–105], in which long strings reconnect, shed loops and radiate axions so as to keep their number per Hubble volume at an $\mathcal{O}(1)$ constant. The length per Hubble volume is then $\sim H^{-1}$, the length density $\sim H^2$, and

$$\rho_{\text{str}}(t) \simeq \mu H^2(t) \sim \frac{\mu}{t^2} \propto R^{-4}(t) \quad (\text{scaling}), \quad (2.91)$$

the same dilution as radiation.

Maintaining scaling requires the network to shed its excess energy continuously, which it does by radiating relativistic axions from oscillating loops and kinks. Book-keeping the transferred energy and the resulting axion number gives the source equations [106]

$$\dot{\rho}_{\text{str}} + 2H(t) \rho_{\text{str}} = -\Gamma_{\text{str} \rightarrow a}(t), \quad \dot{n}_{\text{str} \rightarrow a} + 3H(t) n_{\text{str} \rightarrow a} = \frac{\Gamma_{\text{str} \rightarrow a}(t)}{\omega(t)}, \quad (2.92)$$

where $\Gamma_{\text{str} \rightarrow a} = d\rho_{\text{str} \rightarrow a}/dt$ is the energy radiated per unit time and $\omega(t)$ the mean energy of the emitted quanta. The string-radiated axions of (2.92) are a second cold-axion source alongside vacuum realignment, and because $\omega(t)$ is set by the small-scale structure of the network, which is numerically costly to resolve, their spectrum, and hence the total post-inflationary abundance, remains actively debated. Strings, though, are only the first half of the defect story: the instant the axion potential, generated by QCD effects, switches on, every string becomes the rim of a second, far more dangerous defect.

2.5.5 Domain walls and the domain wall problem

When the temperature reaches T_{osc} and the potential (2.72) switches on, the N degenerate minima of Section 2.3.4 become physically realized vacua, and every patch relaxes into whichever minimum lies nearest its local θ_{in} . Two patches that settle into adjacent minima are separated by a surface across which the field interpolates from one vacuum to the next: a *domain wall*. Its profile is the static kink of the

potential (2.72), and minimizing the wall energy per unit area gives the sine-Gordon solution $a(z) = 4f_a \arctan e^{m_a z}$ with surface tension [45]

$$\sigma = \int dz \left[\frac{1}{2} (\partial_z a)^2 + V_a(a) \right] = 8 m_a(0) f_a^2, \quad (2.93)$$

evaluated with the zero-temperature mass, since the walls form as $m_a(T)$ saturates. Because the field winds once around the *entire* vacuum circle on encircling a string, it must pass through all N minima in doing so, so every string becomes the common boundary of exactly $N_{\text{DW}} = N$ walls.

The fate of the resulting string-wall network turns entirely on whether $N_{\text{DW}} = 1$ or $N_{\text{DW}} \geq 2$. If $N_{\text{DW}} = 1$, a single wall ends on each string. Each wall is then bounded on all sides by string, and its own tension pulls the bounding string taut and reels it in, collapsing the wall-string system within about a Hubble time of formation. Essentially all of the network energy is converted into axion radiation, with a small fraction into gravitational waves, and no cosmological catastrophe results.

If $N_{\text{DW}} \geq 2$, by contrast, N_{DW} walls attach to every string and pull on it from different directions; their tensions balance, and no wall can collapse its neighbors. Here we must be careful about how the surviving network dilutes, because this is exactly where a naive estimate misleads.

Treating the walls as *non-interacting* sheets merely stretched by the expansion, one would fix the number of walls per comoving volume and let each physical area grow as R^2 , giving an area-per-volume of $R^2/R^3 = R^{-1}$ and

$$\rho_{\text{DW}}^{\text{free}} = \sigma \times \frac{\text{area}}{\text{volume}} \propto R^{-1}(t) \quad (\text{free stretching}). \quad (2.94)$$

This is the two-dimensional analog of the free-string estimate (2.90), and it is *not* what a real axion-wall network does, for exactly the reason the strings did not dilute as R^{-2} : the walls interact. They self-intersect and straighten under their own tension, driving the network into a scaling regime with roughly one wall of horizon-scale curvature per Hubble volume. The wall area per Hubble volume is then $\sim H^{-2}$, the area density $\sim H$, and

$$\rho_{\text{DW}}(t) \simeq \sigma H(t) \sim \frac{\sigma}{t} \propto R^{-2}(t) \quad (\text{scaling, radiation era}). \quad (2.95)$$

The interacting network therefore dilutes as R^{-2} , faster than the frozen estimate (2.94) but still far slower than radiation (R^{-4}) or matter (R^{-3}), so a stable wall network is guaranteed to overtake them.

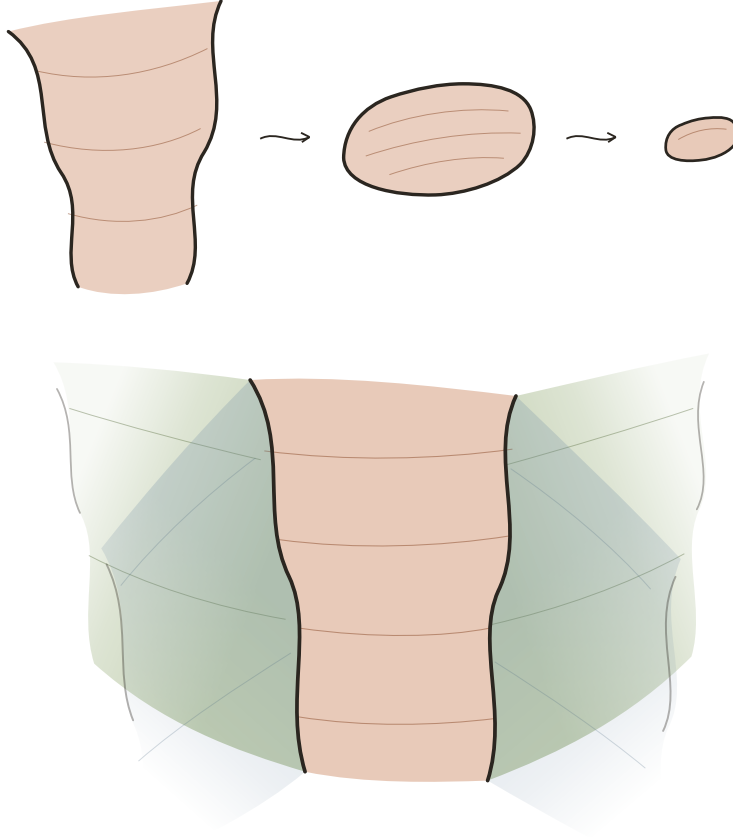


Figure 2.4: The two fates of the string-wall network. Top: for $N_{\text{DW}} = 1$, each wall is bounded by string; the wall tension reels the string in and the system collapses into axion radiation within roughly a Hubble time. Bottom: for $N_{\text{DW}} \geq 2$, several walls pull on every string, their tensions balance, and the network survives.

This is in gross conflict with the successful predictions of standard Big Bang cosmology. This is the *domain wall problem*: any axion model with $N_{\text{DW}} \geq 2$ in a post-inflationary cosmology is, without further structure, ruled out. Three logically distinct escapes exist. The first, and most commonly invoked, is to place the PQ transition before inflation (Section 2.5.2), so that no strings or walls ever form, whatever the value of N_{DW} ; the price is the loss of the scenario’s one sharp prediction, the parameter-free relic abundance, in exchange for an unknown initial angle θ_{in} . The second is to arrange $N_{\text{DW}} = 1$ so that the network self-destructs even after inflation, which requires understanding, specialized to the UV model, why the naive vacuum count overstates the number of physically distinct vacua; this is the Lazarides–Shafi mechanism, taken up next and made the organizing principle of Chapter 3. The third is to add a small explicit breaking of the degeneracy-protecting discrete symmetry, biasing one vacuum so that its walls collapse before t_w ; this works only in a narrow window, since the bias must be large enough to destroy the walls

in time yet small enough not to reintroduce an unacceptable shift in $\bar{\theta}$ through the same quality-problem logic of Section 2.3.5. We will focus on the second option for the rest of this part.

2.6 Composite axion models

The reader now holds three separate liabilities of the elementary invisible axion: the hierarchy problem of a fundamental PQ scalar, the quality problem that Planck-suppressed operators inflict on any global $U(1)_{\text{PQ}}$ (Section 2.3.5), and the cosmological domain-wall problem of any $N_{\text{DW}} \geq 2$ post-inflationary scenario (Section 2.5.5).

This section seeks to see how the chiral gauge structure of a new confining sector addresses these problems. Confinement generates f_a by dimensional transmutation, with no elementary scalar anywhere, dispatching the hierarchy problem outright. If the confining gauge theory is chiral, the low-dimensional PQ-violating operators are forbidden by gauge invariance rather than by assumption, parametrically protecting the quality. And the unbroken discrete gauge structure that any confining group carries supplies exactly the identification of vacua that the Lazarides–Shafi mechanism uses to reduce N_{DW} . We develop these three payoffs in turn: the Choi–Kim model, which secures the first; the chiral moose, which secures the second; and the Lazarides–Shafi mechanism, which secures the third and launches the constructions of Chapter 3.

2.6.1 The Choi–Kim model

The first composite axion model was proposed by Choi and Kim [21, 22], by direct analogy with QCD. Take an $SU(N)$ gauge theory with four Dirac fermions in the fundamental representation. Exactly as for two-flavor QCD, this theory has the chiral symmetry breaking pattern $SU(4)_L \times SU(4)_R \rightarrow SU(4)_V$ below its confinement scale. Embed $SU(3)_c$ in the diagonal $SU(4)_V$, so that the fundamental quartet decomposes into a colored triplet Q and a colorless singlet η , and take $U(1)_{\text{PQ}}$ to act only on the left-handed fields, with the charges of Table 2.1 chosen traceless within $SU(4)_L$ so that the current is non-anomalous under the confining $SU(N)$ while remaining anomalous under color. Fixing the right-handed charges to zero in this way is the most convenient normalization for the domain-wall counting of Chapter 3.¹¹

Below the $SU(N)$ confinement scale, the chiral condensate forms. Once the QCD

¹¹The domain-wall number is a physical prediction, and it does not depend on the convention. The convention was used merely to simplify the arguments.

	$[SU(3)]_{\text{QCD}}$	$SU(4)_L$	$[SU(N)]$	$SU(4)_R$	$U(1)_{\text{PQ}}$
Q_L	3	4	N		+1
η_L	1		N		-3
Q_R	3		N	4	0
η_R	1		N		0

Table 2.1: Matter content of the Choi–Kim model, where $[G]$ denotes a gauged factor G .

coupling is switched on, the condensate aligns with the unbroken color group,

$$\frac{1}{3}\langle\bar{Q}_L Q_R\rangle = \langle\bar{\eta}_L \eta_R\rangle, \quad \langle\bar{Q}_L \eta_R\rangle = \langle\bar{\eta}_L Q_R\rangle = 0. \quad (2.96)$$

Fifteen pseudo-Goldstone bosons appear in the adjoint of the approximate $SU(4)_V$, decomposing under $[SU(3)]_{\text{QCD}}$ as $\mathbf{15} \rightarrow \mathbf{8} + \bar{\mathbf{3}} + \mathbf{3} + \mathbf{1}$. The colored states acquire mass at one loop from gluon exchange, just as the charged pion is heavier than the neutral one; these heavy colored states become charged massive particles (CHAMPs) and overclose the universe if not attended for, which is what will be done in Section 3.4.2. The singlet state is the composite axion who remains massless until the QCD anomaly switches on, exactly as in Section 2.3. The axion decay constant is the analog of f_π , set by the confinement scale of the new sector, $f_a \sim \Lambda_N$: the scale is generated by dimensional transmutation, and the hierarchy problem of an elementary PQ-breaking scalar is solved.

The quality of PQ symmetry. The underlying $SU(N)$ interaction is vector-like, so the bilinear $\bar{Q}_L Q_R$ is a gauge-invariant, PQ-violating Dirac mass term, and nothing forbids it from appearing with an unsuppressed coefficient. The PQ symmetry is protected only by an assumption of approximate global flavor symmetry, no better justified here than in the elementary models of Section 2.3. What is needed is a construction in which the dangerous mass term is forbidden by gauge invariance itself, and that is the subject of the next subsection.

The domain-wall number. The quarks of Choi–Kim model can be interpreted as N KSVZ quarks; therefore, the model has a domain-wall number $N_{\text{DW}} = N$. The derivation of this result is fairly simple, and is analogous to what we have done before. We will come back to this result when we develop the master formula to compute N_{DW} in Section 3.1.

2.6.2 Forbidding the mass term: chiral composite axions

The goal is a confining sector in which $U(1)_{\text{PQ}}$ is not merely embedded in an accidental flavor symmetry but to have its violating operators parametrically suppressed. This is done through using gauge symmetries to forbid the low-dimensional symmetry-breaking operators; this is precisely the logic for $U(1)_B$ in the Standard Model: no gauge-invariant, renormalizable operator violates baryon number, simply because no such operator can be built from the available fields. The Choi–Kim mass term $\bar{Q}_L Q_R$ is dangerous specifically because Q_L and Q_R are both charged under the *same* confining $SU(N)$ and the same representation, so a gauge-singlet bilinear exists. If Q_L and Q_R are instead charged under *different* gauge groups, no such bilinear can be written, and the mass term is forbidden by gauge invariance rather than by assumption.

This can be engineered with Georgi’s moose (deconstruction) technology [107–109]: arrange the confining dynamics along a linear chain of gauge groups, with matter fields living only on the links between adjacent nodes, so that no single gauge group ever sees both halves of a would-be Dirac mass term. A concrete two-site realization [110] doubles the Choi–Kim construction, with gauge groups $SU(3)_0 \times SU(N)_1 \times SU(4)_1 \times SU(N)_2$ and matter arranged as in Table 2.2 and Figure 2.5.

	$[SU(3)]_0$	$SU(4)_L$	$[SU(N)]_1$	$[SU(4)]_1$	$[SU(N)]_2$	$SU(4)_R$	$U(1)_{\text{PQ}}$
Q_L	3	4	$\bar{\mathbf{N}}$				+1
η_L	1		$\bar{\mathbf{N}}$				−3
χ_R			$\bar{\mathbf{N}}$	4			0
χ_L				4	$\bar{\mathbf{N}}$		0
Q_R	3				$\bar{\mathbf{N}}$	4	0
η_R	1				$\bar{\mathbf{N}}$		0

Table 2.2: Matter content of the moose-deconstructed extension of the Choi–Kim model.

Gauging $SU(4)_1$ fully, and only an $SU(3)_0$ subgroup of $SU(4)_L \times SU(4)_R$, the would-be mass term $\bar{Q}_L Q_R$ is simply not gauge invariant: Q_L is charged under $SU(N)_1$, Q_R under $SU(N)_2$, and these are now genuinely different groups. The lowest PQ-violating operator allowed by the full gauge symmetry is the four-fermion condensate $\bar{Q}_L \chi_R \bar{\chi}_L Q_R$, of dimension six rather than three. **The quality of the PQ symmetry** has thus improved from “arbitrarily bad” to “parametrically suppressed”: the two-site model realizes the mechanism, and each additional interme-

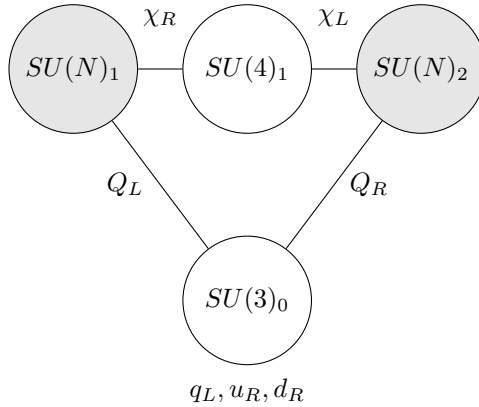


Figure 2.5: Moose diagram of the deconstructed Choi–Kim extension. Link fermions sit in bifundamental representations of adjacent gauge groups, and the Standard Model quarks are charged solely under $SU(3)_0$.

mediate site raises the dimension of the leading PQ-violating operator by three, to $3(d+1)$ for d intermediate sites. Meeting the bound of Section 2.3.5, $n \geq 9$ in operator dimension for $f_a \sim 10^9$ GeV, therefore requires a chain of moderate length; the quality problem is solved by structure, at the cost of replication.

The chiral gauge structure has addressed the quality and hierarchy problems of the model. However, it does nothing to reduce **the domain-wall number** of Choi–Kim model, $N_{\text{DW}} = N$. Remarkably, the use of another kind of discrete gauge structure can reduce this domain-wall number, through the Lazarides–Shafi mechanism to which we finally turn.

2.6.3 The Lazarides–Shafi mechanism

The domain wall problem of Section 2.5.5 is a statement about the number of *physically distinct* vacua of the axion potential, not merely about the bare periodicity of the $U(1)_{\text{PQ}}$ charge lattice computed in Section 2.3.4. Lazarides and Shafi observed that these two countings need not agree [23]. After the QCD anomaly has done its work, the surviving symmetry is the discrete $\mathbb{Z}_N$ subgroup of $U(1)_{\text{PQ}}$ identified in Section 2.3.4. Suppose this $\mathbb{Z}_N$ (or even a discrete subgroup of it) coincides with, or sits inside, some *continuous gauge symmetry* already present in the theory. Then field configurations related by the discrete transformation are not merely degenerate in energy but *gauge equivalent*: literally the same physical state, related by a redundancy of description rather than a genuine distinct vacuum. Figure 2.6 illustrates the idea for the simplest nontrivial case, $N = 3$. Identifying and exploiting such a coincidence reduces the number of physically inequivalent vacua, and hence N_{DW} , below the naive anomaly-coefficient count, potentially all the way to one.

This mechanism is the single thread connecting every other topic of the chapter

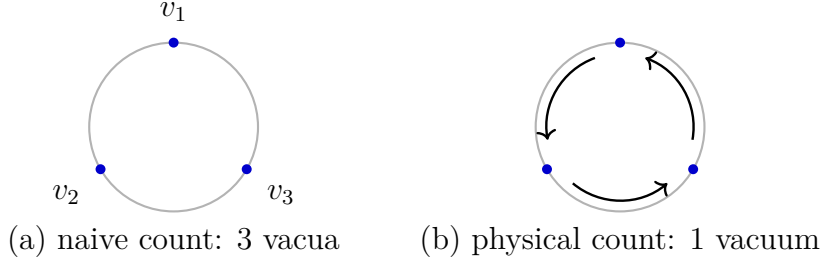


Figure 2.6: The Lazarides–Shafi mechanism, illustrated for the simplest nontrivial case. **(a)** The QCD anomaly alone leaves $N = 3$ discrete, equally spaced minima v_1, v_2, v_3 on the $U(1)_{\text{PQ}}$ vacuum circle – the naive counting of Section 2.3.4, which by itself would give $N_{\text{DW}} = 3$. **(b)** If a continuous gauge symmetry already present in the theory rotates $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_1$ into one another, the three configurations are not physically distinct vacua but a single gauge orbit. Only $N_{\text{DW}} = 1$ physically inequivalent vacuum survives, and the domain-wall problem of Section 2.5.5 is absent identically.

into the research program pursued in the remainder of this thesis. Composite axion models (Section 2.6) are built from confining gauge theories, which always possess a center symmetry or other unbroken discrete gauge subgroup of exactly the kind the Lazarides–Shafi mechanism requires. The domain wall problem (Section 2.5.5) is the specific cosmological catastrophe that makes finding such an identification not an aesthetic nicety but a strict prerequisite for any composite axion model that hopes to support a post-inflationary cosmology. And quantifying precisely how large an identification a given confining gauge group can supply, and whether it is large enough to reach $N_{\text{DW}} = 1$, requires a systematic, general formula for counting gauge-equivalent vacua, which is the first order of business of Chapter 3.

2.7 Summary

This chapter has traced a logical chain from the structural vacuum of non-Abelian gauge theories to the concrete model-building program of the rest of Part I. Non-trivial gauge topology forces the existence of θ -vacua (Section 2.1); quarks lock the vacuum angle to the flavor sector, producing the physical parameter $\bar{\theta}$ (Section 2.1.3); chiral perturbation theory carries $\bar{\theta}$ down to the neutron electric dipole moment, yielding the extraordinarily small empirical bound $|\bar{\theta}| \lesssim 10^{-10}$ (Section 2.2) that constitutes the strong CP problem. The Peccei–Quinn mechanism resolves this by promoting $\bar{\theta}$ to a dynamical field, the axion, whose relaxation to the CP-conserving minimum is automatic and whose low-energy phenomenology is controlled by the single scale f_a (Section 2.3).

Every axion, elementary or composite, must then survive three independent gauntlets. Astrophysics and the laboratory bound f_a from below, at roughly 10^8 – 10^9 GeV, through the stellar-cooling and supernova arguments of Section 2.4. Cosmology puts an upper bound on f_a from the abundance of axions with an undetermined initial angle in the pre-inflationary scenario or without in the post-inflationary scenario, and in the latter case cosmology also forces the theory to confront strings, domain walls, and the wall problem of Section 2.5.5. And quantum gravity is expected to bound f_a indirectly, by demanding that the leading PQ-violating operator have dimension $n \geq 9$ for $f_a \sim 10^9$ GeV, on pain of reopening the very problem the axion was invented to solve (Section 2.3.5).

Compositeness, developed in Section 2.6, addresses two of these three challenges by construction: the Choi–Kim model generates f_a dynamically, solving the hierarchy problem, and the chiral moose forbids the dangerous PQ-violating mass term by gauge invariance, solving the quality problem down to whatever operator dimension the chain length demands. Neither construction, on its own, touches the domain-wall number: both inherit $N_{\text{DW}} = N$ directly from the confining gauge group, unprotected and generically large. The Lazarides–Shafi mechanism is the missing piece, and it is a genuinely different kind of solution: not a suppression, but an exact identification of naively-distinct vacua by an unbroken gauge symmetry the confining sector already possesses.

Chapter 3 inherits exactly three ingredients from this review: the operator-dimension threshold $n \geq 9$ that any UV completion must meet for a phenomenologically viable f_a (Section 2.3.5); the requirement $N_{\text{DW}} = 1$, without which no composite construction can support a post-inflationary cosmology (Section 2.5.5); and the lower bound $f_a \gtrsim 10^8$ GeV from astrophysics (Section 2.4), against which every candidate construction must ultimately be checked.

Chapter 3

Composite Axion Models with Domain Wall Number One

Chapter 2 closed by identifying the Lazarides–Shafi mechanism, the possibility that a discrete subgroup of $U(1)_{\text{PQ}}$, left unbroken by the QCD anomaly, coincides with a subgroup of a gauge symmetry, as the one mechanism capable of reducing the domain wall number N_{DW} below the naive anomaly-coefficient count, and hence of rescuing a post-inflationary cosmology for a composite axion. This chapter, which is based on a published collaboration [1], makes that idea precise and constructive. Section 3.1 states the general formula used to compute N_{DW} for an arbitrary confining gauge theory. Sections 3.2 and 3.3 apply it to two structurally different families of composite axion models, in each case reviewing the relevant pre-existing construction before presenting our own extension to $N_{\text{DW}} = 1$. Section 3.4 then confronts both resulting models with a second, independent cosmological obstacle, the stable, electromagnetically charged composite relics they predict, and shows how a brief secondary inflationary epoch reconciles them with the post-inflationary scenario that motivated the whole chapter.

3.1 A master formula for the domain wall number

The domain wall number N_{DW} is the single quantity that decides whether a composite axion can survive a post-inflationary cosmology, so computing it for a given confining theory is the first technical task of this chapter. In this section we derive a compact master formula, Eq. (3.5), that writes N_{DW} as the greatest common divisor of a set of identification coefficients, one for the Peccei–Quinn symmetry itself and one for possible unbroken gauged discrete symmetries that the theory happens to possess, which is exactly the freedom exploited by the Lazarides–Shafi mecha-

nism [23]. We build the formula in two steps: the identification of θ -vacua forced by a trivial $U(1)_{\text{PQ}}$ rotation, and the further identifications supplied by any unbroken discrete symmetry, most importantly the center of the confining gauge group, that reduce this naive count.

3.1.1 Identification by a trivial rotation

Consider a confining gauge theory whose elementary fermions ψ_j carry Peccei–Quinn charges p_j , normalized so that they are mutually prime integers, $\gcd(p_1, \dots, p_n) = 1$ (any common factor could simply be absorbed into the definition of the charge unit and carries no physical content). The transformation

$$\psi_j \longrightarrow \psi_j \times \exp(i 2\pi p_j) \quad (3.1)$$

is, term by term, manifestly the identity on every field: each ψ_j is simply multiplied by $e^{2\pi i p_j} = 1$. The action is therefore exactly invariant under (3.1). Yet, the transformation (3.1) is at the same time a finite $U(1)_{\text{PQ}}$ rotation, by angle $\alpha = 2\pi$, of exactly the kind that shifted the θ -vacuum in Section 2.1.2. Because $U(1)_{\text{PQ}}$ is anomalous, performing this transformation inside the path integral picks up the anomalous Jacobian of Section 2.1.2 for *every* confining or QCD-like gauge group G under which the ψ_j are charged,

$$\theta \longrightarrow \theta + 2\pi k_{\text{PQ}}, \quad k_{\text{PQ}} = \sum_j p_j \times \dim(r_j) \times 2 I_G(s_j), \quad (3.2)$$

where $\dim(r_j)$ is the dimension of the representation of ψ_j under the *confining* group and $I_G(s_j)$ is its Dynkin index under the color group G , normalized to 1/2 for the fundamental representation.

Since (3.1) is the identity transformation, and since the θ -dependence of the physics is exactly 2π -periodic (Section 2.1.2), the only way to reconcile the statement that the action is invariant under (3.1) with the statement that the action depends on θ , which has just shifted by $2\pi k_{\text{PQ}}$, is to conclude that

$$\theta \equiv \theta + 2\pi k_{\text{PQ}} \quad (3.3)$$

are physically identical configurations, literally the same point in field space, reached by a transformation that does nothing at all to any field. This is the precise sense in which k_{PQ} counts the number of vacua of the axion potential: it is the number of distinct copies of the fundamental 2π period of θ that the trivial rotation (3.1)

sweeps out before returning to its starting point, and hence it recovers the naive guess one might make for N_{DW} , equal to the anomaly coefficient $\mathcal{A}(U(1)_{\text{PQ}} \cdot G^2)$.

3.1.2 Identification by an unbroken gauge symmetry

The same logic applies when there is a gauge symmetry which intersects with the PQ symmetry, either a center symmetry or a subgroup of an Abelian gauge symmetry. We will be interested in the discrete intersection, so we will use a discrete gauge symmetry like the center symmetry. The gauge redundancy of the discrete symmetry offers another identification for θ . If the center is given by $\mathbb{Z}_N$, then the fermion charges will be given by c_j/N as $\psi_j \rightarrow \psi_j e^{2\pi i c_j/N}$. Repeating the argument of (3.1)–(3.3) for this transformation gives the identification,

$$\theta \equiv \theta + 2\pi k_{\mathcal{G}}, \quad k_{\mathcal{G}} = \sum_j \frac{c_j}{N} \times \dim(r_j) \times 2 I_G(s_j). \quad (3.4)$$

Both (3.3) and the center identification (3.4) must hold simultaneously, and more generally so must any further identification supplied by other gauge symmetries that the model happens to possess, which is exactly the freedom exploited by the Lazarides–Shafi mechanism of Section 2.6.3. The number of vacua that survive *all* such identifications at once is the greatest common divisor of every independent identification coefficient [1],

$$\boxed{N_{\text{DW}} = \text{gcd} \left(k_{\text{PQ}}, k_{\mathcal{G}}, \dots \right)}, \quad (3.5)$$

which is the master formula used throughout the remainder of this chapter. The rest of this chapter is, in essence, a search for confining gauge theories generous enough in their available center and discrete symmetries that (3.5) can be driven all the way down to $N_{\text{DW}} = 1$, the borderline, maximally favorable case in which the string-wall network is expected to be unstable and to collapse, exactly as discussed in Section 2.5.5.

3.2 Automatic PQ symmetry

The first of our two model families realizes $U(1)_{\text{PQ}}$ as the baryon number of a chiral confining sector, an idea that makes the Lazarides–Shafi identification of Section 3.1 automatic rather than accidental. We begin by reviewing the original single-sector proposal [25] and showing, with the master formula, that it stalls at $N_{\text{DW}} = 2$. We then present our own two-sector chiral moose [1], which reaches $N_{\text{DW}} = 1$, and

close with a critical examination of the strong-coupling assumption on which both constructions rest.

3.2.1 The original automatic model

The composite constructions reviewed in Section 2.6 all embedded $U(1)_{\text{PQ}}$ inside an *approximate flavor* symmetry of a vector-like confining theory. An entirely different strategy [25] is to identify $U(1)_{\text{PQ}}$ with the *baryon number* of the confining gauge theory itself. The appeal of this identification is immediate from the discussion of Section 3.1: the center of the confining $SU(N)$ acts non-trivially on baryon number, so identifying $U(1)_{\text{PQ}}$ with $U(1)_B$ automatically guarantees that the center symmetry supplies a non-trivial k_G in (3.5), with no further model-building required to arrange the coincidence. This is exactly the kind of accidental alignment the Lazarides–Shafi mechanism asks for. There is, however, an immediate obstruction: vector-like confining theories, of the kind used for Choi–Kim, do not spontaneously break baryon number at all (Vafa–Witten theorem) [111, 112], so realizing $U(1)_{\text{PQ}} = U(1)_B$ as a *spontaneously broken*, axion-bearing symmetry requires the confining sector to be chiral.¹

The minimal realization is a chiral $SU(5)$ gauge theory, with Weyl fermions $\chi \sim \mathbf{10}$ and $\psi \sim \bar{\mathbf{5}}$, the same anomaly-free combination familiar from $SU(5)$ grand unification [113], additionally placed in a real representation $\mathbf{r}$ of $SU(3)_{\text{QCD}}$ (the minimal choice being $\mathbf{r} = \mathbf{3} \oplus \bar{\mathbf{3}}$) so that the mixed $SU(5)$ – $SU(3)$ and the pure $SU(3)$ gauge anomalies cancel automatically, as in Table 3.1.

	$[SU(5)]$	$SU(3)_c$	$U(1)_{\text{PQ}} (= U(1)_B)$	$\mathbb{Z}_5$
χ	$\mathbf{10}$	$\mathbf{3} \oplus \bar{\mathbf{3}}$	+1	+2
ψ	$\bar{\mathbf{5}}$	$\mathbf{3} \oplus \bar{\mathbf{3}}$	−3	−1

Table 3.1: Matter content of the automatic PQ model. $U(1)_{\text{PQ}}$ is identified with the baryon symmetry of the $SU(5)$ theory, and $\mathbb{Z}_5$ is the center gauge symmetry.

The condensates that break global symmetries. This model, being a chiral gauge theory, does not possess the two-fermion gauge invariant condensates that QCD has. Therefore, which fermion operators gets a condensate must be assumed. this assumption is a dynamical question we re-examine critically in Section 3.2.3.

¹The price we pay for this is to lose the comfort we have in QCD-like theories and the ability to state results that have been compared against natural realizations and lattice results.

The four-fermion operators $\chi\chi\chi\psi$ is the lowest dimension gauge-invariant operator available to this theory: being built out of four fermions, it can acquire a Lorentz-invariant vacuum expectation value. It is energetically favored that this VEV be invariant under QCD, hence the global symmetry of the model $SU(6)_\chi \times SU(6)_\psi$ is spontaneously broken to a subgroup $H \supseteq SO(6) \supset SU(3)_c$. However, this operator is neutral under PQ symmetry, so this VEV does not spontaneously break PQ and does not provide the necessary axion.

