The Standard Model of Particle Physics

Uwe-Jens Wiese
Institute for Theoretical Physics
University of Bern

March 17, 2010

Contents

Chapter 1

Introduction

The standard model of particle physics summarizes all we know about the fundamental forces of electromagnetism, as well as the weak and strong interactions (but not gravity). It has been tested in great detail up to energies in the hundred GeV range and has passed all these tests very well. The standard model is a relativistic quantum field theory that incorporates the basic principles of quantum mechanics and special relativity. Like quantum electrodynamics (QED) the standard model is a gauge theory, however, with the non-Abelian gauge group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ instead of the simple Abelian $U(1)_{em}$ gauge group of QED. The gauge bosons are the photons mediating the electromagnetic interactions, the W- and Z-bosons mediating the weak interactions, as well as the gluons mediating the strong interactions. Gauge theories can exist in several phases: in the Coulomb phase with massless gauge bosons (like in QED), in the Higgs-phase with spontaneously broken gauge symmetry and with massive gauge bosons (e.g. the W- and Z-bosons), and in the confinement phase, in which the gauge bosons do not appear in the spectrum (like the gluons in quantum chromodynamics (QCD)). All these different phases are indeed realized in Nature and hence in the standard model that describes it.

In particle physics symmetries play a central role. One distinguishes global and local symmetries. Global symmetries are usually only approximate. Exact symmetries, on the other hand, are locally realized, and require the existence of a gauge field. Our world is not quite as symmetric as the theories we use to describe it. This is because many symmetries are broken. The simplest form of symmetry breaking is explicit breaking which is due to non-invariant symmetry breaking terms in the classical Lagrangian of the theory. On the other hand, the quantization of the theory may also lead to explicit symmetry breaking, even if the classical Lagrangian is invariant. In that case one has an anomaly which is due to an explicit symmetry breaking in the measure of the Feynman path integral. Only global symmetries can be explicitly broken (either in the Lagrangian or via an anomaly). Theories with explicitly broken gauge symmetries, on the other hand, are inconsistent (perturbatively and even non-perturbatively non-renormalizable). For example, in the standard model all gauge anomalies are canceled due to the properly arranged fermion content of each generation.

A more interesting form of symmetry breaking is spontaneous symmetry breaking which is a dynamical effect. When a continuous global symmetry breaks spontaneously, massless Goldstone bosons appear in the spectrum. If there is, in addition, a weak explicit symmetry breaking, the Goldstone bosons pick up a small mass. This is the case for the pions, which arise as a consequence of the spontaneous breaking of the approximate global chiral symmetry in QCD. When a gauge symmetry is spontaneously broken one has the so-called Higgs mechanism which gives mass to the gauge bosons. This gives rise to an additional helicity state. This state has the quantum numbers of a Goldstone boson that would arise if the symmetry were global. One says that the gauge boson eats the Goldstone boson and thus becomes massive.

The fermions in the standard model are either leptons or quarks. Leptons participate only in the electromagnetic and weak gauge interactions, while quarks also participate in the strong interactions. Quarks and leptons also pick up their masses through the Higgs mechanism. The values of these masses are free parameters of the standard model that are presently not understood on the basis of a more fundamental theory. There are six quarks: up, down, strange, charm, bottom, and top, and six leptons: the electron, muon, tau, as well as their corresponding neutrinos. The weak interaction eigenstates are mixed to form the mass eigenstates. The quark mixing Cabbibo-Kobayashi-Maskawa (CKM) matrix contains several more free parameters of the standard model. Recently, convincing experimental evidence for non-zero neutrino masses has been accumulated. This implies that there are not only additional free mass parameters for the electron-, muon-, and tau-neutrinos, but an entire lepton mixing matrix. Altogether, the fermion sector of the standard model has so many free parameters that it is hard to believe that there should not be a more fundamental theory that will be able to explain the values of these parameters.

There is a very interesting parameter in the standard model — the CP violating QCD θ-vacuum angle — which seems to be zero in the real world. The strong CP problem is to understand why this is the case. The θ-angle is related to the topology of the gluon field which manifests itself e.g. in so-called instanton field configurations. The standard model can be extended by the introduction of a second Higgs field which gives rise to an additional $U(1)_{PQ}$ symmetry as first suggested by Peccei and Quinn, and it naturally leads to $\theta = 0$. The spontaneous breaking of the Peccei-Quinn symmetry leads to an almost massless Goldstone boson — the axion. If this particle would be found in experimental searches, it could be a first concrete hint to the physics beyond the standard model.

Non-trivial topology also arises for the electroweak gauge field. This leads to an anomaly in the fermion number — or more precisely in the $U(1)_{B+L}$ global symmetry of baryon plus lepton number. In particular, baryon number itself is not strictly conserved in the standard model. This has been discussed as a possible explanation of the baryon asymmetry in the universe — the fact that there is more matter than anti-matter. It is more likely that baryon number violating processes beyond the standard model are responsible for the baryon asymmetry. For example, in the $SU(5)$ grand unified theory (GUT) baryon number violating processes appear naturally at extremely high energies close to the GUT scale. Although the $U(1)_{B+L}$ symmetry is explicitly broken by an anomaly, the global $U(1)_{B-L}$ symmetry remains intact both in the standard model and in the $SU(5)$ GUT, at least if the neutrinos were massless. This would, in fact, be quite strange (an exact symmetry should be local, not global) and we now know that neutrinos are indeed massive. A grand unified theory that naturally incorporates massive neutrinos is based on the symmetry group $SO(10)$. In this model $B - L$ is also violated and all exact symmetries are locally realized. In addition, the so-called see-saw mechanism gives a natural explanation for very small neutrino masses.

Despite these successes of grand unified theories, they suffer from the hierarchy problem — the puzzle how to stabilize the electroweak scale against the much higher GUT scale. This may be achieved using supersymmetric theories which would lead us to questions beyond the scope of this course. Another attempt to avoid the hierarchy problem is realized in technicolor models which have their own problems and are hence no longer popular. Still, they are interesting from a theoretical point of view and will therefore also be discussed.

At the end of the course we will study the implications of standard model physics for the evolution of the Universe immediately after the big bang with an emphasis on the electroweak and strong phase transitions.

In this course we will not put much emphasis on the rich and successful detailed phenomenology resulting from the standard model. Of course, this is very interesting as well, but would be the subject for another course. Instead, we will concentrate on the general structure and the symmetries of the standard model and some theories that go beyond it.

Chapter 2

Higgs and Gauge Bosons

In this chapter we introduce the bosonic part of the standard model. Even without fermions (leptons or quarks) there is a lot of nontrivial physics of Higgs and gauge bosons alone. In the pure Higgs sector a global $O(4) = SU(2)_L \otimes SU(2)_R$ symmetry gets spontaneously broken down to $O(3) = SU(2)_{L=R}$. According to the Goldstone theorem, this gives rise to three massless bosons. When electroweak gauge fields are included, the $SU(2)_L$ symmetry as well as a $U(1)_Y$ subgroup of $SU(2)_R$ become local symmetries. The electroweak $SU(2)_L \otimes U(1)_Y$ gauge symmetry breaks spontaneously to $U(1)_{em}$ — the gauge group of electromagnetism. Due to the Higgs mechanism the W and Z gauge bosons become massive. The additional third polarization states of the three massive vector bosons W^+, W^-, and Z are provided by the three Goldstone modes, which also become massive. One says that the gauge bosons "eat" the Goldstone bosons and thus get a mass. The photon, on the other hand, remains massless as a consequence of the unbroken $U(1)_{em}$ gauge symmetry of electromagnetism. Before quarks are added, the bosonic gluons do not interact with Higgs bosons, W and Z bosons, or photons. Hence, we will add the gluons only later when we discuss the strong interactions. To illustrate the basic ideas behind the Higgs mechanism, we first turn to a simpler toy model — electrodynamics with charged scalar fields.

2.1 Scalar Electrodynamics

The prototype of a gauge theory is quantum electrodynamics, the theory of the electromagnetic interaction between charged electrons and positrons via photon exchange. We want to make our life easy, and consider charged particles without spin — so-called scalars. For example, we can think of the Cooper pairs in a superconductor. A charged scalar particle is described by a complex field $\Phi(x) \in \mathbb{C}$. One needs two real degrees of freedom in order to describe both particles and anti-particles. As we have discussed before, a quantum field theory is defined by a Euclidean path integral over all field configurations

$$Z = \int \mathcal{D}\Phi \ \exp(-\frac{1}{\hbar}S[\Phi]). \tag{2.1.1}$$

Here

$$S[\Phi] = \int d^4x \ \mathcal{L}(\Phi, \partial_\mu \Phi) \tag{2.1.2}$$

is the Euclidean action of the field $\Phi(x) = \Phi_1(x) + i\Phi_2(x)$ with the Lagrange density

$$\mathcal{L}(\Phi, \partial_\mu \Phi) = \frac{1}{2}\partial_\mu \Phi^* \partial_\mu \Phi - V(\Phi) = \frac{1}{2}\partial_\mu \Phi_1 \partial_\mu \Phi_1 + \frac{1}{2}\partial_\mu \Phi_2 \partial_\mu \Phi_2 - V(\Phi_1, \Phi_2). \tag{2.1.3}$$

The simplest form for the potential is that of a harmonic oscillator

$$V(\Phi) = \frac{m^2}{2}\Phi^*\Phi = \frac{m^2}{2}|\Phi|^2 = \frac{m^2}{2}(\Phi_1^2 + \Phi_2^2). \tag{2.1.4}$$

The classical field equations then take the form

$$\partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu \Phi_i} - \frac{\delta \mathcal{L}}{\delta \Phi_i} = \partial_\mu \partial_\mu \Phi_i + m^2 \Phi_i = 0. \tag{2.1.5}$$

This is the 2-component Klein-Gordon equation for a free charged scalar field. The Lagrange density has a symmetry. It is invariant under global $U(1)$ transformations

$$\Phi(x)' = \exp(ie\varphi)\Phi(x) \ \Rightarrow \ \Phi^*(x)' = \exp(-ie\varphi)\Phi^*(x), \tag{2.1.6}$$

because then

$$\partial_\mu \Phi(x)' = \exp(ie\varphi)\partial_\mu \Phi(x), \ \partial_\mu \Phi^*(x)' = \exp(-ie\varphi)\partial_\mu \Phi^*(x). \tag{2.1.7}$$

This invariance is related to the conservation of the charge e. We now also want to allow interactions by generalizing the potential $V(\Phi)$. Using the $U(1)$ invariance as a guiding principle, we come to a $|\Phi|^4$ theory

$$V(\Phi) = \frac{m^2}{2}|\Phi|^2 + \frac{\lambda}{4}|\Phi|^4. \tag{2.1.8}$$

The coupling constant λ must be positive for the potential to be bounded from below. One can, however, choose $m^2 < 0$. Hence, we distinguish two cases. For

$m^2 > 0$ the potential has a single minimum at $\Phi = 0$. The classical solution of lowest energy (the classical vacuum) is simply the constant field $\Phi(x) = 0$. When $m^2 < 0$, this trivial vacuum configuration is unstable because it corresponds to a maximum of the potential. The condition for a minimum now reads

$$\frac{\partial V}{\partial \Phi_i} = m^2 \Phi_i + \lambda |\Phi|^2 \Phi_i = 0 \;\Rightarrow\; |\Phi|^2 = -\frac{m^2}{\lambda}. \tag{2.1.9}$$

The true vacuum is no longer unique. Instead, there is a whole class of degenerate vacua

$$\Phi(x) = v \exp(i\chi), \;\; v = \sqrt{-\frac{m^2}{\lambda}}, \tag{2.1.10}$$

parameterized by an angle $\chi \in [0, 2\pi[$. The quantity v is the vacuum expectation value of the field Φ. Let us choose the vacuum state with $\chi = 0$ and let us expand around the corresponding minimum

$$\begin{aligned} &\Phi(x) = v + \sigma(x) + i\pi(x) \;\Rightarrow\; \Phi^*(x) = v + \sigma(x) - i\pi(x), \\ &|\Phi(x)|^2 = (v + \sigma(x))^2 + \pi(x)^2, \\ &\partial_\mu \Phi(x) = \partial_\mu \sigma(x) + i\partial_\mu \pi(x), \\ &\partial_\mu \Phi^*(x) = \partial_\mu \sigma(x) - i\partial_\mu \pi(x). \end{aligned} \tag{2.1.11}$$

The Lagrangian then takes the form

$$\begin{aligned} \mathcal{L}(\sigma, \partial_\mu\sigma, \pi, \partial_\mu\pi) &= \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{2}\partial_\mu\pi\partial_\mu\pi - \frac{m^2}{2}(v+\sigma)^2 - \frac{m^2}{2}\pi^2 \\ &- \frac{\lambda}{4}((v+\sigma)^2 + \pi^2)^2 \\ &\approx \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{2}\partial_\mu\pi\partial_\mu\pi - \frac{m^2}{2}v^2 - m^2 v\sigma - \frac{m^2}{2}\sigma^2 \\ &- \frac{m^2}{2}\pi^2 - \frac{\lambda}{4}(v^4 + 4v^3\sigma + 6v^2\sigma^2 + 2v^2\pi^2) \\ &= \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma - \frac{1}{2}(m^2 + 3\lambda v^2)\sigma^2 + \frac{1}{2}\partial_\mu\pi\partial_\mu\pi. \end{aligned} \tag{2.1.12}$$

One finds a σ-particle of mass

$$m_\sigma^2 = m^2 + 3\lambda v^2 = m^2 - 3\lambda\frac{m^2}{\lambda} = -2m^2 > 0, \tag{2.1.13}$$

as well as a massless π-particle with $m_\pi = 0$. This massless particle is a so-called Goldstone boson. Its presence is characteristic for the spontaneous breaking of the global $U(1)$ symmetry.

Now we want to promote the global $U(1)$ symmetry to a local one. We demand $U(1)_{em}$ gauge invariance

$$\Phi(x)' = \exp(ie\varphi(x))\Phi(x), \tag{2.1.14}$$

where $\varphi(x)$ now is a space-time dependent transformation parameter. The potential is gauge invariant ($V(\Phi') = V(\Phi)$) because

$$|\Phi(x)'|^2 = \Phi^*(x)'\Phi(x)' = \Phi^*(x)\exp(-ie\varphi(x))\exp(ie\varphi(x))\Phi(x) = |\Phi(x)|^2. \tag{2.1.15}$$

The kinetic term, on the other hand, is not invariant, because

$$\partial_\mu\Phi(x)' = \exp(ie\varphi(x))(\partial_\mu\Phi(x) + ie\partial_\mu\varphi(x)\Phi(x)). \tag{2.1.16}$$

We introduce a gauge field $A_\mu(x)$, that transforms such that the $\partial_\mu\varphi$ term is eliminated, i.e.

$$A_\mu(x)' = A_\mu(x) + \partial_\mu\varphi(x). \tag{2.1.17}$$

Then

$$(\partial_\mu - ieA_\mu(x)')\Phi(x)' = \exp(ie\varphi(x))(\partial_\mu - ieA_\mu(x))\Phi(x) \tag{2.1.18}$$

is gauge covariant. The above quantity is the covariant derivative

$$D_\mu\Phi(x) = (\partial_\mu - ieA_\mu(x))\Phi(x). \tag{2.1.19}$$

The Lagrangian can now be written in a gauge invariant way

$$\mathcal{L}(\Phi, \partial_\mu\Phi, A_\mu) = \frac{1}{2}D_\mu\Phi^* D_\mu\Phi - V(\Phi). \tag{2.1.20}$$

Up to now, the gauge field A_μ appeared only as an external field. We have not yet written a kinetic term for it. From classical electrodynamics we know how to write such a term. We construct the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{2.1.21}$$

which is gauge invariant

$$F'_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu = \partial_\mu A_\nu + \partial_\mu\partial_\nu\varphi - \partial_\nu A_\mu - \partial_\nu\partial_\mu\varphi = F_{\mu\nu}. \tag{2.1.22}$$

The Lagrange density of the free electromagnetic field is

$$\mathcal{L}(A_\mu) = -\frac{1}{4}F_{\mu\nu}F_{\mu\nu}, \tag{2.1.23}$$

which leads to Maxwell's equations

$$\partial_\mu F_{\mu\nu} = 0. \qquad (2.1.24)$$

The total Lagrangian of scalar QED takes the form

$$\mathcal{L}(\Phi, \partial_\mu \Phi, A_\mu) = -\frac{1}{4} F_{\mu\nu} F_{\mu\nu} + \frac{1}{2} D_\mu \Phi^* D_\mu \Phi - V(\Phi). \qquad (2.1.25)$$

An explicit mass term $\frac{m_\gamma^2}{2} A_\mu A_\mu$ is not allowed, because it would violate gauge invariance.

Like for the global $U(1)$ symmetry, we distinguish two cases. For $m^2 > 0$ the symmetry is unbroken, and we have a Coulomb phase with scalar particles of charge e and massless photons. The broken phase at $m^2 < 0$ is more interesting. Again, there are degenerate vacua, but they are now related by gauge transformations and thus represent the same physical state. Hence, it is helpful to fix the gauge. We choose a physical unitary gauge

$$\mathrm{Im}\Phi(x) = \Phi_2(x) = 0, \ \mathrm{Re}\Phi(x) = \Phi_1(x) \geq 0. \qquad (2.1.26)$$

Let us again consider fluctuation around the vacuum value v

$$\Phi(x) = v + \sigma(x). \qquad (2.1.27)$$

Now there is no π excitation, because in the unitary gauge the field Φ is real and positive. We find

$$\begin{aligned} V(\Phi) &= \frac{m^2}{2}(v+\sigma)^2 + \frac{\lambda}{4}(v+\sigma)^4 \\ &= \frac{m^2}{2} v^2 + m^2 v\sigma + \frac{m^2}{2}\sigma^2 + \frac{\lambda}{4}(v^4 + 4v^3\sigma + 6v^2\sigma^2) + ... \\ &= \frac{1}{2}(m^2 + 3\lambda v^2)\sigma^2 + ... = -m^2\sigma^2 + ..., \end{aligned} \qquad (2.1.28)$$

i.e., again there is a massive σ particle, but no longer a massless π. What happened to this degree of freedom? Let us consider the gauge field

$$\begin{aligned} \frac{1}{2} D_\mu \Phi^* D_\mu \Phi &= \frac{1}{2}(\partial_\mu + ieA_\mu)(v+\sigma)(\partial_\mu - ieA_\mu)(v+\sigma) \\ &= \frac{1}{2}(\partial_\mu \sigma + ieA_\mu v + ieA_\mu \sigma)(\partial_\mu \sigma - ieA_\mu v - ieA_\mu \sigma) \\ &= \frac{1}{2}\partial_\mu \sigma \partial_\mu \sigma + \frac{1}{2} e^2 v^2 A_\mu A_\mu + ... \end{aligned} \qquad (2.1.29)$$

We obtain a massive photon with

$$m_\gamma = ev. \quad (2.1.30)$$

This mechanism of mass generation is called the Higgs mechanism. It is based on the spontaneous breakdown of the gauge symmetry. A phase, in which the gauge symmetry is spontaneously broken, and in which the gauge bosons hence are massive, is called a Higgs phase. In our Universe, the $U(1)$ gauge symmetry of electrodynamics is unbroken, the photon is massless, and we are in a Coulomb phase. Still, intelligent human beings can create to coolest spot of the Universe here on earth and generate superconductivity in condensed matter. Then the $U(1)$ gauge symmetry indeed gets spontaneously broken, and the photon becomes massive. The mass can be measured, because it is related to the penetration depth of magnetic fields in the superconductor.

2.2 The Electroweak Interaction

The electroweak interaction is described by an $SU(2)_L \otimes U(1)_Y$ gauge theory, whose symmetry gets spontaneously broken to the $U(1)_{em}$ symmetry of electromagnetism. Again, a scalar Higgs field Φ plays a central role in the dynamics. However, the field Φ now has two complex components (complex doublet)

$$\Phi(x) = \begin{pmatrix} \Phi_+(x) \\ \Phi_0(x) \end{pmatrix}, \ \Phi_+(x), \Phi_0(x) \in \mathbb{C}. \quad (2.2.1)$$

The indexes + and 0 will turn out to denote electric charges. Again, let us first discuss a model with global symmetry

$$\mathcal{L}(\Phi, \partial_\mu \Phi) = \frac{1}{2}\partial_\mu \Phi^\dagger \partial_\mu \Phi - V(\Phi), \quad (2.2.2)$$

where

$$V(\Phi) = \frac{m^2}{2}|\Phi|^2 + \frac{\lambda}{4}|\Phi|^4. \quad (2.2.3)$$

Of course, now

$$|\Phi|^2 = \Phi^\dagger \Phi = \Phi_+^* \Phi_+ + \Phi_0^* \Phi_0 = \Phi_{+1}^* \Phi_{+1} + \Phi_{+2}^* \Phi_{+2} + \Phi_{01}^* \Phi_{01} + \Phi_{02}^* \Phi_{02}. \quad (2.2.4)$$

The Lagrange density is invariant under $SU(2)_L$ transformations.

$$\Phi(x)' = g\Phi(x), \ g \in SU(2)_L. \quad (2.2.5)$$

$SU(2)$ is the group of unitary 2×2 matrices with determinant 1

$$g^\dagger g = g g^\dagger = 1, \ g^\dagger = g^{T*}, \ \det g = 1. \tag{2.2.6}$$

A general $SU(2)$ matrix can be written as

$$g = \begin{pmatrix} g_1 & -g_2^* \\ g_2 & g_1^* \end{pmatrix} \Rightarrow g^\dagger = \begin{pmatrix} g_1^* & g_2^* \\ -g_2 & g_1 \end{pmatrix} \Rightarrow$$
$$g g^\dagger = \begin{pmatrix} |g_1|^2 + |g_2|^2 & 0 \\ 0 & |g_1|^2 + |g_2|^2 \end{pmatrix} = 1 \ \Rightarrow$$
$$|g_1|^2 + |g_2|^2 = 1, \ \det g = |g_1|^2 + |g_2|^2 = 1. \tag{2.2.7}$$

The space of $SU(2)$ matrices is isomorphic to the 3-dimensional sphere S^3. The global $SU(2)$ invariance of the above Lagrange density follows from

$$|\Phi'(x)|^2 = \Phi^\dagger(x)'\Phi(x)' = (g\Phi(x))^\dagger g\Phi(x) = \Phi^\dagger(x) g^\dagger g \Phi(x) = |\Phi(x)|^2,$$
$$\partial_\mu \Phi^\dagger(x)' \partial_\mu \Phi(x)' = \partial_\mu \Phi^\dagger(x) g^\dagger g \partial_\mu \Phi(x) = \partial_\mu \Phi^\dagger(x) \partial_\mu \Phi(x). \tag{2.2.8}$$

In fact, the Lagrange density is invariant even under $O(4) = SU(2)_L \otimes SU(2)_R$ transformations. However, only the $SU(2)_L$ symmetry as well as a $U(1)_Y$ subgroup of $SU(2)_R$ are gauged in the standard model. The custodial $SU(2)_R$ symmetry is easy to understand when one rewrites the Higgs field as a 2×2 matrix

$$\Phi = \begin{pmatrix} \Phi_+ & -\Phi_0^* \\ \Phi_0 & \Phi_+^* \end{pmatrix}. \tag{2.2.9}$$

In this notation the Lagrangian takes the form

$$\mathcal{L}(\Phi, \partial_\mu \Phi) = \frac{1}{4}\mathrm{Tr}\partial_\mu \Phi^\dagger \partial_\mu \Phi - \frac{m^2}{4}\mathrm{Tr}\Phi^\dagger \Phi - \frac{\lambda}{16}(\mathrm{Tr}\Phi^\dagger\Phi)^2, \tag{2.2.10}$$

and the $SU(2)_L$ symmetry again acts as $\Phi(x)' = g\Phi(x)$, except that $\Phi(x)$ is now a matrix. The $SU(2)_R$ symmetry, on the other hand, acts as

$$\Phi(x)' = \Phi(x) h^\dagger, \ h \in SU(2)_R. \tag{2.2.11}$$

Another equivalent notation sets up $\vec{\Phi} = (\Phi_{+1}, \Phi_{+2}, \Phi_{01}, \Phi_{02})$ as a 4-component vector. In this representation the Lagrangian is manifestly $O(4)$ invariant

$$\mathcal{L}(\vec{\Phi}, \partial_\mu \vec{\Phi}) = \frac{1}{2}\partial_\mu \vec{\Phi} \cdot \partial_\mu \vec{\Phi} - \frac{m^2}{2}\vec{\Phi} \cdot \vec{\Phi} - \frac{\lambda}{4}(\vec{\Phi} \cdot \vec{\Phi})^2. \tag{2.2.12}$$

Again, we must distinguish between the symmetric phase at $m^2 > 0$ and the broken phase at $m^2 < 0$. For $m^2 > 0$ there is a unique vacuum state

$$\Phi(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.2.13}$$

For $m^2 < 0$ the vacuum is degenerate and we make the choice

$$\Phi(x) = \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v \in \mathbb{R}_{>0}. \tag{2.2.14}$$

Again, we expand around the vacuum state

$$\Phi(x) = \begin{pmatrix} \pi_1(x) + i\pi_2(x) \\ v + \sigma(x) + i\pi_3(x) \end{pmatrix}, \tag{2.2.15}$$

and we obtain

$$\frac{1}{2}\partial_\mu \Phi^\dagger \partial_\mu \Phi = \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{2}\partial_\mu\pi_1\partial_\mu\pi_1 + \frac{1}{2}\partial_\mu\pi_2\partial_\mu\pi_2 + \frac{1}{2}\partial_\mu\pi_3\partial_\mu\pi_3, \tag{2.2.16}$$

and

$$\begin{aligned} V(\Phi) &= \frac{m^2}{2}((v+\sigma)^2 + \pi_1^2 + \pi_2^2 + \pi_3^2) + \frac{\lambda}{4}((v+\sigma)^2 + \pi_1^2 + \pi_2^2 + \pi_3^2)^2 \\ &= \frac{m^2}{2}(v^2 + 2v\sigma + \sigma^2 + \pi_1^2 + \pi_2^2 + \pi_3^2) \\ &+ \frac{\lambda}{4}(v^4 + 4v^3\sigma + 6v^2\sigma^2 + 2v^2(\pi_1^2 + \pi_2^2 + \pi_3^2)) + ... \\ &= -m^2\sigma^2 + ... \end{aligned} \tag{2.2.17}$$

Again, we find a σ particle of mass

$$m_\sigma^2 = -2m^2, \tag{2.2.18}$$

and in this case three massless Goldstone bosons π_1, π_2, π_3.

Now we want to promote the global $SU(2)_L$ symmetry to a local one, i.e. we demand invariance under

$$\Phi(x)' = g(x)\Phi(x). \tag{2.2.19}$$

The potential $V(\Phi)$ is trivially gauge invariant. The kinetic term, on the other hand, is not invariant, because

$$\partial_\mu\Phi(x)' = g(x)\partial_\mu\Phi(x) + \partial_\mu g(x)\Phi(x) = g(x)(\partial_\mu\Phi(x) + g^\dagger(x)\partial_\mu g(x)\Phi(x)). \tag{2.2.20}$$

Again, we want to compensate the additional term by a gauge field. We introduce

$$W_\mu(x)' = g(x)(W_\mu(x) + \frac{1}{g}\partial_\mu)g^\dagger(x), \tag{2.2.21}$$

where $W_\mu(x)$ is an anti-Hermitean matrix, that can be written as

$$W_\mu(x) = iW_\mu^a(x)\frac{\sigma_a}{2}, \ a = 1,2,3. \tag{2.2.22}$$

The gauge coupling strength g in the above expression should not be confused with the gauge transformation function $g(x)$. The covariant derivative takes the form

$$D_\mu \Phi(x) = (\partial_\mu + g W_\mu(x))\Phi(x), \tag{2.2.23}$$

and one finds

$$\begin{aligned} D_\mu \Phi(x)' &= (\partial_\mu + g W_\mu(x)')\Phi(x)' \\ &= g(x)[\partial_\mu \Phi(x) + g^\dagger(x)\partial_\mu g(x)\Phi(x) \\ &+ g W_\mu(x) g^\dagger(x) g(x)\Phi(x) + \partial_\mu g^\dagger(x) g(x)\Phi(x)] \\ &= g(x)(\partial_\mu + g W_\mu(x))\Phi(x) = g(x) D_\mu \Phi(x). \end{aligned} \tag{2.2.24}$$

As before, we introduce the gauge invariant Lagrange density

$$\mathcal{L}(\Phi, \partial_\mu \Phi, W_\mu) = \frac{1}{2} D_\mu \Phi^\dagger D_\mu \Phi - V(\Phi). \tag{2.2.25}$$

Here the gauge field is an external field. We still have to add the kinetic term. The field strength of a non-Abelian gauge field is given by

$$W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu + g[W_\mu, W_\nu], \tag{2.2.26}$$

and it transforms as

$$W'_{\mu\nu}(x) = g(x) W_{\mu\nu}(x) g^\dagger(x). \tag{2.2.27}$$

In analogy with Abelian gauge theory we write

$$\mathcal{L}(W_\mu) = -\frac{1}{4} W^a_{\mu\nu} W^a_{\mu\nu} = -\frac{1}{2} \mathrm{Tr} W_{\mu\nu} W_{\mu\nu}, \tag{2.2.28}$$

which is gauge invariant. In contrast to Abelian gauge fields, non-Abelian gauge fields are themselves charged, and hence interact even without other charged matter fields present.

Thus far, we have limited ourselves to the $SU(2)_L$ transformations with determinant 1. Now we want to discuss the extra $U(1)$ transformations related to the determinant. The scalar field then transforms as

$$\Phi(x)' = \exp(-i\frac{g'}{2}\varphi(x))\Phi(x). \tag{2.2.29}$$

Here g' is a new coupling constant — the weak hypercharge. Note that the factor $-\frac{1}{2}$ is purely conventional. The $U(1)_Y$ symmetry is in fact a subgroup of $SU(2)_R$ with

$$h(x) = \begin{pmatrix} \exp(-i\frac{g'}{2}\varphi(x)) & 0 \\ 0 & \exp(i\frac{g'}{2}\varphi(x)) \end{pmatrix}. \tag{2.2.30}$$

It should be emphasized again that only the $U(1)_Y$ subgroup and not the whole $SU(2)_R$ symmetry is gauged in the standard model. In fact, gauging only the $U(1)_Y$ subgroup leads to an explicit breaking of the global custodial $SU(2)_R$ symmetry. The $U(1)_Y$ gauge field transforms as

$$B_\mu(x)' = B_\mu(x) + \partial_\mu \varphi(x). \tag{2.2.31}$$

The new field contributes an additional term to the covariant derivative

$$\begin{aligned} D_\mu \Phi(x) &= (\partial_\mu + gW_\mu(x) + i\frac{g'}{2}B_\mu(x))\Phi(x) \\ &= (\partial_\mu + igW_\mu^a(x)\frac{\sigma_a}{2} + i\frac{g'}{2}B_\mu(x)) \left(\begin{array}{c} \Phi_+(x) \\ \Phi_0(x) \end{array} \right). \end{aligned} \tag{2.2.32}$$

Once again, we introduce a gauge invariant field strength

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.2.33}$$

and we obtain the total Lagrange density

$$\mathcal{L}(\Phi, W_\mu, B_\mu) = \frac{1}{2} D_\mu \Phi^\dagger D_\mu \Phi - V(\Phi) - \frac{1}{2}\mathrm{Tr} W_{\mu\nu} W_{\mu\nu} - \frac{1}{4} B_{\mu\nu} B_{\mu\nu}. \tag{2.2.34}$$

Let us consider the symmetry breaking case $m^2 < 0$, again in the unitary gauge

$$\Phi(x) = \left(\begin{array}{c} 0 \\ v \end{array} \right), \quad v \in \mathbb{R}_{>0}. \tag{2.2.35}$$

This vacuum state is still invariant under $U(1)_{em}$ gauge transformations of the type

$$\Phi(x)' = \left(\begin{array}{cc} \exp(ie\varphi(x)) & 0 \\ 0 & 1 \end{array} \right) \Phi(x), \tag{2.2.36}$$

which have a $U(1)_Y$ hypercharge part, but also a diagonal $SU(2)_L$ part. Hence, the vacuum state does not break the $SU(2)_L \otimes U(1)_Y$ symmetry completely. Instead, there is a remaining $U(1)_{em}$ symmetry, that we will soon identify with the one of electromagnetism. Since that symmetry remains unbroken, despite the Higgs mechanism, there will be one massless gauge boson — the photon. All other gauge bosons eat a Goldstone boson and become massive. To see this, we again consider the fluctuations in unitary gauge

$$\Phi(x) = \left(\begin{array}{c} 0 \\ v + \sigma(x) \end{array} \right). \tag{2.2.37}$$

Expanding in powers of the real field $\sigma(x)$, we obtain

$$\frac{1}{2}D_\mu\Phi^\dagger D_\mu\Phi = \frac{1}{2}|(\partial_\mu + igW_\mu^a\frac{\sigma_a}{2} + i\frac{g'}{2}B_\mu)\begin{pmatrix} 0 \\ v+\sigma \end{pmatrix}|^2$$
$$= \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{2}(v+\sigma)^2(01)(gW_\mu^a\frac{\sigma_a}{2} + \frac{g'}{2}B_\mu)(gW_\mu^a\frac{\sigma_a}{2} + \frac{g'}{2}B_\mu)\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$= \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{8}(v+\sigma)^2(g^2W_\mu^1W_\mu^1 + g^2W_\mu^2W_\mu^2$$
$$+\frac{1}{8}(gW_\mu^3 - g'B_\mu)(gW_\mu^3 - g'B_\mu)). \tag{2.2.38}$$

Also we have

$$V(\Phi) = \frac{m^2}{2}(v+\sigma)^2 + \frac{\lambda}{4}(v+\sigma)^4 = -m^2\sigma^2 + ..., \tag{2.2.39}$$

and hence there is a Higgs particle of mass

$$m_\sigma^2 = -2m^2. \tag{2.2.40}$$

Also, there are two W-bosons of mass

$$m_W = \frac{1}{2}gv. \tag{2.2.41}$$

Furthermore, we introduce the linear combination

$$Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}}, \tag{2.2.42}$$

which has the mass

$$m_Z = \frac{1}{2}\sqrt{g^2+g'^2}v. \tag{2.2.43}$$

The other linear independent combination

$$A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2+g'^2}}, \tag{2.2.44}$$

remains massless and describes the photon. We introduce the Weinberg angle θ_W via

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \tag{2.2.45}$$

such that

$$\frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_W, \ \frac{g'}{\sqrt{g^2+g'^2}} = \sin\theta_W, \tag{2.2.46}$$

and hence

$$\frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2 + g'^2}} = \cos\theta_W. \tag{2.2.47}$$

The experimental values for the masses are $m_W = 80$ GeV and $m_Z = 91$ GeV, such that $\sin^2\theta_W = 0.229$. The coupling constant of the photon is the charge e. On the other hand, the corresponding covariant derivative of the scalar field is

$$\begin{aligned}
D_\mu\Phi &= (\partial_\mu + igW_\mu^1\frac{\sigma_1}{2} + igW_\mu^2\frac{\sigma_2}{2} + igW_\mu^3\frac{\sigma_3}{2} + i\frac{g'}{2}B_\mu)\begin{pmatrix}\Phi_+\\ \Phi_0\end{pmatrix} \\
&= (\partial_\mu + igW_\mu^1\frac{\sigma_1}{2} + igW_\mu^2\frac{\sigma_2}{2} \\
&+ \frac{i}{2}\begin{pmatrix} gW_\mu^3 + g'B_\mu & 0 \\ 0 & -gW_\mu^3 + g'B_\mu \end{pmatrix})\begin{pmatrix}\Phi_+\\ \Phi_0\end{pmatrix} \\
&= (\partial_\mu + igW_\mu^1\frac{\sigma_1}{2} + igW_\mu^2\frac{\sigma_2}{2} \\
&+ i\begin{pmatrix} \frac{(g^2-g'^2)}{2\sqrt{g^2+g'^2}}Z_\mu + \frac{gg'}{\sqrt{g^2+g'^2}}A_\mu & 0 \\ 0 & \sqrt{g^2+g'^2}Z_\mu \end{pmatrix})\begin{pmatrix}\Phi_+\\ \Phi_0\end{pmatrix}.
\end{aligned} \tag{2.2.48}$$

We identify the electric charge as

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}}. \tag{2.2.49}$$

Now we see that indeed only Φ_+ couples to the electric field. It has charge e, while Φ_0 is neutral.

2.3 Triviality of the Standard Model

As we have seen, the Higgs sector of the standard model is a 4-component Φ^4 theory. The Lagrangian contains the dimensionful mass parameter m^2 as well as the dimensionless scalar self-coupling λ. These parameters determine the vacuum expectation value

$$v^2 = -m^2/\lambda \tag{2.3.1}$$

of the scalar field as well as the Higgs mass

$$m_\sigma^2 = -2m^2. \tag{2.3.2}$$

When gauge fields are introduced they pick up a mass proportional to v. For example, the W-boson mass is given by $m_W = \frac{1}{2} g v$ so that

$$\frac{m_\sigma^2}{m_W^2} = \frac{8\lambda}{g^2}. \tag{2.3.3}$$

Hence, a heavy Higgs particle requires a strongly coupled scalar field. We have obtained these results just using classical field theory. When the theory is quantized using perturbative methods, the bare parameters must be renormalized, but then again the same results emerge with the renormalized λ as a free parameter.

When Φ^4 theory is quantized beyond perturbation theory, on the other hand, an unexpected subtlety arises. In the lattice regularization there is overwhelming evidence (but no rigorous proof) that Φ^4 theory (and hence the standard model) is trivial. This means that the renormalized self-coupling λ approaches zero when one insists on completely removing the ultraviolet cut-off. In other words, the continuum limit of a lattice Φ^4 theory is just a free field theory. How can we then define the standard model as an interacting field theory beyond perturbation theory? Indeed, one should not insist on completely removing the ultraviolet cut-off. This means that the standard model cannot possibly make sense at arbitrarily high energies (beyond the finite cut-off). Hence, it is at best a low-energy effective theory and it must necessarily be replaced by something else at sufficiently high energies. In other words, the standard model could not even in principle be the theory of everything. This is actually a nice property of the standard model: it knows about its own limitations and tells us that it will eventually break down. Nontrivial theories (like e.g. QCD), on the other hand, remain interacting even when the cut-off is removed completely. These theories could, in principle, be valid at arbitrarily high energy scales.

The triviality of the standard model leads to an interesting upper bound on the Higgs boson mass. A heavy Higgs boson requires a large renormalized λ. On the other hand, when we remove the cut-off completely, we get a trivial free theory with $\lambda = 0$. Only when we leave the cut-off finite, we can get a heavy Higgs. On the other hand, the theory would clearly not make sense if the Higgs mass itself were larger than the ultraviolet cut-off. In the lattice regularization this puts an upper limit on the Higgs mass of around 500 GeV. Although this limit is not universal (it depends on the chosen regularization) the triviality bound suggests that the standard model Higgs should be found below 500 GeV or that the standard model is cut-off and replaced by some new physics at that energy scale. Indeed, some recent preliminary evidence for the Higgs indicated a mass around 115 GeV.

2.4 The Goldstone Theorem

In this chapter we have encountered a number of Goldstone bosons (which were then eaten by gauge bosons thus generating mass via the Higgs mechanism). Let us finally prove the Goldstone theorem that predicts the number of these massless particles for any given pattern of spontaneous breakdown of a continuous global symmetry.

Let us consider an N-component real scalar field $\vec{\Phi} = (\Phi_1, \Phi_2, ..., \Phi_N)$ with a self-interaction potential $V(\vec{\Phi})$ that is invariant under transformations of a symmetry group G. The n_G generators of the symmetry T^a ($a \in \{1, 2, ..., n_G\}$) are $N \times N$ matrices. An infinitesimal symmetry transformation of the field takes the form

$$\vec{\Phi}' = (1 + i\alpha_a T^a)\vec{\Phi}. \tag{2.4.1}$$

Now let us assume that the potential has a set of degenerate vacua. We pick one spontaneously, $\vec{\Phi} = \vec{v}$, and ask about the masses of excitations above this vacuum. First of all, since $\vec{v}$ is a minimum of the potential, we have

$$\frac{\partial V}{\partial \Phi_i}|_{\vec{\Phi}=\vec{v}} = 0. \tag{2.4.2}$$

The second derivatives of the potential define a mass squared matrix

$$\frac{\partial^2 V}{\partial \Phi_i \partial \Phi_j}|_{\vec{\Phi}=\vec{v}} = M^2_{ij}. \tag{2.4.3}$$

The eigenvalues of the matrix M^2 are the squared masses of the physical particle excitations above the vacuum $\vec{v}$. Now we assume that the vacuum is invariant under the transformations in a subgroup H that are generated by T^b with $b \in \{1, 2, ..., n_H\}$, i.e.

$$(1 + i\alpha_b T^b)\vec{v} = \vec{v} \;\Rightarrow\; T^b \vec{v} = 0. \tag{2.4.4}$$

Invariance of the potential implies

$$0 = V(\vec{\Phi}') - V(\vec{\Phi}) = \frac{\partial V}{\partial \Phi_i} i\alpha_a T^a_{ij} \Phi_j. \tag{2.4.5}$$

We differentiate this equation with respect to Φ_k and evaluate it at $\vec{\Phi} = \vec{v}$ to obtain

$$0 = \frac{\partial^2 V}{\partial \Phi_i \partial \Phi_k}|_{\vec{\Phi}=\vec{v}} i\alpha_a T^a_{ij} v_j + \frac{\partial V}{\partial \Phi_i}|_{\vec{\Phi}=\vec{v}} i\alpha_a T^a_{ik} \;\Rightarrow\; M^2_{ij}(T^a \vec{v})_j = 0. \tag{2.4.6}$$

For the unbroken subgroup (i.e. for $a = b$) this is trivially satisfied because $T^b \vec{v} = 0$. For the remaining generators with $a \in \{n_H+1, n_H+2, ..., n_G\}$, however, the equation implies that $T^a \vec{v}$ is an eigenvector of the matrix M^2 with eigenvalue zero. Hence, there are $n_G - n_H$ massless particle excitations above the vacuum $\vec{v}$.

It should be noted that massless Goldstone bosons can only exist in more than two space-time dimensions. Due to the Coleman-Hohenberg-Mermin-Wagner theorem spontaneous breakdown of a continuous global symmetry cannot happen in two dimensions because the Goldstone modes then pick up a mass nonperturbatively.

Since they are massless, the Goldstone bosons dominate the low-energy physics of any physical system with a spontaneously broken continuous global symmetry. There is a general low-energy effective Lagrangian technique — known as chiral perturbation theory — that describes the low-energy dynamics of the Goldstone bosons. Chiral perturbation theory was first developed for pions — the Goldstone bosons of the spontaneously broken chiral symmetry of QCD — but it is, in fact, generally applicable to any Goldstone boson. Let us illustrate this technique for the Goldstone bosons that arise in the Higgs sector of the standard model (before gauging the symmetry). As we have seen, the global symmetry of the Higgs sector is $G = SU(2)_L \otimes SU(2)_R = O(4)$ which then gets spontaneously broken down to a single $H = SU(2) = O(3)$ symmetry. In general, in chiral perturbation theory Goldstone bosons are described by fields in the coset space G/H — the manifold of the group G with points being identified if they are connected by a symmetry transformation in the unbroken subgroup H. The dimension of the coset space, $n_G - n_H$, corresponds to the number of Goldstone bosons predicted by Goldstone's theorem. In the Higgs sector of the standard model the coset space is

$$G/H = SU(2)_L \otimes SU(2)_R/SU(2) = SU(2) = S^3, \qquad (2.4.7)$$

and hence the Goldstone bosons are described by $SU(2)$ matrix fields, i.e. fields that live on the surface of a 3-dimensional sphere S^3. Denoting the Goldstone boson field by an $SU(2)$ matrix $U(x)$, one can write down a low-energy effective Lagrangian

$$\mathcal{L}(U) = \frac{\rho}{4} \text{Tr} \partial_\mu U^\dagger \partial_\mu U. \qquad (2.4.8)$$

This is the lowest order contribution to the chiral perturbation theory Lagrangian. Higher order contributions involve higher derivatives and thus contribute less than the leading term in the low-energy limit. The above Lagrangian is invariant against global $SU(2)_L \otimes SU(2)_R$ transformations

$$U(x)' = gU(x)h^\dagger. \qquad (2.4.9)$$

Note that while the Higgs sector of the standard model itself is a linear σ-model, its low-energy effective theory is a nonlinear one. It is also interesting to note that the same effective Lagrangian describes the low-energy pion dynamics in QCD with two massless quarks. Only the value of the coupling constant ρ is then very different. The chiral perturbation theory Lagrangian only depends on the symmetry and how it breaks spontaneously, and the chiral symmetry of two flavor QCD is again $SU(2)_L \otimes SU(2)_R$ spontaneously broken to $SU(2)$.

Chapter 3

Leptons and Quarks

In this chapter we add the fermions to the Lagrangian of the standard model. These fermions are leptons and quarks. The leptons participate only in the electroweak interactions while quarks participate both in the electroweak and in the strong interactions. At this point we have not added the gluons to the standard model Lagrangian. Hence, strictly speaking, at this level we only have electroweak but no strong interactions. We will add the gluons (as the last remaining fields of the standard model) in the next chapter. It is interesting that we need to add leptons and quarks at the same time because a variant of the standard model without quarks would be inconsistent. This is because the quarks cancel an anomaly of the leptons that would render a purely leptonic model non-renormalizable. Cancellation of anomalies in gauge symmetries is absolutely necessary both perturbatively and beyond perturbation theory. Anomalies in global symmetries, on the other hand, are a perfectly acceptable form of explicit symmetry breaking. In fact, they are necessary to correctly describe some of the physics. Until recently, the lepton fields of the standard model would have included only left-handed electrons and neutrinos as well as right-handed electrons, but no right-handed neutrinos. By now we know that neutrinos have a small non-zero mass which makes the inclusion of a right-handed neutrino field necessary. Still, we will follow our strategy of adding fields one by one, and so we will first work with massless neutrinos. Initially, we will also limit ourselves to a single generation of fermions. Once we add the two remaining generations, an interesting additional effect emerges — the explicit breaking of CP.