In order to break the baryon symmetry in this mode, we have to resort to the six-fermion operator $\langle \chi\psi\psi\chi\psi\psi \rangle \neq 0$, built from a pair of baryon-like composites $\chi\psi\psi$, each of them has a baryon charge -5 , then $U(1)_{\text{PQ}} = U(1)_B$ is spontaneously broken and the composite axion appears as the phase of this condensate. Applying (3.2) to Table 3.1,

$$k_{PQ} = \underbrace{(-3)}_{p_\psi} \times \underbrace{5}_{\dim(\mathbf{5})_{SU(5)}} \times \underbrace{2}_{2I_{SU(3)}(\mathbf{3} \oplus \bar{\mathbf{3}})} + \underbrace{(+1)}_{p_\chi} \times \underbrace{10}_{\dim(\mathbf{10})_{SU(5)}} \times \underbrace{2}_{2I_{SU(3)}(\mathbf{3} \oplus \bar{\mathbf{3}})} = -10, \quad (3.6)$$

naively suggesting ten degenerate vacua. The unbroken $\mathbb{Z}_5$ center of $SU(5)$, however, also acts on the fermions such that $\chi \rightarrow e^{2\pi i \cdot (2k/5)}\chi$, $\psi \rightarrow e^{2\pi i \cdot (-k/5)}\psi$ for $k = 0, \dots, 4$, the N -alities of $\mathbf{10}$ and $\bar{\mathbf{5}}$ under the $\mathbb{Z}_5$ center being 2 and -1 respectively, giving, via (3.4),

$$k_{\mathcal{G}} = \frac{2}{5} \times 10 \times 2 + \frac{-1}{5} \times 5 \times 2 = 6, \quad (3.7)$$

so that both identifications ultimately give $N_{\text{DW}} = \gcd(k_{PQ}, k_{\mathcal{G}}) = \gcd(10, 6) = 2$. As a first non-trivial test of Eq. (3.5), it reproduces the result of Ref. [25], obtained there by direct enumeration of gauge-inequivalent vacua.

The domain wall problem. By exactly the Lazarides–Shafi mechanism this model was designed around, the number of inequivalent vacua decreased from ten to two, but not all the way to one. The obstruction is structural: canceling the $SU(3)$ anomaly with a *single* $SU(5)$ multiplet forced χ, ψ into the real representation $\mathbf{3} \oplus \bar{\mathbf{3}}$, and this particular choice of QCD representation ties k_{PQ} to an even integer that cannot be brought to coprimality with $k_{\mathcal{G}}$ no matter how the $\mathbb{Z}_5$ center is exploited. The original proposal accepts the resulting $N_{\text{DW}} = 2$ domain wall problem and falls back on the standard, pre-inflationary escape of Section 2.5.5.

The quality of PQ symmetry. The leading operators that explicitly break PQ are the same operators that spontaneously break it. So the first contribution that can affect the quality of the PQ symmetry of this model comes from a dimension-nine operator saturating the bound $d \geq 9$ derived in Section 2.3.5 for $f_a \sim 10^9$ GeV.

3.2.2 A chiral moose realizing $N_{\text{DW}} = 1$ [1]

The obstruction identified in Section 3.2.1 is specific to needing ψ, χ to have fundamental and antifundamental content in color. The natural fix, in direct analogy with the moose construction that generalized the Choi–Kim model in Section 2.6.2, is to deconstruct the automatic-PQ idea across *two* confining $SU(5)$ sectors rather than one. This allows the content of one sector $SU(5)_1$, which will be used to carry the PQ charges, to be only fundamental in color; this reduces the factor $2I_G(SU(3))$ from 2 to 1, so k_{PQ} is no longer forced even. The QCD gauge anomaly that they will introduce will be canceled by the content of the other sector $SU(5)_2$.

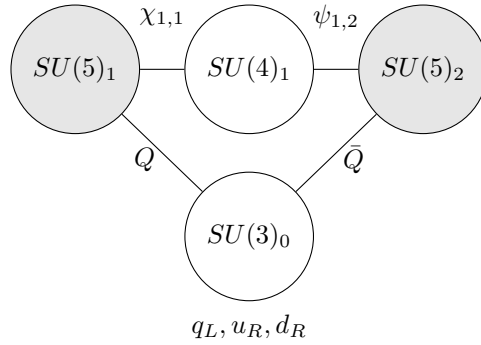


Figure 3.1: Moose diagram of the chiral $SU(5)$ moose model. $Q, \bar{Q}$ link the two confining $SU(5)$ sectors to $SU(3)_0$, while $\chi_{1,1}, \psi_{1,2}$ link them to $SU(4)_1$.

Field	$[SU(3)_0]$	$SU(4)_L$	$[SU(5)_1]$	$[SU(4)_1]$	$[SU(5)_2]$	$SU(4)_R$	$U(1)_{\text{PQ}}$
Q	3	4	$\bar{5}$				-3
η	1		$\bar{5}$				-3
$\chi_{1,1}$			10	4			1
$\psi_{1,2}$				$\bar{4}$	10		0
$\bar{Q}$	$\bar{3}$				$\bar{5}$	$\bar{4}$	0
$\bar{\eta}$	1				$\bar{5}$		0

Table 3.2: Matter content of the chiral $SU(5)$ moose model.

The gauge group of the model is $[SU(3)_0] \times [SU(5)_1] \times [SU(4)_1] \times [SU(5)_2]$ (Figure 3.1, Table 3.2). The two confining $SU(5)$ sectors have each four flavors of chiral fermions in $\bar{5} \oplus 10$ representation, and consequently, there is an approximate flavor symmetry $SU(4)_{\bar{5}} \times SU(4)_{10}$ per sector. A diagonal $SU(3)_0$ subgroup of the two $SU(4)_{\bar{5}}$ factors, and a diagonal $SU(4)_1$ subgroup of the two $SU(4)_{10}$ factors, are weakly gauged, linking the two $SU(5)$ sectors together exactly as the two $SU(N)$

sectors of Section 2.6.2 were linked. The QCD gauge group is the unbroken subgroup of $SU(3)_0 \times SU(4)_1$.

As in Section 3.2.1, we assume the four-fermion condensates $\langle \chi_{1,1} \chi_{1,1} \chi_{1,1} Q \rangle$ and $\langle \psi_{1,2} \psi_{1,2} \psi_{1,2} \bar{Q} \rangle$ form in each sector, breaking $SU(3)_0 \times SU(4)_1 \rightarrow SU(3)$, as well as the PQ-charged six-fermion condensates $QQ\chi_{1,1}\psi_{1,2}\bar{Q}\bar{Q}$, and the analogues that replace one or both pairs of $(Q, \bar{Q})$ with $(\eta, \bar{\eta})$, that break the PQ symmetry and produces the composite axion.

After integrating out the heavy gauge bosons of $SU(3)_0 \times SU(4)_1$, the tree level matching conditions gives $g_s^{-2} = g_3^{-2} + g_4^{-2}$. The surviving low-energy θ -angle of QCD is the appropriately diagonal $\theta = \theta_3 + \theta_4$ since the instanton of the diagonal group is an instanton in each factor simultaneously. The identification (3.2) of the master formula must be applied to *both* factors and added,

$$\begin{aligned} k_{PQ}|_{SU(3)_0} &= \underbrace{(-3)}_{p_Q} \times \underbrace{5}_{\dim(\mathbf{5})_{SU(5)_1}} \times \underbrace{1}_{2I_{SU(3)_0}(\mathbf{3})} , \\ k_{PQ}|_{SU(4)_1} &= \underbrace{(+1)}_{p_{\chi_{1,1}}} \times \underbrace{10}_{\dim(\mathbf{10})_{SU(5)_1}} \times \underbrace{1}_{2I_{SU(4)_1}(\mathbf{4})} , \\ k_{PQ} &= k_{PQ}|_{SU(3)_0} + k_{PQ}|_{SU(4)_1} = -15 + 10 = -5 . \end{aligned} \quad (3.8)$$

The domain-wall number. Since only $Q, \chi_{1,1}$ are charged under $U(1)_{\text{PQ}}$, only the center symmetry of $SU(5)_1$ allows for identification of vacua. Using Eq. (3.4), we get

$$k_{\mathcal{G}}|_{SU(5)_1} = \underbrace{-\frac{1}{5} \times 5 \times 1}_Q + \underbrace{\frac{2}{5} \times 10 \times 1}_{\chi_{1,1}} = -1 + 4 = 3 . \quad (3.9)$$

This identification coefficient $k_{\mathcal{G}} = 3$ [1] lets the domain wall number be

$$\boxed{N_{\text{DW}} = \gcd(k_{PQ}, k_{\mathcal{G}}) = \gcd(5, 3) = 1 .} \quad (3.10)$$

This is the central result of the section. Doubling the confining sector removes the obstruction that left the single-sector automatic model of Section 3.2.1 stuck at $N_{\text{DW}} = 2$: with the PQ-charged fermions in fundamental representation rather than in $\mathbf{3} \oplus \bar{\mathbf{3}}$, k_{PQ} is no longer forced to be even, and the $\mathbb{Z}_5$ center of $SU(5)_1$ then supplies an identification coprime to it.

The quality of PQ symmetry. This model has exactly the same quality as the original model demonstrated in the previous section. The leading operators that explicitly break PQ are the nine-dimensional operators, $QQ\chi_{1,1}\psi_{1,2}\bar{Q}\bar{Q}$ and its

analogues.²

At this point the construction is complete at the level of symmetry and anomaly bookkeeping. There is, however, a separate dynamical question: does the assumed chiral $SU(5)$ theory actually realize the assumed pattern on which the construction relies?

3.2.3 Critique: does the assumed confinement pattern actually occur?

Everything in Sections 3.2.1–3.2.2 rests on an assumption about strongly coupled, chiral $SU(5)$ dynamics that cannot be checked in perturbation theory: the spontaneous flavor symmetry breaking is realized through specific four- and six-fermion condensates, which were posited above, while the confining group $SU(5)$ remains unbroken through this development. Then, and only then, the axion is produced due to the baryon symmetry breaking from the six-fermion condensate. This is far from guaranteed, and it deserves to be examined critically rather than simply assumed, since the entire phenomenological payoff of Sections 3.2.1–3.2.2 depends on it.

't Hooft anomaly matching does not decide the question. A first, model-independent consistency check is whether the high-energy 't Hooft anomalies of $U(1)_B$ [117], namely $\mathcal{A}(U(1)_B^3)$ and $\mathcal{A}(U(1)_B \cdot \text{grav}^2)$, can be matched in the infrared. For the single-sector model of Table 3.1, they can: a single massless composite baryon $\chi\psi\psi$, of baryon charge -5 , reproduces both anomalies exactly. This means that a scenario in which $U(1)_B$ remains *entirely unbroken*, with no axion at all and the spectrum populated instead by a massless, baryon-charged composite fermion, is also a logically consistent possibility, on equal footing with the symmetry-breaking scenario assumed above. Anomaly matching alone, in other words, cannot distinguish between these two qualitatively different infrared phases; it is a necessary consistency condition on either, not a selection principle between them.

The MAC analysis favors a different pattern altogether. A more quantitative, but still heuristic, handle comes from the most-attractive-channel (MAC) hypothesis [118, 119]. A condensation channel $r_1 \otimes r_1 \rightarrow s$ is energetically favored

²The axion can acquire a potential from small instanton effects associated with the broken part of $SU(3)_0 \times SU(4)_1$ [114, 115]. Recent work on a related two-sector composite axion construction identifies a symmetry condition under which such contributions remain parametrically small [116]; we have not checked whether it is satisfied here. The impact of the small instanton potential on the domain wall problem in composite axion models is left for future work.

to form if its associated change in quadratic Casimir,

$$\Delta C = C(r_1) + C(r_2) - C(s), \quad (3.11)$$

from single-gauge-boson exchange is maximal. For $\chi \sim \mathbf{10}$ and $\psi \sim \bar{\mathbf{5}}$ of $SU(5)$, with $C(\mathbf{10}) = 36/5$ and $C(\bar{\mathbf{5}}) = 12/5$,

$$\mathbf{10} \otimes \mathbf{10} \rightarrow \bar{\mathbf{5}} : \Delta C = \frac{60}{5}, \quad \bar{\mathbf{5}} \otimes \mathbf{10} \rightarrow \mathbf{5} : \Delta C = \frac{36}{5}, \quad (3.12)$$

we can identify the *two*-fermion channel $\mathbf{10} \otimes \mathbf{10} \rightarrow \bar{\mathbf{5}}$ as the MAC, contrary to the four- and six-fermion condensates assumed in Section 3.2.1.

A non-zero $\langle \chi \chi \rangle$ condensate of this kind would *Higgs* $SU(5) \rightarrow SU(4)$ directly, via $\langle \epsilon^{abcde} \chi_{bc} \chi_{de} \rangle \propto \delta^{a5}$, rather than leave $SU(5)$ confined and unbroken: the theory would become a strongly coupled $SU(4)$ gauge theory below this Higgsing, with the original $\mathbf{10}, \bar{\mathbf{5}}$ multiplets decomposing under the unbroken $SU(4)$ as $\mathbf{10} \rightarrow \mathbf{6}_2 \oplus \mathbf{4}_{-3}$ and $\bar{\mathbf{5}} \rightarrow \bar{\mathbf{4}}_{-1} \oplus \mathbf{1}_4$ (where subscripts denote the charge under the residual $U(1) \subset SU(5)/SU(4)$, which mixes with baryon number into a combination $B' = B - R/2$).

The Majorana fermions in $\mathbf{6}_0$ acquire mass directly from the Higgsing condensate. The Dirac fermion $(\mathbf{4}, \bar{\mathbf{4}})$ condenses through ordinary QCD-like $SU(4)$ confinement. The last surviving singlet $\mathbf{1}_{-5}$ happens to carry exactly the baryon charge of the composite state that would have matched the 't Hooft anomalies in the alternative, unbroken- $U(1)_B$ scenario above. This is a non-trivial, reassuring consistency check between the two pictures, even though they describe physically inequivalent infrared phases: one with a Higgsed $SU(5)$ and no axion at all in this minimal field content, the other with confined $SU(5)$ and the composite axion of Section 3.2.1.

Functional renormalization group results. This concern is no longer purely a matter of weak-coupling extrapolation. A recent functional renormalization group study of Georgi–Glashow-type chiral gauge theories [120], tracking the renormalization group flow of the relevant four-fermion operators without relying on the perturbative MAC heuristic logic, finds a genuine, dynamically generated resonance in precisely the two-fermion, antisymmetric-representation channel identified by the MAC analysis above, for $N_c < 11$, squarely including the relevant case $N_c = 5$ used throughout this section. This is a substantially stronger statement than the MAC analysis alone: it indicates that the symmetry-breaking pattern actually generated by the renormalization-group flow of the theory, not merely the pattern favored by a static, weak-coupling estimate, is the Higgsing pattern, in direct tension with the confinement with broken $U(1)_B$ assumption underlying Sections 3.2.1–3.2.2.

Assessment. Neither the MAC analysis nor the functional renormalization group result constitutes a rigorous proof that the automatic-PQ condensation pattern is wrong: both are, in the end, extrapolations of weak-coupling or truncated-flow intuition into a regime where the gauge coupling is, by construction, of order unity, and a fully non-perturbative, lattice-based determination of the true infrared phase of this chiral $SU(5)$ theory does not yet exist. What can be said with confidence is that the central dynamical assumption of Sections 3.2.1–3.2.2 is a genuinely open question under active theoretical scrutiny, not a settled technical detail, and that the best available quantitative tools beyond the original proposal’s own assumption point, if anything, the other way. We proceed in the rest of this chapter exactly as the original proposal does, on the assumption that the stated condensation pattern is realized, because establishing the existence of an $N_{\text{DW}} = 1$ solution *conditional on that assumption* is already a non-trivial and useful result, and because doing so does not require resolving the underlying non-perturbative dynamics, which is left for future lattice work.

3.3 Gauged PQ symmetry

The first construction solved the domain-wall problem through an unbroken discrete symmetry embedded in the confining dynamics. We now pursue a qualitatively different route: to make the PQ symmetry intersect with a gauge $U(1)$ symmetry. Because a single anomalous $U(1)_{\text{PQ}}$ cannot itself be gauged (its mixed anomaly with QCD is precisely what makes it a PQ symmetry), so one needs a second $U(1)$ so that an anomaly-free orthogonal combination exists to gauge.

We first explain why gauging a combination of the two PQ symmetries solves the quality problem by forbidding low-dimensional operators, then we review what is known about the cosmological string networks that a broken gauge symmetry produces. Finally, we build an explicit composite realization from two gauged Choi–Kim sectors and show that it, too, achieves $N_{\text{DW}} = 1$.

3.3.1 The Barr–Seckel idea

An entirely different way to suppress dangerous Planck-suppressed PQ-violating operators was proposed early on by Barr and Seckel [26, 27]: rather than relying on an accidental symmetry of the matter content, *gauge* an orthogonal combination of the Peccei–Quinn symmetry. If $U(1)_{\text{PQ}}$ is linked to a local gauge symmetry, unbroken down to the PQ scale, then no operator that violates the gauge symmetry can appear in the Lagrangian at all.

This gauged $U(1)$ does not come for free. It requires the duplication of the sector that gives rise to PQ symmetry. In addition, having two sectors introduces two types of topological defects, i.e., strings. Not all of the strings are going to be global with divergent gradient energy as those introduced in Section 2.5.3. A *local* (gauge) string has its long-range gradient energy screened by the massive gauge boson [121]. In addition, such a string will not see the axion direction, so it attaches to no axion domain walls, i.e., $N_{\text{DW}} = 0$. The cosmological string network produced when a gauged PQ symmetry breaks is therefore, in general, qualitatively different from the global-string networks discussed throughout Chapter 2, and its consequences for the domain wall problem of Section 2.5.5 are not automatically the same.

3.3.2 Local versus global strings

Whether a network formed by breaking a gauged PQ symmetry ends up dominated by local (gauge) strings or by global strings is a genuinely dynamical question, and it was addressed numerically in Refs. [122, 123] for a simulation of the cosmological formation of strings in a model with two Abelian symmetries that get broken at comparable scales, and a specific combination of the two symmetries is gauged.

These studies found that the network is *not* a uniform population of strings all carrying the same domain wall number. The dominant population, by a clear margin, consists of local strings, which carry $N_{\text{DW}} = 0$ in the present context: such a string is not the boundary of any domain wall, since there is no surviving global symmetry for a wall to interpolate across. The next most common population consists of a type of global strings with $N_{\text{DW}} = 1$, which, as established in Section 2.5.5, are themselves cosmologically safe, collapsing on their own even in a fully post-inflationary setting.

This result does not, by itself, eliminate every dangerous string from the network: a subdominant population of strings with $N_{\text{DW}} \geq 2$ is still produced in the simulation, and these remain just as catastrophic, in principle, as the uniform $N_{\text{DW}} = N$ network of an ordinary global-symmetry-breaking model. What the gauging of PQ symmetry achieves is therefore not a complete, first-principles solution to the domain wall problem by itself, but a substantial *improvement in the odds*: whereas an ordinary composite axion model with $N_{\text{DW}} = N > 1$ produces a network in which *every* string carries the same dangerous winding number N , a gauged PQ model produces a network in which the large majority of strings are either entirely harmless ($N_{\text{DW}} = 0$) or self-resolving ($N_{\text{DW}} = 1$). This observation is the cosmological motivation for constructing an explicit *composite* realization of the gauged PQ idea, to which we turn next.

3.3.3 A composite two-sector realization [1]

We now construct a genuinely composite version of the Barr–Seckel idea, combining *two* otherwise-ordinary, vector-like Choi–Kim sectors (Section 2.6.1), with confining groups $SU(N_A)$ and $SU(N_B)$, $N_A \neq N_B$, coupled only through an additional gauged $U(1)$ under which both sectors’ quarks are charged. Crucially, this $U(1)$ gauging is precisely what forbids the dangerous Choi–Kim mass terms $\bar{Q}_L Q_R$ and $\bar{\eta}_L \eta_R$ of Section 2.6.1: both fields are charged under the new $U(1)$ in a way that makes the theory chiral with respect to it, even though each $SU(N_A)$ (or $SU(N_B)$) sector remains, by itself, perfectly vector-like.

	$[SU(3)]$	$[SU(N_A)]$	$[U(1)_{\text{gPQ}}]$	$U(1)_{\text{PQA}}$
Q_L	3	N_A	$2N_B$	1
η_L	1	N_A	$2N_B$	−3
Q_R	3	N_A	$5N_B$	0
η_R	1	N_A	$−7N_B$	0

	$[SU(3)]$	$[SU(N_B)]$	$[U(1)_{\text{gPQ}}]$	$U(1)_{\text{PQB}}$
Q'_L	3	N_B	$−2N_A$	1
η'_L	1	N_B	$−2N_A$	−3
Q'_R	3	N_B	$−5N_A$	0
η'_R	1	N_B	$7N_A$	0

Table 3.3: Matter content and $U(1)_{\text{gPQ}}$ charge assignment of the gauged composite PQ model. The charge assignment follows the anomaly-free set of [124].

Each sector has its own anomalous PQ symmetry, denoted as $U(1)_{\text{PQ}_{1,2}}$ and defined as usual. The construction hinges on finding a particular linear combination $U(1)_{\text{gPQ}}$ of $U(1)_{\text{PQ}_{1,2}}$ that cancels all gauge anomalies. We choose the charges of fermions under $U(1)_{\text{gPQ}}$ such that its mixed gauge anomaly with *both* confining groups $SU(N_A), SU(N_B)$ cancels out separately. Since all $U(1)_{\text{gPQ}}$ -charged fermions transform in the fundamental of $SU(N)$, the gravitational and mixed anomalies are proportional,

$$\mathcal{A}(U(1)_{\text{gPQ}} \cdot \text{grav}^2) = N \cdot \mathcal{A}(U(1)_{\text{gPQ}} \cdot SU(N)^2) = 0. \quad (3.13)$$

The charge assignment, following the anomaly-free sets classified in Ref. [124], is given in Table 3.3; it is not the unique possible choice, but it is simple enough to

work with.³ The cubic gauge anomaly cancels per sector for this charge assignment,

$$\mathcal{A}(U(1)_{\text{gPQ}}^3) \propto 3 \times 2^3 + 2^3 - 3 \times 5^3 - (-7)^3 = 0. \quad (3.14)$$

The mixed anomaly of $U(1)_{\text{gPQ}}$ with QCD for one sector successfully cancels out the mixed anomaly of the other,

$$\mathcal{A}(U(1)_{\text{gPQ}} \cdot SU(3)^2) \propto N_A \times (2 - 5)N_B + N_B \times (-2 + 5)N_A = 0. \quad (3.15)$$

Identifying the axion. Below the two confinement scales, we are interested in the phase fields of the condensates $\langle \bar{Q}_L Q_R \rangle \sim e^{ia_A/f_A}$ and $\langle \bar{Q}'_L Q'_R \rangle \sim e^{-ia_B/f_B}$. Any $U(1)_{\text{gPQ}}$ gauge transformation with some angle α is a gauge redundancy of the configurations, and thus it is an identification of the phases,

$$\frac{a_A}{f_A} \equiv \frac{a_A}{f_A} + 3N_B\alpha, \quad \frac{a_B}{f_B} \equiv \frac{a_B}{f_B} + 3N_A\alpha \quad (3.16)$$

The axion direction is the gauge invariant combination of these two phase fields

$$\frac{a}{f_a} \equiv N_A \frac{a_A}{f_A} - N_B \frac{a_B}{f_B}, \quad f_a = \frac{f_A f_B}{\{N_B^2 f_A^2 + N_A^2 f_B^2\}^{1/2}}. \quad (3.17)$$

The phase along the direction orthogonal to the axion is eaten by the gauge boson.

The domain-wall number. To prove the possible existence of $N_{\text{DW}} = 1$ strings in this model, we consider a string with winding numbers (n_A, n_B) . Depending on the winding numbers, the physical axion possesses a definite domain wall number.

$$\Delta\theta_A = 2\pi n_A, \quad \Delta\theta_B = 2\pi n_B \implies \Delta\theta = 2\pi(N_A n_A - N_B n_B). \quad (3.18)$$

The number of domain walls that are going to attach themselves to a string of the winding (n_A, n_B) is

$$N_{\text{DW}} = N_A n_A - N_B n_B. \quad (3.19)$$

This has an immediate consequence: the strings whose windings are given by $n_A \propto N_B$, $n_B \propto N_A$ give $N_{\text{DW}} = 0$, so the axion field does not wind around such a string at all, and thus no domain wall needs to attach to this string in order to interpolate between different axion vacua. It is also possible to see the existence of strings with $N_{\text{DW}} = 1$ when N_A, N_B are coprime; we will take $(N_A = 3, N_B = 4)$ for a benchmark.

³The main result about the existence of $N_{\text{DW}} = 1$ strings is independent of the choice of gauged-PQ charges.

The spectrum of strings also contains strings with higher values of N_{DW} .

Cross-check in the limit of well-separated scales. The conclusion about strings with $N_{\text{DW}} = 0, 1$ may also be obtained, somewhat more laboriously but with the virtue of being directly comparable to Sections 3.2.1–3.2.2, from Eq. (3.5). Suppose, for definiteness, that $\Lambda_A \gg \Lambda_B$, so that $SU(N_A)$ confines first; its condensates $\langle \bar{Q}_L Q_R \rangle, \langle \bar{\eta}_L \eta_R \rangle$ carry $U(1)_{\text{gPQ}}$ charge $3N_B, -9N_B$, thus they break the gauge symmetry $U(1)_{\text{gPQ}} \rightarrow \mathbb{Z}_{3N_B}$. The surviving discrete subgroup acts on the fermions of $SU(N_B)$ sector as $Q'_L \rightarrow e^{2\pi i \cdot (-2N_A/3N_B)} Q'_L$, and likewise for the other fermions. This surviving $\mathbb{Z}_{3N_B}$ supplies its own identification of vacua, computed exactly as in Eq. (3.4) but for the discrete group $\mathbb{Z}_{3N_B}$ rather than a center symmetry,

$$k_{PQ}|_{SU(N_B)} = N_B, \quad k_{\mathcal{G}}|_{\mathbb{Z}_{3N_B}} = N_B \times \left(-\frac{2N_A}{3N_B} + \frac{5N_A}{3N_B} \right) = N_A. \quad (3.20)$$

The domain wall number of strings in this case reads as

$$N_{\text{DW}} = \gcd(k_{PQ}, k_{\mathcal{G}}) = \gcd(N_B, N_A) = 1. \quad (3.21)$$

The separation of the scales signifies that the two sectors do not confine simultaneously. As the A -sector confines first, it forms local-string structures, which are not relevant for the axion direction, and thus are the $N_{\text{DW}} = 0$ strings in our previous discussion. Later on, when the B -sector confines, it starts to form global axion strings, and the domain wall number with $N_{\text{DW}} = 1$ is given by Eq. (3.21) as predicted in the previous discussion for coprime (N_A, N_B) .

The simulation results [122, 123] find that gauging $U(1)_{\text{gPQ}}$ favors the formation of the local strings by a huge proportion (it was $\sim 60\%$ of the string population). Out of global strings, the population of $N_{\text{DW}} = 1$ strings was found to be of large proportion compared to $N_{\text{DW}} \geq 2$ strings (which were suppressed to below $\sim 1\%$). The dynamics can be seen to favor cosmologically benign strings, and it confirms the results argued for above.

The quality of PQ symmetry. We write the generic gauge-invariant operator that breaks PQ symmetry as

$$\mathcal{O} = (\bar{Q}_L Q_R)^{\alpha_A} (\bar{\eta}_L \eta_R)^{\beta_A} (\bar{Q}'_L Q'_R)^{\alpha_B} (\bar{\eta}'_L \eta'_R)^{\beta_B}, \quad (3.22)$$

where the exponent $\alpha_{A,B}, \beta_{A,B} \in \mathbb{Z}$.⁴ The gauge invariance of this operator under $SU(3), SU(N_A), SU(N_B)$ is evident; however, the gauge-invariance under $U(1)_{\text{gPQ}}$ requires that

$$3N_B(\alpha_A - 3\beta_A) - 3N_A(\alpha_B - 3\beta_B) = 0. \quad (3.23)$$

Thus, it is necessary that $\alpha_A - 3\beta_A \propto N_A$, $\alpha_B - 3\beta_B \propto N_B$ with the same proportionality constant. The PQ charge of this operator is obtained by applying

$$Q_L(\eta_L) \rightarrow e^{-ip_A a_A/f_A} Q_L(\eta_L), \quad Q'_L(\eta'_L) \rightarrow e^{+ip_B a_B/f_B} Q'_L(\eta'_L), \quad p_{Q(\eta)} = 1(-3). \quad (3.24)$$

Thus, the phase of the operator in (3.22) is

$$\arg \mathcal{O} = (\alpha_A - 3\beta_A) \frac{a_A}{f_A} - (\alpha_B - 3\beta_B) \frac{a_B}{f_B} \propto \frac{a}{f_a}, \quad (3.25)$$

i.e., it breaks PQ symmetry. The leading operator⁵ that breaks PQ symmetry minimizes $(|\alpha_A| + |\beta_A|) + (|\alpha_B| + |\beta_B|)$ subject to the conditions from gauge invariance. For our benchmark, this operator is given by the dimension-nine operator,

$$\mathcal{O} = (\bar{\eta}_L \eta_R)(\bar{Q}'_R Q'_L)(\bar{\eta}'_L \eta'_R). \quad (3.27)$$

This operator, similar to the construction in the previous section, also saturates the bound $d \geq 9$ derived in Section 2.3.5 for $f_a \sim 10^9$ GeV.

This composite construction is, to our knowledge, the first explicit, ultraviolet-complete model combining a solution to the quality problem (via gauging, Section 3.3.1), a demonstrated $N_{\text{DW}} = 1$ solution and a non-trivial dynamical reason to expect the cosmologically dangerous higher-winding strings to be subdominant.

Each of the two sectors in the model has four flavors of Dirac fermions, hence their asymptotic freedom is evident. There is no such requirement for QCD or the Abelian $U(1)_{\text{gPQ}}$; however, it is affordable to preserve the asymptotic freedom of QCD like in the SM by demanding $N_A + N_B \leq 10$. This, together with the requirement for $N_{\text{DW}} = 1$ given by $\text{gcd}(N_A, N_B) = 1$, provides a finite list of choices.

⁴A negative value of the exponent means the conjugate factor raised to the absolute value of the exponent.

⁵There is a six-dimensional gauge-invariant operator that is PQ-charged, given by

$$\mathcal{O}_6 = (Q_L \eta'_L)(\eta_L Q'_L)^\dagger. \quad (3.26)$$

However, this operator is charged under the vector-like symmetries $U(1)_X^{A(B)} \subset SU(4)_V^{A(B)}$ where $SU(4)_V$ is the unbroken global symmetry of the confining sectors. Since each sector is vector-like under its confining group, the vector-like flavor symmetries are unbroken by Vafa–Witten theorem. Therefore, the VEV of this operator must vanish $\langle \mathcal{O}_6 \rangle = 0$ and does not contribute to the axion potential. It may contribute at next order via $\mathcal{O}_6^\dagger \mathcal{O}_6$ but this is subleading to the operator in (3.27).

The quality of PQ symmetry depends on the choice; we chose our benchmark merely as a representative. The last gauge group in the model, $U(1)_{\text{gPQ}}$, develops a Landau pole; however, this is not a crippling condition. The coupling of the model can be adjusted to push the Landau pole beyond M_{Pl} , or it can be assumed that the Abelian gauge group is embedded in a simple group at a higher scale.

3.4 Heavy colored relics

Both constructions of Sections 3.2 and 3.3 solve the domain wall problem of Section 2.5.5 in the sense made precise above, removing what Chapter 2 identified as the principal obstacle to a post-inflationary cosmology for a composite axion. They do not, however, solve a second problem, common to essentially *every* composite axion model regardless of the strategy used to fix the domain wall number: the new confining sector responsible for generating f_a dynamically also generates new hadron-like bound states, the baryons and mesons of the new strong dynamics, some of which are charged under QCD. After the QCD phase transition, these heavy colored states bind to ordinary quarks to form exotic, electromagnetically charged heavy states (CHAMPs). This section first weighs the two generic strategies for disposing of such relics, then works out in detail the dilution strategy we adopt and the joint constraints it must satisfy.