3.1 One Generation with Massless Neutrinos

The leptons of the first generation are electrons and their neutrinos. We start with left-handed neutrinos and right-handed anti-neutrinos only. Before we introduce right-handed neutrino fields, the neutrinos are massless, while the electrons will pick up a mass through spontaneous symmetry breaking. The left-handed neutrinos and the left-handed components of the electrons form an $SU(2)_L$ doublet

$$\begin{pmatrix} \nu_{eL}(x) \\ e_L(x) \end{pmatrix}, \tag{3.1.1}$$

while the right-handed component of the electron $e_R(x)$ is an $SU(2)$ singlet. The left-handed doublet is assigned a weak hypercharge $-g'/2$ and the right-handed singlet carries $-g'$. It is characteristic of a chiral gauge theory that the left- and right-handed fermion components carry different charges and thus transform differently under the gauge group. Before we introduce the electron mass term, the Lagrange function of the leptons of the first generation takes the form

$$\begin{aligned}\mathcal{L}(e_L, \nu_{eL}, e_R, W_\mu, B_\mu) &= (\bar{\nu}_{eL}\bar{e}_L) i\gamma_\mu(\partial_\mu + igW_\mu^a \frac{\sigma_a}{2} - i\frac{g'}{2}B_\mu) \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \\ &+ \bar{e}_R i\gamma_\mu(\partial_\mu - ig'B_\mu)e_R. \end{aligned} \tag{3.1.2}$$

As before, we introduce the Z-boson and photon fields

$$A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2 + g'^2}}, \ Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}}, \tag{3.1.3}$$

that are natural to consider after spontaneous symmetry breaking. Also, we introduce the charged W-boson fields

$$W_\mu^\pm = W_\mu^1 \pm iW_\mu^2. \tag{3.1.4}$$

Using

$$W_\mu^3 = \frac{g'A_\mu + gZ_\mu}{\sqrt{g^2 + g'^2}}, \ B_\mu = \frac{gA_\mu - g'Z_\mu}{\sqrt{g^2 + g'^2}}, \tag{3.1.5}$$

one can write the lepton-gauge coupling terms as

$$\begin{aligned} &\mathcal{L}(e_L, \nu_{eL}, e_R, A_\mu, Z_\mu) \\ = \ &(\bar{\nu}_{eL}\bar{e}_L) i\gamma_\mu[\partial_\mu + i \begin{pmatrix} \frac{\sqrt{g^2+g'^2}}{2} Z_\mu & gW_\mu^- \\ gW_\mu^+ & \frac{g'^2-g^2}{2\sqrt{g^2+g'^2}} Z_\mu - \frac{gg'}{\sqrt{g^2+g'^2}} A_\mu \end{pmatrix}] \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \\ + \ &\bar{e}_R i\gamma_\mu(\partial_\mu + i\frac{g'^2}{\sqrt{g^2 + g'^2}} Z_\mu - i\frac{gg'}{\sqrt{g^2 + g'^2}} A_\mu)e_R. \end{aligned} \tag{3.1.6}$$

Again, identifying

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}}, \tag{3.1.7}$$

as the electric charge, we see that indeed both the left- and right-handed electron carry the electric charge $-e$, while the neutrino is neutral (it does not couple to A_μ). Of course, the $U(1)_Y$ weak hypercharge assignments have been made to make exactly this happen. One finds that the electric charge Q, the weak hypercharge Y and the third component of weak isospin $T_3 = \frac{1}{2}\sigma_3$ are related by

$$Q = Y + T_3. \tag{3.1.8}$$

For the left-handed neutrino this equation takes the form

$$Q_{\nu_{eL}} = Y_{\nu_{eL}} + T_{3\nu_{eL}} = -\frac{1}{2} + \frac{1}{2} = 0, \tag{3.1.9}$$

for the left-handed electron it reads

$$Q_{e_L} = Y_{e_L} + T_{3e_L} = -\frac{1}{2} - \frac{1}{2} = -1, \tag{3.1.10}$$

and for the right-handed electron it takes the form

$$Q_{e_R} = Y_{e_R} + T_{3e_R} = -1 + 0 = -1. \tag{3.1.11}$$

Besides the couplings to the photon, we also have so-called weak neutral and charged current interactions. The weak neutral current couples to the Z-boson and is given by

$$j^0_\mu = \frac{\sqrt{g^2 + g'^2}}{2}\bar{\nu}_{eL}\gamma_\mu\nu_{eL} + \frac{g'^2 - g^2}{2\sqrt{g^2 + g'^2}}\bar{e}_L\gamma_\mu e_L + \frac{g'^2}{\sqrt{g^2 + g'^2}}\bar{e}_R\gamma_\mu e_R. \tag{3.1.12}$$

The charged currents, on the other hand, couple to the W-bosons and take the form

$$j^-_\mu = \frac{g}{2}\bar{\nu}_{eL}\gamma_\mu e_L, \; j^+_\mu = \frac{g}{2}\bar{e}_L\gamma_\mu\nu_{eL}. \tag{3.1.13}$$

The various neutral and charged current interactions give rise to a number of physical processes. For example, a W^- boson can decay into an electron and an anti-neutrino. Similarly, the Z-boson can decay into a neutrino and anti-neutrino or into an electron-positron pair. The observed width of the Z-boson has been used to constrain the number of generations (with neutrino masses lighter than one half of the Z-boson mass). As a result, we know that there are three generations. At this point we are more interested in the general structure of the theory than in the details of the resulting particle interactions. Hence, we will not calculate the Z-boson width at this point.

As it stands, the standard model with just electrons and electron-neutrinos is inconsistent because it suffers from anomalies in its gauge interactions. First, there is a triangle anomaly in the $U(1)_Y$ gauge interaction which manifests itself already in perturbation theory (but, of course, also beyond perturbation theory). The anomaly is proportional to the difference of $U(1)_Y$ charges cubed of the left- and right-handed sectors

$$A = \sum_L Y_L^3 - \sum_R Y_R^3 = Y_{\nu_{eL}}^3 + Y_{e_L}^3 - Y_{e_R}^3 = -\frac{1}{8} - \frac{1}{8} + 1 = \frac{3}{4} \neq 0. \qquad (3.1.14)$$

There is a similar quantity that results from a triangle diagram with one external $U(1)_Y$ boson and two external $SU(2)_L$ bosons

$$B = \sum_L (T_{3L})^2 Y_L - \sum_R (T_{3R})^2 Y_R = \frac{1}{4}\sum_L Y_L = -\frac{1}{4} \neq 0. \qquad (3.1.15)$$

Besides this, there is a so-called global anomaly in the $SU(2)_L$ gauge interactions, which manifests itself only beyond perturbation theory, i.e. it is not visible in a perturbative triangle diagram. The global anomaly has originally been discovered by Witten. It is related to the fact that the homotopy group $\Pi_4(SU(2)) = \Pi_4(S^3) = \mathbb{Z}(2)$ is non-trivial. We will discuss the topology of gauge fields later in more detail. The global topological structure of the gauge group is responsible for the name "global" anomaly. This should not be confused with anomalies that may arise in global (non-gauge) symmetries which need not to be canceled. Witten's global anomaly arises for theories with an odd number of $SU(2)_L$ doublets. Since, we have thus far introduced one doublet, we must add an odd number of doublets to cancel the global anomaly.

In the first generation there are also the up and down quarks which will come to our rescue and cancel both the triangle and the global anomaly. The quarks are massive and thus require the introduction of left- and right-handed fields. There is a controversy if the up-quark mass might be zero (which would solve the strong CP problem) but even then one would need to introduce both left- and right-handed up quark fields in order to achieve anomaly cancellation. The quarks participate in the strong interactions and thus carry an $SU(3)$ color charge. The color index on a quark field takes values $c \in \{1, 2, 3\}$. Sometimes one can read that anomaly cancellation in the standard model requires exactly three colors. However, we will see later that one could cancel all anomalies with any odd number of colors N_c. Of course, there is still enough experimental evidence that the quarks in our world have indeed three colors. The left-handed up and down quark fields form three $SU(2)_L$ doublets

$$\left(\begin{array}{c} u_L^1(x) \\ d_L^1(x) \end{array} \right), \left(\begin{array}{c} u_L^2(x) \\ d_L^2(x) \end{array} \right), \left(\begin{array}{c} u_L^3(x) \\ d_L^3(x) \end{array} \right), \tag{3.1.16}$$

or equivalently two $SU(3)_c$ color triplets. The right-handed quarks

$$u_R^1(x),\ u_R^2(x),\ u_R^3(x),\ d_R^1(x),\ d_R^2(x),\ d_R^3(x), \tag{3.1.17}$$

again form two $SU(3)_c$ color triplets, but they are $SU(2)_L$ singlets. Since there is the same number of left- and right-handed color triplets, the QCD part of the standard model is a non-chiral vector-like theory in which all $SU(3)_c$ anomalies cancel in a trivial way. We still need to assign the weak hypercharges of the various fields. As in the lepton sector, the electric charge will result from $Q = Y + T_3$. The up and down quarks have the electric charges

$$Q_u = \frac{2}{3},\ Q_d = -\frac{1}{3}. \tag{3.1.18}$$

As we will see later, two up quarks and one down quark form a proton with electric charge 1, while one up quark and two down quarks form a neutron with electric charge 0. Denoting the weak hypercharges of the $SU(2)_L$ doublets by Y_L we must hence obey the equations

$$\frac{2}{3} = Y_L + \frac{1}{2},\ -\frac{1}{3} = Y_L - \frac{1}{2} \Rightarrow\ Y_L = \frac{1}{6}. \tag{3.1.19}$$

Similarly, the weak hypercharges of the right-handed up and down quark result as

$$Y_{u_R} = \frac{2}{3},\ Y_{d_R} = -\frac{1}{3}. \tag{3.1.20}$$

At this point we can again calculate the triangle anomaly. The quarks contribute

$$A = \sum_{c\in\{1,2,3\}} [\sum_L Y_L^3 - \sum_R Y_R^3] = 3[2Y_L^3 - Y_{u_R}^3 - Y_{d_R}^3] = 3[\frac{2}{216} - \frac{8}{27} + \frac{1}{27}] = -\frac{3}{4}, \tag{3.1.21}$$

as well as

$$B = \frac{1}{4} \sum_{c\in\{1,2,3\}} \sum_L Y_L = \frac{1}{4} 3 \frac{2}{6} = \frac{1}{4}, \tag{3.1.22}$$

and thus cancel the anomalies arising in the lepton sector.

Sometimes it is argued that in the standard model the number of colors must be $N_c = 3$ in order to achieve anomaly cancellation. In contrast to this claim, we will now show that the standard model would indeed be consistent for any odd value of N_c. Even values of N_c are excluded by Witten's global anomaly. Of course, when $N_c \neq 3$ we should no longer expect the same values for the electric charges Q_u and Q_d. However, we still have

$$Q_u = Y_L + \frac{1}{2},\ Q_d = Y_L - \frac{1}{2} \ \Rightarrow \ Y_L = \frac{1}{2}(Q_u + Q_d), \tag{3.1.23}$$

as well as

$$Y_{u_R} = Q_u,\ Y_{d_R} = Q_d, \tag{3.1.24}$$

which implies

$$A = N_c[\frac{1}{4}(Q_u+Q_d)^3 - Q_u^3 - Q_d^3] = -\frac{3}{4} \ \Rightarrow \ (Q_u+Q_d)(Q_u-Q_d)^2 = \frac{1}{N_c}, \tag{3.1.25}$$

as well as

$$B = \frac{1}{4} \sum_{c\in\{1,\ldots,N_c\}} \sum_L Y_L = \frac{1}{4}N_c(Q_u + Q_d) = \frac{1}{4} \ \Rightarrow \ Q_u + Q_d = \frac{1}{N_c}. \tag{3.1.26}$$

Anomaly cancellation hence implies

$$Q_u = \frac{N_c + 1}{2N_c},\ Q_d = -\frac{N_c - 1}{2N_c}. \tag{3.1.27}$$

In a world with N_c colors, quarks are confined inside baryons which consist of N_c quarks bound together by gluons. The electric charge of a baryon consisting of N_c up quarks (the analog of a Δ isobar) then is

$$N_c Q_u = \frac{N_c + 1}{2}, \tag{3.1.28}$$

which indeed is an integer for odd N_c. When we replace one up quark by a down quark the charge changes by $Q_u - Q_d = 1$. When all up quarks are replaced by down quarks the charge of the resulting baryon is

$$N_c Q_d = -\frac{N_c - 1}{2}, \tag{3.1.29}$$

which is again an integer. It is interesting that the absence of fractional charges is a consequence not just of confinement but also of the cancellation of Witten's global anomaly.

It has been argued by 't Hooft that anomaly cancellation should take place even if one considers only the low-energy limit of a given theory. Anomalies must therefore be cancelled properly even in a low-energy effective theory for the standard model. This so-called anomaly matching condition puts nontrivial constraints on the possible low-energy dynamics. For example, at low energies quarks are confined inside protons and neutrons, and so the anomalies should also cancel between leptons and nucleons. To convince ourselves that this is indeed the case, let us consider the first generation now expressed in terms of low-energy (nucleon rather than quark) degrees of freedom

$$\left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right), e_R, \left(\begin{array}{c} p_L \\ n_L \end{array}\right), p_R, n_R. \tag{3.1.30}$$

Indeed, the global anomaly is still cancelled because the left-handed nucleons form one $SU(2)_L$ doublet. The weak hypercharge assignments for the nucleons are

$$Q_p = Y_L + \frac{1}{2} = 1, \; Q_n = Y_L - \frac{1}{2} = 0 \;\Rightarrow\; Y_L = \frac{1}{2}, \tag{3.1.31}$$

as well as

$$Y_{p_R} = Q_p = 1, \; Y_{n_R} = Q_n = 0. \tag{3.1.32}$$

Hence, one obtains

$$A = \sum_L Y_L^3 - \sum_R Y_R^3 = 2\frac{1}{8} - 1 = -\frac{3}{4}, \tag{3.1.33}$$

as well as

$$B = \frac{1}{4}\sum_L Y_L = \frac{1}{4}2\frac{1}{2} = \frac{1}{4}, \tag{3.1.34}$$

which indeed again cancels the anomalies of the leptons.

3.2 Electron and Quark Masses

So far we have not introduced any mass terms for the leptons. An electron mass term would have the form $m_e[\bar{e}_R e_L + \bar{e}_L e_R]$. Such a term is not gauge invariant, because the left- and right-handed components transform differently under $SU(2)_L$ and $U(1)_Y$. Hence, explicit mass terms are forbidden in chiral gauge theories. The scalar field Φ gave mass to the gauge bosons via spontaneous symmetry breaking. Similarly, it can give mass to fermions. Let us write down a Yukawa coupling

$$\mathcal{L}(\nu_{eL}, e_L, e_R, \Phi) = f_e[\bar{e}_R(\Phi_+^* \Phi_0^*) \left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right) + (\bar{\nu}_{eL}\bar{e}_L) \left(\begin{array}{c} \Phi_+ \\ \Phi_0 \end{array}\right) e_R]. \tag{3.2.1}$$

This Lagrangian is obviously $SU(2)_L$ gauge invariant. Interestingly, it is also $U(1)_Y$ invariant because

$$Y_{e_R} + Y_\Phi = -1 + \frac{1}{2} = -\frac{1}{2} = Y_{e_L}. \quad (3.2.2)$$

Again, using the vacuum state from before, we obtain

$$\begin{aligned} \mathcal{L}(e_L, \nu_{eL}, e_R, \Phi) &= f_e[\bar{e}_R(0v)\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} + (\bar{\nu}_{eL}\bar{e}_L)\begin{pmatrix} 0 \\ v \end{pmatrix} e_R] \\ &= f_e v[\bar{e}_R e_L + \bar{e}_L e_R]. \end{aligned} \quad (3.2.3)$$

Hence, the electron obtains the mass

$$m_e = f_e v, \quad (3.2.4)$$

while the neutrino indeed remains massless.

It should be noted that — like the scalar self-coupling λ — the Yukawa coupling f_e is an unpleasant thing to have in a fundamental theory. It is not a gauge coupling, and thus it is not highly constrained. In fact, we have just introduced another free parameter into the theory which determines the electron mass. The standard model itself does not make any predictions about this parameter. If we wanted to understand the value of the electron mass, we would clearly need to go beyond the standard model. In fact, at present nobody understands why the electron has its particular mass. Hence, we see again that the standard model contains more than just electroweak and strong interactions. Every Yukawa coupling parameterizes a fundamental force that is not often emphasized on the same level as the gauge forces. In fact, we don't really believe that the Yukawa coupling is truly fundamental. For example, it would be nice to replace it by some gauge force of a new kind. In this way, we would perhaps gain predictive power and finally understand the value of the electron mass. As we continue to add mass terms, the number of free parameters in the standard model will increase quickly. This tell us that the origin of mass is not at all understood. The often celebrated Higgs mechanism is in fact a big mess.

Since the down quark appears in the same position of an $SU(2)$ doublet as the electron and since again

$$Y_{d_R} + Y_\Phi = -\frac{1}{3} + \frac{1}{2} = \frac{1}{6} = Y_{d_L}, \quad (3.2.5)$$

we can give the down quarks a mass $m_d = f_d v$ by adding a term

$$\mathcal{L}(u_L, d_L, d_R, \Phi) = f_d[\bar{d}_R(\Phi_+^* \Phi_0^*)\begin{pmatrix} u_L \\ d_L \end{pmatrix} + (\bar{u}_L \bar{d}_L)\begin{pmatrix} \Phi_+ \\ \Phi_0 \end{pmatrix} d_R], \quad (3.2.6)$$

to the standard model Lagrangian. We cannot give mass to the up quark in the same way, just as we did not get a massive neutrino. This could make us happy, because a massless up quark would solve the strong CP problem. However, we will soon add two generations of heavier fermions, and the charm and top quarks, which replace the up quark in the second and third generation, clearly have a non-zero mass. In fact, it would be possible to construct a mass term for the up quark if we had a scalar field

$$\Phi'(x) = \begin{pmatrix} \Phi'_0(x) \\ \Phi'_-(x) \end{pmatrix}, \quad (3.2.7)$$

which would take the vacuum value

$$\Phi'(x) = \begin{pmatrix} v' \\ 0 \end{pmatrix}, \quad (3.2.8)$$

because then we could add a term

$$\mathcal{L}(u_L, d_L, u_R, \Phi') = f_u[\bar{u}_R(\Phi_0'^* \Phi_-'^*) \begin{pmatrix} u_L \\ d_L \end{pmatrix} + (\bar{u}_L \bar{d}_L) \begin{pmatrix} \Phi'_0 \\ \Phi'_- \end{pmatrix} u_R]. \quad (3.2.9)$$

To make this term gauge invariant, the weak hypercharge of the field Φ' must obey

$$Y_{u_R} + Y_{\Phi'} = \frac{2}{3} + Y_{\Phi'} = Y_L = \frac{1}{6} \ \Rightarrow \ Y_{\Phi'} = -\frac{1}{2}. \quad (3.2.10)$$

At this point, we could just add another Higgs field Φ' with the desired features. In fact, this is exactly what Peccei and Quinn have proposed in order to solve the strong CP problem, which we will discuss later. Here we follow the way things are done in the standard model. It may come as a surprise that a field Φ' can be constructed directly from the standard model Higgs field Φ as

$$\Phi'(x) = \begin{pmatrix} \Phi'_0(x) \\ \Phi'_-(x) \end{pmatrix} = \begin{pmatrix} -\Phi_0^*(x) \\ \Phi_+^*(x) \end{pmatrix}. \quad (3.2.11)$$

While it is clear that this field indeed has $Y_{\Phi'} = -\frac{1}{2}$, it may be less clear that it also transforms as an $SU(2)_L$ doublet. To see this, it is useful to return to the matrix form

$$\Phi(x) = \begin{pmatrix} \Phi_+(x) & -\Phi_0^*(x) \\ \Phi_0(x) & \Phi_+^*(x) \end{pmatrix}. \quad (3.2.12)$$

We have seen earlier that under $SU(2)_L$ transformations $g(x)$ the matrix $\Phi(x)$ transforms into $g(x)\Phi(x)$. Since the field $\Phi'(x)$ is nothing but the second column vector of the matrix $\Phi(x)$, it is clear that it indeed transforms as an $SU(2)_L$ doublet. The above construction implies $v' = -v$ and hence the up quark mass is given by $m_u = -f_u v$.

3.3 One Generation with Massive Neutrinos

When we can give mass to the up quark, why should we not give mass to the neutrino? In fact, the only thing that prevents us from doing this is that we have not even included a field for right-handed neutrinos. Of course, adding such a field might screw up the anomaly cancellation, so we have to be careful. Assume we would add an $SU(2)_L$ singlet field $\nu_{eR}(x)$ and then write a Yukawa coupling term

$$\mathcal{L}(\nu_{eL}, e_L, \nu_{eR}, \Phi') = f_{\nu_e}[\bar{\nu}_{eR}(\Phi_0'^* \Phi_-'^*) \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} + (\bar{\nu}_{eL} \bar{e}_L) \begin{pmatrix} \Phi_0' \\ \Phi_-' \end{pmatrix} \nu_{eR}]. \quad (3.3.1)$$

This term would be gauge invariant only if

$$Y_{\nu_{eR}} + Y_{\Phi'} = Y_{\nu_{eR}} - \frac{1}{2} = Y_{\nu_{eL}} = -\frac{1}{2} \Rightarrow Y_{\nu_{eR}} = 0. \quad (3.3.2)$$

On the one hand, this is quite pleasant because it means that a right-handed neutrino field is completely neutral under all standard model gauge forces. As a consequence, it would not contribute to any anomalies and anomaly cancellation would work as before. On the other hand, the interactions of a completely neutral particle are not very much constrained by the gauge principle. In fact, we will see that this leads to some more parameters in the standard model.

It has been argued that adding neutrino masses is really beyond the standard model. While this is clearly a matter of taste, I personally think that the standard model without right-handed neutrinos has always looked unnatural. In fact, with massless neutrinos the standard model has an exact global symmetry $B - L$ (baryon minus lepton number conservation). As I have argued before, exact symmetries should be locally realized, or alternatively, global symmetries should be only approximate. With right-handed neutrinos present, one can construct Majorana mass terms which indeed explicitly violate $B - L$.

A Majorana spinor can be constructed from the right-handed neutrino field as

$$\nu_e = \nu_{eR} + C\bar{\nu}_{eR}^T, \quad (3.3.3)$$

where C is the charge conjugation matrix. The charge conjugate of a Majorana neutrino is identical with itself

$$C\bar{\nu}_e^T = C\bar{\nu}_{eR}^T + \nu_{eR} = \nu_e, \quad (3.3.4)$$

and hence it is its own anti-particle. This is possible only for a neutral gauge singlet. A Majorana mass term takes the form $M\bar{\nu}_e\nu_e$ and changes lepton number by two. It is automatically gauge invariant and does not require the inclusion of the Higgs field. Consequently, the Majorana mass M is not tied to the electroweak scale v. In fact, M is a second dimensionful parameter that enters the standard model once we introduce right-handed neutrino fields. A renormalizable Majorana mass term cannot be constructed from the left-handed neutrino field because it would not be gauge invariant. However, as described above, we can construct Dirac mass terms that couple left- and right-handed neutrino fields through a Yukawa coupling to the Higgs field. Altogether we can write a neutrino mass matrix

$$(\bar{\nu}_{eL}\ C\nu_{eR}^T)\left(\begin{array}{cc} 0 & f_{\nu_e}v \\ f_{\nu_e}v & M \end{array}\right)\left(\begin{array}{c} C\bar{\nu}_{eL}^T \\ \nu_{eR} \end{array}\right). \quad (3.3.5)$$

For $M \ll f_{\nu_e}$ the eigenvalues of the mass matrix are

$$m_1 = M,\ m_2 = \frac{f_{\nu_e}^2 v^2}{M}, \quad (3.3.6)$$

i.e. there is a large mass m_1 and a much smaller mass m_2. When we will discuss grand unified theories (GUT) we will see that it is reasonable to assume that $M \ll f_{\nu_e}$. Then m_2 naturally gives light neutrinos. In GUT theories this is called the see-saw mechanism.

3.4 Three Generations and the CKM Matrix

Let us now add the remaining two generations of fermions. Let us first consider the case of massless neutrinos. Then we don't need to introduce fields for the right-handed neutrinos. For the first generation we have

$$\left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right), e_R, \left(\begin{array}{c} u_L'^c \\ d_L'^c \end{array}\right), u_R'^c, d_R'^c. \quad (3.4.1)$$

We have denoted the weak interaction eigenstates for the up and down quarks by u' and d'. Once we add the other generations, u' and d' mix with the other quarks to form the mass eigenstates u and d. In the second generation, we have muons (as heavier copies of the electrons) and their neutrinos, as well as charm and strange quarks (as the heavier copies of up and down quarks). The lepton and quark multiplets of the second generation then take the form

$$\left(\begin{array}{c} \nu_{\mu L} \\ \mu_L \end{array}\right), \mu_R, \left(\begin{array}{c} c_L'^c \\ s_L'^c \end{array}\right), c_R'^c, s_R'^c. \quad (3.4.2)$$

The charge assignments are exactly the same as in the first generation. Similarly, in the third generation we have tau leptons, their neutrinos, as well as top and bottom quarks

$$\left(\begin{array}{c} \nu_{\tau L} \\ \tau_L \end{array}\right), \tau_R, \left(\begin{array}{c} t'^c_L \\ b'^c_L \end{array}\right), t'^c_R, b'^c_R. \tag{3.4.3}$$

Up, charm, and top quarks are indistinguishable from the point of view of the electroweak and strong interactions, and they can therefore mix together. The down, strange and bottom quarks can do the same thing. Mixing between up and strange quarks, on the other hand, is forbidden because they sit in different positions of $SU(2)_L$ doublets. After spontaneous symmetry breaking, the most general quark mass term takes the form

$$(\bar{d}'_L \ \bar{s}'_L \ \bar{b}'_L) M^D \left(\begin{array}{c} d'_R \\ s'_R \\ b'_R \end{array}\right) + (\bar{u}'_L \ \bar{c}'_L \ \bar{t}'_L) M^U \left(\begin{array}{c} u'_R \\ c'_R \\ t'_R \end{array}\right). \tag{3.4.4}$$

The mass matrices M^D and M^U are general complex 3×3 matrices whose elements are products of Yukawa couplings and the vacuum value v of the Higgs field. A general complex matrix can be diagonalized by a bi-unitary transformation

$$U_L^{D\dagger} M^D U_R^D = \mathrm{diag}(m_d, m_s, m_b), \ U_L^{U\dagger} M^U U_R^U = \mathrm{diag}(m_u, m_c, m_t), \tag{3.4.5}$$

where the quark masses $m_u, m_d, ..., m_t$ are real and positive. We should keep in mind that the quarks are confined. Hence, their mass parameters do not appear as eigenvalues of the standard model Hamiltonian. The bi-unitary transformations relate the weak interaction eigenstates to the mass eigenstates

$$\left(\begin{array}{c} d'_L \\ s'_L \\ b'_L \end{array}\right) = U_L^D \left(\begin{array}{c} d_L \\ s_L \\ b_L \end{array}\right), \ \left(\begin{array}{c} d'_R \\ s'_R \\ b'_R \end{array}\right) = U_R^D \left(\begin{array}{c} d_R \\ s_R \\ b_R \end{array}\right),$$
$$\left(\begin{array}{c} u'_L \\ c'_L \\ t'_L \end{array}\right) = U_L^U \left(\begin{array}{c} u_L \\ c_L \\ t_L \end{array}\right), \ \left(\begin{array}{c} u'_R \\ c'_R \\ t'_R \end{array}\right) = U_R^U \left(\begin{array}{c} u_R \\ c_R \\ t_R \end{array}\right). \tag{3.4.6}$$

Let us now express the weak interaction currents in the basis of mass eigenstates. The quark neutral current contains terms such as $\bar{u}'_L \gamma_\mu u'_L + \bar{c}'_L \gamma_\mu c'_L + \bar{t}'_L \gamma_\mu t'_L$ or $\bar{d}'_R \gamma_\mu d'_R + \bar{s}'_R \gamma_\mu s'_R + \bar{b}'_R \gamma_\mu b'_R$. When one rotates these terms into the basis of mass eigenstates one can simply drop the primes, because $U_L^{U\dagger} U_L^U = \mathbb{1}$, $U_R^{D\dagger} U_R^D = \mathbb{1}$. Hence, the neutral current interactions do not lead to changes among different quark flavors. One says that the standard model is free of flavor-changing neutral currents. This characteristic of the standard model has been verified with high precision in numerous experiments. The charged current, on the other hand, takes the form

$$\begin{aligned} j_\mu^- &= \bar{u}'_L \gamma_\mu d'_L + \bar{c}'_L \gamma_\mu s'_L + \bar{t}'_L \gamma_\mu b'_L = (\bar{u}_L \ \bar{c}_L \ \bar{t}_L) \gamma_\mu U_L^{U\dagger} U_L^D \left(\begin{array}{c} d_L \\ s_L \\ b_L \end{array} \right) \\ &= (\bar{u}_L \ \bar{c}_L \ \bar{t}_L) \gamma_\mu V \left(\begin{array}{c} d_L \\ s_L \\ b_L \end{array} \right). \end{aligned} \tag{3.4.7}$$

We have introduced the Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix

$$V = U_L^{U\dagger} U_L^D. \tag{3.4.8}$$

This matrix describes the amount of flavor-changing in the charged current interactions in the standard model.

Let us count the number of physical parameters in the mixing matrix V for a given number of generations N. Since V is unitary, one would naively expect N^2 parameters. However, one can change the matrices U_L^U and U_L^D to

$$U_L'^U = U_L^U D^U, \ U_L'^D = U_L^D D^D, \tag{3.4.9}$$

by multiplying them with diagonal unitary matrices D^U and D^D from the right. This still leaves the resulting mass matrices diagonal, and turns V into

$$V' = U_L'^{U\dagger} U_L'^D = D^{U\dagger} V D^D. \tag{3.4.10}$$

The matrices D^U and D^D have $2N$ parameters (the complex phases on their diagonals) which should not be counted as physical parameters in V. However, an overall phase common to both D^U and D^D would not even affect V', and hence the correct counting of physical parameters in V' is

$$N^2 - 2N + 1 = (N-1)^2. \tag{3.4.11}$$

With a single generation there is no mixing and hence no free parameter. With two generations there is one physical parameter — the so-called Cabibbo angle θ_C. A general unitary 2×2 matrix could be written as

$$V = \left(\begin{array}{cc} A e^{i\varphi} & B e^{i\varphi} \\ -B^* & A^* \end{array} \right) = \left(\begin{array}{cc} |A| e^{i\alpha} e^{i\varphi} & |B| e^{i\beta} e^{i\varphi} \\ -|B| e^{-i\beta} & |A| e^{-i\alpha} \end{array} \right), \tag{3.4.12}$$

where A and B are arbitrary complex numbers with $|A|^2 + |B|^2 = 1$. We can now use the freedom to introduce D^U and D^D in order to change any given V to

$$V' = D^{U\dagger} V D^D = \left(\begin{array}{cc} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{array} \right). \tag{3.4.13}$$

Choosing

$$D^U = \mathrm{diag}(e^{i(\varphi+\beta)}, e^{-i\alpha}),\ D^D = \mathrm{diag}(e^{i(\beta-\alpha)}, 1), \tag{3.4.14}$$

one indeed turns the general $U(2)$ matrix V (which seems to have four parameters) into the special form V' which only depends on the Cabibbo angle and indeed belongs to $SO(2)$.

As we have seen, for a general number of generations N the number of physical parameters in the matrix V is $(N-1)^2$. For $N > 2$ the matrix V' will in general not belong to $SO(N)$ (which has only $N(N-1)/2$ parameters) because

$$(N-1)^2 - \frac{N(N-1)}{2} = \frac{(N-1)(N-2)}{2} > 0. \tag{3.4.15}$$

For example, for the physical case of $N = 3$ generations, the CKM matrix contains $(N-1)(N-2)/2 = 1$ complex phase in addition to $N(N-1)/2 = 3$ Cabbibo-type Euler angles. The complex phase is a source of explicit CP symmetry breaking which manifests itself, for example, in the neutral kaon system. In the quark sector alone the Yukawa couplings give rise to ten free parameters of the standard model — the six quark masses, three mixing angles, and one complex CP violating phase.

As long as we consider massless neutrinos, we need not worry about lepton mixing. The lepton analog of the CKM matrix would be given by

$$W = U_L^{\nu\dagger} U_L^E. \tag{3.4.16}$$

However, if all neutrinos are massless one can replace U_L^ν by $U_L'^\nu = U_L^\nu D^\nu$ — now with a general unitary (and not necessarily diagonal) matrix D^ν — and still keep the neutrino mass matrix unchanged. Choosing $D^\nu = U_L^{\nu\dagger} U_L^E$ one simply obtains $W = \mathbb{1}$. Once we introduce neutrino mass terms, we will have an analog of the CKM matrix in the lepton sector. Without Majorana mass terms constructed from the right-handed neutrinos, there would simply be $(N-1)^2$ additional lepton mixing parameters. With Majorana mass terms present, the situation is more complicated.

Chapter 4

The Strong Interactions

In this chapter we add the gluons as the last ingredient to the standard model. Without the gluons present, we had created a world of Higgs particles, W and Z bosons, photons, as well as charged leptons and quarks. In particular, in this world there would be particles with fractional electric charges (the quarks). Single quarks have never been observed despite numerous experimental efforts (people have even looked inside oyster shells). In fact, the strong interactions are so strong that quarks are permanently confined. The confining force is mediated by the gluons, whose presence thus totally changes the low-energy physics. As a consequence of confinement colored quarks and gluons form color-neutral hadrons which have integer electric charges. Hadrons are baryons (three quark states), mesons (quark-anti-quark states), or more exotic creatures like glueballs. Confinement is a complicated dynamical phenomenon that is presently only poorly understood. We are far from a satisfactory quantitative understanding of the properties of hadrons. Fortunately, when we want to understand hadrons, we need not consider the entire standard model. At low energies, the strong interactions are much stronger than the electroweak forces which can be neglected to a first approximation. Still, leptons can be used as electromagnetic or weak probes to investigate the complex interior of hadrons. The part of the standard model that is most relevant at low energies is just quantum chromodynamics (QCD), the $SU(3)_c$ gauge theory of quarks and gluons. The original Yukawa couplings of the standard model enter QCD in the form of quark mass parameters. However, due to confinement, these parameters no longer represent the masses of physical observable particles. The light quarks up, down and strange have mass terms below the typical QCD energy scale $\Lambda_{QCD} \approx 0.2$ GeV, while the quarks charm, bottom and top are much heavier. These quarks do not play a role in QCD at low energies.

In general, quark mass terms do not play a dominant role in the QCD dynamics. In particular, the masses of hadrons are not at all the sum of the masses of the quarks within them. Even with exactly massless quarks, due to confinement the hadrons (except for the Goldstone bosons among them) would still have masses of the order of Λ_{QCD}. The strong binding energy of quarks and gluons manifests itself as the mass of hadrons. Often one can read that the origin of mass in the universe is the Higgs mechanism, and indeed we have seen that the quark masses would be zero if the Higgs potential would not have the Mexican hat shape. However, the dominant contribution to the mass of the matter that surrounds us is due to protons and neutrons, and thus due to QCD binding energy.

Despite the fact that quarks do not exist as free particles, there is a lot of indirect experimental evidence for quarks, thanks to another fundamental property of the strong interactions. At high energies QCD is asymptotically free, i.e. quarks and gluons behave more like free particles, which can be observed in deep inelastic lepton-hadron scattering processes. The high-energy physics of QCD is accessible to perturbative calculations. Lattice gauge theory allows us to perform nonperturbative QCD calculations from first principles. In practice, these calculations require a very large computational effort and suffer from numerous technical problems. Still, there is little doubt that QCD will eventually be solved quantitatively using lattice methods. Even without deriving hadron properties using lattice methods, one can deduce some aspects just by using group theoretical arguments.

When the mass terms of the light u and d quarks are neglected, the QCD Lagrangian has a global $SU(2)_L \otimes SU(2)_R$ chiral symmetry. Hence, one would at first expect corresponding degeneracies in the hadron spectrum. Since this is not what is actually observed, one concludes that chiral symmetry is spontaneously broken. After spontaneous symmetry breaking only a $SU(2)_{L=R}$ symmetry remains intact. When a global, continuous symmetry breaks spontaneously, the Goldstone phenomenon gives rise to a number of massless particles. In QCD these Goldstone bosons are the three pions π^+, π^0 and π^-. Due to the small but nonzero quark masses chiral symmetry is also explicitly broken, and the pions are not exactly massless. Chiral symmetry leads to interesting predictions about the low energy dynamics of QCD. A systematic method to investigate this is provided by chiral perturbation theory, which is based on low energy pion effective Lagrangians.

4.1 Quantum Chromodynamics

We introduce the gluons via an algebra-valued gauge potential

$$G_\mu(x) = iG_\mu^a(x)\lambda_a, \ a \in \{1, 2, ..., 8\}. \tag{4.1.1}$$

The gluon field strength is

$$G_{\mu\nu}(x) = \partial_\mu G_\nu(x) - \partial_\nu G_\mu(x) + g_s[G_\mu(x), G_\nu(x)]. \tag{4.1.2}$$

where g_s is the dimensionless gauge coupling constant of the strong interactions. We postulate the usual behavior under gauge transformations

$$G'_\mu(x) = g(x)(G_\mu(x) + \frac{1}{g_s}\partial_\mu)g^+(x). \tag{4.1.3}$$

In contrast to an Abelian gauge theory the field strength is not gauge invariant. It transforms as

$$G'_{\mu\nu}(x) = g(x)G_{\mu\nu}(x)g^+(x). \tag{4.1.4}$$

The QCD Lagrange function takes the gauge invariant form

$$\begin{aligned}\mathcal{L}_{QCD}(\bar{\Psi}, \Psi, G_\mu) &= \sum_f \bar{\Psi}_f(x)(i\gamma^\mu(\partial_\mu + g_s G_\mu(x)) - m_f)\Psi_f(x) \\ &- \frac{1}{4}\text{Tr}G^{\mu\nu(x)}G_{\mu\nu(x)}. \end{aligned} \tag{4.1.5}$$

The quark field $\Psi_f = \Psi_{fL} + \Psi_{fR}$ with the flavor index $f \in \{u, d, s\}$ is just a collection of the quark fields we had already introduced in the previous chapter. For example, $\Psi_u = u_L + u_R$.

An important difference between Abelian and non-Abelian gauge theories is that in a non-Abelian gauge theory the gauge fields are themselves charged. The non-Abelian charge of the gluons leads to a self interaction, that is not present for the Abelian photons. The interaction results from the commutator term in the gluon field strength. It gives rise to three and four gluon vertices in the QCD Feynman rules. We will not derive the QCD Feynman rules here, we discuss them only qualitatively. The terms in the Lagrange function that are quadratic in G_μ give rise to the free gluon propagator. Due to the commutator term, however, there are also terms cubic and quartic in G_μ, that lead to the gluon self interaction vertices. Correspondingly, there is a free quark propagator and a quark-gluon vertex. The perturbative quantization of a non-Abelian gauge theory requires to fix the gauge. In the Landau gauge $\partial^\mu G_\mu = 0$ this leads to so-called ghost fields, which are scalars, but still anticommute. Correspondingly, there is a ghost propagator and a ghost-gluon vertex. In QCD the ghost fields are also color octets. They are only a mathematical tool arising in the loops of a Feynman diagram, not in external legs. Strictly speaking one could say the same about quarks and gluons, because they also cannot exist as asymptotic states.

The objects in the classical QCD Lagrange function do not directly correspond to observable quantities. Both fields and coupling constants get renormalized. In particular, the formal expression

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi\mathcal{D}G \exp(-i \int d^4x \; \mathcal{L}_{QCD}(\bar{\Psi}, \Psi, G_\mu)) \tag{4.1.6}$$

for the QCD path integral is undefined, i.e. divergent, until it is regularized and appropriately renormalized. In gauge theories it is essential that gauge invariance is maintained in the regularized theory. A regularization scheme that allows nonperturbative calculations defines the path integral on a space-time lattice with spacing ε. The renormalization of the theory corresponds to performing the continuum limit $\varepsilon \to 0$ in a controlled way, such that ratios of particle masses — i.e. the physics — remains constant. A perturbative regularization scheme works with single Feynman diagrams. The loop integrations in the corresponding mathematical expressions can be divergent in four dimensions. In dimensional regularization one works in d dimensions (by analytic continuation in d) and one performs the limit $\varepsilon = 4 - d \to 0$ again such that the physics remains constant. To absorb the divergences, quark and gluon fields are renormalized

$$\Psi(x) = Z_\Psi(\varepsilon)^{1/2}\Psi^R(x), \; G_\mu(x) = Z_G(\varepsilon)^{1/2}G_\mu^R(x), \tag{4.1.7}$$

and also the coupling constant is renormalized via

$$g_s = \frac{Z(\varepsilon)}{Z_\Psi(\varepsilon)Z_G(\varepsilon)^{1/2}} g_s^R. \tag{4.1.8}$$

Here the unrenormalized quantities as well as the Z-factors are divergent, but the renormalized quantities are finite in the limit $\varepsilon \to 0$. Correspondingly, one renormalizes the n-point Green's functions and the resulting vertex functions

$$\Gamma^R_{n_\Psi,n_G}(k_i, p_j) = \lim_{\varepsilon \to 0} Z_\Psi(\varepsilon)^{n_\Psi/2} Z_G(\varepsilon)^{n_G/2} \Gamma_{n_\Psi,n_G}(k_i, p_j, \varepsilon). \tag{4.1.9}$$

Demanding convergence of the renormalized vertex function fixes the divergent part of the Z-factors. To fix the finite part as well one must specify renormalization conditions. In QCD this can be done using the vertex functions $\Gamma_{0,2}$, $\Gamma_{2,0}$ and $\Gamma_{2,1}$, i.e. the inverse gluon and quark propagators and the quark-gluon vertex.

As opposed to QED, where mass and charge of the electron are directly observable, in QCD one chooses an arbitrary scale $\mathcal{M}$ to formulate the renormalization conditions

$$\begin{aligned}
&\Gamma^R_{0,2}(p,-p)^{\mu\nu}_{ab}|_{p^2=-\mathcal{M}^2} = i(-g_{\mu\nu}p^2 + p^\mu p^\nu)\delta_{ab}, \\
&\Gamma^R_{2,0}(k,k)|_{k^2=-\mathcal{M}^2} = i\gamma^\mu k_\mu, \\
&\Gamma^R_{2,1}(k,k,k)^\mu_a|_{k^2=-\mathcal{M}^2} = -ig_s^R \frac{\lambda_a}{2}\gamma^\mu.
\end{aligned} \tag{4.1.10}$$

The renormalized vertex functions are functions of the renormalized coupling constant g_s^R and of the renormalization scale $\mathcal{M}$, while the unrenormalized vertex functions depend on the bare coupling g_s and on the regularization parameter ε (the cut-off). Hence, there is a hidden relation

$$g_s^R = g_s^R(g,\varepsilon,\mathcal{M}). \tag{4.1.11}$$

This relation defines the β-function

$$\beta(g_s^R) = \lim_{\varepsilon\to 0} \mathcal{M}\frac{\partial}{\partial\mathcal{M}} g_s^R(g,\varepsilon,\mathcal{M}). \tag{4.1.12}$$

The β-function can be computed in QCD perturbation theory. To leading order in the coupling constant one obtains

$$\beta(g_s^R) = -\frac{(g_s^R)^3}{16\pi^2}(11 - \frac{2}{3}N_f). \tag{4.1.13}$$

Here N_f is the number of quark flavors. Fixed points g_s^* of the renormalization group are of special interest. They are invariant under a change of the arbitrarily chosen renormalization scale $\mathcal{M}$, and hence they correspond to zeros of the β-function. In QCD there is a single fixed point at $g_s^* = 0$. For

$$11 - \frac{2}{3}N_f > 0 \;\Rightarrow\; N_f \leq 16, \tag{4.1.14}$$

i.e. for not too many flavors, the β-function is negative close to the fixed point. This behavior is known as asymptotic freedom. It is typical for non-Abelian gauge theories in four dimensions, as long as there are not too many fermions or scalars. Asymptotic freedom is due to the self interaction of the gauge field, that is not present in an Abelian theory. We now use

$$\begin{aligned}
&\beta(g_s^R) = \mathcal{M}\frac{\partial}{\partial\mathcal{M}} g_s^R = -\frac{(g_s^R)^3}{16\pi^2}(11 - \frac{2}{3}N_f) \;\Rightarrow \\
&\frac{\partial g_s^R}{\partial\mathcal{M}}/(g_s^R)^3 = \frac{1}{2}\frac{\partial (g_s^R)^2}{\partial\mathcal{M}}/(g_s^R)^4 = -\frac{11-\frac{2}{3}N_f}{16\pi^2}\frac{1}{\mathcal{M}} \;\Rightarrow \\
&\frac{\partial (g_s^R)^2}{(g_s^R)^4} = -\frac{33-2N_f}{24\pi^2}\frac{\partial\mathcal{M}}{\mathcal{M}} \;\Rightarrow\; \frac{1}{(g_s^R)^2} = \frac{33-2N_f}{24\pi^2}\log\frac{\mathcal{M}}{\Lambda_{QCD}}.
\end{aligned} \tag{4.1.15}$$

Here Λ_{QCD} is an integration constant. Introducing the renormalized strong fine structure constant

$$\alpha_s^R = \frac{(g_s^R)^2}{4\pi}, \tag{4.1.16}$$

we obtain

$$\alpha_s^R(\mathcal{M}) = \frac{6\pi}{33 - 2N_f} \frac{1}{\log(\mathcal{M}/\Lambda_{QCD})}. \tag{4.1.17}$$

At high energy scales $\mathcal{M}$ the renormalized coupling constant slowly (i.e. logarithmically) goes to zero. Hence the quarks then behave like free particles.

The classical Lagrange function for QCD with massless fermions has no dimensionful parameter. Hence the classical theory is scale invariant, i.e. to each solution with energy E correspond other solutions with scaled energy λE for any arbitrary scale parameter λ. Scale invariance, however, is anomalous. It does not survive the quantization of the theory. This explains why there is a proton with a very specific mass $E = M_p$, but no scaled version of it with mass λM_p. We now understand better why this is the case. In the process of quantization the dimensionful scale $\mathcal{M}$ (and related to this Λ_{QCD}) emerged, leading to an explicit breaking of the scale invariance of the classical theory. Scale transformations are therefore no symmetry of QCD.

4.2 Isospin Symmetry

Proton and neutron have almost the same masses

$$M_p = 0.938 \text{ GeV}, \; M_n = 0.940 \text{ GeV}. \tag{4.2.1}$$

While the proton seems to be absolutely stable, a free neutron decays radioactively into a proton, an electron and an electron-anti-neutrino $n \rightarrow p + e + \bar{\nu}_e$. Protons and neutrons (the nucleons) are the constituents of atomic nuclei. Originally, Yukawa postulated a light particle mediating the interaction between protons and neutrons. This π-meson or pion is a boson with spin 0, which exist in three charge states π^+, π^0 and π^-. The corresponding masses are

$$M_{\pi^+} = M_{\pi^-} = 0.140 \text{ GeV}, \; M_{\pi^0} = 0.135 \text{ GeV}. \tag{4.2.2}$$

In pion-nucleon scattering a resonance occurs in the total cross section as a function of the pion-nucleon center of mass energy. The resonance energy is interpreted as the mass of an unstable particle — the so-called Δ-isobar. One may

Hadron	Representation	I	I_3	Q	S
p, n	{2}	$\frac{1}{2}$	$\frac{1}{2}, -\frac{1}{2}$	1, 0	$\frac{1}{2}$
Δ^{++}, Δ^{+}, Δ^{0}, Δ^{-}	{4}	$\frac{3}{2}$	$\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$	2, 1, 0, -1	$\frac{3}{2}$
π^{+}, π^{0}, π^{-}	{3}	1	1, 0, -1	1, 0, -1	0
ρ^{+}, ρ^{0}, ρ^{-}	{3}	1	1, 0, -1	1, 0, -1	1

Table 4.1: *The isospin classification of hadrons.*

view the Δ-particle as an excited state of the nucleon. It exists in four charge states Δ^{++}, Δ^{+}, Δ^{0} and Δ^{-} with masses

$$M_{\Delta^{++}} \approx M_{\Delta^{+}} \approx M_{\Delta^{0}} \approx M_{\Delta^{-}} \approx 1.232 \text{ GeV} \tag{4.2.3}$$

Similar to pion-nucleon scattering there is also a resonance in pion-pion scattering. This so-called ρ-meson comes in three charge states ρ^{+}, ρ^{0} and ρ^{-} with masses

$$M_{\rho^{+}} \approx M_{\rho^{0}} \approx M_{\rho^{-}} \approx 0.768 \text{ GeV}. \tag{4.2.4}$$

Particles with different electric charges have (almost) degenerate masses, and it is natural to associate this with an (approximate) symmetry. This so-called isospin symmetry is similar to the ordinary spin $SU(2)$ rotational symmetry. Isospin is, however, not related to space-time transformations, it is an intrinsic symmetry. As we know each total spin $S = 0, 1/2, 1, 3/2, \ldots$ is associated with an irreducible representation of the $SU(2)_S$ rotation group containing $2S+1$ states distinguished by their spin projection

$$S_z = -S, -S+1, \ldots, S-1, S. \tag{4.2.5}$$

In complete analogy the representations of the $SU(2)_I$ isospin symmetry group are characterized by their total isospin $I = 0, 1/2, 1, 3/2, \ldots$. The states of an isospin representation are distinguished by their isospin projection

$$I_3 = -I, -I+1, \ldots, I-1, I. \tag{4.2.6}$$

A representation with isospin I contains $2I+1$ states and is denoted by $\{2I+1\}$. We can classify the hadrons by their isospin. This is done in table 4.1. For the baryons (nucleon and Δ) isospin projection and electric charge are related by $Q = I_3 + \frac{1}{2}$, and for the mesons (π and ρ) $Q = I_3$.

Isospin is an (approximate) symmetry of the strong interactions. For example, the proton-pion scattering reaction $p+\pi \rightarrow \Delta$ is consistent with isospin symmetry because the coupling of the isospin representations of nucleon and pion

$$\{2\} \otimes \{3\} = \{2\} \oplus \{4\} \tag{4.2.7}$$

does indeed contain the quadruplet isospin 3/2 representation of the Δ-isobar. The isospin symmetry of the hadron spectrum indicates that the strong interactions are charge independent. This is no surprise because the charge Q is responsible for the electromagnetic but not for the strong interactions.

4.3 Nucleon and Δ-Isobar in the Quark Model

We want to approach the question of the hadronic constituents by investigating various symmetries. First we consider isospin. Since the hadrons form isospin multiplets the same should be true for their constituents. The only $SU(2)$ representation from which we can generate all others is the fundamental representation — the isospin doublet $\{2\}$ with $I = 1/2$ and $I_3 = \pm 1/2$. We identify the two states of this multiplet with the constituent quarks up ($I_3 = 1/2$) and down ($I_3 = -1/2$). A constituent quark is a quasiparticle carrying the same quantum numbers as a fundamental (current) quark, but also containing numerous gluons. After all, a constituent quark is not a very well defined object. We can view it as a basic building block for hadrons that plays a role in some simple phenomenological models for the strong interactions. Still, the concept of constituent quarks leads to a very successful group theoretical classification scheme for hadrons.