3.4.1 Decay or dilution

There are, in principle, two strategies for dealing with a thermally produced population of CHAMPs. The first is to render them *unstable*: assign $U(1)_{\text{EM}}$ charges to the new fermions such that some of the resulting composite states can decay into ordinary Standard Model particles, removing them from the late-universe inventory entirely. This is not, in practice, as straightforward as it sounds, because the new fermions' gauge representations were chosen throughout Sections 3.2–3.3 specifically to cancel a delicate web of $SU(5)_{1,2}$, or $SU(N_{A,B})$ and $U(1)_{\text{gPQ}}$, and gravitational anomalies. Introducing additional electromagnetic charge assignments generically reopens some of these anomaly conditions, forcing the introduction of yet further new fields to re-cancel them, which can in turn introduce *their own* new stable relics, trading one heavy-relic problem for another rather than solving it.

The second, more model-independent strategy, and the one we pursue is *dilution*: arrange for a brief secondary period of cosmic inflation [125], occurring after the relics have already frozen out thermally but before the present day, and let the entropy production dilute their comoving number density to an acceptable level. The price of

this approach is that the same secondary inflation that dilutes the relics will, if it is not carefully timed, also dilute away the axion string network of Sections 3.2–3.3, so we require the construction below must simultaneously dilute the relics *and* preserve a genuinely post-inflationary axion cosmology, a non-trivial joint requirement rather than two independent ones.

3.4.2 Diluting the relic abundance

We model the relic abundance using ordinary thermal freeze-out: a heavy composite state of mass m_B (the scale of the new confining dynamics) annihilates with a cross section set by its own size, $\langle\sigma v\rangle \sim 1/m_B^2$, exactly as for any strongly-interacting thermal relic with no further small parameters. Standard freeze-out kinetics fixes its abundance, at freeze-out temperature T_f to the estimate [126, 127]

$$\frac{m_B n_B(t_I)}{s(t_I)} \sim \frac{x_f}{M_{\text{Pl}} \langle\sigma v\rangle} \sim \frac{x_f m_B^2}{M_{\text{Pl}}}, \quad x_f \equiv \frac{m_B}{T_f}. \quad (3.28)$$

At a time t_I after freeze-out, the secondary inflationary epoch under discussion starts (the precise value of t_I relative to the freeze-out will not matter, since both n_B and s separately redshift as R^{-3} between freeze-out and t_I , leaving their ratio fixed).

Now let the universe undergo a second, brief inflationary epoch, beginning at t_I and temperature T_I , ending at t_{end} , followed by a matter-dominated period of inflaton oscillation (the generic post-inflationary reheating behavior, $R(t) \propto t^{2/3}$) that completes reheating at t_R , and temperature T_R . Because the comoving relic number $n_B R^3$ is conserved throughout (the relics are far too heavy and rare to re-annihilate or re-thermalize once frozen out), the dilution factor is purely geometric, the ratio of comoving volumes before and after the combined inflation-plus-reheating epoch,

$$\frac{n_B(t_R)}{n_B(t_I)} = \left[\frac{R(t_I)}{R(t_{\text{end}})} \right]^3 \left[\frac{R(t_{\text{end}})}{R(t_R)} \right]^3. \quad (3.29)$$

The first factor is the dilution during inflation itself, $[R(t_I)/R(t_{\text{end}})]^3 \sim e^{-3N_e}$ for N_e e-folds, by definition of the e-fold number; the second is fixed by energy conservation during the matter-dominated reheating phase that follows any period of inflaton oscillation, during which $\rho \propto R^{-3}$ while the temperature of the eventually-thermalized radiation bath redshifts more slowly, giving the standard result $[R(t_{\text{end}})/R(t_R)]^3 \sim (t_{\text{end}}/t_R)^2 \sim T_R^4/T_I^4$ (the temperature at the start of inflation is given by $T_I^4 \sim M_{\text{Pl}}^2 H_I^2$). Multiplying (3.28) by both dilution factors of (3.29), and trading the explicit N_e for the more directly meaningful quantity introduced next, gives the diluted relic abundance we will impose the phenomenological bounds on.

3.4.3 Axion string reentry

The construction of Section 3.4.2 must not be allowed to simply erase the axion strings of Sections 2.5.3 and 3.2–3.3 along with the CHAMPs. Since the secondary inflationary epoch stretches every physical length scale by the same factor e^{N_e} , the typical inter-string separation, of order the Hubble length at the time the strings were last in their scaling regime, is stretched along with everything else, to a comoving scale that temporarily exceeds the now much smaller post-inflationary horizon. The strings do not, however, disappear: as in ordinary post-inflationary cosmology, the horizon subsequently grows faster than this stretched comoving separation, and the strings are assumed to eventually *reenter* the horizon at some later temperature $T_{\text{ent}} < T_R$, after which the network resumes ordinary scaling behavior exactly as if no secondary inflation had ever occurred. The number of e-folds can be written in terms of the T_{ent} [128], by comparing the physical distance at T_{ent} with the horizon

$$\frac{d_{\text{inter-string}}}{H_I^{-1}} = \underbrace{e^{N_e}}_{\text{inflation}} \underbrace{\frac{T_I^{4/3}}{T_R^{4/3}}}_{\text{inf. oscillations}} \underbrace{\frac{T_R}{T_{\text{ent}}}}_{\text{radiation}} \sim \frac{H_{\text{ent}}^{-1}}{H_I^{-1}} = \frac{T_I^2}{T_{\text{ent}}^2} \implies e^{3N_e} \sim \frac{T_I^2 T_R}{T_{\text{ent}}^3}, \quad (3.30)$$

eliminating N_e entirely in favor of the four directly physical scales f_a (equivalently the string-formation scale), T_I , T_R , and T_{ent} . The number density of the CHAMPs is diluted and the entropy is injected back at T_R

$$\frac{n_B(t_R)}{n_B(t_I)} = e^{-3N_e} \frac{T_R^4}{T_I^4}, \quad \frac{s(t_R)}{s(t_I)} = \frac{T_R^3}{T_I^3}, \quad (3.31)$$

giving the diluted CHAMP yield

$$\frac{m_B n_B}{s} \sim \frac{x_f m_B^2}{M_{\text{Pl}}} \frac{T_{\text{ent}}^3}{T_I^3}. \quad (3.32)$$

3.4.4 Axion abundance from the string-wall network

The question whether the string-wall network resumes ordinary post-inflationary dynamics smoothly, or dilution has qualitatively changed the outcome, depends on the relative ordering of T_{ent} and the axion-oscillation temperature T_{osc} . If the strings reenter *well before* domain walls would have formed, $T_{\text{ent}} > T_{\text{osc}}$, the network simply resumes the standard post-inflationary scaling behavior of Section 2.5.3, and the present axion abundance is, to a good approximation, the same as in the standard (undiluted) post-inflationary scenario evaluated with $T_{\text{ent}} \rightarrow T_{\text{osc}}$.

An interesting and qualitatively new regime is $T_{\text{ent}} \lesssim T_{\text{osc}}$: the strings reenter the horizon *after* the domain walls (of which there is, by construction, only $N_{\text{DW}} = 1$ attached to each string) have already formed. While the domain wall is outside the horizon, it pulls the string inward with acceleration,

$$a \sim \frac{\sigma_{\text{wall}}}{\mu_{\text{string}}} = \frac{8}{\pi} \frac{m_a(T)}{\log(f_a/H)}. \quad (3.33)$$

The earliest possible collapse is precisely horizon re-entry, and it depends on the ratio of the time-scales the wall's pull on the strings and the Hubble expansion's push against them. Collapse occurs when

$$\epsilon \equiv \frac{t_{\text{push}}}{t_{\text{pull}}} \sim \frac{\sigma}{\mu H} \gtrsim 1. \quad (3.34)$$

At T_{osc} , this ratio is given by $\epsilon \sim 8/70 \approx 0.11$. However, this grows quickly with $\sigma/\mu H \propto T^{-6}$, so it is particularly sensitive to the re-entry temperature. The resulting boost, obtained from $a = \sigma/\mu$, $\Delta t = H^{-1}$, is $\gamma \sim \sqrt{1 + \epsilon^2}$; therefore, if the walls enter the horizon at $T_{\text{ent}} < T_{\text{osc}}$ such that the boost is large enough $\gamma \sim \epsilon \gg 1$, then the emitted axions from the collapse will also be boosted.⁶

For the entering network before or at T_{osc} , the axion abundance is

$$\left. \frac{\rho_a}{s} \right|_0 \sim \frac{m_a(0)}{m_a(T_{\text{ent}})} \frac{\rho_{\text{string}}(T_{\text{ent}}) + \rho_{\text{wall}}(T_{\text{ent}})}{s(T_{\text{ent}})}, \quad (3.35)$$

while the emitted axions take boosted values when the re-entry is later than T_{osc} , $E_a \sim \gamma m_a(T_{\text{ent}})$.

3.4.5 The viable parameter space

Four scales (f_a , T_I , T_R , T_{ent}) parametrize this construction, and a viable point must simultaneously satisfy several conditions.

- (i) The diluted CHAMP abundance of (3.32) must not exceed the percent-level cosmic-microwave-background bound on any exotic charged relic and, more restrictively in the mass range of interest here, the direct flux bounds from underground and cosmic-ray detectors sensitive to anomalously ionizing particles, principally the ICRR [129, 130] and Baksan [131, 132] experiments, which together limit $\Omega_{\text{CHAMP}} h^2 / 0.12 \lesssim 10^{-10} - 10^{-8}$ for $10^8 \text{ GeV} \lesssim m_B \lesssim 10^{11} \text{ GeV}$ [24].

⁶In the standard scenario, since the string-wall network is inside the horizon, the collapse occurs with $\epsilon \gtrsim 1$, so the emitted axions are not significantly boosted, $\langle E_a \rangle \sim m_a(T)$.

- (ii) The axion abundance of (3.35) must not exceed the observed dark matter density, $\Omega_a h^2 / 0.12 < 1$.
- (iii) Thermal regeneration of CHAMPs during the secondary inflaton-oscillation epoch must remain suppressed, which, using the maximum temperature reached during that epoch, $T_{\max} \sim \sqrt{T_I T_R}$ [133], and $m_B \sim 4\pi f_a$ for the typical confinement-scale baryon relics, requires $T_{\max} \lesssim 10^{-3} - 10^{-2} m_B$ so as not to thermally produce them after inflation.
- (iv) The hierarchy $f_a > T_I > T_R > T_{\text{ent}}$ must hold, simply to ensure the four scales are used in a physically sensible order: the PQ-breaking scale above the onset of the secondary inflation, which is above the reheating temperature it produces, which is above the temperature at which the diluted strings eventually reenter the horizon.

Scanning these constraints, one finds that condition (i) is the binding one throughout most of the relevant parameter space, and that satisfying it while also satisfying (ii)–(iv) consistently pushes the viable region toward T_{ent} of order a few GeV to a few hundred MeV, that is, closer to the regime $T_{\text{ent}} \sim T_{\text{osc}}$ identified in Section 3.4.4. Figure 3.2 shows the resulting viable region in the (T_R, f_a) plane for a representative grid of T_I/T_R and T_{ent} values, confirming that an open, non-empty parameter space exists in which both the heavy-relic problem of Sections 3.4.2–3.4.3 and the domain-wall problem of Section 2.5.5 are simultaneously resolved, for either of the two $N_{\text{DW}} = 1$ models constructed in Sections 3.2–3.3.

This parameter space comes with two practical caveats worth flagging explicitly, beyond the open dynamical question of Section 3.2.3. First, the axions radiated directly by the string network during its pre-dilution and post-reentry scaling phases (Section 2.5.3) have been folded into (3.35) only through the string energy density at reentry, $\rho_{\text{string}}(T_{\text{ent}})$, rather than through a dedicated simulation of the radiated spectrum; numerical studies of ordinary (undiluted) post-inflationary axion strings suggest that this radiated component can be comparable to, or somewhat larger than, the wall-collapse estimate used here, which would shift the viable f_a in Figure 3.2 down by an $\mathcal{O}(1)$ factor without changing the qualitative shape of the allowed region. Second, the premise of Section 3.3.2, that the subdominant $N_{\text{DW}} > 1$ string population produced alongside the dominant, benign one is small enough to neglect, has not been verified for the specific gauge-coupling and symmetry-breaking-scale hierarchies of the model constructed in Section 3.3.3, since the only existing simulations of this kind were performed for simpler, non-composite gauged-PQ models with different charge assignments; extending those simulations to the present set-

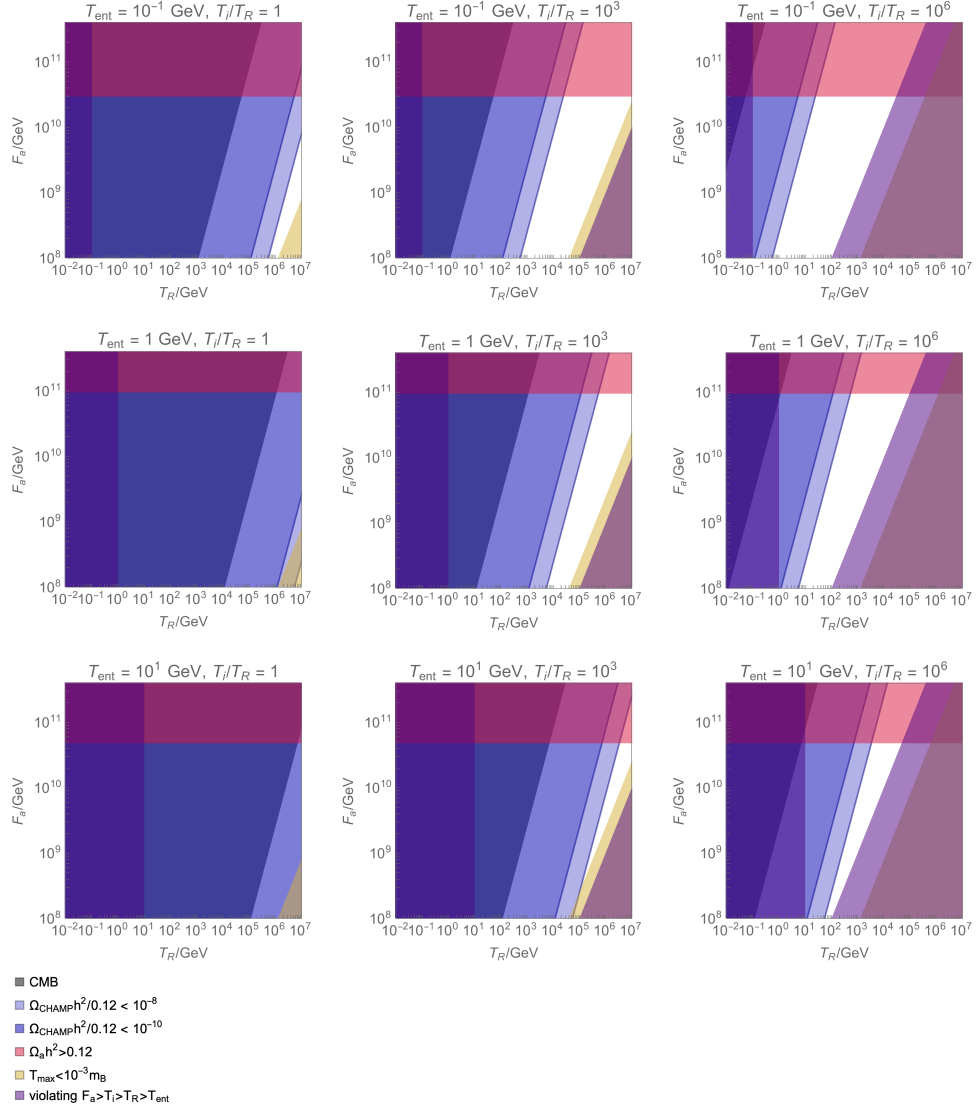


Figure 3.2: Viable parameter space in the T_R - f_a plane, for a representative grid of T_I/T_R and T_{ent} values. Gray/blue shadings show CHAMP overabundance, red shows axion overabundance, brown shows insufficient suppression of thermal CHAMP regeneration, and purple shows violation of the scale ordering $f_a > T_I > T_R > T_{\text{ent}}$; uncolored regions are simultaneously viable. Reproduced from our Ref. [1].

ting is, together with the confinement-dynamics question of Section 3.2.3, among the most pressing pieces of follow-up work motivated by this chapter.

3.5 Summary

This chapter turned the obstruction identified at the end of Chapter 2 into two solved instances of it. The master formula $N_{\text{DW}} = \text{gcd}(k_{PQ}, k_G, \dots)$ of Section 3.1 reduced the search for a working composite axion to a question of representation theory, and two structurally different answers followed from it. Identifying $U(1)_{PQ}$ with the baryon number of a chiral confining sector (Section 3.2) makes the required identification automatic, at the price of a single-sector construction that stalls at $N_{\text{DW}} = 2$; doubling the confining sector into a chiral moose removes the obstruction and reaches $N_{\text{DW}} = 1$. Gauging an orthogonal combination of two ordinary Choi–Kim sectors (Section 3.3) reaches the same destination by a different route, and does so while also disposing of the quality problem by gauge invariance rather than by a large operator dimension. Both constructions were then shown to share a second obstacle common to essentially all composite axion models, thermally produced colored relics, and a shared fix, dilution by a brief secondary inflation timed to preserve the axion string network, was shown in Section 3.4 to open a genuine, non-empty region of viable parameter space for either model.

Two open questions follow both constructions into Section 3.4 rather than being resolved by it: whether the assumed condensation pattern of the automatic-PQ construction is actually realized by the strongly coupled dynamics, rather than the Higgsing pattern favored by both the MAC estimate and the functional renormalization group result of Section 3.2.3; and whether the subdominant population of dangerous $N_{\text{DW}} \geq 2$ strings in the gauged construction is small enough to neglect for its specific scales, which existing simulations have not directly tested. None of these bear on the central group-theoretic result of this chapter, that $N_{\text{DW}} = 1$ is achievable by either route; they bear on how much confidence to place in the models built to demonstrate it, and are left, honestly, as such. These threads are picked up again in the concluding chapter.

Chapter 4

Conclusions

4.1 Summary of results

This part of the thesis set out from a single, sharply posed question, framed in our introduction 1: can a composite axion, a construction that solves both the hierarchy problem of an elementary PQ-breaking scalar and, through chiral confinement, the axion quality problem, support a genuinely *post-inflationary* cosmology?

Chapter 2 built up the machinery needed to make this question precise: the topological origin of the QCD θ -term and the strong CP problem (Section 2.1), the PQ mechanism and its concrete realizations together with a quantitative statement of the quality problem (Section 2.3), the astrophysical and laboratory constraints that bound the viable decay-constant window (Section 2.4), and the cosmology of the axion field, its strings and domain walls (Sections 2.5–2.5.5). Then we considered the Choi–Kim composite axion and its chiral, moose-deconstructed descendants (Section 2.6) culminating in the identification of the Lazarides–Shafi mechanism as the one available route to a domain wall number of one (Section 2.6.3).

Chapter 3 answered the question affirmatively in two structurally independent ways. First, we provided a general master formula (Section 3.1) that makes the Lazarides–Shafi identification of vacua easily computable directly from the representation theory of a confining gauge theory, without relying solely on the generally overcounting anomaly coefficient. It is model-independent and can be applied to any confining axion sector, as we have checked against direct enumeration in examples of the literature. Applying it to a chiral moose extension of the automatic PQ model (Section 3.2), and to a new composite realization of a gauged PQ symmetry built from two Choi–Kim sectors (Section 3.3), gives, in both cases, $N_{\text{DW}} = 1$: the existence of an axion string solution whose attached domain wall is unstable and collapses on its own, with no need to assume that the PQ transition predates

inflation.

Both constructions simultaneously improve the axion quality problem, through precisely the same chirality (in the first case) or gauging (in the second) that controls the domain wall number, underscoring that these are not two independent model-building choices but two faces of the same underlying structure. A critical re-examination of the confinement dynamics underlying the automatic-PQ construction (Section 3.2.3), using both the most-attractive-channel heuristic and a recent functional renormalization group analysis, showed that the assumed condensation pattern is a real, open dynamical question rather than a settled technical detail, a caveat that applies to the entire automatic-PQ literature and not merely to the extension constructed here.-

Finally, Section 3.4 showed that the new colored composite relics generic to both constructions, a second cosmological obstacle independent of the domain wall problem, can be diluted to acceptable levels by a brief secondary inflationary epoch, provided the axion string-wall network reenters the horizon close to the QCD transition, identifying an explicit, non-empty viable parameter space in which both problems are resolved together.

4.2 Limitations

Two open dynamical questions limit the constructions of Chapter 3, one for each.

For the automatic-PQ construction of Section 3.2, the result rests on an assumed confinement pattern for the chiral $SU(5)$ sector: that $SU(5)$ confines while baryon number is spontaneously broken, rather than Higgsing directly via a two-fermion condensate. Section 3.2.3 showed, using both the most-attractive-channel heuristic and a recent functional renormalization group study [120], that the latter may instead occur for the phenomenologically relevant case $N = 5$. This is a genuinely open, non-perturbative dynamics question that neither this thesis nor the existing literature resolves; it can likely only be settled by a dedicated lattice simulation of the chiral gauge theory itself, which remains technically challenging for chiral, as opposed to vector-like, confining matter.

For the gauged-PQ construction of Section 3.3, the gap is cosmological rather than dynamical. Its confining sectors are ordinary, vector-like Choi–Kim theories whose infrared dynamics is comparatively well understood, so it does not share the liability above. But the relative abundance of $N_{\text{DW}} = 0$, $N_{\text{DW}} = 1$, and $N_{\text{DW}} > 1$ strings formed in the cosmological network, which is central to the argument of Section 3.3.2 that the benign $N_{\text{DW}} \leq 1$ population dominates, has, to date, only been

simulated for models in which the PQ-breaking field is an elementary scalar. Our construction generates that field compositely, through confinement, and whether the same relative string abundances carry over to the composite case has not been checked.

Part II

Q-Ball Superradiance: Validity Regime of Linear Analysis and Extension to Many-Field Models

Chapter 5

Introduction

5.1 Solitons in field theory

Classical field theories with nonlinear interactions admit smooth, localized, finite-energy solutions that do not disperse under time evolution. These solutions, collectively known as *solitons*, behave as particle-like objects despite being extended configurations of classical fields. Their defining property is stability: once formed, they cannot continuously relax into the vacuum [134].

Different solitons owe this stability to different conserved quantities. For topological solitons, stability is guaranteed by topology due to the presence of multiple degenerate vacua; the vacuum and the soliton belong to different homotopy classes, preventing any continuous deformation between them. Non-topological solitons, in contrast, possess the same topology as the vacuum and therefore require no degenerate vacuum structure [135]. Their stability instead originates from a conserved Noether charge associated with a continuous global symmetry. The existence of such objects is therefore an energetic question: can the theory lower its energy by storing a fixed conserved charge inside a localized field configuration rather than in a collection of free particles?

The simplest realization of this idea is the *Q-ball*, a localized configuration of a single complex scalar field carrying a conserved global charge [136]. Unlike an elementary particle, however, a Q-ball behaves as a finite-density object. Its conserved charge is distributed throughout an extended classical configuration, allowing it to exchange both charge and energy with its environment while remaining a coherent soliton. Understanding this exchange in processes of wave scattering is the central theme of this part of the thesis.

5.2 Q-balls in cosmological environment

The cosmological significance of Q-balls stems not only from their formation, but also from their subsequent interactions with their environment. In many extensions of the Standard Model, scalar condensates carrying a global charge naturally arise during inflation and may fragment into large populations of Q-balls through dynamical instabilities [137–140]. Similar objects may also form during cosmological phase transitions through the concentration of charge by expanding bubble walls [141, 142]. These mechanisms illustrate that non-topological solitons can arise under a broad range of cosmological conditions.

Once formed, however, the evolution of a Q-ball is determined by how it exchanges charge and energy with its environment. Thermal particles may be absorbed, allowing the Q-ball to grow through *solitosynthesis* [143–146]. Conversely, charge and energy may be released into the surrounding plasma through evaporation or other scattering processes. The efficiency of these exchanges directly affects the lifetime, relic abundance, and cosmological impact of any Q-ball population [147–155].

Recently, a particularly striking example of such interactions has attracted considerable attention. Under suitable conditions, incident scalar waves are amplified after scattering from a Q-ball, extracting both charge and energy from the soliton in a process known as *Q-ball superradiance* [156, 157]. This phenomenon provides a direct example of charge and energy transfer between a soliton and an external excitation, where the scattering process itself encodes the exchange of conserved quantities.

5.3 Objectives and Contributions

Existing analyses of Q-ball superradiance have established the perturbative scattering framework and demonstrated its predictions through comparisons with fully nonlinear lattice simulations. In these studies, however, the nonlinear evolution was typically examined at incident-wave amplitudes sufficiently small to remain within the linear regime, confirming the consistency between the perturbative description and the full dynamics. The range of amplitudes over which the linear approximation remains quantitatively reliable has not been systematically determined. Since perturbative calculations of charge and energy transfer provide the basis for assessing possible cosmological consequences, establishing the domain of validity of this approximation is an essential step toward making quantitative predictions.

Alongside establishing the regime of validity of the perturbative description, we

develop a clearer physical understanding of the scattering formalism itself. The mode conversion appearing in the linearized equations can be interpreted as a microscopic charge-exchange process between the incident wave and the Q-ball. By relating the charge and energy carried by the perturbative modes to the energy difference between neighboring Q-ball configurations, we clarify the thermodynamic origin of the amplification factors.

Building upon this understanding, we perform a systematic scan over incident-wave amplitudes using nonlinear lattice simulations. This allows us to determine the range of amplitudes over which the linear scattering description accurately captures the nonlinear dynamics and to identify the onset of genuinely nonlinear effects.

Finally, we extend this framework beyond the one-field Q-ball to the Friedberg–Lee–Sirlin soliton [158–160]. The additional field introduces a richer perturbation spectrum with multiple scattering channels, providing a non-trivial test of the generality of the scattering formalism. We construct the corresponding multichannel perturbation theory, and analyze the dynamics in the linear regime.

Together, these results provide both a physical interpretation and a quantitative validation of perturbative scattering off non-topological solitons, establishing when the linear description can be reliably used to study charge and energy exchange.

5.4 Outline of this part

The remainder of Part II is organized as follows:

- Chapter 6 reviews the theory of non-topological solitons and constructs the stationary solutions relevant for the subsequent analysis, including the Q-ball and the Friedberg–Lee–Sirlin soliton. Then, we derive the properties needed for the perturbative analysis that follows.
- Chapter 7 is the technical core of this part: it develops the linear perturbation theory and scattering formalism for Q-balls, determines the range of validity of the perturbative description through comparison with nonlinear lattice simulations, and extends the same framework to the Friedberg–Lee–Sirlin soliton.
- Chapter 8 summarizes the main results and discusses possible directions for future work.
- Appendix A details the numerical methods to solve the background solutions of the solitons. Appendix B expands on the s-wave analysis in Chapter 7 and derives the cross section of a plane wave scattering off the Q-ball rather than a spherical wave.

Chapter 6

Non-topological Solitons and Their Stationary Solutions

This chapter constructs the stationary solutions introduced qualitatively in Chapter 5: the one-field Q-ball and its Friedberg–Lee–Sirlin (FLS) extension. We begin with the Q-ball, arising from a complex scalar field with an unbroken global $U(1)$ symmetry. By minimizing the energy at fixed charge, we derive the stationary solution, establish its thermodynamic interpretation, and determine the properties of the background configuration relevant for perturbations analyzed in Chapter 7.

The discussion is then extended to the FLS soliton, in which the charged field is coupled to an additional real scalar field. This additional degree of freedom modifies the structure of the soliton and, as anticipated in Section 5.3, leads to a richer perturbation spectrum with multiple scattering channels. The two constructions share a common variational structure which we develop once, for the Q-ball, and reuse directly for the FLS soliton.

6.1 The one-field Q-ball

The simplest non-topological soliton arises in a theory containing a single complex scalar field with a global $U(1)$ symmetry. The corresponding Lagrangian is

$$\mathcal{L} = |\partial_\mu \phi|^2 - V(\phi), \tag{6.1}$$

where the potential function $V(\phi) \geq 0$ has a minimum at $\phi = 0$. This vacuum therefore preserves the global $U(1)$ symmetry $\phi \rightarrow e^{-i\theta} \phi$. By Noether's theorem,

this symmetry results in the conserved current

$$J_\mu = -i(\partial_\mu \phi^* \phi - \phi^* \partial_\mu \phi), \quad \partial^\mu J_\mu = 0, \quad (6.2)$$

corresponding to a conserved charge Q . Likewise, the translation symmetry results in the conservation of the stress–energy–momentum tensor current whose conserved quantities are the momentum components P_ν ,

$$T_{\mu\nu} = \partial_\mu \phi^* \partial_\nu \phi + \partial_\mu \phi \partial_\nu \phi^* - g_{\mu\nu} \mathcal{L}, \quad \partial^\mu T_{\mu\nu} = 0. \quad (6.3)$$

6.1.1 Energy minimization at fixed charge

The Q-ball is defined as the lowest-energy configuration that carries a fixed conserved charge Q . To find such a configuration, we solve a constrained minimization problem. We begin by enforcing the charge constraint onto the energy through a Lagrange multiplier ω such that the energy of configurations of charge Q becomes

$$E_Q[\phi, Q] = E[\phi] + \omega(Q - Q[\phi]) = \int d^3\mathbf{x} T_{00} + \omega \left(Q - \int d^3\mathbf{x} J_0 \right). \quad (6.4)$$

The constrained energy functional takes the form of the internal energy of a thermodynamical system at zero-temperature [161],

$$E_Q[\phi, Q] = \omega Q + \Omega[\phi, \omega], \quad (6.5)$$

where the grand canonical potential $\Omega[\phi, \omega]$ is

$$\Omega[\phi, \omega] \equiv \int d^3\mathbf{x} |\dot{\phi} + i\omega\phi|^2 + |\nabla\phi|^2 + V(\phi) - \omega^2|\phi|^2. \quad (6.6)$$

Thus, the multiplier ω is interpreted as the chemical potential of this system.

The system's state of equilibrium minimizes the grand potential,

$$\delta\Omega[\phi, \omega] = \delta E[\phi] - \omega \delta Q[\phi] = 0. \quad (6.7)$$

This equilibrium state corresponds to the lowest energy configuration that the field can take at fixed charge, i.e., the Q-ball. Since the equilibrium solution satisfies the variational equation $\delta E = \omega \delta Q$, along the family of equilibrium solutions parameterized by Q , this becomes $dE = \omega dQ$, i.e., the change of Q-ball energy per unit

change in charge follows from (6.7),

$$\boxed{\frac{dE}{dQ} = \omega.} \quad (6.8)$$

This is in accordance with the interpretation of ω as the chemical potential: adding one particle to the similarly-charged Q-ball (increase the charge by one unit) increases the energy by an amount of ω .¹ This is a key result that will be used later in Chapter 7 when discussing the exchange of energy and charge between perturbative excitations and the Q-ball.

6.1.2 Stationary solutions

The first term in the grand potential (6.6) is manifestly positive-definite, so setting it to zero identically, $\dot{\phi} = -i\omega\phi$, is a necessary step to minimize the grand potential. This is naturally realized by the ansatz

$$\phi_Q(\mathbf{x}, t) = \phi_Q(\mathbf{x}) e^{-i\omega t}, \quad \phi_Q(\mathbf{x}) \in \mathbb{R}. \quad (6.9)$$

By a slight abuse of notation, ϕ_Q will denote both the full solution and its radial profile whenever no confusion arises. We see that the multiplier ω acquires a complementary dynamical interpretation: it is the internal phase-rotation frequency of the Q-ball solution.

The Q-ball profile $\phi_Q(\mathbf{x})$ is determined equally from minimizing $\Omega[\phi, \omega]$ directly, or by substituting (6.9) into the classical field equation that follows from (6.1),

$$\square \phi + \frac{\partial V}{\partial \phi^*} = 0. \quad (6.10)$$

Since spherically symmetric profiles can be shown to minimize the energy [162], we restrict ourselves to $\phi_Q(\mathbf{x}) = \phi_Q(r)$, a spherical non-rotating nodeless profile, for which the field equation reduces to the radial profile equation²

$$-\phi_Q''(r) - \frac{2}{r}\phi_Q'(r) - \omega^2\phi_Q(r) + \frac{1}{2}\frac{\partial V}{\partial \phi_Q} = 0. \quad (6.11)$$

¹The free energy of the Q-ball thermodynamical system is exactly equivalent to the internal energy since temperature is zero.

²The factor of 1/2 in the last term of (6.11) arises from converting the Wirtinger derivative $\partial/\partial\phi^*$ into a derivative with respect to the real profile ϕ_Q .

Regularity at the origin and finiteness of the energy require

$$\phi_Q'(0) = 0, \quad \lim_{r \rightarrow \infty} \phi_Q(r) = 0. \quad (6.12)$$

Equation (6.11) admits a standard and very useful mechanical analogy [135, 136]: it is the equation of motion of a unit-mass fictitious body at “position” ϕ_Q , moving in “time” r under the *Newtonian potential*³

$$V_N(\phi_Q; \omega) \equiv \omega^2 \phi_Q^2 - V(\phi_Q), \quad (6.13)$$

subject to a time-dependent friction with coefficient $2/r$. Regularity at the origin means the body starts at rest, $\phi_Q'(0) = 0$, at some *a priori* unknown position $\phi_Q(0) \equiv \phi_0$; the friction then dissipates its mechanical energy as r grows, until the friction coefficient itself becomes negligible at large r . Finiteness of the energy requires the body to come to rest exactly at $\phi_Q = 0$ as $r \rightarrow \infty$.

6.1.3 Existence window for the Q-ball frequency

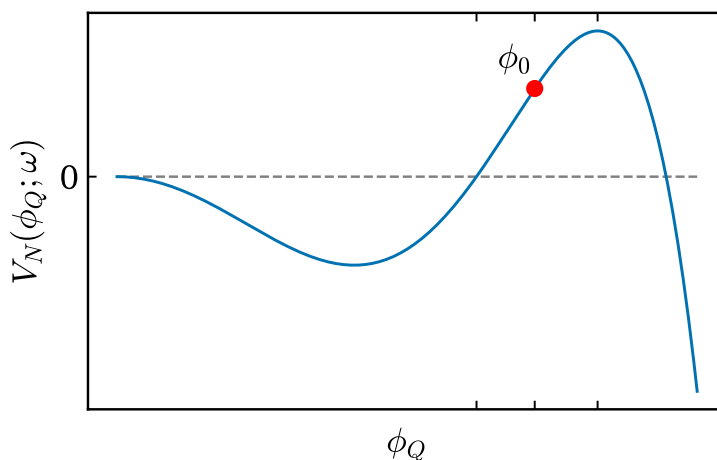


Figure 6.1: Schematic Newtonian potential $V_N(\phi_Q)$ of (6.13) for a frequency in the allowed window [163]. The origin $\phi_Q = 0$ is a local maximum, as required by (6.15), while $V_N > 0$ over a finite range of $\phi_Q > 0$, as required by (6.14).

The mechanical analogy immediately constrains which frequencies ω *may* allow for a soliton solution. Since $V_N(0; \omega) = 0$, the body can only dissipate a positive amount of mechanical energy on its way to $\phi_Q = 0$ if it starts in a region where

³The same factor of $1/2$ is absorbed into the normalization of the force, $F = -\frac{1}{2}\partial V_N/\partial \phi_Q$, and reflects the non-canonical normalization of the kinetic term for the real variable ϕ_Q ; the metric on this line is $ds^2 = 2d\phi_Q^2$.

$V_N(\phi_0; \omega) > 0$; this requires

$$\omega^2 > \min_{\phi \neq 0} \frac{V(\phi)}{|\phi|^2} \equiv \omega_-^2. \quad (6.14)$$

Having a region with $V_N > 0$ is necessary but not sufficient: if $\phi_Q = 0$ were a local minimum of V_N rather than a local maximum, the solution would approach the origin through oscillations, producing radially excited configurations with nodes, which are not the subject of this analysis. Restricting to the nodeless ground-state profile therefore requires $\phi_Q = 0$ to be a local maximum of V_N , i.e.

$$\omega^2 < \left. \frac{\partial^2 V}{\partial \phi^* \partial \phi} \right|_{\phi=0} \equiv m^2. \quad (6.15)$$

The mass of the elementary scalar sets an upper bound on the Q-ball frequency [136].

The derived conditions (6.14), (6.15) are model-independent: any complex field theory that can form Q-balls can only form Q-balls whose frequency lies in the range $\omega_- < \omega < \omega_+ \equiv m$. The qualitative shape of V_N required, where it vanishes at the origin, and becomes positive somewhere, with the origin a local maximum, is sketched in Figure 6.1.

6.1.4 Q-ball stability

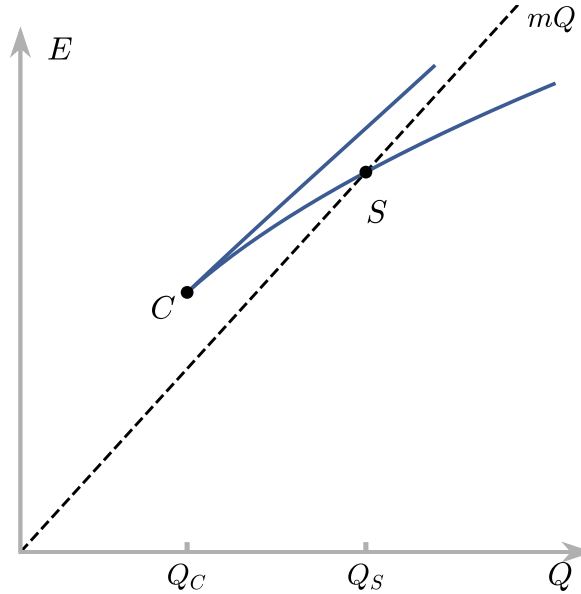


Figure 6.2: Schematic plot for the energy curve of Q-balls in 3 + 1 [164]. The upper energy branch describes unstable small Q-balls with $\omega \rightarrow m$. The lower energy branch describes stable large Q-balls $\omega \rightarrow \omega_-$ and metastable Q-balls of charges $Q_C < Q < Q_S$.

The stability of the Q-ball is determined by its energy–charge curve (Figure 6.2), where the energy and charge contained in a Q-ball are given by

$$E = 4\pi \int_0^\infty dr r^2 \left[\omega^2 \phi_Q^2 + (\phi'_Q)^2 + V(\phi_Q) \right], \quad Q = 4\pi \int_0^\infty dr r^2 \cdot 2\omega \phi_Q^2. \quad (6.16)$$

There are two distinct notions of stability that follow from this curve, which we are going to demonstrate. The first is classical stability against small perturbations that preserve the charge: a Q-ball is a genuine local minimum of E at fixed Q only where $dQ/d\omega < 0$ [135, 136],

$$\frac{d^2 E}{dQ^2} = \frac{d\omega}{dQ} < 0. \quad (6.17)$$

When it is $dQ/d\omega > 0$ instead, the configuration is a saddle point of the energy at fixed charge, not a minimum, and is unstable to charge-conserving perturbations. This defines a critical frequency ω_C through

$$\left. \frac{dQ}{d\omega} \right|_{\omega=\omega_C} = 0, \quad (6.18)$$

which separates the two branches of Figure 6.2: small Q-balls, $\omega_C < \omega < m$, sit on the upper, unstable branch, asymptoting to $E \rightarrow mQ$ as $\omega \rightarrow m$; large Q-balls, $\omega_- < \omega < \omega_C$, sit on the lower branch and are classically stable in this restricted sense.

Classical stability along the lower branch does not yet make a Q-ball the true ground state of its charge sector, since Q widely separated free quanta of mass m are always an available, charge-conserving alternative configuration. The Q-ball is *absolutely* stable only if its energy is below the mass of Q free scalars carrying the same charge,

$$\boxed{E < mQ}. \quad (6.19)$$

The lower branch crosses $E = mQ$ at a single intermediate frequency ω_S , defined by

$$E(\omega_S) = mQ(\omega_S), \quad \omega_- < \omega_S < \omega_C. \quad (6.20)$$

This splits the lower branch into two physically distinct regimes: for $\omega_- < \omega < \omega_S$, Q-balls are absolutely stable, lower in energy than any collection of free quanta of the same charge; for $\omega_S < \omega < \omega_C$, they remain classically stable against small perturbations but are only *metastable*, since decay into Q free particles is energetically allowed, and can proceed only through a finite-size fluctuation over the barrier separating the two configurations [135, 136, 165].

6.1.5 Symmetries broken by the Q-balls

The Q-ball solution (6.9) spontaneously breaks time translations and the global $U(1)$ individually, but leaves an unbroken diagonal combination of them,

$$\phi_Q(r) e^{-i\omega t} \xrightarrow{t \rightarrow t-t_0} \phi_Q(r) e^{-i\omega(t-t_0)} \xrightarrow{\theta=\omega t_0} e^{-i\theta} \phi_Q(r) e^{-i\omega(t-t_0)} = \phi_Q(r) e^{-i\omega t}. \quad (6.21)$$

Together with the three broken spatial translations, each Q-ball solution belongs to a four-parameter family related by the broken generators that live in the space $\mathbb{R}^3 \times S^1$ (the Q-ball's position and its internal $U(1)$ phase at some reference time). These broken symmetries manifest themselves as zero modes in the perturbation spectrum constructed in Chapter 7.

6.2 Q-balls of polynomial potentials

We will consider a specific class of potentials that can support such Q-ball solutions. A polynomial potential capable of satisfying both (6.14) and (6.15) must grow more slowly than $m^2|\phi|^2$ for some range of $|\phi|$ before eventually turning over; the simplest choice with this property [135] is the ϵ -deformed sextic potential

$$V(\phi) = \frac{g^2}{1+\epsilon^2} |\phi|^2 \left[(|\phi|^2 - f^2)^2 + \epsilon^2 f^4 \right] = g^2 f^4 |\phi|^2 - \frac{2g^2 f^2}{1+\epsilon^2} |\phi|^4 + \frac{g^2}{1+\epsilon^2} |\phi|^6, \quad (6.22)$$

shown in Figure 6.3. The elementary excitations around $\phi = 0$ have mass $m^2 = g^2 f^4$ and unit charge. Of the three parameters (g, f, ϵ) , only ϵ is considered a free parameter and will be specified, $\epsilon = 0.4$, for our construction. The other two parameters combine into dimensionless quantities: f sets the field scale and can be absorbed into the dimensionless ratio ϕ/f , while the coupling g , or equivalently the scalar mass m , sets the length scale via the dimensionless quantity mr . On the other hand, ϵ controls the *shape* of the potential, specifically how far the would-be second minimum at $|\phi| = f$ is lifted above $V = 0$, with larger ϵ corresponding to a shallower, more single-welled potential. Substituting (6.22) into (6.14), (6.15) gives the existence window in closed form,

$$\frac{\epsilon^2 m^2}{1+\epsilon^2} < \omega^2 < m^2. \quad (6.23)$$

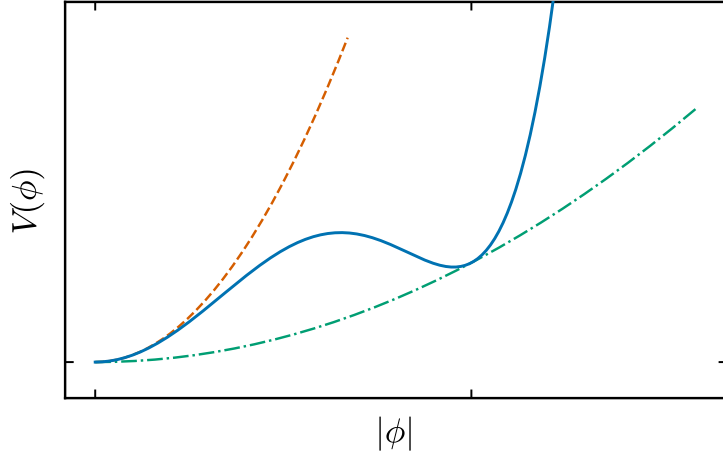


Figure 6.3: The sextic potential (blue) given by (6.22), together with the quadratic mass term $m^2|\phi|^2$ (orange) that sets the upper bound (6.15) on the Q-ball frequency and the envelope quadratic (green) that sets the lower bound (6.14) on the Q-ball frequency.

6.2.1 Numerical construction: the shooting method

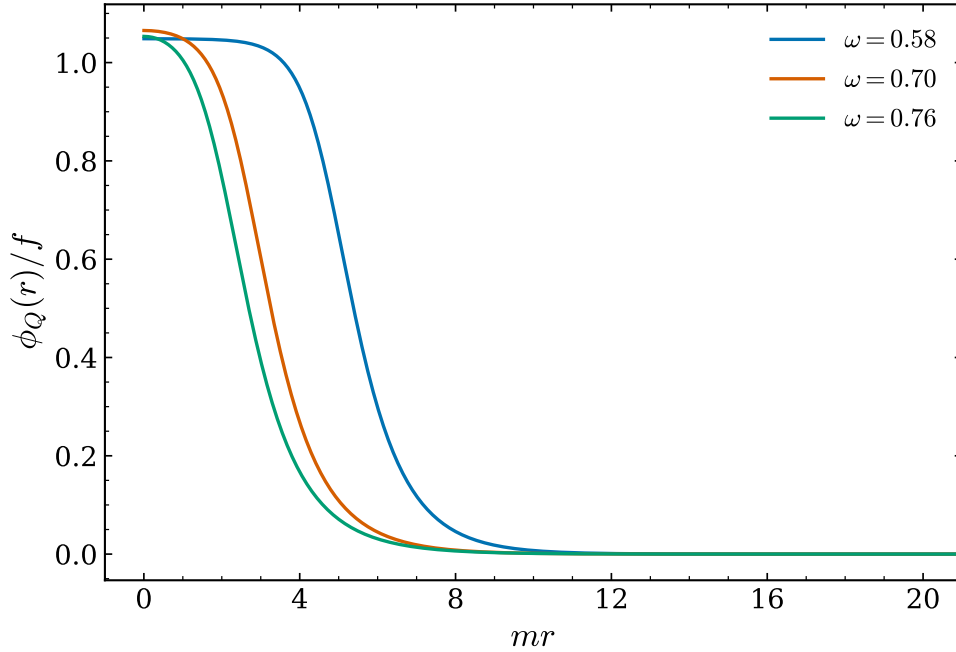


Figure 6.4: Q-ball profiles obtained by shooting, for $\epsilon = 0.4$ and three representative frequencies: $\omega = 0.58$ (blue), $\omega = 0.70$ (orange), $\omega = 0.76$ (green).

For a fixed ω in the existence window, (6.11) together with the boundary conditions (6.12) is a two-point boundary value problem. We solve it by the *shooting method*: rather than imposing the asymptotic condition directly, we solve the initial value problem $\phi_Q(0) = \phi_0$, $\phi'_Q(0) = 0$ for a trial ϕ_0 in the window identified above, integrate (6.11) outward, and adjust ϕ_0 until the numerical solution approaches

$\phi_Q(r \rightarrow \infty) = 0$ to the desired accuracy. Values of ϕ_0 that are too large cause ϕ_Q to overshoot past zero and run off to large negative values; values that are too small cause ϕ_Q to turn around before reaching zero. Bisecting between an overshooting and an undershooting trial value converges geometrically to the unique soliton profile.

Figure 6.4 shows the resulting profiles for $\epsilon = 0.4$ and three representative frequencies $\omega = 0.58, 0.70, 0.76$ (in units of mass m). Smaller ω (closer to ω_-) gives more extended profiles: from (6.23), decreasing ω lowers the maximum height of $V_N(\phi_0)$ that the fictitious body can access, so it carries less mechanical energy to dissipate and the friction term must act over a correspondingly longer “time” r before the body can settle at the origin. The limit $\omega \rightarrow \omega_-$ is the *thin-wall* limit, in which the profile approaches a step function with an increasingly sharp edge [136]; the opposite limit $\omega \rightarrow m$ is the *thick-wall* limit, in which the profile becomes increasingly Gaussian-like until no normalizable solution survives [149].

6.3 Friedberg–Lee–Sirlin (FLS) soliton

The construction of the previous section relied on a non-renormalizable sextic self-interaction for a single complex scalar field. Friedberg, Lee and Sirlin showed that non-topological solitons can instead be supported using only renormalizable interactions by coupling the charged field to a second, real scalar field [158–160]. The resulting model retains the conserved global $U(1)$ symmetry responsible for the soliton charge, while the additional scalar dynamically modifies the effective mass of the charged field inside the soliton. Throughout this thesis we refer to the corresponding stationary solution as the Friedberg–Lee–Sirlin (FLS) soliton.

The Lagrangian of the theory that admits such solution is given by

$$\mathcal{L} = |\partial_\mu \phi|^2 + \frac{1}{2}(\partial_\mu \sigma)^2 - V(\phi, \sigma), \quad V(\phi, \sigma) = g^2 \sigma^2 |\phi|^2 + \frac{h^2}{8}(\sigma^2 - f^2)^2, \quad (6.24)$$

which is invariant under the unbroken global $U(1)$ rotation of ϕ and the spontaneously broken discrete symmetry $\sigma \rightarrow -\sigma$. The potential has two degenerate vacua at $(\phi, \sigma) = (0, \pm f)$, both of which preserve the $U(1)$ symmetry, as required by non-topological solitons to exist. Expanding around either vacuum, the elementary excitations consist of a charged scalar of mass $m_\phi^2 = g^2 f^2 \equiv m^2$ and unit charge, and a neutral scalar of mass $m_\sigma^2 = \kappa^2 m^2 = h^2 f^2$. As in the previous section, m and f simply define the natural length and field scales through the dimensionless combinations mr , ϕ/f , and σ/f , leaving the mass ratio $\kappa = m_\sigma/m_\phi$ as the only relevant free parameter.

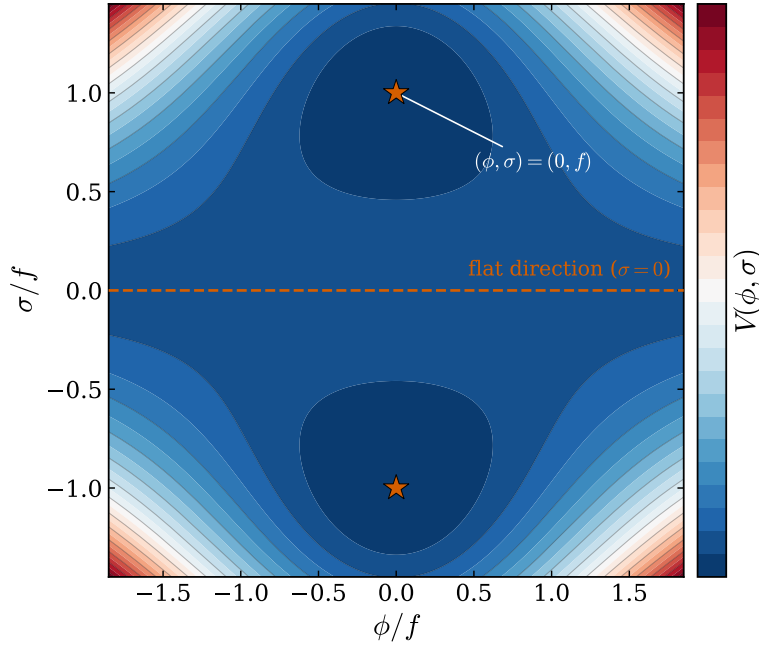


Figure 6.5: Contour plot of the FLS potential $V(\phi, \sigma)$ (6.24). The two minima at $(\phi, \sigma) = (0, \pm f)$ are degenerate; $\sigma = 0$ is a flat direction for ϕ at the classical level.

The physical mechanism differs qualitatively from that of the Q-ball presented in the previous section, even though both support localized configurations carrying the same conserved charge. Far from the soliton the real scalar approaches its vacuum expectation value, $\sigma \rightarrow f$, so that the charged field acquires its full vacuum mass m_ϕ . Inside the soliton, where the amplitude of ϕ becomes large, the portal interaction energetically favors smaller values of σ . This locally reduces the effective mass of the charged field, allowing the conserved charge to be stored more efficiently inside the soliton than in the vacuum. The resulting configuration therefore consists of a charge-carrying core surrounded by a transition region in which σ relaxes back to its vacuum value.

Unlike the polynomial potential of the previous section, the classical FLS potential possesses a flat direction. The most general renormalizable potential compatible with the symmetries also allows an intrinsic mass term and quartic self-interaction for the complex scalar, both of which have been omitted so that the portal interaction provides its only source of mass. Consequently,

$$\min V(\phi, \sigma)/|\phi|^2 \rightarrow 0 \quad (6.25)$$

as $\sigma \rightarrow 0$, implying that the classical lower edge of the existence window is $\omega_- = 0$. Radiative corrections generate an effective self-interaction for ϕ , lifting this flat

direction and restoring a finite lower threshold [166]. Since the present work concerns the classical equations of motion, we retain the simplified potential throughout.

6.3.1 Field equations and the two-field mechanical analogy

The construction of stationary FLS solitons follows exactly the variational procedure developed for the one-field Q-ball. The conserved $U(1)$ charge is fixed by introducing the same Lagrange multiplier ω , which again plays the dual role of chemical potential and phase-rotation frequency. Consequently, the stationary solution is obtained from the harmonic ansatz

$$\phi_Q(\mathbf{x}, t) = \phi_Q(r) e^{-i\omega t}, \quad \sigma_Q(\mathbf{x}, t) = \sigma_Q(r), \quad (6.26)$$

where the complex field carries the conserved charge while the real scalar remains static. Rather than repeating the derivation of such an ansatz, we simply use it to obtain the resulting coupled profile equations,

$$-\phi_Q''(r) - \frac{2}{r}\phi_Q'(r) - \omega^2\phi_Q(r) + \frac{1}{2}\frac{\partial V}{\partial\phi_Q} = 0, \quad -\sigma_Q''(r) - \frac{2}{r}\sigma_Q'(r) + \frac{\partial V}{\partial\sigma_Q} = 0. \quad (6.27)$$

Together with the regularity conditions at the origin and the vacuum boundary conditions at spatial infinity,

$$\phi_Q'(0) = 0, \quad \sigma_Q'(0) = 0, \quad \lim_{r \rightarrow \infty} \phi_Q(r) = 0, \quad \lim_{r \rightarrow \infty} \sigma_Q(r) = f. \quad (6.28)$$

these equations determine the stationary FLS soliton for each frequency ω .

As in the one-field case, the coupled profile equations admit a useful mechanical interpretation. The fictitious particle now moves in the two-dimensional field space (ϕ, σ) under the Newtonian potential

$$V_N(\phi_Q, \sigma_Q; \omega) = \omega^2\phi_Q^2 - V(\phi_Q, \sigma_Q), \quad (6.29)$$

while experiencing the same radial friction term $2/r$. The trajectory begins at an unknown point (ϕ_0, σ_0) with vanishing velocity, loses mechanical energy through friction, and must come to rest precisely at one of the maxima $(0, \pm f)$ of the Newtonian potential as $r \rightarrow \infty$. Throughout this thesis we choose the vacuum with $\sigma = f$ without loss of generality.

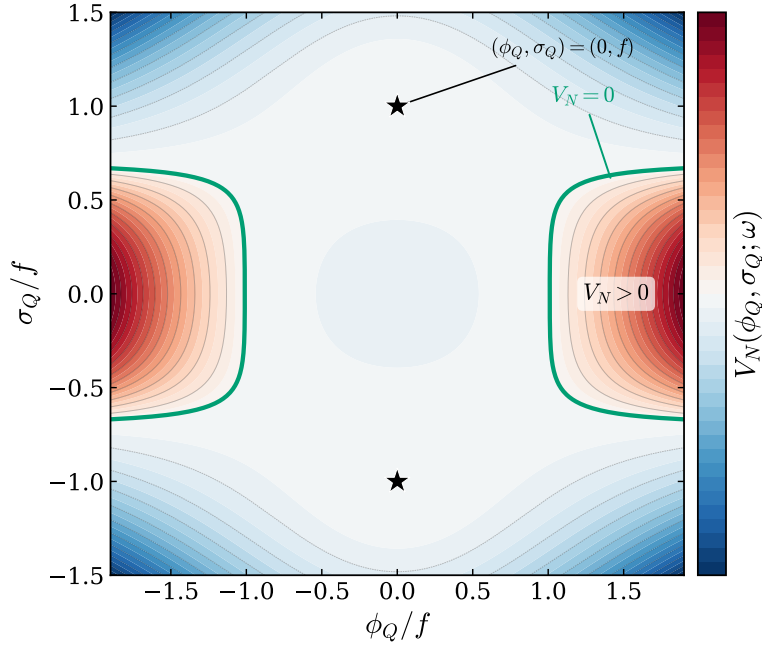


Figure 6.6: Contour plot of the Newtonian potential $V_N(\phi_Q, \sigma_Q; \omega)$. The green contour marks $V_N = 0$; the body's trajectory starts in the region $V_N > 0$ and ends at one of the local maxima $(\phi_Q, \sigma_Q) = (0, \pm f)$.

6.3.2 Numerical construction: multiple shooting

In contrast to the polynomial Q-ball, whose profile is determined by a single shooting parameter, the FLS soliton requires the simultaneous determination of two unknown central values, $\phi(0)$ and $\sigma(0)$. The resulting two-dimensional boundary value problem is highly sensitive to the initial conditions, making conventional shooting numerically unreliable. Throughout this thesis the profiles are instead constructed using the method of multiple shooting [167], in which the integration domain is divided into several subintervals whose solutions are matched iteratively. This substantially improves numerical stability while preserving direct control over the matching accuracy. The details of the implementation are presented in Appendix A.

Figure 6.7 shows representative FLS soliton profiles for $\kappa = 2.0$ and the same three frequencies considered for the polynomial Q-balls. As the frequency decreases, the charged field develops a broader core while the real scalar is increasingly driven away from its vacuum expectation value inside the soliton. The opposite limit, $\omega \rightarrow m$, corresponds to increasingly diffuse configurations that smoothly disappear into the vacuum, analogous to the thick-wall limit of the one-field Q-ball.

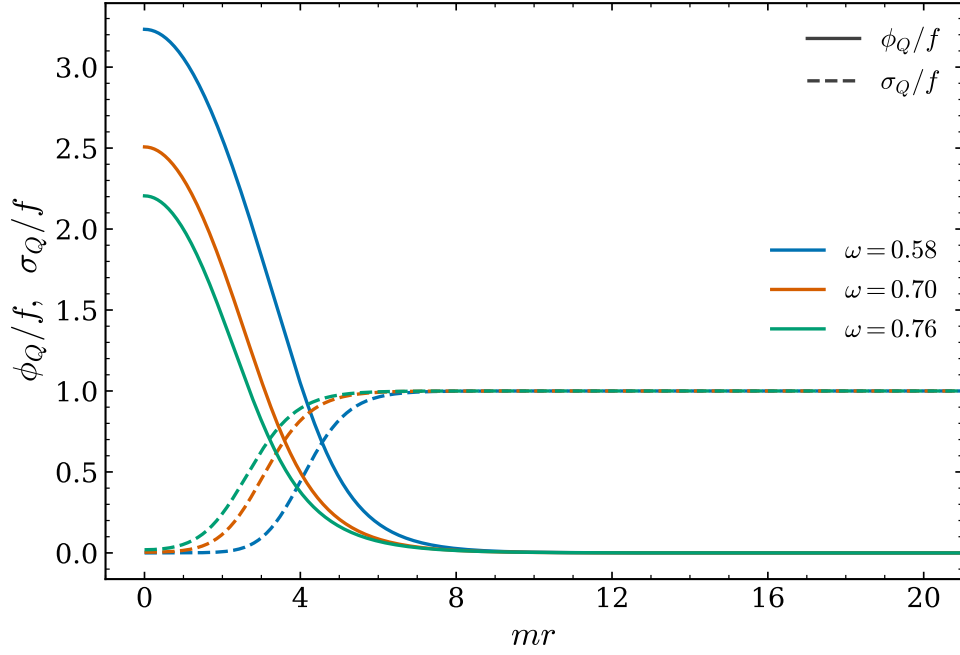


Figure 6.7: FLS soliton profiles obtained by multiple shooting, for $\kappa = 2.0$ for the three representative frequencies: $\omega = 0.58$ (blue), $\omega = 0.70$ (orange), $\omega = 0.76$ (green).

6.4 Summary

With the stationary profiles of both the one-field Q-ball and the FLS soliton now constructed, we have assembled the static backgrounds required for the perturbative analysis of Chapter 7. Two results from this construction carry forward directly. First, the identity $dE/dQ = \omega$ (Section 6.1) ties a Q-ball’s energy and charge to a single frequency label, and governs the exchange of energy and charge between perturbative excitations and the soliton. Second, the existence window $\omega_- < \omega < \omega_+$ and its stability subdivision into absolutely stable ($\omega_- < \omega < \omega_S$), metastable ($\omega_S < \omega < \omega_C$), and unstable ($\omega_C < \omega < m$) regimes (Section 6.1.4) fixes which backgrounds are physically sensible scattering targets.

The FLS soliton is built from the same variational principle and the same mechanical analogy, extended to the two-field plane (ϕ, σ) , but the physics is qualitatively different: the effective mass of the charged field is lowered inside the soliton by the portal coupling to σ , rather than by the self-interaction of a single field. It is this distinction that produces the richer perturbation spectrum of Chapter 7, where the two constructions are treated in parallel throughout.

Chapter 7

Scattering against Non-topological Solitons

Having established the existence and properties of the stationary Q-ball and FLS backgrounds in Chapter 6, we now turn to the central dynamical question of Part II: how these solitons interact with their surroundings. While their internal structure is encoded in the classical background solution itself, it is only through scattering that one can determine how they exchange charge and energy with external radiation, and ultimately how they respond to perturbations away from equilibrium.

For the Q-ball, this problem is qualitatively different from ordinary potential scattering. The time-dependent phase of the rotating background couples particle and antiparticle excitations into a genuinely two-channel system, allowing an incident mode to change its identity as it reflects off the soliton. This conversion process is the microscopic mechanism through which charge and energy are transferred between the radiation field and the Q-ball, and understanding it is the central objective of this chapter.

We proceed as follows. Section 7.1 derives the linearized two-channel equations for perturbations of a rotating Q-ball background and classifies the resulting spectrum. Section 7.2 builds the associated scattering matrix from first principles, states its unitarity and reciprocity properties, and settles precisely how the two channels relate to one another under $\lambda \rightarrow -\lambda$. Section 7.3 presents the resulting reflection and conversion probabilities. Section 7.4 returns to the Noether currents of Chapter 6 to derive the charge and energy exchanged during a scattering event, and shows that the result admits an exact thermodynamic interpretation in terms of the Q-ball's chemical potential. Section 7.5 verifies the regime of validity of the linear theory against fully non-linear lattice simulations. Section 7.6 extends the entire formalism to the two-field FLS soliton, whose additional scalar channel introduces a second,

higher scattering threshold and a genuinely new charge-conserving radiative process.

The results presented in this chapter are based on our work [168].