Since the Δ-isobar has isospin 3/2 it contains at least three constituent quarks. We couple

$$\{2\} \otimes \{2\} \otimes \{2\} = (\{1\} \oplus \{3\}) \otimes \{2\} = \{2\} \oplus \{2\} \oplus \{4\}, \tag{4.3.1}$$

and we do indeed find a quadruplet. For the charges of the baryons we found

$$Q = I_3 + \frac{1}{2} = \sum_{q=1}^{3} (I_{3q} + \frac{1}{6}) = \sum_{q=1}^{3} Q_q, \tag{4.3.2}$$

and hence we obtain for the charges of the quarks

$$Q_q = I_{3q} + \frac{1}{6},\ Q_u = \frac{1}{2} + \frac{1}{6} = \frac{2}{3},\ Q_d = -\frac{1}{2} + \frac{1}{6} = -\frac{1}{3}. \tag{4.3.3}$$

The quarks have fractional electric charges. Using Clebsch-Gordon coefficients of

$SU(2)$ one finds

$$
\begin{aligned}
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,3/2} &= uuu \equiv \Delta^{++},\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,1/2} &= \frac{1}{\sqrt{3}}(uud+udu+duu) \equiv \Delta^{+},\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-1/2} &= \frac{1}{\sqrt{3}}(udd+dud+ddu) \equiv \Delta^{0},\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-3/2} &= ddd \equiv \Delta^{-}.
\end{aligned}
\tag{4.3.4}
$$

These isospin states are completely symmetric against permutations of the constituent quarks.

We write the general coupling of the three quarks as

$$
\begin{array}{|c|}\hline 1\\\hline\end{array} \otimes \begin{array}{|c|}\hline 2\\\hline\end{array} \otimes \begin{array}{|c|}\hline 3\\\hline\end{array} = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array} \oplus \begin{array}{|c|c}\hline 1&\multicolumn{1}{c|}{2}\\\hline 3\\\cline{1-1}\end{array} \oplus \begin{array}{|c|c}\hline 1&\multicolumn{1}{c|}{3}\\\hline 2\\\cline{1-1}\end{array} \oplus \begin{array}{|c|}\hline 1\\\hline 2\\\hline 3\\\hline\end{array}\,. \tag{4.3.5}
$$

Translated into $SU(2)$ language this equation reads

$$
\{2\}\otimes\{2\}\otimes\{2\} = \{4\}\oplus\{2\}\oplus\{2\}\oplus\{0\}. \tag{4.3.6}
$$

Here $\{0\}$ denotes an empty representation — one that cannot be realized in $SU(2)$ because the corresponding Young tableau has more than two rows. We identify the totally symmetric representation as the four charge states of the Δ-isobar, and we write as before $\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,I_3}$.

Before we can characterize the state of the Δ-isobar in more detail we must consider the other symmetries of the problem. The Δ-isobar is a resonance in the scattering of spin 1/2 nucleons and spin 0 pions. The experimentally observed spin of the resonance is 3/2. To account for this we associate a spin 1/2 with the constituent quarks. Then, in complete analogy to isospin, we can construct a totally symmetric spin representation for the Δ-particle

$$
\begin{aligned}
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,3/2} &= \uparrow\uparrow\uparrow,\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,1/2} &= \frac{1}{\sqrt{3}}(\uparrow\uparrow\downarrow \;+\; \uparrow\downarrow\uparrow \;+\; \downarrow\uparrow\uparrow),\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-1/2} &= \frac{1}{\sqrt{3}}(\uparrow\downarrow\downarrow \;+\; \downarrow\uparrow\downarrow \;+\; \downarrow\downarrow\uparrow),\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-3/2} &= \downarrow\downarrow\downarrow\,.
\end{aligned}
\tag{4.3.7}
$$

The isospin-spin part of the Δ-isobar state hence takes the form

$$|\Delta I_3 S_z\rangle = \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline \end{array}_{\,I_3} \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline \end{array}_{\,S_z}. \tag{4.3.8}$$

This state is symmetric with respect to both isospin and spin. Consequently, it is symmetric under simultaneous isospin-spin permutations. For illustrative purposes we write down the state for a Δ^+ particle with spin projection $S_z = 1/2$

$$\begin{aligned} |\Delta \tfrac{1}{2}\tfrac{1}{2}\rangle &= \frac{1}{3}(u\uparrow u\uparrow d\downarrow + u\uparrow u\downarrow d\uparrow + u\downarrow u\uparrow d\uparrow \\ &+ u\uparrow d\uparrow u\downarrow + u\uparrow d\downarrow u\uparrow + u\downarrow d\uparrow u\uparrow \\ &+ d\uparrow u\uparrow u\downarrow + d\uparrow u\downarrow u\uparrow + d\downarrow u\uparrow u\uparrow). \end{aligned} \tag{4.3.9}$$

One sees explicitly that this state is totally symmetric.

As we have seen, the Young tableau $\begin{array}{|c|c|}\hline \ & \ \\ \hline \ \\ \cline{1-1}\end{array}$ is associated with the isodoublet $\{2\}$. Hence, it is natural to expect that the nucleon state can be constructed from it. Now we have two possibilities $\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}_{\,I_3}$ and $\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}_{\,I_3}$ corresponding to symmetric or antisymmetric couplings of the quarks 1 and 2. Using Clebsch-Gordon coefficients one finds

$$\begin{aligned} \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}_{\,1/2} &= \frac{1}{\sqrt{6}}(2uud - udu - duu), \\ \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}_{\,-1/2} &= \frac{1}{\sqrt{6}}(udd + dud - 2ddu), \\ \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}_{\,1/2} &= \frac{1}{\sqrt{2}}(udu - duu), \\ \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}_{\,-1/2} &= \frac{1}{\sqrt{2}}(udd - dud). \end{aligned} \tag{4.3.10}$$

Proton and neutron have spin 1/2. Hence, we have two possible coupling schemes $\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}_{\,S_z}$ and $\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}_{\,S_z}$. We now want to combine the mixed isospin and spin permutation symmetries to an isospin-spin representation of definite permutation symmetry. This requires to reduce the inner product

$$\begin{array}{|c|c|}\hline \ & \ \\ \hline \ & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{I_3} \times \begin{array}{|c|c|}\hline \ & \ \\ \hline \ & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{S_z} = \begin{array}{|c|c|c|}\hline \ & \ & \ \\ \hline\end{array}_{I_3 S_z} \oplus \begin{array}{|c|c|}\hline \ & \ \\ \hline \ & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{I_3 S_z} \oplus \begin{array}{|c|}\hline \ \\ \hline \ \\ \hline \ \\ \hline\end{array}_{I_3 S_z} \tag{4.3.11}$$

in S_3. The two isospin and spin representations can be coupled to a symmetric, mixed or antisymmetric isospin-spin representation. As for the Δ-isobar we want to couple isospin and spin symmetrically. To do this explicitly, we need the Clebsch-Gordon coefficients of the group S_3. One finds

$$|N I_3 S_z\rangle = \frac{1}{\sqrt{2}}\left(\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{I_3} \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{S_z} + \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{I_3} \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}_{S_z} \right). \tag{4.3.12}$$

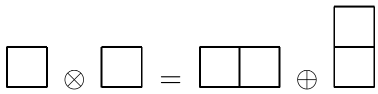
In our construction we have implicitly assumed that the orbital angular momentum of the constituent quarks inside a hadron vanishes. Then the orbital state is completely symmetric in the coordinates of the quarks. The orbital part of the baryon wave function therefore is described by the Young tableau $\begin{array}{|c|c|c|}\hline \ & \ & \ \\ \hline\end{array}$. Since also the isospin-spin part is totally symmetric, the baryon wave function is completely symmetric under permutations of the quarks. Since we have treated constituent quarks as spin 1/2 fermions, this contradicts the Pauli principle which requires a totally antisymmetric fermion wave function, and hence the Young tableau $\begin{array}{|c|}\hline \ \\ \hline \ \\ \hline \ \\ \hline\end{array}$. To satisfy the Pauli principle the color symmetry comes to our rescue. In $SU(3)_c$, $\begin{array}{|c|}\hline \ \\ \hline \ \\ \hline \ \\ \hline\end{array}$ corresponds to a singlet representation, which means that baryons are color-neutral. Since we have three colors we can now completely antisymmetrize three quarks

$$\begin{array}{|c|}\hline \ \\ \hline \ \\ \hline \ \\ \hline\end{array} = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr). \tag{4.3.13}$$

The color symmetry is the key to the fundamental understanding of the strong interactions. As opposed to isospin, color is an exact and even local symmetry.

4.4 Anti-Quarks and Mesons

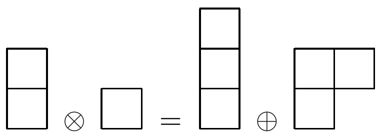
We have seen that the baryons (nucleon and Δ) consist of three constituent quarks (isospin doublets, spin doublets, color triplets). Now we want to construct the mesons (pion and ρ) in a similar manner. Since these particles have spin 0 and 1 respectively, they must contain an even number of constituent quarks. When we use two quarks, i.e. when we construct states like uu, ud or dd, the resulting electric charges are $4/3$, $1/3$ and $-2/3$ in contradiction to experiment. Also the coupling of two color triplets



$$\{3\} \otimes \{3\} = \{6\} \oplus \{\bar{3}\}, \tag{4.4.1}$$

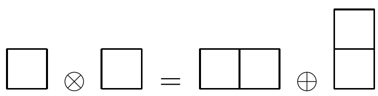
does not contain a singlet as desired by the confinement hypothesis.

We have seen already that a representation together with its anti-representation can always be coupled to a singlet. In $SU(3)$ this corresponds to



$$\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}, \tag{4.4.2}$$

Hence it is natural to work with anti-quarks. Anti-quarks are isospin doublets, spin doublets and color anti-triplets. We have quarks $\bar{u}$ and $\bar{d}$ with electric charges $Q_{\bar{u}} = -2/3$ and $Q_{\bar{d}} = 1/3$. Now we consider combinations of quark and anti-quark $u\bar{d}$, $u\bar{u}$, $d\bar{d}$ and $\mathrm{d}\bar{u}$, which have charges 1, 0 and -1 as we need them for the mesons. First we couple the isospin wave function



$$\{2\} \otimes \{2\} = \{3\} \oplus \{1\}, \tag{4.4.3}$$

and we obtain

$$
\begin{aligned}
\begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,1} &= u\bar{d}, \\
\begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,0} &= \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}), \\
\begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,-1} &= d\bar{u}, \\
\begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}_{\,0} &= \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}).
\end{aligned} \tag{4.4.4}
$$

We proceed analogously for the spin and we obtain

$$
\begin{aligned}
|\pi I_3 S_z\rangle &= \begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,I_3} \begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}_{\,S_z}, \\
|\rho I_3 S_z\rangle &= \begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,I_3} \begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,S_z}.
\end{aligned} \tag{4.4.5}
$$

Since quarks and anti-quarks are distinguishable particles (for example they have different charges) we don't have to respect the Pauli principle in this case. As opposed to the baryons here the coupling to color singlets follows only from the confinement hypothesis.

Of course, we can combine isospin and spin wave functions also in a different way

$$
\begin{aligned}
|\omega I_3 S_z\rangle &= \begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}_{\,I_3} \begin{array}{|c|c|}\hline & \\ \hline\end{array}_{\,S_z}, \\
|\eta' I_3 S_z\rangle &= \begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}_{\,I_3} \begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}_{\,S_z}.
\end{aligned} \tag{4.4.6}
$$

Indeed one observes mesons with these quantum numbers with masses $M_\omega = 0.782\text{GeV}$ and $M_{\eta'} = 0.958\text{GeV}$.

4.5 Strange Hadrons

Up to now we have considered hadrons that consist of up and down quarks and their anti-particles. However, one also observes hadrons containing strange quarks. The masses of the scalar ($S = 0$) mesons are given by

$$
M_\pi = 0.138\text{GeV}, M_K = 0.496\text{GeV}, M_\eta = 0.549\text{GeV}, M_{\eta'} = 0.958\text{GeV}, \tag{4.5.1}
$$

while the vector ($S = 1$) meson masses are

$$M_\rho = 0.770\text{GeV}, M_\omega = 0.783\text{GeV}, M_{K^*} = 0.892\text{GeV}, M_\varphi = 1.020\text{GeV}. \quad (4.5.2)$$

Altogether we have nine scalar and nine vector mesons. In each group we have so far classified four (π^+, π^0, π^-, η' and ρ^+, ρ^0, ρ^-, ω). The number four resulted from the $SU(2)_I$ isospin relation

$$\{\bar{2}\} \otimes \{2\} = \{1\} \oplus \{3\}. \quad (4.5.3)$$

The number nine then suggests to consider the corresponding $SU(3)$ identity

$$\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}. \quad (4.5.4)$$

Indeed we obtain nine mesons if we generalize isospin to a larger symmetry $SU(3)_F$. This so-called flavor group has nothing to with with the color symmetry $SU(3)_c$. It is only an approximate symmetry of QCD, with $SU(2)_I$ as a subgroup. In $SU(3)_F$ we have another quark flavor s — the strange quark.

The generators of $SU(3)$ can be chosen as follows

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \ \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$
$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \ \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \ \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (4.5.5)$$

Two of the generators commute with each other $[\lambda_3, \lambda_8] = 0$. We say that the group $SU(3)$ has rank 2. One can now identify the generators of the isospin subgroup $SU(2)_I$

$$I_1 = \frac{1}{2}\lambda_1, \ I_2 = \frac{1}{2}\lambda_2, \ I_3 = \frac{1}{2}\lambda_3. \quad (4.5.6)$$

Also it is convenient to introduce the so-called strong hypercharge

$$Y = \frac{1}{\sqrt{3}}\lambda_8, \quad (4.5.7)$$

(not to be confused with the generator of $U(1)_Y$ gauge transformations in the standard model). Then I^2, I_3 and Y commute with each other, and we can characterize the states of an $SU(3)_F$ multiplet by their isospin quantum numbers and by their hypercharge. Starting with the $SU(3)_F$ triplet we have

$$\begin{aligned} I^2 u &= \frac{1}{2}(\frac{1}{2}+1)u = \frac{3}{4}u, \ I_3 u = \frac{1}{2}u, \ Y u = \frac{1}{3}u, \\ I^2 d &= \frac{1}{2}(\frac{1}{2}+1)d = \frac{3}{4}d, \ I_3 d = -\frac{1}{2}d, \ Y d = \frac{1}{3}d, \\ I^2 s &= 0, \ I_3 s = 0, \ Y s = -\frac{2}{3}s. \end{aligned} \tag{4.5.8}$$

The electric charge is now given by

$$Q = I_3 + \frac{1}{2}Y, \tag{4.5.9}$$

such that

$$Q_u = \frac{2}{3}, \ Q_d = -\frac{1}{3}, \ Q_s = -\frac{1}{3}, \tag{4.5.10}$$

i.e. the charge of the strange quark is the same as the one of the down quark. If $SU(3)_F$ would be a symmetry as good as $SU(2)_I$ the states in an $SU(3)_F$ multiplet should be almost degenerate. This is, however, not quite the case, and $SU(3)_F$ is only approximately a symmetry of QCD.

Of course, we can also include the s quark in baryons. Then we have

$$\{3\} \otimes \{3\} \otimes \{3\} = \{10\} \oplus 2\{8\} \oplus \{1\} \tag{4.5.11}$$

compared to the old $SU(2)_I$ result

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus 2\{2\} \oplus \{0\}. \tag{4.5.12}$$

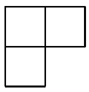
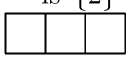
Indeed one observes more baryons than just nucleon and Δ-isobar.

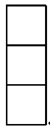
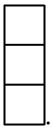
The baryon masses for the spin 1/2 baryons are

$$M_N = 0.939\text{GeV}, \ M_\Lambda = 1.116\text{GeV}, \ M_\Sigma = 1.193\text{GeV}, \ M_\Xi = 1.318\text{GeV}, \tag{4.5.13}$$

while the spin 3/2 baryon masses are

$$M_\Delta = 1.232\text{GeV}, \ M_{\Sigma^*} = 1.385\text{GeV}, \ M_{\Xi^*} = 1.530\text{GeV}, \ M_\Omega = 1.672\text{GeV}. \tag{4.5.14}$$

Proton and neutron are part of an octet:  is $\{2\}$ in $SU(2)_I$ and $\{8\}$ in $SU(3)_F$. The Δ-isobar is part of a decouplet:  is $\{4\}$ in $SU(2)_I$ and

$\{10\}$ in $SU(3)_F$. One does not find an $SU(3)_F$ singlet . This is because a spatially symmetric color singlet wave function is totally antisymmetric. To obtain a totally antisymmetric wave function also the spin part should transform as . Of course, in $SU(2)_S$ this is impossible.

We want to assume that the $SU(3)_F$ symmetry is explicitly broken because the s quark is heavier than the u and d quarks. Based on the quark content one would expect

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*} = M_s - M_q. \tag{4.5.15}$$

In fact one finds experimentally

$$M_{\Sigma^*} - M_\Delta = 0.153\text{GeV},\ M_{\Xi^*} - M_{\Sigma^*} = 0.145\text{GeV},\ M_\Omega - M_{\Xi^*} = 0.142\text{GeV}. \tag{4.5.16}$$

4.6 Gellman-Okubo Baryon Mass Formula

We have seen that the constituent quark model leads to a successful classification of hadron states in terms of flavor symmetry. The results about the hadron dynamics are, however, of more qualitative nature, and the assumption that a hadron is essentially a collection of a few constituent quarks is certainly too naive. The fundamental theory of the strong interactions is QCD. Here we want to use very basic QCD physics together with group theory to describe patterns in the hadron spectrum. The interaction between quarks and gluons is flavor independent, and therefore $SU(3)_F$ symmetric. Also the gluon self-interaction is flavor symmetric because the gluons are flavor singlets. A violation of flavor symmetry results only from the quark mass matrix

$$\mathcal{M} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix}. \tag{4.6.1}$$

We want to assume that u and d quark have the same mass m_q, while the s quark is heavier ($m_s > m_q$). The quark mass matrix can be written as

$$\begin{aligned}\mathcal{M} &= \frac{2m_q + m_s}{3}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} + \frac{m_q - m_s}{3}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix}\\ &= \frac{2m_q + m_s}{3}1 + \frac{m_q - m_s}{\sqrt{3}}\lambda_8.\end{aligned} \tag{4.6.2}$$

The mass matrix contains an $SU(3)_F$ singlet as well as an octet piece. Correspondingly, the QCD Hamilton operator can be written as

$$H_{QCD} = H_1 + H_8. \tag{4.6.3}$$

We want to assume that H_8 is small and can be treated as a perturbation. Then we first consider H_1 alone. This is justified if the mass difference $m_q - m_s$ is small. Since H_1 is $SU(3)_F$ symmetric we expect degenerate states in $SU(3)_F$ multiplets — the hadron octets and decouplets. Here we assume that the flavor symmetry is not spontaneously broken. This should indeed be correct for QCD.

Let us start with the baryons. The eigenstates of H_1 are denoted by $|B_1YII_3\rangle$

$$H_1|B_1YII_3\rangle = M_{B_1}|B_1YII_3\rangle. \tag{4.6.4}$$

We use degenerate perturbation theory to first order in H_8 and obtain

$$M_B = M_{B_1} + \langle B_1YII_3|H_8|B_1YII_3\rangle. \tag{4.6.5}$$

A diagonalization in the space of degenerate states is not necessary, since H_8 transforms as the λ_8 component of an octet, and can therefore not change Y, I and I_3. Next we will compute the required matrix elements using group theory. Starting with the baryon decouplet, we obtain a nonzero value only if $\{8\}$ and $\{10\}$ can couple to $\{10\}$. Indeed the decouplet appears in the reduction. Using the Wigner-Eckart theorem we obtain

$$\langle B_1YII_3|H_8|B_1YII_3\rangle = \langle B_1||H_8||B_1\rangle\langle\{10\}YII_3|\{8\}000\{10\}YII_3\rangle, \tag{4.6.6}$$

where $\langle B_1||H_8||B_1\rangle$ is a reduced matrix element, and the second factor is an $SU(3)_F$ Clebsch-Gordon coefficient given by

$$\langle\{10\}YII_3|\{8\}000\{10\}YII_3\rangle = \frac{Y}{\sqrt{8}}. \tag{4.6.7}$$

Then we obtain for the baryon masses in the decouplet

$$M_B = M_{B_1} + \langle B_1||H_8||B_1\rangle\frac{Y}{\sqrt{8}}, \tag{4.6.8}$$

and hence

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*} = -\frac{1}{\sqrt{8}}\langle B_1||H_8||B_1\rangle. \qquad (4.6.9)$$

Indeed, as we saw before, the three mass differences are almost identical. In view of the fact that we have just used first order perturbation theory, this is quite remarkable.

Next we consider the mass splittings in the baryon octet. Here we must ask if $\{8\}$ and $\{8\}$ can couple to $\{8\}$. One finds

$$\{8\} \otimes \{8\} = \{27\} \oplus \{10\} \oplus \{\bar{10}\} \oplus 2\{8\} \oplus \{1\}. \qquad (4.6.10)$$

Hence there are even two ways to couple two octets to an octet. One is symmetric, the other is antisymmetric under the exchange of the two octets. We can write

$$\begin{aligned} \langle B_1 Y I I_3|H_8|B_1 Y I I_3\rangle &= \langle B_1||H_8||B_1\rangle_s \langle\{8\}Y I I_3|\{8\}000\{8\}Y I I_3\rangle_s \\ &+ \langle B_1||H_8||B_1\rangle_a \langle\{8\}Y I I_3|\{8\}000\{8\}Y I I_3\rangle_a. \end{aligned} \qquad (4.6.11)$$

The Clebsch-Gordon coefficients are given by

$$\begin{aligned} \langle\{8\}Y I I_3|\{8\}000\{8\}Y I I_3\rangle_s &= \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1), \\ \langle\{8\}Y I I_3|\{8\}000\{8\}Y I I_3\rangle_a &= \sqrt{\frac{3}{4}}Y, \end{aligned} \qquad (4.6.12)$$

and we obtain for the baryon octet

$$M_B = M_{B_1} + \langle B_1||H_8||B_1\rangle_s \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1) + \langle B_1||H_8||B_1\rangle_a \sqrt{\frac{3}{4}}Y. \qquad (4.6.13)$$

These formulas for the baryon masses were first derived by Gellman and Okubo. From the octet formula one obtains

$$\begin{aligned} 2M_N + 2M_\Xi &= 4M_{B_1} + \langle B_1||H_8||B_1\rangle_s \frac{4}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1), \\ M_\Sigma + 3M_\Lambda &= 4M_{B_1} + \langle B_1||H_8||B_1\rangle_s \frac{1}{\sqrt{5}}((2-1) + 3(-1)), \\ 2M_N + 2M_\Xi &= M_\Sigma + 3M_\Lambda. \end{aligned} \qquad (4.6.14)$$

Experimentally the two sides of the last equation give 1.129 GeV and 1.135 GeV in excellent agreement with the theory.

4.7 Meson Mixing

Similar to the baryons the explicit $SU(3)_F$ symmetry breaking due to the larger s quark mass leads to mass splittings also for the mesons. There, however, one has in addition a mixing between flavor octet and flavor singlet states. For the baryons a mixing between octet and decouplet is excluded because they have different spins. First we consider eigenstates of H_1 again

$$H_1|M_1 Y I I_3\rangle = M_{M_1}|M_1 Y I I_3\rangle. \quad (4.7.1)$$

The following analysis applies both to scalar and to vector mesons. In both cases we have an $SU(3)_F$ octet and a singlet. In perturbation theory we must now diagonalize a 9×9 matrix. Similar to the baryons the matrix is, however, already almost diagonal. Let us first consider the seven meson states with $Y, I, I_3 \neq 0, 0, 0$. These are π and K for the scalar and ρ and K^* for the vector mesons. One has

$$M_M = M_{M_1} + \langle M_1 Y I I_3|H_8|M_1 Y I I_3\rangle. \quad (4.7.2)$$

In complete analogy to the baryon octet we obtain

$$M_M = M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1) + \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}Y. \quad (4.7.3)$$

As opposed to the baryons the mesons and their anti-particles are in the same multiplet. For example we have

$$\begin{aligned} M_{K^+} &= M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1) + \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}, \\ M_{K^-} &= M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1) - \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}. \end{aligned} \quad (4.7.4)$$

According to the CPT theorem particles and anti-particles have exactly the same masses in a relativistic quantum field theory, and therefore

$$\langle M_1||H_8||M_1\rangle_a = 0. \quad (4.7.5)$$

Now we come to the issue of mixing between the mesons η_1 and η_8 and between ω_1 and ω_8. We concentrate on the vector mesons. Then we need the following matrix elements

$$\begin{aligned} \langle\omega_1|H_8|\omega_1\rangle &= 0, \\ \langle\omega_8|H_8|\omega_8\rangle &= \langle M_1||H_8||M_1\rangle_s \langle\{8\}000|\{8\}000\{8\}000\rangle_s \\ &= \langle M_1||H_8||M_1\rangle_s(-\frac{1}{\sqrt{5}}). \end{aligned} \quad (4.7.6)$$

The actual meson masses are the eigenvalues of the matrix

$$\mathcal{M} = \begin{pmatrix} M_{\omega_1} & \langle\omega_1|H_8|\omega_8\rangle \\ \langle\omega_8|H_8|\omega_1\rangle & M_{\omega_8} - \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}} \end{pmatrix}. \tag{4.7.7}$$

The particles φ and ω that one observes correspond to the eigenstates

$$\begin{aligned} |\varphi\rangle &= \cos\theta|\omega_1\rangle - \sin\theta|\omega_8\rangle, \\ |\omega\rangle &= \sin\theta|\omega_1\rangle + \cos\theta|\omega_8\rangle. \end{aligned} \tag{4.7.8}$$

Here θ is the meson mixing angle. One obtains

$$\begin{aligned} M_\varphi + M_\omega &= M_{\omega_1} + M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s, \\ M_\varphi M_\omega &= M_{\omega_1}(M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s) - |\langle\omega_1|H_8|\omega_8\rangle|^2. \end{aligned} \tag{4.7.9}$$

Also we have

$$\begin{aligned} M_\rho &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(2-1), \\ M_{K^*} &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1), \end{aligned} \tag{4.7.10}$$

and hence

$$\begin{aligned} \frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{4}{3}(-\frac{1}{2}) - \frac{1}{3}) \\ &= M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s, \end{aligned} \tag{4.7.11}$$

such that

$$\begin{aligned} M_{\omega_1} &= M_\varphi + M_\omega - \frac{4}{3}M_{K^*} + \frac{1}{3}M_\rho = 0.870\text{GeV}, \\ |\langle\omega_1|H_8|\omega_8\rangle|^2 &= M_{\omega_1}(\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho) - M_\varphi M_\omega = (0.113\text{GeV})^2. \end{aligned} \tag{4.7.12}$$

The mixing angle is now determined from

$$\begin{aligned} &M_{\omega_1}\cos\theta - \langle\omega_1|H_8|\omega_8\rangle\sin\theta = M_\varphi\cos\theta, \\ &\langle\omega_8|H_8|\omega_1\rangle\cos\theta - (M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s)\sin\theta = -M_\varphi\sin\theta. \end{aligned} \tag{4.7.13}$$

and we obtain

$$(M_{\omega_1} + M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s)\sin\theta\cos\theta - \langle\omega_1|H_8|\omega_8\rangle = 2M_\varphi \sin\theta\cos\theta, \quad (4.7.14)$$

and hence

$$\frac{1}{2}\sin(2\theta) = \pm\frac{\sqrt{(M_\varphi + M_\omega - \frac{4}{3}M_{K^*} + \frac{1}{3}M_\rho)(\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho) - M_\varphi M_\omega}}{M_\varphi - M_\omega}. \quad (4.7.15)$$

Numerically one obtains $\theta = \pm 52.6^0$ and therefore $\cos\theta \approx 1/\sqrt{3}$, $\sin\theta \approx \pm\sqrt{2/3}$, such that

$$\begin{aligned} |\varphi\rangle &\approx s\bar{s} \text{ or } \frac{1}{3}(2u\bar{u} + 2d\bar{d} - s\bar{s}), \\ |\omega\rangle &\approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}) \text{ or } -\frac{1}{\sqrt{18}}(u\bar{u} + d\bar{d} + 4s\bar{s}). \end{aligned} \quad (4.7.16)$$

The φ mesons decays in 84 percent of all cases into kaons ($\varphi \rightarrow K^+ + K^-, K^0 + \bar{K}^0$) and only in 16 percent of all cases into pions ($\varphi \rightarrow \pi^+ + \pi^0 + \pi^-$). Hence one concludes that the φ meson is dominated by s quarks, such that one has ideal mixing

$$|\varphi\rangle \approx s\bar{s}, \quad |\omega\rangle \approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}). \quad (4.7.17)$$

It is instructive to repeat the calculation of meson mixing for the scalar mesons η and η'.

4.8 Spontaneous Chiral Symmetry Breaking

Chiral symmetry is an approximate global symmetry of the QCD Lagrange density that results from the fact that the u and d quark masses are small compared to the typical QCD scale Λ_{QCD}. Neglecting the quark masses, the QCD Lagrange density is invariant against separate $U(2)$ transformations of the left- and right-handed quarks, such that we have a $U(2)_L \otimes U(2)_R$ symmetry. We can decompose each $U(2)$ symmetry into an $SU(2)$ and a $U(1)$ part, and hence we obtain $SU(2)_L \otimes SU(2)_R \otimes U(1)_L \otimes U(1)_R$. The $U(1)_B$ symmetry related to baryon number conservation corresponds to simultaneous rotations of left- and right-handed quarks, i.e. $U(1)_B = U(1)_{L=R}$. The remaining so-called axial $U(1)$ is affected by the Adler-Bell-Jackiw anomaly. It is explicitly broken by quantum effects, and hence it is not a symmetry of QCD. Later we will return to the

$U(1)$ problem related to this symmetry. Here we are interested in the ordinary (non-anomalous) symmetries of QCD — the $SU(2)_L \otimes SU(2)_R \otimes U(1)_B$ chiral symmetry. Based on this symmetry one would expect corresponding degeneracies in the QCD spectrum. Indeed we saw that the hadrons can be classified as isospin multiplets. The isospin transformations are $SU(2)_I$ rotations, that act on left- and right-handed fermions simultaneously, i.e. $SU(2)_I = SU(2)_{L=R}$. The symmetry that is manifest in the spectrum is hence $SU(2)_I \otimes U(1)_B$, but not the full chiral symmetry $SU(2)_L \otimes SU(2)_R \otimes U(1)_B$. One concludes that chiral symmetry must be spontaneously broken. The order parameter of chiral symmetry breaking is the so-called chiral condensate $\langle\bar{\Psi}\Psi\rangle$. When a continuous global symmetry breaks spontaneously, massless particles — the Goldstone bosons — appear in the spectrum. According to the Goldstone theorem the number of Goldstone bosons is the difference of the number of generators of the full symmetry group and the subgroup remaining after spontaneous breaking. In our case we hence expect $3 + 3 + 1 - 3 - 1 = 3$ Goldstone bosons. In QCD they are identified as the pions π^+, π^0 and π^-. Of course, the pions are light, but they are not massless. This is due to a small explicit chiral symmetry breaking related to the small but nonzero masses of the u and d quarks. Chiral symmetry is only an approximate symmetry, and the pions are only pseudo-Goldstone bosons. It turns out that the pion mass squared is proportional to the quark mass. When we also consider the s quark as being light, chiral symmetry can be extended to $SU(3)_L \otimes SU(3)_R \otimes U(1)_B$, which then breaks spontaneously to $SU(3)_F \otimes U(1)_B$. Then one expects $8+8+1-8-1 = 8$ Goldstone bosons. The five additional bosons are identified as the four kaons K^+, K^0, $\bar{K}^0$, K^- and the η-meson. Since the s quark mass is not really negligible, these pseudo Goldstone bosons are heavier than the pion.

The Goldstone bosons are the lightest particles in QCD. Therefore they determine the dynamics at small energies. One can construct effective theories that are applicable in the low energy regime, and that are formulated in terms of Goldstone boson fields. At low energies the Goldstone bosons interact only weakly and can hence be treated perturbatively. This is done systematically in chiral perturbation theory.

Let us consider the quark part of the QCD Lagrange density

$$\mathcal{L}(\bar{\Psi}, \Psi, G_\mu) = \bar{\Psi}(x)(i\gamma^\mu(\partial_\mu + g_s G_\mu(x)) - \mathcal{M})\Psi(x). \tag{4.8.1}$$

We now decompose the quark fields in right- and left-handed components

$$\Psi_R(x) = \frac{1}{2}(1+\gamma_5)\Psi(x),\ \Psi_L(x) = \frac{1}{2}(1-\gamma_5)\Psi(x),\ \Psi(x) = \Psi_R(x)+\Psi_L(x). \tag{4.8.2}$$

Here we have used

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3, \ \{\gamma^\mu, \gamma^\nu\} = 2g_{\mu\nu}, \ \{\gamma^\mu, \gamma_5\} = 0. \tag{4.8.3}$$

Next we consider the adjoint spinors

$$\begin{aligned}\bar{\Psi}_R(x) &= \Psi_R(x)^+\gamma^0 = \Psi(x)^+\frac{1}{2}(1+\gamma_5^+)\gamma^0 = \Psi(x)^+\gamma^0\frac{1}{2}(1-\gamma_5) \\ &= \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5), \\ \bar{\Psi}_L(x) &= \Psi_L(x)^+\gamma^0 = \Psi(x)^+\frac{1}{2}(1-\gamma_5^+)\gamma^0 = \Psi(x)^+\gamma^0\frac{1}{2}(1+\gamma_5) \\ &= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5).\end{aligned} \tag{4.8.4}$$

Here we used

$$\gamma^0\gamma_5^+\gamma^0 = -\gamma_5. \tag{4.8.5}$$

Inserting the decomposed spinors in the Lagrange density we obtain

$$\mathcal{L}(\bar{\Psi}, \Psi, G_\mu) = (\bar{\Psi}_R(x)+\bar{\Psi}_L(x))(i\gamma^\mu(\partial_\mu+g_sG_\mu(x))-\mathcal{M})(\Psi_R(x)+\Psi_L(x)). \tag{4.8.6}$$

First, we investigate the γ^μ term

$$\begin{aligned}&\bar{\Psi}_R(x)i\gamma^\mu(\partial_\mu + g_sG_\mu(x))\Psi_L(x) \\ &= \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)i\gamma^\mu(\partial_\mu + g_sG_\mu(x))\frac{1}{2}(1-\gamma_5)\Psi(x) \\ &= \bar{\Psi}(x)\frac{1}{4}(1-\gamma_5)(1+\gamma_5)i\gamma^\mu(\partial_\mu + g_sG_\mu(x))\Psi(x) = 0.\end{aligned} \tag{4.8.7}$$

On the other hand, for the mass term one finds

$$\begin{aligned}\bar{\Psi}_R(x)\mathcal{M}\Psi_R(x) &= \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\mathcal{M}\frac{1}{2}(1+\gamma_5)\Psi(x) \\ &= \bar{\Psi}(x)\frac{1}{4}(1-\gamma_5)(1+\gamma_5)\mathcal{M}\Psi(x) = 0.\end{aligned} \tag{4.8.8}$$

Hence, we can write

$$\begin{aligned}\mathcal{L}(\bar{\Psi}, \Psi, G_\mu) &= \bar{\Psi}_R(x)i\gamma^\mu(\partial_\mu + g_sG_\mu(x))\Psi_R(x) \\ &+ \bar{\Psi}_L(x)i\gamma^\mu(\partial_\mu + g_sG_\mu(x))\Psi_L(x) \\ &- \bar{\Psi}_R(x)\mathcal{M}\Psi_L(x) - \bar{\Psi}_L(x)\mathcal{M}\Psi_R(x).\end{aligned} \tag{4.8.9}$$

The γ^μ term decomposes into two decoupled contributions from right- and left-handed quarks. This part of the Lagrange density is invariant against separate $U(N_f)$ transformations of the right- and left-handed components in flavor space

$$\begin{aligned}&\Psi'_R(x) = R\Psi_R(x), \ \bar{\Psi}'(x) = \bar{\Psi}_R(x)R^+, \ R \in U(N_f)_R, \\ &\Psi'_L(x) = L\Psi_L(x), \ \bar{\Psi}'(x) = \bar{\Psi}_L(x)L^+, \ L \in U(N_f)_L.\end{aligned} \tag{4.8.10}$$

Without the mass term the classical QCD Lagrange density hence has a $U(N_f)_L \otimes U(N_f)_R$ symmetry. Due to the anomaly in the axial $U(1)$ symmetry the symmetry of the quantum theory is reduced to

$$SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_{L=R} = SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B. \quad (4.8.11)$$

Of course, the chiral symmetry is only approximate, because the mass term couples right- and left-handed fermions. In addition, the mass matrix does not commute with R and L. If all quarks had the same mass, i.e. if $\mathcal{M} = m\mathbb{1}$, one would have

$$\bar{\Psi}'_R(x)\mathcal{M}\Psi'_L(x) = \bar{\Psi}_R(x)R^+ m\mathbb{1}L\Psi_L(x) = \bar{\Psi}_R(x)R^+ L\mathcal{M}\Psi_L(x). \quad (4.8.12)$$

Then the mass term is invariant only against simultaneous transformations $R = L$ such that $R^+L = R^+R = \mathbb{1}$. Hence, chiral symmetry is then explicitly broken to

$$SU(N_f)_{L=R} \otimes U(1)_{L=R} = SU(N_f)_F \otimes U(1)_B, \quad (4.8.13)$$

which corresponds to the flavor and baryon number symmetry. In reality the quark masses are different, and the symmetry is in fact explicitly broken to

$$\otimes_f U(1)_f = U(1)_u \otimes U(1)_d \otimes U(1)_s. \quad (4.8.14)$$

It is, however, much more important that the u and d quark masses are small, and can hence almost be neglected. Therefore, in reality the chiral $SU(2)_L \otimes SU(2)_R \otimes U(1)_B \otimes U(1)_s$ symmetry is almost unbroken explicitly. Since the s quark is heavier, $SU(3)_L \otimes SU(3)_R \otimes U(1)_B$ is a more approximate chiral symmetry, because it is explicitly more strongly broken.

Since the masses of the u and d quarks are so small, the $SU(2)_L \otimes SU(2)_R$ chiral symmetry should work very well. Hence, one would expect that the hadron spectrum shows corresponding degeneracies. Let us neglect quark masses and consider the then conserved currents

$$\begin{aligned} J_\mu^{La}(x) &= \bar{\Psi}_L(x)\gamma_\mu \frac{\sigma^a}{2}\Psi_L(x), \\ J_\mu^{Ra}(x) &= \bar{\Psi}_R(x)\gamma_\mu \frac{\sigma^a}{2}\Psi_R(x), \end{aligned} \quad (4.8.15)$$

where $a \in \{1, 2, 3\}$. From the right- and left-handed currents we now construct vector and axial currents

$$\begin{aligned}
V^a_\mu(x) &= J^{La}_\mu(x) + J^{Ra}_\mu(x) \\
&= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\frac{1}{2}(1-\gamma_5)\Psi(x) \\
&+ \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\frac{1}{2}(1+\gamma_5)\Psi(x) \\
&= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\Psi(x) + \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\Psi(x) \\
&= \bar{\Psi}(x)\gamma_\mu\frac{\sigma^a}{2}\Psi(x), \\
A^a_\mu(x) &= J^{La}_\mu(x) - J^{Ra}_\mu(x) = \bar{\Psi}(x)\gamma_5\gamma_\mu\frac{\sigma^a}{2}\Psi(x). \qquad (4.8.16)
\end{aligned}$$

Let us consider an $SU(2)_L \otimes SU(2)_R$ invariant state $|\Phi\rangle$ as a candidate for the QCD vacuum. Then

$$\langle\Phi|J^{La}_\mu(x)J^{Rb}_\nu(y)|\Phi\rangle = \langle\Phi|J^{Ra}_\mu(x)J^{Lb}_\nu(y)|\Phi\rangle = 0, \qquad (4.8.17)$$

and hence

$$\langle\Phi|V^a_\mu(x)V^b_\nu(y)|\Phi\rangle = \langle\Phi|A^a_\mu(x)A^b_\nu(y)|\Phi\rangle. \qquad (4.8.18)$$

On both sides of the equation one can insert complete sets of states between the two operators. On the left hand side states with quantum numbers $J^P = 0^+, 1^-$ contribute, while on the right hand side the nonzero contributions come from states $0^-, 1^+$. The two expressions can be equal only if the corresponding parity partners are energetically degenerate. In the observed hadron spectrum there is no degeneracy of particles with even and odd parity, not even approximately. We conclude that the $SU(2)_L \otimes SU(2)_R$ invariant state $|\Phi\rangle$ is not the real QCD vacuum. The true vacuum $|0\rangle$ cannot be chirally invariant. The same is true for all other eigenstates of the QCD Hamiltonian. This means that chiral symmetry must be spontaneously broken.

Let us now consider the states

$$\begin{aligned}
Q^a_V|0\rangle &= \int d^3x V^a_0(\vec{x},0)|0\rangle, \\
Q^a_A|0\rangle &= \int d^3x A^a_0(\vec{x},0)|0\rangle, \qquad (4.8.19)
\end{aligned}$$

constructed from the vacuum by acting with the vector and axial charge densities. If the vacuum were chirally symmetric we would have

$$Q^a_V|\Phi\rangle = Q^a_A|\Phi\rangle = 0. \qquad (4.8.20)$$

The real QCD vacuum is not chirally invariant because

$$Q_A^a|0\rangle \neq 0. \tag{4.8.21}$$

Since the axial current is conserved (for massless quarks $\partial^\mu A_\mu^a(x) = 0$) we have

$$[H_{QCD}, Q_A^a] = 0. \tag{4.8.22}$$

Hence the new state $Q_A^a|0\rangle$ is again an eigenstate of the QCD Hamilton operator

$$H_{QCD} Q_A^a|0\rangle = Q_A^a H_{QCD}|0\rangle = 0 \tag{4.8.23}$$

with zero energy. This state corresponds to a massless Goldstone boson with quantum numbers $J^P = 0^-$. These pseudoscalar particles are identified with the pions of QCD.

If one would also have $Q_V^a|0\rangle \neq 0$, the vector flavor symmetry would also be spontaneously broken, and there would be another set of scalar Goldstone bosons with $J^P = 0^+$. Such particles do not exist in the hadron spectrum, and we conclude that the isospin symmetry $SU(2)_I = SU(2)_{L=R}$ is not spontaneously broken. As we have seen before, the isospin symmetry is indeed manifest in the hadron spectrum.

One can also detect spontaneous chiral symmetry breaking by investigating the chiral order parameter

$$\langle\bar\Psi\Psi\rangle = \langle 0|\bar\Psi(x)\Psi(x)|0\rangle = \langle 0|\bar\Psi_R(x)\Psi_L(x) + \bar\Psi_L(x)\Psi_R(x)|0\rangle. \tag{4.8.24}$$

The order parameter is invariant against simultaneous transformations $R = L$, but not against general chiral rotations. If chiral symmetry would be intact the chiral condensate would vanish. When the symmetry is spontaneously broken, on the other hand, $\langle\bar\Psi\Psi\rangle \neq 0$.

Being almost massless, the Goldstone bosons are the lightest particles in QCD. Therefore they dominate the dynamics of the strong interactions at low energies, and it is possible to switch to a low-energy effective description that only involves the Goldstone bosons. This is not only the case for QCD but also for any other model with a continuous global symmetry G breaking spontaneously to a subgroup H, provided that there are no other massless particles besides the Goldstone bosons. The Goldstone bosons are described by fields in the coset space G/H, in which points are identified if they are connected by symmetry transformations of the remaining subgroup H. In QCD we have

$$\begin{aligned} G &= SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B, \\ H &= SU(N_f)_{L=R} \otimes U(1)_B, \end{aligned} \tag{4.8.25}$$

and the corresponding coset space is

$$G/H = SU(N_f). \tag{4.8.26}$$

Hence the Goldstone boson fields can be represented as special unitary matrices $U(x) \in SU(N_f)$. Under chiral rotations they transform as

$$U(x)' = LU(x)R^+. \tag{4.8.27}$$

Now we must construct an effective Lagrange function which is chirally symmetric. For this purpose we consider

$$\partial_\mu U(x)' = L\partial_\mu U(x)R^+, \tag{4.8.28}$$

and we form a chirally invariant Lorentz scalar

$$\mathcal{L}(U) = \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)^+\partial_\mu U(x)). \tag{4.8.29}$$

The coupling constant f_π determines the strength of the interaction between the Goldstone bosons. It also plays a role in the weak decay of the pion and is therefore known as the pion decay constant. The above Lagrange density is chiral invariant because

$$\begin{aligned}\mathcal{L}(U') &= \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)'^+\partial_\mu U(x)') \\ &= \frac{f_\pi^2}{4}\mathrm{Tr}(R\partial^\mu U(x)^+L^+L\partial_\mu U(x)R^+) = \mathcal{L}(U).\end{aligned} \tag{4.8.30}$$

We still must introduce the chiral symmetry breaking mass terms, which enter via the quark mass matrix. We write

$$\mathcal{L}(U) = \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)^+\partial_\mu U(x)) + c\mathrm{Tr}(\mathcal{M}(U + U^+)). \tag{4.8.31}$$

Under chiral transformations the additional term transforms into

$$\mathrm{Tr}(\mathcal{M}(U(x)' + U(x)'^+)) = \mathrm{Tr}(\mathcal{M}(LU(x)R^+ + RU(x)^+L^+)). \tag{4.8.32}$$

If all quark masses are equal, i.e. for $\mathcal{M} = m\mathbb{1}$, the Lagrange density is again invariant against rotations $R^+L = \mathbb{1}$ and hence for $R = L$. For a general mass matrix the symmetry is reduced to $\otimes_f U(1)_f/U(1)_B$.