7.1 Linear perturbations of a rotating background

The perturbations on top of a static background are conceptually straightforward: since the background is time-independent, perturbations of different frequencies evolve independently, and the problem reduces to the familiar scattering of a wave from a localized potential. A Q-ball is stationary in a subtler sense. Although its radial profile is fixed, the field itself rotates uniformly in the complex plane with frequency ω . This seemingly simple difference fundamentally changes the structure of the perturbation theory: rather than evolving independently, particle and antiparticle excitations become coupled to one another through the background's rotation, turning what would otherwise be an ordinary one-channel problem into a genuinely coupled two-channel one. As we shall see, this coupling is precisely what allows the Q-ball to exchange charge and energy with the surrounding radiation.

7.1.1 The dichromatic ansatz and the two-channel equations

An exact treatment of the full non-linear scattering problem is generally out of reach. Throughout most of this chapter we therefore restrict ourselves to perturbations whose amplitude is small enough that their backreaction on the background can be neglected; the regime of validity of this approximation is tested explicitly in Section 7.5. We perturb the background as

$$\phi(\mathbf{x}, t) = \phi_Q(r) e^{-i\omega t} + \delta\phi(\mathbf{x}, t). \quad (7.1)$$

Substituting into the field equation (6.10) and retaining only terms linear in the perturbation [169] gives

$$\square \delta\phi + U(r) \delta\phi + M(r) e^{-2i\omega t} \delta\phi^* = 0, \quad (7.2)$$

$$\left. \frac{\partial^2 V}{\partial \phi^* \partial \phi} \right|_{\phi_Q} \equiv U(r), \quad \left. \frac{\partial^2 V}{\partial \phi^* \partial \phi^*} \right|_{\phi_Q} \equiv M(r) e^{-2i\omega t}, \quad (7.3)$$

where the second derivative with respect to ϕ^* carries the phase $e^{-2i\omega t}$ due to the profile, so $M(r)$ is the real, time-independent remainder. The two functions $U(r)$ and $M(r)$ play distinct physical roles. The first is the effective scattering potential familiar from ordinary linear perturbation theory; the second is unique to the

rotating background, coupling the perturbation directly to its own complex conjugate. Both are localized within the Q-ball core, approaching their free-theory values $U \rightarrow m^2$, $M \rightarrow 0$ at large distances (Figure 7.1), so that far from the Q-ball the two channels decouple into free Klein–Gordon waves.

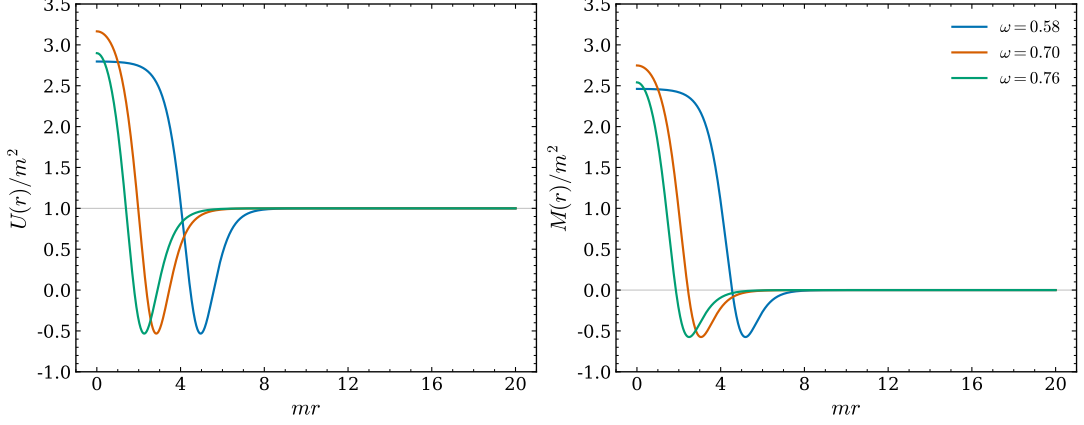


Figure 7.1: The potentials $U(r)$ (left), $M(r)$ (right) of the linearized equation (7.3), associated with the three backgrounds of Figure 6.4 ($\epsilon = 0.4$; $\omega = 0.58, 0.70, 0.76$). Both approach their free-theory values $(1, 0)$ outside the soliton core.

For an ordinary static potential, a disturbance of definite frequency remains at that frequency for all time: the linear equation is diagonal in Fourier space. Equation (7.3) does not have this property. The term $M(r)e^{-2i\omega t}\delta\phi^*$ ties $\delta\phi$ to its own complex conjugate through a phase rotating twice as fast as the background, so a single Fourier mode $\delta\phi \propto \varphi(\mathbf{x})e^{-iEt}$ cannot solve the equation on its own: substitution produces a term $\propto \varphi^*(\mathbf{x})e^{-i(2\omega-E)t}$, an oscillation at a different frequency with nothing already present to cancel it. This failure has nothing to do with the shape of U or M and everything to do with the fact that the background itself rotates.

The minimal fix allows $\delta\phi$ two frequencies rather than one, tied together in a specific way [156, 157],

$$\delta\phi(\mathbf{x}, t) = \sum_{\lambda} a_{\lambda} \varphi_{+, \lambda}(\mathbf{x}) e^{-iE_+ t} + a_{\lambda}^* \varphi_{-, \lambda}^*(\mathbf{x}) e^{+iE_- t}, \quad E_{\pm} \equiv \lambda \pm \omega. \quad (7.4)$$

This dichromatic decomposition resolves the problem exactly, for a reason worth stating plainly, since it is the observation on which the rest of this chapter turns,

$$E_+ - E_- = 2\omega, \quad (7.5)$$

identically, for every λ . Substituting (7.4) into (7.3), the troublesome potential term $M(r)e^{-2i\omega t}\delta\phi^*$ produces oscillations at $e^{i(E_+-2\omega)t}$ and $e^{-i(E_-+2\omega)t}$; by (7.5) these are exactly e^{iE_-t} and e^{-iE_+t} , the two frequencies the ansatz already contains. The

hierarchy stops after two terms because the background's own rotation frequency is exactly the gap between them: feed the equation two frequencies separated by 2ω , and the coupling closes the system on itself rather than generating a third.

Collecting the coefficients of the two phases e^{-iE_+t} , e^{iE_-t} defines the two-channel Schrödinger-like system

$$-\nabla^2 \Psi(\mathbf{x}) + \mathcal{U}(r) \Psi(\mathbf{x}) = \mathcal{E}^2 \Psi(\mathbf{x}), \quad (7.6)$$

$$\Psi(\mathbf{x}) = \begin{pmatrix} \varphi_+ \\ \varphi_- \end{pmatrix}, \quad \mathcal{U}(r) = \begin{pmatrix} U(r) & M(r) \\ M(r) & U(r) \end{pmatrix}, \quad \mathcal{E}^2 = \begin{pmatrix} E_+^2 & 0 \\ 0 & E_-^2 \end{pmatrix}. \quad (7.7)$$

The two labels $+$ and $-$ of the wavefunction components are, for now, purely formal: they distinguish which of the two coupled equations a mode function solves, and carry no physical (particle/antiparticle) content yet until Section 7.4. The two channels are independent far from the soliton, where $M \rightarrow 0$, and mixed only inside the core, where the background's rotation is felt.

Before turning to the scattering problem, it is worth asking what (7.6) conserves. A direct computation shows that although neither channel's own Schrödinger current is separately conserved, since M pushes probability from one channel into the other, their sum is:

$$\mathbf{J} \equiv \mathbf{J}_+ + \mathbf{J}_-, \quad \mathbf{J}_\pm \equiv \frac{1}{2i} (\varphi_\pm^* \nabla \varphi_\pm - \varphi_\pm \nabla \varphi_\pm^*), \quad \nabla \cdot \mathbf{J} = 0. \quad (7.8)$$

Indeed $\nabla \cdot \mathbf{J}_+ \propto M \operatorname{Im}(\varphi_+^* \varphi_-)$ and $\nabla \cdot \mathbf{J}_- \propto -M \operatorname{Im}(\varphi_+^* \varphi_-)$ are equal and opposite and cancel in the sum. This single conserved current is what turns (7.6) into an honest scattering problem with a well-defined notion of incoming and outgoing flux, and Section 7.2 builds the formalism directly on it.

7.1.2 Mode classification

Far outside the soliton, $U \rightarrow m^2$ and $M \rightarrow 0$, so the two channels decouple into free waves with wavenumbers $k_\pm^2 = E_\pm^2 - m^2$. The signs of $k_\pm^2$ classify the spectrum of $\delta\phi$ into four kinds of mode, according to whether each channel is free to propagate or bound to the soliton.

- *Zero modes*, $\lambda = 0$, encode the perturbations along the moduli mentioned in Section 6.1.5. For the phase mode, $\delta\phi = -i\phi_Q(r) e^{-i\omega t}$, substituting into (7.3) gives us

$$U \delta\phi + M e^{-2i\omega t} \delta\phi^* = (U - M) \delta\phi, \quad (7.9)$$

$$U - M = \left(\frac{\partial^2 V}{\partial \phi^* \partial \phi} - \left| \frac{\partial^2 V}{\partial \phi^* \partial \phi^*} \right| \right)_{\phi_Q} = \frac{1}{\phi} \frac{\partial V}{\partial \phi^*} \Big|_{\phi_Q}, \quad (7.10)$$

which reduces the equation exactly to the background profile equation (6.11). For the translation modes, $\delta\phi = \nabla\phi_Q(r) e^{-i\omega t}$, the substitution gives

$$U \delta\phi + M e^{-2i\omega t} \delta\phi^* = (U + M) \delta\phi = \nabla \left(\frac{\partial V}{\partial \phi^*} \Big|_{\phi_Q} \right), \quad (7.11)$$

and the equation reduces to the gradient of the profile equation (6.11).

- *Bound modes*, $0 < \lambda < m - \omega$: the energies of the two channels are $E_-^2 < E_+^2 < m^2$, so both partner modes decay exponentially and exist only at discrete λ , more readily for thinner-walled, larger Q-balls; a deeper, wider potential well supports more bound states.
- *Half-scattering modes*, $m - \omega < \lambda < m + \omega$: one channel is free while the other remains bound, since we have $E_-^2 < m^2 < E_+^2$. An incident free mode then reflects with unit probability, having no way to convert; we return to this simple but necessary limit in Section 7.3.
- *Scattering modes*, $\lambda > m + \omega$: both channels are free, and the soliton can convert one into the other. Only in this regime can charge and energy actually be exchanged with the radiation field, and it is the only regime the remainder of this chapter is concerned with.

The decomposition (7.4) double-counts: sending $\lambda \rightarrow -\lambda$ maps the $+$ and $-$ equations into one another, so λ and $-\lambda$ describe the same physical content with the two channels exchanged $\varphi_+ \leftrightarrow \varphi_-$. We take $\lambda \geq 0$ without loss of generality when classifying modes, and return to the full two-sided domain in Sections 7.2.4–7.3, where it becomes a convenient way of visualizing the charge-conjugate scattering process without solving anything new.

7.2 The two-channel S-matrix

Having established the coupled two-channel equations and the current they conserve, we now build the scattering matrix they define.

7.2.1 Partial waves and the s-wave reduction

The potentials $U(r), M(r)$ inherit the spherical symmetry of the Q-ball profile, so a general perturbation separates into partial waves, $\Psi(\mathbf{x}) = \sum_{\ell} \psi_{\ell}(r) Y_{\ell m}(\hat{\mathbf{x}})$, where

$\psi_\ell = (\varphi_{+, \ell}, \varphi_{-, \ell})$, with each angular momentum sector obeying its own decoupled radial system,

$$-\psi_\ell'' - \frac{2}{r}\psi_\ell' + \frac{\ell(\ell+1)}{r^2}\psi_\ell + \mathcal{U}(r)\psi_\ell = \mathcal{E}^2\psi_\ell. \quad (7.12)$$

We restrict ourselves to s-waves throughout the remainder of this chapter, and thus, we drop the subscript ℓ accordingly. The construction below applies unchanged, channel by channel, to every $\ell > 0$; we develop the resulting cross section, needed for a genuine thermal accretion-rate calculation, separately in Appendix B.

For the s-wave components of the modes $\ell = 0$, the system (7.12) becomes

$$-\psi'' - \frac{2}{r}\psi' + \mathcal{U}(r)\psi = \mathcal{E}^2\psi, \quad (7.13)$$

and the regularity at the origin imposes the initial conditions, $\varphi'_\pm(0) = 0$, while keeping $\varphi_\pm(0) \neq 0$ (Only for the s-wave). In the scattering regime, $\lambda > m + \omega$, both channels admit the flux-normalized asymptotic form

$$\varphi_\pm(r) \xrightarrow{r \rightarrow \infty} \frac{1}{\sqrt{4\pi k_\pm} r} \left(A_\pm^{\text{in}} e^{-ik_\pm r} + A_\pm^{\text{out}} e^{+ik_\pm r} \right), \quad k_\pm \equiv \sqrt{E_\pm^2 - m^2}, \quad (7.14)$$

normalized so that $|A_\pm|^2$ is directly the particle flux carried by that channel, with no extra kinematic prefactor to track.

7.2.2 Construction of the S-matrix

The propagating behavior of the modes in the far region is better studied by solving (7.12) twice, with two linearly independent regular initial data at $r = 0$, and integrating both solutions outward to obtain two solution vectors $\psi^{(1)}(r), \psi^{(2)}(r)$. Matching each onto the asymptotic form (7.14) defines two 2×2 matrices,

$$\left(A^{\text{in}} \right)_{sj} \equiv A_s^{\text{in},(j)}, \quad \left(A^{\text{out}} \right)_{sj} \equiv A_s^{\text{out},(j)}, \quad s = \pm, \quad j = 1, 2, \quad (7.15)$$

whose columns record the incoming and outgoing content of the j -th regular solution. Because *every* regular solution is a linear combination $\psi = c_1\psi^{(1)} + c_2\psi^{(2)}$, and no other solution is admissible (a singular piece at the origin is excluded by regularity), the incoming and outgoing amplitudes of any physical scattering solution are related by $A^{\text{out}} \cdot c = S A^{\text{in}} \cdot c$ for every $c = (c_1, c_2)$, i.e.,

$$S(\lambda) \equiv A^{\text{out}}(\lambda) \left[A^{\text{in}}(\lambda) \right]^{-1}, \quad (7.16)$$

with $(A_{\pm}^{\text{out}}) = S(\lambda)(A_{\pm}^{\text{in}})$. If the two regular solutions are chosen (or subsequently recombined) so that A^{in} is already the identity matrix, S is simply read off as A^{out} ; this is the case used for the numerics of Section 7.3, where each column of S is obtained directly by solving the scattering problem with unit incoming amplitude in one channel and none in the other. In either case, the physical content of S is the same: its first column is the pair of amplitudes (reflection, conversion) produced by a unit-amplitude incoming $+$ mode, and its second column the same pair for an incoming $-$ mode,

$$\begin{pmatrix} A_+^{\text{out}} \\ A_-^{\text{out}} \end{pmatrix} = \begin{pmatrix} S_{++} & S_{+-} \\ S_{-+} & S_{--} \end{pmatrix} \begin{pmatrix} A_+^{\text{in}} \\ A_-^{\text{in}} \end{pmatrix}, \quad (7.17)$$

so that $|S_{++}|^2$ is the probability that an incoming $+$ mode reflects as a $+$ mode, $|S_{-+}|^2$ the probability that it converts into an outgoing $-$ mode, and correspondingly for the second column.

7.2.3 Properties of the S-matrix

If we write $\psi = r^{-1}u$, then writing Eq. (7.13) in terms of u removes the friction term $-2/r$. The Wronskian defined on two independent solutions $u^{(1)}, u^{(2)}$ is constant,

$$W[u^{(1)} \equiv r\psi^{(1)}, u^{(2)} \equiv r\psi^{(2)}] = \sum_s (r\varphi_s^{(1)})(r\varphi_s^{(2)})' - (r\varphi_s^{(1)})'(r\varphi_s^{(2)}). \quad (7.18)$$

The regularity at the origin ensures that $W = 0$. Eq. (7.13) is built from real operators, then we can consider the two independent solutions are conjugate to each other, $\psi^{(2)} = \psi^{(1)*} = \psi^*$, then

$$W = \sum_s (r\varphi_s)(r\varphi_s^*)' - (r\varphi_s)'(r\varphi_s^*) = r^2 \sum_s \varphi_s \varphi_s'^* - \varphi_s' \varphi_s^* = 0. \quad (7.19)$$

The r -independence of the Wronskian is equivalent to the conservation of $\mathbf{J}$, since we can write $\partial_r W = -2i r^2 \nabla \cdot \mathbf{J} = -2i \partial_r (r^2 J_r) = 0$. Furthermore, since $W = 0$, then we can prove the S-matrix is *unitary*,

$$W = r^2 \sum_s \varphi_s \varphi_s'^* - \varphi_s' \varphi_s^* = \frac{i}{2\pi} \sum_s |A_s^{\text{out}}|^2 - |A_s^{\text{in}}|^2 = 0 \implies S^\dagger S - \mathbf{1} = 0. \quad (7.20)$$

Now, we can take two genuinely independent solutions, then the S-matrix can be shown to be symmetric; this property is called *reciprocity*,

$$W = \frac{i}{2\pi} \sum_s A_s^{\text{in},(1)} A_s^{\text{out},(2)} - A_s^{\text{in},(2)} A_s^{\text{out},(1)} = 0 \implies S^T - S = 0. \quad (7.21)$$

Eqs. (7.20) and (7.21) are the two properties of $S(\lambda)$ used throughout the rest of this chapter, and both generalize without modification to the 3×3 FLS S-matrix of Section 7.6.

7.2.4 The $\lambda \rightarrow -\lambda$ exchange

We can now make precise the redundancy of the decomposition (7.4), in the unambiguous form needed for Section 7.3. Since $U(r)$ and $M(r)$ are properties of the static background alone and carry no dependence on λ , while $E_+^2(-\lambda) = (-\lambda + \omega)^2 = E_-^2(\lambda)$ and $E_-^2(-\lambda) = (-\lambda - \omega)^2 = E_+^2(\lambda)$, the radial system (7.12) evaluated at $-\lambda$ for (φ_+, φ_-) is *identical* to the system evaluated at $+\lambda$ for (φ_-, φ_+) . Since the boundary conditions (regularity at the origin, the asymptotic form (7.14)) are themselves symmetric under exchanging the two channels, the mode functions and the S-matrix inherit the exact relation

$$S(-\lambda) = P S(\lambda) P, \quad P \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (7.22)$$

Solving the “+” scattering problem at $-\lambda$ computes precisely $S_{--}(|\lambda|)$ and $S_{+-}(|\lambda|)$, which is exactly the reflection and conversion data of an incoming $-$ mode at $|\lambda|$, with the two channels simply relabelled. What Eq. (7.22) does *not* yet say is which of the two channels is the physical particle and which the antiparticle; that identification requires the charge each channel carries, which we introduce only in Section 7.4, precisely because it is not fixed by (7.22) alone. Until then we consistently use the labels “+” and “−”, not “particle” and “antiparticle”.

7.3 Reflection and conversion probabilities

We solve (7.12) numerically for unit incoming + amplitude, $A_+^{\text{in}} = 1, A_-^{\text{in}} = 0$, following the construction of Section 7.2.2. Figure 7.2 shows the resulting reflection probability $|A_+^{\text{out}}|^2 = |S_{++}|^2$ over $\lambda > m + \omega$, for our three representative backgrounds. For visualization we extend the plot to $\lambda < -(m + \omega)$ using the exact exchange relation (7.22): a point on the negative branch is not an independent calculation but, by that relation, exactly the reflection probability $|A_-^{\text{out}}|^2 = |S_{--}(-\lambda)|^2$ of an incoming $-$ mode.

Between the two scattering windows, $|\lambda| < m + \omega$, the plot is omitted rather than populated with a curve. This region contains both the bound modes ($|\lambda| < m - \omega$) and the half-scattering modes ($m - \omega < |\lambda| < m + \omega$) of Section 7.1.2. In the half-

scattering window only one of the two asymptotic momenta $k_{\pm}$ is real, so an incident mode in the open channel has no free channel to convert into; unitarity (7.20) then forces $|A_+^{\text{out}}|^2 = |S_{++}|^2 = 1$ trivially, by kinematics alone, with no dynamical content to plot. The gap is therefore a genuine kinematic exclusion zone, and its edges at $|\lambda| = m + \omega$ mark the true onset of the physics this chapter is about.

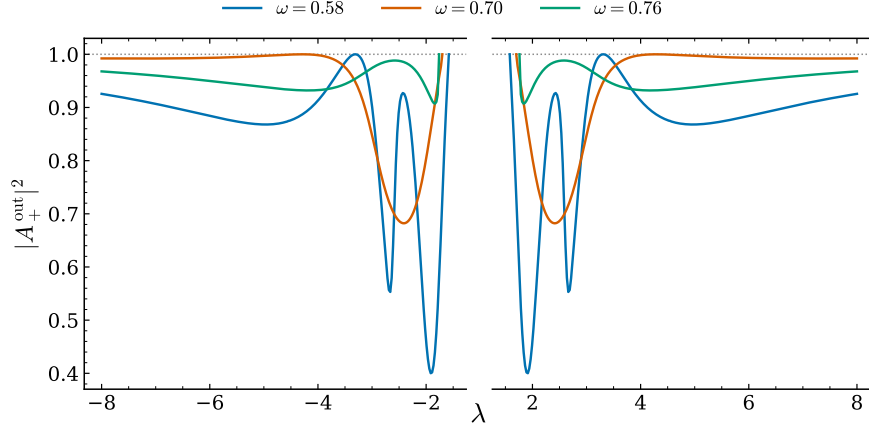


Figure 7.2: The reflection probability $|A_+^{\text{out}}|^2 = |S_{++}|^2$ over the two-sided λ domain, for an incoming $+$ mode. Larger Q -balls, $\omega \rightarrow \omega^-$ allow for more conversion probability for the incident mode. As the energy of the incoming mode increases, the reflection process becomes more dominant $|A_+^{\text{out}}|^2 \rightarrow 1$.

7.4 Charge and energy transfer

We now return to the conserved currents of Section 6.1. The number current $\mathbf{J}$ of Section 7.1 resulted in the conservation relation (7.20). However, the Noether currents: $U(1)$ charge current (6.2) and stress-energy tensor (6.3) quantify the charge and energy fluxes exchanged in the scattering. We expand (6.2), (6.3) to quadratic order in $\delta\phi$, then by averaging over time, the linear terms average out to zero since they oscillate with frequency λ . The fate of quadratic terms depends on the mode content of the term: same-mode terms are time-independent so they survive, while cross terms oscillate with frequency 2λ so they average out to zero as well. The time-averaged radial components of the currents will be used to get the flux of charge J_Q^r and energy $J_E^r \equiv T^{rt}$,

$$\langle J_Q^r \rangle = -i \sum_s s (\varphi_s^* \partial_r \varphi_s - \partial_r \varphi_s^* \varphi_s), \quad (7.23)$$

$$\langle J_E^r \rangle = -i \sum_s E_s (\varphi_s^* \partial_r \varphi_s - \partial_r \varphi_s^* \varphi_s). \quad (7.24)$$

We substitute with the asymptotic form (7.14), we get

$$\langle J_Q^r \rangle \approx \frac{2}{4\pi r^2} \sum_s s (|A_s^{\text{out}}|^2 - |A_s^{\text{in}}|^2), \quad (7.25)$$

$$\langle J_E^r \rangle \approx \frac{2}{4\pi r^2} \sum_s E_s (|A_s^{\text{out}}|^2 - |A_s^{\text{in}}|^2). \quad (7.26)$$

By integrating over all directions, we get the charge and energy fluxes

$$\mathcal{F}_Q = 2 \sum_s q_s (|A_s^{\text{out}}|^2 - |A_s^{\text{in}}|^2), \quad (7.27)$$

$$\mathcal{F}_E = 2 \sum_s E_s (|A_s^{\text{out}}|^2 - |A_s^{\text{in}}|^2), \quad (7.28)$$

where $q_s = s (= \pm 1)$ is the $U(1)$ charge carried by mode s . This weight is essential for the charge flux, since a $-$ mode genuinely carries the opposite charge to a $+$ mode; it is unnecessary for the energy flux, since E_s already carries its own sign, and energy does not flip sign between particle and antiparticle. Unlike $\mathbf{J}$, neither current vanishes at the origin, since both are conserved for the full field rather than for the perturbation alone.

Throughout the rest of the chapter, we take $\omega > 0$, so that $Q > 0$ in Eq. (6.16): the Q-ball is positively charged, and $q_+ = +1$ is genuinely the charge of the mode that is a particle at $\lambda > 0$. The charge-conjugate, negatively charged soliton is obtained by $\omega \rightarrow -\omega$ throughout, and is not treated separately here.

With $\mathcal{F}_Q^{\text{in}}, \mathcal{F}_E^{\text{in}}$ and $\mathcal{F}_Q^{\text{out}}, \mathcal{F}_E^{\text{out}}$ built from the incoming and outgoing amplitudes separately, we can define the amplification factors

$$1 + Z_Q \equiv \frac{\mathcal{F}_Q^{\text{out}}}{\mathcal{F}_Q^{\text{in}}}, \quad 1 + Z_E \equiv \frac{\mathcal{F}_E^{\text{out}}}{\mathcal{F}_E^{\text{in}}}, \quad (7.29)$$

where a positive sign signals the extraction of charge or energy from the Q-ball, and a negative signals the deposition of charge or energy into it. ¹

With $A_+^{\text{in}} = 1, A_-^{\text{in}} = 0$, an incoming $+$ mode either reflects, with probability $|A_+^{\text{out}}|^2$, or converts into an outgoing $-$ mode, with probability $|A_-^{\text{out}}|^2 = 1 - |A_+^{\text{out}}|^2$. At $\lambda > 0$, where the $+$ mode is a genuine particle of charge $+1$, this is the process

$$(Q) + \phi \rightarrow (Q) + \phi, \quad (Q) + \phi \rightarrow (Q + 2) + \bar{\phi}. \quad (7.30)$$

¹This is trivial for the particle $\lambda > 0$ being represented by a $+$ modes. When, we consider an antiparticle $\lambda < 0$ being represented by a $+$ mode as well, the sign of Z_Q needs to be corrected to account for $q = -1$ of the mode.

Substituting the unitarity relation (7.20) into (7.29) gives, exactly,

$$Z_Q = -2|A_-^{\text{out}}|^2, \quad Z_E = -\frac{2\omega}{E_+}|A_-^{\text{out}}|^2. \quad (7.31)$$

Both are exact consequences of unitarity alone, with no approximation yet made; Figure 7.3 shows them over the same two-sided λ domain as Figure 7.2.

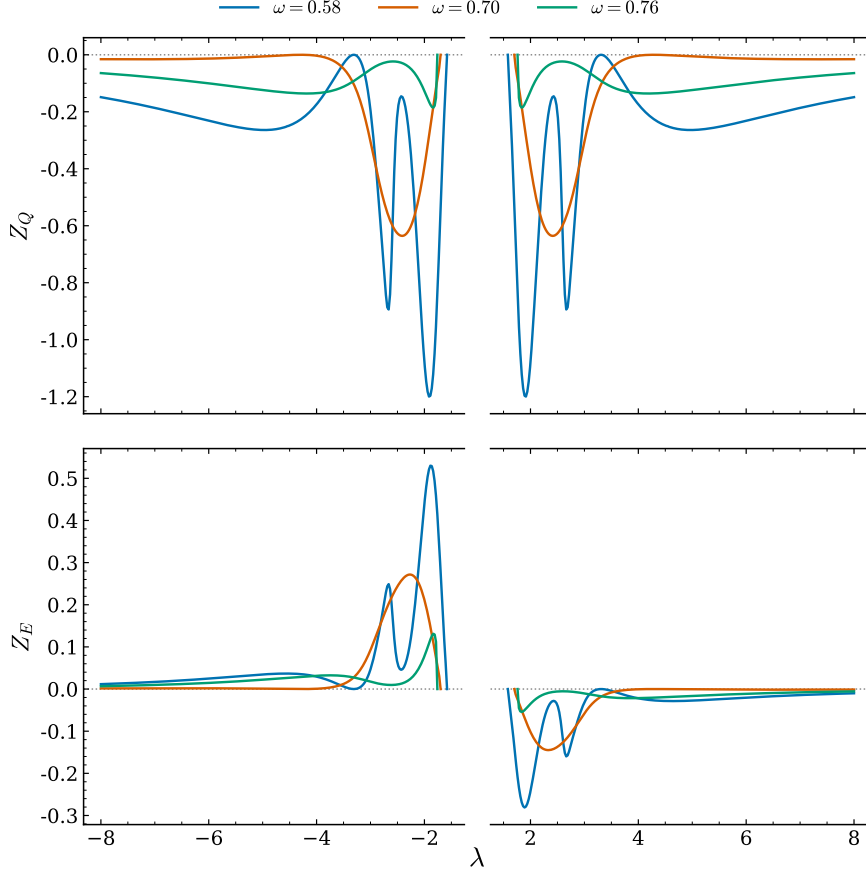


Figure 7.3: Charge and energy amplification factors Z_Q, Z_E of Eq. (7.31) over the two-sided domain, for an incoming $+$ mode. When incoming is a particle state, $\lambda > 0$, it deposits charge and energy into the Q-ball $Z_Q < 0, Z_E < 0$. When it is an antiparticle state, $\lambda < 0$, it extracts charge and energy from the positively-charged Q-ball, $Z_Q < 0$ (we need to correct for $q_{\text{in}} = -1$), and $Z_E > 0$ (which is readily correct since $E_{\text{in}} < 0$).

The sign of Z_Q on the two branches. With $A_+^{\text{in}} = 1$ fixed, Z_Q in (7.31) is manifestly negative for every λ , on both branches of Figure 7.3. This does not mean the Q-ball only ever absorbs charge. By Eq. (7.22), the result of the negative branch $|S_{-+}(\lambda < 0)|^2$ is the same $-2|A_+^{\text{out}}|^2 = -2|S_{+-}(\lambda > 0)|^2$. Solving the “incoming $+$ ” problem at $\lambda < 0$ computes exactly the scattering data of an incoming $-$ mode (an antiparticle, of charge -1) at $\lambda > 0$, and the conversion process there is the charge

conjugate of (7.30),

$$(Q) + \bar{\phi} \rightarrow (Q - 2) + \phi, \quad (7.32)$$

in which the Q-ball genuinely *loses* charge. The reason Z_Q does not display this by its sign is that its normalization in (7.27) fixes $q_+ = +1$ regardless of which physical species is actually incoming; Z_Q therefore reports the probability of converting the incoming species, not the sign of the Q-ball's own charge change, and that sign must instead be read off from which branch, and hence which physical species, is being examined. Energy has no analogous ambiguity, since $\mathcal{F}_E^{\text{in}}$ is built from the actual signed energy E_+ of whatever is incoming, and E_+ itself changes sign between the two branches; consequently Z_E flips sign correctly on its own, negative on the branch of (7.30) and positive on its conjugate (7.32).

7.4.1 Kinematic interpretation

The result (7.31) has a purely kinematic reading. Recall the established thermodynamical relation (6.8) where $\omega = dE/dQ$ is the chemical potential of the Q-ball, the energy needed to add one unit of charge; transferring $\Delta Q = 2$ should therefore cost $\Delta E \approx 2\omega$. Concretely,

$$Q - (Q + 2) = -2, \quad E(Q) - E(Q + 2) = -2 \frac{dE}{dQ} = -2\omega. \quad (7.33)$$

Weighting the conversion probability $|S_{-+}|^2$ by these two differences reproduces (7.31) exactly provided we divide by the charge/energy of the incident particle: the dynamical S -matrix result and the equilibrium thermodynamics of the background agree, because ω is nothing but the energy cost of the two units of charge the process actually transfers. This is, to our knowledge, the first time this dynamical amplification factor has been shown to coincide with the thermodynamic data of the background alone, and it is the direct fulfillment of the objective set out in Section 5.3.