We still have to determine the prefactor c. For this purpose we determine the vacuum value of $\partial/\partial m_f \mathcal{L}|_{\mathcal{M}=0}$ first in QCD

$$\langle 0|\frac{\partial}{\partial m_f}\mathcal{L}|_{\mathcal{M}=0}|0\rangle = \langle 0| - \bar{\Psi}_f\Psi_f|0\rangle = -\frac{1}{N_f}\langle\bar{\Psi}\Psi\rangle. \tag{4.8.33}$$

In the effective theory the same value should arise. The classical vacuum of the effective theory with $\mathcal{M} = 0$ corresponds to a constant field

$$U(x) = \mathbb{1}. \tag{4.8.34}$$

Hence, in the effective theory we have

$$\langle 0| \frac{\partial}{\partial m_f} \mathcal{L}|_{\mathcal{M}=0}|0\rangle = c\text{Tr}(\text{diag}(1, 0, ..., 0)(1 + 1)) = 2c, \tag{4.8.35}$$

and therefore

$$c = -\frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle. \tag{4.8.36}$$

The constants f_π and $\langle\bar{\Psi}\Psi\rangle$ determine the low-energy dynamics of QCD. They can only be determined from QCD itself, and must be inserted in the effective theory as a priori unknown parameters. Up to these two low-energy constants the Goldstone boson dynamics is completely determined by chiral symmetry. At higher energies additional terms arise in the effective theory. Again, they are restricted by chiral symmetry requirements, but they contain additional parameters. In fact, chiral perturbation theory is a systematic low-energy expansion, in which the higher order terms contain a larger number of derivatives. Here we restrict ourselves to lowest order, and hence to the Lagrange density

$$\mathcal{L}(U) = \frac{f_\pi^2}{4}\text{Tr}(\partial^\mu U(x)^+ \partial_\mu U(x)) - \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\text{Tr}(\mathcal{M}(U + U^+)). \tag{4.8.37}$$

Chiral perturbation theory is an expansion around the classical vacuum solution $U(x) = \mathbb{1}$. One writes

$$U(x) = \exp(i\pi^a(x)\eta_a/f_\pi),\ a \in \{1, 2, ..., N_f^2 - 1\}, \tag{4.8.38}$$

where η_a are the generators of $SU(N_f)$, and one expands in powers of $\pi^a(x)$. To lowest order we have

$$U(x) = 1 + i\pi^a(x)\eta_a/f_\pi,\ \partial_\mu U(x) = i\partial_\mu\pi^a(x)\eta_a/f_\pi, \tag{4.8.39}$$

and hence

$$\begin{aligned}\mathcal{L}(U) &= \frac{1}{4}\text{Tr}(\partial^\mu\pi^a(x)\eta_a\partial_\mu\pi^b(x)\eta_b) - \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\text{Tr}(\mathcal{M}(1 + 1)) \\ &= \frac{1}{2}\partial^\mu\pi^a(x)\partial_\mu\pi^b(x)\delta_{ab} - \frac{1}{N_f}\langle\bar{\Psi}\Psi\rangle\text{Tr}\mathcal{M}. \end{aligned}\tag{4.8.40}$$

The last term is an irrelevant constant. We have to expand consistently to order π^2

$$U(x) = 1 + i\pi^a(x)\eta_a/f_\pi + \frac{1}{2}(i\pi^a(x)\eta_a/f_\pi)^2, \tag{4.8.41}$$

such that for quarks with equal masses, i.e. for $\mathcal{M} = m\mathbb{1}$,

$$\begin{aligned}\mathrm{Tr}(\mathcal{M}(U(x) + U(x)^+)) &= m\mathrm{Tr}(U(x) + U(x)^+) \\ &= 2mN_f - m\frac{1}{f_\pi^2}\pi^a(x)\pi^b(x)\mathrm{Tr}(\eta_a\eta_b) \\ &= 2mN_f - 2m\frac{1}{f_\pi^2}\pi^a(x)\pi^a(x).\end{aligned} \tag{4.8.42}$$

Altogether we obtain

$$\mathcal{L}(U) = \frac{1}{2}\partial^\mu\pi^a(x)\partial_\mu\pi^a(x) - \frac{1}{N_f}\langle\bar{\Psi}\Psi\rangle(mN_f - m\frac{1}{f_\pi^2}\pi^a(x)\pi^a(x)). \tag{4.8.43}$$

The resulting equation of motion is given by

$$\partial^\mu\frac{\delta\mathcal{L}}{\delta\partial^\mu\pi^a} - \frac{\delta\mathcal{L}}{\delta\pi^a} = \partial^\mu\partial_\mu\pi^a(x) - \frac{2m\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}\pi^a(x) = 0. \tag{4.8.44}$$

This is the Klein-Gordon equation for a pseudoscalar particle with mass

$$M_\pi^2 = \frac{2m\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}. \tag{4.8.45}$$

This behavior is typical for the mass of a pseudo Goldstone boson: it is proportional to the square root of the explicit symmetry breaking term. It is interesting to derive a corresponding mass formula for a nondegenerate mass matrix, e.g. for light quark masses $m_u = m_d = m_q$, but a heavier strange quark mass m_s. For general quark masses one obtains

$$\begin{aligned}M_{K^+}^2 &= M_{K^-}^2 = \frac{(m_u + m_s)\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}, \\ M_{K^0}^2 &= M_{\bar{K}^0}^2 = \frac{(m_d + m_s)\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}, \\ M_\eta^2 &= \frac{1}{3}\frac{(m_u + m_d + 4m_s)\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}.\end{aligned} \tag{4.8.46}$$

This leads to the following mass relation

$$3M_\eta^2 + M_\pi^2 = 2M_{K^+}^2 + 2M_{K^0}^2, \tag{4.8.47}$$

which is well satisfied experimentally. The left hand side has a value of $0.923\mathrm{GeV}^2$ and the right hand side is $0.984\mathrm{GeV}^2$. Introducing the average light quark mass

$$m_q = \frac{1}{2}(m_u + m_d) \tag{4.8.48}$$

one obtains

$$\frac{M_{K^+}^2 + M_{K^0}^2}{M_\pi^2} = \frac{m_s + m_q}{m_q}, \tag{4.8.49}$$

which yields $m_s/m_q = 24.2$. Similarly

$$\frac{M_\eta^2}{M_\pi^2} = \frac{2m_s + m_q}{3m_q}, \tag{4.8.50}$$

which leads to $m_s/m_q = 22.7$. Still the quark masses themselves are not directly measurable. Using various models one obtains

$$m_q \approx 0.007\text{GeV}, \;\; m_s \approx 0.16\text{GeV}. \tag{4.8.51}$$

Chapter 5

Topology of Gauge Fields

In this chapter we investigate the topological structure of non-Abelian gauge fields. In the standard model, the non-trivial topology of $SU(2)_L$ gauge fields gives rise to baryon number violating processes. Similarly, in QCD a non-trivial topology of the gluon field leads to an explicit breaking of the flavor-singlet axial symmetry. This offers an explanation for the $U(1)_A$ problem in QCD — the question why the η'-meson is not a pseudo-Goldstone boson. The gauge field topology also gives rise to a new parameter in QCD — the vacuum angle θ. This confronts us with the strong CP problem: why is θ consistent with zero in nature? We will return to the $U(1)_A$ and the strong CP problem in the next chapter. Here we concentrate on first understanding the topology of the gauge field itself.

5.1 The Anomaly

Let us consider the baryon number current in the standard model

$$J_\mu(x) = \sum_f \bar{\Psi}^f(x)\gamma_\mu \Psi^f(x). \tag{5.1.1}$$

The Lagrange density of the standard model is invariant under global $U(1)_B$ baryon number transformations. The corresponding Noether current J_μ is hence conserved at the classical level

$$\partial^\mu J_\mu(x) = 0. \tag{5.1.2}$$

At the quantum level, however, the symmetry cannot be maintained because it is violated by the Adler-Bell-Jackiw anomaly

$$\partial^\mu J_\mu(x) = N_g P(x). \tag{5.1.3}$$

Here N_g is the number of generations, and

$$P(x) = -\frac{1}{32\pi^2}\varepsilon^{\mu\nu\rho\sigma}\mathrm{Tr}(W_{\mu\nu}(x)W_{\rho\sigma}(x)) \tag{5.1.4}$$

is the so-called Chern-Pontryagin density. Here $W_{\mu\nu}$ is the field strength tensor of the $SU(2)_L$ gauge field.

Let us also consider the flavor-singlet axial current in QCD

$$J^5_\mu(x) = \sum_f \bar{\Psi}^f(x)\gamma_5\gamma_\mu\Psi^f(x). \tag{5.1.5}$$

Again, the Lagrange density of QCD with massless quarks is invariant under global $U(1)_A$ transformations, and hence $J^5_\mu(x)$ is conserved at the classical level

$$\partial^\mu J^5_\mu(x) = 0. \tag{5.1.6}$$

However, at the quantum level the symmetry is again explicitly broken by an anomaly

$$\partial^\mu J^5_\mu(x) = 2N_f P(x). \tag{5.1.7}$$

Now N_f is the number of quark flavors, and

$$P(x) = -\frac{1}{32\pi^2}\varepsilon^{\mu\nu\rho\sigma}\mathrm{Tr}(G_{\mu\nu}(x)G_{\rho\sigma}(x)) \tag{5.1.8}$$

now is the Chern-Pontryagin density of the gluon field.

In the following we consider the topology of a general non-Abelian gauge potential G_μ. The anomaly equation can be derived in perturbation theory and it follows from a triangle diagram. The Chern-Pontryagin density can be written as a total divergence

$$P(x) = \partial^\mu\Omega^{(0)}_\mu(x), \tag{5.1.9}$$

where $\Omega^{(0)}_\mu(x)$ is the so-called Chern-Simons density or 0-cochain, which is given by

$$\Omega^{(0)}_\mu(x) = -\frac{1}{8\pi^2}\varepsilon^{\mu\nu\rho\sigma}\mathrm{Tr}[G_\nu(x)(\partial_\rho G_\sigma(x) + \frac{2}{3}G_\rho(x)G_\sigma(x))]. \tag{5.1.10}$$

It is a good exercise to convince oneself that this satisfies eq.(5.1.9). We can now formally construct a conserved current

$$\tilde{J}^5_\mu(x) = J^5_\mu(x) - 2N_f\Omega^{(0)}_\mu(x), \tag{5.1.11}$$

because then

$$\partial^\mu \tilde{J}_\mu(x) = \partial^\mu J_\mu(x) - 2N_f P(x) = 0. \qquad (5.1.12)$$

One might think that we have found a new $U(1)$ symmetry which is free of the anomaly. This is, however, not the case, because the current $\tilde{J}_\mu(x)$ contains $\Omega_\mu^{(0)}(x)$ which is not gauge invariant. Although the gauge variant current is formally conserved, this has no gauge invariant physical consequences.

5.2 The Topological Charge

For the rest of this chapter we will leave Minkowski space-time and Wick rotate ourselves into a Euclidean world with an imaginary (or Euclidean) time. It is then difficult to interpret space-time processes, because we have to perform an analytic continuation to make contact with the real world. Still, it is mathematically advantageous to use Euclidean time, and physical results like particle masses remain unaffected when we go back to Minkowski space-time. From now on we stop distinguishing co- and contravariant indices and we write the topological charge as

$$\begin{aligned} Q &= -\frac{1}{32\pi^2} \int d^4x \; \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}(G_{\mu\nu}(x) G_{\rho\sigma}(x)) = \int d^4x \; P(x) \\ &= \int d^4x \; \partial_\mu \Omega_\mu^{(0)}(x) = \int_{S^3} d^3\sigma_\mu \; \Omega_\mu^{(0)}(x). \end{aligned} \qquad (5.2.1)$$

We have used Gauss' law to reduce the integral over Euclidean space-time to an integral over its boundary, which is topologically a 3-sphere S^3. We will restrict ourselves to gauge field configurations with a finite action. Hence, their field strength should vanish at infinity, and consequently the gauge potential should then be a pure gauge (a gauge transformation of a zero field)

$$G_\mu(x) = g(x) \partial_\mu g(x)^\dagger. \qquad (5.2.2)$$

Of course, this expression is only valid at space-time infinity. Inserting it in the expression for the 0-cochain we obtain

$$
\begin{aligned}
Q &= -\frac{1}{8\pi^2}\int_{S^3} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(g(x)\partial_\nu g(x)^\dagger)(\partial_\rho(g(x)\partial_\sigma g(x)^\dagger) \\
&+ \frac{2}{3}(g(x)\partial_\rho g(x)^\dagger)(g(x)\partial_\sigma g(x)^\dagger))] \\
&= -\frac{1}{8\pi^2}\int_{S^3} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma} \\
&\times \mathrm{Tr}[-(g(x)\partial_\nu g(x)^\dagger)(g(x)\partial_\rho g(x)^\dagger)(g(x)\partial_\sigma g(x)^\dagger) \\
&+ \frac{2}{3}(g(x)\partial_\nu g(x)^\dagger)(g(x)\partial_\rho g(x)^\dagger)(g(x)\partial_\sigma g(x)^\dagger)] \\
&= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(g(x)\partial_\nu g(x)^\dagger)(g(x)\partial_\rho g(x)^\dagger)(g(x)\partial_\sigma g(x)^\dagger)].
\end{aligned}
\tag{5.2.3}
$$

The gauge transformation $g(x)$ defines a mapping of the sphere S^3 at space-time infinity to the gauge group $SU(N)$

$$g : S^3 \to SU(N). \tag{5.2.4}$$

Such mappings have topological properties. They fall into equivalence classes — the so-called homotopy classes — which represent topologically distinct sectors. Two mappings are equivalent if they can be deformed continuously into one another. The homotopy properties are described by so-called homotopy groups. In our case the relevant homotopy group is

$$\Pi_3[SU(N)] = \mathbb{Z}. \tag{5.2.5}$$

Here the index 3 indicates that we consider mappings of the 3-dimensional sphere S^3. The third homotopy group of $SU(N)$ is given by the integers. This means that for each integer Q there is a class of mappings that can be continuously deformed into one another, while mappings with different Q are topologically distinct. The integer Q that characterizes the mapping topologically is the topological charge. Now we want to show that the above expression for Q is exactly that integer. For this purpose we decompose

$$
g = VW,\ W = \left(\begin{array}{ccccc}
1 & 0 & 0 & ... & 0 \\
0 & \tilde{g}_{11} & \tilde{g}_{12} & ... & \tilde{g}_{1N} \\
0 & \tilde{g}_{21} & \tilde{g}_{22} & ... & \tilde{g}_{1N} \\
. & . & . & & . \\
. & . & . & & . \\
0 & \tilde{g}_{N1} & \tilde{g}_{N2} & ... & \tilde{g}_{NN}
\end{array}\right), \tag{5.2.6}
$$

where the embedded matrix $\tilde{g}$ is in $SU(N-1)$. It is indirectly defined by

$$V = \begin{pmatrix} g_{11} & -g_{21}^* & -\frac{g_{31}^*(1+g_{11})}{1+g_{11}^*} & \dots & -\frac{g_{N1}^*(1+g_{11})}{1+g_{11}^*} \\ g_{21} & \frac{1+g_{11}^*-|g_{21}|^2}{1+g_{11}} & -\frac{g_{31}^* g_{21}}{1+g_{11}^*} & \dots & -\frac{g_{N1}^* g_{21}}{1+g_{11}^*} \\ g_{31} & -\frac{g_{21}^* g_{31}}{1+g_{11}} & \frac{1+g_{11}^*-|g_{31}|^2}{1+g_{11}^*} & \dots & -\frac{g_{N1}^* g_{31}}{1+g_{11}^*} \\ . & . & . & & . \\ . & . & . & & . \\ g_{N1} & -\frac{g_{21}^* g_{N1}}{1+g_{11}} & -\frac{g_{31}^* g_{N1}}{1+g_{11}^*} & \dots & \frac{1+g_{11}^*-|g_{N1}|^2}{1+g_{11}^*} \end{pmatrix} \in SU(N). \quad (5.2.7)$$

The matrix V is constructed entirely from the elements $g_{11}, g_{21}, ..., g_{N1}$ of the first column of the matrix g. One should convince oneself that V is indeed an $SU(N)$ matrix, and that the resulting matrix $\tilde{g}$ is indeed in $SU(N-1)$. The idea now is to reduce the expression for the topological charge from $SU(N)$ to $SU(N-1)$ by using the formula

$$\begin{aligned} &\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(VW)\partial_\nu(VW)^\dagger(VW)\partial_\rho(VW)^\dagger(VW)\partial_\sigma(VW)^\dagger] = \\ &\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(V\partial_\nu V^\dagger)(V\partial_\rho V^\dagger)(V\partial_\sigma V^\dagger) \\ &+\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(W\partial_\nu W^\dagger)(W\partial_\rho W^\dagger)(W\partial_\sigma W^\dagger)] \\ &+3\partial_\nu\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[((V\partial_\rho V^\dagger)(W\partial_\sigma W^\dagger)]. \end{aligned} \quad (5.2.8)$$

Again, it is instructive to prove this formula. Applying the formula to the expression for the topological charge and using $g = VW$ we obtain

$$\begin{aligned} Q &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(g(x)\partial_\nu g(x)^\dagger)(g(x)\partial_\rho g(x)^\dagger)(g(x)\partial_\sigma g(x)^\dagger)] \\ &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(V(x)\partial_\nu V(x)^\dagger)(V(x)\partial_\rho V(x)^\dagger)(V(x)\partial_\sigma V(x)^\dagger) \\ &+ (W(x)\partial_\nu W(x)^\dagger)(W(x)\partial_\rho W(x)^\dagger)(W(x)\partial_\sigma W(x)^\dagger)]. \end{aligned} \quad (5.2.9)$$

The ∂_ν term of the formula eq.(5.2.8) drops out using Gauss' law together with the fact that S^3 has no boundary. It follows that the topological charge of a product of two gauge transformations V and W is the sum of the topological charges of V and W. Since V only depends on $g_{11}, g_{21}, ..., g_{N1}$, it can be viewed as a mapping of S^3 into the sphere S^{2N-1}

$$V : S^3 \to S^{2N-1}. \quad (5.2.10)$$

This is because $|g_{11}|^2 + |g_{21}|^2 + ... + |g_{N1}|^2 = 1$. Remarkably the corresponding homotopy group is trivial for $N > 2$, i.e.

$$\Pi_3[S^{2N-1}] = \{0\}. \quad (5.2.11)$$

All mappings of S^3 into the higher dimensional sphere S^{2N-1} are topologically equivalent (they can be deformed into each other). This can be understood better in a lower dimensional example

$$\Pi_1[S^2] = \{0\}. \tag{5.2.12}$$

Each closed curve on an ordinary sphere can be constricted to the north pole, and hence is topologically trivial. In fact,

$$\Pi_m[S^n] = \{0\}, \tag{5.2.13}$$

for $m < n$, while

$$\Pi_n[S^n] = \mathbb{Z}. \tag{5.2.14}$$

Still, $\Pi_m[S^n]$ with $m > n$ is not necessarily trivial, for example

$$\Pi_4[S^3] = \mathbb{Z}(2). \tag{5.2.15}$$

Since the mapping V is topologically trivial its contribution to the topological charge vanishes. The remaining W term reduces to the $SU(N-1)$ contribution

$$Q = \frac{1}{24\pi^2} \int_{S^3} d^3\sigma_\mu \ \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(\tilde{g}(x)\partial_\nu \tilde{g}(x)^\dagger)(\tilde{g}(x)\partial_\rho \tilde{g}(x)^\dagger)(\tilde{g}(x)\partial_\sigma \tilde{g}(x)^\dagger)]. \tag{5.2.16}$$

The separation of the V contribution works only if the decomposition of g into V and $\tilde{g}$ is non-singular. In fact, the expression for V is singular for $g_{11} = -1$. This corresponds to a $((N-1)^2-1)$-dimensional subspace of the (N^2-1)-dimensional $SU(N)$ group space. The mapping g itself covers a 3-d subspace of $SU(N)$. Hence it is arbitrarily improbable to hit a singularity (it is of zero measure). Since we have now reduced the $SU(N)$ topological charge to the $SU(N-1)$ case, we can go down all the way to $SU(2)$. It remains to be shown that the $SU(2)$ expression is actually an integer. First of all

$$\tilde{g} : S^3 \rightarrow SU(2) = S^3, \tag{5.2.17}$$

and indeed

$$\Pi_3[SU(2)] = \Pi_3[S^3] = \mathbb{Z}. \tag{5.2.18}$$

The topological charge specifies how often the $SU(2)$ group space (which is isomorphic to the 3-sphere) is covered by $\tilde{g}$ as we go along the boundary of Euclidean space-time (which is also topologically S^3). Again, it is useful to consider a lower dimensional example, mappings from the circle S^1 to the group $U(1)$ which is topologically also a circle

$$g = \exp(i\varphi) : S^1 \rightarrow U(1) = S^1. \tag{5.2.19}$$

The relevant homotopy group is

$$\Pi_1[U(1)] = \Pi_1[S^1] = \mathbb{Z}. \tag{5.2.20}$$

Again, for each integer there is an equivalence class of mappings that can be continuously deformed into one another. Going over the circle S^1 the mapping may cover the group space $U(1)$ any number of times. In $U(1)$ the expression for the topological charge is analogous to the one in $SU(N)$

$$\begin{aligned} Q &= -\frac{1}{2\pi}\int_{S^1} d\sigma_\mu \varepsilon_{\mu\nu}(g(x)\partial_\nu g(x)^\dagger) = \frac{1}{2\pi}\int_{S^1} d\sigma_\mu \varepsilon_{\mu\nu}\partial_\nu \varphi(x) \\ &= \frac{1}{2\pi}(\varphi(2\pi) - \varphi(0)). \end{aligned} \tag{5.2.21}$$

If $g(x)$ is continuous over the circle $\varphi(2\pi)$ and $\varphi(0)$ must differ by 2π times an integer. That integer is the topological charge. It counts how many times the mapping g covers the group space $U(1)$ as we move along the circle S^1. We are looking for an analogous expression in $SU(2)$. For this purpose we parametrize the mapping $\tilde{g}$ as

$$\begin{aligned} &\tilde{g}(x) = \exp(i\vec{\alpha}(x)\cdot\vec{\sigma}) = \cos\alpha(x) + i\sin\alpha(x)\vec{e}_\alpha(x)\cdot\vec{\sigma}, \\ &\vec{e}_\alpha(x) = (\sin\theta(x)\sin\varphi(x), \sin\theta(x)\cos\varphi(x), \cos\theta(x)). \end{aligned} \tag{5.2.22}$$

It is a good exercise to convince oneself that

$$\begin{aligned} &\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(\tilde{g}(x)\partial_\nu\tilde{g}(x)^\dagger)(\tilde{g}(x)\partial_\rho\tilde{g}(x)^\dagger)(\tilde{g}(x)\partial_\sigma\tilde{g}(x)^\dagger)] \\ &= 12\sin^2\alpha(x)\sin\theta(x)\varepsilon_{\mu\nu\rho\sigma}\partial_\nu\alpha(x)\partial_\rho\theta(x)\partial_\sigma\varphi(x). \end{aligned} \tag{5.2.23}$$

This is exactly the volume element of a 3-sphere (and hence of the $SU(2)$ group space). Thus we can now write

$$Q = \frac{1}{2\pi^2}\int_{S^3} d^3\sigma_\mu \sin^2\alpha(x)\sin\theta(x)\varepsilon_{\mu\nu\rho\sigma}\partial_\nu\alpha(x)\partial_\rho\theta(x)\partial_\sigma\varphi(x) = \frac{1}{2\pi^2}\int_{S^3} d\tilde{g}. \tag{5.2.24}$$

The volume of the 3-sphere is given by $2\pi^2$. When the mapping $\tilde{g}$ covers the sphere Q times the integral gives Q times the volume of S^3. This finally explains why the prefactor $1/32\pi^2$ was introduced in the original expression for the topological charge.

5.3 Topology of a Gauge Field on a Compact Manifold

Imagine our Universe was closed both in space and time, and hence had no boundary. Our previous discussion, for which the value of the gauge field at the boundary was essential, would suggest that in a closed Universe the topology is trivial. On the other hand, we think that topology has local consequences. For example, baryon number is violated because the topological charge does not vanish. To resolve this apparent contradiction we will now discuss the topology of a gauge field on a compact Euclidean space-time manifold M, and we will see that nontrivial topology is still present. Let us again consider the topological charge

$$Q = \int_M d^4x \ P(x). \tag{5.3.1}$$

Writing the Chern-Pontryagin density as the total divergence of the 0-cochain

$$P(x) = \partial_\mu \Omega_\mu^{(0)}(x), \tag{5.3.2}$$

and using Gauss' law we obtain

$$Q = \int_M d^4x \ \partial_\mu \Omega_\mu^{(0)}(x) = \int_{\partial M} d^3\sigma_\mu \Omega_\mu^{(0)}(x) = 0. \tag{5.3.3}$$

Here we have used that M has no boundary, i.e. ∂M is an empty set. A gauge field whose Chern-Pontryagin density can globally be written as a total divergence is indeed topologically trivial on a compact manifold. The important observation is that eq.(5.3.2) may be valid only locally. In other words, gauge singularities may prevent us from using Gauss' law as we did above. In general, it will be impossible to work in a gauge that makes the gauge field nonsingular everywhere on the space-time manifold. Instead we must subdivide space-time into local patches in which the gauge field is smooth, and glue the patches together by nontrivial gauge transformations, which form a fibre bundle of transition functions. A topologically nontrivial gauge field will contain singularities at some points $x_i \in M$. We cover the manifold M by closed sets c_i such that $x_i \in c_i \backslash \partial c_i$, i.e. each singularity lies in the interior of a set c_i. Also $M = \cup_i c_i$ with $c_i \cap c_j = \partial c_i \cap \partial c_j$.

The next step is to remove the gauge singularities x_i by performing gauge transformations g_i in each local patch

$$G_\mu^i(x) = g_i(x)(G_\mu(x) + \partial_\mu) g_i^\dagger(x). \tag{5.3.4}$$

After the gauge transformation the gauge potential $G_\mu^i(x)$ is free of singularities in the local region c_i. Hence we can now use Gauss' law and obtain

$$\begin{aligned} Q &= \sum_i \int_{c_i} d^4x \ P(x) = \sum_i \int_{\partial c_i} d^3\sigma_\mu \ \Omega_\mu^{(0)}(i) \\ &= \frac{1}{2} \sum_{ij} \int_{c_i \cap c_j} d^3\sigma_\mu \ [\Omega_\mu^{(0)}(i) - \Omega_\mu^{(0)}(j)]. \end{aligned} \tag{5.3.5}$$

The argument i of the 0-cochain indicates that we are in the region c_i. At the intersection of two regions $c_i \cap c_j$ the gauge field G^i_μ differs from G^j_μ, although the original gauge field $G_\mu(x)$ was continuous there. In fact, the two gauge fields are related by a gauge transformation v_{ij}

$$G^i_\mu(x) = v_{ij}(x)(G^j_\mu(x) + \partial_\mu)v_{ij}(x)^\dagger, \tag{5.3.6}$$

which is defined only on $c_i \cap c_j$. The gauge transformations v_{ij} form a fibre bundle of transition functions given by

$$v_{ij}(x) = g_i(x)g_j(x)^\dagger. \tag{5.3.7}$$

This equation immediately implies a consistency equation. This so-called cocycle condition relates the transition functions in the intersection $c_i \cap c_j \cap c_k$ of three regions

$$v_{ik}(x) = v_{ij}(x)v_{jk}(x). \tag{5.3.8}$$

The above difference of two 0-cochains in different gauges is given by the so-called coboundary operator Δ

$$\Delta\Omega^{(0)}_\mu(i,j) = \Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j). \tag{5.3.9}$$

It is straight forward to show that

$$\begin{aligned}\Delta\Omega^{(0)}_\mu(i,j) = &-\frac{1}{24\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^\dagger v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)\partial_\sigma v_{ij}(x)^\dagger] \\ &-\frac{1}{8\pi^2}\varepsilon_{\mu\nu\rho\sigma}\partial_\nu \mathrm{Tr}[\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)G^i_\sigma(x)].\end{aligned} \tag{5.3.10}$$

The above equation for the topological charge then takes the form

$$\begin{aligned}Q &= -\frac{1}{48\pi^2}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma} \\ &\times \mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^\dagger v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)\partial_\sigma v_{ij}(x)^\dagger] \\ &- \frac{1}{16\pi^2}\sum_{ij}\int_{\partial(c_i\cap c_j)} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)G^i_\sigma(x)].\end{aligned} \tag{5.3.11}$$

Using the cocycle condition this can be rewritten as

$$\begin{aligned}Q &= -\frac{1}{48\pi^2}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma} \\ &\times \mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^\dagger v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)\partial_\sigma v_{ij}(x)^\dagger] \\ &-\frac{1}{48\pi^2}\sum_{ijk}\int_{c_i\cap c_j\cap c_k} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{jk}(x)\partial_\rho v_{jk}(x)^\dagger].\end{aligned} \tag{5.3.12}$$

This shows that the topology of the fibre bundle is entirely encoded in the transition functions.

In the appropriate mathematical language the gauge transformations g_i form sections of the fibre bundle. Using formula (5.2.8) together with eq.(5.3.7) one can show that the topological charge is expressed in terms of the section in the following way

$$\begin{aligned} Q &= \sum_i Q_i \\ &= \frac{1}{24\pi^2} \sum_i \int_{\partial c_i} d^3\sigma_\mu \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[g_i(x)\partial_\nu g_i(x)^\dagger g_i(x)\partial_\rho g_i(x)^\dagger g_i(x)\partial_\sigma g_i(x)^\dagger]. \end{aligned} \tag{5.3.13}$$

We recognize the integer winding number Q_i that characterizes the mapping g_i topologically. In fact, the boundary ∂c_i is topologically a 3-sphere, such that

$$g_i : \partial c_i \to SU(3), \tag{5.3.14}$$

and hence

$$Q_i \in \Pi_3[SU(3)] = \mathbb{Z}. \tag{5.3.15}$$

The topological charge Q is a sum of local winding numbers $Q_i \in \mathbb{Z}$, which are associated with the regions c_i. In general, the Q_i are not gauge invariant. Hence, individually they have no physical meaning. Still, the total charge — as the sum of all Q_i — is gauge invariant. It is instructive to show this explicitly by performing a gauge transformation on the original gauge field

$$G_\mu(x)' = g(x)(G_\mu(x) + \partial_\mu)g(x)^\dagger. \tag{5.3.16}$$

Deriving the gauge transformation properties of the section and using formula (5.2.8) this is again straightforward.

5.4 Cochain Reduction in $SU(2)$

To complete the mathematical investigation of the topological charge let us finally investigate its expression eq.(5.3.12) in terms of transition functions. To understand why this expression also is an integer we concentrate on $SU(2)$ and we go back to the equation

$$Q = \sum_i \int_{\partial c_i} d^3\sigma_\mu \Omega_\mu^{(0)}(i) = \frac{1}{2!} \sum_{ij} \int_{c_i \cap c_j} d^3\sigma_\mu [\Omega_\mu^{(0)}(i) - \Omega_\mu^{(0)}(j)]. \tag{5.4.1}$$

The coboundary operator

$$\Delta\Omega^{(0)}_\mu(i,j) = \Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j) = \partial_\nu \Omega^{(1)}_{\mu\nu}(i,j) \tag{5.4.2}$$

can be written as a total divergence of the 1-cochain

$$\begin{aligned} \Omega^{(1)}_{\mu\nu}(i,j) &= -\frac{1}{8\pi^2}(\alpha - \sin\alpha\cos\alpha)\varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha \cdot (\partial_\rho \vec{e}_\alpha \times \partial_\sigma \vec{e}_\alpha) \\ &- \frac{1}{8\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[\partial_\rho v_{ij} v^\dagger_{ij} G^i_\sigma]. \end{aligned} \tag{5.4.3}$$

Here we have parametrized the $SU(2)$ transition function as

$$v_{ij} = \exp(i\vec{\alpha}\cdot\vec{\sigma}) = \cos\alpha + i\sin\alpha\vec{e}_\alpha\cdot\vec{\sigma},\ \alpha \in [0,\pi]. \tag{5.4.4}$$

When $v_{ij} = -1$, i.e. when $\alpha = \pi$, the above parametrization is singular because then the unit vector $\vec{e}_\alpha$ is not well defined. The singularities $v_{ij} = -1$ occur at isolated points $x \in c_i \cap c_j$. In their neighborhood the 1-cochain has a directional singularity

$$\Omega^{(1)}_{\mu\nu}(i,j) = -\frac{1}{8\pi}\varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha \cdot (\partial_\rho \vec{e}_\alpha \times \partial_\sigma \vec{e}_\alpha). \tag{5.4.5}$$

Using Gauss' law the topological charge can now be written as

$$Q = Q^{(1)} + Q^{(1)}_\Sigma, \tag{5.4.6}$$

where

$$Q^{(1)} = \frac{1}{2!}\sum_{ij}\sum_{x\in c_i\cap c_j}\frac{1}{8\pi}\int_{S^2_\varepsilon(x)} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha \cdot (\partial_\rho \vec{e}_\alpha \times \partial_\sigma \vec{e}_\alpha). \tag{5.4.7}$$

Here $S^2_\varepsilon(x)$ is a 2-sphere of infinitesimal radius ε around the singularity x. It appears as an internal boundary in the application of Gauss' law because the integrand is singular. Performing the 2-d integral gives

$$\frac{1}{8\pi}\int_{S^2_\varepsilon(x)} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha \cdot (\partial_\rho \vec{e}_\alpha \times \partial_\sigma \vec{e}_\alpha) = n^{(1)}(x;i,j), \tag{5.4.8}$$

where $n^{(1)}(x;i,j) \in \mathbb{Z}$ is a local winding number associated with the singularity. It is an element of $\Pi_2[S^2] = \mathbb{Z}$ and it counts how often the unit vector $\vec{e}_\alpha$ covers S^2 as we integrate over $S^2_\varepsilon(x)$. Hence

$$Q^{(1)} = \sum_{\Lambda^{(1)}} n^{(1)}(x;i,j) \tag{5.4.9}$$

is a sum of local winding numbers associated with the singular points that form the set

$$\Lambda^{(1)} = \cup_{ij}\{x \in c_i \cap c_j\}. \qquad (5.4.10)$$

The remaining contribution to the topological charge is given by

$$\begin{aligned} Q_\Sigma^{(1)} &= \frac{1}{2!}\sum_{ij}\int_{\partial(c_i\cap c_j)} d^2\sigma_{\mu\nu}\ \Omega_{\mu\nu}^{(1)}(i,j) \\ &= \frac{1}{3!}\sum_{ijk}\int_{c_i\cap c_j\cap c_k} d^2\sigma_{\mu\nu}\ [\Omega_{\mu\nu}^{(1)}(i,j) - \Omega_{\mu\nu}^{(1)}(i,k) + \Omega_{\mu\nu}^{(1)}(j,k)]. \end{aligned} \qquad (5.4.11)$$

We identify the coboundary operator

$$\Delta\Omega_{\mu\nu}^{(1)}(i,j,k) = \Omega_{\mu\nu}^{(1)}(i,j) - \Omega_{\mu\nu}^{(1)}(i,k) + \Omega_{\mu\nu}^{(1)}(j,k) = \partial_\rho\Omega_{\mu\nu\rho}^{(2)}(i,j,k), \qquad (5.4.12)$$

which again is a total divergence. The 2-cochain is given by

$$\begin{aligned} &\Omega_{\mu\nu\rho}^{(2)}(i,j,k) = \\ &-\frac{1}{8\pi^2}\varepsilon_{\mu\nu\rho\sigma}(1 + 2\cos\alpha\cos\beta\cos\gamma - \cos^2\alpha - \cos^2\beta - \cos^2\gamma)^{-1} \\ &\times\{(\vec\alpha + \vec\beta - \vec\gamma)\cdot\sin\alpha\vec e_\alpha[\partial_\sigma(\sin\beta\vec e_\beta)\cdot\sin\gamma\vec e_\gamma - \sin\beta\vec e_\beta\cdot\partial_\sigma(\sin\gamma\vec e_\gamma)] \\ &+(\vec\alpha + \vec\beta - \vec\gamma)\cdot\sin\beta\vec e_\beta[\partial_\sigma(\sin\gamma\vec e_\gamma)\cdot\sin\alpha\vec e_\alpha - \sin\gamma\vec e_\gamma\cdot\partial_\sigma(\sin\alpha\vec e_\alpha)] \\ &+(\vec\alpha + \vec\beta - \vec\gamma)\cdot\sin\gamma\vec e_\gamma[\partial_\sigma(\sin\alpha\vec e_\alpha)\cdot\sin\beta\vec e_\beta - \sin\alpha\vec e_\alpha\cdot\partial_\sigma(\sin\beta\vec e_\beta)]\}. \end{aligned} \qquad (5.4.13)$$

Here we have parametrized the transition functions as

$$v_{ij} = \exp(i\vec\alpha\cdot\sigma),\ v_{jk} = \exp(i\vec\beta\cdot\sigma),\ v_{ik} = \exp(i\vec\gamma\cdot\sigma). \qquad (5.4.14)$$

The cocycle condition $v_{ik} = v_{ij}v_{jk}$ then takes the form

$$\begin{aligned} \cos\gamma &= \cos\alpha\cos\beta - \sin\alpha\sin\beta\vec e_\alpha\cdot\vec e_\beta, \\ \sin\gamma\vec e_\gamma &= \sin\alpha\cos\beta\vec e_\alpha + \cos\alpha\sin\beta\vec e_\beta - \sin\alpha\sin\beta\vec e_\alpha\times\vec e_\beta. \end{aligned} \qquad (5.4.15)$$

These equations define a spherical triangle on the sphere S^3 with side lengths α, β and γ.

The 2-cochain also has singularities. They occur when the spherical triangle defined by the three transition functions degenerates to a great circle, i.e. when

$$(\vec\alpha + \vec\beta - \vec\gamma)\cdot\vec e_\alpha = (\vec\alpha + \vec\beta - \vec\gamma)\cdot\vec e_\beta = -(\vec\alpha + \vec\beta - \vec\gamma)\cdot\vec e_\gamma = 2\pi. \qquad (5.4.16)$$

In the neighborhood of such points the 2-cochain has a directional singularity

$$\Omega^{(2)}_{\mu\nu\rho}(i,j,k) = -\frac{1}{4\pi^2}\varepsilon_{\mu\nu\rho\sigma}(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot(\vec{e}_s\times\partial_\sigma\vec{e}_s). \quad (5.4.17)$$

Here we have defined another unit vector

$$\vec{e}_s = \cos\alpha\sin\beta\sin\gamma\vec{e}_\beta\times\vec{e}_\gamma + \cos\beta\sin\gamma\sin\alpha\vec{e}_\gamma\times\vec{e}_\alpha + \cos\gamma\sin\alpha\sin\beta\vec{e}_\alpha\times\vec{e}_\beta. \quad (5.4.18)$$

Using Gauss' law we now obtain

$$Q^{(1)}_\Sigma = Q^{(2)} + Q^{(2)}_\Sigma, \quad (5.4.19)$$

with

$$Q^{(2)} = \frac{1}{3!}\sum_{ijk}\sum_{x\in c_i\cap c_j\cap c_k}\frac{1}{4\pi^2}\int_{S^1_\varepsilon(x)} d\sigma_{\mu\nu\rho}\varepsilon_{\mu\nu\rho\sigma}(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot(\vec{e}_s\times\partial_\sigma\vec{e}_s). \quad (5.4.20)$$

Now $S^1_\varepsilon(x)$ is an infinitesimal circle around the singularity. Performing the integral yields

$$\frac{1}{4\pi^2}\int_{S^1_\varepsilon(x)} d\sigma_{\mu\nu\rho}\varepsilon_{\mu\nu\rho\sigma}(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot(\vec{e}_s\times\partial_\sigma\vec{e}_s) = n^{(2)}(x;i,j,k), \quad (5.4.21)$$

where $n^{(2)}(x;i,j,k)$ is a winding number from $\Pi_1[S^1] = \mathbb{Z}$, which counts how often the unit vector $\vec{e}_s$ (that is perpendicular to $\vec{\alpha}+\vec{\beta}-\vec{\gamma}$) covers the circle S^1 as we integrate over $S^1_\varepsilon(x)$. Hence we obtain

$$Q^{(2)} = \sum_{\Lambda^{(2)}} n^{(2)}(x;i,j,k), \quad (5.4.22)$$

where the set of singular points is now denoted by

$$\Lambda^{(2)} = \cup_{ijk}\{x\in c_i\cap c_j\cap c_k\}. \quad (5.4.23)$$

The remaining contributions to the topological charge take the form

$$\begin{aligned} Q^{(2)}_\Sigma &= \frac{1}{3!}\sum_{ijk}\int_{\partial(c_i\cap c_j\cap c_k)} d\sigma_{\mu\nu\rho}\Omega^{(2)}_{\mu\nu\rho}(i,j,k) \\ &= \frac{1}{4!}\sum_{ijkl}\int_{c_i\cap c_j\cap c_k\cap c_l} d\sigma_{\mu\nu\rho}[\Omega^{(2)}_{\mu\nu\rho}(i,j,k) - \Omega^{(2)}_{\mu\nu\rho}(i,j,l) \\ &+ \Omega^{(2)}_{\mu\nu\rho}(i,k,l) - \Omega^{(2)}_{\mu\nu\rho}(j,k,l)]. \end{aligned} \quad (5.4.24)$$

It comes as no surprise that the coboundary operator

$$\begin{aligned}\Delta\Omega^{(2)}_{\mu\nu\rho}(i,j,k,l) &= \Omega^{(2)}_{\mu\nu\rho}(i,j,k) - \Omega^{(2)}_{\mu\nu\rho}(i,j,l) + \Omega^{(2)}_{\mu\nu\rho}(i,k,l) - \Omega^{(2)}_{\mu\nu\rho}(j,k,l) \\ &= \partial_\sigma \Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l) \qquad (5.4.25)\end{aligned}$$

yields the total divergence of a 3-cochain. It is remarkable that the 3-cochain has a geometric meaning. It is given by the volume $V(i,j,k,l)$ of the spherical tetrahedron consisting of four spherical triangles defined by the participating transition functions. It is quite tedious to show that

$$\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l) = \frac{1}{2\pi^2}\varepsilon_{\mu\nu\rho\sigma} V(i,j,k,l). \qquad (5.4.26)$$

The key observation is that the variation of the volume of the spherical tetrahedron is given by Schläfli's differential form

$$\partial_\sigma V(i,j,k,l) = \frac{1}{2}(\alpha\partial_\sigma A + \beta\partial_\sigma B + \gamma\partial_\sigma \Gamma + \delta\partial_\sigma \Delta + \varepsilon\partial_\sigma E + \zeta\partial_\sigma Z), \qquad (5.4.27)$$

where $A, B, ..., Z$ are the angles between the faces of the spherical tetrahedron defined by the transition functions. Schläfli was a mathematician in Bern (Switzerland) around the time when Einstein worked there in the patent office. We can now perform the final integration and get

$$Q^{(2)}_\Sigma = Q^{(3)} + Q^{(3)}_\Sigma. \qquad (5.4.28)$$

The first contribution has the form

$$Q^{(3)} = \sum_{\Lambda^{(3)}} n^{(3)}(x; i,j,k,l), \qquad (5.4.29)$$

where

$$\Lambda^{(3)} = \cup_{ijkl}\{x \in c_i \cap c_j \cap c_k \cap c_l\} \qquad (5.4.30)$$

is a set of singular points at which the surface of the spherical tetrahedron degenerates to a 2-sphere. At the singularities the volume of the spherical tetrahedron changes by an integer times $2\pi^2$ — the volume of the sphere S^3. The local winding number

$$n^{(3)}(x; i,j,k,l) = \frac{1}{2\pi^2}\Delta V(i,j,k,l) \qquad (5.4.31)$$

measures the change in volume $\Delta V(i,j,k,l)$ at the singularity. The remaining contribution takes the form

$$\begin{aligned}
Q_\Sigma^{(3)} &= \frac{1}{4!}\sum_{\partial(c_i\cap c_j\cap c_k\cap c_l)} \sigma_{\mu\nu\rho\sigma}\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l) \\
&= \frac{1}{5!}\sum_{c_i\cap c_j\cap c_k\cap c_l\cap c_m} \sigma_{\mu\nu\rho\sigma}[\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l) - \Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,m) \\
&+ \Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,l,m) - \Omega^{(3)}_{\mu\nu\rho\sigma}(i,k,l,m) + \Omega^{(3)}_{\mu\nu\rho\sigma}(j,k,l,m)] \\
&= \frac{1}{5!}\sum_{c_i\cap c_j\cap c_k\cap c_l\cap c_m} \sigma\frac{1}{2\pi^2}[V(i,j,k,l) - V(i,j,k,m) \\
&+ V(i,j,l,m) - V(i,k,l,m) + V(j,k,l,m)]. \qquad (5.4.32)
\end{aligned}$$

Here $\sigma = \sigma_{\mu\nu\rho\sigma}\varepsilon_{\mu\nu\rho\sigma}$ determines the orientation of $\partial(c_i \cap c_j \cap c_k \cap c_l)$. The intersection of four 4-dimensional regions determines a point x. All these points form a set

$$\Lambda^{(4)} = \{x \in \cup_{ijklm} c_i \cap c_j \cap c_k \cap c_l \cap c_m\}, \qquad (5.4.33)$$

such that

$$Q_\Sigma^{(3)} = Q^{(4)} = \sum_{\Lambda^{(4)}} n^{(4)}(x;i,j,k,l,m) \qquad (5.4.34)$$

with

$$\begin{aligned}
n^{(4)}(x;i,j,k,l,m) &= \sigma\frac{1}{2\pi^2}[V(i,j,k,l) - V(i,j,k,m) + V(i,j,l,m) \\
&- V(i,k,l,m) + V(j,k,l,m)]. \qquad (5.4.35)
\end{aligned}$$

This is always an integer because the five spherical tetrahedra together are compact and cover S^3 a certain number of times — their total volume is an integer multiple of $2\pi^2$. Altogether the topological charge is a sum of local winding numbers of different topological origin

$$\begin{aligned}
Q &= \sum_{\Lambda^{(1)}} n^{(1)}(x;i,j) + \sum_{\Lambda^{(2)}} n^{(2)}(x;i,j,k) \\
&+ \sum_{\Lambda^{(3)}} n^{(3)}(x;i,j,k,l) + \sum_{\Lambda^{(4)}} n^{(4)}(x;i,j,k,l,m). \qquad (5.4.36)
\end{aligned}$$

Again, one should note that the local winding numbers are not gauge invariant. Only their sum — the total topological charge — is gauge invariant and has a physical meaning.

5.5 The Instanton in $SU(2)$

We have argued mathematically that gauge field configurations fall into topologically distinct classes. Now we want to construct concrete examples of topologically nontrivial field configurations. Here we consider instantons, which have $Q = 1$ and are solutions of the Euclidean classical field equations. The instanton occurs at a given instant in Euclidean time. Since these solutions do not live in Minkowski space-time they have no direct interpretation in terms of real time events. Also it is unclear which role they play in the quantum theory. Instantons describe tunneling processes between degenerate classical vacuum states. Their existence gives rise to the θ-vacuum structure of non-Abelian gauge theories.

Here we concentrate on $SU(2)$. This is sufficient, because we have seen that the $SU(N)$ topological charge can be reduced to the $SU(2)$ case. In this section we go back to an infinite space with a boundary sphere S^3, and we demand that the gauge field has finite action. Then at space-time infinity the gauge potential is in a pure gauge

$$G_\mu(x) = g(x)\partial_\mu g(x)^\dagger. \tag{5.5.1}$$

Provided the gauge field is otherwise smooth, the topology resides entirely in the mapping g. We want to construct a field configuration with topological charge $Q = 1$, i.e. one in which the mapping g covers the group space $SU(2) = S^3$ once as we integrate over the boundary sphere S^3. The simplest mapping of this sort is the identity, i.e. each point at the boundary of space-time is mapped into the corresponding point in the group space such that

$$g(x) = \frac{x_0 + i\vec{x}\cdot\vec{\sigma}}{|x|},\ |x| = \sqrt{x_0^2 + |\vec{x}|^2}. \tag{5.5.2}$$

Next we want to extend the gauge field to the interior of space-time without introducing singularities. We cannot simply maintain the form of eq.(5.5.1) because g is singular at $x = 0$. To avoid this singularity we make the ansatz

$$G_\mu(x) = f(|x|)g(x)\partial_\mu g(x)^\dagger, \tag{5.5.3}$$

where $f(\infty) = 1$ and $f(0) = 0$. For any smooth function f with these properties the above gluon field configuration has $Q = 1$. Still, this does not mean that we have constructed an instanton. Instantons are field configurations with $Q \neq 0$ that are in addition solutions of the Euclidean classical equations of motion, i.e. they are minima of the Euclidean action

$$S[G_\mu] = \int d^4x\ \frac{1}{2g^2}\mathrm{Tr}[G_{\mu\nu}(x)G_{\mu\nu}(x)]. \tag{5.5.4}$$

Let us consider the following integral

$$\int d^4x \ \mathrm{Tr}[(G_{\mu\nu}(x) \pm \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}G_{\rho\sigma}(x))(G_{\mu\nu}(x) \pm \frac{1}{2}\varepsilon_{\mu\nu\kappa\lambda}G_{\kappa\lambda}(x)) =$$
$$\int d^4x \ \mathrm{Tr}[G_{\mu\nu}(x)G_{\mu\nu}(x) \pm \varepsilon_{\mu\nu\rho\sigma}G_{\mu\nu}(x)G_{\rho\sigma}(x) + G_{\mu\nu}(x)G_{\mu\nu}(x)] =$$
$$4g_s^2 S[G_\mu] \pm 32\pi^2 Q[G_\mu]. \quad (5.5.5)$$

We have integrated a square. Hence it is obvious that

$$S[G_\mu] \pm \frac{8\pi^2}{g^2}Q[G_\mu] \geq 0 \ \Rightarrow \ S[G_\mu] \geq \frac{8\pi^2}{g^2}|Q[G_\mu]|, \quad (5.5.6)$$

i.e. a topologically nontrivial field configuration costs at least a minimum action proportional to the topological charge. Instantons are configurations with minimum action, i.e. for them

$$S[G_\mu] = \frac{8\pi^2}{g^2}|Q[G_\mu]|. \quad (5.5.7)$$

From the above argument it is clear that a minimum action configuration arises only if

$$G_{\mu\nu}(x) = \pm\frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}G_{\rho\sigma}(x). \quad (5.5.8)$$

Configurations that obey this equation with a plus sign are called selfdual. The ones that obey it with a minus sign are called anti-selfdual. It is instructive to convince oneself that the above gluon field with

$$f(|x|) = \frac{|x|^2}{|x|^2 + \rho^2} \quad (5.5.9)$$

is indeed an instanton for any value of ρ. The instanton configuration hence takes the form

$$G_\mu(x) = \frac{|x|^2}{|x|^2 + \rho^2} g(x)\partial_\mu g(x)^\dagger. \quad (5.5.10)$$

There is a whole family of instantons with different radii ρ. As a consequence of scale invariance of the classical action they all have the same action $S[G_\mu] = 8\pi^2/g^2$.

5.6 θ-Vacua

The existence of topologically nontrivial gauge transformations has drastic consequences for non-Abelian gauge theories. In fact, there is not just one classical vacuum state, but there is one for each topological winding number. Instantons describe tunneling transitions between topologically distinct vacua. Due to tunneling the degeneracy of the classical vacuum states is lifted, and the true quantum vacuum turns out to be a θ-state, i.e. one in which configurations of different winding numbers are mixed.

In the following we fix to $G_4(x) = 0$ gauge, and we consider space to be compactified from $\mathbb{R}^3$ to S^3. This is just a technical trick which makes life easier. Using transition functions one could choose any other compactification, e.g. on a torus T^3, or one could choose appropriate boundary conditions on $\mathbb{R}^3$ itself. The classical vacuum solutions of such a theory are the pure gauge fields

$$G_i(x) = g(x)\partial_i g(x)^\dagger. \tag{5.6.1}$$

Since we have compactified space, the classical vacua can be classified by their winding number

$$n \in \Pi_3[SU(3)] = \mathbb{Z}, \tag{5.6.2}$$

which is given by

$$n = \frac{1}{24\pi^2} \int_{S^3} d^3x \; \varepsilon_{ijk} \mathrm{Tr}[g(x)\partial_i g(x)^\dagger g(x)\partial_j g(x)^\dagger g(x)\partial_k g(x)^\dagger]. \tag{5.6.3}$$

One might think that one can construct a quantum vacuum $|n\rangle$ just by considering small fluctuations around a classical vacuum with given n. Quantum tunneling, however, induces transitions between the various classical vacua. Imagine the system is in a classical vacuum state with winding number m at early times $t = -\infty$, then it changes continuously (now deviating from a pure gauge), and finally at $t = \infty$ it returns to a classical vacuum state with a possibly different winding number n. The time evolution corresponds to one particular path in the Feynman path integral. The corresponding gauge field smoothly interpolates between the initial and final classical vacua. When we calculate its topological charge, we can use Gauss' law, which yields an integral of the 0-cochain over the space-time boundary, which consists of the spheres S^3 at $t = -\infty$ and at $t = \infty$. At each boundary sphere the gauge field is in a pure gauge, and the integral yields the corresponding winding number such that

$$Q = n - m. \tag{5.6.4}$$

Hence, a configuration with topological charge Q induces a transition from a classical vacuum with winding number m to one with winding number $n = m + Q$. In other words, the Feynman path integral that describes the amplitude for transitions from one classical vacuum to another is restricted to field configurations in the topological sector Q, such that

$$\langle n|U(\infty,-\infty)|m\rangle = \int \mathcal{D}G_\mu^{(n-m)} \exp(-S[G_\mu]). \tag{5.6.5}$$

Here $G_\mu^{(Q)}$ denotes a gauge field with topological charge Q, and $U(t',t)$ is the time evolution operator.

It is crucial to note that the winding number n is not gauge invariant. In fact, as we perform a gauge transformation with winding number 1 the winding number of the pure gauge field changes to $n+1$. In the quantum theory such a gauge transformation g is implemented by a unitary operator T that acts on wave functionals $\Psi[G_i]$ by gauge transforming the field G_i, i.e.

$$T\Psi[G_i] = \Psi[g(G_i+\partial_i)g^\dagger]. \tag{5.6.6}$$

In particular, acting on a state that describes small fluctuations around a classical vacuum one finds

$$T|n\rangle = |n+1\rangle, \tag{5.6.7}$$

i.e. T acts as a ladder operator. Since the operator T implements a special gauge transformation, it commutes with the Hamiltonian, just the theory is gauge invariant. This means that the Hamiltonian and T can be diagonalized simultaneously, and each eigenstate can be labelled by an eigenvalue of T. Since T is a unitary operator its eigenvalues are complex phases $\exp(i\theta)$, such that an eigenstate — for example the vacuum — can be written as $|\theta\rangle$ with

$$T|\theta\rangle = \exp(i\theta)|\theta\rangle. \tag{5.6.8}$$

On the other hand, we can construct the θ-vacuum as a linear combination

$$|\theta\rangle = \sum_n c_n|n\rangle. \tag{5.6.9}$$

Using

$$\begin{aligned} T|\theta\rangle &= \sum_n c_n T|n\rangle = \sum_n c_n|n+1\rangle \\ &= \sum_n c_{n-1}|n\rangle = \exp(i\theta)\sum_n c_n|n\rangle, \end{aligned} \tag{5.6.10}$$

one obtains $c_{n-1} = \exp(i\theta)c_n$ such that $c_n = \exp(-in\theta)$ and

$$|\theta\rangle = \sum_n \exp(-in\theta)|n\rangle. \tag{5.6.11}$$

The true vacuum of a non-Abelian gauge theory is a linear combination of classical vacuum states of different winding numbers. For each value of θ there is a corresponding vacuum state. This is analogous to the energy bands in a solid. There a state is labelled by a Bloch momentum as a consequence of the discrete translation symmetry. In non-Abelian gauge theories T induces discrete translations between classical vacua, with analogous mathematical consequences.

Now let us consider the quantum transition amplitude between different θ-vacua

$$\begin{aligned}\langle\theta|U(\infty,-\infty)|\theta'\rangle &= \sum_{m,n}\exp(in\theta)\exp(-im\theta')\langle n|U(\infty,-\infty)|m\rangle \\ &= \sum_{n,Q=n-m}\exp(in\theta - i(n-Q)\theta')\int\mathcal{D}G_\mu^{(Q)}\exp(-S[G_\mu]) \\ &= \delta(\theta-\theta')\sum_Q\int\mathcal{D}G_\mu^{(Q)}\exp(-S[G_\mu])\exp(i\theta Q[G_\mu]) \\ &= \int\mathcal{D}G_\mu\exp(-S_\theta[G_\mu]).\end{aligned} \tag{5.6.12}$$

There is no transition between different θ-vacua, which confirms that they are eigenstates. Also we can again identify the action in a θ-vacuum as

$$S_\theta[G_\mu] = S[G_\mu] - i\theta Q[G_\mu]. \tag{5.6.13}$$

Finally, let us consider the theory with at least one massless fermion. In that case the Dirac operator $\gamma_\mu(G_\mu(x)+\partial_\mu)$ has a zero mode. This follows from an index theorem due to Atiyah and Singer. They considered the eigenvectors of the Dirac operator with zero eigenvalue

$$\gamma_\mu(G_\mu(x)+\partial_\mu)\Psi(x) = 0. \tag{5.6.14}$$

These eigenvectors have a definite handedness, i.e.

$$\frac{1}{2}(1\pm\gamma_5)\Psi(x) = \Psi(x), \tag{5.6.15}$$

because

$$\gamma_5\gamma_\mu(G_\mu(x)+\partial_\mu)\Psi(x) = -\gamma_\mu(G_\mu(x)+\partial_\mu)\gamma_5\Psi(x) = 0. \tag{5.6.16}$$

The Atiyah-Singer index theorem states that

$$Q = n_L - n_R, \tag{5.6.17}$$

where n_L and n_R are the numbers of left- and right-handed zero modes. Hence, a topologically nontrivial gauge field configuration necessarily has at least one zero mode. This zero mode of the Dirac operator eliminates topologically nontrivial field configurations from theories with a massless fermions, i.e. then $Q[G_\mu] = 0$ for all configurations that contribute to the Feynman path integral. In that case the θ-term in the action has no effect, and all θ-vacua would be physically equivalent. This scenario has been suggested as a possible solution of the strong CP problem. If the lightest quark (the u quark) would be massless, θ would not generate an electric dipole moment for the neutron. There is still no agreement on this issue. Some experts of chiral perturbation theory claim that a massless u-quark is excluded by experimental data. However, the situation is not clear. For example, the pion mass depends only on the sum $m_u + m_d$, and one must look at more subtle effects. Most likely the solution of the strong CP problem is beyond the standard model. We will soon discuss extensions of the standard model with an additional $U(1)_{PQ}$ Peccei-Quinn symmetry, which will allow us to rotate θ to zero. As a consequence of spontaneous $U(1)_{PQ}$ breaking, we will also find a new light pseudo-Goldstone boson — the axion.

5.7 The U(1)-Problem

The topological properties of the gluon field give rise to several questions in the standard model. One is the strong CP problem related to the presence of the θ-vacuum angle. A naive hope to avoid this problem might be to assume that gluon field configurations with non-vanishing topological charge are negligible in the QCD path integral. This, however, does not work because there is also the so-called $U(1)$-problem in QCD. The problem is to explain why the η'-meson has a large mass and hence is not a Goldstone boson. This is qualitatively understood based on the Adler-Bell-Jackiw anomaly — the axial $U(1)$ symmetry of QCD is simply explicitly broken. To solve the $U(1)$-problem quantitatively — i.e. to explain the large value of the η'-mass — requires gluon field configurations with non-zero topological charge to appear frequently in the path integral. This is confirmed by lattice calculations and indeed offers a nice explanation of the $U(1)$-problem. However, if we use topologically non-trivial configurations to solve the $U(1)$-problem, we cannot ignore these configurations when we face the strong CP-problem.

The chiral symmetry of the classical QCD Lagrange function is $U(N_f)_L \otimes U(N_f)_R$, while in the spectrum only the flavor and baryon number symmetries $SU(N_f)_{L+R} \otimes U(1)_{L=R} = U(N_f)_{L=R}$ are manifest. According to the Goldstone theorem one might hence expect $N_f^2 + N_f^2 - N_f^2 = N_f^2$ Goldstone bosons, while in fact one finds only $N_f^2 - 1$ Goldstone bosons in QCD. The missing Goldstone boson should be a pseudoscalar, flavorscalar particle. The lightest particle with these quantum numbers is the η'-meson. However, its mass is $M_{\eta'} = 0.958$ GeV, which is far too heavy for a Goldstone boson. The question why the η'-meson is so heavy is the so-called $U(1)$-problem of QCD. At the end the question is why the axial $U(1)$ symmetry is not spontaneously broken, although it is also not manifest in the spectrum. It took a while before 't Hooft realized that axial $U(1)$ is not a symmetry of QCD. Although the symmetry is present in the classical Lagrange density, it cannot be maintained under quantization because it has an anomaly. This explains qualitatively why the η'-meson is not a Goldstone boson. To understand the problem more quantitatively, one must consider the origin of the quantum mechanical symmetry breaking in more detail. It turns out that topologically non-trivial configurations of the gluon field — for example instantons — give mass to the η'-meson. If the color symmetry would be $SU(N_c)$ instead of $SU(3)$, the explicit axial $U(1)$ breaking via the anomaly would disappear in the large N_c limit. In this limit the η'-meson does indeed become a Goldstone boson. For large but finite N_c the η'-meson gets a mass proportional to the topological susceptibility — the vacuum value of the topological charge squared per space-time volume — evaluated in the pure glue theory.

Qualitatively one understands why the η'-meson is not a Goldstone boson, because the axial $U(1)$-symmetry is explicitly broken by the Adler-Bell-Jackiw anomaly

$$\partial_\mu J_\mu^5(x) = 2N_f P(x), \tag{5.7.1}$$

where P is the Chern-Pontryagin density. However, the question arises how strong this breaking really is, and how it affects the η'-mass quantitatively. To understand this issue we consider QCD with a large number of colors, i.e we replace the gauge group $SU(3)$ by $SU(N_c)$.

It is interesting that large N_c QCD is simpler than real QCD, but still it is too complicated to solve it analytically. Still, one can classify the subset of Feynman diagrams that contribute in the large N_c limit. An essential observation is that for many colors the distinction between $SU(N_c)$ and $U(N_c)$ becomes irrelevant. Then each gluon propagator in a Feynman diagram may be replaced formally by the color flow of a quark-antiquark pair. In this way any large N_c QCD diagram can be represented as a quark diagram. For the gluon self-energy diagram, for example, one finds an internal quark loop which yields a color factor N_c and each vertex gives a factor g_s, such that the diagram diverges as $g_s^2 N_c$. We absorb this divergence in a redefinition of the coupling constant by defining

$$g^2 = g_s^2 N_c, \tag{5.7.2}$$

and we perform the large N_c limit such that g_s goes to zero but g remains finite. Let us now consider a planar 2-loop diagram contributing to the gluon self-energy. There are two internal loops and hence there is a factor N_c^2. Also there are four vertices contributing factors $g_s^4 = g^4/N_c^2$ and the whole diagram is proportional to g^4 and hence it is finite. Let us also consider a planar 4-loop diagram. It has a factor N_c^4 together with six 3-gluon vertices that give a factor $g_s^6 = g^6/N_c^3$ and a 4-gluon vertex that gives a factor $g_s^2 = g^2/N_c$. Altogether the diagram is proportional to g^8 and again it is finite as N_c goes to infinity. Next let us consider a non-planar 4-loop diagram. The color flow is such that now there is only one color factor N_c but there is a factor $g_s^6 = g^6/N_c^3$ from the vertices. Hence the total factor is g^6/N_c^2 which vanishes in the large N_c limit. In general any non-planar gluon diagram vanishes in the large N_c limit. Planar diagrams, on the other hand, survive in the limit. In particular, if we add another propagator to a planar diagram such that it remains planar, we add two 3-gluon vertices and hence a factor $g_s^2 = g^2/N_c$, and we cut an existing loop into two pieces, thus introducing an extra loop color factor N_c. The total weight remains of order 1. Now consider the quark contribution to the gluon propagator. There is no color factor N_c for this diagram, and still there are two quark-gluon vertices contributing a factor $g_s^2 = g^2/N_c$. Hence this diagram disappears in the large N_c limit. Similarly, all diagrams with internal quark loops vanish at large N_c. Even though this eliminates a huge class of diagrams, the remaining planar gluon diagrams are still too complicated to be summed up analytically. Still, the above N_c counting allows us to understand some aspects of the QCD dynamics.

In the large N_c limit, QCD reduces to a theory of mesons and glueballs, while the baryons disappear. This can be understood in the constituent quark model. In $SU(N_c)$ a color singlet baryon consists of N_c quarks, each contributing the constituent quark mass to the total baryon mass. Hence the baryon mass is proportional to N_c such that baryons are infinitely heavy (and hence disappear) in the large N_c limit. Mesons, on the other hand, still consist of a quark and an anti-quark, such that their mass remains finite.

Also the topology of the gluon field is affected in the large N_c limit. We have derived the instanton action bound

$$S[G_\mu] \geq \frac{8\pi^2}{g_s^2}|Q[G_\mu]| = \frac{8\pi^2 N_c}{g^2}|Q[G_\mu]|, \tag{5.7.3}$$

which is valid for all $SU(N_c)$. In the large N_c limit the action of an instanton diverges, and topologically non-trivial field configurations are eliminated from the Feynman path integral. This means that the source of quantum mechanical symmetry breaking via the anomaly disappears, and the η'-meson should indeed become a Goldstone boson in the large N_c limit. In that case one should be able to derive a mass formula for the η'-meson just like for the Goldstone pion. The pion mass resulted from an explicit chiral symmetry breaking due to a finite quark mass. Similarly, the η'-mass results from an explicit axial $U(1)$ breaking via the anomaly due to finite N_c. This can be computed as a $1/N_c$ effect.

Let us consider the so-called topological susceptibility as the integrated correlation function of two Chern-Pontryagin densities

$$\chi_t = \int d^4x \ {}_{pg}\langle 0|P(0)P(x)|0\rangle_{pg} = \frac{\langle Q^2\rangle}{V} \tag{5.7.4}$$

in the pure gluon theory (without quarks). Here $|0\rangle_{pg}$ is the vacuum of the pure gluon theory, and V is the volume of space-time. When we add massless quarks, the Atiyah-Singer index theorem implies that the topological charge — and hence χ_t — vanishes, because the zero-modes of the Dirac operator eliminate topologically nontrivial field configurations. Therefore in full QCD (with massless quarks)

$$\int d^4x \ \langle 0|P(0)P(x)|0\rangle = 0, \tag{5.7.5}$$

where $|0\rangle$ is the full QCD vacuum. In the large N_c limit the effects of quarks are $1/N_c$ suppressed. Therefore it is unclear how they can eliminate the topological susceptibility of the pure gluon theory. In the large N_c limit the quark effects manifest themselves entirely in terms of mesons. One finds

$$\chi_t - \sum_m \frac{\langle 0|P|m\rangle\langle m|P|0\rangle}{M_m^2} = 0, \tag{5.7.6}$$

where the sum runs over all meson states and M_m are the corresponding meson masses. Using large N_c techniques one can show that $|\langle 0|P|m\rangle|^2$ is of order $1/N_c$, while χ_t is of order 1. If also all meson masses would be of order 1 there would be a contradiction. The puzzle gets resolved when one assumes that the lightest flavorscalar, pseudoscalar meson — the η' — has in fact a mass of order $1/N_c$, such that

$$\chi_t = \frac{|\langle 0|P|\eta'\rangle|^2}{M_{\eta'}^2}. \tag{5.7.7}$$

Using the anomaly equation one obtains

$$\langle 0|P|\eta'\rangle = \frac{1}{2N_f}\langle 0|\partial_\mu A_\mu|\eta'\rangle = \frac{1}{\sqrt{2N_f}} M_{\eta'}^2 f_{\eta'}. \tag{5.7.8}$$

In the large N_c limit $f_{\eta'} = f_\pi$ and we arrive at the Witten-Veneziano formula

$$\chi_t = \frac{f_\pi^2 M_{\eta'}^2}{2N_f}. \tag{5.7.9}$$

In this equation χ_t is of order 1, f_π^2 is of order N_c and $M_{\eta'}^2$ is of order $1/N_c$. This means that the η'-meson is indeed a Goldstone boson in a world with infinitely many colors. At finite N_c the anomaly arises leading to an explicit axial $U(1)$ symmetry breaking proportional to $1/N_c$. The pseudo-Goldstone boson mass squared is hence proportional to $1/N_c$. So far we have assumed that all quarks are massless. When a nonzero s quark mass is taken into account, the formula changes to

$$\chi_t = \frac{1}{6} f_\pi^2 (M_{\eta'}^2 + M_\eta^2 - 2M_K^2) = (0.180\text{GeV})^4. \tag{5.7.10}$$

Lattice calculations are at least roughly consistent with this value, which supports this solution of the $U(1)$-problem.

Chapter 6

Topology of Goldstone Boson Fields

We have seen that the topology of gauge fields gives rise to interesting effects in QCD. These arise through the Adler-Bell-Jackiw anomaly and are hence quite subtle quantum effects. We have also seen that the low-energy physics of QCD is dominated by almost massless pseudo-Goldstone bosons — the pions, kaons and the η-meson. Via 't Hooft's anomaly matching mechanism, all QCD (and in fact all standard model) anomalies should also manifest themselves in the low-energy effective theory. It may not be too surprising that anomalies in the low-energy effective theory arise via a nontrivial topology of the Goldstone boson fields. In QCD, the Goldstone boson fields live in the coset space $SU(N_f)_L \otimes SU(N_f)_R/SU(N_f)_{L=R} = SU(N_f)$. For any number of flavors $\Pi_3[SU(N_f)] = \mathbb{Z}$. This homotopy gives rise to a winding number of the Goldstone boson field which can be identified with baryon number. It is quite surprising that meson fields, which don't carry baryon number themselves, can give rise to topological excitations that are baryons. These baryons are the so-called Skyrmions. Depending on the number of colors N_c, the baryons of QCD are either fermions (for odd N_c) or bosons (for even N_c). On the level of the effective theory, the statistics of the Skyrmions is determined via the so-called Wess-Zumino-Witten term. This term is related to the nontrivial topology of $\Pi_5[SU(N_f)] = \mathbb{Z}$ for $N_f > 2$, and its prefactor is proportional to N_c. The Wess-Zumino term contains the information about the standard model anomalies on the level of the effective theory. For $N_f = 2$ we have $\Pi_5[SU(2)] = \{0\}$, and a Wess-Zumino term does not exist. Instead one has $\Pi_4[SU(2)] = \mathbb{Z}(2)$. The corresponding winding number appears in the low-energy pion effective theory when N_c is odd and the Skyrmions need to be quantized as fermions. The standard model anomalies are now encoded in the so-called Goldstone-Wilczek current. Among other things, this term gives rise to the decay of the neutral pion into a pair of photons.

6.1 Skyrmions

As we have seen earlier, the lowest order term in the chiral perturbation theory action for the Goldstone bosons of QCD takes the form

$$S[U] = \int d^4x \; \frac{f_\pi^2}{4} \mathrm{Tr}[\partial^\mu U^\dagger \partial_\mu U]. \tag{6.1.1}$$

Here U is a field in the coset space $SU(N_f)_L \otimes SU(N_f)_R / SU(N_f)_{L=R} = SU(N_f)$. At each moment in time, the Goldstone boson field can be viewed as a map of 3-d space into the group $SU(N_f)$. The nontrivial homotopy $\Pi_3[SU(N_f)] = \mathbb{Z}$ implies that such maps fall in topologically distinct classes characterized by an integer winding number

$$B = \frac{1}{24\pi^2} \int d^3x \; \varepsilon_{ijk} \mathrm{Tr}[(U\partial^i U^\dagger)(U\partial^j U^\dagger)(U\partial^k U^\dagger)] \in \mathbb{Z}. \tag{6.1.2}$$

Amazingly, the winding number can be identified with baryon number and the topologically nontrivial Goldstone boson fields hence describe baryons. It is far from obvious that B is indeed the baryon number, but we will understand this only later. While B is defined for each instant of time, it actually doesn't change with time — baryon number is conserved. On the level of the effective theory, this is a consequence of topological current conservation. The baryon number current

$$J_\mu = \frac{1}{24\pi^2} \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(U\partial^\nu U^\dagger)(U\partial^\rho U^\dagger)(U\partial^\sigma U^\dagger)] \tag{6.1.3}$$

is conserved, i.e.

$$\partial^\mu J_\mu = 0, \tag{6.1.4}$$

independent of the equations of motion. Consequently, the baryon number

$$B = \int d^3x \; J_0 \tag{6.1.5}$$

does not change with time, i.e.

$$\partial^0 B = \int d^3x \; \partial^0 J_0 = \int d^3x \; \partial^i J_i = \int d^2\sigma^i \; J_i = 0, \tag{6.1.6}$$

provided the baryon number current vanishes at spatial infinity.

The lowest order effective action of eq.(6.1.1) does not give rise to stable classical solutions with $B = 1$. In order to stabilize baryonic solitons, Skyrme added a specific higher order term with four derivatives to the Lagrangian. The emerging soliton is known as a Skyrmion. For $N_f = 2$ it takes the form

$$U(\vec{x}) = \exp(iF(|\vec{x}|)\frac{\vec{x}}{|\vec{x}|} \cdot \vec{\sigma}), \tag{6.1.7}$$

where the function $F(|\vec{x}|)$ describes the radial shape of the soliton. One has $F(\infty) = \pi$ and $F(0) = 0$, which ensures that the soliton has $B = 1$. The Skyrme model has been used to describe baryons in QCD. This is certainly not rigorous but has worked reasonably well. However, it should be noted that, strictly speaking, the Skyrmion is beyond the applicability of the low-energy effective theory. Skyrmions have a mass at the high-energy scale f_π (they are heavy baryons) where low-order chiral perturbation breaks down. Even if the exact quartic derivative terms in the chiral Lagrangian for QCD were known exactly, and if they would also stabilize the soliton, in order to describe the actual QCD baryons, one would need the chiral Lagrangian to all orders. Here we will not ask any detailed questions about Skyrmions, but concentrate on rather qualitative aspects of their dynamics.

First of all, let us consider $N_f = 2$. Then $\Pi_4[SU(2)] = \mathbb{Z}(2)$, i.e. the space-time dependent Goldstone boson field histories fall into two topologically distinct classes. We can assign a winding number

$$\text{Sign}[U] = \pm 1 \tag{6.1.8}$$

to each field configuration. Fields that can be deformed into a trivial configuration $U = \mathbb{1}$ have $\text{Sign}[U] = 1$ while the others have $\text{Sign}[U] = -1$. It is interesting that the $\Pi_4[SU(2)]$ winding number appears in the path integral of the low-energy effective theory as

$$Z = \int \mathcal{D}U \ \exp(iS[U])\text{Sign}[U]^{N_c}. \tag{6.1.9}$$

When the number of colors N_c is even, the baryons are bosons and the winding number drops out of the path integral. When N_c is odd, on the other hand, the baryons are fermions and $\text{Sign}[U]$ enters the path integral. Indeed $\text{Sign}[U]$ is a fermion permutation sign that encodes the Pauli principle for Skyrmions in the pion effective theory. One can show this using various U field configurations. For example, a configuration with $B = 2$, in which two Skyrmions interchange their position as time evolves, has $\text{Sign}[U] = -1$ in accordance with the Pauli principle for fermionic baryons. In addition, a configuration with $B = 1$ in which a single Skyrmion rotates by 2π as time evolves also has $\mathrm{Sign}[U] = -1$. This ensures that Skyrmions have spin 1/2. In other words, the $\Pi_4[SU(2)] = \mathbb{Z}(2)$ winding number $\mathrm{Sign}[U] = \pm 1$ ensures that Skyrmions behave properly as fermions. It should be noted that in four dimensions particles can only be quantized as bosons or fermions because $\Pi_4[SU(2)] = \Pi_4[S^3] = \mathbb{Z}(2)$. This is in contrast to theories in three space-time dimensions where particles can exist with any spin or statistics (anyons) because then $\Pi_3[S^2] = \mathbb{Z}$.

6.2 Anomaly Matching for $N_f = 2$

It is interesting to note that the term $\mathrm{Sign}[U]^{N_c}$ also ensures the proper cancellation of Witten's global anomaly at the level of the effective theory. To see this, let us gauge $SU(2)_L$ in the chiral Lagrangian. This is simple to do, because the transformation $U' = LU$ with $L \in SU(2)_L$ is exactly the symmetry that is gauged in the standard model. To obtain a low-energy effective theory of the entire standard model, the pion Lagrangian should then be extended to include Higgs fields, W-bosons, as well as one generation of leptons. We have learned earlier that the global anomaly of the lepton $SU(2)_L$ doublet is canceled by the one of $N_c = 3$ quark doublets. On the level of the low-energy effective theory, the quarks are replaced by pions and the question arises how the global anomaly is now canceled. When we perform an $SU(2)_L$ gauge transformation L with a winding number $\mathrm{Sign}[L] = -1$, the fermionic measure of the lepton fields is not invariant, but spits out a minus sign. In the standard model this minus sign is canceled by $N_c = 3$ minus signs spit out by the fermionic measure of the quarks, so that the full theory is gauge invariant. In the effective theory there are no quarks, but the path integral now contains the factor $\mathrm{Sign}[U]^{N_c}$, which transforms as

$$\mathrm{Sign}[U']^{N_c} = \mathrm{Sign}[LU]^{N_c} = \mathrm{Sign}[L]^{N_c}\mathrm{Sign}[U]^{N_c}. \tag{6.2.1}$$

This is exactly what we need in order to cancel the global anomaly of the leptons.

We have also learned that baryon number is anomalous in the standard model because electroweak instantons lead to baryon number violating processes via the Adler-Bell-Jackiw anomaly

$$\partial^\mu J_\mu = -\frac{1}{32\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}(W^{\mu\nu}W^{\rho\sigma}). \tag{6.2.2}$$

If the Skyrme current indeed represents baryon number, after gauging the $SU(2)_L$ symmetry we should obtain the same result in the effective theory. When we gauge $SU(2)_L$ we need to replace ordinary derivatives $\partial_\mu U$ by covariant derivatives

$$D_\mu U = (\partial_\mu + gW_\mu)U, \ D_\mu U^\dagger = \partial_\mu U^\dagger - gU^\dagger W_\mu. \tag{6.2.3}$$

When one only replaces ordinary derivatives by covariant derivatives in eq.(6.1.3) one does not obtain eq.(6.2.2). Instead one should consider the so-called Goldstone-Wilczek current which represents the baryon number current in the presence of $SU(2)_L$ gauge fields

$$\begin{aligned} J_\mu &= \frac{1}{24\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)(UD^\rho U^\dagger)(UD^\sigma U^\dagger)] \\ &+ \frac{g}{32\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)W^{\rho\sigma}]. \end{aligned} \tag{6.2.4}$$

Here

$$W^{\mu\nu} = \partial^\mu W^\nu - \partial^\nu W^\mu + [W^\mu, W^\nu] \tag{6.2.5}$$

is the field strength of the W-boson field. It can now be shown that the Goldstone-Wilczek current indeed obeys eq.(6.2.2). Note that in the absence of $SU(2)_L$ gauge fields the Goldstone-Wilczek current reduces to Skyrme's baryon number current of eq.(6.1.3). Since eq.(6.2.2) is now satisfied on the level of the effective theory, just like the standard model it will contain baryon number violating processes. If we consider a Skyrmion propagating in an electroweak instanton background (a W-field configuration with topological charge $Q = 1$) the baryon number will change by one unit

$$\begin{aligned} B(t = \infty) - B(t = -\infty) = \int d^3x \ J_0(t = \infty) - \int d^3x \ J_0(t = -\infty) = \\ \int d^4x \ \partial^\mu J_\mu = -\frac{1}{32\pi^2}\int d^4x \ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}(W^{\mu\nu}W^{\rho\sigma}) = Q = 1. \end{aligned} \tag{6.2.6}$$

In the same way, the lepton number also changes by one unit so that $B - L$ remains conserved.

Of course, in the standard model we have also gauged $U(1)_Y$. The electroweak hypercharge of the left-handed up and down quarks is $Y_L = 1/6$, while the hypercharges of the right-handed quarks are $Y_{u_R} = 2/3$ and $Y_{d_R} = -1/3$. Hence on the level of the effective theory, the corresponding transformations act both on the left and on the right, i.e. $U' = LUR^\dagger$, with

$$L = \begin{pmatrix} \exp(ig'\varphi/6) & 0 \\ 0 & \exp(ig'\varphi/6) \end{pmatrix}, \ R = \begin{pmatrix} \exp(2ig'\varphi/3) & 0 \\ 0 & \exp(-ig'\varphi/3) \end{pmatrix}. \tag{6.2.7}$$

However, one should note that L and R, written in this form, are not elements of $SU(2)$ but of $U(2)$. Still, since L is proportional to the unit matrix and hence commutes with any matrix U we can pull it over to the right and combine it with R to form a new $SU(2)$ matrix that multiplies U only from the right. Calling the new $SU(2)_R$ matrix R again we then obtain

$$U' = UR^\dagger, \ R = \begin{pmatrix} \exp(ig'\varphi/2) & 0 \\ 0 & \exp(-ig'\varphi/2) \end{pmatrix}. \tag{6.2.8}$$

In this way we have identified $U(1)_Y$ as the Abelian subgroup of $SU(2)_R$. The total covariant derivative hence takes the form

$$D_\mu U = \partial_\mu U + igW_\mu^a \frac{\sigma_a}{2} U - ig'UB_\mu \frac{\sigma_3}{2}. \tag{6.2.9}$$

With $U(1)_Y$ also being gauged, we need to modify the definition of the Goldstone-Wilczek current to

$$\begin{aligned} J_\mu &= \frac{1}{24\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)(UD^\rho U^\dagger)(UD^\sigma U^\dagger)] \\ &+ \frac{g}{32\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)W^{\rho\sigma}] \\ &+ \frac{g'}{32\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)B^{\rho\sigma}\frac{i\sigma_3}{2}], \end{aligned} \tag{6.2.10}$$

where

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu \tag{6.2.11}$$

is the field strength of the $U(1)_Y$ gauge field.

Let us now identify the way electromagnetism is coupled into the effective theory. Inserting

$$W_\mu^3 = \frac{g'A_\mu + gZ_\mu}{\sqrt{g^2+g'^2}}, \ B_\mu = \frac{gA_\mu - g'Z_\mu}{\sqrt{g^2+g'^2}}, \tag{6.2.12}$$

into the covariant derivative and putting $W_\mu^\pm = Z_\mu = 0$, one obtains the coupling to the photon as

$$D_\mu U = \partial_\mu U + i\frac{gg'}{\sqrt{g^2+g'^2}}(A_\mu\frac{\sigma_3}{2}U - UA_\mu\frac{\sigma_3}{2}) = \partial_\mu U + i\frac{e}{2}A_\mu[\sigma_3, U]. \tag{6.2.13}$$

Here we have identified the electric charge as

$$e = \frac{gg'}{\sqrt{g^2+g'^2}}. \tag{6.2.14}$$

Let us also identify the $U(1)_{em}$ gauge symmetry of electromagnetism in another way. The electric charge of the left- and right-handed up-quarks is 2/3 while that of the down-quarks is −1/3. Hence, we can again write $U' = LUR^\dagger$, now with

$$L = \left(\begin{array}{cc} \exp(2ie\varphi/3) & 0 \\ 0 & \exp(-ie\varphi/3) \end{array}\right), \ R = \left(\begin{array}{cc} \exp(2ie\varphi/3) & 0 \\ 0 & \exp(-ie\varphi/3) \end{array}\right). \tag{6.2.15}$$

As before, these L and R are elements of $U(2)$ — not of $SU(2)$. However, by commuting a multiple of the unit matrix through, we can again replace these by new matrices

$$L = \left(\begin{array}{cc} \exp(ie\varphi/2) & 0 \\ 0 & \exp(-ie\varphi/2) \end{array}\right), \ R = \left(\begin{array}{cc} \exp(ie\varphi/2) & 0 \\ 0 & \exp(-ie\varphi/2) \end{array}\right), \tag{6.2.16}$$

which are indeed in $SU(2)$. Hence, on the level of the low-energy effective theory the covariant derivative of electromagnetism again takes the form

$$D_\mu U = \partial_\mu U + i\frac{e}{2}A_\mu[\sigma_3, U]. \tag{6.2.17}$$

In the presence of just electromagnetic gauge fields the Goldstone-Wilczek current takes the form

$$\begin{aligned} J_\mu &= \frac{1}{24\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)(UD^\rho U^\dagger)(UD^\sigma U^\dagger)] \\ &+ \frac{e}{16\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(UD^\nu U^\dagger)F^{\rho\sigma}\frac{i\sigma_3}{2}], \end{aligned} \tag{6.2.18}$$

where

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{6.2.19}$$

is the field strength of the electromagnetic field.

Even though the previous arguments have allowed us to identify the correct expressions for the covariant derivatives and the Goldstone-Wilczek current, we have not yet coupled the $U(1)_Y$ or $U(1)_{em}$ gauge fields in the right form. For example, the anomalous decay of the neutral pion into two photons is not yet contained in our effective theory. From our previous discussion we know that the full $U(1)_Y$ or $U(1)_{em}$ gauge transformations of the standard model are described by $U(2)_L \otimes U(2)_R$ matrices. Still, on the level of the pion effective theory, these are indistinguishable from equivalent $SU(2)_L \otimes SU(2)_R$ transformations because one can commute a multiple of the unit matrix from the left to the right. Based on this observation, it might seem that the low-energy effective theory of the standard model with $Y_L = 1/6$, $Y_{u_R} = 2/3$ and $Y_{d_R} = -1/3$ would be the same as the one of a theory with $Y_L = 0$, $Y_{u_R} = 1/2$ and $Y_{d_R} = -1/2$. In such a theory the quarks would have electric charges $Q_u = 1/2$ and $Q_d = -1/2$ and the neutral pion would indeed not decay into two photons. In the standard model the charge assignments are

$$Y_L = \frac{1}{6} + 0, \ Y_{u_R} = Q_u = \frac{1}{6} + \frac{1}{2}, \ Y_{d_R} = Q_d = \frac{1}{6} - \frac{1}{2}, \tag{6.2.20}$$

i.e. they differ from that of the other theory by a shift 1/6 which is just one half of the baryon number $B = 1/3$ of the quarks. In other words, the $U(1)_Y$ and $U(1)_{em}$ gauge fields of the standard model have a coupling to baryon number. On the level of the effective pion theory this coupling enters the action via the Goldstone-Wilczek current which represents the baryon number. Indeed, there is a term $\frac{g'}{2} B^\mu J_\mu$ that must be added to the effective pion Lagrangian. When one limits oneself to just electromagnetic gauge fields this term takes the form

$$\begin{aligned} \frac{e}{2} A^\mu J_\mu &= \frac{e}{48\pi^2} A^\mu \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(U D^\nu U^\dagger)(U D^\rho U^\dagger)(U D^\sigma U^\dagger)] \\ &+ \frac{e^2}{32\pi^2} A^\mu \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(U D^\nu U^\dagger) F^{\rho\sigma} \frac{i\sigma_3}{2}]. \end{aligned} \tag{6.2.21}$$

Let us now identify a vertex for the decay of a neutral pion into two photons from this expression. For this purpose we write

$$U = \exp(i \frac{1}{f_\pi} \pi^0 \sigma_3) \approx 1 + i \frac{1}{f_\pi} \pi^0 \sigma_3. \tag{6.2.22}$$

Then the electromagnetic covariant derivative takes the form

$$D_\mu U = \partial_\mu U + i \frac{e}{2} A_\mu [\sigma_3, U] \approx i \frac{1}{f_\pi} \partial_\mu \pi^0 \sigma_3. \tag{6.2.23}$$

Hence, expanding to leading order in the pion field, we obtain

$$\frac{e}{2} A^\mu J_\mu = \frac{e^2}{64\pi^2 f_\pi} \pi^0 \varepsilon_{\mu\nu\rho\sigma} F^{\mu\nu} F^{\rho\sigma}, \tag{6.2.24}$$

which is indeed the vertex we were looking for.

6.3 The Wess-Zumino-Witten Term

For $N_f > 2$ the topological properties of the Goldstone bosons are qualitatively different from the two flavor case. In particular, for more than two flavors $\Pi_4[SU(N_f)] = \{0\}$. Instead, $\Pi_5[SU(N_f)] = \mathbb{Z}$ for $N_f > 2$ while $\Pi_5[SU(2)] = \{0\}$.

One might think that this implies that in the $N_f > 2$ case Goldstone bosons living in four space-time dimensions are topologically trivial. If this were so, how would the low-energy effective theory know if the Skyrmion is quantized as a boson or a fermion? How would it know about baryon violating processes or about the decay of a neutral pion into two photons? Indeed, by introducing an additional unphysical fifth dimension, the homotopy $\Pi_5[SU(N_f)] = \mathbb{Z}$ can be used to explain all these effects. This leads to the so-called Wess-Zumino-Witten term.

The effective action of eq.(6.1.1) has more symmetries than QCD itself. In particular, it is invariant under the so-called intrinsic parity, while QCD is invariant only against the full parity transformation. As pseudoscalar particles, under the full parity transformation P the Goldstone pions transform as

$$^P\pi^a(\vec{x},t) = -\pi^a(-\vec{x},t), \tag{6.3.1}$$

while under the intrinsic parity P_0 they would only transform as

$$^{P_0}\pi^a(\vec{x},t) = -\pi^a(\vec{x},t). \tag{6.3.2}$$

Writing $U = \exp(i\pi^a\eta_a/f_\pi)$ these transformations take the form

$$^PU(\vec{x},t) = U^\dagger(-\vec{x},t), \ ^{P_0}U(\vec{x},t) = U^\dagger(\vec{x},t). \tag{6.3.3}$$

The effective action of eq.(6.1.1) is invariant under both P and P_0. If P_0 were a symmetry of QCD, the number of Goldstone bosons would never change from even to odd. However, there are processes in QCD in which a single kaon decays into a pair of pions. Such effects are indeed also described by the Wess-Zumino-Witten term.

Let us first introduce a term similar to the Wess-Zumino-Witten term that would appear in a 5-d theory of Goldstone bosons. The winding number of the homotopy group $\Pi_5[SU(N_f)] = \mathbb{Z}$ for $N_f > 2$ might play a role in a 5-d theory in which space-time is compactified to S^5 and it takes the form

$$\begin{aligned} W[U] &= \frac{i}{480\pi^3}\int_{S^5} d^5x \ \varepsilon_{\mu\nu\rho\sigma\tau} \\ &\times \ \mathrm{Tr}[(U\partial^\mu U^\dagger)(U\partial^\nu U^\dagger)(U\partial^\rho U^\dagger)(U\partial^\sigma U^\dagger)(U\partial^\tau U^\dagger)]. \end{aligned} \tag{6.3.4}$$

How can an object like this play a role in a 4-d theory? Let us compactify 4-d space-time to a sphere S^4 and let this sphere be the boundary of a 5-d world B^5 inside S^4. The physical Goldstone boson field $U(\vec{x},t)$ lives at the 4-d boundary of B^5 but it should also be defined in the interior of B^5. On the other hand, the 4-d physics should be independent of what happens in the bulk of the fifth dimension. There is an arbitrary number of ways to extend the 4-d map $U(\vec{x},t)$, which is a map of S^4 into $SU(N_f)$, to the 5-d bulk B^5. How can the 4-d physics be independent of what map we choose? To understand this, let us construct the Wess-Zumino-Witten action

$$\begin{aligned} S_{WZW}[U] &= \frac{i}{480\pi^3}\int_{B^5} d^5x\ \varepsilon_{\mu\nu\rho\sigma\tau} \\ &\times \ \mathrm{Tr}[(U\partial^\mu U^\dagger)(U\partial^\nu U^\dagger)(U\partial^\rho U^\dagger)(U\partial^\sigma U^\dagger)(U\partial^\tau U^\dagger)]. \end{aligned} \quad (6.3.5)$$

Although the integrand is the same as in eq.(6.3.4), the Wess-Zumino-Witten action is not an integer because we are no longer integrating over the compact space S^5. Instead we now integrate over the 5-d manifold B^5 with boundary S^4. If this boundary would be absent, the action $S_{WZW}[U]$ would indeed be an integer. One can show that the integrand in eq.(6.3.5) is a total divergence. Using Stokes' theorem one might then conclude that the integral is indeed determined just by the value of U at the boundary of B^5, i.e. in the physical space-time S^4. However, one must be careful because there may be singularities within the bulk of B^5. Such singularities must be exiled from the space and surrounded by a small sphere S^4_ε. The integral over S^4_ε yields an integer which does indeed depend on the behavior of U within the bulk of S^5. Hence, $S_{WZW}[U]$ in fact not only depends on the values of U in the physical space-time S^4 — it also receives integer-valued contributions from the 5-d bulk.

To confirm that the bulk contributions are indeed integers, let us now consider the space B'^5 which forms the 5-d exterior of the physical space-time sphere S^4, as well as another Wess-Zumino-Witten action

$$\begin{aligned} S'_{WZW}[U] &= \frac{i}{480\pi^3}\int_{B'^5} d^5x\ \varepsilon_{\mu\nu\rho\sigma\tau} \\ &\times \ \mathrm{Tr}[(U\partial^\mu U^\dagger)(U\partial^\nu U^\dagger)(U\partial^\rho U^\dagger)(U\partial^\sigma U^\dagger)(U\partial^\tau U^\dagger)]. \end{aligned} \quad (6.3.6)$$

Since we have compactified the 5-d bulk to S^5, we have $B^5 \cup B'^5 = S^5$ and hence

$$S_{WZW}[U] + S'_{WZW}[U] = W[U]. \quad (6.3.7)$$

The space-time S^4 surface contributions to $S_{WZW}[U]$ and $S'_{WZW}[U]$ cancel each other because S^4 forms the boundary of both B^5 and B'^5, but it is oppositely oriented in the two cases. Hence, only the bulk contributions survive in the sum $S_{WZW}[U] + S'_{WZW}[U]$ which is indeed nothing but the integer-valued winding number $W[U]$. This again confirms that $S_{WZW}[U]$ is a sum of a (non-integer) surface contribution from the physical space-time S^4 and integer-valued contributions from the unphysical bulk B^5.

How can we get rid of the unwanted integer-valued contribution from the bulk? We should keep in mind that the Wess-Zumino-Witten action enters the path integral of the effective theory as

$$Z = \int \mathcal{D}U \ \exp(iS[U]) \exp(icS_{WZW}[U]), \tag{6.3.8}$$

where c is some constant. For arbitrary values of c the path-integral would indeed know about the singularities within the bulk of B^5. However, if c were a multiple of 2π, the integer bulk ambiguity of $S_{WZW}[U]$ would in fact become invisible. In other words, for the whole construction to work, the coefficient c must be quantized. Indeed, for the low-energy effective theory of QCD with $N_f > 2$ Witten found

$$c = 2\pi N_c, \tag{6.3.9}$$

where N_c is the number of colors. It is remarkable that the low-energy theory of pions, kaons and η-mesons knows about the number of colors, which appears as a quantized parameter in its effective action.

Chapter 7

The Strong CP-Problem

We have seen that non-Abelian $SU(N)$ gauge fields have nontrivial topological structure. Classical vacuum (pure gauge) field configurations are characterized by an integer winding number from the homotopy group $\Pi_3[SU(N)] = \mathbb{Z}$. Instantons are examples of Euclidean field configurations with topological charge Q that describe tunneling between topologically distinct classical vacua. Due to tunneling, the quantum vacuum is a linear superposition of classical vacua characterized by a vacuum angle $\theta \in [-\pi, \pi]$. In the Euclidean action the vacuum angle manifests itself as an additional term $i\theta Q$. For $\theta \neq 0, \pi$ this term explicitly breaks the CP symmetry. As a consequence, the neutron would have an electric dipole moment proportional to θ, while without CP violation the dipole moment vanishes. Indeed, the observed electric dipole moment of the neutron is indistinguishable from zero. This puts a stringent bound on the vacuum angle $|\theta| < 10^{-9}$. The question arises why in Nature $\theta = 0$ to such a high accuracy. This is the strong CP-problem.

Within QCD itself, one could "solve" the strong CP-problem simply by demanding CP symmetry. In the standard model, however, the Yukawa couplings already lead to CP violation which is indeed observed in the neutral kaon system. This effect is rather subtle and requires the presence of at least three generations. If there were CP violation in the strong interactions, it would give rise to much more drastic effects. Naively, one might hope to solve the strong CP-problem by the assumption that gluon fields with $Q \neq 0$ are very much suppressed. However, this probably does not work. First of all, the quantitative solution of the $U(1)$-problem relies on the fact that gluon fields with $Q \neq 0$ appear frequently in the pure gauge theory. Of course, this need not necessarily be the case in full QCD with quarks. Indeed, if the up quark would be massless, the Atiyah-Singer index theorem would imply that fermionic zero-modes of the Dirac operator completely eliminate gluon fields with $Q \neq 0$ from the path integral. In that case, the θ-vacuum term would vanish and all θ-vacua would be physically equivalent and thus CP conserving. It is a controversial issue if the up quark might indeed be massless, but most experts of chiral perturbation theory believe that this possibility is excluded. In any case, if the up quark would indeed be massless, and we would solve the CP problem in that way, we would immediately face the m_u-problem: why is the up quark massless?

We have seen already that the Chern-Pontryagin topological charge density is intimately connected with the divergence of the flavor-singlet axial current. This implies that the vacuum angle can be rotated using an axial $U(1)$ transformation. In this way, one can indeed get rid of any hypothetical θ'-angle in the electroweak $SU(2)_L$ gauge field. The strong $SU(3)_c$ θ-vacuum angle, on the other hand, cannot be rotated away in this fashion, because it just gets transformed into a complex phase of the determinant of the quark mass matrix. Still, such a transformation can be quite useful, for example, because we can then investigate the θ-vacuum dynamics using chiral Lagrangians. For example, for unequal up and down quark masses, one finds a phase transition at $\theta = \pi$ at which CP is spontaneously broken. Hence, despite the fact that $\theta = \pi$ does not break CP explicitly, the CP symmetry is now broken dynamically. This means that θ cannot be π in Nature and must indeed be zero.

The chiral Lagrangian method also allows us to study θ-vacuum effects in the large N_c limit. In this limit, the axial $U(1)$ anomaly vanishes and the η'-meson becomes a massless Goldstone boson. In fact, the η'-meson couples to the complex phase of the quark mass matrix — and hence to θ — and can indeed be used to rotate θ away. Hence, there is no strong CP-problem at $N_c = \infty$. Of course, we know that in our world $N_c = 3$ (although some of the textbook arguments for this (anomaly cancellation, π^0 decay) are incorrect), and we indeed face the strong CP-problem.

A very appealing solution of the strong CP-problem was suggested by Peccei and Quinn. They suggested an extension of the standard model with two Higgs doublets. This situation also naturally arises in supersymmetric extensions of the standard model. As a consequence of the presence of the second Higgs field, there is an extra $U(1)_{PQ}$ — a so-called Peccei-Quinn symmetry — which allows one to rotate θ away even at finite N_c. When $SU(2)_L \otimes U(1)_Y$ breaks down to $U(1)_{em}$, the Peccei-Quinn $U(1)_{PQ}$ symmetry also gets spontaneously broken. It was first pointed out by Weinberg and Wilczek that this leads to a new pseudo-Goldstone boson — the axion. Unfortunately, so far nobody has ever detected an axion despite numerous experimental efforts and it is still unclear if this is indeed the correct solution of the strong CP problem. Although the original Peccei-Quinn model was soon ruled out by experiments, the symmetry breaking scale of the model can be shifted to higher energy scales making the axion more or less invisible.