7.5 Lattice verification

The linear theory of Sections 7.1–7.4.1 assumes the perturbation stays small enough that its own backreaction on the background is negligible. We test this by evolving the exact, non-linear field equation of Chapter 6 on a spatial lattice: a fourth-order finite-difference discretization in r , evolved in time with a fourth-order Runge–Kutta integrator.

The initial data is the background profile plus an inward-directed, Gaussian-enveloped wave packet of carrier frequency $E_+ = \lambda + \omega$, centered outside the soliton core at radius r_0 with width σ_r ,

$$\delta\phi(r, 0) = \delta \left(\frac{r_0}{r} \right) \exp \left[-\frac{(r - r_0)^2}{2\sigma_r^2} \right] e^{-ik_+ r}, \quad \partial_t \delta\phi(r, 0) = -iE_+ \delta\phi(r, 0). \quad (7.34)$$

The wave-packet is set to start approaching the Q-ball starting from $t = 0$ (with r_0, σ_r chosen so that the packet does not yet overlap the soliton core, and σ_r large enough compared to $1/k_+$ to resolve the carrier frequency, but small compared to r_0 so that the packet is well localized on arrival). Figure 7.4 shows the resulting scattering event directly, as a contour of the perturbation amplitude in the (r, t) plane: the incoming and outgoing wavefronts are both clearly resolved.

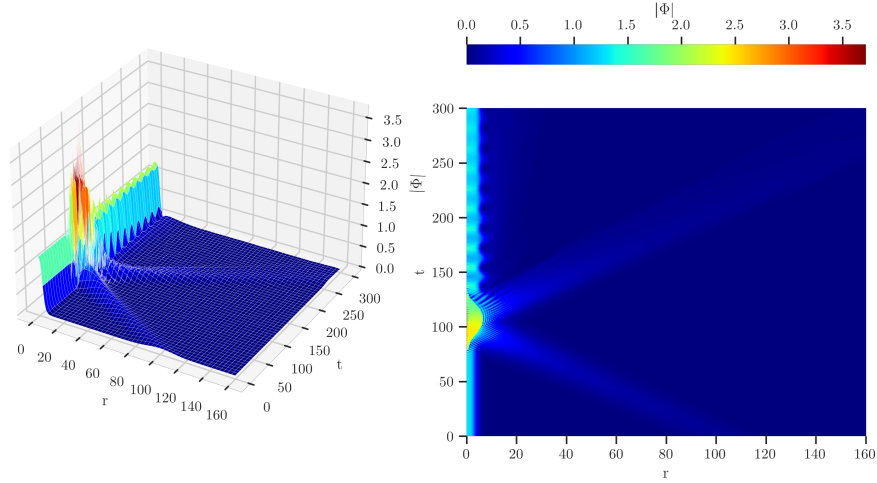


Figure 7.4: Non-linear lattice evolution of the perturbation amplitude in the (r, t) plane, for an incoming wave packet of the form (7.34), with $\omega = 0.70$, $\lambda = 2.0$, $\delta = 0.05$.

To extract Z_E from a single run, we track the energy current J_E^r through a detector sphere at fixed radius, and its flux integrated up to time t , shown in Figure 7.5.

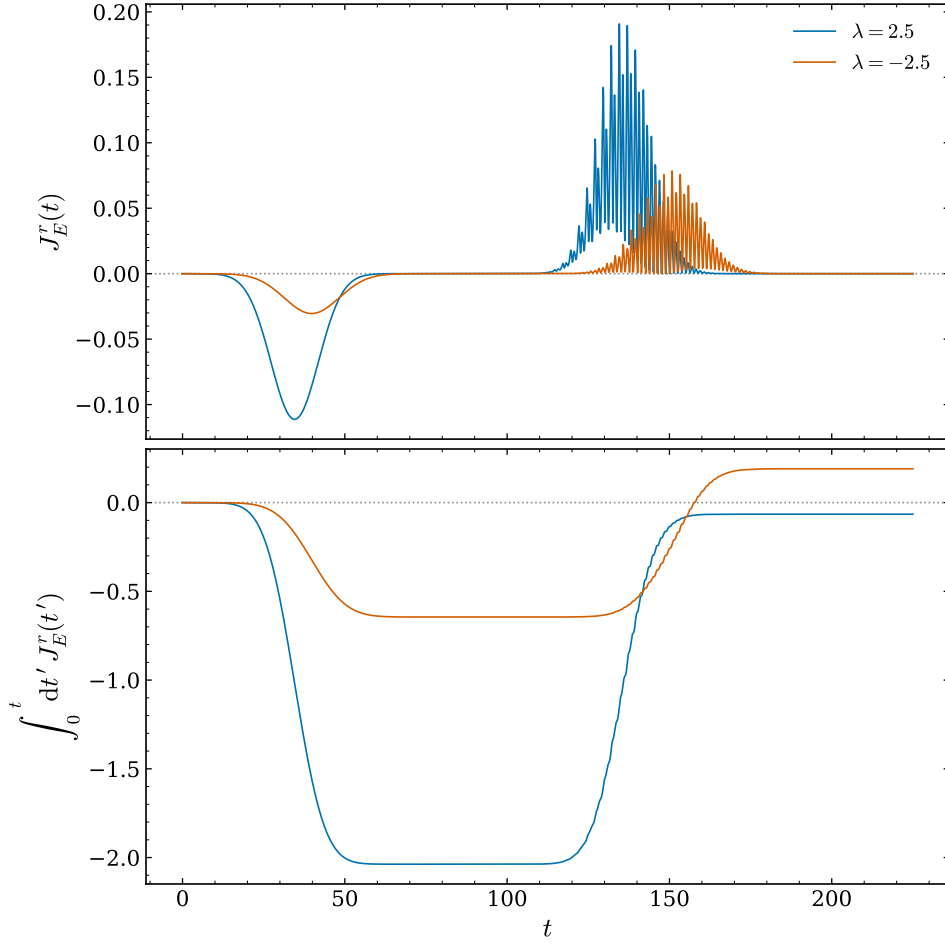


Figure 7.5: Energy current $J_E^r(r, t)$ and its flux accumulated through a fixed detector sphere, with the mid-time and late-time values used in (7.35) marked.

At an intermediate time, once the incoming packet has crossed the detector but before it has reached the soliton, the accumulated flux is due entirely to the incoming wave crossing inward, opposite to the outward normal, so $F_E(\text{mid}) = -\mathcal{F}_E^{\text{in}}$. At late times, once the scattering event is complete and all outgoing radiation has crossed back out, the accumulated flux is the full difference $F_E(\text{late}) = \mathcal{F}_E^{\text{out}} - \mathcal{F}_E^{\text{in}} = \mathcal{F}_E$. Combining these with the definition (7.29) gives

$$Z_E = \frac{\mathcal{F}_E}{\mathcal{F}_E^{\text{in}}} = -\frac{F_E(\text{late})}{F_E(\text{mid})}, \quad (7.35)$$

a measurement requiring only the two values of a single accumulated flux, with no need to isolate the incoming and outgoing packets in the field data directly.

The amplitude scan. The effective amplitude of the wavepacket grows as it gets closer to the Q-ball such that its amplitude when it is about to hit the Q-ball is

given by

$$\delta_{\text{impact}} = \delta \times \frac{r_0}{R_Q}, \quad (7.36)$$

where R_Q is the Q-ball radius. We therefore report the validity range in terms of the ratio of the wavepacket amplitude upon collision to the Q-ball profile itself,

$$\delta_r \equiv \frac{\delta \times r_0}{R_Q \times \phi_Q(0)}. \quad (7.37)$$

Repeating the flux measurement over a scan of δ (equivalently δ_r) at fixed λ, ω , and comparing the lattice-extracted Z_E of (7.35) against the linear prediction of (7.31) at the same λ, ω , gives Figure 7.6.

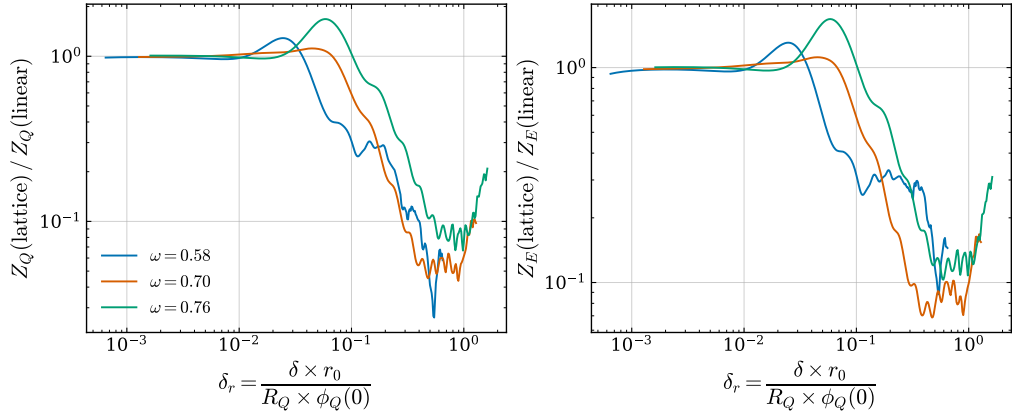


Figure 7.6: Energy amplification factor Z_E from the non-linear lattice evolution, (7.35), against the linear prediction (7.31), as a function of the rescaled amplitude δ_r of (7.37). Agreement holds for $\delta_r \lesssim 10^{-2}$.

The two agree closely for $\delta_r \lesssim 10^{-2}$, and diverge above it, marking the amplitude range over which the linear theory of this chapter can be trusted; the residual discrepancies below this scale are dominated by finite wave-packet width close to threshold (where one of $k_{\pm}$ is small and a correspondingly broader packet is needed to resolve it), the numerical precision of the background profile itself, which otherwise radiates a small amount of spurious charge and energy that contaminates the measurement, and ordinary discretization error.

Physical implications. The bound on the relative impact amplitude $\delta_r \lesssim 10^{-2}$ is only useful once translated into a statement about an actual radiation bath rather than a single wave packet. For a localized packet, the amplitude is not affected by the geometric focusing that a converging spherical shell undergoes, so the condition on the relative amplitude is simply $\delta/\phi_Q(0)$ rather than the focused amplitude used for the shell; the packet's coherence volume is correspondingly σ^3 . Imposing that

the packet carries unit charge gives

$$Q = 1 \sim E \times \delta^2 \sigma^3, \quad (7.38)$$

so that the bound reads

$$\frac{\delta}{\phi_Q(0)} \sim \frac{E^{-1/2} \sigma^{-3/2}}{\phi_Q(0)} \lesssim 10^{-2}. \quad (7.39)$$

The coherence length σ here is a single-particle quantity, the size of one incident quantum's own wave packet, not a collective plasma scale such as the mean free path or inter-particle spacing; for a thermal particle in the nonrelativistic regime it is set by the spread in thermal momentum, $\sigma \sim (mT)^{-1/2}$, with energy dominated by the rest mass, $E \sim m$, since $E = \sqrt{p^2 + m^2} \approx m$ for $p \sim \sqrt{mT} \ll m$. Together with $E \sim m \sim f \sim \phi_Q(0)$, this gives

$$T \lesssim 10^{-8/3} m \sim 2 \times 10^{-3} m. \quad (7.40)$$

Above this temperature, the wave packets available in the thermal bath are too short-lived for the linear description to apply at all.

7.6 Perturbations of the FLS soliton

We now extend the formalism of Sections 7.1–7.2 to the two-field FLS soliton of Section 6.3. The construction is identical in spirit to the Q-ball perturbation with the novelty of a third channel and an indirect mixing of $+$, $-$ modes. The third channel sets its own threshold, so that the mode classification of Section 7.1.2 acquires additional windows, and the S-matrix of Section 7.2 grows from 2×2 to 3×3 (in the window where all channels are open).

7.6.1 Linearized equations and mode classification

We follow the same procedure we have carried out for the Q-ball; we perturb the FLS background

$$\phi(\mathbf{x}, t) = \phi_Q(r) e^{-i\omega t} + \delta\phi(\mathbf{x}, t), \quad (7.41)$$

$$\sigma(\mathbf{x}, t) = \sigma_Q(r) + \delta\sigma(\mathbf{x}, t), \quad (7.42)$$

and by this we linearize the field equations of the FLS theory Eq. (6.24),

$$\square \delta\phi + U_\phi(r) \delta\phi + M(r) e^{-i\omega t} \delta\sigma = 0, \quad (7.43)$$

$$\square \delta\sigma + U_\sigma(r) \delta\sigma + M(r) (e^{+i\omega t} \delta\phi + e^{-i\omega t} \delta\phi^*) = 0, \quad (7.44)$$

$$\left. \frac{\partial^2 V}{\partial \phi^* \partial \phi} \right|_{(\phi, \sigma)_Q} \equiv U_\phi(r), \quad \left. \frac{\partial^2 V}{\partial \sigma \partial \sigma} \right|_{(\phi, \sigma)_Q} \equiv U_\sigma(r), \quad \left. \frac{\partial^2 V}{\partial \phi^* \partial \sigma} \right|_{(\phi, \sigma)_Q} \equiv M(r) e^{-i\omega t}. \quad (7.45)$$

The mixing of the modes of the FLS soliton needs to go through σ modes.

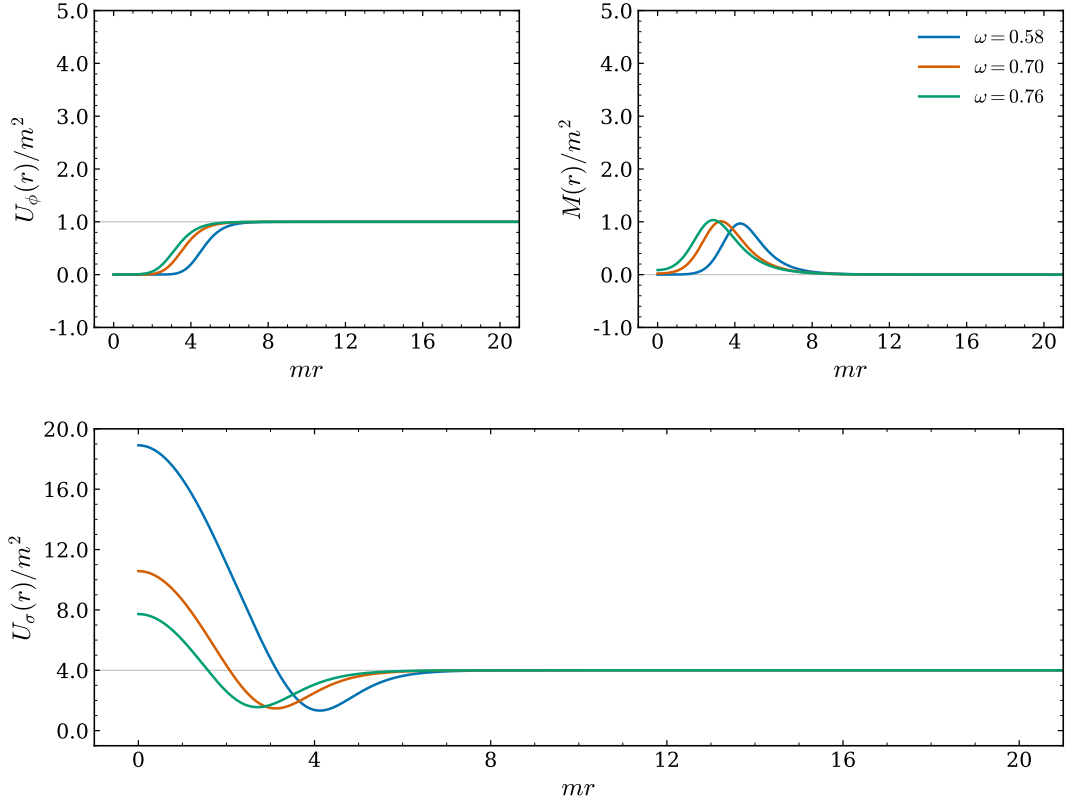


Figure 7.7: The potentials $U_\phi(r)$ (top left), $M(r)$ (top right), $U_\sigma(r)$ (bottom) of the linearized system (7.43), (7.44), associated with the three FLS backgrounds of Figure 6.7 ($\kappa = 2.0$; $\omega = 0.58, 0.70, 0.76$). All of them approach their free-theory values $(1, 0, \kappa^2 = 4)$ outside the soliton core.

Following the same reasoning as for the Q-ball, we choose to decompose the perturbations through a dichromatic ansatz for the charged field and a monochromatic ansatz for the neutral field,

$$\delta\phi(\mathbf{x}, t) = \sum_{\lambda} a_{\lambda} \varphi_{+, \lambda}(\mathbf{x}) e^{-iE_+ t} + a_{\lambda}^* \varphi_{-, \lambda}^*(\mathbf{x}) e^{+iE_- t}, \quad E_{\pm} \equiv \lambda \pm \omega, \quad (7.46)$$

$$\delta\sigma(\mathbf{x}, t) = \sum_{\lambda} a_{\lambda} \chi_{0, \lambda}(\mathbf{x}) e^{-iE_0 t} + a_{\lambda}^* \chi_{0, \lambda}^*(\mathbf{x}) e^{+iE_0 t}, \quad E_0 \equiv \lambda. \quad (7.47)$$

Consequently, we can define the three-channel Schrödinger-like system

$$-\nabla^2 \Psi(\mathbf{x}) + \mathcal{U}(r) \Psi(\mathbf{x}) = \mathcal{E}^2 \Psi(\mathbf{x}), \quad (7.48)$$

$$\Psi(\mathbf{x}) = \begin{pmatrix} \varphi_+ \\ \varphi_- \\ \chi_0 \end{pmatrix}, \quad \mathcal{U}(r) = \begin{pmatrix} U_\phi(r) & 0 & M(r) \\ 0 & U_\phi(r) & M(r) \\ M(r) & M(r) & U_\sigma(r) \end{pmatrix}, \quad \mathcal{E}^2 = \begin{pmatrix} E_+^2 & 0 & 0 \\ 0 & E_-^2 & 0 \\ 0 & 0 & E_0^2 \end{pmatrix}. \quad (7.49)$$

Similar to the two-channel system of the perturbations of the Q-ball, this three-channel system of the FLS perturbations allows for a conserved current,

$$\mathbf{J} \equiv \sum_{s=\pm,0} \mathbf{J}_s, \quad \nabla \cdot \mathbf{J} = 0, \quad (7.50)$$

$$\mathbf{J}_\pm = \frac{1}{2i}(\varphi_\pm^* \nabla \varphi_\pm - \varphi_\pm \nabla \varphi_\pm^*), \quad \mathbf{J}_0 = \frac{1}{2i}(\chi_0^* \nabla \chi_0 - \chi_0 \nabla \chi_0^*). \quad (7.51)$$

Far outside the FLS soliton, Figure 7.7, $U_\phi \rightarrow m_\phi^2 = m^2$, $U_\sigma \rightarrow m_\sigma^2 = \kappa^2 m^2$, $M \rightarrow 0$, so the channels decouple from each other. Comparing the spectrum parameter λ to the three thresholds of the system: $m - \omega, m + \omega, \kappa m$, identifies whether some of the channels becomes bound. This introduces much more variation than it was for the Q-ball. We conveniently choose to have $\kappa = 2$ where the neutral scalar is twice as massive as the charged scalar, and thus the threshold of the neutral scalar is always higher than any of the charged scalars. Restricting to $\lambda > 0$ for definiteness, the windows are:

- $0 < \lambda < m - \omega$: all three channels are bound. Zero and non-zero bound modes live within this window (the direct generalization of the Q-ball bound-mode window).
- $m - \omega < \lambda < m + \omega$: only the $+$ mode is a free mode in this window.
- $m + \omega < \lambda < \kappa m$: both charged modes, $\pm$, are free in this window, while the neutral mode is bound.
- $\lambda > \kappa m$: all three modes are free.

The numerical results below are restricted to the two- and three-channel scattering windows, in direct analogy with restricting the Q-ball analysis to $\lambda > m + \omega$.

7.6.2 Reflection and conversion probabilities

We shall consider the s-wave content of (7.48),

$$-\psi'' - \frac{2}{r}\psi + \mathcal{U}(r)\psi = \mathcal{E}^2\psi, \quad (7.52)$$

repeating all steps for solving this system. However, unlike the Q-ball perturbations where an incoming $+$ mode was physically related to an incoming $-$ mode once we carry out $\lambda \leftrightarrow -\lambda$, in the FLS perturbations, we need to consider an incoming $+$ mode and an incoming 0 mode separately.

First, we solve the three-channel system for an incoming $+$ mode: $A_+^{\text{in}} = 1$, $A_-^{\text{in}} = A_0^{\text{in}} = 0$, scanned over both signs of λ (with the negative branch related to an incoming $-$ mode exactly as for the Q-ball). Figure 7.8 shows the outgoing

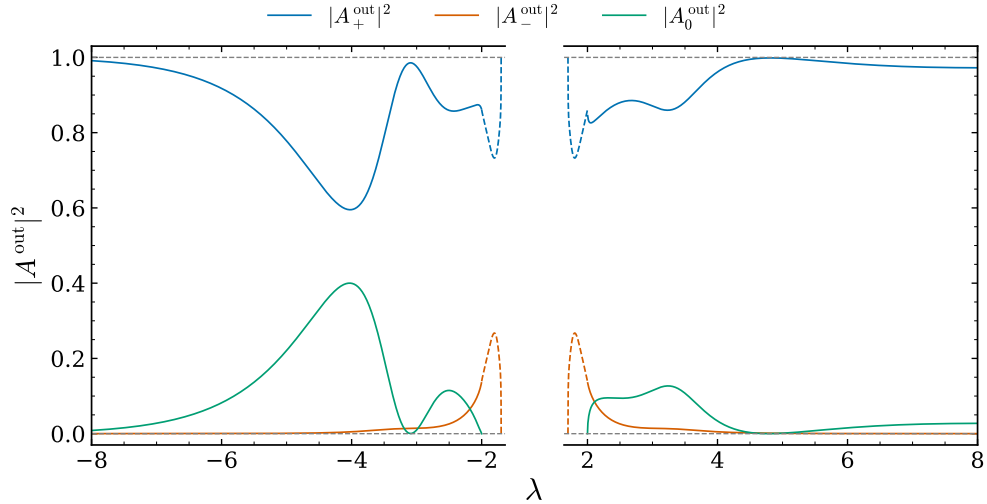


Figure 7.8: Reflection and conversion probabilities $|S_{s+}|^2 = |A_s^{\text{out}}|^2$ for the FLS soliton, $\kappa = 2.0$, for an incoming $+$ mode. The cusp marks the opening of the χ_0 channel at $|\lambda| = \kappa m$, separating the two-channel and three-channel scattering windows of Section 7.6.1.

amplitudes squared for the three channels for each λ . The reflection probability $|S_{++}|^2 = |A_+^{\text{out}}|^2$ dominates over conversion probabilities (as it was for the Q-ball perturbations). The conversion to $-$ modes is symmetric across $\lambda \leftrightarrow -\lambda$ due to the indirect coupling between $+$ and $-$ modes. At $|\lambda| = \kappa m$ the neutral mode can become a free propagating mode, so we get a cusp behavior of the probabilities.² Below the threshold, we see parity across $\lambda \leftrightarrow -\lambda$ in the behavior of both $|S_{++}|^2 = |A_+^{\text{out}}|^2$ and $|S_{-+}|^2 = |A_-^{\text{out}}|^2$ which is expected from two-channel systems.

²This is a clear signature of the threshold; even though it was not shown in the Q-ball, the half-scattering window had a cusp where the reflection probability becomes exactly 1. This is Wigner's threshold law [170].

Next, we consider an incoming 0 mode, $A_0^{\text{in}} = 1$, $A_+^{\text{in}} = A_-^{\text{in}} = 0$, for which both signs of λ are independently informative. Figure 7.9 shows the reflection probability

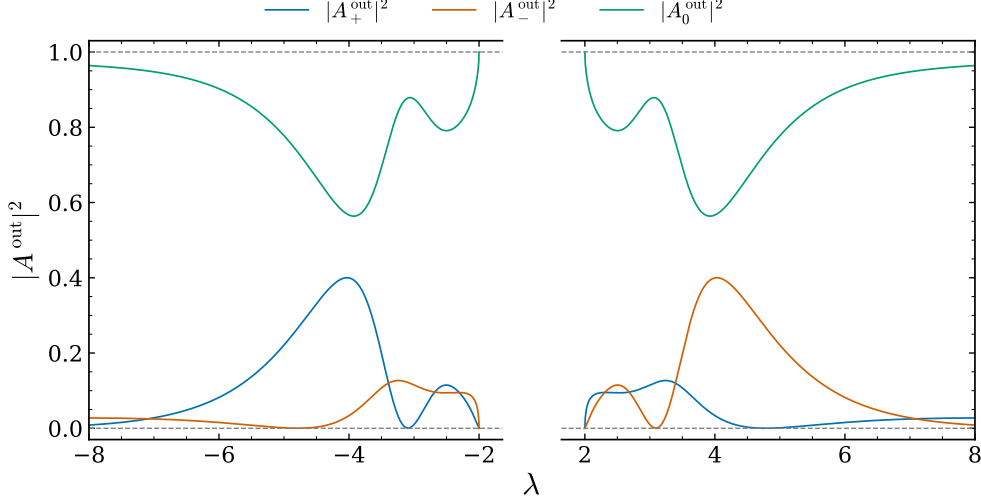


Figure 7.9: Reflection and conversion probabilities $|S_{s0}|^2 = |A_s^{\text{out}}|^2$ for the FLS soliton, $\kappa = 2.0$ for an incoming neutral 0 mode.

dominate the process once again. The conversion probabilities to $\pm$ modes switch identity when we have $\lambda \leftrightarrow -\lambda$, a result that stems from $\varphi_+ \leftrightarrow \varphi_-$,

$$|S_{+0}(\lambda)|^2 = |S_{-0}(-\lambda)|^2. \quad (7.53)$$

Furthermore, we find that the conversion probability of a 0 mode to a + mode is exactly the same as the conversion probability of a + mode to a 0 mode due to the reciprocity of the system,

$$|S_{+0}(\lambda)|^2 = |S_{0+}(\lambda)|^2. \quad (7.54)$$

7.6.3 Charge and energy transfer

The conversion probabilities that we see in the perturbation system of the FLS soliton lead to charge and energy transfer in direct analogy to the analysis performed for the Q-ball. When the incident quantum is a charged particle, we can derive the amplification of charge and energy from Eqs. (7.27) and (7.28),

$$Z_Q = -2|A_-^{\text{out}}|^2 - |A_0^{\text{out}}|^2, \quad Z_E = -\frac{2\omega}{E_+}|A_-^{\text{out}}|^2 - \frac{\omega}{E_+}|A_0^{\text{out}}|^2. \quad (7.55)$$

This also follows directly from the kinematic understanding of the process,

$$(Q) + \phi \rightarrow (Q) + \phi, \quad (Q) + \phi \rightarrow (Q+2) + \bar{\phi}, \quad (Q) + \phi \rightarrow (Q+1) + \sigma. \quad (7.56)$$

The transfer of charge and energy is related to the conversion probabilities; the incoming flux loses ω (the chemical potential) of energy for each one unit of charge it loses to the soliton, and we divide by the energy of the incoming flux to calculate the amplification.

The same argument holds when the incident is antiparticle, just instead of depositing energy and charge into the soliton, the flux extracts out of it

$$(Q) + \bar{\phi} \rightarrow (Q - 2) + \phi, \quad (Q) + \bar{\phi} \rightarrow (Q - 1) + \sigma. \quad (7.57)$$

When the incident is the neutral σ particle, we can define the transfer of energy as usual by dividing the outgoing energy flux by the incoming energy flux,

$$Z_E = \frac{\omega}{E_0} |A_+^{\text{out}}|^2 - \frac{\omega}{E_0} |A_-^{\text{out}}|^2, \quad (7.58)$$

in association with the conversion processes,

$$(Q) + \sigma \rightarrow (Q - 1) + \phi, \quad (Q) + \sigma \rightarrow (Q + 1) + \bar{\phi}. \quad (7.59)$$

However, it is not possible to define the charge transfer via dividing by the incoming charge flux. Therefore, we use kinematics to calculate the charge transfer,

$$Z_Q = |A_+^{\text{out}}|^2 - |A_-^{\text{out}}|^2. \quad (7.60)$$

7.7 Summary

This chapter asked what happens when radiation scatters off a Q -ball, and answered it by building an exact scattering theory around one structural fact: because the background rotates, a perturbation does not stay at a single frequency but splits into two channels locked together by $E_+ - E_- = 2\omega$ (7.5). Section 7.1 derived the resulting two-channel equations and classified their modes; Section 7.2 built the associated S-matrix and proved its unitarity, reciprocity, and the exchange relation $S(-\lambda) = PS(\lambda)P$ (7.22) relating the two channels.

The central result followed in Section 7.4: unitarity alone fixes the charge and energy amplification factors, $Z_Q = -2|A_-^{\text{out}}|^2$ and $Z_E = -(2\omega/E_+)|A_-^{\text{out}}|^2$ (7.31), and these agree exactly with what the Q-ball's thermodynamics predicts through its chemical potential $\omega = dE/dQ$. Scattering microphysics and soliton thermodynamics turn out to be the same statement seen twice, which is the fulfillment of the objective set out in Section 5.3.

Section 7.5 confirmed this linear picture against full non-linear lattice evolution, finding agreement for $\delta_r \lesssim 10^{-2}$ and translating this into a bath temperature bound $T \lesssim 10^{-3}m$ below which the scattering of particles off the Q -ball is described linearly. Section 7.6 then showed the formalism extends directly to the two-field FLS soliton: a third, neutral channel opens at κm , enabling the new process $(Q)+\phi \rightarrow (Q+1)+\sigma$, with the same unitarity and reciprocity structure carrying over unchanged.

Chapter 8

Conclusions

8.1 Summary of results

This part of the thesis studied how Q-balls and Friedberg-Lee-Sirlin solitons exchange charge and energy with scattered radiation, and asked over what range of amplitudes a linear description of that exchange can be trusted. Chapter 6 built the stationary backgrounds needed for this analysis: the one-field Q-ball, obtained by the shooting method, and the two-field FLS soliton, obtained by multiple shooting, along with the thermodynamic identity $dE/dQ = \omega$ and the stability windows that determine which backgrounds are physically sensible.

Chapter 7 developed the scattering theory itself. Because the Q-ball background rotates, a perturbation splits into two coupled channels linked by the closure relation $E_+ - E_- = 2\omega$. We built the S-matrix for this two-channel system, proved its unitarity and reciprocity, and derived the charge and energy amplification factors Z_Q and Z_E from unitarity alone. These factors agree exactly with the prediction obtained from the background's own thermodynamics through $\omega = dE/dQ$, so the scattering result and the equilibrium description of the Q-ball turn out to be the same statement. We then tested the linear theory against a non-linear lattice evolution of the full field equation, at $(\omega, \lambda) = (0.70, 2.0)$ scanned over incident amplitude, and found agreement up to $\delta_r \lesssim 10^{-2}$. Translating this bound to a thermal bath gives the temperature range over which the linear treatment of Q-ball scattering applies.

Finally, we extended the formalism to the FLS soliton. The two-field structure introduces a third, neutral channel with its threshold at κm , and charge-changing scattering proceeds only indirectly through this channel. We built the resulting three-channel S-matrix and computed its reflection and conversion probabilities for both a charged and a neutral incident mode.

8.2 Limitations

No lattice check was performed for the FLS soliton. Whether the same threshold $\delta_r \lesssim 10^{-2}$ holds there, or is shifted by the extra neutral channel, is not known.

The partial-wave extension in Appendix B was carried out for the Q-ball only; the same steps apply to the FLS soliton but were not repeated here. More broadly, the treatment throughout this part is classical: the incident radiation is treated as a wave rather than a quantum state, and spontaneous emission, which has no classical counterpart, is not included in any of the results above.