Axions are very interesting players in the Universe. They couple only weakly to ordinary matter, but they still have interesting effects. First of all, they are massive and could provide enough energy to close the Universe. If it exists, the axion can also shorten the life-time of stars. Stars live so long, because they cannot get rid of their energy by radiation very fast. For example, a photon that is generated in a nuclear reaction in the center of the sun spends 10^7 years before it reaches the sun's surface, simply because its electromagnetic cross section with the charged matter in the sun is large. An axion, on the other hand, interacts weakly and can thus get out much father. Like neutrinos, axions can therefore act as a super coolant for stars. The observed life-time of stars can thus be used to put astrophysical limits on axion parameters like the axion mass. Axions can be generated in the early Universe in multiple ways. First, they can simply be thermally produced. Then they can be generated by a disoriented $U(1)_{PQ}$ condensate. This mechanism is similar to the recently discussed pion production via a disoriented chiral condensate in a heavy ion collision generating a quark-gluon plasma. Also, the spontaneous breakdown of a $U(1)$ symmetry is accompanied by the generation of cosmic strings. Indeed, if the axion exists, axionic cosmic strings should exist as well. A network of such fluctuating strings could radiate energy by emitting the corresponding Goldstone bosons, namely axions.

7.1 Rotating θ into the Mass Matrix

Let us assume that there is a θ-vacuum term $i\theta Q$ in the Euclidean action of QCD. We have seen already that such a term is intimately connected with the flavor-singlet axial $U(1)$ symmetry. Indeed, due to the axial anomaly, the fermionic measure is not invariant under axial $U(1)$ transformations. Let us discuss this in the theory with N_f quark flavors. Under an axial $U(1)$ transformation

$$q_L' = \exp(-i\theta/2N_f)q_L, \ q_R' = \exp(i\theta/2N_f)q_R, \tag{7.1.1}$$

the fermion determinant in the background of a gluon field with topological charge Q is not invariant. In fact, it changes by $\exp(i\theta Q)$. Hence, the above axial transformation can be used to cancel any pre-existing θ-vacuum term in the QCD action. Of course, the transformation must be applied consistently everywhere.

It cancels out in the quark-gluon gauge interactions which are chirally invariant, but not in mass terms. In fact, the mass matrix $\mathcal{M} = \text{diag}(m_u, m_d, ..., m_{N_f})$ now turns into

$$\mathcal{M}' = \text{diag}(m_u \exp(i\theta/N_f), m_d \exp(i\theta/N_f), ..., m_{N_f} \exp(i\theta/N_f)), \tag{7.1.2}$$

i.e. θ turns into the complex phase of the determinant of the quark mass matrix. If one of the quarks is massless, the determinant vanishes and its phase becomes physically irrelevant. Interestingly, strong CP violation manifests itself by a complex phase in the quark mass matrix, while the CP violation due to the Yukawa couplings leads to the complex phase in the Cabibbo-Kobayashi-Maskawa quark mixing matrix.

The strong interaction θ-angle cannot be completely rotated away, because both left- and right-handed quarks are coupled to the gluons. A potential electroweak interaction θ'-angle, on the other hand, can simply be rotated away, because only the left-handed fermions couple to the $SU(2)_L$ gauge field. For example, in order to remove a θ'-angle, one just performs a left-handed $U(1)$ transformation

$$\begin{pmatrix} \nu'_{eL} \\ e'_L \end{pmatrix} = \exp(-i\theta') \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}. \tag{7.1.3}$$

The change of the fermion measure under the transformation cancels against the θ'-term, the gauge interactions remain unchanged, but the mass terms are again affected. However, we can now simply rotate the right-handed fields as well

$$\nu'_{eR} = \exp(-i\theta')\nu_{eR}, \; e'_R = \exp(-i\theta')e_R, \tag{7.1.4}$$

which then leaves the mass term invariant. Unlike in the QCD case, this does not regenerate the θ'-term because the right-handed fermions do not couple to the $SU(2)_L$ gauge field.

7.2 θ-Angle in Chiral Perturbation Theory

Let us now discuss how the vacuum angle affects the low-energy QCD dynamics. Since we know how the quark mass matrix enters the chiral Lagrangian, and since θ is just the complex phase in that matrix, it is clear how to include θ in chiral perturbation theory. To lowest order the chiral perturbation theory action then takes the form

$$S[U] = \int d^4x \; \{\frac{F_\pi^2}{4}\text{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\text{Tr}[\mathcal{M}'U^\dagger + U\mathcal{M}'^\dagger]\}, \tag{7.2.1}$$

where $\mathcal{M}'$ is the θ-dependent quark mass matrix of eq.(7.1.2). The above action is not 2π-periodic in θ. Instead, it is only $2\pi N_f$-periodic. Still, it is easy to show that the resulting path integral is indeed 2π-periodic. The situation in QCD itself is similar. While the contribution $i\theta Q$ to the action itself is not periodic in θ, it enters the path integral through the 2π-periodic Boltzmann factor $\exp(i\theta Q)$. Hence, the path integral is periodic while the action itself is not. Let us check that a non-zero vacuum angle indeed breaks CP. On the level of the chiral Lagrangian charge conjugation corresponds to ${}^C U = U^T$, while parity corresponds to ${}^P U(\vec{x},t) = U(-\vec{x},t)^\dagger$. The action from above breaks P while it leaves C invariant, and hence it indeed violates CP.

Let us now examine the effect of θ on the vacuum of the pion theory in the $N_f = 2$ case. Then the mass matrix takes the form

$$\mathcal{M}' = \mathrm{diag}(m_u \exp(i\theta/2), m_d \exp(i\theta/2)). \tag{7.2.2}$$

In order to find the vacuum configuration, we must minimize the potential energy, and hence we must maximize

$$\mathrm{Tr}[\mathcal{M}'U^\dagger + U\mathcal{M}'^\dagger] = m_u \cos(\frac{\theta}{2} + \varphi) + m_d \cos(\frac{\theta}{2} - \varphi). \tag{7.2.3}$$

Here we have parametrized $U = \mathrm{diag}(\exp(i\varphi), \exp(-i\varphi))$. The minimum energy configuration has

$$\tan\varphi = \frac{m_d - m_u}{m_u + m_d} \tan\frac{\theta}{2}. \tag{7.2.4}$$

As expected, for $\theta = 0$ one obtains $\varphi = 0$ and hence $U = \mathbb{1}$. It is interesting that the θ-angle affects the pion vacuum only for nondegenerate quark masses. At $\theta = \pm\pi$ the pion vacuum configuration has $\varphi = \pm\pi/2$, i.e. $U = \pm\mathrm{diag}(i, -i)$. These two vacua are both not CP invariant. Instead, they are CP images of one another. This indicates that, at $\theta = \pm\pi$, the CP symmetry is spontaneously broken. Hence, despite the fact that for $\theta = \pm\pi$ there is no explicit CP violation, the symmetry is still not intact. This means that in Nature we have $\theta = 0$, not $\theta = \pi$.

7.3 Vacuum Angle at Large N_c

We have seen that the $U(1)$ problem can be understood quantitatively in the limit of many colors N_c. At $N_c = \infty$ the anomalous axial $U(1)$ symmetry is restored and the η'-meson becomes a Goldstone boson. At large but finite N_c the

η'-meson is a pseudo-Goldstone boson with a mass

$$M_{\eta'}^2 = \frac{N_f \chi_t}{F_\pi^2}, \tag{7.3.1}$$

proportional to $1/N_c$ (note that F_π^2 is of order N_c). Here $\chi_t = \langle Q^2 \rangle / V$ is the topological susceptibility of the pure gauge theory which is of order one in the large N_c limit.

Since for large N_c the η'-meson becomes light, it must be included in the low-energy chiral Lagrangian. Since the axial $U(1)$ symmetry is restored at $N_c = \infty$ and is then spontaneously broken, the chiral symmetry is now $U(N_f)_L \otimes U(N_f)_R$ broken to $U(N_f)_{L=R}$. Consequently, the Goldstone bosons now live in the coset space $U(N_f)_L \otimes U(N_f)_R / U(N_f)_{L=R} = U(N_f)$. Hence, now there are N_f^2 Goldstone bosons. The additional η' Goldstone boson is described by the complex phase of the determinant of a unitary matrix $\tilde{U}$, which would have determinant one if the η'-meson were heavy. For large N_c the chiral perturbation theory action takes the form

$$\begin{aligned} S[U] &= \int d^4x \, \{ \frac{F_\pi^2}{4} \mathrm{Tr}[\partial^\mu \tilde{U}^\dagger \partial_\mu \tilde{U}] + \frac{1}{2N_f} \langle \bar{\Psi} \Psi \rangle \mathrm{Tr}[\mathcal{M}' \tilde{U}^\dagger + \tilde{U} \mathcal{M}'^\dagger] \\ &+ N_f \chi_t (i \log \det \tilde{U})^2 \}. \end{aligned} \tag{7.3.2}$$

If there is a θ-angle in the quark mass matrix, this angle can now be absorbed into the η'-meson field, i.e. in the complex phase of the determinant of the Goldstone boson field $\tilde{U}$. Then the action turns into

$$\begin{aligned} S[U] &= \int d^4x \, \{ \frac{F_\pi^2}{4} \mathrm{Tr}[\partial^\mu \tilde{U}^\dagger \partial_\mu \tilde{U}] + \frac{1}{2N_f} \langle \bar{\Psi} \Psi \rangle \mathrm{Tr}[\mathcal{M} \tilde{U}^\dagger + \tilde{U} \mathcal{M}^\dagger] \\ &+ N_f \chi_t (i \log \det \tilde{U} - \theta)^2 \}. \end{aligned} \tag{7.3.3}$$

In the large N_c limit the last term which is of order one can be neglected compared to the other terms which are of order N_c. Hence, at $N_c = \infty$, the vacuum angle drops out of the theory, and all θ-vacua become physically equivalent. Hence, for infinitely many colors there is no CP problem. Essentially, the restored $U(1)$ symmetry then allows us to rotate θ away, despite the fact that it is still explicitly broken by the quark masses.

7.4 The Peccei-Quinn Symmetry

At finite N_c, the axial $U(1)$ symmetry is inevitably broken by the anomaly. Hence, we will not be able to rotate θ away using that symmetry. The idea of Peccei and

Quinn was to introduce another $U(1)_{PQ}$ symmetry — now known as a Peccei-Quinn symmetry — that will allow us to get rid of θ despite the fact that the axial $U(1)$ symmetry is explicitly anomalously broken. We will discuss the Peccei-Quinn symmetry in the context of the single generation standard model. The generalization to more generations is straightforward. Of course, it should be noted that with less than three generations, there is no CP violating phase in the quark mixing matrix and θ would be the only source of CP violation. Let us first remind ourselves how the up and down quarks get their masses in the standard model. As we have seen earlier, the mass of the down quark $m_d = f_d v$ is due to the Yukawa coupling

$$\mathcal{L}(u_L, d_L, d_R, \Phi) = f_d[\bar{d}_R(\Phi_+^* \Phi_0^*) \begin{pmatrix} u_L \\ d_L \end{pmatrix} + (\bar{u}_L \bar{d}_L) \begin{pmatrix} \Phi_+ \\ \Phi_0 \end{pmatrix} d_R], \qquad (7.4.1)$$

while the mass of the up quark $m_u = f_u v$ is due to the term

$$\mathcal{L}(u_L, d_L, d_R, \Phi) = f_d[\bar{u}_R(\Phi_0'^* \Phi_-'^*) \begin{pmatrix} u_L \\ d_L \end{pmatrix} + (\bar{u}_L \bar{d}_L) \begin{pmatrix} \Phi_0' \\ \Phi_-' \end{pmatrix} u_R]. \qquad (7.4.2)$$

In the standard model the Higgs field Φ' is constructed out of the Higgs field Φ as

$$\Phi' = \begin{pmatrix} \Phi_0' \\ \Phi_-' \end{pmatrix} = \begin{pmatrix} \Phi_0^* \\ -\Phi_+^* \end{pmatrix}. \qquad (7.4.3)$$

When we write the Higgs field as a matrix

$$\Phi = \begin{pmatrix} \Phi_0^* & \Phi_+ \\ -\Phi_+^* & \Phi_0 \end{pmatrix}, \qquad (7.4.4)$$

both Yukawa couplings can be combined into one expression

$$\mathcal{L}(u_L, d_L, u_R, d_R, \Phi) = (\bar{u}_R \bar{d}_R) \mathcal{F}^\dagger \Phi^\dagger \begin{pmatrix} u_L \\ d_L \end{pmatrix} + (\bar{u}_L \bar{d}_L) \Phi \mathcal{F} \begin{pmatrix} u_R \\ d_R \end{pmatrix}, \qquad (7.4.5)$$

where $\mathcal{F} = \mathrm{diag}(f_u, f_d)$ is the diagonal matrix of Yukawa couplings. When a θ-term is present in the QCD Lagrangian, it can be rotated into the matrix of Yukawa couplings by the transformation

$$\begin{pmatrix} u_L' \\ d_L' \end{pmatrix} = \exp(-i\theta/4) \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad \begin{pmatrix} u_R' \\ d_R' \end{pmatrix} = \exp(-i\theta/4) \begin{pmatrix} u_R \\ d_R \end{pmatrix}. \qquad (7.4.6)$$

This turns the matrix of Yukawa couplings into

$$\mathcal{F}' = \mathrm{diag}(f_u \exp(i\theta/2), f_d \exp(i\theta/2)). \qquad (7.4.7)$$

Since the Higgs field matrix Φ is proportional to an $SU(2)$ matrix, the complex phase $\exp(i\theta)$ of the matrix of Yukawa couplings cannot be absorbed into it, and hence θ cannot be rotated away. Here we have assumed that f_u and f_d are real. Otherwise, the effective vacuum angle would still be the complex phase of the determinant of $\mathcal{F}'$.

It is instructive to include the Yukawa couplings in the chiral perturbation theory action

$$S[U,\Phi] = \int d^4x \ \{\frac{F_\pi^2}{4}\text{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle \text{Tr}[\Phi\mathcal{F}'U^\dagger + U\mathcal{F}'^\dagger\Phi^\dagger]\}. \quad (7.4.8)$$

Again, the complex phase in $\mathcal{F}'$ cannot be absorbed into the Higgs field matrix Φ because it is proportional to an $SU(2)$ matrix. The Goldstone boson matrix U is also an $SU(2)$ matrix, and hence θ cannot be rotated away. As we have seen, θ can actually be rotated away if the Goldstone boson matrix is in $U(2)$ and contains the η'-meson field as a complex phase of its determinant. This, however, is the case only at large N_c.

The basic idea of Peccei and Quinn can be boiled down to extending the standard model Higgs field to a matrix proportional to $U(2)$ — not just to $SU(2)$. The extra $U(1)_{PQ}$ Peccei-Quinn symmetry then allows us to rotate θ away. The actual proposal of Peccei and Quinn does a bit more. It introduces two completely independent Higgs doublets Φ and Φ' which can be combined to form a $GL(2,\mathbb{C})$ matrix. Working with $GL(2,\mathbb{C})$ rather than with $U(2)$ matrices ensures that the Higgs sector is described by a perturbatively renormalizable linear σ-model, instead of a perturbatively nonrenormalizable nonlinear σ-model. Still, this is not too relevant since, as we have discussed earlier, both the linear and the nonlinear σ-model are trivial in the continuum limit, and physically equivalent. For simplicity, we will not follow Peccei and Quinn all the way and introduce two Higgs fields. Instead we will just extend the standard Higgs field to a matrix $\tilde{\Phi}$ proportional to a $U(2)$ matrix. This means that we introduce just one additional degree of freedom, while Peccei and Quinn introduced four. The complex phase in $\mathcal{F}'$ can then be absorbed in a redefinition of $\tilde{\Phi}$ and one obtains

$$S[U,\tilde{\Phi}] = \int d^4x \ \{\frac{F_\pi^2}{4}\text{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle \text{Tr}[\tilde{\Phi}\mathcal{F}U^\dagger + U\mathcal{F}^\dagger\tilde{\Phi}^\dagger]\}. \quad (7.4.9)$$

Since this expression now contains the original real Yukawa coupling matrix $\mathcal{F}$, all signs of the vacuum angle have completely disappeared from the theory. Instead the complex phase $\exp(ia/v)$ of $\tilde{\Phi}$ now plays the role of θ. In particular, the axion field $a(x)/v$ behaves like a space-time dependent θ-vacuum angle.

7.5 $U(1)_{PQ}$ Breaking and the Axion

The scalar potential $V(\tilde{\Phi})$ in the extension of the standard model is invariant against $SU(2)_L \otimes SU(2)_R \otimes U(1)_{PQ}$ transformations. In the vacuum the Higgs field takes the value $\tilde{\Phi} = \mathrm{diag}(v, v)$, which breaks this symmetry down spontaneously to $SU(2)_{L=R}$. Hence, there are $3+3+1-3=4$ massless Goldstone bosons. As usual, we then gauge $SU(2)_L$ as well as the $U(1)_Y$ subgroup of $SU(2)_R$, which amounts to a partial explicit breaking of $SU(2)_R$. The unbroken subgroup of $SU(2)_{L=R}$ then is just $U(1)_{em}$. Via the Higgs mechanism, three of the four Goldstone bosons are eaten by the gauge bosons and become the longitudinal components of Z_0 and $W^{\pm}$. Since $U(1)_{PQ}$ remains a global symmetry, the fourth Goldstone boson does not get eaten. This Goldstone boson is the axion.

Let us construct a low-energy effective theory that contains all Goldstone bosons of the extended standard model, namely the pions and the axion. This is easy to do, because we have already included $\tilde{\Phi}$ in the chiral Lagrangian. After spontaneous symmetry breaking at the electroweak scale v, we can write

$$\tilde{\Phi} = v\mathrm{diag}(\exp(ia/v), \exp(ia/v)), \tag{7.5.1}$$

where a parametrizes the axion field. Similarly, we can write

$$U = \mathrm{diag}(\exp(i\pi^0/F_\pi), \exp(-i\pi^0/F_\pi)). \tag{7.5.2}$$

Of course, the field U also contains the charged pions. At this point, we are interested in axion-pion mixing. Since the axion is electrically neutral, it cannot mix with the charged pions and we thus ignore them. Let us first search for the vacuum of the axion-pion system. Minimizing the energy implies maximizing

$$\mathrm{Tr}[\tilde{\Phi}\mathcal{F}U^\dagger + U\mathcal{F}^\dagger\tilde{\Phi}^\dagger] = m_u \cos(a/v + \pi^0/F_\pi) + m_d \cos(a/v - \pi^0/F_\pi). \tag{7.5.3}$$

Obviously, this expression is maximized for $a = \pi^0 = 0$. Next we expand around this vacuum to second order in the fields. The resulting mass squared matrix takes the form

$$M^2 = \frac{\langle\bar{\Psi}\Psi\rangle}{4} \begin{pmatrix} (m_u + m_d)/F_\pi^2 & (m_u - m_d)/F_\pi v \\ (m_u - m_d)/F_\pi v & (m_u + m_d)/v^2 \end{pmatrix}. \tag{7.5.4}$$

In the limit $v \to \infty$ this matrix turns into

$$M^2 = \frac{\langle\bar{\Psi}\Psi\rangle}{4} \begin{pmatrix} (m_u + m_d)/F_\pi^2 & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.5.5}$$

from which we read off the familiar mass squared of the pion

$$M_\pi^2 = \frac{\langle\bar{\Psi}\Psi\rangle(m_u + m_d)}{4F_\pi^2}. \quad (7.5.6)$$

In this limit the axion remains massless and there is no axion-pion mixing. Next, we keep v finite, but we still use $v \gg F_\pi$. Then there is a small amount of mixing between the axion and the pion, but the pion mass is to leading order unaffected. The determinant of the mass squared matrix is given by

$$\frac{\langle\bar{\Psi}\Psi\rangle^2}{16}[\frac{(m_u + m_d)^2}{F_\pi^2 v^2} - \frac{(m_u - m_d)^2}{F_\pi^2 v^2}] = \frac{4m_u m_d}{F_\pi^2 v^2} = M_\pi^2 M_a^2. \quad (7.5.7)$$

Hence, the axion mass squared is given by

$$M_a^2 = \frac{\langle\bar{\Psi}\Psi\rangle m_u m_d}{(m_u + m_d)v^2}. \quad (7.5.8)$$

It vanishes in the chiral limit, and even if just one of the quark masses is zero. The ratio of the axion and pion mass squares is

$$\frac{M_a^2}{M_\pi^2} = \frac{4m_u m_d F_\pi^2}{(m_u + m_d)^2 v^2}. \quad (7.5.9)$$

Hence, for $m_u = m_d$ we have

$$\frac{M_a}{M_\pi} = \frac{F_\pi}{v} \approx \frac{250\text{GeV}}{0.1\text{GeV}} = 2500 \ \Rightarrow \ M_a \approx \frac{0.14\text{GeV}}{2500} \approx 50\text{keV}. \quad (7.5.10)$$

This is an unusually light particle that should have observable effects. Indeed, there have been several experimental searches for this "standard" axion, but they did not find anything. The only exception was an experiment performed in Aachen (Germany). The signature they saw was called the "Aachion" but it was not confirmed by other experiments, and the standard axion with a mass around 50keV has actually been ruled out. Still, by pushing the $U(1)_{PQ}$ breaking scale far above the electroweak scale one can make the axion more weakly coupled and thus make it invisible to all experiments performed so far. Presently, there are still searches going on that attempt to detect the "invisible" axion. Before they are successful, we cannot be sure that Peccei and Quinn's elegant solution of the strong CP problem is actually correct.

Invisible axions are light and interact only weakly, because the axion coupling constants are proportional to the mass. Still, unless they are too light — and thus too weakly interacting — axions can cool stars very efficiently. Their low interaction cross section allows them to carry away energy more easily than the more strongly coupled photon. A sufficiently interacting axion could shorten the life-time of stars by a substantial amount. From the observed life-time one can hence infer an upper limit on the axion mass. In this way invisible axions heavier than 1 eV have been ruled out. This implies that the Peccei-Quinn symmetry breaking scale must be above 10^7 GeV. If they exist, axions would also affect the cooling of a neutron star that forms after a supernova explosion. There would be less energy taken away by neutrinos. The observed neutrino burst of the supernova SN 1987A would have consisted of fewer neutrinos if axions had also cooled the neutron star. This astrophysical observation excludes axions of masses between 10^{-3} and 0.02 MeV — a range that cannot be investigated in the laboratory.

There are various mechanisms in the early Universe that can lead to the generation of axions. The simplest is via thermal excitation. One can estimate that thermally generated axions must be rather heavy in order to contribute substantially to the energy density of the Universe. In fact, thermal axions cannot close the Universe, because the required mass is already ruled out by the astrophysical limits. Another interesting mechanism for axion production relies on the fact that the axion potential is not completely flat but has a unique minimum. At high temperatures the small axion mass is irrelevant, the potential is practically flat and corresponds to a family of degenerate minima related to one another by $U(1)_{PQ}$ symmetry transformations. Hence, there are several degenerate vacua labeled by different values of the axion field (hence with different values of θ) and all values of θ are equally probable. Then different regions of the hot early Universe must have been in different θ-vacua. When the temperature decreases, $a = 0$ is singled out as the unique minimum. In order to minimize its energy, the scalar field then "rolls" down to this minimum, and oscillates about it. The oscillations are damped by axion emission, and finally $a = 0$ is reached everywhere in the Universe. The axions produced in this way would form a Bose condensate that could close the Universe for an axion mass in the 10^{-5} eV range. This makes the axion an attractive candidate for dark matter in the Universe. This axion production mechanism via a disoriented Peccei-Quinn condensate is very similar to the pion-production mechanism that has been discussed via disorienting the chiral condensate in a heavy ion collision. At temperatures high above the QCD scale, $U(1)_{PQ}$ is almost an exact global symmetry, which gets spontaneously broken at some high scale. This necessarily leads to the generation of a network of cosmic strings. Such a string network can lower its energy by radiating axions. Once the axion mass becomes important, the string solutions become unstable, and the string network disappears, again leading to axion emission. This production mechanism may also lead to enough axions to close the Universe.

7.6 Axionic Strings and Domain Walls

When a $U(1)$ symmetry breaks spontaneously, topological excitations may arise. In a type II superconductor, for example, the $U(1)$ gauge symmetry of electromagnetism is spontaneously broken. When the superconductor is placed in an external magnetic field, magnetic flux tubes — so-called Abrikosov strings — are formed, which are 1-dimensional topological excitations. Whenever a $U(1)$ symmetry (global or local) is spontaneously broken, this phenomenon arises. In the framework of particle physics this was first discussed by Nielsen and Olesen. The corresponding 1-dimensional topological excitations are therefore known as Nielsen-Olesen strings. When a $U(1)$ symmetry is spontaneously broken in the early Universe, 1-dimensional topological excitations arise, which may pass through the entire Universe. Networks of these so-called cosmic strings are discussed as a seed for galaxy formation. Cosmic strings also act as gravitational lenses, and they have been used in Gedanken experiments related to time machines. Still, cosmic strings are purely theoretical objects, which have not yet been observed in the sky.

Let us first consider a simple scalar field theory with a global $U(1)$ symmetry, which is related to phase transformations of a complex scalar field $\Phi(x) \in \mathbb{C}$. The corresponding Lagrange density is

$$\mathcal{L}(\Phi, \partial_\mu \Phi) = \partial^\mu \Phi^* \partial_\mu \Phi - V(\Phi), \ V(\Phi) = m^2 |\Phi|^2 + \lambda |\Phi|^4. \tag{7.6.1}$$

When $m^2 < 0$ the $U(1)$ symmetry gets spontaneously broken, and we obtain one massless Goldstone boson. Let us now consider cylindrically symmetric field configurations

$$\Phi(x) = \Phi(\vec{x}, t) = \Phi(\vec{x}) = \Phi(\rho, \varphi), \tag{7.6.2}$$

i.e. in cylindrical coordinates ρ, φ, z the field is independent of z. At $\rho = \infty$ the field must assume its vacuum value, i.e. $|\Phi| = v$, in order to have finite energy. However, it may have a φ-dependent complex phase $\chi(\varphi)$

$$\Phi(\rho = \infty, \varphi) = v \exp(i\chi(\varphi)). \tag{7.6.3}$$

Points at infinity are parametrized by the polar angle φ, which is topologically a circle S^1. In this case the vacuum manifold is also a circle, because $\mathcal{M} = U(1) = S^1$. It is parametrized by the angle χ. Hence, the field values at infinity $\Phi(\rho = \infty, \varphi)$ can be viewed as a map

$$\Phi : S^1 \to \mathcal{M}. \tag{7.6.4}$$

We know that such maps are characterized by winding numbers in the homotopy group

$$\Pi_1[\mathcal{M}] = \Pi_1[S^1] = \mathbb{Z}. \tag{7.6.5}$$

Cosmic strings are field configurations with non-vanishing winding number. They are topological stable since the winding cannot be eliminated by continuous deformations.

The simplest cosmic string has winding number 1. One can then use the identity map

$$\chi(\varphi) = \varphi. \tag{7.6.6}$$

Let us make the ansatz

$$\Phi(\rho, \varphi) = f(\rho)\exp(i\varphi), \;\; f(\infty) = v. \tag{7.6.7}$$

At the z-axis at $\rho = 0$ the angle φ is not well-defined (coordinate singularity). For the field Φ to be well-defined, we must therefore demand $f(0) = 0$. Inserting this ansatz in the classical equations of motion

$$\partial_\mu \partial^\mu \Phi + m^2\Phi + 2\lambda|\Phi|^2\Phi = 0, \tag{7.6.8}$$

one obtains

$$\begin{aligned}
&\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho\frac{\partial\Phi}{\partial\rho}) + \frac{1}{\rho^2}\frac{\partial^2\Phi}{\partial\rho^2} = m^2\Phi + 4\lambda|\Phi|^2\Phi \Rightarrow \\
&\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho f')\exp(i\varphi) - \frac{1}{\rho^2} f\exp(i\varphi) = m^2 f\exp(i\varphi) + 2\lambda f^3\exp(i\varphi) \Rightarrow \\
&f'' + \frac{1}{\rho}f' - \frac{1}{\rho^2}f = m^2 f + 2\lambda f^3.
\end{aligned} \tag{7.6.9}$$

Although this equation cannot be solved analytically, it is easy to solve it numerically. The result is a monotonically rising function with $f(0) = 0$ and $f(\infty) = v = \sqrt{-m^2/2\lambda}$. The energy of the solution is concentrated around the z-axis, where $|\Phi|$ deviates from the vacuum value. The energy per unit length in the z-direction is known as the string tension μ. Due to its coupling to the massless Goldstone boson the tension of the string diverges logarithmically with its length.

The solution that we just discussed represents a so-called global cosmic string. Such a string cannot simply end. Either it is closed, or it extends through the whole Universe. When a $U(1)$ gauge symmetry gets spontaneously broken, string-like solutions still exist. They are then called local cosmic strings. Local strings interact with the massive gauge bosons that result when the massless Goldstone boson is eaten. As a consequence, the infrared logarithmic divergence of the string tension of global strings disappears. Indeed the string tension of a local cosmic string is finite.

The relevant homotopy group that decides about the existence of topologically stable cosmic string solutions is $\Pi_1[\mathcal{M}]$. In general, a group G may get spontaneously broken to a subgroup H. Then $\mathcal{M} = G/H$ and the corresponding homotopy group is

$$\Pi_1[\mathcal{M}] = \Pi_1[G/H]. \tag{7.6.10}$$

In our example we had $G = U(1)$ and $H = \{\mathbb{1}\}$ such that

$$\Pi_1[\mathcal{M}] = \Pi_1[U(1)/\{\mathbb{1}\}] = \Pi_1[U(1)] = \Pi_1[S^1] = \mathbb{Z}. \tag{7.6.11}$$

In the standard model, on the other hand, $G = SU(3) \otimes SU(2) \otimes U(1)$ and $H = SU(3) \otimes U(1)$, such that

$$\Pi_1[\mathcal{M}] = \Pi_1[SU(3) \otimes SU(2) \otimes U(1)/SU(3) \otimes U(1)] = \Pi_1[SU(2)] = \Pi_1[S^3] = \{0\}. \tag{7.6.12}$$

Hence, there are no stable cosmic string solutions in the standard model. In the Peccei-Quinn extension of the standard model cosmic strings appear naturally because the $U(1)_{PQ}$ symmetry breaks spontaneously. Then a network of cosmic strings emerges that extends throughout the whole Universe.

Cosmic strings have interesting gravitational effects, because they curve space to a cone. The resulting deficit angle leads to a gravitational lensing effect. This may eventually lead to their observation. Due to their gravitational effects, cosmic strings may act as seeds for galaxy formation. Just like the normal conducting state is restored inside an Abrikosov flux string in a superconductor, the vacuum can become superconducting inside a cosmic string. Such superconducting cosmic strings occur in models with two coupled $U(1)$ symmetries, one of them the symmetry of electromagnetism, which is unbroken in the vacuum. The other $U(1)$ is broken and gives rise to the cosmic string. Inside the string the unbroken vacuum of this $U(1)$ is restored, but this may now lead to a breaking of electromagnetism. Superconducting cosmic strings can lead to large electric fields. They are dangerous objects, because they can explode. People have even tried to "build" time-machines using colliding cosmic strings. Unfortunately, in the examples studied so far the Universe always disappeared in a big crunch before the travel around a closed time-like curve could be completed. Clearly, cosmic strings are interesting objects for theoretical study. If they exist in our Universe is an open question that can only be decided by observation.

We have seen that non-zero quark masses lead to a non-zero axion mass. Hence, the axion potential is not completely flat. In fact, the unique minimum of the potential is at $a = 0$. However, a/v behaves like a space-time dependent θ-vacuum angle and hence as a periodic variable. This gives rise to cosmic axionic domain wall solutions. A domain wall is a 2-dimensional topological defect that interpolates between two degenerate vacuum states. In the case of axionic domain walls, the two states are actually physically indistinguishable and just correspond to a shift of a/v by 2π. If axionic cosmic strings exist at some high temperature in the early Universe, they will disappear as the Universe expands and cools. This is because an axionic domain wall will form between the strings as the temperature drops to the QCD scale. Then the axion potential can no longer be viewed as essentially flat, and axionic strings become unstable.

Chapter 8

Grand Unified Theories

Although the standard model of particle physics agrees very well with experiments, from a theoretical point of view it is not completely satisfactory. In particular, the large number of free parameters like fermion masses (Yukawa couplings) or gauge boson masses (gauge couplings) suggests the existence of a more fundamental underlying theory, which would allow to explain these parameters. Also gravity is not included in the standard model, which again points to its incompleteness. Even electromagnetism and the weak force are not really unified in the standard model, because there are two independent coupling constants g and g'. QCD has yet another coupling constant g_s, which is unrelated to g and g'. In the framework of grand unified theories (GUT) one embeds the electroweak and strong interactions in one simple gauge group (e.g. $SU(5)$), which leads to a relation between g, g' and g_s.

Of course, the $SU(5)$ symmetry is too big, in order to be realized at low temperatures. It must be spontaneously broken to the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ symmetry of the standard model. In grand unified theories this happens at about 10^{14} GeV at about 10^{-34} sec after the Big Bang. At present (and in the foreseeable future) these energy scales cannot be probed experimentally. Hence, we must now rely on theoretical arguments, and sometimes on speculation.

The unification of the electroweak and strong interactions naturally leads to the decay of strongly interacting particles into particles which participate in the electroweak interactions only, for example, quarks can decay into leptons. This inevitably leads to the decay of the proton, and hence to a violation of baryon number conservation. Experimentally, the proton is extremely long lived (life-time $> 10^{32}$ years). This has already ruled out the simplest version of the

$SU(5)$ theory. Still, we will discuss that model in some detail, because it is the simplest representative in a larger class of GUTs. Other GUT theories allow larger life-times of the proton, and are not ruled out experimentally. On the other hand, proton decay is a very attractive feature of GUTs, because it may offer an explanation of the baryon asymmetry — the fact that our Universe consists of matter — not of anti-matter.

8.1 The minimal $SU(5)$ Model

The group $SU(n)$ has rank $n-1$, i.e. $n-1$ of the n^2-1 generators commute with each other. The rank of a simple $U(1)$ group is 1. Thus, the rank of the standard model group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ is $2+1+1=4$. Hence, if we want to embed that group in a simple Lie group, its rank must be at least 4. The simplest Lie group with that property is $SU(5)$, which has rank 4 and $5^2-1=24$ generators. Consequently, in an $SU(5)$ gauge theory there are 24 gauge bosons. When the standard model is embedded in $SU(5)$, half of the gauge bosons can be identified with known particles: $SU(3)$ has $3^2-1=8$ gluons, $SU(2)$ has $2^2-1=3$ W-bosons, and $U(1)$ has one B-boson, which, together with W^3, forms the Z-boson and the photon. The remaining 12 gauge bosons of $SU(5)$ are new particles, which are called X and Y. To make these particles heavy, the $SU(5)$ symmetry must be spontaneously broken to $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$. Again, this is achieved via the Higgs mechanism, here with a scalar field transforming under the 24-dimensional adjoint representation of $SU(5)$. We write

$$\Phi(x) = \Phi^a(x)\eta_a, \ a \in \{1, 2, ..., 24\}. \tag{8.1.1}$$

The η_a are the 24 generators of $SU(5)$. They are Hermitean, traceless 5×5 matrices, analogous to the 3 Pauli matrices of $SU(2)$. Under gauge transformations $g \in SU(5)$ the scalar field transforms as

$$\Phi(x)' = g(x)\Phi(x)g(x)^\dagger. \tag{8.1.2}$$

We introduce a Φ^4 potential, which takes the form

$$V(\Phi) = \frac{1}{2}m^2 \mathrm{Tr}(\Phi^2) + \lambda_1(\mathrm{Tr}(\Phi^2))^2 + \lambda_2 \mathrm{Tr}(\Phi^4). \tag{8.1.3}$$

The potential is gauge invariant due to the cyclic nature of the trace. We can choose a unitary gauge, in which the scalar field is diagonal (one uses the unitary transformation $g(x)$ to diagonalize the Hermitean matrix $\Phi(x)$)

$$\Phi = \begin{pmatrix} \Phi_1 & 0 & 0 & 0 & 0 \\ 0 & \Phi_2 & 0 & 0 & 0 \\ 0 & 0 & \Phi_3 & 0 & 0 \\ 0 & 0 & 0 & \Phi_4 & 0 \\ 0 & 0 & 0 & 0 & \Phi_5 \end{pmatrix}, \ \Phi_i \in \mathbb{R}, \ \sum_i \Phi_i = 0. \tag{8.1.4}$$

The potential then takes the form

$$V(\Phi) = \frac{1}{2} \sum_i \Phi_i^2 + \lambda_1 (\sum_i \Phi_i^2)^2 + \lambda_2 \sum_i \Phi_i^4. \tag{8.1.5}$$

The minima of the potential are characterized by

$$\frac{\partial V}{\partial \Phi_i} = m^2 \Phi_i + 4\lambda_1 \sum_j \Phi_j^2 \Phi_i + 4\lambda_2 \Phi_i^3 = C. \tag{8.1.6}$$

Here C is a Lagrange multiplier that implements the constraint $\sum_i \Phi_i = 0$. We are interested in minima with an unbroken $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ symmetry, for which

$$\Phi_1 = \Phi_2 = \Phi_3, \ \Phi_4 = \Phi_5. \tag{8.1.7}$$

Hence, we can write

$$\Phi = v \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 \end{pmatrix}, \tag{8.1.8}$$

such that

$$\begin{aligned} & m^2 v + 4\lambda_1 v^2 (3 + 2\frac{9}{4}) v + 4\lambda_2 v^3 = C, \\ & -\frac{3}{2} m^2 v - 4\lambda_1 v^2 (3 + 2\frac{9}{4}) \frac{3}{2} v - 4\lambda_2 v^3 \frac{27}{8} = C \ \Rightarrow \\ & C = \frac{4}{5} \lambda_2 v^3 (3 - \frac{27}{4}) = -3\lambda_2 v^3 \ \Rightarrow \\ & m^2 v + \lambda_1 30 v^3 + \lambda_2 7 v^3 = 0 \ \Rightarrow \ v = \sqrt{-\frac{m^2}{30\lambda_1 + 7\lambda_2}}. \end{aligned} \tag{8.1.9}$$

The value of the potential at the minimum is given by

$$\begin{aligned} V(\Phi) &= \frac{1}{2}m^2v^2(3+2\frac{9}{4}) + \lambda_1 v^4(3+2\frac{9}{4})^2 + \lambda_2 v^4(3+2\frac{81}{16}) \\ &= \frac{1}{2}m^2v^2\frac{15}{2} + \lambda_1 v^4\frac{225}{4} + \lambda_2 v^4\frac{105}{8}) \\ &= -\frac{15}{4}v^4(30\lambda_1+7\lambda_2) + \lambda_1 v^4\frac{225}{4} + \lambda_2 v^4\frac{105}{8}) \\ &= v^4(-\frac{225}{4}\lambda_1 - \frac{105}{8}\lambda_2) = -m^4\frac{15}{8}\frac{1}{30\lambda_1+7\lambda_2}. \end{aligned} \tag{8.1.10}$$

For $\lambda_1, \lambda_2 > 0$ the value of the potential is negative, indicating that the $SU(5)$ symmetric phase at $\Phi = 0$ with $V(\Phi) = 0$ is not the true vacuum. It is instructive to convince oneself that other symmetry breaking patterns — for example to $SU(4) \otimes U(1)$ — are not dynamically prefered over $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ breaking.

Let us now consider the gauge fields

$$V_\mu = ig_5 V_\mu^a(x)\eta_a. \tag{8.1.11}$$

Under non-Abelian gauge transformations we have

$$V_\mu(x)' = g(x)(V_\mu(x) + \partial_\mu)g(x)^\dagger. \tag{8.1.12}$$

For an adjoint Higgs field the covariant derivative takes the form

$$D_\mu\Phi(x) = \partial_\mu\Phi(x) + [V_\mu(x), \Phi(x)]. \tag{8.1.13}$$

It is instructive to show that this indeed transforms covariantly. Introducing the field strength tensor

$$V_{\mu\nu} = \partial_\mu V_\nu(x) - \partial_\nu V_\mu(x) + [V_\mu(x), V_\nu(x)], \tag{8.1.14}$$

the bosonic part of the $SU(5)$ GUT Lagrange density takes the form

$$\mathcal{L}(\Phi, V_\mu) = \frac{1}{2}\text{Tr}D^\mu\Phi D_\mu\Phi - V(\Phi) - \frac{1}{4}\text{Tr}V^{\mu\nu}V_{\mu\nu}. \tag{8.1.15}$$

Next we insert the vacuum value of the scalar field to obtain the mass terms for the gauge field

$$\frac{1}{2}\text{Tr}D^\mu\Phi D_\mu\Phi = \text{Tr}[V^\mu, \Phi][V_\mu, \Phi]. \tag{8.1.16}$$

We introduce the X and Y-bosons via

$$V_\mu = \begin{pmatrix} & G_\mu & & X_\mu^r & Y_\mu^r \\ & & & X_\mu^g & Y_\mu^g \\ & & & X_\mu^b & Y_\mu^b \\ X_\mu^{r*} & X_\mu^{g*} & X_\mu^{b*} & & \\ Y_\mu^{r*} & Y_\mu^{g*} & Y_\mu^{b*} & & W_\mu \end{pmatrix}. \tag{8.1.17}$$

X and Y-bosons are color triplets and electroweak doublets. They are the fields that become massive after the spontaneous breakdown of $SU(5)$ to $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, because one obtains

$$\begin{aligned} [V_\mu, \Phi] &= v \begin{pmatrix} & G_\mu \mathbb{1} & & -\frac{3}{2}X_\mu^r & -\frac{3}{2}Y_\mu^r \\ & & & -\frac{3}{2}X_\mu^g & -\frac{3}{2}Y_\mu^g \\ & & & -\frac{3}{2}X_\mu^b & -\frac{3}{2}Y_\mu^b \\ X_\mu^{r*} & X_\mu^{g*} & X_\mu^{b*} & & \\ Y_\mu^{r*} & Y_\mu^{g*} & Y_\mu^{b*} & & -\frac{3}{2}W_\mu \mathbb{1} \end{pmatrix} \\ &- v \begin{pmatrix} & G_\mu \mathbb{1} & & X_\mu^r & Y_\mu^r \\ & & & X_\mu^g & Y_\mu^g \\ & & & X_\mu^b & Y_\mu^b \\ -\frac{3}{2}X_\mu^{r*} & -\frac{3}{2}X_\mu^{g*} & -\frac{3}{2}X_\mu^{b*} & & \\ -\frac{3}{2}Y_\mu^{r*} & -\frac{3}{2}Y_\mu^{g*} & -\frac{3}{2}Y_\mu^{b*} & & -\frac{3}{2}W_\mu \mathbb{1} \end{pmatrix} \\ &= v \begin{pmatrix} & 0 & & -\frac{5}{2}X_\mu^r & -\frac{5}{2}Y_\mu^r \\ & & & -\frac{5}{2}X_\mu^g & -\frac{5}{2}Y_\mu^g \\ & & & -\frac{5}{2}X_\mu^b & -\frac{5}{2}Y_\mu^b \\ -\frac{5}{2}X_\mu^{r*} & -\frac{5}{2}X_\mu^{g*} & -\frac{5}{2}X_\mu^{b*} & & \\ -\frac{5}{2}Y_\mu^{r*} & -\frac{5}{2}Y_\mu^{g*} & -\frac{5}{2}Y_\mu^{b*} & & 0 \end{pmatrix}, \end{aligned} \tag{8.1.18}$$

and hence

$$\mathrm{Tr}[V^\mu, \Phi][V_\mu, \Phi] = -\frac{25}{2} v^2 (X^{\mu *} X_\mu + Y^{\mu *} Y_\mu). \tag{8.1.19}$$

The X and Y-fields thus pick up the mass

$$m_X^2 = m_Y^2 = \frac{25}{2} v^2 g_5^2. \tag{8.1.20}$$

The 12 gauge bosons become massive by eating 12 Goldstone bosons. Indeed, due to the Goldstone theorem, in this case their are $24-8-3-1=12$ Goldstone bosons.

8.2 The Fermion Multiplets

How can we arrange quarks and leptons in representations of $SU(5)$? Let us consider the fermions of the first generation

$$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \; e_R, \; \begin{pmatrix} u_L^r \\ d_L^r \end{pmatrix}, \; u_R^r, \; d_R^r, \; \begin{pmatrix} u_L^g \\ d_L^g \end{pmatrix}, \; u_R^g, \; d_R^g, \; \begin{pmatrix} u_L^b \\ d_L^b \end{pmatrix}, \; u_R^b, \; d_R^b. \quad (8.2.1)$$

These are 15 fermionic degrees of freedom. The fundamental representation of $SU(5)$ is 5-dimensional. It decomposes into an $SU(3)$ triplet and an $SU(2)$ doublet with corresponding $U(1)$ quantum numbers

$$\{5\} = \{3,1\}_{-2/3} \oplus \{1,2\}_1. \quad (8.2.2)$$

Let us couple two fundamental representations in $SU(2)$, $SU(3)$ and $SU(5)$

$$\begin{aligned} \{2\} \otimes \{2\} &= \{3\} \oplus \{1\}, \\ \{3\} \otimes \{3\} &= \{6\} \oplus \{\bar{3}\}, \\ \{5\} \otimes \{5\} &= \{15\} \oplus \{10\}. \end{aligned} \quad (8.2.3)$$

Indeed, there is a 15-dimensional representation. Can we host the fermions of a generation in that representation? We investigate the $SU(3)$ and $SU(2)$ content

$$\begin{aligned} \{5\} \otimes \{5\} &= (\{3,1\}_{-2/3} \oplus \{1,2\}_1) \otimes (\{3,1\}_{-2/3} \oplus \{1,2\}_1) \\ &= \{3,1\}_{-2/3} \otimes \{3,1\}_{-2/3} \oplus \{1,2\}_1 \otimes \{3,1\}_{-2/3} \\ &\oplus \{3,1\}_{-2/3} \otimes \{1,2\}_1 \oplus \{1,2\}_1 \otimes \{1,2\}_1 \\ &= \{6,1\}_{-4/3} \oplus \{\bar{3},1\}_{-4/3} \oplus \{3,2\}_{1/3} \oplus \{3,2\}_{1/3} \\ &\oplus \{1,3\}_2 \oplus \{1,1\}_2 = \{15\} \oplus \{10\}. \end{aligned} \quad (8.2.4)$$

The symmetric combination $\{15\}$ contains

$$\{15\} = \{6,1\}_{-4/3} \oplus \{3,2\}_{1/3} \oplus \{1,3\}_2, \quad (8.2.5)$$

while the anti-symmetric combination $\{10\}$ is

$$\{10\} = \{\bar{3},1\}_{-4/3} \oplus \{3,2\}_{1/3} \oplus \{1,1\}_2. \quad (8.2.6)$$

The standard model does not contain fermions in a sextet representation of $SU(3)$ (the quarks are triplets and the leptons are singlets). Hence the 15 fermions of a generation do not form a 15-plet of $SU(5)$. We have

$$\begin{aligned} &\{\begin{pmatrix} u_L^r \\ d_L^r \end{pmatrix}, \begin{pmatrix} u_L^g \\ d_L^g \end{pmatrix}, \begin{pmatrix} u_L^b \\ d_L^b \end{pmatrix}\} = \{3,2\}_{1/3}, \; \{u_R^r, u_R^g, u_R^b\} = \{3,1\}_{4/3}, \\ &\{d_R^r, d_R^g, d_R^b\} = \{3,1\}_{-2/3}, \; \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} = \{1,2\}_{-1}, \; e_R = \{1,1\}_{-2}. \end{aligned} \quad (8.2.7)$$

Both $\{3,2\}$ and $\{1,1\}$ are contained in $\{10\}$. However, there is also a $\{\bar{3},1\}$. Furthermore, we cannot mix left and right-handed fermion components. We use charge conjugation to write everything in terms of left-handed fields

$$\begin{aligned}
&\{\left(\begin{array}{c} u_L^r \\ d_L^r \end{array}\right), \left(\begin{array}{c} u_L^g \\ d_L^g \end{array}\right), \left(\begin{array}{c} u_L^b \\ d_L^b \end{array}\right)\} = \{3,2\}_{1/3}, \\
&\{{}^C u_R^r, {}^C u_R^g, {}^C u_R^b\} = \{\bar{3},1\}_{-4/3}, \ \{{}^C d_R^r, {}^C d_R^g, {}^C d_R^b\} = \{\bar{3},1\}_{2/3}, \\
&\left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right) = \{1,2\}_{-1}, {}^C e_R = \{1,1\}_2.
\end{aligned} \tag{8.2.8}$$

Now we can identify a decouplet

$$\begin{aligned}
&\{\left(\begin{array}{c} u_L^r \\ d_L^r \end{array}\right), \left(\begin{array}{c} u_L^g \\ d_L^g \end{array}\right), \left(\begin{array}{c} u_L^b \\ d_L^b \end{array}\right)\} \oplus \{{}^C u_R^r, {}^C u_R^g, {}^C u_R^b\} \oplus {}^C e_R = \\
&\{3,2\}_{1/3} \oplus \{\bar{3},1\}_{-4/3} \oplus \{1,1\}_2 = \{10\}.
\end{aligned} \tag{8.2.9}$$

The remaining five fermions are

$$\{{}^C d_R^r, {}^C d_R^g, {}^C d_R^b\} = \{\bar{3},1\}_{2/3}, \ \left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right) = \{1,2\}_{-1}. \tag{8.2.10}$$

They naturally fit into an anti-quintet

$$\{\bar{5}\} = \{\bar{3},1\}_{2/3} \oplus \{1,\bar{2}\}_{-1}. \tag{8.2.11}$$

In $SU(2)$ the representations $\{2\}$ and $\{\bar{2}\}$ are equivalent, such that

$$\{{}^C d_R^r, {}^C d_R^g, {}^C d_R^b\} \oplus \left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right) = \{\bar{3},1\}_{2/3} \oplus \{1,2\}_{-1} = \{\bar{5}\}. \tag{8.2.12}$$

Hence, the fermions of one generation form a reducible 15-dimensional representation of $SU(5)$, which decomposes into $\{10\}$ and $\{\bar{5}\}$. In $SU(5)$ quarks and leptons are in the same irreducible representation. This immediately explains why proton and positron have the same electric charge — a fact that remains unexplained in the standard model.

Let us also determine the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ properties of the X and Y-bosons. They transform under the adjoint representation — the 24-plet, that is obtained from

$$\{5\} \otimes \{\bar{5}\} = \{24\} \oplus \{1\}. \tag{8.2.13}$$

On the other hand, we have

$$\begin{aligned}\{5\} \otimes \{\bar{5}\} &= (\{3,1\}_{-2/3} \oplus \{1,2\}_1) \otimes (\{\bar{3},1\}_{2/3} \oplus \{1,2\}_{-1}) \\ &= \{3,1\}_{-2/3} \otimes \{\bar{3},1\}_{2/3} \oplus \{1,2\}_1 \otimes \{\bar{3},1\}_{2/3} \\ &\oplus \{3,1\}_{-2/3} \otimes \{1,2\}_{-1} \oplus \{1,2\}_1 \otimes \{1,2\}_{-1} \\ &= \{8,1\}_0 \oplus \{1,1\}_0 \oplus \{\bar{3},2\}_{5/3} \oplus \{3,2\}_{-5/3} \oplus \{1,3\}_0 \oplus \{1,1\}_0.\end{aligned} \tag{8.2.14}$$

Hence, we identify

$$\begin{aligned}\{24\} &= \{8,1\}_0 \oplus \{1,3\}_0 \oplus \{1,1\}_0 \oplus \{\bar{3},2\}_{5/3} \oplus \{3,2\}_{-5/3} \\ &= \{\text{Gluons}\} \oplus \{\text{W-bosons}\} \oplus \{\text{B-boson}\} \oplus \{\text{X,Y-bosons}\}.\end{aligned} \tag{8.2.15}$$

The X and Y-bosons are $SU(3)$ triplets, $SU(2)$ doublets, and have hypercharge 5/3.

Up to this point all fermions and all gauge bosons of the standard model are still massless. Of course, we could write Yukawa couplings and try to give mass to the fermions by coupling them to the adjoint scalar field. However, it turns out that this is inconsistent with the group theory of $SU(5)$. In principle, this is good news, because otherwise the fermions would naturally get masses at the GUT scale. However, this also means that we will need more scalar fields, in particular the $SU(2)$ complex doublet of the standard model, which is naturally contained in the fundamental representation of $SU(5)$. Hence, we introduce a 5-plet of scalar fields

$$H = \begin{pmatrix} H_1 \\ H_2 \\ H_3 \\ H_4 \\ H_5 \end{pmatrix}, \ H_4 = \Phi_+, \ H_5 = \Phi_0. \tag{8.2.16}$$

Now we can write down various Yukawa couplings. The group theory of $SU(5)$ implies

$$\{\bar{5}\} \otimes \{10\} = \{5\} \oplus \{45\}, \ \{10\} \otimes \{10\} = \{\bar{5}\} \oplus \{\bar{45}\} \oplus \{50\}. \tag{8.2.17}$$

Hence, the fermions in the $\{\bar{5}\}$ and $\{10\}$ representations can be coupled to $\{5\}$ and $\{\bar{5}\}$ — i.e. together with the scalar 5-plet they can be coupled in a gauge invariant way (form an $SU(5)$ singlet).

8.3 Predictions of Grand Unified Theories

In the $SU(5)$ GUT there is only one gauge coupling g_5. Hence, it must be possible to relate the standard model gauge couplings g, g' and g_s to that coupling. In an $SU(5)$ symmetric phase one has

$$g = g_s = g_5, \; g' = \sqrt{\frac{3}{5}} g_5. \tag{8.3.1}$$

Hence, the Weinberg angle takes the form

$$\sin^2 \theta_W = \frac{g'^2}{g^2 + g'^2} = \frac{3g_5^2}{5g_5^2 + 3g_5^2} = \frac{3}{8}. \tag{8.3.2}$$

Furthermore, the structure of the Yukawa couplings implies

$$m_d = m_e, \; m_s = m_\mu, \; m_b = m_\tau. \tag{8.3.3}$$

This is not in agreement with experiments. However, we do not live in an $SU(5)$ symmetric world. One can use the renormalization group to run the above relations from the GUT scale, where they apply, down to our low energy scales. One obtains realistic values for particle masses and coupling constants when one puts the GUT scale at about $v = 10^{15}$ GeV. The masses of the X and Y-bosons are also in that range. The GUT scale is significantly below the Planck scale 10^{19} GeV, which justifies neglecting gravity in the above considerations.

An important prediction of GUTs is the instability of the proton. Proton decay proceeds via the X and Y-boson channel and is hence suppressed with the large mass of these particles. The following decays are possible

$$p \rightarrow e^+ + \pi^0, \; p \rightarrow \bar{\nu}_e + \pi^+. \tag{8.3.4}$$

The resulting lifetime of the proton is given by

$$\tau_p \propto \frac{m_X^4}{m_p^5} \approx 10^{32} \text{years}. \tag{8.3.5}$$

Although the predicted lifetime is much larger than the age of the Universe, the minimal $SU(5)$ model has been ruled out experimentally, because the proton indeed lives longer than the model predicts. More complicated GUTs based on $SO(10)$ or E_6 are not yet ruled out experimentally.

8.4 Evidence for a Baryon Asymmetry

Why is there a baryon asymmetry in the Universe? In other words, why does the Universe consist of matter, and not also of anti-matter? The ratio of baryon to photon density $n_B/n_\gamma \approx 10^{-10}$ is an initial condition of the standard Big Bang model, which cannot be derived within it. Grand unified theories, however, allow us to compute a baryon asymmetry based on the presence of baryon number violating processes (like proton decay). This does alone does not guarantee a baryon asymmetry. Besides baryon number, also C and CP must be violated, and the Universe must get out of thermal equilibrium. All these additional conditions are indeed satisfied in our Universe. Since the standard model is a chiral gauge theory, C is maximally violated, and also CP is explicitly broken (at least weakly). Furthermore, the expansion of the Universe implies that a true thermal equilibrium is never achieved. The Universe cools down and certain processes (for example those that violate baryon number) get out of equilibrium.

Matter and anti-matter annihilate each other, for example, into photons. Regions in the Universe, in which matter and anti-matter systems collide, should thus be a source of very intense x-ray radiation. Such a thing has not been observed, indicating that the Universe consists of matter only, and not also of anti-matter. This is confirmed by the composition of cosmic rays, which contain 10000 times more protons than anti-protons, and even those few are due to secondary processes.

Of course, in the early Universe anti-matter has also been present. In particular, at temperatures above the GeV range, about the same number of quarks and anti-quarks have been around. During cooling quarks and anti-quarks annihilated each other, and all matter in the Universe today is a tiny fraction that survived the mass extinction. Comparing observed abundances of light nuclei with theoretical calculations of primordial nucleosynthesis leads to $n_B/n_\gamma \approx 10^{-10}$.

Since our Universe is electrically neutral, it contains as many electrons as protons, but almost no positrons. Correspondingly, there should also be a lepton asymmetry. However, since also the weakly interacting neutrinos contribute to it, it can at present not be detected experimentally.

8.5 Necessary Conditions for a Baryon Asymmetry

A trivial condition for the explanation of the baryon asymmetry is the existence of baryon number violating processes. Only then an initial state with unknown baryon number (for example $B = 0$), may turn into the situation that we observe now. GUTs indeed give rise to these processes. At temperatures in the 10^{14} GeV range they occur frequently, while today they are extremely suppressed. In the ultra-early Universe baryon number violating processes have been in thermal equilibrium, such that the effects of even earlier initial conditions are eliminated. In the $SU(5)$ theory the following processes are possible

$$\begin{aligned} &u + u \to X \to e^+ + \bar{d}, \\ &u + d \to Y \to \bar{\nu}_e + \bar{d}, \\ &u + d \to Y \to \bar{u} + e^+. \end{aligned} \tag{8.5.1}$$

All these processes violate baryon number ($\Delta B = 1$) and lepton number ($\Delta L = 1$) but $B - L$ remains unchanged. Indeed, $B - L$ is conserved in the $SU(5)$ GUT (but not necessarily in other GUTs), and it is also conserved in the standard model.

If the theory were C or CP invariant, baryon number violating processes would generate anti-baryons at the same rate as baryons, and no net baryon asymmetry could be generated. The charge conjugate of a left-handed quark is a right-handed anti-quark, i.e. baryon number changes sign under charge conjugation. Similarly, under CP a left-handed quark is turned into a left-handed anti-quark, and again baryon number changes sign. Both the standard model and the $SU(5)$ GUT are chiral gauge theories, in which C is maximally violated. In the standard model CP is weakly broken ($K - \bar{K}$ system), which manifests itself in the complex phase of the Kobayashi-Maskawa matrix. This is possible only with at least three generations. The situation in the $SU(5)$ GUT is similar.

In thermal equilibrium baryon number violating processes proceed in both directions, i.e. baryons are generated but also annihilated. Hence, a net baryon asymmetry cannot be generated in thermal equilibrium. Formally, this results from CPT invariance (a symmetry of all consistent quantum field theories). Since the Universe expands and cools, baryon number violating processes get out of equilibrium, such that indeed all necessary conditions for the generation of a baryon asymmetry are satisfied in our Universe.

At very high temperatures ($\gg 10^{14}$ GeV) baryon number violating processes are in thermal equilibrium and $n_X = n_Y = n_\gamma$. When the temperature drops below m_X the X and Y-bosons get out of thermal equilibrium — they decay into quarks and leptons, but are not re-generated. Let us denote the rate for $X \to \bar{q}\bar{l}$ by r. Then the rate for $X \to qq$ is $1-r$. Correspondingly, the rate for $\bar{X} \to ql$ is $\bar{r}$, and the rate for $\bar{X} \to \bar{q}\bar{q}$ is $1-\bar{r}$. Under the decay of an X and an $\bar{X}$ one generates the baryon asymmetry

$$\Delta B = -\frac{1}{3}r + \frac{2}{3}(1-r) + \frac{1}{3}\bar{r} - \frac{2}{3}(1-\bar{r}) = \bar{r} - r = \varepsilon, \tag{8.5.2}$$

where ε is a measure of CP violation. The entropy density of the Universe is

$$s = \frac{2\pi^2}{45} g_* T^3, \tag{8.5.3}$$

where g_* is the number of relativistic degrees of freedom — for a typical GUT a number between 100 and 1000. The number density of X and $\bar{X}$ bosons is

$$n_X = g_X \frac{1}{\pi^2}\zeta(3) T^3, \tag{8.5.4}$$

where g_X counts the degrees of freedom of X and $\bar{X}$ bosons. After the decay of these particles a baryon asymmetry

$$n_B = n_X \varepsilon = g_X \frac{1}{\pi^2}\zeta(3)\frac{45}{2\pi^2}\frac{s}{g_*}\varepsilon \tag{8.5.5}$$

is generated. During the expansion of the Universe both the entropy and the baryon number are conserved. Today the entropy is almost entirely in the cosmic background radiation (as well as in the neutrinos, in case they are light enough) and we can estimate

$$s = \frac{2\pi^2}{45}\pi^2 \frac{1}{\zeta(3)} n_\gamma, \tag{8.5.6}$$

such that today the baryon-photon ratio is

$$\frac{n_B}{n_\gamma} = \frac{g_X}{g_*}\varepsilon. \tag{8.5.7}$$

The observed ratio is $n_B/n_\gamma = 10^{-10}$, such that we need $\varepsilon \approx 10^{-8}$. This is not an unnatural number for typical GUTs. On the other hand, one has some free parameters that allow to adjust the right amount of baryon asymmetry. The predictive power of GUTs for the baryon asymmetry is therefore limited. Also the above arguments are only qualitatively correct. A more detailed investigation would require the numerical solution of rate equations similar to nucleosynthesis calculations.

8.6 Baryon Number Violation in the Standard Model

The classical Lagrange density of the standard model does not contain baryon number violating interactions. However, this does not imply that the standard model conserves baryon number after quantization. Indeed, due to the chiral couplings of the fermions, the baryon number current has an anomaly in the standard model. Although the Lagrange density has a global $U(1)$ baryon number symmetry, this symmetry is explicitly broken in the quantum theory. The same is true for lepton number. The difference, $B - L$, on the other hand, remains conserved. The existence of baryon number violating processes at the electroweak scale may change the baryon asymmetry that has been generated at the GUT scale.

Let us consider the vacuum structure of a non-Abelian gauge theory (like the $SU(2)$ sector of the standard model). A classical vacuum solution is

$$\Phi(\vec{x}) = \left(\begin{array}{c} \Phi_+(\vec{x}) \\ \Phi_0(\vec{x}) \end{array} \right) = \left(\begin{array}{c} 0 \\ v \end{array} \right), \ A_i(\vec{x}) = 0. \tag{8.6.1}$$

Of course, gauge transformations of this solution are also vacua. However, states that are related by a gauge transformation are physically equivalent, and one should not consider the other solutions as additional vacua. Still, there is a subtlety, because there are gauge transformations with different topological properties. First of all, there are the so-called small gauge transformations, which can be continuously deformed into the identity, and one should indeed not distinguish between states related by small gauge transformations. However, there are also large gauge transformations — those that can not be deformed into a trivial gauge transformation — and they indeed give rise to additional vacuum states. The gauge transformations

$$g : \mathbb{R}^3 \rightarrow SU(2) \tag{8.6.2}$$

can be viewed as mappings from coordinate space into the group space. When one identifies points at spatial infinity $\mathbb{R}^3$ is compactified to S^3. On the other hand, the group space of $SU(2)$ is also S^3. Hence, the gauge transformations are mappings

$$g : S^3 \rightarrow S^3. \tag{8.6.3}$$

Such mappings are known to fall into topologically distinct classes characterized by a winding number

$$n[g] \in \Pi_3[SU(2)] = \mathbb{Z} \tag{8.6.4}$$

from the third homotopy group of the gauge group. In this case, mappings with any integer winding number are possible. Denoting a mapping with winding number n by g_n we can thus construct a set of topologically inequivalent vacuum states

$$\Phi^{(n)}(\vec{x}) = g_n(\vec{x}) \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad A_i^{(n)}(\vec{x}) = g_n(\vec{x}) \partial_i g_n(\vec{x})^\dagger. \tag{8.6.5}$$

Topologically distinct vacua are separated by energy barriers, and thus there is a periodic potential in the space of field configurations.

Classically, the system is in one of the degenerate vacuum states. Quantum mechanically, however, the system can tunnel from one vacuum to another. It turns out that a transition from the vacuum (m) to the vacuum (n) is accompanied by a baryon number violating process of strength $\Delta B = N_g(n - m)$, where N_g is the number of generations of quarks and leptons. Also the lepton number changes by $\Delta L = N_g(n-m)$, such that $B-L$ is conserved. The tunnel amplitude — and hence the rate of baryon number violating processes — is controlled by the barrier height between adjacent classical vacua. The unstable field configuration at the to of the barrier is known as a sphaleron (meaning ready to decay). In the standard model the height of the barrier (the sphaleron energy) is given by $4\pi v/g$ and the resulting tunneling rate is

$$\exp(-\frac{8\pi^2}{g^2}) \approx \exp(-200), \tag{8.6.6}$$

which is totally negligible. Hence, for some time people assumed that baryon number violation in the standard model is only of academic interest. However, it was overlooked that in the early Universe one need not tunnel through the barrier — one can simply step over it classically due to large thermal fluctuations. Then one must assume that in the TeV range baryon number violating processes are un-suppressed in the standard model. This means that any pre-existing baryon asymmetry — carefully created at the GUT scale — will be washed out, because baryon number violating processes are again in thermal equilibrium. Since the electroweak phase transition is of second or of weakly first order, it is unlikely (but not excluded) that a sufficient baryon asymmetry is re-generated at the electroweak scale.

However, we should not forget that $B-L$ is conserved in the standard model. This means that this mode is not thermalized. When baryon and lepton asymmetries ΔB_i and ΔL_i have been initially generated at the GUT scale, equilibrium sphaleron processes will imply that finally

$$\Delta(B_f + L_f) = 0, \tag{8.6.7}$$

but still

$$\Delta(B_f - L_f) = \Delta(B_i - L_i) = 0. \tag{8.6.8}$$

Hence, the present baryon and lepton asymmetries then are

$$\Delta B_f = -\Delta L_f = \frac{1}{2}\Delta(B_i - L_i). \tag{8.6.9}$$

This again leads to a problem, because also the minimal $SU(5)$ model conserves $B - L$. An asymmetry $\Delta(B_i - L_i)$ must hence be due to processes in the even earlier Universe. Then we would know as much as before. Fortunately, there is a way out. Other GUTs like $SO(10)$ and E_6 are not ruled out via proton decay and indeed do not conserve $B - L$. The reason for $B - L$ violation in these models is related to the existence of massive neutrinos. The so-called "see-saw" mechanism gives rise to one heavy neutrino of mass 10^{14} GeV and one light neutrino of mass in the eV range, that is identified with the neutrinos that we observe. Hence, we can explain the baryon asymmetry using GUTs only if the neutrinos are massive. Otherwise, we must assume that it was generated at times before 10^{-34} sec after the Big Bang, or we must find a way to go sufficiently out of thermal equilibrium around the electroweak phase transition and generate the baryon asymmetry via sphaleron processes.

Chapter 9

Technicolor

We have seen that grand unified theories have several attractive features. Especially the $SO(10)$ GUT appears as a very natural extension of the standard model to high energies. The GUT scale is already close to the Planck scale — the scale where quantum effects of gravity become important — which may well be the ultimate high-energy cut-off. GUTs may therefore play an important role in the final "theory of everything" that people seek in order to unify all interactions including gravity. Of course, we have many reasons to believe that the standard model itself could not possibly be the "theory of everything". First, it does not contain quantum gravity. Second, it has far too many unexplained parameters, and third, it is trivial and must be cut-off at some high-energy scale. If the Planck scale is a relevant scale for particle physics, we must understand why the electroweak scale is so much smaller. The electroweak scale is determined by the only dimensionful parameter in the standard model — the m^2 mass parameter in the Higgs potential. This parameter in the bare Lagrangian is renormalized and is indeed extremely sensitive to the ultra-violet cut-off. If that cut-off is the Planck scale, it requires an enormous amount of fine-tuning of the bare m^2 parameter to keep the electroweak scale at its low value. Such fine-tuning appears extremely unnatural. The pain of fine-tuning is especially well-known to people who have some experience with lattice field theory. Removing the cut-off, and hence taking the continuum limit of the lattice standard model, requires to fine-tune the lattice m^2 parameter to several digits accuracy in order to approach the second order phase transition that corresponds to the continuum limit. This difficulty is known as the hierarchy problem: it appears very difficult to naturally maintain a large separation between the electroweak and the Planck scale.

The same is not true for lattice QCD. Due to the gauge symmetry, light excitations are natural in QCD without fine-tuning. Thanks to asymptotic freedom, the nonperturbatively generated mass scale in QCD is exponentially insensitive to the cut-off. Hence keeping the QCD scale far below the Planck scale does not require any fine-tuning of the bare gauge coupling. All one needs to do is to pick a not too large bare coupling at the cut-off scale. This does not appear unnatural at all. Thanks to chiral symmetry, the masses of fermions can also be kept small without unnatural fine-tuning of bare parameters. It is interesting that chiral symmetry has been understood nonperturbatively (on the lattice) only a few years ago. Until then, and even in most numerical calculations today, one suffers from the pain of fine-tuning in order to keep the quark masses far below the cut-off. The natural nonperturbative solution of the lattice fermion problem involves D-branes and a fifth dimension. In fact, D-branes were introduced by David Kaplan in his lattice domain wall fermion construction long before they became popular in string theory. The lattice fermion problem illustrates nicely how a hierarchy problem has been solved thanks to the nonperturbative realization of a symmetry — chiral symmetry in that case. There are other examples of naturally light particles. A gauge symmetry, for example, protects the masses of gauge bosons from running to the cut-off scale. Even scalar particles can be light naturally without fine-tuning when they arise as Goldstone bosons of a spontaneously broken continuous global symmetry. In these cases, the symmetries that protect the particle masses from running to the cut-off are easy to maintain beyond perturbation theory, i.e. on the lattice. A longstanding puzzle that has now been solved, was to do the same for lattice fermions.

In order to solve the hierarchy problem in the standard model, one must be able to keep the electroweak energy scale associated with the scalar Higgs particle naturally small. The most popular way to achieve this is by using supersymmetry. Supersymmetry relates bosons to fermions. In this way a scalar mass that is related to a fermion mass can be indirectly protected from running to the cut-off by the fermion's chiral symmetry. At present, supersymmetry has the same status as chiral symmetry a few years ago. It is well understood perturbatively, but has not been implemented beyond perturbation theory, i.e. on the lattice. It may very well be that supersymmetry will eventually find a natural satisfactory implementation on the lattice, just as it happened recently for chiral symmetry. In that case, one could feel completely comfortable using supersymmetry to solve the hierarchy problem. However, this is less clear than in the case of chiral symmetry. In that case, Nature had shown us through experiments that it is possible to nonperturbatively regularize chiral fermions: they naturally appear as light particles. It took lattice physicists many years to understand how the hierarchy problem for lattice chiral fermions can be solved, but it was clear by Nature's example that it should somehow be possible. This is less clear for supersymmetry, because so far not a single supersymmetric partner particle has been found. An optimist and true believer in supersymmetry might, of course, say that at least one half of all particles has already been found. Many people hope that superpartners will eventually be found, and that supersymmetry will eventually be better understood beyond perturbation theory. This would, indeed, be very exciting because supersymmetry is certainly a truly remarkable theoretical idea. It deserves an entire course, and is beyond what will be discussed here.

What if supersymmetry is a perturbative illusion that cannot be formulated naturally beyond perturbation theory? Then even Nature will not be able to make use of it. In that case, one would need other ideas to understand why the electroweak scale is far below the Planck scale. One of these ideas goes under the name of technicolor. The central rather attractive idea of technicolor is to replace the Higgs sector of the standard model by some gauge dynamics. In that way one hopes to predict some of the free parameters of the standard model, and to explain why the electroweak scale (just as the QCD scale) is naturally small. As we will see, the chiral symmetry breaking of QCD gives rise to a small contribution to the W and Z boson masses proportional to F_π even without the Higgs mechanism. In technicolor models the Higgs sector is replaced with a QCD like technicolor gauge theory whose chiral symmetry breaking corresponds to electroweak symmetry breaking. Just like the QCD scale, thanks to asymptotic freedom the technicolor scale is protected from running to the Planck scale, and can naturally agree with the electroweak scale. The breaking of technicolor's global $SU(2)_L \otimes SU(2)_R$ chiral symmetry gives rise to three Goldstone bosons — the techni-pions. When we switch on the electroweak interactions, these particles are eaten and become the longitudinal components of the W and Z bosons. Of course, technicolor theories also predict a techni-rho as well as techni-baryons. The masses of these particles are in the TeV range.

While technicolor works nicely to give masses to the W and Z bosons, it works much less well when it comes to giving masses to the fermions. Of course, since technicolor models get rid of the elementary Higgs field, they can no longer generate fermion masses from Yukawa couplings. Since the physical Higgs particle is now a techni-quark-techni-anti-quark bound state, the Yukawa couplings are now replaced by four-fermion couplings between two techni-quarks and two ordinary standard model fermions. Unfortunately, it is very difficult to achieve realistic fermion masses and mixing parameters in this way. In particular, technicolor models have a severe problem with flavor-changing neutral currents which are absent in the standard model at tree level. Experimental bounds on flavor-changing neutral currents have ruled out most, if not all, technicolor models. Therefore, these models are no longer popular. Still, they are quite interesting theoretically, and one can sometimes learn a lot from a failed attempt. There may even by some truth in these models, and perhaps somebody will eventually have the missing idea that will make it all work naturally.

9.1 W and Z Boson Masses from QCD

Let us once again consider the low-energy chiral action for QCD with two massless quarks

$$S[U] = \int d^4x \, \frac{F_\pi^2}{4} \mathrm{Tr}[\partial_\mu U^\dagger \partial_\mu U], \tag{9.1.1}$$

where $U(x) \in SU(2)$ is the matrix field describing the Goldstone pions. Under global chiral rotations the field $U(x)$ transforms as

$$U(x)' = L^\dagger U(x) R. \tag{9.1.2}$$

In the next step, we are gauging $SU(2)_L$ by replacing ordinary with covariant derivatives

$$D_\mu U = (\partial_\mu + W_\mu) U, \tag{9.1.3}$$

where $W_\mu(x) = igW_\mu^a T^a$. Under $SU(2)_L$ gauge transformations the fields now transform as

$$U(x)' = L(x)^\dagger U(x) R, \ W_\mu(x)' = L(x)^\dagger (W_\mu(x) + \partial_\mu) L(x). \tag{9.1.4}$$

The total action now takes the form

$$S[U, W_\mu] = \int d^4x \, \{\frac{F_\pi^2}{4} \mathrm{Tr}[D_\mu U^\dagger D_\mu U] + \frac{1}{4g^2} \mathrm{Tr}[W_{\mu\nu} W_{\mu\nu}]\}. \tag{9.1.5}$$

In chapter 2 we have calculated the W and Z boson masses resulting from the Higgs mechanism. In that case, we have chosen the unitary gauge for the Higgs field. In the same way, we now choose the unitary gauge for the pion field, i.e. we arrange $L(x)$ such that $U(x) = \mathbb{1}$. Then the above action turns into

$$S[\mathbb{1}, W_\mu] = \int d^4x \, \{\frac{F_\pi^2}{4} \mathrm{Tr}[W_\mu W_\mu] + \frac{1}{4g^2} \mathrm{Tr}[W_{\mu\nu} W_{\mu\nu}]\}. \tag{9.1.6}$$

Hence, we can read off the W mass

$$M_W = gF_\pi, \tag{9.1.7}$$

which results just from chiral symmetry breaking in QCD. Of course, numerically $F_\pi \approx 0.09$ GeV is negligible compared to the electroweak scale $v = 245$ GeV which provides by far the dominant contribution to the physical W mass. Still, it is very interesting that QCD alone breaks the electroweak symmetry dynamically. In technicolor models QCD is replaced by another confining gauge theory that operates at much higher energy scales and has a value of F_π that corresponds to the electroweak scale v of the standard model.

Let us also gauge $U(1)_Y$ as a subgroup of $SU(2)_R$ in order to see what happens to the Z boson. Then the covariant derivative takes the form

$$D_\mu U = \partial_\mu U + W_\mu U - U B_\mu, \tag{9.1.8}$$

where $B_\mu = ig' B_\mu^3 T^3$, and the action now reads

$$\begin{aligned} S[U, W_\mu, B_\mu] &= \int d^4x \, \{ \frac{F_\pi^2}{4} \mathrm{Tr}[D_\mu U^\dagger D_\mu U] \\ &+ \frac{1}{4g^2} \mathrm{Tr}[W_{\mu\nu} W_{\mu\nu}] + \frac{1}{4g'^2} \mathrm{Tr}[B_{\mu\nu} B_{\mu\nu}] \}. \end{aligned} \tag{9.1.9}$$

In unitary gauge $U(x) = \mathbb{1}$ this expression takes the form

$$\begin{aligned} S[\mathbb{1}, W_\mu, B_\mu] &= \int d^4x \, \{ \frac{F_\pi^2}{4} \mathrm{Tr}[(W_\mu - B_\mu)(W_\mu - B_\mu)] \\ &+ \frac{1}{4g^2} \mathrm{Tr}[W_{\mu\nu} W_{\mu\nu}] + \frac{1}{4g'^2} \mathrm{Tr}[B_{\mu\nu} B_{\mu\nu}] \}. \end{aligned} \tag{9.1.10}$$

When one writes

$$W_\mu^3 = \frac{g' A_\mu + g Z_\mu}{\sqrt{g^2 + g'^2}}, \ B_\mu^3 = \frac{g A_\mu - g' Z_\mu}{\sqrt{g^2 + g'^2}}, \tag{9.1.11}$$

one sees that Z picks up a mass

$$M_Z = \sqrt{g^2 + g'^2} F_\pi, \tag{9.1.12}$$

while the photon A_μ remains massless. In particular, like in the standard model

$$\frac{M_W}{M_Z} = \frac{g}{\sqrt{g^2 + g'^2}} = \cos\theta_W. \tag{9.1.13}$$

This implies that a technicolor model is indeed capable of replacing the Higgs sector of the standard model as long as the value of F_π for the technicolor interaction is at the electroweak scale v. When we gauged $U(1)_Y$ in the chiral Lagrangian for QCD we also had to introduce the Goldstone-Wilczek term in order to account for the decay $\pi^0 \rightarrow \gamma\gamma$. When the neutral techni-pion is eaten, it becomes the longitudinal component of the Z boson, which might hence also have this anomalous decay. Of course, the width of the physical Z boson is well described by the standard model, in which this process is absent. Hence, in the construction of technicolor models we should make sure that the techni-pion cannot decay into two photons. As we have seen earlier, this implies that the electric charges of the techni-up and techni-down quarks should be equal and opposite.

9.2 An Anomaly Free Technicolor Model

Let us construct a concrete technicolor model, in particular, to show explicitly that such constructions are at all possible. In addition to the standard model fermions, we want to add a techni-up and a techni-down quark U and D whose left-handed components form an $SU(2)_L$ doublet and whose right-handed components are $SU(2)_L$ singlets. The technicolor gauge group is chosen to be $SU(N_{tc})$ and both the left and the right-handed techni-quarks transform in the fundamental representation of $SU(N_{tc})$. The number of technicolors N_{tc} will be restricted by anomaly cancellation conditions. All the standard model fermions are assumed to be technicolor singlets. We will not introduce any techni-leptons. Hence, anomaly cancellation works differently from the one in the standard model. For simplicity we choose the techni-quarks to be $SU(3)_c$ color singlets — they are still confined by technicolor interactions. Let us denote the $U(1)_Y$ quantum numbers of the techni-quarks by Y_L, Y_{U_R} and Y_{D_R}. These numbers will be determined by anomaly cancellation conditions.

The gauge group of our technicolor model is given by $SU(N_{tc}) \otimes SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$. Let us now check anomaly cancellation. Since, like QCD, technicolor is a vector-like theory, the triangle diagram with three external techni-gauge bosons automatically vanishes. Triangle diagrams with external techni-gauge bosons and external gluons also vanish for the same reason. Triangle diagrams with only one external techni-gauge boson vanish because the generators of $SU(N_{tc})$ are traceless. The triangle diagram with two techni-gauge bosons and one $SU(2)_L$ boson vanishes because the generators of $SU(2)_L$ are traceless. Similarly, the triangle diagram with two external techni-gauge bosons and one external $U(1)_Y$ boson vanishes only if

$$2Y_L = Y_{U_R} + Y_{D_R}. \tag{9.2.1}$$

The techni-quarks also contribute to the anomaly cancellation conditions of the standard model itself. For example, the triangle diagram with two $SU(2)_L$ bosons and one $U(1)_Y$ boson now vanishes only if

$$Y_L = 0, \tag{9.2.2}$$

while the diagram with three external $U(1)_Y$ bosons vanishes only if

$$2Y_L^3 = Y_{U_R}^3 + Y_{D_R}^3. \tag{9.2.3}$$

Anomaly cancellation hence implies

$$Y_L = 0,\ Y_{U_R} = -Y_{D_R}. \tag{9.2.4}$$

We also want to be able to show our new theory to gravity, which is possible only if we cancel the gravitational anomaly. This again requires

$$2Y_L = Y_{U_R} + Y_{D_R}, \tag{9.2.5}$$

and is hence automatically satisfied.

In order to reproduce the physics of the standard model, we want to maintain $U(1)_{em}$ as a good symmetry. This means that the breaking of techni-chiral symmetry should leave $U(1)_{em}$ intact. This requires the electric charges of the left- and right-handed techni-quarks to be equal, because otherwise the techni-chiral condensate would carry an electric charge and would make the vacuum superconducting. Since we have

$$Q = T_L^3 + Y, \tag{9.2.6}$$

we obtain

$$Q_{U_L} = \frac{1}{2},\ Q_{D_L} = -\frac{1}{2},\ Q_{U_R} = Y_{U_R},\ Q_{D_R} = Y_{D_R}. \tag{9.2.7}$$

Hence, in order to have equal charges for left- and right-handed techni-quarks we demand

$$Y_{U_R} = -Y_{D_R} = \frac{1}{2}. \tag{9.2.8}$$

Then $Q_U = -Q_D$ and the techni-pion will indeed not decay into two photons.

In order to also cancel Witten's global anomaly, the total number of $SU(2)_L$ doublets must be even and hence N_{tc} must also be even. The naive simplest choice $N_{tc} = 2$ is problematic. Due to the real nature of $SU(2)$, techni-quarks and techni-anti-quarks would then be indistinguishable and the chiral symmetry would be $SU(4)$ instead of $SU(2)_L \otimes SU(2)_R$. The actual simplest choice therefore is $N_{tc} = 4$. The gauge symmetry of our extended standard model then is $SU(4)_{tc} \otimes SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$.

Chapter 10

Particle Physics of the Early Universe

General relativity predicts that our expanding Universe emerged from a space-time singularity — the big bang — and that it will either continue to expand forever, or ultimately collapse in a big crunch. It is an observational fact that the Universe is indeed expanding. Using the observed expansion rate one estimates that the big bang occurred about 10^{10} years ago. Immediately after the bang the energy density of the Universe was so high that particle physics effects at arbitrarily high energy occurred naturally. As time passes, the Universe expands and cools down thus going through all energy scales down to the few degrees Kelvin that we observe nowadays in the cosmic background radiation, which was emitted about 10^5 years after the big bang. Long before that at about 10^{-5} seconds after the big bang the temperature was at the typical QCD scale $\Lambda_{QCD} \approx 0.2$GeV. At around that time the Universe went through a phase transition (or crossover) after which confinement set in and chiral symmetry broke spontaneously. In the early high temperature phase quarks and gluons were deconfined, and chiral symmetry was still intact. Similarly, at temperatures around 250 GeV, i.e. at times around 10^{-12} sec after the bang, a phase transition must have happened in the electroweak sector of the standard model. To understand these epochs better let us first discuss the standard big bang cosmology.

10.1 The Standard Big Bang Cosmology

A basic assumption of the standard big bang cosmology comes from the observation that on the largest scales our Universe is spatially rather homogeneous and isotropic. Also the Universe is expanding, which is described by a time dependent scale parameter $R(t)$. This leads to the Friedmann-Lemaitre-Robertson-Walker ansatz for the space-time metric

$$ds^2 = dt^2 - R(t)^2(\frac{d\rho^2}{1-k\rho^2} + \rho^2(d\theta^2 + \sin^2\theta d\varphi^2)) = g_{\mu\nu}dx^\mu dx^\nu. \tag{10.1.1}$$

Note that we are in a space-time with negative signature in the metric. In the above expression $k = 0$ for a flat space, and $k = \pm 1$ for a space with constant positive or negative curvature. This metric selects a preferred cosmic rest frame, in which the Universe (e.g. the cosmic background radiation) appears homogeneous and isotropic. From the Christoffel symbols

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \tag{10.1.2}$$

one derives the Riemann tensor

$$R^\sigma_{\mu\nu\rho} = \partial_\nu\Gamma^\sigma_{\mu\rho} - \partial_\rho\Gamma^\sigma_{\mu\nu} + \Gamma^\sigma_{\lambda\nu}\Gamma^\lambda_{\mu\rho} - \Gamma^\sigma_{\lambda\rho}\Gamma^\lambda_{\mu\nu}, \tag{10.1.3}$$

which then leads to the Ricci tensor

$$R_{\mu\sigma} = R^\nu_{\mu\nu\sigma}, \tag{10.1.4}$$

the curvature scalar

$$\mathcal{R} = R^\mu_\mu = g^{\mu\nu}R_{\nu\mu}, \tag{10.1.5}$$

and finally to the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R}. \tag{10.1.6}$$

For the above metric one finds

$$\begin{aligned}
\mathcal{R} &= -6(\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2}), \\
G_{ij} &= (2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2})g_{ij}, \\
G_{00} &= 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}), \\
G_{0i} &= G_{i0} = 0.
\end{aligned} \tag{10.1.7}$$

Here i, j denote spatial indices. Einstein's field equation (without a cosmological constant) takes the form

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{10.1.8}$$

where G is Newton's constant, and $T_{\mu\nu}$ is the energy-momentum tensor of matter. In the early Universe it takes the ideal gas form

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu - p g_{\mu\nu}, \tag{10.1.9}$$

where ρ is the energy density and p is the pressure of matter, and u_μ is its four-velocity. To be consistent with the assumption of spatial homogeneity and isotropy we must assume that the matter is at rest in the cosmic rest frame, i.e. $u_0 = 1$, $u_i = 0$, such that

$$T_{ii} = -p g_{ij}, \ T_{00} = \rho. \tag{10.1.10}$$

Using Einstein's field equations we now obtain

$$2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = -8\pi G p, \ 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) = 8\pi G \rho. \tag{10.1.11}$$

From this we infer

$$\begin{aligned} &\frac{d}{dt}(8\pi G \rho R^3) + 8\pi G p \frac{d}{dt} R^3 = \\ &\frac{d}{dt}(3R\dot{R}^2 + 3kR) - (2\frac{\ddot{R}}{R}\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) 3R^2 \dot{R} = 0. \end{aligned} \tag{10.1.12}$$

We identify this as the first law of thermodynamics

$$\frac{d}{dt}(\rho R^3) = -p \frac{d}{dt} R^3, \tag{10.1.13}$$

or equivalently

$$dE = -p dV, \tag{10.1.14}$$

where $E = \rho R^3$ is the energy and R^3 is the volume. For a relativistic gas of particles (as it existed in the very early Universe) the equation of state is

$$p = \frac{1}{3}\rho, \tag{10.1.15}$$

such that

$$\frac{\ddot{R}}{R} = -\frac{8\pi G}{3}\rho < 0. \tag{10.1.16}$$

This shows that the expansion of the Universe is decelerating. Going backwards in time this implies the existence of a big bang. Using

$$\begin{aligned} &\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}R^3 = -\frac{1}{3}\rho 3R^2\dot{R} = -\rho R^2 \dot{R} \Rightarrow \\ &\dot{\rho} R^3 + 3\rho R^2 \dot{R} + \rho R^2 \dot{R} = \dot{\rho} R^3 + 4\rho R^2 \dot{R} = 0, \end{aligned} \tag{10.1.17}$$

one finds

$$\rho R^4 = N, \tag{10.1.18}$$

where N is a constant. We then obtain

$$\frac{\ddot{R}}{R} = -\frac{8\pi GN}{3R^4} \Rightarrow \frac{1}{2}\dot{R}^2 = \frac{4\pi GN}{3R^2} + C. \tag{10.1.19}$$

The constant C follows from eq.(10.1.16) to be

$$C = -\frac{k}{2}, \tag{10.1.20}$$

such that

$$\dot{R} = \sqrt{\frac{8\pi GN}{3R^2} - k}. \tag{10.1.21}$$

The assumption of a relativistic gas is valid only for the very early Universe, i.e. as long as R is small. Then we can neglect the curvature k and we obtain

$$t = \int_0^R dRR\sqrt{\frac{3}{8\pi GN}} = \frac{1}{2}R^2\sqrt{\frac{3}{8\pi GN}}, \tag{10.1.22}$$

such that

$$\frac{32\pi GN}{3R^4}t^2 = \frac{32\pi G\rho}{3}t^2 = 1, \tag{10.1.23}$$

i.e. the energy density decreases as t^{-2} and the size of the Universe increases as $t^{1/2}$. For a relativistic gas of bosons at temperature T one has

$$\rho = \frac{\pi^2}{15}T^4, \tag{10.1.24}$$

and for fermions one has

$$\rho = \frac{7\pi^2}{120}T^4. \tag{10.1.25}$$

In both cases ρ is proportional to T^4 such that the temperature decreases proportional to $t^{-1/2}$ in the very early Universe. It is convenient to express things in units of the Planck mass

$$m_P = G^{-1/2} = 1.2211 \times 10^{19}\text{GeV}, \tag{10.1.26}$$

or the Planck time

$$t_p = G^{1/2} = 5.3904 \times 10^{-44}\text{sec}. \tag{10.1.27}$$

This allows us to write

$$\frac{32\pi\rho/m_P^4}{3}(t/t_P)^2 = 1. \tag{10.1.28}$$

Typical QCD energy densities are of the order of GeV^4 such that $\rho/m_P^4 \approx 10^{-76}$. This then corresponds to times $t/t_P \approx 10^{38}$ and hence to $t \approx 10^{-5}$sec. When we come to typical nuclear physics energy densities MeV^4 this corresponds to $t \approx 10$sec. Indeed during the first few minutes after the big bang the first light nuclei have been formed in the so-called primordial nucleosynthesis. Typical energy densities of atomic physics are in the eV^4 range such that the corresponding time is now $t \approx 10^{13}$sec which is roughly 100000 years. At that time electrons and nuclei combined to neutral atoms, the Universe become transparent, and the majority of photons decoupled from the matter forming the cosmic background radiation. From that moment on the Universe was no longer radiation dominated, so the above description using a relativistic ideal gas no longer applies. Instead nonrelativistic matter (like stars) takes over and the corresponding calculation is slightly modified.

10.2 The QCD Phase Transition

Quantum systems (in particular quantum field theories) at finite temperature are characterized by a partition function

$$Z = \mathrm{Tr}\exp(-\beta H), \tag{10.2.1}$$

that can be represented by a Euclidean path integral with a Euclidean time extent given by the inverse temperature $\beta = 1/T$. Due to the trace one integrates over periodic field configurations. At high temperatures the system is very short in the Euclidean time direction, and it effectively reduces to a 3-d system in the infinite temperature limit. Even at finite temperature, if the system undergoes a second order phase transition, the finite size in Euclidean time is negligible compared to the diverging correlation length, such that again dimensional reduction occurs. In QCD with N_f massless quarks the $SU(N_f)_L \otimes SU(N_f)_R$ chiral symmetry is spontaneously broken at low temperatures. At high temperatures, on the other hand, one expects that due to asymptotic freedom quarks and gluons behave like free particles, and hence the theory deconfines and chiral symmetry should be restored. Lattice simulations indeed support this expectation. In the early Universe the deconfinement-confinement phase transition must have happened at a temperature of about 0.15 GeV. For two massless quarks the chiral symmetry group is $SU(2)_L \otimes SU(2)_R$, which is isomorphic to $O(4)$. At low temperatures the symmetry breaks spontaneously to $SU(2)_{L=R}$ which is isomorphic to $O(3)$. From condensed matter physics investigations of ferro- and antiferromagnets it is well known that 3-d systems with symmetry breaking $O(4) \rightarrow O(3)$ have second order phase transitions. Hence, based on universality and the fact that at high temperature QCD is dimensionally reduced to a 3-d theory, one expects that with two massless quarks the QCD phase transition would be second order. In fact, QCD would behave very similar to a magnet, e.g. all critical exponents defined at the phase transition would be the same. For three massless quarks, on the other hand, one expects a first order phase transition, because systems with symmetry breaking $SU(3)_L \otimes SU(3)_R \rightarrow SU(3)_{L=R}$ don't have second order transitions. In that case there is no universal behavior. In reality the quarks are not completely massless. Especially the s-quark has an intermediate mass, and it is unclear if the two or three flavor argument applies to our Universe. In any case, a true second order phase transition would only happen in the limit of vanishing quark mass. Therefore, in our Universe we either had a first order phase transition or merely a cross over.

A first order phase transition proceeds via bubble nucleation, just as in boiling water. At the phase transition both phases coexist with each other, and they are separated by an interface with a finite interface tension. In case of a first order QCD phase transition bubbles of confined phase would appear from the hot quark-gluon plasma. The bubbles would expand and would ultimately fill the whole Universe. Witten has suggested that a strange state of matter — so-called quark nuggets — may be created during the QCD phase transition provided it is strongly first order. He argued that the expanding bubble walls of confined phase may compress the remaining quark-gluon plasma to superlarge hadrons, which would be stable due to a large s-quark content that would be favored by the Pauli principle. These quark nuggets would have masses of about a ton. They would not participate in primordial nucleosynthesis and would actually be exotic candidates for baryonic dark matter. Still, if such objects existed in the Universe, one would expect them to collide with the earth sometimes causing major damage. The fact, that we don't experience catastrophes like that too often puts an upper limit on the number of quark nuggets. In fact, they seem to be a very exotic form of matter, for which we have no observational evidence.

10.3 The Electroweak Phase Transition

At temperatures in the 300 GeV range, i.e. at times around 10^{-12} sec after the bang, a phase transition must have happened in the electroweak sector of the standard model. At very early times the $SU(2)_L \otimes U(1)_Y$ gauge symmetry was not yet spontaneously broken to $U(1)_{em}$. Correspondingly, at that time all fermions and gauge bosons have been massless. Only the scalar field Φ had a mass. The order of the phase transition (and hence the strength of cosmological effects) depends on the mass of the Higgs particle via the temperature dependence of the effective potential of the scalar field. For a heavy Higgs particle (mass > 100 GeV) as it seems to exist in our Universe, the phase transition is second order, and cosmological effects are very weak, at least as long as the gauge couplings g, g' and g_s can be neglected. When one includes the effect of the gauge couplings, the phase transition becomes weakly first order.

In case of a light Higgs particle, which is already ruled out experimentally, the phase transition would be strongly first order, and would lead to a strong super-cooling of the symmetric phase. This could have delayed the electroweak phase transition until the QCD transition. In that case the generated entropy would be unacceptably large, such that cosmological effects alone lead to a lower bound on the Higgs mass of about 10 GeV.

Appendix A

Quantum Field Theory

This chapter provides a brief summary of the mathematical structure of quantum field theory. Classical field theories are discussed as a generalization of point mechanics to systems with infinitely many degrees of freedom — a given number per space point. Similarly, quantum field theories are just quantum mechanical systems with infinitely many degrees of freedom. In the same way as point mechanics systems, classical field theories can be quantized with path integral methods. The quantization of field theories at finite temperature leads to path integrals in Euclidean time. This provides us with an analogy between quantum field theory and classical statistical mechanics. We also mention the lattice regularization which has recently provided a mathematically satisfactory formulation of the standard model beyond perturbation theory.

A.1 From Point Mechanics to Classical Field Theory

Point mechanics describes the dynamics of classical nonrelativistic point particles. The coordinates of the particles represent a finite number of degrees of freedom. In the simplest case — a single particle moving in one spatial dimension — we are dealing with a single degree of freedom: the x-coordinate of the particle. The dynamics of a particle of mass m moving in an external potential $V(x)$ is described by Newton's equation

$$m\partial_t^2 x = ma = F(x) = -\frac{dV(x)}{dx}. \tag{A.1.1}$$

Once the initial conditions are specified, this ordinary second order differential equation determines the particle's path $x(t)$, i.e. its position as a function of time. Newton's equation results from the variational principle to minimize the action

$$S[x] = \int dt\ L(x, \partial_t x), \tag{A.1.2}$$

over the space of all paths $x(t)$. The action is a functional (a function whose argument is itself a function) that results from the time integral of the Lagrange function

$$L(x, \partial_t x) = \frac{m}{2}(\partial_t x)^2 - V(x). \tag{A.1.3}$$

The Euler-Lagrange equation

$$\partial_t \frac{\delta L}{\delta(\partial_t x)} - \frac{\delta L}{\delta x} = 0, \tag{A.1.4}$$

is nothing but Newton's equation.