Appendices

Appendix A

The Shooting and Multiple Shooting Methods

In Section 6.2.1 we presented the numerical solution of the one-field Q-ball profile equation (6.11) obtained by the ordinary shooting method. In Section 6.3.2 it was noted that the same strategy is numerically unreliable for the two-field FLS soliton, and the presented numerical solution was obtained by the multiple shooting method, which will be demonstrated in this appendix. We work through both methods here: Section A.1 makes the Q-ball shooting procedure explicit and illustrates the overshoot/undershoot behavior that drives the bisection method of refining the initial condition; Section A.2 then builds the multiple shooting solver for the FLS soliton, including the sensitivity matrix that measures the solution dependence on the initial conditions and the Newton solver system for refining the solution.

A.1 Single shooting: the one-field Q-ball

The second-order profile equation (6.11) can be written as a first-order system

$$\phi'_Q \equiv \pi_Q, \quad \pi'_Q = -\frac{2}{r} \pi_Q - \omega^2 \phi_Q + \frac{1}{2} \frac{\partial V}{\partial \phi_Q}. \quad (\text{A.1})$$

The regularity at the origin leads to the initial condition $\pi_Q(0) = 0$, while the other initial condition remains an indeterminate parameter, $\phi_Q(0) \equiv \phi_0$. The mechanical analogy of Section 6.1 already tells us where to look for it: ϕ_0 must lie in the range where the Newtonian potential $V_N(\phi_0; \omega)$ of Eq. (6.13) is positive, bounded in the interval from the onset of $V_N(\phi_{\min}) \geq 0$ to the local maximum $V'_N(\phi_{\max}) = 0$; both bounds are quadratic in ϕ_0^2 for the sextic potential (6.22) and are solved for in closed form to bracket the bisection search.

For a trial ϕ_0 inside this bracket, integrating (A.1) outward produces one of three behaviors, exactly as anticipated in Section 6.2.1:

- ϕ_0 too large: the fictitious body carries more mechanical energy than the friction can dissipate before reaching the intended boundary condition $\phi_Q = 0$, so it overshoots and runs off to negative values.
- ϕ_0 too small: the body loses its energy too early, turns around before reaching $\phi_Q = 0$ (π_Q changes sign from negative back to positive), and diverges toward $\phi_Q > 0$ on the way back out.
- ϕ_0 tuned exactly: the body approaches $\phi_Q = 0$ asymptotically as $r \rightarrow \infty$, giving the genuine soliton profile.

Bisecting on this overshoot/undershoot classification converges geometrically to the correct initial condition ϕ_0 . Figure A.1 shows this directly for the Q-ball with the frequency $\omega = 0.7$, $\epsilon = 0.4$ of Figure 6.4: the converged profile sits between two trial profiles differing from it by less than one percent in ϕ_0 , one already visibly overshooting past zero, the other visibly turning back upward.

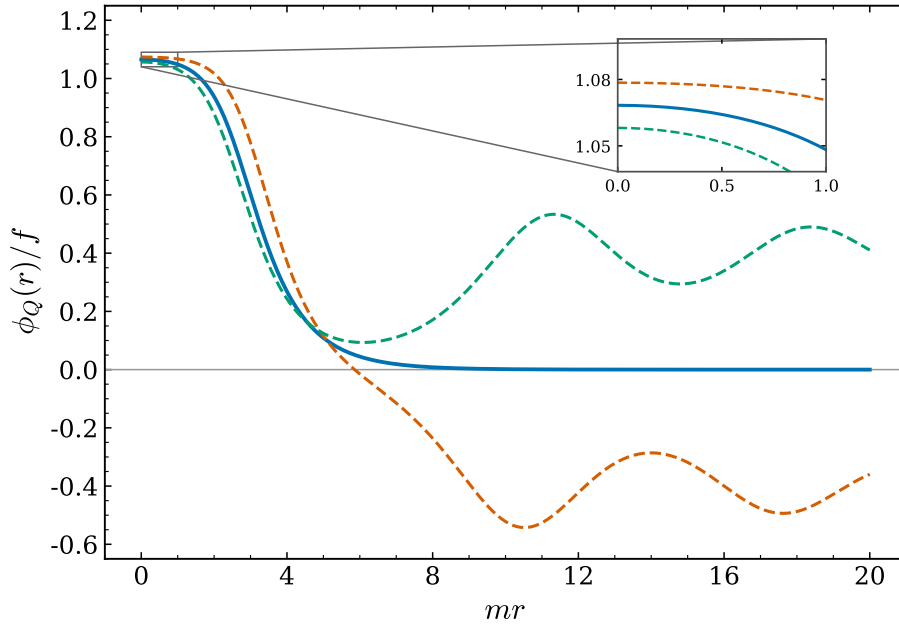


Figure A.1: Illustration of the shooting method for the Q-ball of frequency $\omega = 0.7$, $\epsilon = 0.4$. The converged profile (solid) is bracketed by an overshooting trial with $\phi_Q(0)$ increased by 0.8% (dashed orange, crosses $\phi_Q = 0$ and continues to oscillate around a different, unphysical extremum) and an undershooting trial with $\phi_Q(0)$ decreased by 0.8% (dashed green, turns around before reaching zero and likewise fails to settle). The inset zooms in on the three trial values of $\phi_Q(0)$ themselves, indistinguishable at the scale of the main plot.

A.2 Multiple shooting: the FLS soliton

The FLS soliton requires two indeterminate initial values, $\phi_Q(0)$ and $\sigma_Q(0)$, but there is no analogue of the one-dimensional Newtonian-potential bracket to localize them: the search is genuinely two-dimensional. What is worse is that the sensitivity of the endpoint behavior to the initial data grows with the length of the interval integrated, so a single shot across the full domain amplifies any error in the trial $(\phi_Q(0), \sigma_Q(0))$ so severely that a Newton search on the full interval fails to converge except from initial guesses too accurate to be available in practice. Multiple shooting sidesteps this by never integrating across the full domain in one piece: the interval is cut into N short segments, each mild enough that its own sensitivity to its own starting point stays under control, and the pieces are then stitched together by a Newton iteration acting on all of them simultaneously.

A.2.1 Discretization and the augmented state

The second-order FLS profile equation system (6.27) is cast as a first-order system in $\mathbf{y} \equiv (\phi_Q, \sigma_Q, \pi_\phi, \pi_\sigma)$, with $\pi_\phi \equiv \phi'_Q$, $\pi_\sigma \equiv \sigma'_Q$,

$$\mathbf{y}' = \mathcal{F}(r, \mathbf{y}), \quad \mathcal{F}(r, \mathbf{y}) = \begin{pmatrix} \pi_\phi \\ \pi_\sigma \\ -\frac{2}{r}\pi_\phi + F_1(\phi_Q, \sigma_Q) \\ -\frac{2}{r}\pi_\sigma + F_2(\phi_Q, \sigma_Q) \end{pmatrix}, \quad (\text{A.2})$$

where $F_1 \equiv -\frac{1}{2}\partial V_N/\partial\phi_Q = (\sigma_Q^2 - \omega^2)\phi_Q$ and $F_2 \equiv -\partial V_N/\partial\sigma_Q = 2\sigma_Q[\phi_Q^2 + \frac{1}{4}\kappa^2(\sigma_Q^2 - 1)]$ follow directly from the FLS potential of Eq. (6.24) (in units $m_\phi = f = 1$, so $h = \kappa$).

Choose radial nodes $0 = r_0 < r_1 < r_2 < \dots < r_N = R$, and let $\mathbf{y}_0^{(i)}$ denote the initial conditions for the i^{th} node. These conditions, collected into a single vector $\mathcal{Y} = (\mathbf{y}_0^{(0)}, \mathbf{y}_0^{(1)}, \dots, \mathbf{y}_0^{(N-1)}) \in \mathbb{R}^{4N}$, are the unknowns of the problem; the solution at the final node $r_N = R$ is obtained from integrating the last segment, and is required to satisfy the boundary conditions. On each segment $[r_i, r_{i+1}]$, propagating $\mathbf{y}_0^{(i)}$ forward under (A.2) defines a flow map $\mathbf{y}_0^{(i)} \mapsto \mathbf{y}^{(i)}(r_{i+1}; \mathbf{y}_0^{(i)})$, and multiple shooting demands that consecutive segments satisfy the matching conditions,

$$\mathbf{y}^{(i)}(r_{i+1}; \mathbf{y}_0^{(i)}) = \mathbf{y}_0^{(i+1)}, \quad i = 0, \dots, N-2. \quad (\text{A.3})$$

In addition, the system has boundary conditions at the two ends of the full domain:

the regularity conditions at the origin,

$$\pi_\phi(r_0) = \pi_\sigma(r_0) = 0, \quad (\text{two conditions on } \mathbf{y}_0^{(0)} \text{ alone}), \quad (\text{A.4})$$

and the vacuum conditions at infinity,

$$\phi_Q(R) = 0, \sigma_Q(R) = 1, \quad (\text{two conditions on } \mathbf{y}^{(N-1)}(r_N; \mathbf{y}_0^{(N-1)}) . \quad (\text{A.5})$$

Eq. (A.3) supplies $4(N-1)$ conditions, and the two boundary conditions (A.4), (A.5) supply another four, so we have exactly as many equations as the $4N$ unknowns in $\mathcal{Y}$; the multiple shooting requires solving this Newton system.

The sensitivity matrix. The Jacobian of the equation system (A.3)–(A.5) with respect to $\mathcal{Y}$ can be written in the sparse matrix format,

$$\mathcal{J} = \left(\begin{array}{cc|c|c|c|c} \mathbf{0}_2 & \mathbf{1}_2 & & & & \\ -S_0(r_1) & \mathbf{1}_4 & & & & \\ & -S_1(r_2) & \mathbf{1}_4 & & & \\ & & \ddots & \ddots & & \\ & & & -S_{N-2}(r_{N-1}) & \mathbf{1}_4 & \\ & & & & -S_{N-1}(r_N)|_{\text{row}=1,2} & \end{array} \right), \quad (\text{A.6})$$

where $S_i(r)$ is the sensitivity matrix of the solution $\mathbf{y}^{(i)}(r; \mathbf{y}_0^{(i)})$ on $\mathbf{y}_0^{(i)}$. It satisfies the variational equation obtained from taking the derivative of (A.2) on the i^{th} segment with respect to $\mathbf{y}_0^{(i)}$,

$$S'_i = \frac{\partial \mathcal{F}}{\partial \mathbf{y}_0^{(i)}} \cdot S_i, \quad S_i(r_i) = \mathbf{1}_4, \quad \frac{\partial \mathcal{F}}{\partial \mathbf{y}_0^{(i)}} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \partial_\phi F_1^{(0)} & \partial_\sigma F_1^{(0)} & -2/r & 0 \\ \partial_\phi F_2^{(0)} & \partial_\sigma F_2^{(0)} & 0 & -2/r \end{pmatrix}. \quad (\text{A.7})$$

The first and last two rows of the Jacobian (A.6) follow from (A.4), (A.5), and are so arranged to make apparent the sparseness of the matrix. The variational equation can be integrated together with the profile equation in terms of an augmented state to accelerate the method.

A.2.2 The Newton system and its block structure

The equation system is solved by enforcing the $4N$ matching and boundary condition mismatch, collected into a single vector $\mathcal{R}(\mathcal{Y})$, to be zero. Thus, the mismatch residual vector needs to be driven to zero,

$$\mathcal{R}(\mathcal{Y}) = \begin{pmatrix} \pi_\phi^{(0)} \\ \pi_\sigma^{(0)} \\ \mathbf{y}_0^{(1)} - \mathbf{y}^{(0)}(r_1; \mathbf{y}_0^{(0)}) \\ \mathbf{y}_0^{(2)} - \mathbf{y}^{(1)}(r_2; \mathbf{y}_0^{(1)}) \\ \vdots \\ \mathbf{y}_0^{(N-1)} - \mathbf{y}^{(0)}(r_{N-1}; \mathbf{y}_0^{(N-2)}) \\ 0 - \phi_Q^{(N-1)}(r_N; \mathbf{y}_0^{(N-1)}) \\ 1 - \sigma_Q^{(N-1)}(r_N; \mathbf{y}_0^{(N-1)}) \end{pmatrix} \rightarrow 0, \quad (\text{A.8})$$

by evaluating its value after using a choice of unknowns $\mathcal{Y}$, and then updating their value $\mathcal{Y} \rightarrow \mathcal{Y} + \alpha \delta \mathcal{Y}$ with $J \delta \mathcal{Y} = -\mathcal{R}$ where $J \equiv \partial \mathcal{R} / \partial \mathcal{Y}$, repeated until $\|\mathcal{R}\|$ falls below a chosen tolerance; the damping factor $\alpha < 1$ keeps a single Newton step from overcorrecting on early iterations, where the trial profile can still be far from the true one.

A.2.3 Convergence of the Newton iteration

It is worth seeing directly what the matching conditions (A.3) are doing during the iteration, rather than only trusting the residual norm to fall. Figure A.2 shows the piecewise trajectory obtained from *one* Newton step past the initial guess of Section A.2, for the $\omega = 0.7$, $\kappa = 2.0$ background with $N = 40$ segments: every one of the 39 interior nodes is still visibly a separate, independently-integrated segment, stitched into a ragged sawtooth rather than a smooth curve, with residual $\|\mathcal{R}\| \approx 8.1$. Four damped Newton steps later, Figure A.3 shows the same construction at $\|\mathcal{R}\| \approx 3.6$: the ϕ_Q segments already agree with one another closely enough to look like a single smooth curve at this resolution, while σ_Q , which is still relaxing into its own matching condition more slowly, retains a visible residual sawtooth. This asymmetry is not a numerical accident: σ_Q is the slower of the two fields to settle because the outer boundary condition constrains it only weakly relative to ϕ_Q near the core, so its segments carry more of the mismatch for longer into the iteration. By the time $\|\mathcal{R}\|$ has fallen below the tolerance used throughout this thesis, both fields are smooth to the eye, exactly as in Figure 6.7.

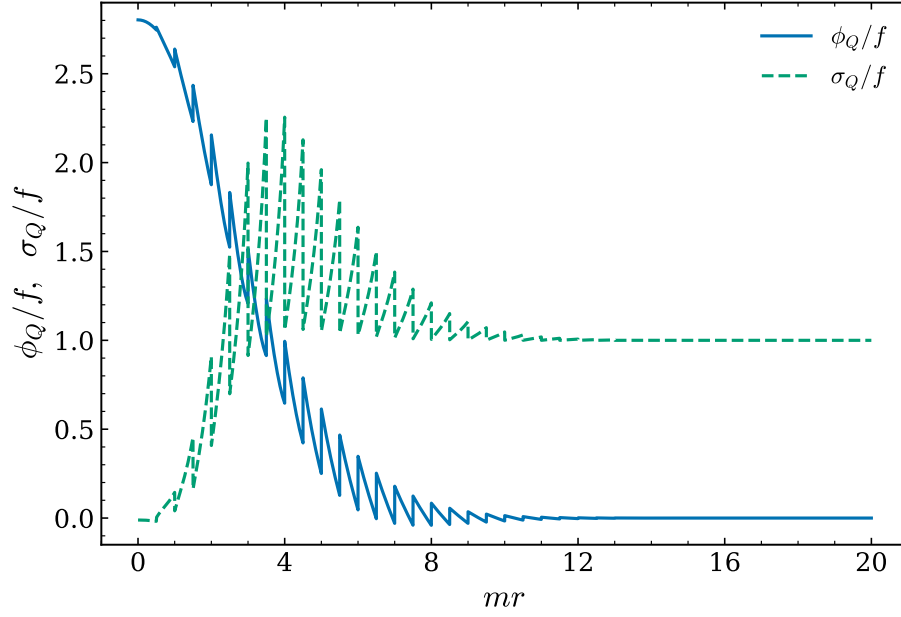


Figure A.2: The multiple-shooting trajectory after one Newton step, $\omega = 0.7$, $\kappa = 2.0$, $N = 40$, $\|\mathcal{R}\| \approx 8.1$. Every node still marks a visible discontinuity between adjacent segments.

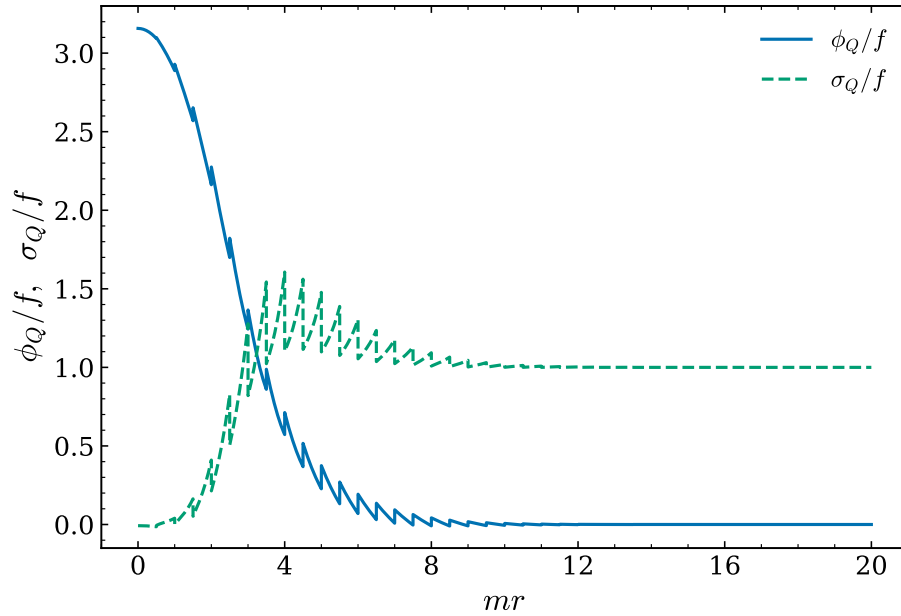


Figure A.3: The same construction after five Newton steps (the damping factor was kept small intentionally to generate this plot), $\|\mathcal{R}\| \approx 3.6$. The ϕ_Q segments have approximately converged to a smooth profile; σ_Q still shows a residual sawtooth, reflecting its slower relaxation into the matching and boundary conditions.

Appendix B

Partial-wave scattering off the Q-ball

Chapter 7 restricted the two-channel scattering problem to the s-wave $\ell = 0$, since it shows what we need to know about the reflection and conversion *probabilities* $|S_{ss'}(\lambda)|^2$ to understand the charge- and energy-exchange mechanism in the super-radiance. The incident particles on a Q-ball in realistic processes are not s-waves but rather plane waves, so we need to pick up the higher partial waves $\ell > 0$ that we omitted.

This appendix supplies that step. Section B.1 extends the construction of Section 7.2 to general ℓ ; the extension is immediate, since every argument there was already stated for the radial system (7.12) rather than its $\ell = 0$ reduction. Section B.2 then relates the resulting $S_\ell(\lambda)$ to the cross section a physical incident beam would see, and Section B.3 presents the resulting cross sections numerically.

B.1 The partial-wave S-matrix

Because the background potentials $U(r)$ and $M(r)$ depend on the radial coordinate alone, the matrix $\mathcal{U}(r)$ of Eq. (7.6) commutes with orbital angular momentum, and the general perturbation decomposes as $\Psi(\mathbf{x}) = \sum_{\ell,m} \psi_\ell(r) Y_{\ell m}(\hat{\mathbf{x}})$, with each (ℓ, m) sector obeying the radial system already written down in Eq. (7.12),

$$-\psi_\ell'' - \frac{2}{r}\psi_\ell' + \frac{\ell(\ell+1)}{r^2}\psi_\ell + \mathcal{U}(r)\psi_\ell = \mathcal{E}^2\psi_\ell, \quad \psi_\ell = \begin{pmatrix} \varphi_{+, \ell} \\ \varphi_{-, \ell} \end{pmatrix}. \quad (\text{B.1})$$

Since $\mathcal{U}(r)$ and $\mathcal{E}^2$ carry no dependence on the azimuthal number m , neither does ψ_ℓ or the scattering matrix built from it; m only re-enters when a specific physical

beam is projected onto the partial-wave basis, in Section B.2. Chapter 7 used the $\ell = 0$ case of this same equation,

$$-\psi_0'' - \frac{2}{r}\psi_0' + \mathcal{U}(r)\psi_0 = \mathcal{E}^2\psi_0, \quad (\text{B.2})$$

recovered from (B.1) by dropping the centrifugal term.

Regularity. Near $r = 0$ the centrifugal term dominates the two finite pieces $\mathcal{U}(r)\psi_\ell$ and $\mathcal{E}^2\psi_\ell$, so the indicial equation for each component of ψ_ℓ is the free one, with roots r^ℓ and $r^{-\ell-1}$. The regular solution therefore has the boundary behavior

$$\varphi_{s,\ell}(r) \xrightarrow{r \rightarrow 0} c_s r^\ell, \quad s = \pm, \quad (\text{B.3})$$

independently for each channel s at leading order: the off-diagonal coupling $M(r)$ only sources a subleading correction at this order, since it multiplies a quantity that is itself already $O(r^\ell)$. As in the $\ell = 0$ case, this leaves a two-dimensional space of regular solutions, spanned for instance by the two choices $(c_+, c_-) = (1, 0)$ and $(0, 1)$.

The S-matrix at fixed ℓ . In the scattering window $|\lambda| > m + \omega$, both channels admit the same flux-normalized asymptotic form as Eq. (7.14), now labelled by ℓ ,

$$\varphi_{\pm,\ell}(r) \xrightarrow{r \rightarrow \infty} \frac{1}{\sqrt{4\pi k_\pm} r} \left(A_{\pm,\ell}^{\text{in}} e^{-ik_\pm r} + A_{\pm,\ell}^{\text{out}} e^{+ik_\pm r} \right), \quad (\text{B.4})$$

and the construction of Section 7.2.2 carries over verbatim: two regular solutions of (B.1), matched onto (B.4), define $A_\ell^{\text{in}}, A_\ell^{\text{out}}$ and hence

$$S_\ell(\lambda) \equiv A_\ell^{\text{out}}(\lambda) \left[A_\ell^{\text{in}}(\lambda) \right]^{-1}. \quad (\text{B.5})$$

Unitarity and reciprocity follow exactly as in Sections 7.2.3–7.2.4, since neither the flux argument behind (7.20) nor the Wronskian argument behind (7.21) used the centrifugal term at any step; it appears identically in both channels and drops out of $\mathbf{J}$ and W by the same cancellation as before:

$$S_\ell^\dagger(\lambda) S_\ell(\lambda) = 1, \quad S_\ell(\lambda)^T = S_\ell(\lambda), \quad S_\ell(-\lambda) = P S_\ell(\lambda) P, \quad (\text{B.6})$$

with P the same channel-exchange matrix as in Eq. (7.22). Equation (B.5) is solved numerically for each ℓ exactly as $S(\lambda)$ was solved at $\ell = 0$; a representative ℓ -dependence of S_ℓ is discussed in Section B.3.

B.2 From S_ℓ to a cross section

The quantity a thermal or astrophysical incident flux couples to is not S_ℓ at fixed angular momentum but the total cross section for a directed beam. We derive it here for an incident $+$ (particle) beam; the incident $-$ (antiparticle) beam follows by the same steps with $+$ $\leftrightarrow$ $-$ exchanged, and independently by the exchange relation of Eq. (B.6).

Setup. Take the incident beam along $\hat{\mathbf{z}}$ with unit coefficient, e^{ik_+z} , and allow both an elastically reflected and a converted outgoing spherical wave,

$$\Phi_+(r, \theta) = e^{ik_+z} + f_{++}(\theta) \frac{e^{ik_+r}}{r}, \quad \Phi_-(r, \theta) = f_{-+}(\theta) \frac{e^{ik_-r}}{r}, \quad (\text{B.7})$$

with no incoming piece in the $-$ channel, since the beam contains no antiparticles. Expanding the plane wave and each scattering amplitude in Legendre polynomials,

$$e^{ik_+z} = \sum_{\ell=0}^{\infty} (2\ell+1) i^\ell j_\ell(k_+r) P_\ell(\cos \theta), \quad f_{s+}(\theta) = \sum_{\ell=0}^{\infty} (2\ell+1) f_{s+, \ell} P_\ell(\cos \theta), \quad (\text{B.8})$$

and using $P_\ell(\cos \theta) = \sqrt{4\pi/(2\ell+1)} Y_{\ell 0}(\theta)$ (the beam is azimuthally symmetric, so only $m=0$ appears), Eq. (B.7) is exactly the $m=0$ partial-wave decomposition $\Phi_s = \sum_\ell \varphi_{s, \ell}(r) Y_{\ell 0}(\theta)$ used in Section B.1, with

$$\varphi_{+, \ell}(r) = \sqrt{4\pi(2\ell+1)} \left[i^\ell j_\ell(k_+r) + f_{++, \ell} \frac{e^{ik_+r}}{r} \right], \quad \varphi_{-, \ell}(r) = \sqrt{4\pi(2\ell+1)} f_{-, \ell} \frac{e^{ik_-r}}{r}. \quad (\text{B.9})$$

Matching to S_ℓ . Using the large-argument form $j_\ell(x) \sim (2ix)^{-1} [(-i)^\ell e^{ix} - i^\ell e^{-ix}]$ in (B.9) and comparing to the definition (B.4) of $A_{\pm, \ell}^{\text{in, out}}$ fixes both amplitudes for the $+$ channel,

$$\frac{A_{+, \ell}^{\text{in}}}{4\pi} = \sqrt{k_+(2\ell+1)} \frac{(-1)^{\ell+1}}{2ik_+}, \quad \frac{A_{+, \ell}^{\text{out}}}{4\pi} = \sqrt{k_+(2\ell+1)} \left[\frac{1}{2ik_+} + f_{++, \ell} \right], \quad (\text{B.10})$$

while the $-$ channel, having no incoming piece by construction, gives $A_{-, \ell}^{\text{in}} = 0$ and $A_{-, \ell}^{\text{out}}/4\pi = \sqrt{k_-(2\ell+1)} f_{-, \ell}$. Forming the two ratios $S_{s+, \ell} = A_{s+, \ell}^{\text{out}}/A_{+, \ell}^{\text{in}}$, and solving for the scattering amplitudes gives

$$\boxed{f_{++, \ell} = \frac{i}{2k_+} \left[1 + (-1)^\ell S_{++, \ell}(\lambda) \right], \quad f_{-, \ell} = \frac{(-1)^{\ell+1}}{2i\sqrt{k_+k_-}} S_{-, \ell}(\lambda).} \quad (\text{B.11})$$

The factor $(-1)^\ell$ is not an extra convention layered on top of S_ℓ ; it is the direct consequence of matching the flux-normalized S_ℓ , which for a genuinely free channel tends to $S_\ell \rightarrow (-1)^{\ell+1}$ rather than 1, to the plane-wave phase convention of (B.8). Correspondingly, for a free (unscattered) channel $f_{++,\ell} \rightarrow 0$ exactly, as it must.

Total cross sections. A reaction $i \rightarrow f$ with a unit-amplitude incident beam in channel i has differential cross section $d\sigma_{i \rightarrow f}/d\Omega = (k_f/k_i) |f_{fi}(\theta)|^2$, the k_f/k_i weight converting the outgoing spherical-wave flux into a rate per unit incident flux. Using $\int d\Omega P_\ell P_{\ell'} = \frac{4\pi}{2\ell+1} \delta_{\ell\ell'}$ to integrate (B.8) over solid angle and inserting (B.11), the sums over ℓ close into

$$\sigma_{++}(E) = \frac{\pi}{E^2 - m^2} \sum_{\ell=0}^{\infty} (2\ell+1) \left| 1 + (-1)^\ell S_{++,\ell}(\lambda) \right|^2, \quad (\text{B.12})$$

$$\sigma_{-+}(E) = \frac{\pi}{E^2 - m^2} \sum_{\ell=0}^{\infty} (2\ell+1) \left| S_{-+,\ell}(\lambda) \right|^2, \quad (\text{B.13})$$

where $E = E_+ = \lambda + \omega$ is the physical energy of the incident particle, so that $\lambda = E - \omega$ labels the S_ℓ entering both sums, and $E^2 - m^2 = k_+^2$. The corresponding formulas for an incident $-$ (antiparticle) beam of physical energy $E = E_- = \lambda - \omega$ follow by relabelling $+ \leftrightarrow -$ throughout the derivation above, equivalently by evaluating the *same* $S_\ell(\lambda)$ at $\lambda = E + \omega$ and reading off its second column, exactly as the reflection probability of an incoming $-$ mode was read off the same $S(\lambda)$ used for the incoming $+$ mode in Section 7.3:

$$\sigma_{--}(E) = \frac{\pi}{E^2 - m^2} \sum_{\ell} (2\ell+1) \left| 1 + (-1)^\ell S_{--, \ell}(\lambda) \right|^2, \quad (\text{B.14})$$

$$\sigma_{+-}(E) = \frac{\pi}{E^2 - m^2} \sum_{\ell} (2\ell+1) \left| S_{+-, \ell}(\lambda) \right|^2. \quad (\text{B.15})$$

Because S_ℓ is symmetric, $S_{-+,\ell} = S_{+-,\ell}$, and comparing (B.13) and (B.15) *at the same* λ , i.e. at $E_+ = \lambda + \omega$ for the first and $E_- = \lambda - \omega$ for the second, gives a cross-section reciprocity relation with no further input,

$$\sigma_{+-}(E_-) = \frac{E_+^2 - m^2}{E_-^2 - m^2} \sigma_{-+}(E_+), \quad E_\pm = \lambda \pm \omega. \quad (\text{B.16})$$

No such relation constrains σ_{++} and σ_{--} individually, since both carry the forward-scattering interference term $1 + (-1)^\ell S_{\pm\pm,\ell}$, sensitive to the phase of $S_{\pm\pm,\ell}$ and not merely its modulus.

B.3 Numerical results

Equations (B.12)–(B.15) are evaluated by solving (B.5) at each ℓ up to a truncation $\ell_{\max}$, chosen adaptively from the impact-parameter estimate $\ell_{\max} \sim R_Q k + (\text{padding})$, with R_Q the charge radius of the Q-ball of Section 6.1 and k the relevant asymptotic momentum; each partial-wave contribution to (B.12)–(B.15) falls off once ℓ exceeds this scale, since a mode of angular momentum ℓ classically misses the soliton core once its impact parameter ℓ/k exceeds R_Q .

Figure B.1 shows the cumulative sum $\sum_{\ell'=0}^{\ell} (2\ell'+1) |\cdots|^2 \times \pi/(E^2 - m^2)$ against the partial-wave ℓ , for an incoming $+$ mode on the $\omega = 0.7$ background at $\lambda = 2, 4$, confirming clean saturation for both the elastic channel $\sigma(\phi \rightarrow \phi)$ and the charge-converting channel $\sigma(\phi \rightarrow \bar{\phi})$, well before the adaptively chosen $\ell_{\max}$ is reached.

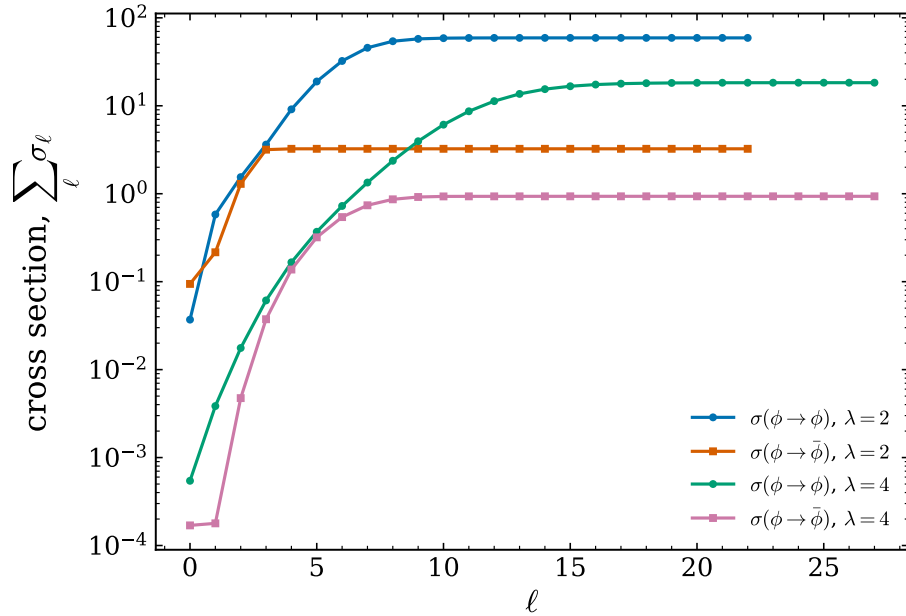


Figure B.1: Cumulative partial-wave sum for the total cross section, incoming $+$ mode, $\omega = 0.7$ background, at $\lambda = 2$ and $\lambda = 4$. Both the elastic channel $\sigma(\phi \rightarrow \phi)$ and the charge-converting channel $\sigma(\phi \rightarrow \bar{\phi})$ saturate well within the plotted range of ℓ .