Classical field theories are a generalization of point mechanics to systems with infinitely many degrees of freedom — a given number for each space point $\vec{x}$. In this case, the degrees of freedom are the field values $\Phi(\vec{x})$, where Φ is some generic field. In case of a neutral scalar field, Φ is simply a real number representing one degree of freedom per space point. A charged scalar field, on the other hand, is described by a complex number and hence represents two degrees of freedom per space point. The scalar Higgs field $\Phi^a(\vec{x})$ (with $a \in \{1, 2\}$) in the standard model is a complex doublet, i.e. it has four real degrees of freedom per space point. An Abelian gauge field $A_i(\vec{x})$ (with a spatial direction index $i \in \{1, 2, 3\}$) — for example, the photon field in electrodynamics — is a neutral vector field with 3 real degrees of freedom per space point. One of these degrees of freedom is redundant due to the $U(1)$ gauge symmetry. Hence, an Abelian gauge field has two physical degrees of freedom per space point which correspond to the two polarization states of the massless photon. Note that the time-component $A_0(\vec{x})$ does not represent a physical degree of freedom. It is just a Lagrange multiplier field that enforces the Gauss law. A non-Abelian gauge field $A_i^a(\vec{x})$ is charged and has an additional index a. For example, the gluon field in chromodynamics with a color index $a \in \{1, 2, ..., 8\}$ represents $2 \times 8 = 16$ physical degrees of freedom per space point, again because of some redundancy due to the $SU(3)_c$ color gauge symmetry. The field that represents the W- and Z-bosons in the standard model has an index $a \in \{1, 2, 3\}$ and transforms under the gauge group $SU(2)_L$. Thus, it represents $2 \times 3 = 6$ physical degrees of freedom. However, in contrast to the photon, the W- and Z-bosons are massive due to the Higgs mechanism and have three (not just two) polarization states. The extra degree of freedom is provided by the Higgs field.

The analog of Newton's equation in field theory is the classical field equation of motion. For example, for a neutral scalar field this is the Klein-Gordon equation

$$\partial^\mu \partial_\mu \Phi = -\frac{dV(\Phi)}{d\Phi}. \tag{A.1.5}$$

Again, after specifying appropriate initial conditions it determines the classical field configuration $\Phi(x)$, i.e. the values of the field Φ at all space-time points $x = (t, \vec{x})$. Hence, the role of time in point mechanics is played by space-time in field theory, and the role of the point particle coordinates is now played by the field values. As before, the classical equation of motion results from minimizing the action

$$S[\Phi] = \int d^4x \ \mathcal{L}(\Phi, \partial_\mu \Phi). \tag{A.1.6}$$

The integral over time in eq.(A.1.2) is now replaced by an integral over space-time and the Lagrange function of point mechanics gets replaced by the Lagrange density function (or Lagrangian)

$$\mathcal{L}(\Phi, \partial_\mu \Phi) = \frac{1}{2} \partial^\mu \Phi \partial_\mu \Phi - V(\Phi). \tag{A.1.7}$$

A simple interacting field theory is the Φ^4 theory with the potential

$$V(\Phi) = \frac{m^2}{2} \Phi^2 + \frac{\lambda}{4} \Phi^4. \tag{A.1.8}$$

Here m is the mass of the scalar field and λ is the coupling strength of its self-interaction. Note that the mass term corresponds to a harmonic oscillator potential in the point mechanics analog, while the interaction term corresponds to an anharmonic perturbation. As before, the Euler-Lagrange equation

$$\partial_\mu \frac{\delta L}{\delta(\partial_\mu \Phi)} - \frac{\delta L}{\delta \Phi} = 0, \tag{A.1.9}$$

is the classical equation of motion, in this case the Klein-Gordon equation. The analogies between point mechanics and field theory are summarized in table A.1.

A.2 Path Integral in Real Time

The quantization of field theories is most conveniently performed using the path integral approach. Here we first discuss the path integral in quantum mechanics

Point Mechanics	Field Theory
time t	space-time $x = (t, \vec{x})$
particle coordinate x	field value Φ
particle path $x(t)$	field configuration $\Phi(x)$
action $S[x] = \int dt\ L(x, \partial_t x)$	action $S[\Phi] = \int d^4x\ \mathcal{L}(\Phi, \partial_\mu \Phi)$
Lagrange function $L(x, \partial_t x) = \frac{m}{2}(\partial_t x)^2 - V(x)$	Lagrangian $\mathcal{L}(\Phi, \partial_\mu \Phi) = \frac{1}{2}\partial^\mu \Phi \partial_\mu \Phi - V(\Phi)$
equation of motion $\partial_t \frac{\delta L}{\delta(\partial_t x)} - \frac{\delta L}{\delta x} = 0$	field equation $\partial_\mu \frac{\delta L}{\delta(\partial_\mu \Phi)} - \frac{\delta L}{\delta \Phi} = 0$
Newton's equation $\partial_t^2 x = -\frac{dV(x)}{dx}$	Klein-Gordon equation $\partial^\mu \partial_\mu \Phi = -\frac{dV(\Phi)}{d\Phi}$
kinetic energy $\frac{m}{2}(\partial_t x)^2$	kinetic energy $\frac{1}{2}\partial^\mu \Phi \partial_\mu \Phi$
harmonic oscillator potential $\frac{m}{2}\omega^2 x^2$	mass term $\frac{m^2}{2}\Phi^2$
anharmonic perturbation $\frac{\lambda}{4}x^4$	self-interaction term $\frac{\lambda}{4}\Phi^4$

Table A.1: *The dictionary that translates point mechanics into the language of field theory.*

— quantized point mechanics — using the real time formalism. A mathematically more satisfactory formulation uses an analytic continuation to so-called Euclidean time. This will be discussed in the next section.

The real time evolution of a quantum system described by a Hamilton operator H is given by the time-dependent Schrödinger equation

$$i\hbar \partial_t |\Psi(t)\rangle = H|\Psi(t)\rangle. \tag{A.2.1}$$

For a time-independent Hamilton operator the time evolution operator is given by

$$U(t', t) = \exp(-\frac{i}{\hbar} H(t' - t)), \tag{A.2.2}$$

such that

$$|\Psi(t')\rangle = U(t', t)|\Psi(t)\rangle. \tag{A.2.3}$$

Let us consider the transition amplitude $\langle x'|U(t', t)|x\rangle$ of a nonrelativistic point particle that starts at position x at time t and arrives at position x' at time t'. Using

$$\langle x|\Psi(t)\rangle = \Psi(x, t) \tag{A.2.4}$$

we obtain

$$\Psi(x', t') = \int dx\ \langle x'|U(t', t)|x\rangle \Psi(x, t), \tag{A.2.5}$$

i.e. $\langle x'|U(t',t)|x\rangle$ acts as a propagator for the wave function. The propagator is of physical interest because it contains information about the energy spectrum. When we consider propagation from an initial position x back to the same position we find

$$\begin{aligned}\langle x|U(t',t)|x\rangle &= \langle x|\exp(-\frac{i}{\hbar}H(t'-t))|x\rangle \\ &= \sum_n |\langle x|n\rangle|^2 \exp(-\frac{i}{\hbar}E_n(t'-t)).\end{aligned} \tag{A.2.6}$$

We have inserted a complete set, $\sum_n |n\rangle\langle n| = \mathbb{1}$, of energy eigenstates $|n\rangle$ with

$$H|n\rangle = E_n|n\rangle. \tag{A.2.7}$$

Hence, according to eq.(A.2.6), the Fourier transform of the propagator yields the energy spectrum as well as the energy eigenstates $\langle x|n\rangle$.

Inserting a complete set of position eigenstates we arrive at

$$\begin{aligned}\langle x'|U(t',t)|x\rangle &= \langle x'|\exp(-\frac{i}{\hbar}H(t'-t_1+t_1-t))|x\rangle \\ &= \int dx_1 \langle x'|\exp(-\frac{i}{\hbar}H(t'-t_1))|x_1\rangle \\ &\times \langle x_1|\exp(-\frac{i}{\hbar}H(t_1-t))|x\rangle \\ &= \int dx_1 \langle x'|U(t',t_1)|x_1\rangle\langle x_1|U(t_1,t)|x\rangle.\end{aligned} \tag{A.2.8}$$

It is obvious that we can repeat this process an arbitrary number of times. This is exactly what we do in the formulation of the path integral. Let us divide the time interval $[t,t']$ into N elementary time steps of size ε such that

$$t'-t = N\varepsilon. \tag{A.2.9}$$

Inserting a complete set of position eigenstates at the intermediate times $t_i, i \in \{1,2,...,N-1\}$ we obtain

$$\begin{aligned}\langle x'|U(t',t)|x\rangle &= \int dx_1 \int dx_2 ... \int dx_{N-1} \langle x'|U(t',t_{N-1})|x_{N-1}\rangle ... \\ &\times \langle x_2|U(t_2,t_1)|x_1\rangle\langle x_1|U(t_1,t)|x\rangle.\end{aligned} \tag{A.2.10}$$

In the next step we concentrate on one of the factors and we consider a single nonrelativistic point particle moving in an external potential $V(x)$ such that

$$H = \frac{p^2}{2m} + V(x). \tag{A.2.11}$$

Using the Baker-Campbell-Haussdorff formula and neglecting terms of order ε^2 we find

$$\begin{aligned}\langle x_{i+1}|U(t_{i+1},t_i)|x_i\rangle &= \langle x_{i+1}|\exp(-\frac{i\varepsilon p^2}{2m\hbar})\exp(-\frac{i\varepsilon}{\hbar}V(x))|x_i\rangle \\ &= \frac{1}{2\pi}\int dp\langle x_{i+1}|\exp(-\frac{i\varepsilon p^2}{2m\hbar})|p\rangle\langle p|\exp(-\frac{i\varepsilon}{\hbar}V(x))|x_i\rangle \\ &= \frac{1}{2\pi}\int dp\exp(-\frac{i\varepsilon p^2}{2m\hbar})\exp(-\frac{i}{\hbar}p(x_{i+1}-x_i)) \\ &\times \exp(-\frac{i\varepsilon}{\hbar}V(x_i)). \end{aligned} \qquad (A.2.12)$$

The integral over p is ill-defined because the integrand is a very rapidly oscillating function. To make the expression well-defined we replace the time step ε by $\varepsilon - ia$, i.e. we go out into the complex time plane. After doing the integral we take the limit $a \to 0$. Still, one should keep in mind that the definition of the path integral required an analytic continuation in time. One finds

$$\langle x_{i+1}|U(t_{i+1},t_i)|x_i\rangle = \sqrt{\frac{m}{2\pi i\hbar\varepsilon}}\exp(\frac{i}{\hbar}\varepsilon[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)]). \qquad (A.2.13)$$

Inserting this back into the expression for the propagator we obtain

$$\langle x'|U(t',t)|x\rangle = \int \mathcal{D}x\exp(\frac{i}{\hbar}S[x]). \qquad (A.2.14)$$

The action has been identified in the time continuum limit as

$$\begin{aligned} S[x] &= \int dt\ [\frac{m}{2}(\partial_t x)^2 - V(x)] \\ &= \lim_{\varepsilon\to 0}\sum_i \varepsilon[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)]. \end{aligned} \qquad (A.2.15)$$

The integration measure is defined as

$$\int \mathcal{D}x = \lim_{\varepsilon\to 0}\sqrt{\frac{m}{2\pi i\hbar\varepsilon}}^{N-1}\int dx_1\int dx_2 ... \int dx_{N-1}. \qquad (A.2.16)$$

This means that we integrate over the possible particle positions for each intermediate time t_i. In this way we integrate over all possible paths of the particle starting at x and ending at x'. Each path is weighted with an oscillating phase factor $\exp(\frac{i}{\hbar}S[x])$ determined by the action. The classical path of minimum action has the smallest oscillations, and hence the largest contribution to the path integral. In the classical limit $\hbar \to 0$ only that contribution survives.

A.3 Path Integral in Euclidean Time

As we have seen, it requires a small excursion into the complex time plane to make the path integral mathematically well-defined. Now we will make a big step into that plane and actually consider purely imaginary so-called Euclidean time. The physical motivation for this, however, comes from quantum statistical mechanics. Let us consider the quantum statistical partition function

$$Z = \mathrm{Tr}\exp(-\beta H), \tag{A.3.1}$$

where $\beta = 1/T$ is the inverse temperature. It is mathematically equivalent to the time interval we discussed in the real time path integral. In particular, the operator $\exp(-\beta H)$ turns into the time evolution operator $U(t', t)$ if we identify

$$\beta = \frac{i}{\hbar}(t' - t). \tag{A.3.2}$$

In this sense the system at finite temperature corresponds to a system propagating in purely imaginary (Euclidean) time. By dividing the Euclidean time interval into N time steps, i.e. by writing $\beta = Na/\hbar$, and again by inserting complete sets of position eigenstates we now arrive at the Euclidean time path integral

$$Z = \int \mathcal{D}x \exp(-\frac{1}{\hbar}S_E[x]). \tag{A.3.3}$$

The action now takes the Euclidean form

$$\begin{aligned} S_E[x] &= \int dt\ [\frac{m}{2}(\partial_t x)^2 + V(x)] \\ &= \lim_{a\rightarrow 0}\sum_i a[\frac{m}{2}(\frac{x_{i+1} - x_i}{a})^2 + V(x_i)]. \end{aligned} \tag{A.3.4}$$

In contrast to the real time case the measure now involves N integrations

$$\int \mathcal{D}x = \lim_{a\rightarrow 0}\sqrt{\frac{m}{2\pi\hbar a}}^N \int dx_1 \int dx_2 ... \int dx_N. \tag{A.3.5}$$

The extra integration over $x_N = x'$ is due to the trace in eq.(A.3.1). Note that there is no extra integration over $x_0 = x$ because the trace implies periodic boundary conditions in the Euclidean time direction, i.e. $x_0 = x_N$.

The Euclidean path integral allows us to evaluate thermal expectation values. For example, let us consider an operator $\mathcal{O}(x)$ that is diagonal in the position state basis. We can insert this operator in the path integral and thus compute its expectation value

$$\langle \mathcal{O}(x) \rangle = \frac{1}{Z} \mathrm{Tr}[\mathcal{O}(x) \exp(-\beta H)] = \frac{1}{Z} \int \mathcal{D}x \ \mathcal{O}(x(0)) \exp(-\frac{1}{\hbar} S_E[x]). \quad \text{(A.3.6)}$$

Since the theory is translation invariant in Euclidean time one can place the operator anywhere in time, e.g. at $t = 0$ as done here. When we perform the low temperature limit, $\beta \rightarrow \infty$, the thermal fluctuations are switched off and only the quantum ground state $|0\rangle$ (the vacuum) contributes to the partition function, i.e. $Z \sim \exp(-\beta E_0)$. In this limit the path integral is formulated in an infinite Euclidean time interval, and describes the vacuum expectation value

$$\langle \mathcal{O}(x) \rangle = \langle 0|\mathcal{O}(x)|0\rangle = \lim_{\beta \rightarrow \infty} \frac{1}{Z} \int \mathcal{D}x \ \mathcal{O}(x(0)) \exp(-\frac{1}{\hbar} S_E[x]). \quad \text{(A.3.7)}$$

It is also interesting to consider 2-point functions of operators at different instances in Euclidean time

$$\begin{aligned} \langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle &= \frac{1}{Z} \mathrm{Tr}[\mathcal{O}(x) \exp(-Ht)\mathcal{O}(x) \exp(Ht) \exp(-\beta H)] \\ &= \frac{1}{Z} \int \mathcal{D}x \ \mathcal{O}(x(0))\mathcal{O}(x(t)) \exp(-\frac{1}{\hbar} S_E[x]). \quad \text{(A.3.8)} \end{aligned}$$

Again, we consider the limit $\beta \rightarrow \infty$, but we also separate the operators in time, i.e. we also let $t \rightarrow \infty$. Then the leading contribution is $|\langle 0|\mathcal{O}(x)|0\rangle|^2$. Subtracting this, and thus forming the connected 2-point function, one obtains

$$\lim_{\beta, t \rightarrow \infty} \langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle - |\langle \mathcal{O}(x) \rangle|^2 = |\langle 0|\mathcal{O}(x)|1\rangle|^2 \exp(-(E_1 - E_0)t). \quad \text{(A.3.9)}$$

Here $|1\rangle$ is the first excited state of the quantum system with an energy E_1. The connected 2-point function decays exponentially at large Euclidean time separations. The decay is governed by the energy gap $E_1 - E_0$. In a quantum field theory E_1 corresponds to the energy of the lightest particle. Its mass is determined by the energy gap $E_1 - E_0$ above the vacuum. Hence, in Euclidean field theory particle masses are determined from the exponential decay of connected 2-point correlation functions.

A.4 Spin Models in Classical Statistical Mechanics

So far we have considered quantum systems both at zero and at finite temperature. We have represented their partition functions as Euclidean path integrals over configurations on a time lattice of length β. We will now make a completely new start and study classical discrete systems at finite temperature. We will see that their mathematical description is very similar to the path integral formulation of quantum systems. Still, the physical interpretation of the formalism is drastically different in the two cases. In the next section we will set up a dictionary that allows us to translate quantum physics language into the language of classical statistical mechanics.

For simplicity, let us concentrate on simple classical spin models. Here the word spin does not mean that we deal with quantized angular momenta. All we do is work with classical variables that can point in specific directions. The simplest spin model is the Ising model with classical spin variables $s_x = \pm 1$. (Again, these do not represent the quantum states up and down of a quantum mechanical angular momentum 1/2.) More complicated spin models with an $O(N)$ spin rotational symmetry are the XY model ($N = 2$) and the Heisenberg model ($N = 3$). The spins in the XY model are 2-component unit-vectors, while the spins in the Heisenberg model have three components. In all these models the spins live on the sites of a d-dimensional spatial lattice. The lattice is meant to be a crystal lattice (so typically $d = 3$) and the lattice spacing has a physical meaning. This is in contrast to the Euclidean time lattice that we have introduced to make the path integral mathematically well-defined, and that we finally send to zero in order to reach the Euclidean time continuum limit. The Ising model is characterized by its classical Hamilton function (not a quantum Hamilton operator) which simply specifies the energy of any configuration of spins. The Ising Hamilton function is a sum of nearest neighbor contributions

$$\mathcal{H}[s] = J \sum_{\langle xy \rangle} s_x s_y - \mu B \sum_x s_x, \tag{A.4.1}$$

with a ferromagnetic coupling constant $J < 0$ that favors parallel spins, plus a coupling to an external magnetic field B. The classical partition function of this system is given by

$$Z = \int \mathcal{D}s \ \exp(-\mathcal{H}[s]/T) = \prod_x \sum_{s_x = \pm 1} \exp(-\mathcal{H}[s]/T). \tag{A.4.2}$$

The sum over all spin configurations corresponds to an independent summation over all possible orientations of individual spins. Thermal averages are computed by inserting appropriate operators. For example, the magnetization is given by

$$\langle s_x \rangle = \frac{1}{Z} \prod_x \sum_{s_x = \pm 1} s_x \exp(-\mathcal{H}[s]/T). \tag{A.4.3}$$

Similarly, the spin correlation function is defined by

$$\langle s_x s_y \rangle = \frac{1}{Z} \prod_x \sum_{s_x = \pm 1} s_x s_y \exp(-\mathcal{H}[s]/T). \quad \text{(A.4.4)}$$

At large distances the connected spin correlation function typically decays exponentially

$$\langle s_x s_y \rangle - \langle s \rangle^2 \sim \exp(-|x-y|/\xi), \quad \text{(A.4.5)}$$

where ξ is the so-called correlation length. At general temperatures the correlation length is typically just a few lattice spacings. When one models real materials, the Ising model would generally be a great oversimplification, because real magnets, for example, not only have nearest neighbor couplings. Still, the details of the Hamilton function at the scale of the lattice spacing are not always important. There is a critical temperature T_c at which ξ diverges and universal behavior arises. At this temperature a second order phase transition occurs. Then the details of the model at the scale of the lattice spacing are irrelevant for the long range physics that takes place at the scale of ξ. In fact, at their critical temperatures some real materials behave just like the simple Ising model. This is why the Ising model is so interesting. It is just a very simple member of a large universality class of different models, which all share the same critical behavior. This does not mean that they have the same values of their critical temperatures. However, their magnetization goes to zero at the critical temperature with the same power of $T_c - T$, i.e. their critical exponents are identical.

A.5 Quantum Mechanics versus Statistical Mechanics

We notice a close analogy between the Euclidean path integral for a quantum mechanical system and a classical statistical mechanics system like the Ising model. The path integral for the quantum system is defined on a 1-dimensional Euclidean time lattice, just like an Ising model can be defined on a d-dimensional spatial lattice. In the path integral we integrate over all paths, i.e. over all configurations $x(t)$, while in the Ising model we sum over all spin configurations s_x. Paths are weighted by their Euclidean action $S_E[x]$ while spin configurations are weighted with their Boltzmann factors depending on the classical Hamilton function $\mathcal{H}[s]$. The prefactor of the action is $1/\hbar$, and the prefactor of the Hamilton function is $1/T$. Indeed $\hbar$ determines the strength of quantum fluctuations, while the temperature T determines the strength of thermal fluctuations. The kinetic energy $\frac{1}{2}((x_{i+1} - x_i)/a)^2$ in the path integral is analogous to the nearest neighbor spin coupling $s_x s_{x+1}$, and the potential term $V(x_i)$ is analogous to the coupling $\mu B s_x$

Quantum mechanics	Classical statistical mechanics
Euclidean time lattice	d-dimensional spatial lattice
elementary time step a	crystal lattice spacing
particle position x	classical spin variable s
particle path $x(t)$	spin configuration s_x
path integral $\int \mathcal{D}x$	sum over configurations $\prod_x \sum_{s_x}$
Euclidean action $S_E[x]$	classical Hamilton function $\mathcal{H}[s]$
Planck's constant $\hbar$	temperature T
quantum fluctuations	thermal fluctuations
kinetic energy $\frac{1}{2}(\frac{x_{i+1}-x_i}{a})^2$	neighbor coupling $s_x s_{x+1}$
potential energy $V(x_i)$	external field energy $\mu B s_x$
weight of a path $\exp(-\frac{1}{\hbar}S_E[x])$	Boltzmann factor $\exp(-\mathcal{H}[s]/T)$
vacuum expectation value $\langle \mathcal{O}(x) \rangle$	magnetization $\langle s_x \rangle$
2-point function $\langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle$	correlation function $\langle s_x s_y \rangle$
energy gap $E_1 - E_0$	inverse correlation length $1/\xi$
continuum limit $a \to 0$	critical behavior $\xi \to \infty$

Table A.2: *The dictionary that translates quantum mechanics into the language of classical statistical mechanics.*

to an external magnetic field. The magnetization $\langle s_x \rangle$ corresponds to the vacuum expectation value of an operator $\langle \mathcal{O}(x) \rangle$ and the spin-spin correlation function $\langle s_x s_y \rangle$ corresponds to the 2-point correlation function $\langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle$. The inverse correlation length $1/\xi$ is analogous to the energy gap $E_1 - E_0$ (and hence to a particle mass in a Euclidean quantum field theory). Finally, the Euclidean time continuum limit $a \to 0$ corresponds to a second order phase transition where $\xi \to \infty$. The lattice spacing in the path integral is an artifact of our mathematical description which we send to zero while the physics remains constant. In classical statistical mechanics, on the other hand, the lattice spacing is physical and hence fixed, while the correlation length ξ goes to infinity at a second order phase transition. All this is summarized in the dictionary of table A.2.

A.6 Lattice Field Theory

So far we have restricted ourselves to quantum mechanical problems and to classical statistical mechanics. The former were defined by a path integral on a 1-d Euclidean time lattice, while the latter involved spin models on a d-dimensional spatial lattice. When we quantize field theories on the lattice, we formulate the theory on a d-dimensional space-time lattice, i.e. usually the lattice is 4-dimensional. Just as we integrate over all configurations (all paths) $x(t)$ of a quantum particle, we now integrate over all configurations $\Phi(x)$ of a quantum field defined at any Euclidean space-time point $x = (\vec{x}, x_4)$. Again the weight factor in the path integral is given by the action. Let us illustrate this for a free neutral scalar field $\Phi(x) \in R$. Its Euclidean action is given by

$$S_E[\Phi] = \int d^4x \, [\frac{1}{2}\partial_\mu \Phi \partial_\mu \Phi + \frac{m^2}{2}\Phi^2]. \tag{A.6.1}$$

Interactions can be included, for example, by adding a $\frac{\lambda}{4}\Phi^4$ term to the action. The Feynman path integral for this system is formally written as

$$Z = \int \mathcal{D}\Phi \exp(-S_E[\Phi]). \tag{A.6.2}$$

(Note that we have put $\hbar = c = 1$.) The integral is over all field configurations, which is a divergent expression if no regularization is imposed. One can make the expression mathematically well-defined by using dimensional regularization of Feynman diagrams. This approach is, however, limited to perturbation theory. The lattice allows us to formulate field theory beyond perturbation theory, which is very essential for strongly interacting theories like QCD, but also for the standard model in general. For example, due to the heavy mass of the top quark, the Yukawa coupling between the Higgs and top quark field is rather strong. The above free scalar field theory, of course, does not really require a nonperturbative treatment. We use it only to illustrate the lattice quantization method in a simple setting. On the lattice the continuum field $\Phi(x)$ is replaced by a lattice field Φ_x, which is restricted to the points x of a d-dimensional space-time lattice. From now on we will work in lattice units, i.e. we put $a = 1$. The above continuum action can be approximated by discretizing the continuum derivatives such that

$$S_E[\Phi] = \sum_{x,\mu} \frac{1}{2}(\Phi_{x+\hat{\mu}} - \Phi_x)^2 + \sum_x \frac{m^2}{2}\Phi_x^2. \tag{A.6.3}$$

Here $\hat{\mu}$ is the unit vector in the μ-direction. The integral over all field configurations now becomes a multiple integral over all values of the field at all lattice points

$$Z = \prod_x \int_{-\infty}^{\infty} d\Phi_x \exp(-S_E[\Phi]). \tag{A.6.4}$$

For a free field theory the partition function is just a Gaussian integral. In fact, one can write the lattice action as

$$S_E[\Phi] = \frac{1}{2}\sum_{x,y} \Phi_x \mathcal{M}_{xy} \Phi_y, \tag{A.6.5}$$

where the matrix $\mathcal{M}$ describes the couplings between lattice points. Diagonalizing this matrix by an orthogonal transformation $\mathcal{O}$ one has

$$\mathcal{M} = \mathcal{O}^T \mathcal{D} \mathcal{O}. \tag{A.6.6}$$

Introducing

$$\Phi'_x = \mathcal{O}_{xy} \Phi_y \tag{A.6.7}$$

one obtains

$$Z = \prod_x \int d\Phi'_x \exp(-\frac{1}{2} \sum_x \Phi'_x \mathcal{D}_{xx} \Phi'_x) = (2\pi)^{N/2} \mathrm{det}\mathcal{D}^{-1/2}, \tag{A.6.8}$$

where N is the number of lattice points.

To extract the energy values of the corresponding quantum Hamilton operator we need to study the 2-point function of the lattice field

$$\langle \Phi_x \Phi_y \rangle = \frac{1}{Z} \int \mathcal{D}\Phi \ \Phi_x \Phi_y \exp(-S_E[\Phi]). \tag{A.6.9}$$

This is most conveniently done by introducing a source field in the partition function, such that

$$Z[J] = \int \mathcal{D}\Phi \exp(-S_E[\Phi] + \sum_x J_x \Phi_x). \tag{A.6.10}$$

Then the connected 2-point function is given by

$$\langle \Phi_x \Phi_y \rangle - \langle \Phi \rangle^2 = \frac{\partial^2 \log Z[J]}{\partial J_x \partial J_y} |_{J=0}. \tag{A.6.11}$$

The Boltzmann factor characterizing the problem with the external sources is given by the exponent

$$\frac{1}{2} \Phi \mathcal{M} \Phi - J\Phi = \frac{1}{2} \Phi' \mathcal{M} \Phi' - \frac{1}{2} J \mathcal{M}^{-1} J. \tag{A.6.12}$$

Here we have introduced

$$\Phi' = \Phi - \mathcal{M}^{-1} J. \tag{A.6.13}$$

Integrating over Φ' in the path integral we obtain

$$Z[J] = (2\pi)^{N/2} \mathrm{det}\mathcal{D}^{-1/2} \exp(\frac{1}{2} J \mathcal{M}^{-1} J), \tag{A.6.14}$$

and hence

$$\langle \Phi_x \Phi_y \rangle = \frac{1}{2} \mathcal{M}^{-1}_{xy}. \tag{A.6.15}$$

It is instructive to invert the matrix $\mathcal{M}$ by going to Fourier space, i.e. by writing

$$\Phi_x = \frac{1}{(2\pi)^d} \int_B d^d p \ \Phi(p) \exp(ipx). \tag{A.6.16}$$

The momentum space of the lattice is given by the Brillouin zone $B =]-\pi, \pi]^d$. For the 2-point function in momentum space one then finds

$$\langle \Phi(-p)\Phi(p) \rangle = [\sum_\mu (2\sin(p_\mu/2))^2 + m^2]^{-1}. \tag{A.6.17}$$

This is the lattice version of the continuum propagator

$$\langle \Phi(-p)\Phi((p)) \rangle = (p^2 + m^2)^{-1}. \tag{A.6.18}$$

From the lattice propagator we can deduce the energy spectrum of the lattice theory. For this purpose we construct a lattice field with definite spatial momentum $\vec{p}$ located in a specific time slice

$$\Phi(\vec{p})_t = \sum_x \Phi_{\vec{x},t} \exp(-i\vec{p}\cdot\vec{x}), \tag{A.6.19}$$

and we consider its 2-point function

$$\langle \Phi(-\vec{p})_0 \Phi(\vec{p})_t \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} dp_d \langle \Phi(-p)\Phi(p) \rangle \exp(ip_d t). \tag{A.6.20}$$

Inserting the lattice propagator of eq.(A.6.17) one can perform the integral. One encounters a pole in the propagator when $p_d = iE$ with

$$(2\sinh(E/2))^2 = \sum_i (2\sin(p_i/2))^2 + m^2. \tag{A.6.21}$$

The 2-point function then takes the form

$$\langle \Phi(-\vec{p})_0 \Phi(\vec{p})_t \rangle = C \exp(-Et), \tag{A.6.22}$$

i.e. it decays exponentially with slope E. This allows us to identify E as the energy of the lattice scalar particle with spatial momentum $\vec{p}$. In general, E differs from the correct continuum dispersion relation

$$E^2 = \vec{p}^2 + m^2. \tag{A.6.23}$$

Only in the continuum limit, i.e. when E, $\vec{p}$ and m are small in lattice units, the lattice dispersion relation agrees with the one of the continuum theory.

We have defined the path integral by using the action of the classical theory. Theories with fermions have no immediate classical limit, and the definition of the path integral needs special care. The first step is to define a so-called Grassmann algebra, which works with anticommuting classical variables η_i with $i \in 1, 2, ..., N$. A Grassmann algebra is characterized by the anticommutation relations

$$\{\eta_i, \eta_j\} = \eta_i \eta_j + \eta_j \eta_i = 0. \tag{A.6.24}$$

An element of the Grassmann algebra is a polynomial in the generators

$$f(\eta) = f + \sum_i f_i \eta_i + \sum_{ij} f_{ij} \eta_i \eta_j + \sum_{ijk} f_{ijk} \eta_i \eta_j \eta_k + ... \tag{A.6.25}$$

The $f_{ij...l}$ are ordinary complex numbers antisymmetric in $i, j, ..., l$. One also defines formal differentiation and integration procedures. The differentiation rules are

$$\frac{\partial}{\partial \eta_i} \eta_i = 1, \ \frac{\partial}{\partial \eta_i} \eta_i \eta_j = \eta_j, \ \frac{\partial}{\partial \eta_i} \eta_j \eta_i = -\eta_j, \tag{A.6.26}$$

and integration is defined by

$$\int d\eta_i = 0, \ \int d\eta_i \ \eta_i = 1, \ \int d\eta_i d\eta_j \ \eta_i \eta_j = -1. \tag{A.6.27}$$

These integrals are formal expressions. It has no meaning to ask over which range of η_i values we actually integrate.

The Grassmann algebra we use to define fermion fields is generated by Grassmann numbers Ψ_x and $\bar{\Psi}_x$, which are completely independent. The index x runs over all space-time points as well as over all spin, flavor, color or other indices. Let us consider the simplest (completely unrealistic) case of just two degrees of freedom Ψ and $\bar{\Psi}$, and let us perform the Gaussian integral

$$\int d\bar{\Psi} d\Psi \ \exp(-m\bar{\Psi}\Psi) = \int d\bar{\Psi} d\Psi \ (1 - m\bar{\Psi}\Psi) = m. \tag{A.6.28}$$

Note that the expansion of the exponential terminates because $\Psi^2 = \bar{\Psi}^2 = 0$. When we enlarge the Grassmann algebra to an arbitrary number of elements the above formula generalizes to

$$\prod_x \int d\bar{\Psi}_x d\Psi_x \ \exp(-\bar{\Psi}_x \mathcal{M}_{xy} \Psi_y) = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \exp(-\bar{\Psi}\mathcal{M}\Psi) = \det\mathcal{M}. \tag{A.6.29}$$

In the two variable case we have

$$\int d\bar{\Psi} d\Psi \ \bar{\Psi}\Psi \exp(-m\bar{\Psi}\Psi) = 1, \tag{A.6.30}$$

which generalizes to

$$\int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \bar{\Psi}_x \Psi_y \exp(-\bar{\Psi}\mathcal{M}\Psi_y) = \mathcal{M}_{ij}^{-1} \mathrm{det}\mathcal{M}. \qquad \text{(A.6.31)}$$

Lattice fermions have several technical problems that have prevented the non-perturbative formulation of the standard model for many years. For example, chiral fermions — like neutrinos — suffer from the lattice fermion doubling problem. Every left-handed neutrino necessarily comes with a right-handed partner. Until recently, it was not known how to couple only the left-handed particles to an electroweak lattice gauge field. Thanks to a recent breakthrough in lattice gauge theory, the standard model is now consistently defined beyond perturbation theory. Even the perturbative definition of the standard model has been incomplete beyond one loop, due to ambiguities in treating γ_5 in dimensional regularization. All these ambiguities are now eliminated, thanks to the new lattice result. In the rest of this course, we will not need the lattice regularization much further. Still, it is good to know that the standard model now stands on a firm mathematical basis and that the path integral expressions we write down for it are completely well-defined even beyond perturbation theory.

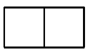
Appendix B

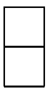
Group Theory of S_N and $SU(n)$

We will soon complete the formulation of the standard model by adding the gluons as the last remaining field, thus introducing the strong interactions which are governed by an $SU(3)_c$ gauge symmetry. To first familiarize ourselves a bit with the relevant group theory, we will now make a short mathematical detour. Once we add the strong interactions to the standard model, the quarks will get confined inside hadrons. In the so-called constituent quark model (which is at best semi-quantitative) baryons are made of three quarks, while mesons consist of a quark and an anti-quark. In the group theoretical construction of baryon states the permutation group S_3 of three quarks plays an important role. In general, the permutation group S_N of N objects is very useful when one wants to couple arbitrary $SU(n)$ representations together.

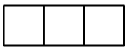
B.1 The Permutation Group S_N

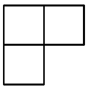
Let us consider the permutation symmetry of N objects — for example the fundamental representations of $SU(n)$. Their permutations form the group S_N. The permutation group has $N!$ elements — all permutations of N objects. The group S_2 has two elements: the identity and the pair permutation. The representations of S_2 are represented by Young tableaux

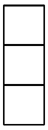
 1-dimensional symmetric representation,

 1-dimensional antisymmetric representation. (B.1.1)

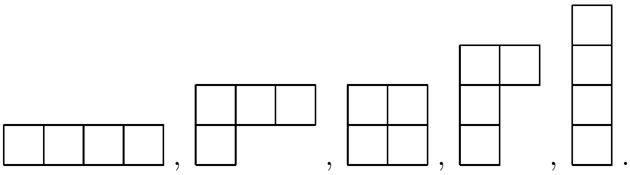
To describe the permutation properties of three objects we need the group S_3. It has $3! = 6$ elements: the identity, 3 pair permutations and 2 cyclic permutations. The group S_3 has three irreducible representations

 1-dimensional symmetric representation,

 2-dimensional representation of mixed symmetry,

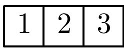
 1-dimensional antisymmetric representation. (B.1.2)

The representations of the group S_N are given by the Young tableaux with N boxes. The boxes are arranged in left-bound rows, such that no row is longer than the one above it. For example, for the representations of S_4 one finds

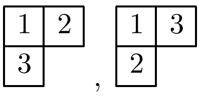


(B.1.3)

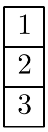
The dimension of a representation is determined as follows. The boxes of the corresponding Young tableau are enumerated from 1 to N such that the numbers grow as one reads each row from left to right, and each column from top to bottom. The number of possible enumerations determines the dimension of the representation. For example, for S_3 one obtains



1-dimensional,



2-dimensional,



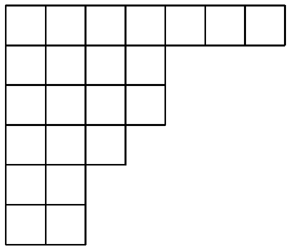
1-dimensional. (B.1.4)

The squares of the dimensions of all representations add up to the order of the group, i.e.

$$\sum_\Gamma d_\Gamma^2 = N! \; . \tag{B.1.5}$$

In particular, for S_2 we have $1^2 + 1^2 = 2 = 2!$ and for S_3 one obtains $1^2 + 2^2 + 1^2 = 6 = 3!$.

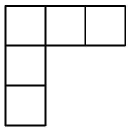
A general Young tableau can be characterized by the number of boxes m_i in its i-th row. For example the Young tableau



has $m_1 = 7$, $m_2 = 4$, $m_3 = 4$, $m_4 = 3$, $m_5 = 2$ and $m_6 = 2$. The dimension of the corresponding representation is given by

$$d_{m_1,m_2,\ldots,m_n} = N! \frac{\prod_{i<k}(l_i - l_k)}{l_1! l_2! \ldots l_n!}, \ l_i = m_i + n - i. \quad \text{(B.1.6)}$$

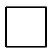
Applying this formula to the following Young tableau from S_5



with $m_1 = 3$, $m_2 = 1$, $m_3 = 1$ and $n = 3$ yields $l_1 = 3+3-1 = 5$, $l_2 = 1+3-2 = 2$, $l_3 = 1 + 3 - 3 = 1$ and hence

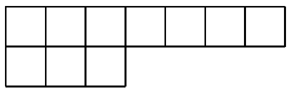
$$d_{3,1,1} = 5! \frac{(l_1 - l_2)(l_1 - l_3)(l_2 - l_3)}{l_1! l_2! l_3!} = 5! \frac{3 \cdot 4 \cdot 1}{5!2!1!} = 6. \quad \text{(B.1.7)}$$

The permuted objects can be the fundamental representations of $SU(n)$. For $SU(2)$ we identify



$$= \{2\}. \quad \text{(B.1.8)}$$

To each Young tableau with no more than two rows one can associate an $SU(2)$ representation. Such a Young tableau is characterized by m_1 and m_2, e.g.



has $m_1 = 7$ and $m_2 = 3$. The corresponding $SU(2)$ representation has

$$S = \frac{1}{2}(m_1 - m_2), \quad \text{(B.1.9)}$$

which is also denoted by $\{m_1 - m_2 + 1\}$. The above Young tableau hence represents $S = 2$ — a spin quintet $\{5\}$. Young tableaux with more than two rows have no realization in $SU(2)$ since among just two distinguishable objects no more than two can be combined anti-symmetrically.

B.2 The Group $SU(n)$

The unitary $n \times n$ matrices with determinant 1 form a group under matrix multiplication — the special unitary group $SU(n)$. This follows immediately from

$$\begin{aligned} UU^\dagger &= U^\dagger U = 1,\ \det U = 1. \\ \det UV &= \det U \det V = 1. \end{aligned} \tag{B.2.1}$$

Associativity ($(UV)W = U(VW)$) holds for all matrices, a unit element 1 exists (the unit matrix), the inverse is $U^{-1} = U^\dagger$, and finally the group property

$$(UV)^\dagger UV = V^\dagger U^\dagger UV = 1,\ UV(UV)^\dagger = UVV^\dagger U^\dagger = 1 \tag{B.2.2}$$

also holds. The group $SU(n)$ is non-Abelian because in general $UV \neq VU$. Each element $U \in SU(n)$ can be represented as

$$U = \exp(iH), \tag{B.2.3}$$

where H is Hermitean and traceless. The matrices H form the $su(n)$ algebra. One has $n^2 - 1$ free parameters, and hence $n^2 - 1$ generators η_i, and one can write

$$H = \alpha_i \eta_i,\ \alpha_i \in R. \tag{B.2.4}$$

The structure of the algebra results from the commutation relations

$$[\eta_i, \eta_j] = 2ic_{ijk}\eta_k, \tag{B.2.5}$$

where c_{ijk} are the so-called structure constants.

The simplest nontrivial representation of $SU(n)$ is the fundamental representation. It is n-dimensional and can be identified with the Young tableau $\square$. Every irreducible representation of $SU(n)$ can be obtained from coupling N fundamental representations. In this way each $SU(n)$ representation is associated with a Young tableau with N boxes, which characterizes the permutation symmetry of the fundamental representations in the coupling. Since the fundamental representation is n-dimensional, there are n different fundamental properties (e.g. u and d in $SU(2)_L$ and $c \in \{1, 2, 3\}$ in $SU(3)_c$). Hence, we can maximally antisymmetrize n objects, and the Young tableaux for $SU(n)$ representations are therefore restricted to no more than n rows.

The dimension of an $SU(n)$ representation can be obtained from the corresponding Young tableau by filling it with factors as follows

n	$n+1$	$n+2$	$n+3$	$n+4$	$n+5$	$n+6$
$n-1$	n	$n+1$	$n+2$			
$n-2$	$n-1$	n	$n+1$			
$n-3$	$n-2$	$n-1$				
$n-4$	$n-3$					
$n-5$	$n-4$					

.

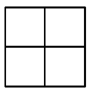
The dimension of the $SU(n)$ representation is given as the product of all factors divided by $N!$ and multiplied with the S_N dimension $d_{m1,m_2,...,m_n}$ of the Young tableau

$$\begin{aligned} D^n_{m1,m_2,...,m_n} &= \frac{(n+m_1-1)!}{(n-1)!}\frac{(n+m_2-2)!}{(n-2)!}...\frac{m_n!}{0!}\frac{1}{N!}N!\frac{\prod_{i<k}(l_i-l_k)}{l_1!l_2!...l_n!} \\ &= \frac{\prod_{i<k}(m_i-m_k-i+k)}{(n-1)!(n-2)!...0!}. \end{aligned} \tag{B.2.6}$$

We see that the dimension of a representation depends only on the differences $q_i = m_i - m_{i+1}$. In particular, for $SU(2)$ we find

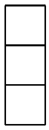
$$D^2_{m_1,m_2} = \frac{m_1 - m_2 - 1 + 2}{1!0!} = m_1 - m_2 + 1 = q_1 + 1 \tag{B.2.7}$$

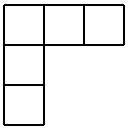
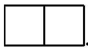
in agreement with our previous result. For a rectangular Young tableau with n rows, e.g. in $SU(2)$ for


,

all $q_i = 0$, and we obtain

$$D^n_{m,m,...,m} = \frac{\prod_{i<k}(m_i-m_k-i+k)}{(n-1)!(n-2)!...0!} = \frac{(n-1)!(n-2)!...0!}{(n-1)!(n-2)!...0!} = 1, \tag{B.2.8}$$

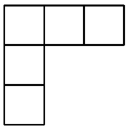
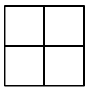
and therefore a singlet. This shows that in $SU(3)$  corresponds to a singlet. It also explains why the dimension of an $SU(n)$ representation depends only on the differences q_i. Without changing the dimension we can couple a representation with a singlet, and hence we can always add a rectangular Young tableau with n rows to any $SU(n)$ representation. For example in $SU(3)$

 $\cong$ . (B.2.9)

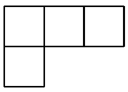
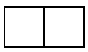
We want to associate an anti-representation with each representation by replacing m_i and q_i with

$$\bar{m}_i = m_1 - m_{n-i+1}, \ \bar{q}_i = \bar{m}_i - \bar{m}_{i+1} = m_{n-i} - m_{n-i+1} = q_{n-i}. \tag{B.2.10}$$

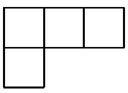
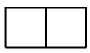
Geometrically the Young tableau of a representation and its anti-representation (after rotation) fit together to form a rectangular Young tableau with n rows. For example, in $SU(3)$

 and 

are anti-representations of one another. In $SU(2)$ each representation is its own anti-representation. For example

 and 

are anti-representations of one another, but

 $\cong$ .

This is not the case for higher n. The dimension of a representation and its anti-representation are identical

$$D^n_{\bar{m}_1,\bar{m}_2,\ldots,\bar{m}_n} = D^n_{m_1,m_2,\ldots,m_n}. \tag{B.2.11}$$

For general n the so-called adjoint representation is given by $q_1 = q_{n-1} = 1$, $q_i = 0$ otherwise, and it is identical with its own anti-representation. The dimension of the adjoint representation is

$$D^n_{2,1,1,\ldots,1,0} = n^2 - 1. \tag{B.2.12}$$

Next we want to discuss a method to couple $SU(n)$ representations by operating on their Young tableaux. Two Young tableaux with N and M boxes are coupled by forming an external product. In this way we generate Young tableaux with $N + M$ boxes that can then be translated back into $SU(n)$ representations. The external product is built as follows. The boxes of the first row of the second Young tableau are labeled with 'a', the boxes of the second row with 'b', etc. Then the boxes labeled with 'a' are added to the first Young tableau in all possible ways that lead to new allowed Young tableaux. Then the 'b' boxes are added to the resulting Young tableaux in the same way. Now each of the resulting tableaux is read row-wise from top-right to bottom-left. Whenever a 'b' or 'c' appears before the first 'a', or a 'c' occurs before the first 'b' etc., the corresponding Young tableau is deleted. The remaining tableaux form the reduction of the external product.

We now want to couple N fundamental representations of $SU(n)$. In Young tableau language this reads

$$\{n\} \otimes \{n\} \otimes \ldots \otimes \{n\} = \square \otimes \square \otimes \ldots \otimes \square. \tag{B.2.13}$$

In this way we generate all irreducible representations of S_N, i.e. all Young tableaux with N boxes. Each Young tableau is associated with an $SU(n)$ multiplet. It occurs in the product as often as the dimension of the corresponding S_N representation indicates, i.e. $d_{m_1,m_2,\ldots,m_n}$ times. Hence we can write

$$\{n\} \otimes \{n\} \otimes \ldots \otimes \{n\} = \sum_\Gamma d_{m_1,m_2,\ldots,m_n} \{D^n_{m_1,m_2,\ldots,m_n}\}. \tag{B.2.14}$$

The sum goes over all Young tableaux with N boxes. For example

$$\square \otimes \square \otimes \square = \begin{array}{|c|c|c|}\hline & & \\ \hline\end{array} \oplus 2 \begin{array}{|c|c|}\hline & \\ \hline \\ \cline{1-1}\end{array} \oplus \begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline\end{array}. \tag{B.2.15}$$

Translated into $SU(n)$ language this reads

$$\begin{aligned}\{n\} \otimes \{n\} \otimes \{n\} &= \{\frac{n(n+1)(n+2)}{6}\} \oplus 2\{\frac{(n-1)n(n+1)}{3}\} \\ &\oplus \{\frac{(n-2)(n-1)n}{6}\}.\end{aligned} \tag{B.2.16}$$

The dimensions test

$$\frac{n(n+1)(n+2)}{6} + 2\frac{(n-1)n(n+1)}{3} + \frac{(n-2)(n-1)n}{6} = n^3 \tag{B.2.17}$$

confirms this result. In $SU(2)$ this corresponds to

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus 2\{2\} \oplus \{0\}, \tag{B.2.18}$$

and in $SU(3)$

$$\{3\} \otimes \{3\} \otimes \{3\} = \{10\} \oplus 2\{8\} \oplus \{1\}. \tag{B.2.19}$$