These results are restricted, as throughout this thesis, to a non-spinning background: because $\mathcal{U}(r)$ in Eq. (B.1) depends on r alone, S_{ℓ} carries no m -dependence, and different values of ℓ do not mix. Both properties fail once the background itself carries angular momentum, since a rotating condensate is no longer spherically symmetric and couples partial waves of adjacent ℓ to one another; the resulting perturbation theory for spinning Q-balls has been worked out in Ref. [171], to which we refer for that extension.

Bibliography

- [1] A. Azatov, M.M. Khalil and M. Suzuki, *Towards a post-inflationary composite axion model*, *JHEP* **03** (2026) 143 [[2510.18538](#)].
- [2] C. Abel et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*, *Phys. Rev. Lett.* **124** (2020) 081803 [[2001.11966](#)].
- [3] PARTICLE DATA GROUP COLLABORATION collaboration, *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001.
- [4] N.J. Ayres, G. Ban, L. Bienstman, G. Bison, K. Bodek, V. Bondar et al., *The design of the n2edm experiment*, *The European Physical Journal C* **81** (2021) .
- [5] R.D. Peccei and H.R. Quinn, *CP conservation in the presence of pseudoparticles*, *Phys. Rev. Lett.* **38** (1977) 1440.
- [6] R.D. Peccei and H.R. Quinn, *Constraints imposed by CP conservation in the presence of pseudoparticles*, *Phys. Rev. D* **16** (1977) 1791.
- [7] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223.
- [8] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279.
- [9] D.J. Marsh, *Axion cosmology*, *Physics Reports* **643** (2016) 1.
- [10] L.D. Luzio, M. Giannotti, E. Nardi and L. Visinelli, *The landscape of qcd axion models*, *Physics Reports* **870** (2020) 1.
- [11] C. O'Hare, *Cosmology of axion dark matter*, *PoS COSMICWISPers* (2024) 040.
- [12] H. Georgi and M.B. Wise, *Hiding the Invisible Axion*, *Phys. Lett. B* **102** (1981) 275.

- [13] M. Kamionkowski and J. March-Russell, *Planck scale physics and the Peccei-Quinn mechanism*, *Phys. Lett. B* **282** (1992) 137 [[hep-th/9202003](#)].
- [14] R. Holman, S.D.H. Hsu, T.W. Kephart, E.W. Kolb, R. Watkins and L.M. Widrow, *Solutions to the strong CP problem in a world with gravity*, *Phys. Lett. B* **282** (1992) 132 [[hep-ph/9203206](#)].
- [15] J. Alvey and M. Escudero, *The axion quality problem: global symmetry breaking and wormholes*, *Journal of High Energy Physics* **2021** (2021) .
- [16] G. 't Hooft, *Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking*, 1979, <https://api.semanticscholar.org/CorpusID:119017162>.
- [17] M.S. Turner, *Cosmic and local mass density of “invisible” axions*, *Phys. Rev. D* **33** (1986) 889.
- [18] G. Raffelt and A. Caputo, *Astrophysical axion bounds: The 2024 edition*, in *Proceedings of 1st General Meeting and 1st Training School of the COST Action COSMIC WSIPers — PoS(COSMICWISPers)*, p. 041, Sissa Medialab, 2024, [DOI](#).
- [19] T.W.B. Kibble, *Topology of cosmic domains and strings*, *Journal of Physics A: Mathematical and General* **9** (1976) 1387.
- [20] P. Sikivie, *Axions, domain walls, and the early universe*, *Phys. Rev. Lett.* **48** (1982) 1156.
- [21] J.E. Kim, *Composite invisible axion*, *Phys. Rev. D* **31** (1985) 1733.
- [22] K. Choi and J.E. Kim, *Dynamical axion*, *Phys. Rev. D* **32** (1985) 1828.
- [23] G. Lazarides and Q. Shafi, *Axion models with no domain wall problem*, *Physics Letters B* **115** (1982) 21.
- [24] D. Dunsby, L.J. Hall and K. Harigaya, *CHAMP Cosmic Rays*, *JCAP* **07** (2019) 015 [[1812.11116](#)].
- [25] M.B. Gavela, M. Ibe, P. Quilez and T.T. Yanagida, *Automatic Peccei-Quinn symmetry*, *Eur. Phys. J. C* **79** (2019) 542 [[1812.08174](#)].
- [26] S.M. Barr and D. Seckel, *Planck scale corrections to axion models*, *Phys. Rev. D* **46** (1992) 539.

- [27] S.M. Barr, *Harmless Axions in Superstring Theories*, *Phys. Lett. B* **158** (1985) 397.
- [28] R.D. Peccei, *The strong cp problem and axions*, in *Axions: Theory, Cosmology, and Experimental Searches*, vol. 741 of *Lecture Notes in Physics*, (Berlin Heidelberg), pp. 3–17, Springer (2008), DOI.
- [29] P. Sikivie, *Axion cosmology*, in *Axions: Theory, Cosmology, and Experimental Searches*, vol. 741 of *Lecture Notes in Physics*, (Berlin Heidelberg), pp. 19–50, Springer (2008), DOI.
- [30] S. Weinberg, *The $u(1)$ problem*, *Phys. Rev. D* **11** (1975) 3583.
- [31] G. 't Hooft, *Symmetry breaking through bell-jackiw anomalies*, *Phys. Rev. Lett.* **37** (1976) 8.
- [32] G. 't Hooft, *Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle*, *Phys. Rev. D* **14** (1976) 3432.
- [33] S.L. Adler, *Axial-vector vertex in spinor electrodynamics*, *Phys. Rev.* **177** (1969) 2426.
- [34] J.S. Bell and R. Jackiw, *A pcac puzzle: π^0 to gamma gamma in the sigma-model*, *Il Nuovo Cimento A* **60** (1969) 47.
- [35] W.A. Bardeen, *Anomalous ward identities in spinor field theories*, *Phys. Rev.* **184** (1969) 1848.
- [36] W.A. Bardeen, *Anomalous currents in gauge field theories*, *Nuclear Physics B* **75** (1974) 246.
- [37] R.J. Crewther, *Effects of topological charge in gauge theories*, in *Facts and Prospects of Gauge Theories*, P. Urban, ed., (Vienna), pp. 47–153, Springer Vienna, 1978.
- [38] C.G. Callan, R.F. Dashen and D.J. Gross, *The structure of the gauge theory vacuum*, *Physics Letters B* **63** (1976) 334.
- [39] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Y.S. Tyupkin, *Pseudoparticle Solutions of the Yang-Mills Equations*, *Phys. Lett. B* **59** (1975) 85.
- [40] R. Jackiw and C. Rebbi, *Vacuum periodicity in a yang-mills quantum theory*, *Phys. Rev. Lett.* **37** (1976) 172.

- [41] K. Fujikawa, *Path-integral measure for gauge-invariant fermion theories*, *Physical Review Letters* **42** (1979) 1195.
- [42] S. Coleman, *The uses of instantons*, in *The Whys of Subnuclear Physics*, A. Zichichi, ed., vol. 15 of *The Subnuclear Series*, (New York), pp. 805–941, Plenum Press (1979), [DOI](#).
- [43] S. Coleman, *The uses of instantons*, in *Aspects of Symmetry: Selected Erice Lectures*, (Cambridge), pp. 265–350, Cambridge University Press (1985).
- [44] A.I. Vainshtein, V.I. Zakharov, V.A. Novikov and M.A. Shifman, *Abc of instantons*, *Soviet Physics Uspekhi* **25** (1982) 195.
- [45] G.G. di Cortona, E. Hardy, J.P. Vega and G. Villadoro, *The qcd axion, precisely*, *Journal of High Energy Physics* **2016** (2016) .
- [46] S. Coleman, J. Wess and B. Zumino, *Structure of Phenomenological Lagrangians. I*, *Phys. Rev.* **177** (1969) 2239.
- [47] J. Callan, Curtis G., S. Coleman, J. Wess and B. Zumino, *Structure of Phenomenological Lagrangians. II*, *Phys. Rev.* **177** (1969) 2247.
- [48] S. Weinberg, *Phenomenological Lagrangians*, *Physica A* **96** (1979) 327.
- [49] J. Gasser and H. Leutwyler, *Chiral Perturbation Theory to One Loop*, *Annals Phys.* **158** (1984) 142.
- [50] J. Gasser and H. Leutwyler, *Chiral Perturbation Theory: Expansions in the Mass of the Strange Quark*, *Nucl. Phys. B* **250** (1985) 465.
- [51] J. Gasser, M. Sainio and A. Švarc, *Nucleons with chiral loops*, *Nuclear Physics B* **307** (1988) 779.
- [52] R.J. Crewther, P. Di Vecchia, G. Veneziano and E. Witten, *Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics*, *Phys. Lett. B* **88** (1979) 123.
- [53] M. Pospelov and A. Ritz, *Electric dipole moments as probes of new physics*, *Annals Phys.* **318** (2005) 119 [[hep-ph/0504231](#)].
- [54] Y. Aoki, T. Blum, G. Colangelo, S. Collins, M.D. Morte, P. Dimopoulos et al., *Flag review 2021*, *The European Physical Journal C* **82** (2022) .

- [55] M. Gorghetto and G. Villadoro, *Topological susceptibility and qcd axion mass: Qcd and nnlo corrections*, *Journal of High Energy Physics* **2019** (2019) .
- [56] W.A. Bardeen, R. Peccei and T. Yanagida, *Constraints on variant axion models*, *Nuclear Physics B* **279** (1987) 401.
- [57] Y. Asano, E. Kikutani, S. Kurokawa, T. Miyachi, M. Miyajima, Y. Nagashima et al., *Search for a rare decay mode $k^+ \rightarrow \pi^+ \nu \bar{\nu}$ and axion*, *Physics Letters B* **107** (1981) 159.
- [58] NA62 collaboration, *Observation of the $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ decay and measurement of its branching ratio*, *JHEP* **02** (2025) 191 [[2412.12015](#)].
- [59] NA62 collaboration, *New measurement of $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ branching ratio at the NA62 experiment*, in *60th Rencontres de Moriond on Electroweak Interactions and Unified Theories: Moriond EW 2026*, 4, 2026 [[2604.12649](#)].
- [60] N. Iwamoto, *Axion emission from neutron stars*, *Phys. Rev. Lett.; (United States)* **53:12** (1984) .
- [61] M. Buschmann, C. Dessert, J.W. Foster, A.J. Long and B.R. Safdi, *Upper Limit on the QCD Axion Mass from Isolated Neutron Star Cooling*, *Phys. Rev. Lett.* **128** (2022) 091102 [[2111.09892](#)].
- [62] M. Dine, W. Fischler and M. Srednicki, *A simple solution to the strong cp problem with a harmless axion*, *Physics Letters B* **104** (1981) 199.
- [63] A.R. Zhitnitsky, *On possible suppression of the axion hadron interactions*, *Sov. J. Nucl. Phys.* **31** (1980) 260.
- [64] J.E. Kim, *Weak-interaction singlet and strong CP invariance*, *Phys. Rev. Lett.* **43** (1979) 103.
- [65] M. Shifman, A. Vainshtein and V. Zakharov, *Can confinement ensure natural cp invariance of strong interactions?*, *Nuclear Physics B* **166** (1980) 493.
- [66] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi and O. Straniero, *Revisiting the bound on axion-photon coupling from globular clusters*, *Phys. Rev. Lett.* **113** (2014) 191302 [[1406.6053](#)].
- [67] F. Capozzi and G. Raffelt, *Axion and neutrino bounds improved with new calibrations of the tip of the red-giant branch using geometric distance determinations*, *Phys. Rev. D* **102** (2020) 083007 [[2007.03694](#)].

- [68] O. Straniero, C. Pallanca, E. Dalessandro, I. Domínguez, F.R. Ferraro, M. Giannotti et al., *The rgb tip of galactic globular clusters and the revision of the axion-electron coupling bound*, *Astron. Astrophys.* **644** (2020) A166 [[2010.03833](#)].
- [69] J. Isern, E. García-Berro, S. Torres and S. Catalán, *Axions and the cooling of white dwarf stars*, *Astrophys. J. Lett.* **682** (2008) L109 [[0806.2807](#)].
- [70] T. Battich, A.H. Córscico, L.G. Althaus and M.M. Miller Bertolami, *First axion bounds from a pulsating helium-rich white dwarf star*, *JCAP* **08** (2016) 062 [[1605.07668](#)].
- [71] G.G. Raffelt, *Astrophysical axion bounds*, in *Axions: Theory, Cosmology, and Experimental Searches*, vol. 741 of *Lecture Notes in Physics*, (Berlin Heidelberg), pp. 51–71, Springer (2008), [DOI](#).
- [72] L.B. Leinson, *Axion mass limit from observations of the neutron star in cassiopeia a*, *JCAP* **08** (2014) 031 [[1405.6873](#)].
- [73] K. Hamaguchi, N. Nagata, K. Yanagi and J. Zheng, *Limit on the axion decay constant from the cooling neutron star in cassiopeia a*, *Phys. Rev. D* **98** (2018) 103015 [[1806.07151](#)].
- [74] M. Buschmann, R.T. Co, C. Dessert and B.R. Safdi, *Axion emission can explain a new hard x-ray excess from nearby isolated neutron stars*, *Phys. Rev. Lett.* **126** (2021) 021102 [[1910.04164](#)].
- [75] KAMIOKANDÉ-II collaboration, *Observation of a neutrino burst from the supernova sn1987a*, *Phys. Rev. Lett.* **58** (1987) 1490.
- [76] R.M. Bionta et al., *Observation of a neutrino burst in coincidence with supernova sn1987a in the large magellanic cloud*, *Phys. Rev. Lett.* **58** (1987) 1494.
- [77] E.N. Alekseev, L.N. Alekseeva, I.V. Krivosheina and V.I. Volchenko, *Detection of the neutrino signal from sn1987a using the INR baksan underground scintillation telescope*, *JETP Lett.* **45** (1987) 589.
- [78] P. Carenza, T. Fischer, M. Giannotti, G. Guo, G. Martínez-Pinedo and A. Mirizzi, *Improved axion emissivity from a supernova via nucleon-nucleon bremsstrahlung*, *JCAP* **10** (2019) 016 [[1906.11844](#)].

- [79] A. Lella, P. Carenza, G. Co', G. Lucente, M. Giannotti, A. Mirizzi et al., *Getting the most on supernova axions*, *Phys. Rev. D* **109** (2024) 023001.
- [80] N. Bar, K. Blum and G. D'Amico, *Is there a supernova bound on axions?*, *Phys. Rev. D* **101** (2020) 123025.
- [81] CAST collaboration, *New cast limit on the axion-photon interaction*, *Nature Phys.* **13** (2017) 584 [[1705.02290](#)].
- [82] IAXO collaboration, *Physics potential of the international axion observatory (iaxo)*, *JCAP* **06** (2019) 047 [[1904.09155](#)].
- [83] ALPS collaboration, *New alps results on hidden-sector lightweights*, *Phys. Lett. B* **689** (2010) 149 [[1004.1313](#)].
- [84] OSQAR collaboration, *New exclusion limits on scalar and pseudoscalar axionlike particles from light shining through a wall*, *Phys. Rev. D* **92** (2015) 092002 [[1506.08082](#)].
- [85] P. Sikivie, *Experimental tests of the invisible axion*, *Phys. Rev. Lett.* **51** (1983) 1415.
- [86] ADMX collaboration, *A search for invisible axion dark matter with the axion dark matter experiment*, *Phys. Rev. Lett.* **120** (2018) 151301 [[1804.05750](#)].
- [87] ADMX collaboration, *Search for invisible axion dark matter in the 3.3–4.2 μeV mass range*, *Phys. Rev. Lett.* **127** (2021) 261803 [[2110.06096](#)].
- [88] ADMX COLLABORATION collaboration, *Admx axion dark matter bounds around 3.3 μeV with dine-fischler-srednicki-zhitnitsky discovery ability*, *Phys. Rev. Lett.* **134** (2025) 111002.
- [89] J. Walter, O. Maliaka, Y. Zhang, J.W. Blanchard, G. Centers, A. Dogan et al., *Search for axionlike dark matter using liquid-state nuclear magnetic resonance*, *Phys. Rev. D* **112** (2025) 052008.
- [90] C.B. Adams et al., *Axion Dark Matter*, in *Snowmass 2021*, 3, 2022 [[2203.14923](#)].
- [91] J. Preskill, M. Wise and F. Wilczek, *Cosmology of the invisible axion*, *Physics Letters B* **120** (1983) 127.
- [92] L.F. Abbott and P. Sikivie, *A cosmological bound on the invisible axion*, *Physics Letters B* **120** (1983) 133.

- [93] M. Dine and W. Fischler, *The not-so-harmless axion*, *Physics Letters B* **120** (1983) 137.
- [94] D.J. Gross, R.D. Pisarski and L.G. Yaffe, *Qcd and instantons at finite temperature*, *Reviews of Modern Physics* **53** (1981) 43.
- [95] S. Borsanyi et al., *Calculation of the axion mass based on high-temperature lattice quantum chromodynamics*, *Nature* **539** (2016) 69 [[1606.07494](#)].
- [96] PLANCK collaboration, *Planck 2018 results. vi. cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [[1807.06209](#)].
- [97] V.F. Mukhanov and G.V. Chibisov, *Quantum Fluctuations and a Nonsingular Universe*, *JETP Lett.* **33** (1981) 532.
- [98] A. Starobinsky, *Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations*, *Physics Letters B* **117** (1982) 175.
- [99] A.H. Guth and S.Y. Pi, *Fluctuations in the New Inflationary Universe*, *Phys. Rev. Lett.* **49** (1982) 1110.
- [100] D.H. Lyth and E.D. Stewart, *Constraining the inflationary energy scale from axion cosmology*, *Physics Letters B* **283** (1992) 189.
- [101] PLANCK collaboration, *Planck 2015 results. xx. constraints on inflation*, *Astron. Astrophys.* **594** (2016) A20.
- [102] A. Vilenkin, *Cosmic Strings and Domain Walls*, *Phys. Rept.* **121** (1985) 263.
- [103] M. Gorghetto, E. Hardy and G. Villadoro, *Axions from strings: the attractive solution*, *Journal of High Energy Physics* **2018** (2018) .
- [104] M. Buschmann, J.W. Foster, A. Hook, A. Peterson, D.E. Willcox, W. Zhang et al., *Dark matter from axion strings with adaptive mesh refinement*, *Nature Communications* **13** (2022) .
- [105] J.N. Benabou, M. Buschmann, S. Kumar, Y. Park and B.R. Safdi, *Signatures of primordial energy injection from axion strings*, *Phys. Rev. D* **109** (2024) 055005.
- [106] D. Harari and P. Sikivie, *On the evolution of global strings in the early universe*, *Physics Letters B* **195** (1987) 361.

- [107] H. Georgi, *A Tool Kit for Builders of Composite Models*, *Nucl. Phys. B* **266** (1986) 274.
- [108] N. Arkani-Hamed, A.G. Cohen and H. Georgi, *(De)constructing dimensions*, *Phys. Rev. Lett.* **86** (2001) 4757 [[hep-th/0104005](#)].
- [109] N. Arkani-Hamed, A.G. Cohen and H. Georgi, *Electroweak symmetry breaking from dimensional deconstruction*, *Phys. Lett. B* **513** (2001) 232 [[hep-ph/0105239](#)].
- [110] M. Redi and R. Sato, *Composite Accidental Axions*, *JHEP* **05** (2016) 104 [[1602.05427](#)].
- [111] C. Vafa and E. Witten, *Restrictions on symmetry breaking in vector-like gauge theories*, *Nuclear Physics B* **234** (1984) 173.
- [112] C. Vafa and E. Witten, *Parity conservation in quantum chromodynamics*, *Physical Review Letters* **53** (1984) 535.
- [113] H. Georgi and S.L. Glashow, *Unity of all elementary particle forces*, *Phys. Rev. Lett.* **32** (1974) 438.
- [114] P. Agrawal and K. Howe, *Factoring the Strong CP Problem*, *JHEP* **12** (2018) 029 [[1710.04213](#)].
- [115] C. Csáki, M. Ruhdorfer and Y. Shirman, *UV Sensitivity of the Axion Mass from Instantons in Partially Broken Gauge Groups*, *JHEP* **04** (2020) 031 [[1912.02197](#)].
- [116] T. Aoki, M. Ibe, S. Shirai and K. Watanabe, *Small instanton effects on composite axion mass*, *JHEP* **07** (2024) 269 [[2404.19342](#)].
- [117] G. 't Hooft, *Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking*, in *Recent Developments in Gauge Theories (Cargèse, 1979)*, G. 't Hooft et al., eds., vol. 59 of *NATO Sci. Ser. B*, pp. 135–157, Plenum Press (1980), [DOI](#).
- [118] S. Raby, S. Dimopoulos and L. Susskind, *Tumbling gauge theories*, *Nuclear Physics B* **169** (1980) 373.
- [119] E. Eichten and F. Feinberg, *Comment on tumbling gauge theories*, *Physics Letters B* **110** (1982) 232.

- [120] H.-L. Li, Álvaro Pastor-Gutiérrez, S. Vatan and L.-X. Xu, *Dynamical symmetry breaking in georgi-glashow chiral-gauge theories*, *Journal of High Energy Physics* **2025** (2025) .
- [121] H.B. Nielsen and P. Olesen, *Vortex-line models for dual strings*, *Nucl. Phys. B* **61** (1973) 45.
- [122] T. Hiramatsu, M. Ibe and M. Suzuki, *New Type of String Solutions with Long Range Forces*, *JHEP* **02** (2020) 058 [[1910.14321](#)].
- [123] T. Hiramatsu, M. Ibe and M. Suzuki, *Cosmic string in Abelian-Higgs model with enhanced symmetry — Implication to the axion domain-wall problem*, *JHEP* **09** (2020) 054 [[2005.10421](#)].
- [124] P. Batra, B.A. Dobrescu and D. Spivak, *Anomaly-free sets of fermions*, *J. Math. Phys.* **47** (2006) 082301 [[hep-ph/0510181](#)].
- [125] D.H. Lyth and E.D. Stewart, *Thermal inflation and the moduli problem*, *Phys. Rev. D* **53** (1996) 1784 [[hep-ph/9510204](#)].
- [126] B.W. Lee and S. Weinberg, *Cosmological lower bound on heavy-neutrino masses*, *Phys. Rev. Lett.* **39** (1977) 165.
- [127] E.W. Kolb and M.S. Turner, *The Early Universe*, vol. 69, Taylor and Francis (5, 2019), [10.1201/9780429492860](#).
- [128] K. Harigaya and L.-T. Wang, *More axions from diluted domain walls*, [2211.08289](#).
- [129] F. Kajino, S. Matsuno, T. Kitamura, T. Aoki, Y.K. Yuan, K. Mitsui et al., *Measurement of the energy spectrum of cosmic-ray muons at 89° zenith angle*, *J. Phys. G* **10** (1984) 447.
- [130] B. Barish, G. Liu and C.E. Lane, *Search for grand unification monopoles and other ionizing heavy particles using a scintillation detector at the Earth’s surface*, *Phys. Rev. D* **36** (1987) 2641.
- [131] E.N. Alekseev, M.M. Boliev, A.E. Chudakov, S.P. Mikheyev and O.Y. Shkvorets, *Search for slowly moving penetrating particles at Baksan Underground Telescope*, in *18th International Cosmic Ray Conference*, vol. HE2.2-1, pp. 52–55, 1983, <https://inspirehep.net/literature/1362537>.

- [132] E.N. Alekseev, M.M. Boliev, A.E. Chudakov and S.P. Mikheyev, *Upper Limit on Magnetic Monopole Flux from Baksan Experiment*, in *19th International Cosmic Ray Conference*, vol. HE-8, p. 250, 1985, <https://inspirehep.net/literature/226938>.
- [133] G.F. Giudice, E.W. Kolb and A. Riotto, *Largest temperature of the radiation era and its cosmological implications*, *Phys. Rev. D* **64** (2001) 023508.
- [134] R. Rajaraman, *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory*, North-Holland, Amsterdam (1987).
- [135] T.D. Lee and Y. Pang, *Nontopological solitons*, *Phys. Rept.* **221** (1992) 251.
- [136] S. Coleman, *Q-balls*, *Nuclear Physics B* **262** (1985) 263.
- [137] I. Affleck and M. Dine, *A New Mechanism for Baryogenesis*, *Nucl. Phys. B* **249** (1985) 361.
- [138] R. Allahverdi and A. Mazumdar, *A mini review on Affleck-Dine baryogenesis*, *New J. Phys.* **14** (2012) 125013 [[1206.2913](#)].
- [139] S. Kasuya and M. Kawasaki, *Q-ball Formation: Obstacle to Affleck-Dine Baryogenesis in the Gauge-mediated SUSY breaking?*, *Phys. Rev. D* **84** (2011) 123528 [[1107.0403](#)].
- [140] S. Kasuya, M. Kawasaki and M. Yamada, *Revisiting Gravitino Dark Matter from Q-ball Decay*, *Phys. Lett. B* **726** (2013) 1 [[1211.4743](#)].
- [141] J.A. Frieman, G.B. Gelmini, M. Gleiser and E.W. Kolb, *Solitogenesis: Primordial Origin of Nontopological Solitons*, *Phys. Rev. Lett.* **60** (1988) 2101.
- [142] K. Griest, E.W. Kolb and A. Massarotti, *Statistical Fluctuations as the Origin of Nontopological Solitons*, *Phys. Rev. D* **40** (1989) 3529.
- [143] K. Griest and E.W. Kolb, *Solitogenesis: Cosmological Evolution of Nontopological Solitons*, *Phys. Rev. D* **40** (1989) 3231.
- [144] M. Postma, *Solitogenesis of Q balls*, *Phys. Rev. D* **65** (2002) 085035 [[hep-ph/0110199](#)].
- [145] K. Griest and D. Seckel, *Three Exceptions in the Calculation of Relic Abundances*, *Phys. Rev. D* **43** (1991) 3191.

- [146] Y. Bai, S. Lu and N. Orlofsky, *Origin of nontopological soliton dark matter: solitosynthesis or phase transition*, *JHEP* **10** (2022) 181 [[2208.12290](#)].
- [147] A. Kusenko, L.C. Loveridge and M. Shaposhnikov, *Astrophysical bounds on the maximum charge of dark matter Q -balls*, *JCAP* **08** (2005) 011 [[astro-ph/0507225](#)].
- [148] A. Kusenko and M.E. Shaposhnikov, *Supersymmetric Q balls as dark matter*, *Phys. Lett. B* **418** (1998) 46 [[hep-ph/9709492](#)].
- [149] A. Kusenko, *Small Q balls*, *Phys. Lett. B* **404** (1997) 285 [[hep-th/9704073](#)].
- [150] K. Enqvist and J. McDonald, *Q balls and baryogenesis in the MSSM*, *Phys. Lett. B* **440** (1998) 59 [[hep-ph/9807269](#)].
- [151] A. Kusenko and P.J. Steinhardt, *Q ball candidates for selfinteracting dark matter*, *Phys. Rev. Lett.* **87** (2001) 141301 [[astro-ph/0106008](#)].
- [152] L. Roszkowski and O. Seto, *Axino dark matter from Q balls in Affleck-Dine baryogenesis and the Ω_b - Ω_{DM} coincidence problem*, *Phys. Rev. Lett.* **98** (2007) 161304 [[hep-ph/0608013](#)].
- [153] I.M. Shoemaker and A. Kusenko, *Gravitino Dark Matter from Q -ball Decays*, *Phys. Rev. D* **80** (2009) 075021 [[0909.3334](#)].
- [154] M. Kawasaki and H. Nakatsuka, *Gravitational waves from type II Affleck-Dine mechanism in NMSSM*, *JCAP* **04** (2020) 017 [[1912.06993](#)].
- [155] E. Pontón, Y. Bai and B. Jain, *Electroweak Symmetric Dark Matter Balls*, *JHEP* **09** (2019) 011 [[1906.10739](#)].
- [156] P.M. Saffin, Q.-X. Xie and S.-Y. Zhou, *Q -ball Superradiance*, *Phys. Rev. Lett.* **131** (2023) 111601 [[2212.03269](#)].
- [157] V. Cardoso, R. Vicente and Z. Zhong, *Energy Extraction from Q -balls and Other Fundamental Solitons*, *Phys. Rev. Lett.* **131** (2023) 111602 [[2307.13734](#)].
- [158] R. Friedberg, T.D. Lee and A. Sirlin, *A Class of Scalar-Field Soliton Solutions in Three Space Dimensions*, *Phys. Rev. D* **13** (1976) 2739.
- [159] R. Friedberg, T.D. Lee and A. Sirlin, *Gauge Field Nontopological Solitons in Three Space Dimensions. 1.*, *Nucl. Phys. B* **115** (1976) 1.

- [160] R. Friedberg, T.D. Lee and A. Sirlin, *Gauge Field Nontopological Solitons in Three Space Dimensions. 2.*, *Nucl. Phys. B* **115** (1976) 32.
- [161] M. Laine and M.E. Shaposhnikov, *Thermodynamics of nontopological solitons*, *Nucl. Phys. B* **532** (1998) 376 [[hep-ph/9804237](#)].
- [162] S. Coleman, V. Glaser and A. Martin, *Action minima among solutions to a class of euclidean scalar field equations*, *Communications in Mathematical Physics* **58** (1978) 211.
- [163] E.Y. Nugaev and A.V. Shkerin, *Review of Nontopological Solitons in Theories with $U(1)$ -Symmetry*, *J. Exp. Theor. Phys.* **130** (2020) 301 [[1905.05146](#)].
- [164] S.-Y. Zhou, *Non-topological solitons and quasi-solitons*, *Reports on Progress in Physics* **88** (2025) 046901.
- [165] M.I. Tsumagari, *The Physics of Q-Balls*, ph.d. thesis, Nottingham University, Oct., 2009. [0910.3845](#).
- [166] E. Kim, E. Nugaev and Y. Shnir, *Large solitons flattened by small quantum corrections*, *Physics Letters B* **856** (2024) 138881.
- [167] U.M. Ascher, R.M.M. Mattheij and R.D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Classics in Applied Mathematics, SIAM, Philadelphia (1995).
- [168] A. Azatov, Q.T. Ho and M.M. Khalil, *Q-ball perturbations with more details: Linear analysis vs lattice*, *Phys. Rev. D* **111** (2025) 096010.
- [169] M.N. Smolyakov, *Perturbations against a Q-ball: Charge, energy, and additivity property*, *Phys. Rev. D* **97** (2018) 045011 [[1711.05730](#)].
- [170] E.P. Wigner, *On the behavior of cross sections near thresholds*, *Physical Review* **73** (1948) 1002.
- [171] G.-D. Zhang, F.-M. Chang, P.M. Saffin, Q.-X. Xie and S.-Y. Zhou, *Spinning Q-ball Superradiance in 3+1D*, [2402.03193](#).